SECOND EDITION

MICROWAVE FILTERS FOR COMMUNICATION SYSTEMS

FUNDAMENTALS, DESIGN, AND APPLICATIONS

RICHARD J. CAMERON
CHANDRA M. KUDSIA
RAAFAT R. MANSOUR

WILEY

Microwave Filters for Communication Systems

Microwave Filters for Communication Systems

Fundamentals, Design, and Applications

Second Edition

Richard J. Cameron
Chandra M. Kudsia
Raafat R. Mansour

WILEY

This second edition first published 2018
© 2018 John Wiley & Sons, Inc.

Edition History
John Wiley & Sons, Inc. (1e, 2007)

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at http://www.wiley.com/go/permissions.

The right of Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour to be identified as the author(s) of this work has been asserted in accordance with law.

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty
The publisher and the authors make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties; including without limitation any implied warranties of fitness for a particular purpose. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for every situation. In view of on-going research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. The fact that an organization or website is referred to in this work as a citation and/or potential source of further information does not mean that the author or the publisher endorses the information the organization or website may provide or recommendations it may make. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. No warranty may be created or extended by any promotional statements for this work. Neither the publisher nor the author shall be liable for any damages arising here from.

Library of Congress Cataloging-in-Publication Data

Names: Cameron, Richard J., author. | Kudsia, Chandra M., author. | Mansour, Raafat R., author.
Title: Microwave filters for communication systems : fundamentals, design and applications / by Richard J. Cameron, Chandra M. Kudsia, Raafat R. Mansour.
Description: Second edition. | Hoboken, NJ, USA : Wiley, 2018. | Includes bibliographical references and index. |
Identifiers: LCCN 2017036462 (print) | LCCN 2017037258 (ebook) | ISBN 9781119292395 (pdf) | ISBN 9781119292388 (epub) | ISBN 9781118274347 (hardback)
Subjects: LCSH: Microwave filters. | Telecommunication systems–Equipment and supplies. | BISAC: TECHNOLOGY & ENGINEERING / Microwaves.
Classification: LCC TK7872.F5 (ebook) | LCC TK7872.F5 C35 2018 (print) | DDC 621.3841/2–dc23
LC record available at https://lccn.loc.gov/2017036462

Cover design by Wiley
Cover image: © John Lund/Gettyimages

Set in 10/12pt WarnockPro by SPi Global, Chennai, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

The authors wish to acknowledge the many engineers, technologists, and manufacturing staff at COM DEV who contributed in advancing the state of the art and in maintaining sustained leadership position for the design and implementation of microwave filters and multiplexing networks for communication satellite systems from 1970s to 2015.[1]

1 COM DEV was sold to Honeywell in 2016.

Contents

Preface *xxiii*

1 Radio Frequency (RF) Filter Networks for Wireless Communications—The System Perspective *1*

Part I Introduction to a Communication System, Radio Spectrum, and Information *1*
1.1 Model of a Communication System *1*
1.1.1 Building Blocks of a Communication System *3*
1.1.1.1 Information Source *3*
1.1.1.2 Transmitter *3*
1.1.1.3 Communication Channel *3*
1.1.1.4 Receiver *5*
1.1.1.5 Information Destination *5*
1.2 Radio Spectrum and its Utilization *6*
1.2.1 Radio Propagation at Microwave Frequencies *6*
1.2.2 Radio Spectrum as a Natural Resource *8*
1.3 Concept of Information *8*
1.4 Communication Channel and Link Budgets *10*
1.4.1 Signal Power in a Communication Link *10*
1.4.1.1 RF Power *11*
1.4.2 Transmit and Receive Antennas *11*
1.4.2.1 Antenna Gain *12*
1.4.2.2 Effective Aperture and Antenna Gain *12*
1.4.2.3 Power Flux Density *13*
1.4.2.4 Power Radiated and Received by Antennas in a Communication System *13*

Part II Noise in a Communication Channel *15*
1.5 Noise in Communication Systems *15*
1.5.1 Adjacent Copolarized Channel Interference *16*
1.5.2 Adjacent Cross-Polarized Channel Interference *16*
1.5.3 Multipath Interference *17*
1.5.4 Thermal Noise *17*
1.5.4.1 Effective Input Noise Temperature *20*
1.5.4.2 Noise Figure *21*
1.5.4.3 Noise of a Lossy Element *22*

1.5.4.4 Attenuators *22*
1.5.5 Noise in Cascaded Networks *23*
1.5.6 Intermodulation (IM) Noise *25*
1.5.7 Distortion Due to Channel Imperfections *26*
1.5.8 RF Link Design *29*
1.5.8.1 Carrier-to-Noise Ratio (*C*/*N*) *29*
1.5.8.2 Multiple RF Links in Tandem *30*
1.6 Modulation–Demodulation Schemes in a Communication System *32*
1.6.1 Amplitude Modulation *32*
1.6.2 Formation of a Baseband Signal *34*
1.6.3 Angle-Modulated Signals *35*
1.6.3.1 Spectra of Analog Angle-Modulated Signals *35*
1.6.4 Comparison of FM and AM Systems *37*
1.7 Digital Transmission *39*
1.7.1 Sampling *40*
1.7.2 Quantization *41*
1.7.3 PCM Systems *41*
1.7.4 Quantization Noise in PCM Systems *41*
1.7.5 Error Rates in Binary Transmission *42*
1.7.6 Digital Modulation and Demodulation Schemes *44*
1.7.6.1 Amplitude Shift Keying (ASK) *44*
1.7.6.2 Frequency Shift Keying (FSK) *44*
1.7.6.3 Phase Shift Keying (PSK) *45*
1.7.7 Advanced Modulation Schemes *45*
1.7.7.1 Bandwidth Efficiency *46*
1.7.7.2 Power Efficiency *48*
1.7.7.3 Bandwidth and Power Efficient Modulation Schemes *48*
1.7.8 Quality of Service and *S/N* Ratio *50*

Part III Impact of System Design on the Requirements of Filter Networks *50*
1.8 Communication Channels in a Satellite System *50*
1.8.1 Receive Section *53*
1.8.2 The Channelizer Section *53*
1.8.3 High-Power Amplifiers (HPAs) *56*
1.8.4 Transmitter Section Architecture *58*
1.8.4.1 Output Multiplexer *58*
1.8.4.2 Harmonic Rejection Filters *60*
1.8.4.3 PIM Requirements for High-Power Output Circuits *60*
1.9 RF Filters in Cellular Systems *62*
1.10 Ultra Wideband (UWB) Wireless Communication *66*
1.11 Impact of System Requirements on RF Filter Specifications *68*
1.11.1 Frequency Plan *68*
1.11.2 Interference Environment *69*
1.11.3 Modulation Schemes *69*
1.11.4 HPA Characteristics *69*
1.11.5 Location of RF Filter in the Communication Link *70*
1.11.6 Operating Environment *70*

1.11.6.1 Space Environment *70*
1.11.6.2 Terrestrial Environment *70*
1.11.7 Limitations of Microwave Filter Technology *70*
1.12 Impact of Satellite and Cellular Communications on Filter Technology *72*
Summary *72*
References *72*
Appendix 1A *74*
Intermodulation Distortion Summary *74*

2 Fundamentals of Circuit Theory Approximation *75*
2.1 Linear Systems *75*
2.1.1 Concept of Linearity *75*
2.2 Classification of Systems *76*
2.2.1 Time-Invariant and Time-Variant Systems *76*
2.2.2 Lumped and Distributed Systems *76*
2.2.3 Instantaneous and Dynamic Systems *77*
2.2.4 Analog and Digital Systems *77*
2.3 Evolution of Electrical Circuits: A Historical Perspective *77*
2.3.1 Circuit Elements *78*
2.4 Network Equation of Linear Systems in the Time Domain *78*
2.5 Network Equation of Linear Systems in the Frequency-Domain Exponential Driving Function *80*
2.5.1 Complex Frequency Variable *80*
2.5.2 Transfer Function *81*
2.5.3 Signal Representation by Continuous Exponentials *82*
2.5.4 Transfer Functions of Electrical Networks *82*
2.6 Steady-State Response of Linear Systems to Sinusoidal Excitations *83*
2.7 Circuit Theory Approximation *84*
Summary *85*
References *86*

3 Characterization of Lossless Lowpass Prototype Filter Functions *87*
3.1 The Ideal Filter *87*
3.1.1 Distortionless Transmission *87*
3.1.2 Maximum Power Transfer in Two-Port Networks *88*
3.2 Characterization of Polynomial Functions for Doubly Terminated Lossless Lowpass Prototype Filter Networks *88*
3.2.1 Reflection and Transmission Coefficients *90*
3.2.2 Normalization of the Transfer Function and Characteristic Polynomials *92*
3.3 Characteristic Polynomials for Idealized Lowpass Prototype Networks *93*
3.4 Lowpass Prototype Characteristics *95*
3.4.1 Amplitude Response *95*
3.4.2 Phase Response *95*
3.4.3 Phase Linearity *96*
3.5 Characteristic Polynomials versus Response Shapes *96*
3.5.1 All-Pole Prototype Filter Functions *97*
3.5.2 Prototype Filter Functions with Finite Transmission Zeros *97*
3.6 Classical Prototype Filters *98*

3.6.1 Maximally Flat Filters *98*
3.6.2 Chebyshev Approximation *99*
3.6.3 Elliptic Function Filters *101*
3.6.4 Odd-Order Elliptic Function Filters *104*
3.6.5 Even-Order Elliptic Function Filters *105*
3.6.5.1 Class A Elliptic Filters *105*
3.6.5.2 Class B Elliptic Filters *105*
3.6.5.3 Class C Elliptic Filters *106*
3.6.6 Filters with Transmission Zeros and a Maximally Flat Passband *107*
3.6.7 Linear Phase Filters *107*
3.6.8 Comparison of Maximally Flat, Chebyshev, and Elliptic Function Filters *107*
3.7 Unified Design Chart (UDC) Relationships *108*
3.8 Lowpass Prototype Circuit Configurations *109*
3.8.1 Scaling of Prototype Networks *110*
3.8.2 Frequency Response of Scaled Networks *111*
3.8.2.1 Amplitude Response *111*
3.8.2.2 Group Delay Response *113*
3.8.2.3 Lowpass Filter Group Delay *113*
3.8.2.4 Bandpass Filter Group Delay *113*
3.9 Effect of Dissipation *113*
3.9.1 Relationship of Dissipation Factor δ and Quality Factor Q_0 *114*
3.10 Asymmetric Response Filters *115*
3.10.1 Positive Functions *116*
Summary *118*
References *119*
Appendix 3A *121*
Unified Design Charts *121*

4 Computer-Aided Synthesis of Characteristic Polynomials *129*
4.1 Objective Function and Constraints for Symmetric Lowpass Prototype Filter Networks *129*
4.2 Analytic Gradients of the Objective Function *131*
4.2.1 Gradient of the Unconstrained Objective Function *132*
4.2.2 Gradient of the Inequality Constraint *133*
4.2.3 Gradient of the Equality Constraint *133*
4.3 Optimization Criteria for Classical Filters *134*
4.3.1 Chebyshev Function Filters *134*
4.3.2 Inverse Chebyshev Filters *134*
4.3.3 Elliptic Function Filters *135*
4.4 Generation of Novel Classes of Filter Functions *136*
4.4.1 Equiripple Passbands and Stopbands *137*
4.4.2 Nonequiripple Stopband with an Equiripple Passband *137*
4.5 Asymmetric Class of Filters *138*
4.5.1 Asymmetric Filters with Chebyshev Passband *140*
4.5.2 Asymmetrical Filters with Arbitrary Response *140*
4.6 Linear Phase Filters *142*
4.7 Critical Frequencies for Selected Filter Functions *143*
Summary *144*

References *144*
Appendix 4A *145*

5 Analysis of Multiport Microwave Networks *147*
5.1 Matrix Representation of Two-Port Networks *147*
5.1.1 Impedance [Z] and Admittance [Y] Matrices *147*
5.1.2 The [$ABCD$] Matrix *149*
5.1.3 The Scattering [S] Matrix *150*
5.1.4 The Transmission Matrix [T] *155*
5.1.5 Analysis of Two-Port Networks *157*
5.2 Cascade of Two Networks *160*
5.3 Multiport Networks *167*
5.4 Analysis of Multiport Networks *169*
Summary *174*
References *175*

6 Synthesis of a General Class of the Chebyshev Filter Function *177*
6.1 Polynomial Forms of the Transfer and Reflection Parameters $S_{21}(S)$ and $S_{11}(S)$ for a Two-port network *177*
6.1.1 Relationship Between ε and ε_R *183*
6.1.2 Relationship Between the Polynomials of the [$ABCD$] Transfer Matrix Parameters and the S-Parameters *184*
6.1.2.1 Short-Circuit Admittance Parameters *186*
6.2 Alternating Pole Method for the Determination of the Denominator Polynomial $E(S)$ *186*
6.3 General Polynomial Synthesis Methods for Chebyshev Filter Functions *189*
6.3.1 Polynomial Synthesis *189*
6.3.2 Recursive Technique *193*
6.3.3 Polynomial Forms for Symmetric and Asymmetric Filtering Functions *196*
6.3.4 Finding the Positions of the In-Band Reflection Maxima and the Out-of-Band Transmission Maxima of the Generalized Chebyshev Prototype *198*
6.3.4.1 In-Band Reflection Maxima *198*
6.3.4.2 Out-of-Band Transmission Maxima *199*
6.4 Predistorted Filter Characteristics *200*
6.4.1 Synthesis of the Predistorted Filter Network *205*
6.5 Transformation for Symmetric Dual-Passband Filters *208*
Summary *211*
References *211*
Appendix 6A *212*
Complex Terminating Impedances in Multiport Networks *212*
6A.1 Change of Termination Impedance *213*
References *213*

7 Synthesis of Network-Circuit Approach *215*
7.1 Circuit Synthesis Approach *216*
7.1.1 Buildup of the [$ABCD$] Matrix for the Third-Degree Network *217*
7.1.2 Network Synthesis *218*
7.2 Lowpass Prototype Circuits for Coupled-Resonator Microwave Bandpass Filters *221*

7.2.1 Synthesis of [*ABCD*] Polynomials for Circuits with Inverters *222*
7.2.1.1 Form *225*
7.2.1.2 Equations *226*
7.2.2 Synthesis of [*ABCD*] Polynomials for the Singly Terminated Filter Prototype *227*
7.3 Ladder Network Synthesis *229*
7.3.1 Parallel-Coupled Inverters: Extraction Process *233*
7.3.2 Extraction of Parallel Admittance Inverters *233*
7.3.3 Summary of the Principal Components of the Prototype Network *235*
7.4 Synthesis Example of an Asymmetric (4–2) Filter Network *235*
7.4.1 Scaling at Resonator Nodes *241*
7.4.2 Scaling of Inverters to Unity *242*
Summary *244*
References *245*

8 Synthesis of Networks: Direct Coupling Matrix Synthesis Methods *247*
8.1 The Coupling Matrix *247*
8.1.1 Bandpass and Lowpass Prototypes *248*
8.1.2 Formation of the General $N \times N$ Coupling Matrix and Its Analysis *250*
8.1.2.1 "$N \times N$" and "$N + 2$" Coupling Matrices *251*
8.1.3 Formation of the Coupling Matrix from the Lowpass Prototype Circuit Elements *253*
8.1.4 Analysis of the Network Represented by the Coupling Matrix *255*
8.1.5 Direct Analysis *257*
8.2 Direct Synthesis of the Coupling Matrix *258*
8.2.1 Direct Synthesis of the $N \times N$ Coupling Matrix *259*
8.3 Coupling Matrix Reduction *261*
8.3.1 Similarity Transformation and Annihilation of Matrix Elements *261*
8.3.1.1 Reduction Procedure: Full Coupling Matrix to Folded Canonical Form *264*
8.3.1.2 Example of Usage *265*
8.4 Synthesis of the $N + 2$ Coupling Matrix *268*
8.4.1 Synthesis of the Transversal Coupling Matrix *269*
8.4.1.1 Synthesis of Admittance Function $[Y_N]$ from the Transfer and Reflection Polynomials *269*
8.4.1.2 Synthesis of Admittance Function $[Y_N]$: Circuit Approach *270*
8.4.1.3 Synthesis of the Two-Port Admittance Matrix $[Y_N]$ *272*
8.4.1.4 Synthesis of the $N + 2$ Transversal Matrix *273*
8.4.2 Reduction of the $N + 2$ Transversal Matrix to the Folded Canonical Form *275*
8.4.3 Illustrative Example *276*
8.4.4 $N + 2$ Coupling Matrix Synthesis for Networks with Complex Terminations *278*
8.4.4.1 Fully Canonical Characteristics with Complex Terminations *280*
8.5 Even- and Odd-Mode Coupling Matrix Synthesis Technique: the Folded Lattice Array *282*
8.5.1 "Straight" Cross-Couplings *284*
8.5.2 Diagonal Cross-Couplings *284*
8.5.3 Odd-Degree Network: Bisection of the Central Resonant Node *285*
8.5.4 Even-Mode and Odd-Mode Circuits of the Symmetric Lattice Network *285*
8.5.5 Development of the Even- and Odd-Mode Admittance Polynomials from the Transfer and Reflection Polynomials *287*
8.5.5.1 Canonical and Bandstop Cases *288*

8.5.5.2 Construction of the Even- and Odd-Mode Admittance Polynomials *289*
8.5.6 Example of the Synthesis of the Coupling Matrix for a Symmetric Lattice Network *289*
Summary *292*
References *293*

9 Reconfiguration of the Folded Coupling Matrix *295*
9.1 Symmetric Realizations for Dual-Mode Filters *295*
9.1.1 Sixth-Degree Filter *297*
9.1.2 Eighth-Degree Filter *298*
9.1.3 10th-Degree Filter *298*
9.1.4 12th-Degree Filter *299*
9.1.4.1 Conditions for Symmetric Realizations *300*
9.2 Asymmetric Realizations for Symmetric Characteristics *300*
9.3 "Pfitzenmaier" Configurations *301*
9.4 Cascaded Quartets (CQs): Two Quartets in Cascade for Degrees Eight and Above *304*
9.5 Parallel-Connected Two-Port Networks *306*
9.5.1 Even-Mode and Odd-Mode Coupling Submatrices *310*
9.6 Cul-de-Sac Configuration *311*
9.6.1 Further Cul-de-Sac Forms *312*
9.6.1.1 Proof-of-Concept Model *314*
9.6.2 Sensitivity Considerations *319*
Summary *321*
References *321*

10 Synthesis and Application of Extracted Pole and Trisection Elements *323*
10.1 Extracted Pole Filter Synthesis *323*
10.1.1 Synthesis of the Extracted Pole Element *324*
10.1.1.1 Removal of Phase Length *324*
10.1.1.2 Removal of Resonant Pair *325*
10.1.2 Example of Synthesis of Extracted Pole Network *327*
10.1.3 Analysis of the Extracted Pole Filter Network *330*
10.1.4 Direct-Coupled Extracted Pole Filters *333*
10.2 Synthesis of Bandstop Filters Using the Extracted Pole Technique *335*
10.2.1 Direct-Coupled Bandstop Filters *338*
10.2.1.1 Cul-de-Sac Forms for the Direct-Coupled Bandstop Matrix *341*
10.3 Trisections *343*
10.3.1 Synthesis of the Trisection: Circuit Approach *344*
10.3.2 Cascade Trisections: Coupling Matrix Approach *349*
10.3.2.1 Synthesis of the Arrow Canonical Coupling Matrix *350*
10.3.2.2 Synthesis of the Arrow Coupling Matrix Form *350*
10.3.2.3 Creation and Repositioning of Trisections within the Arrow Coupling Matrix *351*
10.3.2.4 Illustrative Example *352*
10.3.2.5 Trisections Realizing Complex Transmission Zeros *355*
10.3.3 Techniques Based on the Trisection for Synthesis of Advanced Circuits *356*
10.3.3.1 Synthesis Procedure for Cascade Quartets, Quintets, and Sextets *357*
10.3.3.2 Cascaded Quartets *357*
10.3.3.3 Formal Procedure for Cascade Quartets, Quintets, and Sextets *357*

10.3.3.4 Cascaded Quintets *358*
10.3.3.5 Cascaded Sextets *358*
10.3.3.6 Multiple Cascaded Sections *360*
10.4 Box Section and Extended Box Configurations *361*
10.4.1 Box Sections *362*
10.4.2 Extended Box Sections *364*
Summary *371*
References *371*

11 Microwave Resonators *373*
11.1 Microwave Resonator Configurations *373*
11.2 Calculation of Resonant Frequency *376*
11.2.1 Resonance Frequency of Conventional Transmission-Line Resonators *376*
11.2.2 Resonance Frequency Calculation Using the Transverse Resonance Technique *378*
11.2.3 Resonance Frequency of Arbitrarily Shaped Resonators *380*
11.2.3.1 Eigenmode Analysis *380*
11.2.3.2 *S*-Parameter Analysis *381*
11.3 Resonator Unloaded *Q* Factor *383*
11.3.1 Unloaded *Q* Factor of Conventional Resonators *384*
11.3.1.1 A Coaxial $\lambda/2$ Resonator Terminated in a Short Circuit *384*
11.3.1.2 A $\lambda/2$ Microstrip Resonator *385*
11.3.2 Unloaded *Q* of Arbitrarily Shaped Resonators *386*
11.4 Measurement of Loaded and Unloaded *Q* Factor *387*
11.4.1 Measurement of Unloaded *Q* Using Polar Display of Reflection Coefficient *392*
11.4.2 Measurement of Unloaded *Q* Using Linear Display of Reflection Coefficient *392*
Summary *393*
References *393*

12 Waveguide and Coaxial Lowpass Filters *395*
12.1 Commensurate-Line Building Elements *395*
12.2 Lowpass Prototype Transfer Polynomials *396*
12.2.1 Chebyshev Polynomials of the Second Kind *397*
12.2.2 Achieser–Zolotarev Functions *399*
12.2.2.1 Synthesis of Achieser–Zolotarev Functions *400*
12.3 Synthesis and Realization of the Distributed Stepped Impedance Lowpass Filter *401*
12.3.1 Mapping the Transfer Function S_{21} from the ω Plane to the θ Plane *401*
12.3.1.1 Form of $S_{21}(t)$ for the Circuit Approach *402*
12.3.2 Synthesis of the Stepped Impedance Lowpass Prototype Circuit *404*
12.3.2.1 Network Synthesis *404*
12.3.3 Realization *405*
12.3.3.1 Realization of the Impedance Inverters *408*
12.4 Short-Step Transformers *410*
12.5 Synthesis and Realization of Mixed Lumped/Distributed Lowpass Filters *411*
12.5.1 Formation of the Transfer and Reflection Polynomials *413*
12.5.2 Synthesis of the Tapered–Corrugated Lowpass Prototype Circuit *415*
12.5.2.1 Network Synthesis *416*
12.5.3 Realization *418*
12.5.3.1 Synthesis Accuracy Considerations *419*

12.5.3.2 Spurious Responses *421*
12.5.3.3 Repeat Passbands *421*
12.5.3.4 Higher-Order TE_{n0} Modes in Waveguide *421*
12.5.3.5 Mode Conversion *422*
12.5.3.6 Example of Synthesis *422*
Summary *425*
References *426*

13 Waveguide Realization of Single- and Dual-Mode Resonator Filters *427*
13.1 Synthesis Process *428*
13.2 Design of the Filter Function *428*
13.2.1 Amplitude Optimization *429*
13.2.2 Rejection Lobe Optimization *429*
13.2.3 Group Delay Optimization *431*
13.2.3.1 Group Delay and Attenuation Slope Calculation *432*
13.2.3.2 Group Delay Optimization Procedure *432*
13.3 Realization and Analysis of the Microwave Filter Network *434*
13.4 Dual-Mode Filters *440*
13.4.1 Virtual Negative Couplings *441*
13.5 Coupling Sign Correction *442*
13.6 Dual-Mode Realizations for Some Typical Coupling Matrix Configurations *444*
13.6.1 Folded Array *444*
13.6.2 Pfitzenmaier Configuration *445*
13.6.3 Propagating Forms *446*
13.6.4 Cascade Quartet *446*
13.6.5 Extended Box *447*
13.7 Phase- and Direct-Coupled Extracted Pole Filters *447*
13.7.1 Pole Cavity and Interconnecting Stub *449*
13.7.2 Phase Lengths between the Filter Body and Pole Cavities *450*
13.7.3 Phase Lengths between Adjacent Pole Cavities *450*
13.8 The "Full-Inductive" Dual-Mode Filter *450*
13.8.1 Synthesis of the Equivalent Circuit *452*
Summary *454*
References *454*

14 Design and Physical Realization of Coupled Resonator Filters *457*
14.1 Circuit Models for Chebyshev Bandpass Filters *459*
14.2 Calculation of Interresonator Coupling *463*
14.2.1 The Use of Electric Wall and Magnetic Wall Symmetry *463*
14.2.2 Interresonator Coupling Calculation Using *S*-Parameters *465*
14.3 Calculation of Input/Output Coupling *467*
14.3.1 Frequency Domain Method *467*
14.3.2 Group Delay Method *468*
14.4 Design Example of Dielectric Resonator Filters Using the Coupling Matrix Model *468*
14.4.1 Calculation of Dielectric Resonator Cavity Configuration *470*
14.4.2 Calculation of Iris Dimensions for Interresonator Coupling *471*
14.4.3 Calculation of Input/Output Coupling *472*
14.5 Design Example of a Waveguide Iris Filter Using the Impedance Inverter Model *475*

14.6 Design Example of a Microstrip Filter Using the *J*-Admittance Inverter Model *478*
Summary *483*
References *484*

15 Advanced EM-Based Design Techniques for Microwave Filters *485*
15.1 EM-Based Synthesis Techniques *485*
15.2 EM-Based Optimization Techniques *486*
15.2.1 Optimization Using an EM Simulator *487*
15.2.2 Optimization Using Semi-EM-Based Simulator *488*
15.2.3 Optimization Using an EM Simulator with Adaptive Frequency Sampling *490*
15.2.4 Optimization Using EM-Based Neural Network Models *491*
15.2.5 Optimization Using EM-Based Multidimensional Cauchy Technique *495*
15.2.6 Optimization Using EM-Based Fuzzy Logic *495*
15.3 EM-Based Advanced Design Techniques *496*
15.3.1 Space Mapping Techniques *496*
15.3.1.1 The Original Space Mapping Approach *498*
15.3.1.2 Aggressive Space Mapping (ASM) Approach *500*
15.3.2 Calibrated Coarse Model (CCM) Techniques *504*
15.3.3 Generalized Calibrated Coarse Model Technique for Filter Design *510*
Summary *513*
References *514*

16 Dielectric Resonator Filters *517*
16.1 Resonant Frequency Calculation in Dielectric Resonators *517*
16.2 Rigorous Analyses of Dielectric Resonators *521*
16.2.1 Mode Charts for Dielectric Resonators *521*
16.3 Dielectric Resonator Filter Configurations *524*
16.4 Design Considerations for Dielectric Resonator Filters *528*
16.4.1 Achievable Filter *Q* Value *528*
16.4.1.1 Size of Enclosure *528*
16.4.1.2 Support Structure *528*
16.4.1.3 Effect of Tuning Elements *528*
16.4.2 Spurious Performance of Dielectric Resonator Filters *529*
16.4.3 Temperature Drift *530*
16.4.4 Power Handling Capability *531*
16.5 Other Dielectric Resonator Configurations *531*
16.6 Cryogenic Dielectric Resonator Filters *534*
16.7 Hybrid Dielectric/Superconductor Filters *536*
16.8 Miniature Dielectric Resonators *538*
16.8.1 A Quadruple-Mode Dielectric Resonator Using a Single Traditional Cylindrical Dielectric Resonator *538*
16.8.2 A Dual-Mode Dielectric Resonator Using a Half-Cut Dielectric Resonator *539*
Summary *542*
References *543*

17 Allpass Phase and Group Delay Equalizer Networks *545*
17.1 Characteristics of Allpass Networks *545*
17.2 Lumped-Element Allpass Networks *547*
17.2.1 Resistively Terminated Symmetric Lattice Networks *548*

17.2.2 Network Realizations *549*
17.2.2.1 First-Order Lattice Section *550*
17.2.2.2 Second-Order Lattice Section *550*
17.3 Microwave Allpass Networks *551*
17.4 Physical Realization of Allpass Networks *554*
17.4.1 Transmission-Type Equalizers *555*
17.4.2 Reflection-Type Allpass Networks *556*
17.5 Synthesis of Reflection-Type Allpass Networks *557*
17.6 Practical Narrowband Reflection-Type Allpass Networks *558*
17.6.1 C-Section Allpass Equalizer in Waveguide Structure *559*
17.6.2 D-Section Allpass Equalizer in Waveguide Structure *560*
17.6.3 Narrowband TEM Reactance Networks *560*
17.6.4 Multiple Coupled Cavity Reactance Networks *561*
17.7 Optimization Criteria for Allpass Networks *561*
17.8 Dissipation Loss *566*
17.9 Equalization Tradeoffs *567*
Summary *567*
References *568*

18 Multiplexer Theory and Design *569*
18.1 Background *569*
18.2 Multiplexer Configurations *571*
18.2.1 Hybrid-Coupled Approach *571*
18.2.2 Circulator-Coupled Approach *571*
18.2.3 Directional Filter Approach *571*
18.2.4 Manifold-Coupled Approach *574*
18.3 RF Channelizers (Demultiplexers) *575*
18.3.1 Hybrid Branching Network *576*
18.3.2 Circulator-Coupled MUX *577*
18.3.3 En Passant Distortion *578*
18.3.3.1 En Passant Distortion for a Noncontiguous Multiplexing Scheme *579*
18.3.3.2 En Passant Distortion with Contiguous Channels *579*
18.4 RF Combiners *581*
18.4.1 Circulator-Coupled MUX *582*
18.4.2 Hybrid-Coupled Filter Module (HCFM) Multiplexer *582*
18.4.2.1 En Passant Distortion *584*
18.4.3 Directional Filter Combiner *586*
18.4.4 Manifold Multiplexer *586*
18.4.4.1 Star Junction Multiplexer *587*
18.4.4.2 Manifold-Coupled Multiplexer *587*
18.4.4.3 Design Methodology *588*
18.4.4.4 Analysis of Common Port Return Loss and Channel Transfer Characteristics *589*
18.4.4.5 Common Port Return Loss *592*
18.4.4.6 Channel Transfer Characteristics *592*
18.4.4.7 Channel Filter Design *593*
18.4.4.8 Optimization Strategy *594*
18.4.4.9 Demonstration of the Piecewise Optimization Process and Stages in the Optimization of a Four-Channel Contiguous OMUX *598*
18.5 Transmit–Receive Diplexers *601*

18.5.1 Internal Voltage Levels in Tx/Rx Diplexer Filters *603*
Summary *606*
References *607*

19 Computer-Aided Diagnosis and Tuning of Microwave Filters *609*
19.1 Sequential Tuning of Coupled Resonator Filters *610*
19.1.1 Tuning Steps *612*
19.2 Computer-Aided Tuning Based on Circuit Model Parameter Extraction *615*
19.3 Computer-Aided Tuning Based on Poles and Zeros of the Input Reflection Coefficient *619*
19.4 Time-Domain Tuning *622*
19.4.1 Time-Domain Tuning of Resonator Frequencies *623*
19.4.2 Time-Domain Tuning of Interresonator Coupling *624*
19.4.3 Time-Domain Response of a Golden Filter *627*
19.5 Filter Tuning Based on Fuzzy Logic Techniques *627*
19.5.1 Description of Fuzzy Logic Systems *628*
19.5.2 Steps in Building the FL System *629*
19.5.3 Comparison between Boolean Logic and Fuzzy Logic *632*
19.5.4 Applying Fuzzy Logic to Filter Tuning *634*
19.6 Automated Setups for Filter Tuning *637*
Summary *639*
References *640*

20 High-Power Considerations in Microwave Filter Networks *643*
20.1 Background *643*
20.2 High-Power Requirements in Wireless Systems *643*
20.3 High-Power Amplifiers (HPAs) *645*
20.4 Gas Discharge *645*
20.4.1 Basics *646*
20.4.2 Breakdown in Air *647*
20.4.3 Power Rating of Waveguides and Coaxial Transmission Lines *649*
20.4.4 Derating Factors *650*
20.4.5 Impact of Thermal Dissipation on Power Rating *650*
20.5 Multipaction Breakdown *651*
20.5.1 Dependence on Vacuum Environment *652*
20.5.2 Dependence on Surface Conditions of Materials *653*
20.5.3 Dependence on Signal Amplitude, Signal Frequency, and Geometry of the Device *654*
20.5.4 Dependence on Number of Carriers and Modulation *654*
20.5.5 Two-Surface Theoretical Analysis, Simulation, and Prediction *654*
20.5.5.1 Single-Carrier Theory *654*
20.5.5.2 Multi-Carrier Theory *656*
20.5.6 Multipactor Breakdown Levels *657*
20.5.7 Detection and Prevention of Multipaction *659*
20.5.8 Design Margins in Multipaction *660*
20.6 High-Power Bandpass Filters *662*
20.6.1 Bandpass Filters Limited by Thermal Dissipation *663*
20.6.2 Bandpass Filters Limited by Voltage Breakdown *664*
20.6.3 Filter Prototype Network *664*

20.6.4 Lumped to Distributed Scaling *666*
20.6.5 Resonator Voltages from Prototype Network *666*
20.6.6 Example and Verification via FEM Simulation *667*
20.6.7 Example of High Voltages in a Multiplexer *668*
20.6.8 Summary of Multipactor Breakdown Evaluation Methodologies *669*
20.7 Passive Intermodulation (PIM) Consideration for High-Power Equipment *670*
20.7.1 PIM Measurement *672*
20.7.2 PIM Control Guidelines *673*
Summary *674*
Acknowledgment *675*
References *675*

21 Multiband Filters *679*
21.1 Introduction *679*
21.2 Approach I: Multiband Filters Realized by Having Transmission Zeros Inside the Passband of a Bandpass Filter *681*
21.3 Approach II: Multiband Filters Employing Multimode Resonators *683*
21.3.1 Combline Multiband Filters Employing Multimode Resonators *685*
21.3.2 Waveguide Multiband Filters Employing Multimode Resonators *686*
21.3.3 Dielectric Resonator Multiband Filters Employing Multimode Resonators *693*
21.4 Approach III: Multiband Filters Using Parallel Connected Filters *700*
21.5 Approach IV: Multiband Filter Implemented Using Notch Filters Connected in Cascade with a Wideband Bandpass *701*
21.6 Use of Dual-Band Filters in Diplexer and Multiplexer Applications *703*
21.7 Synthesis of Multiband Filters *705*
21.7.1 Introduction *705*
21.7.2 Multi-passband Characteristics: Prototype Design *707*
21.7.3 Limitations and Restrictions *711*
21.7.4 Coupling Matrix Synthesis *712*
21.7.5 Dual-Band Filters with Transmission Zeros at Zero Frequency *712*
21.7.5.1 Multiple Transmission Zeros at Zero Frequency *713*
21.7.6 Synthesis of Multiband Coupled-Resonator Filters Using Reactance Transformations *718*
21.7.6.1 Synthesis of the Compound Resonator Network *720*
21.7.6.2 Multiband Bandstop Filters *722*
21.7.6.3 Alternative Configurations for the Compound Resonator *724*
Summary *727*
References *728*

22 Tunable Filters *731*
22.1 Introduction *731*
22.2 Major Challenges in Realizing High-Q 3D Tunable Filters *733*
22.2.1 Maintaining Constant Bandwidth and a Reasonable Return Loss Over a Wide Tuning Range *733*
22.2.2 Maintaining a Constant High Q Value Over a Wide Tuning Range *733*
22.2.3 Integration of Tuning Elements with 3D Filters *733*
22.2.4 Linearity and Power Handling Capability *734*
22.3 Combline Tunable Filters *734*
22.3.1 Combline Tunable Filters with a Constant Absolute Bandwidth *734*

22.3.1.1 Coupling Iris Design for Constant Bandwidth *737*
22.3.1.2 Input/Output Coupling Design for Constant Bandwidth *738*
22.3.2 MEMS-Based Combline Tunable Filters *744*
22.4 Tunable Dielectric Resonator Filters *752*
22.4.1 Tunable Dielectric Resonator Filters Employing Mechanical or Piezoelectric Actuators *752*
22.4.2 MEMS-Based Tunable Dielectric Resonator Filters *759*
22.5 Waveguide Tunable Filters *772*
22.6 Filters with Tunable Bandwidth *776*
Summary *778*
References *779*

23 Practical Considerations and Design Examples *785*
Chandra M. Kudsia, Vicente E. Boria, and Santiago Cogollos
23.1 System Considerations for Filter Specifications in Communication Systems *785*
23.1.1 Equivalent Linear Frequency Drift (ELFD) *786*
23.1.1.1 Material Selection Tradeoffs and ELFD *787*
23.1.2 Synchronous versus Asynchronous Resonator Filters *788*
23.1.2.1 Synchronously Tuned Filters *788*
23.1.2.2 Asynchronously Tuned Filters *789*
23.1.3 Filter Tradeoffs in the Characterization of RF Channels *789*
23.2 Filter Synthesis Techniques and Topologies *796*
23.3 Multiplexers *827*
23.4 High-Power Considerations *839*
23.5 Tolerance and Sensitivity Analysis in Filter Design *851*
23.5.1 Manufacturing Tolerance Analysis *851*
23.5.2 Sensitivity Analysis of a Dielectric-Loaded Cavity *853*
23.5.3 Degradation Caused by Tuning Screws in Filter Design *856*
23.5.3.1 Degradation of Quality Factor (Q_0) *856*
23.5.3.2 Degradation of Spurious-Free Window *856*
Summary *858*
Acknowledgments *858*
Appendix 23A *858*
Thermal Expansion *858*
References *859*

A Physical Constants *861*

B Conductivities of Metals *863*

C Dielectric Constants and Loss Tangents of Some Materials *865*

D Rectangular Waveguide Designation *867*

E Impedance and Admittance Inverters *869*
E.1 Filter Realization with Series Elements *869*
E.2 Normalization of the Element Values *872*
E.3 General Lowpass Prototype Case *873*
E.3.1 Coupling Coefficient: Lowpass Prototype *874*

E.4 Bandpass Prototype *874*
E.4.1 Slope Parameter *875*
E.4.2 Coupling Matrix Parameter *M* *875*
E.4.3 Coupling Coefficient: Bandpass Prototype *876*
E.4.4 Slope Parameter of Transmission-Line Resonators *876*
E.4.5 Slope Parameter for Waveguide Resonators *877*
E.4.6 Practical Impedance and Admittance Inverters *878*
References *878*

Index *879*

Preface

Three new chapters are introduced in the second edition. Chapters on multiband filters and tunable filters are added to reflect the emerging markets for wireless systems. The third chapter is devoted to the practical aspects of design and implementation of microwave filters and multiplexing networks. Chapters from edition 1 have undergone a thorough review and minor revisions. New sections have been added in Chapters 1, 6, 8, 16, and 20.

The book begins with a simple model of a communication system. It addresses the issues on: (i) whether there is a limitation on the available bandwidth for a wireless communication system, (ii) what the limitations are for transmitting information in the available bandwidth, and (iii) what the cost-sensitive parameters of a communication system are. Each issue is then addressed to gain understanding of various system parameters with emphasis on the role and requirements of filter networks in different parts of the communication system. This sets the stage to address the fundamentals of filter design based on circuit theory approximation. It continues with a description of classical filters. This is followed by the development of computer-aided techniques to generate a general class of prototype filter functions, exhibiting a symmetrical or asymmetrical frequency response. This general formulation is accomplished by incorporating hypothetical *frequency invariant reactive* (FIR) elements in the lowpass prototype filter design. The FIR elements show up as frequency offsets of resonant circuits in real bandpass or bandstop filters. Absence of FIR elements represents the classical filter function that gives rise to symmetrical frequency response. From this general formulation of the filter function, synthesis techniques are described to realize the equivalent lumped parameter circuit model of filter networks. The next step in the synthesis procedure is to translate the circuit model of the filter into its equivalent microwave structure. As a first approximation, this can be achieved by making use of the extensive existing data that relates circuit models to the physical dimensions and properties of structures used for microwave filters. For more accurate determination of physical dimensions, modern electromagnetic (EM)-based techniques and tools are described to determine filter dimensions with near-arbitrary accuracy. This knowledge is carried through in the design of multiplexing networks having arbitrary bandwidths and channel separations.

Separate chapters are devoted to computer-aided tuning and high-power considerations in filter design. Our goal has been to give the reader a broad view of filter requirements and design and sufficient depth to follow continuing advances in this field. Throughout the book, emphasis has been on fundamentals and practical considerations in filter design. Distinct features of the book include (i) system considerations in the design of filters, (ii) the general formulation and synthesis of filter functions including the FIR elements, (iii) synthesis techniques for lowpass prototype filters exhibiting symmetrical or asymmetrical frequency response in a variety of topologies, (iv) application of EM techniques to optimize physical dimensions of microwave filter structures, (v) design and tradeoffs of various multiplexer configurations,

(vi) computer-aided filter tuning techniques, and (vii) high-power considerations for terrestrial and space applications. The material in the book is organized in 23 chapters:

- Chapter 1 is devoted to an overview of communication systems, more specifically to the relationship between the communication channel and other elements of the system. The intent here is to provide the reader with sufficient background to be able to appreciate the critical role and requirements of radio frequency (RF) filters in communication systems.
  - In the second edition, Digital Transmission, The Channelizer Section, Frequency Plan, and Limitations of Microwave Filter Technology have been revised. The section on RF Filters for Cellular Systems is modified to reflect the requirement of additional frequency bands to meet the explosive growth in wireless services. A section has been added on Ultra Wideband (UWB) Wireless Communication. The summary at the end of the chapter has been revised to reflect the changes.
- The principles that unify communication theory and circuit theory approximations are explained in Chapter 2. It highlights the essential assumptions and the success of the frequency analysis approach that we take for granted in analyzing electrical networks.
- Chapter 3 describes the synthesis of the characteristic polynomials to realize the classical maximally flat, Chebyshev, and elliptic function lowpass prototype filters. It includes a discussion of FIR elements and their inclusion to generate filter functions with asymmetrical frequency response. This leads to transfer function polynomials (with certain restrictions) with complex coefficients, a distinct departure from the more familiar characteristic polynomials with rational and real coefficients. This provides a basis to analyze the most general class of filter functions in the lowpass prototype domain, including minimum and nonminimum phase filters, exhibiting a symmetrical or asymmetrical frequency response.
- Chapter 4 presents the synthesis of characteristic polynomials of lowpass prototype filters with arbitrary amplitude response using computer-aided optimization technique. The key lies in making sure that the optimization procedure is highly efficient. This is accomplished by determining the gradients of the objective function analytically and linking it directly to the desired amplitude response shape. It includes minimum phase and nonminimum phase filters exhibiting a symmetrical or asymmetrical frequency response. To demonstrate the flexibility of this method, examples of some unconventional filters are included.
- Chapter 5 provides a review of the basic concepts used in the analysis of multiport microwave networks. These concepts are important for filter designers since any filter or multiplexer can be divided into smaller two-, three-, or N-port networks connected together. Five matrix representations of microwave networks are described, namely, $[Z]$, $[lY]$, $[ABCD]$, $[S]$, and $[T]$ matrices. These matrices are interchangeable, where the elements of any matrix can be written in terms of those of the other four matrices. Familiarity with the concepts of these matrices is essential in understanding the material presented in this book.
- Chapter 6 begins with a review of some important scattering parameter relations that are relevant for the synthesis of filter networks. This is followed by a discussion of the general kind of Chebyshev function and its application in generating the transfer and reflection polynomials for the equi-ripple class of filter characteristics with an arbitrary distribution of the transmission zeros. In the final part of this chapter, the special cases of predistorted and dual-band filtering functions are discussed.
  - The second edition has two added features: a section for finding the positions of the in-band reflection maxima and the out-of-band transmission maxima of the generalized Chebyshev prototype filter and an appendix extending the two-port S-parameter analysis and synthesis to multiport networks with complex terminations. A section has been added describing the relationship between the characteristic polynomials, S-parameters, short-circuit admittance, and $[ABCD]$ transfer matrix parameters.

- In Chapter 7, filter synthesis based on the [*ABCD*] matrix is presented. The synthesis procedure is broken down into two stages. The first stage involves lumped element lossless inductors, capacitors, and FIR elements. The second stage includes the immitance inverters. Use of such inverters allows for the prototype electrical circuit in a form suitable for realization with intercoupled microwave resonators. The technique is applicable for synthesizing lowpass prototype filters with symmetrical or asymmetrical response, in ladder form, as well as cross-coupled topologies. A further generalization is introduced to allow the synthesis of singly terminated filters. The synthesis process described in this chapter represents the most general technique for synthesizing lumped element, lowpass prototype filter networks.
- In Chapter 8, the concept of $N \times N$ coupling matrix for the synthesis of bandpass prototype filters is introduced. The procedure is modified by including FIR elements to allow synthesis of asymmetric filter response as well. The procedure is then extended to $N + 2$ coupling matrix by separating out the purely resistive and purely reactive portions of the $N \times N$ matrix. The $N + 2$ coupling matrix allows multiple couplings with respect to the input and output ports, in addition to the main input/output couplings to the first and last resonators as envisaged in the $N \times N$ coupling matrix. This allows synthesis of fully canonical filters and simplifies the process of similarity transformations to realize other filter topologies. This synthesis process yields the general coupling matrix with finite entries for all the couplings. The next step in the process is to derive topologies with a minimum number of couplings, referred to as canonical forms. This is achieved by applying similarity transformations to the coupling matrix. Such transformations preserve the eigenvalues and eigenvectors of the matrix, thus ensuring that the desired filter response remains unaltered. There are two principal advantages of this synthesis technique. Once the general coupling matrix with all the permissible couplings has been synthesized, it allows matrix operations on the coupling matrix to realize a variety of filter topologies. The second advantage is that the coupling matrix represents the practical bandpass filter topology. Therefore, it is possible to identify each element of the practical filter uniquely, including its Q value, dispersion characteristics, and sensitivity. This permits a more accurate determination of the practical filter characteristics and an insight into ways to optimize filter performance.
  - In the second edition, two new sections have been added: (i) $N + 2$ Coupling Matrix Synthesis for Networks with Complex Terminations and (ii) Even and Odd Mode Coupling Matrix Synthesis Technique: The Folded Lattice Array.
- Chapter 9 develops methods of similarity transformations to realize a wide range of topologies appropriate for dual-mode filter networks. Dual-mode filters make use of two orthogonally polarized degenerate modes, supported in a single physical resonator, be it a cavity, a dielectric disc, or a planar structure, thereby allowing a significant reduction in the size of filters. Besides the longitudinal and folded configurations, structures referred to as cascade quartets and cul de sac filters are also included. The chapter concludes with examples and a discussion of the sensitivity of the various dual-mode filter topologies.
- In Chapter 10, we introduce two unusual circuit sections: the extracted pole section and the trisection. These sections are capable of realizing one transmission zero each. They can be cascaded with other circuit elements in the filter network. Application of these sections extends the range of topologies for realizing microwave filters. This is demonstrated by synthesizing filters that include cascaded quartet, quintet, and sextet filter topologies. Lastly, the synthesis of the box section, and its derivative, the extended box configuration, are explained. Examples are included to illustrate the intricacies of this synthesis procedure.
- Theoretical and experimental techniques for evaluating the resonant frequency and unloaded Q-factor of microwave resonators are described in Chapter 11. Resonators are the basic building blocks of any bandpass filter. At microwave frequencies, resonators can take

many shapes and forms. The chapter includes two approaches for calculating the resonant frequency of arbitrarily shaped resonators: the eigenmode analysis and the *S*-parameter analysis. Examples are given illustrating the implementation of these two techniques using EM-based commercial software tools such as high frequency system simulator (HFSS). It also includes a step-by-step procedure for measuring the loaded and unloaded *Q* values using either the polar display of a vector network analyzer or the linear display of a scalar network analyzer.

- Chapter 12 addresses the synthesis techniques for the realization of lowpass filters at microwave frequencies. Typical bandwidth requirements for lowpass filters in communication systems are in the gigahertz range. As a consequence, prototype models based on lumped elements are not suitable for realization at microwave frequencies. It requires the use of distributed elements for the prototype filters. The chapter begins with a description of the commensurate line elements and their suitability for realizing distributed lowpass prototype filters. It then goes into a discussion of characteristic polynomials that are best suited for modeling practical lowpass filters and methods to generate such polynomials. This is followed by a detailed description of the synthesis techniques for the stepped impedance and the lumped/distributed lowpass filters.
- Chapter 13 deals with the practical design aspects of dual-mode bandpass filters. It includes the use of dual-mode resonators that operate in the dominant mode, as well as in the higher order propagation modes. A variety of examples are included to illustrate the design procedure. These examples include longitudinal and canonical configurations, the extended box design, the extracted pole filter, and the filters with all inductive couplings. The examples also include symmetrical and asymmetrical response filters. The steps involved in the simultaneous optimization of amplitude and group delay response of a dual-mode linear phase filter are described. Examples in this chapter span the analysis and synthesis techniques described in Chapters 3–11.
- Chapter 14 presents the use of EM simulator tools for designing microwave filters. It is shown how one can couple the filter circuit models with EM simulation tools to synthesize the physical dimensions of microwave filters with near-arbitrary accuracy. The starting point for such computations is usually the physical dimensions derived from the best circuit model of the filter. Methods are described to compute, with much greater accuracy, the input/output and inter-resonator couplings by using the commercially available EM simulator software. The techniques can be adapted for a direct approach to determine the physical dimensions of filters from the elements of the coupling matrix [*M*], using *K*-impedance inverter, or *J*-admittance inverter models. Numerical examples are given in this chapter to illustrate, step by step, the application of this approach to the design of dielectric resonator, waveguide, and microstrip filters. For simple geometries with negligible coupling between nonadjacent resonators, this approach yields excellent results. Use of EM tools represents a major advance in the physical realization of microwave filters.
- Chapter 15 presents several techniques for EM-based design of microwave filters. The most direct approach is to combine an accurate EM simulation tool with an optimization software package and then optimize the physical dimensions of the filter to achieve the desired performance. This is effectively a tuning process where the tuning is done by the optimization package rather than a technologist. The starting point for this technique is the filter dimensions obtained using methods described in Chapter 14. Direct optimization approach, without any simplifying assumptions, can be still very computation intensive. A number of optimization strategies including adaptive frequency sampling, neural networks, and multidimensional Cauchy technique are described to reduce optimization time. Two advanced EM-based techniques, the space mapping technique (SM), and the coarse model technique

(CCM), are described in detail, offering a significant reduction in computation time. The chapter concludes with examples of filter dimensions obtained by using aggressive space mapping (ASM) and CCM techniques.

- Chapter 16 develops the design of dielectric resonator filters in a variety of configurations. Commercial software packages such as HFSS and CST Microwave Studio can be readily utilized to calculate the resonant frequency, field distribution, and resonator Q of dielectric resonators having any arbitrary shape. Using such tools, mode charts, along with plots, illustrating the field distribution of the first four modes in dielectric resonators are included. It also addresses the computation of the resonant frequency and the unloaded Q (Q_0) of cylindrical resonators, including the support structure. Tradeoffs in terms of Q_0, spurious response, temperature drift, and power handling capability are described. The chapter concludes with a detailed description of the design and tradeoffs for cryogenic dielectric resonator filters. Dielectric resonator filters are widely employed in wireless and satellite applications. Continuing advances in the quality of dielectric materials is a good indication of the growing application of this technology.
  - In the second edition, a new section has been added on miniature dielectric resonators illustrating a concept of realizing a quadruple-mode resonator using a traditional cylindrical shape dielectric resonator. It also illustrates how a half-cut dielectric resonator can be used in the realization of dual-mode filters.
- Chapter 17 deals with the analysis and synthesis of allpass networks, often referred to as equalizers. Such external allpass equalizers can be cascaded with filters to improve the phase and group delay response of filter networks. The chapter concludes with a discussion of the practical tradeoffs between the linear phase filters and externally equalized filter networks.
- Chapter 18 presents the design and tradeoffs for multiplexing networks for a variety of applications. It begins with a discussion of tradeoffs among the various types of multiplexing networks, including circulator-coupled, hybrid-coupled, and manifold-coupled multiplexers, employing single-mode or dual-mode filters. It also includes multiplexers based on using directional filters. This is followed by the detailed design considerations for each type of multiplexer. The design methodology and optimization strategy are dealt with in depth for the manifold-coupled multiplexer, by far the most complex microwave network. Numerous examples and photographs are included to illustrate the designs. The chapter is concluded with a brief discussion of the high-power capability of diplexers for cellular applications.
- Chapter 19 is devoted to the computer-aided techniques for tuning microwave filters. From a theoretical standpoint, the physical dimensions of a microwave filter can be perfected using EM-based techniques with near-arbitrary accuracy. In practice, the use of EM-based tools can be very time consuming and prohibitively so for higher-order filters and multiplexing networks. Moreover, owing to manufacturing tolerances and variations in material characteristics, practical microwave filters cannot duplicate the theoretical design. These problems are further exacerbated by the very stringent performance requirements for applications in the wireless and satellite communication systems. As a result, filter tuning is deemed an essential postproduction process. Techniques discussed in this chapter include (i) sequential tuning of coupled resonator filters, (ii) computer-aided tuning based on circuit model parameter extraction, (iii) computer-aided tuning using poles/zeros of the input reflection coefficient, (iv) time domain tuning, and (v) fuzzy logic tuning. The relative advantages of each technique are described.
- Chapter 20 provides an overview of high-power considerations in the design of microwave filters and multiplexing networks for terrestrial and space applications. It describes the phenomena of multipaction and gas discharge based on the classical theory and simple geometries, in particular the two-surface case under single carrier operation on infinite

parallel plates. It highlights the importance of derating factors that can severely degrade the performance of high-power equipment.

– In the second edition, the phenomenon of multipaction is described in depth. In most practical situations, the geometry of high-power RF equipment rarely corresponds to simple parallel-plate conductors, and the high-power equipment must be capable of handling multicarriers. The simple analysis is augmented to include multipaction and gas discharge analysis using numerical techniques. Such methods allow for a more accurate analysis of both single surface and dual surface multipactions in complex structures including nonhomogeneous fields. Although more complex and computer intensive, this analysis is much closer in taking into account the complex geometries of the high-power equipment and the typical operating environment with different numbers of RF carriers and modulations. A description of the classical setup for measuring RF breakdown effects is included. It highlights the efficacy of the industry-accepted methodology of applying the peak power method and the 20-gap crossing rule for the prediction of multipactor discharge in multicarrier operation.
– The chapter also provides guidelines to minimize passive intermodulation (PIM) in the design of high-power equipment.

Chapter 21

The new Chapter 21 provides an overview of the various techniques for the design of multiband filters, presenting and discussing several examples for dual-band and triple-band filters. The focus in this chapter is on high-Q multiband filters realized in coaxial, waveguide, and dielectric resonator structures. It also presents details of synthesis procedures for multiband filters. The chapter also illustrates how dual-band filters can be employed in realizing miniature diplexers and multiplexers.

Chapter 22

The new Chapter 22 provides an overview of tunable filter technology. It addresses the main challenges in realizing high-Q tunable filters, which include (i) maintaining constant bandwidth and a reasonable return loss over a wide tuning range, (ii) maintaining a constant high-Q value over a wide tuning range, (iii) integration of tuning elements with 3D filters, and (iv) linearity and power handling capability. It includes an approach for achieving a constant absolute bandwidth over a wide tuning range using only tuning elements for the resonators. A comparison between the various tuning elements (semiconductor, piezomotors, MEMS (microelectromechanical system), BST (barium strontium titanate), and PCM (phase change material)) is described. The chapter shows various examples for realizing tunable combline, dielectric resonator, and waveguide filters. Several examples were given with a focus on the use of MEMS in the design of such filters. Techniques are also presented for realizing filters that are tunable in both center frequency and bandwidth.

Chapter 23

The new chapter, entitled "Practical Considerations and Design Examples," is aimed at bridging the gap between theory and practical realization of microwave filters and multiplexing networks. A key feature of this chapter is the participation of Professors Vicente Boria and Santiago Cogollos of the Universitat Politecnica de Valencia, Spain, as co-authors. The chapter consists of a series of examples that highlight the methodology for designing and performing tradeoffs for practical filters and multiplexers. This provides a framework to analyze and optimize filtering requirements in communication systems, making use of this information to conduct filter tradeoffs; taking into account the typical operating environment (terrestrial or space), technology limitations, and manufacturing constraints; developing the circuit and quasi-distributed models of the filter topologies; and finally computing the physical dimensions of the structure using EM techniques. The chapter concludes with a brief overview of EM-based tolerance and sensitivity analysis in filter design.

Appendix E

A new appendix on the Impedance and Admittance Inverters has been added. A simple formulation for the application of inverters in filter design is described in this appendix.

The book is aimed at senior undergraduate and graduate students as well as practitioners of microwave technology. In writing this book, we have borrowed heavily from our industrial experience, giving seminars and teaching courses at universities and interactions with the engineering community at large at various conferences. It reflects a lifetime of experiences in advancing the state of the art in microwave filters and multiplexing networks.

1

Radio Frequency (RF) Filter Networks for Wireless Communications—The System Perspective

This chapter is dedicated to an overview of communication systems, especially the relationship between the communication channel and other elements of the system. The intent here is to provide the reader with sufficient background to be able to appreciate the critical role and requirements of radio frequency (RF) filters in communication systems. A number of standard texts [1–8] have been referred to in developing a significant portion of this chapter.

This chapter is divided into three parts: Part I presents a simple model of a communication system: the radio spectrum and its utilization, the concept of information, and system link budgets. Part II describes the noise and interference environment in a communication channel, the nonideal amplitude and phase characteristics of the channel, the choice of modulation–demodulation schemes, and how these parameters affect the efficient use of the allocated bandwidths. Part III discusses the impact of system design on the requirements and the specifications of microwave filter networks in satellite and cellular communication systems.

PART I Introduction to a Communication System, Radio Spectrum, and Information

1.1 Model of a Communication System

Communication refers to the process of conveying information-bearing signals from one point to another that are physically separate. In ancient times, people communicated over long distances by various means, such as smoke signals, drum beating, homing pigeons, and horseback riders. All such means were slow in the transmission of information over any appreciable distance. It was the invention of electricity that changed the means of communication. Communication became almost instantaneous by transmission of electrons through wires or electromagnetic (EM) waves through empty space or fibers, which is limited only by the speed of light—a fundamental constraint of our universe.

At the highest (simplest) level, communication involves an information source, a transmitter, a communication medium (or channel), a receiver, and an information destination (sink), as depicted in Figure 1.1. Until 1980s, most information was communicated in an

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

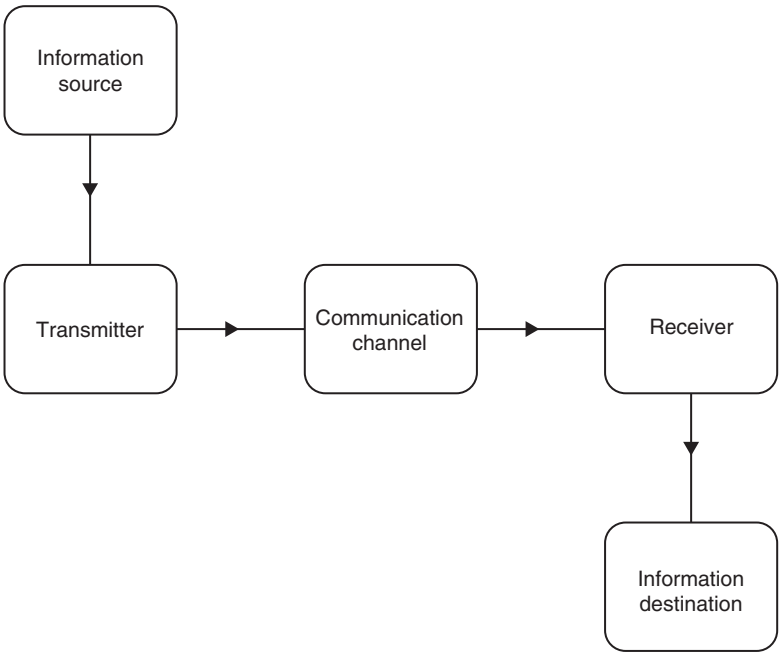


Figure 1.1 Simple model of a communication system.

analog format called *analog communication*. Today, most information is communicated in a digital format called *digital communication*. Even information in analog format is routinely converted into digital format for transmission and then converted back to analog format at the destination.

All communication systems are required to be linear. For such systems, the law of superposition holds. It allows the use of common media for the transmission and reception of an arbitrary number of independent signals, subject only to the constraints of the available bandwidths and adequate power levels. However, all the components of a communication system need not be linear, as long as the overall system is linear over the specified range of bandwidth within an acceptable degree of nonlinearity. In fact, all active devices are inherently nonlinear, which is essential for the purposes of frequency generation, modulation, demodulation, and amplification of signals. However, such intentional nonlinearities can be controlled for a specific application. For broadband wireless communication systems such as line-of-sight (LOS) and satellite systems used for long-distance networks, the frequency spectrum is divided into a number of RF channels, often referred to as *transponders*. Within each RF channel, there can be more than one RF carrier, depending on the system requirements. The channelization of the frequency spectrum provides the flexibility for the communication traffic flow in a multiuser environment. High-power amplifiers (HPAs) can operate with relatively high efficiency, since they are required to amplify a single carrier, or a limited number of signals, on a channel-by-channel basis, incurring a minimum distortion.

Irrespective of the communication format, it should be recognized that the passage of the transmitted signal through the communication channel is a strictly analog operation. The communication channel is a nonideal, lossy medium, and the reception of signals at the receiver involves recovery of the transmitted signal in the presence of impairments, in particular, thermal noise at the receiver, signal distortions (within the channel and from

nonideal transmitter and receiver), and interference from other signals or echoes (multipath) seen by the receiver.

1.1.1 Building Blocks of a Communication System

In this section, the building blocks shown in Figure 1.2 are described for both analog and digital communication systems.

1.1.1.1 Information Source

The information source consists of a large number of individual signals that are combined in a suitable format for transmission over the communication medium. Such a signal is referred to as the *baseband signal*. The transducers shown in Figure 1.2 are required to convert the energy of the individual information sources, either acoustic (voice) or electrical, into an appropriate electrical signal suitable for transmission. For an analog system, all the individual signals, as well as the combined baseband signal, are in an analog format as illustrated in Figure 1.2a.

For digital systems, the baseband signal is a digital datastream, whereas the individual signals constituting the baseband can be digital or analog. Consequently, the individual analog signals need to be converted into their equivalent digital format via analog-to-digital (A/D) converters. Another feature of the information source in a digital system is the use of data compression to conserve bandwidth. A compressor takes the digital data and exploits its redundancy and other features to reduce the amount of data that need to be transmitted but still permits the information to be recovered. The information source for a digital communication system is illustrated in Figure 1.2b.

1.1.1.2 Transmitter

The block diagram of a transmitter is presented in Figure 1.3, and the functionality of each element is described as follows:

Encoder. In digital systems, an encoder introduces error correction data into the baseband information stream that permits recovery of the digital information even after significant impairments are created in the communication channel.

Modulator. It transcribes the baseband signal onto a higher intermediate carrier frequency (IF) as an intermediate step in the transmission and reception of the information bearing signals. Use of IF simplifies the filtering and signal processing circuitry in the modulator. The modulator can shift the signal frequencies, change the bandwidth occupancy, or materially alter the form of the signal, making it more suitable and efficient for transmission over the communication medium.

Upconverter. Also referred to as a mixer, shifts the modulated IF carrier frequency to the microwave range of radio frequencies (RF) within the allocated frequency band for RF transmission.

RF amplifier. Is used to amplify the RF signal. RF power has a direct bearing on the communication capacity of a RF channel.

RF multiplexer. It is used to combine the power of a number of RF channels into a composite broadband signal for transmission via a common antenna.

Transmit antenna. Launches the RF power into space and focuses it toward the receiving station.

1.1.1.3 Communication Channel

For wireless systems, the communication channel is free space. As a consequence, the properties of space, including the atmosphere, play a critical role in system design.

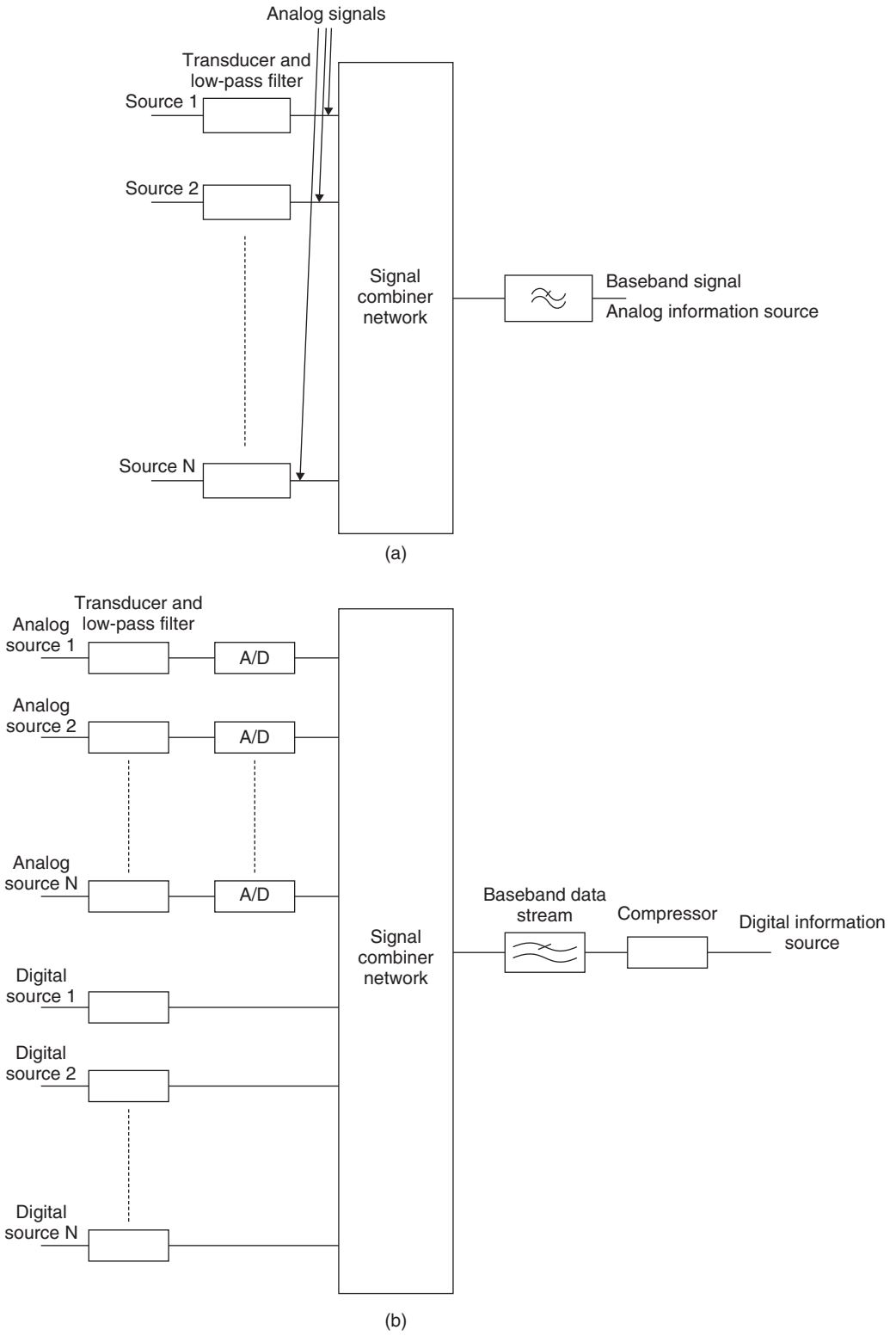


Figure 1.2 Information source: (a) analog system; (b) digital system.

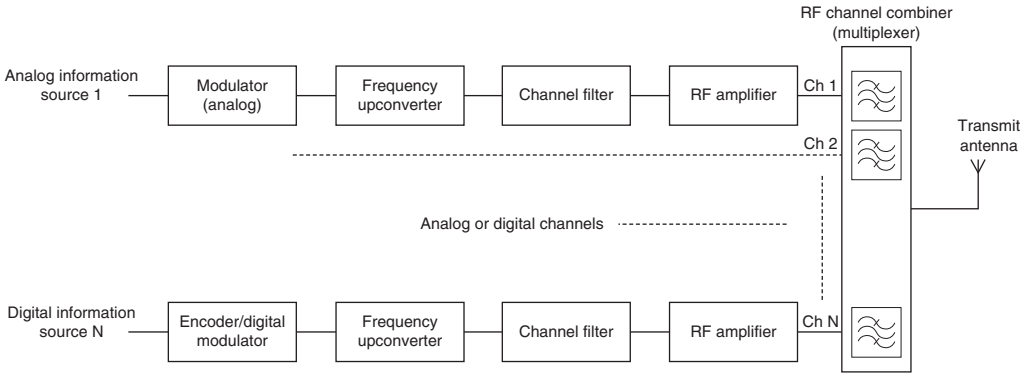


Figure 1.3 Transmitter block diagram.

1.1.1.4 Receiver

The block diagram of a receiver is shown in Figure 1.4, and the functionality of each element is described as follows:

Receive Antenna. It intercepts the RF power and focuses it to a transmission line connected to the low-noise amplifier (LNA).

LNA. It amplifies the very weak received signal with a minimum addition of noise to it.

Downconverter. It serves the function of frequency conversion, which is similar to that required in the transmitter chain. The downconverter converts the uplink frequency band into the downlink frequency band.

Demodulator. It extracts the baseband signal from the RF carrier by a process opposite to that of a modulator.

Decoder. It exploits the error correction data previously inserted into the information stream and uses it to correct the errors made during digital demodulator's recovery of the data.

1.1.1.5 Information Destination

The functionality of the information destination block is opposite to that of the information source. In the case of digital systems, an expander is used to reverse the operation of the compressor as shown in Figure 1.5.

Since a communication system is costly to install, its commercial viability is critically dependent on the number of users who must share the transmission medium. As a result, the

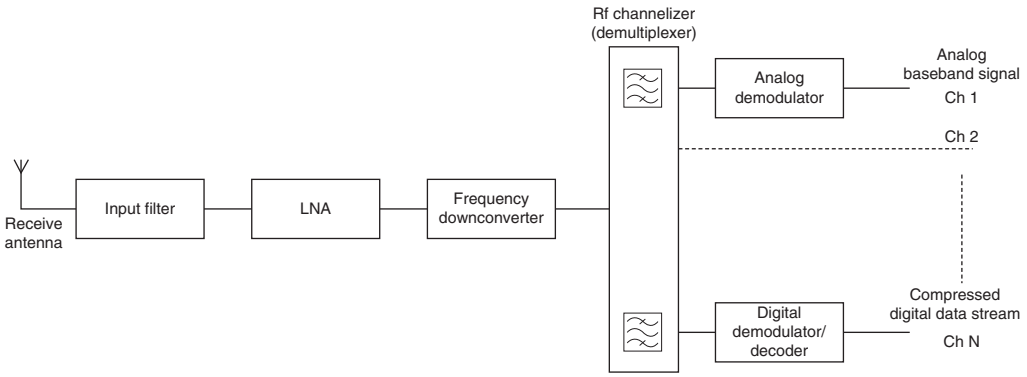


Figure 1.4 Receiver block diagram.

Figure 1.5 Information sink for a digital system.

information source usually consists of a large number of signals, occupying a finite range of frequencies. The width of the range of frequencies is called the *bandwidth* of the system.

Some key questions then arise: Is there a limitation on the available bandwidth for a communication system? What are the limitations of transmitting information over the chosen transmission media in the available bandwidth? What are the cost-sensitive parameters in a communications system?

1.2 Radio Spectrum and its Utilization

To understand the limitations on the available bandwidths for a communication system, it is necessary to understand the radio spectrum and its utilization [4].

EM waves cover an extremely broad spectrum of frequencies, from a few cycles per second to γ rays with frequencies of up to 10^{23} cycles per second. The radio spectrum is that portion of the EM spectrum which can be electronically and effectively radiated from one point in space and received at another. This includes frequencies anywhere from 9 kHz to 400 GHz. Although most of the commercial use takes place between 100 kHz and 44 GHz, some experimental systems reach as high as 100 GHz. The signals employing these same frequencies can also be transmitted over long distances by wire, coaxial cables, and glass fibers. However, since such signals are not intended to be radiated, they are not considered as part of the radio spectrum. A communication system is allowed to access only a portion of the radio spectrum, due to not only technology limitations but also regulatory reasons. The radio spectrum is subdivided into smaller frequency "bands" by national and international agencies, where each band is restricted to a limited set of types of operation. In addition, each band is a controlled commodity that often has a license fee associated with its use. Obviously, this represents a huge incentive to make the most efficient use of the allocated frequency spectrum.

1.2.1 Radio Propagation at Microwave Frequencies

There are many sources of energy loss in the free-space communication medium. The most serious ones include rainfall and presence of oxygen in the atmosphere. The atmospheric losses as a function of frequency are presented in Figure 1.6. The radio energy is absorbed and scattered by the raindrops, and this effect becomes more intense as the wavelength approaches the size of the raindrops. Consequently, rainfall and water vapor produce intense attenuation effects at higher microwave frequencies. The first absorption band, due to water vapor, peaks at approximately 22 GHz, and the first absorption band, due to the presence of oxygen in the atmosphere, peaks at about 60 GHz.

For fixed line of sight (LOS) terrestrial microwave radio links, multipath fading is another major cause of signal loss. Fading results from the variations in the refractive index of air for the first few tenths of a kilometer above the earth's surface. Such gradients in refraction bend the rays, which, on reflection from the ground or other layers, combine with the direct rays, causing coherent interference.

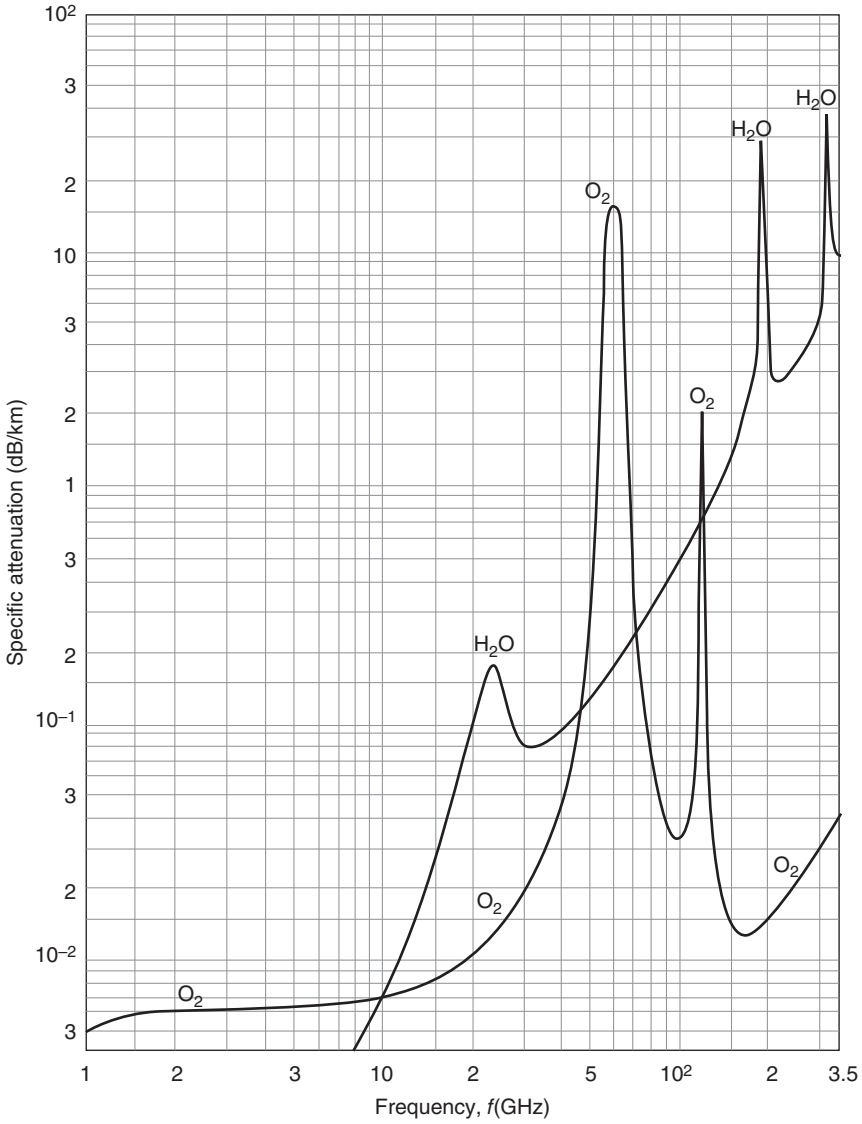


Figure 1.6 Atmospheric losses as a function of frequency. (*Source*: Freeman 2007 [7]. Reproduced with the permission of John Wiley and Sons. CCIR Report., 1990. Reproduced with permission International Radio Consultative Committee.)

Mobile communication adds a new dimension to the propagation problem. Besides the requirements of omnidirectional coverage and the mobility of end users, the communication system must deal with non-LOS and multipath problems caused by the signal reflections from tall buildings, trees, valleys, and other large objects in an urban environment. In addition, the coverage area for mobile services includes the area inside buildings. Also, mobility incurs the *Doppler shift*, a change in the frequency of the received signal, further complicating the issue.

From the foregoing analysis, it is evident that the attenuation due to rain and other atmospheric effects, coupled with multipath fading in an urban environment, severely constrains the available frequency spectrum suitable for commercial communications.

1.2.2 Radio Spectrum as a Natural Resource

The radio spectrum is a natural resource unlike any other communication system. It is a completely renewable resource that can never be permanently depleted. It is also universally available. The limitation comes with its usage. The radio spectrum has a finite capacity that, if exceeded, results in interference, incapacitating the system. For this reason, national governments grant users the privilege of using radio spectra in exchange for agreements to abide by usage rules. Because RF signals often cross national borders where they can interfere with radio frequencies allocated in another country, nations must cooperate to find ways to coordinate their individual allocations. Since communication is vital to all nations, there are national and international agencies that regulate the allocation and usage of the RF spectrum.

The International Telecommunications Union (ITU), part of the United Nations, is the international body that determines the worldwide radio spectrum allocations. The ITU accomplishes this through World Administrative Radio Conferences (WARC), where ITU member nations attempt to reach a consensus on proposals by different countries. These meetings require a consensual approach to decision-making, rendering the process very tedious, and often resulting in delays. Once a consensus is reached, the ITU publishes tables of the frequency allocations, deemed as the *radio regulations* of the ITU. Each country then makes its own detailed frequency allocation plan consistent with the radio regulation tables. In addition, there are other consultative committees such as the International Radio Consultative Committee (CCIR), a part of the ITU assigned to study and recommend the standards for interoperability and guidelines and the control of interference from various services. On the domestic front, most countries have government agencies such as the Federal Communications Commission (FCC) of the United States that regulate all the nonfederal government use of the frequency spectrum. There are other agencies that control frequency allocations for government use, including the military.

In most countries, the priority followed by the regulators for allocating radio spectrum is as follows:

- Military
- Public safety functions such as aeronautical and marine emergency communications, police, fire, and other emergency services
- National telecommunication companies for telephone
- Broadcast radio and television
- Private users such as mobile systems and other services.

Because of the complex web of procedures, priorities, government policies, and so on, once a service has been established and uses a particular portion of the spectrum, it is seldom changed. Incumbency tends to be a big advantage, and often an obstacle in the efficient use of the radio spectrum. The emergence of widespread wireless communications and services has added enormous pressure in both domestic and international regulatory bodies, as well as in original equipment manufacturers (OEMs) to ensure that the frequency allocations and spectrum usage represent the most efficient use of this natural resource.

1.3 Concept of Information

What are the limitations for transmitting information over the chosen transmission media? At the fundamental level, the answer lies in the seminal work of Shannon [9], who developed the concepts of information theory.

In 1948, Claude Shannon from the Bell Labs pointed out that "for a signal to carry information, the signal must be changing; and to convey information, the signal must resolve uncertainty."

Thus, the measure of the amount of information is a probabilistic one. Shannon defined the uncertainty of the outcome between two equally likely message probabilities as a unit of information; thereby, he used, as his measure, the binary digit or bit. Also, he showed that the information capacity of a system is fundamentally limited by a very few parameters. Specifically, he demonstrated that the maximum information capacity C of a channel is limited by the channel bandwidth B and the signal-to-noise ratio (S/N) in the channel as described by the relationship

$$C = B \log_M \left(1 + \frac{S}{N}\right) \quad \text{information messages per second} \tag{1.1}$$

where M is the number of possible source information message states. Here, S/N defines the conditions at the demodulator input, where S is the signal power in watts and N is white Gaussian noise with a mean power of N watts distributed uniformly over the channel bandwidth at the same location. In analog communication, M is difficult to define; however, in digital communication, the binary data ($M = 2$) is considered and the capacity limit is given by

$$C = B \log_2 \left(1 + \frac{S}{N}\right) \quad \text{information bits per second} \tag{1.2}$$

For digital communication, this capacity is defined in the context of information bits, measured at the output of the compressor function before the encoding function, and not at the data source as might be expected; B and S/N are defined within the channel at the input of the receiver section.

Note that the information is not defined at the data source, since there is a significant distinction between "data" and "information":

Data are the raw output from the source, for example, a digitized voice, digitized video, or a sequence of text characters, defining a text document.

Information is the salient content in the raw data. This content is often much less than the data used to represent the raw source output.

For example, consider a voice digitizer. It must continually sample the voice analog source and produce output, even when a person pauses between the words or sentences. Similarly, in a video signal, much of the picture does not change from frame to frame. In a text document, some characters or words are repeated more often than others, but the same number of bits is used to represent each character or word at the data source.

Digital compression exploits the knowledge about the characteristics of the data source type to map the data from one format of representation into another format that reduces the number of digital bits required to provide the salient information. The details of this are not addressed in this book. However, it is important to realize that there can be substantial reductions of raw data prior to their transmission with minimal or no loss of the useful information. For example, the current generation digital TV uses MPEG2 compression that provides at least one order of magnitude reduction in the average data rate required to represent the source data compared with the digitization of the raw video. Digital compression has certainly attracted the interest of the research community over the last few decades, resulting in continual improvement.

Shannon's information theorem proves that as long as the information transmission rate R is less than C, it is possible to limit the error in transmission to an arbitrarily small value. The technique for approaching this limit in digital communication is called *coding*, another favored area of research over the last few decades. The current technology comes within a few

tenths of a decibel of the Shannon limit. The theorem states that for $R < C$, transmission can be accomplished without error in the presence of noise. This is a surprising result in the presence of Gaussian noise, since its probability density extends to infinity.

This theorem, although restricted to the Gaussian channel, is fundamentally important for two reasons: (1) the channels encountered in physical systems are generally Gaussian and (2) the results obtained for the Gaussian channel often provide a *lower bound* on the performance of a system, indicating that it has the highest probability of error. Thus, if a particular encoder–decoder is used with a Gaussian channel and an error probability P_e results, then, with a non-Gaussian channel, another encoder–decoder can be designed so that P_e is smaller. Similar channel capacity equations have been derived for a number of non-Gaussian channels.

Shannon's theorem indicates that a noiseless Gaussian channel ($S/N = \infty$) has an infinite capacity, irrespective of the bandwidth, but the channel capacity does not become infinite as the bandwidth becomes infinite. This is due to the increase in the noise power as the bandwidth is increased. As a result, for a fixed signal power and in the presence of white Gaussian noise, the channel capacity approaches the upper limit with the increasing bandwidth. By using $N = \eta B$ in equation (1.2), where η is the noise density (W/Hz), we obtain

$$C = B\log_2\left(1 + \frac{S}{\eta B}\right) = \frac{S}{\eta}\log_2\left[1 + \frac{S}{\eta B}\right]^{\eta B/S} \tag{1.3}$$

and

$$\lim_{B\to\infty} C \approx \frac{S}{\eta}\log_2 e = 1.44\frac{S}{\eta} \tag{1.4}$$

It is evident that there is also an absolute lower bound on the received signal power in a system for a given capacity, irrespective of the bandwidth utilized. For a given capacity

$$S \geq \frac{C\eta}{1.44} \tag{1.5}$$

In accordance with equation (1.2), once minimum received signal power is exceeded, the bandwidth is traded off with the S/N ratio and vice versa. For example, if $S/N = 7$ and $B = 4$ kHz, $C = 12 \times 10^3$ bps. If the signal-to-noise ratio (SNR) is increased to $S/N = 15$ and B is decreased to 3 kHz, the channel capacity remains the same. With a 3 kHz bandwidth, the noise power is three-fourth as large as with a 4 kHz bandwidth. As a result, the signal power must be increased by the factor $\frac{3}{4} \times \frac{15}{7} = 1.6$. Therefore, this 25% reduction in bandwidth requires a 60% increase in signal power.

Once the basic parameters of radio spectrum and channel bandwidths have been established, the next question is: *What are the cost-sensitive parameters in the communication system?* This leads to an evaluation of the communication channel and system link parameters.

1.4 Communication Channel and Link Budgets

Communication capacity of a channel is bounded by signal power S, bandwidth B, and noise N in the channel. The parameters and the resultant trade-offs are investigated in the subsequent sections.

1.4.1 Signal Power in a Communication Link

Wireless communication systems require transmission of signals as radio frequencies in space. RF signals travel through earth's atmosphere and are collected by the receiver. The direction of

the propagation can vary from a horizontal direction for terrestrial fixed and mobile systems to a vertical direction for satellite systems. The received signal power is primarily a function of four elements of the communication system:

- The transmitter output electrical power at the RF
- The fraction of the transmitter electrical power directed at the receiver as an RF free-space wave, defined by the transmit antenna gain
- The loss of energy in the communication medium, including the loss due to spherical spreading of the energy
- The fraction of the received free-space RF power at the receiver converted into electrical energy, defined by the receive antenna gain.

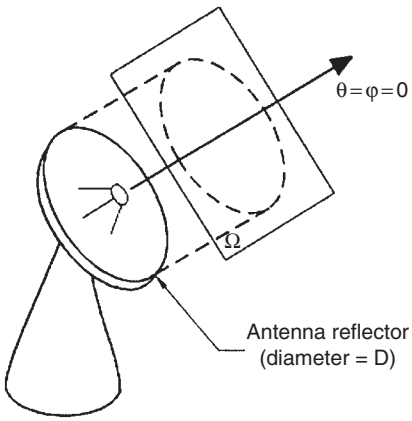
1.4.1.1 RF Power

For a given allocation of bandwidth and a certain level of noise, the communication capacity is governed by the available power for radio transmission. At first glance, it appears that the capacity can be arbitrarily increased by increasing the level of the RF power. There are two problems with this approach: (1) the generation of RF power is expensive and represents a cost-sensitive parameter and (2) a substantial amount of the RF power is wasted in the process of radio transmission, rendering the power the most costly portion of a wireless system. It behooves us to examine the basic limitations of power transmission in a wireless system.

1.4.2 Transmit and Receive Antennas

Antennas are used to launch EM energy from a transmission line into space and vice versa. Antennas are linear, reciprocal elements, and, as a consequence, the properties of an antenna are the same in transmission and in reception. The reciprocity theorem holds for all the antenna characteristics. A physical antenna (reflector type) consists of a reflecting surface and a radiating or absorbing feed network. The reflector is used to focus the energy in a given direction. The feed element converts electric currents into EM waves in a transmitting system or converts EM waves into electric currents in a receiving system. A typical parabolic microwave antenna is depicted in Figure 1.7. A transmit antenna focuses energy in the direction of a receive station or over a specific geographic area. For a receive antenna, the aperture collects the power and focuses it on the input feed element of the receive system. The general properties of the antennas described here can be applied to a broad range of frequency spectra.

Figure 1.7 Parabolic antenna.



1.4.2.1 Antenna Gain

The gain of an antenna is defined with respect to an isotropic antenna. An isotropic transmitting antenna is equivalent to a point source, radiating uniform spherical waves equally in all directions as denoted in Figure 1.8a. The power p_0 in watts, radiated from such a source has a uniform power flux density (pfd) of $(p_0/4\pi r^2)$ watts per square meter at distance r from the source. Assume that the power source is located at the input terminal of an antenna. The power radiated from this source is proportional to $p_0/4\pi$ in any direction (θ, ϕ) of the surrounding space. A directional antenna radiates power $p(\theta, \phi)$ in direction θ, ϕ as shown in Figure 1.8b. The gain of the directional antenna, with respect to an isotropic source, is then given by

$$g(\theta, \phi) = \frac{p(\theta, \phi)}{(p_0/4\pi)} \tag{1.6a}$$

The gain varies, depending on the specific values θ and ϕ. The maximum value of the gain, as determined by the maximum value of power within the (θ, ϕ) envelope, is simply defined as the antenna gain g. It is usually expressed in decibels:

$$G = 10 \log g \quad \text{dBi or simply dB} \tag{1.6b}$$

Note that this definition is independent of the physical attributes of the antenna and depends solely on the geometry of the radiation patterns of the antennas. For most applications employing uniform parabolic antennas, this maximum gain occurs along the boresight of the antenna, where $\theta = \phi = 0$.

1.4.2.2 Effective Aperture and Antenna Gain

Physical antennas are designed to radiate and capture energy in the desired directions with a minimum of loss and spillage of energy outside the region of interest. Effective aperture A_e of an antenna is defined as the equivalent physical area of the antenna that captures or radiates energy within the desired direction of θ and ϕ. It is defined by

$$A_e = \eta A \tag{1.7}$$

where $\eta(< 1)$ is the efficiency of the antenna and A is the physical aperture used to radiate or capture energy. The effective aperture represents the projected area in the direction of the beam

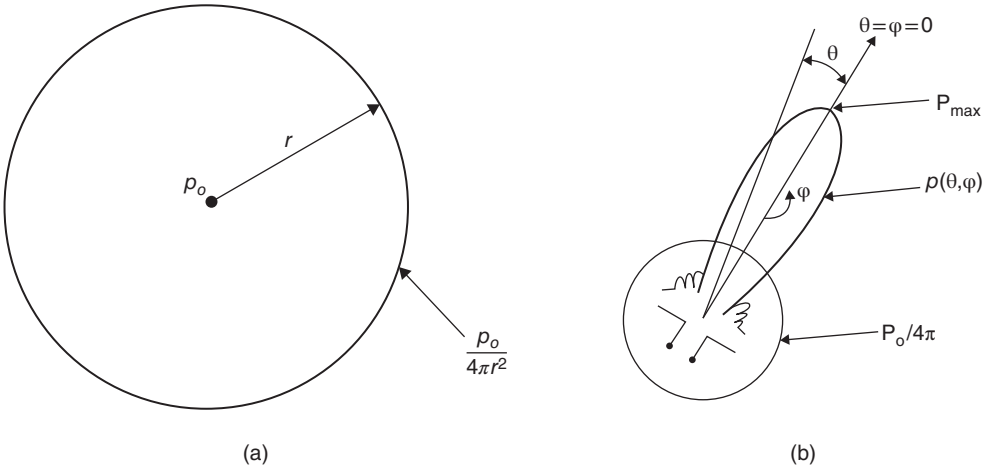


Figure 1.8 Radiated power of a transmit antenna: (a) uniform spherical waves from an isotropic source; (b) a directional antenna beam.

to achieve the maximum gain and includes the degradation due to the losses and nonuniformities of the structure and the nonuniformity in the illumination of the aperture. If an antenna is perfect and the energy propagation is uniform, A_e equals the actual projected area A, and η is unity. For practical applications, η varies between 0.5 and 0.8.

The next question is how to determine the amount of power that an effective aperture is able to focus in a given frequency band. This is a complex problem, but the result is simple and elegant. From the theory of antennas [10], the gain of an antenna is related to the effective aperture by

$$g = \frac{4\pi A_e}{\lambda^2} \tag{1.8}$$

where λ is the wavelength of the RF. This demonstrates that the gain of an antenna depends on its effective aperture and operating frequency. The higher the frequency, the higher the gain of the antenna for a given aperture size. This relationship indicates that there are inherent limitations in terms of the achievable gain for a given size of the aperture. Moreover, the surface of the aperture needs to be accurate to within small fractions of λ, making higher-frequency structures more expensive. Another useful relationship can be derived by considering the solid angle Ω within which an antenna focuses the power. This can be readily derived from equation (1.8) as

$$\Omega = \frac{\lambda^2}{A_e} \text{ rad}^2 \tag{1.9}$$

and

$$g = \frac{4\pi}{\Omega}$$

For a satellite-based system, once the antenna footprint is specified (i.e., the coverage area is delineated), the solid angle to cover that area from the satellite gets fixed. This implies that the antenna gain (and therefore the aperture size) is determined by the coverage area. In other words, the limit on the achievable gain from the satellite is set by the coverage area and not by the design or physical structure of the antenna. This relationship also shows that in order to achieve a high directivity, that is, a smaller solid angle, the area of the aperture must be much larger than the operating wavelength. In the microwave region, the physical apertures required to achieve relatively high gains are reasonable and readily implemented. For such antennas, the radiated fields do not interact strongly with nearby objects that are not in the direction of the transmission. Thus, the ground effects, which play a major role in the design of megahertz region (AM, FM, and TV broadcasts) antennas, are generally not important above 1 GHz.

1.4.2.3 Power Flux Density

The radiation of EM energy follows the fundamental inverse square law. Consequently, energy, radiated by the isotropic antenna to a distance r, is uniformly distributed over a sphere of area $4\pi r^2$, as shown in Figure 1.8a. Note that the energy is independent of the frequency of the radiation. The power received per unit area is simply $p_0/4\pi r^2$, where p_0 is the amount of power at the input of the isotropic source. This value is defined as pfd:

$$\text{pfd} = \frac{p_0}{4\pi r^2} \text{ W/m}^2 \tag{1.10}$$

1.4.2.4 Power Radiated and Received by Antennas in a Communication System

Consider the power transmitted and received via radio transmission in an ideal communication channel. Let

p_T be the power of the transmitter in watts

g_T be the transmit antenna gain

g_R be the receive antenna gain

A_{eT}, A_{eR} be the effective apertures of the transmit and receive antennas in meters

r be the distance between the transmitter and receiver in meters

Then power transmitted $= p_T \times g_T$ W. The pfd in the desired direction at distance r is given by

$$\text{pfd} = \frac{p_T g_T}{4\pi r^2} \text{ W/m}^2 \tag{1.11a}$$

The product $p_T g_T$ is called the *equivalent isotropic radiated power* (EIRP), an important factor in link calculations. The pfd represents the received power that is intercepted by an ideal antenna with an aperture of 1 m^2, expressed in decibels

$$\begin{aligned}\text{PFD} &= P_T + G_T - 10\log 4\pi r^2 \\ &= \text{EIRP} - 10\log 4\pi r^2 \text{ dBW/m}^2\end{aligned} \tag{1.11b}$$

where $P_T = 10\log p_T$, $G_T = 10\log g_T$, and $\text{PFD} = 10\log(\text{pfd})$. This equation relates that no matter what, in a radio transmission, a very large portion of the transmitted energy, represented by the term $10\log 4\pi r^2$, is not seen by the receiver. Also, this loss is independent of the frequency, a point often lost if the equation is manipulated in terms of the gains of the transmit and receive antennas. To clarify this point, equation (1.11) can be expressed in terms of power p_r received at distance r as follows:

$$\begin{aligned}p_r &= (\text{pfd})_t \times A_{eR} = \frac{p_T g_T}{4\pi r^2} g_R \frac{\lambda^2}{4\pi} \\ &= p_T g_T g_R \left(\frac{\lambda}{4\pi r}\right)^2 \text{ W}\end{aligned} \tag{1.12a}$$

In decibels, this is

$$P_R = P_T + G_T + G_R - 20\log\frac{4\pi r}{\lambda} \text{ dBW} \tag{1.12b}$$

This equation indicates that the power received has frequency and distance dependent terms. The combined dependence arises because of the receive antenna, required to focus the received energy to a point source. In other words, the term $20\log(4\pi r/\lambda)$, often referred to as the *range loss*, represents the power loss between two isotropic antennas within a particular range and at a particular frequency. As far as spreading of the radio energy from a point source is concerned, the energy follows the classic inverse square law, independent of the frequency, described by equation (1.11). Regardless, either equation (1.11) or (1.12) can be used for link calculations.

Example 1.1

In a microwave radio relay link, repeater stations are located 50 km apart. They are equipped with 10 W HPAs and antennas with a gain of 30 dB. Assuming the transmission line and filter losses to be 2 dB, compute (1) the EIRP of the transmitter and the pfd at the receive antenna and (2) given an antenna efficiency of 0.55, calculate the diameter of a circular parabolic antenna to realize the 30 dB gain in the 4 GHz frequency band.

Solution:

1) EIRP $= 10\ \text{dBW} + (-2\ \text{dB}) + 30\ \text{dB}$
   $= 38\ \text{dBW}$
   PFD $= \text{EIRP} - 10\log 4\pi r^2$
   $= 38 - 105$
   $= -67\ \text{dBW/m}^2\ (\text{or}\ 200\ \text{nW/m}^2)$
   Thus, the aperture of a square meter area, located 50 km apart from the transmitter, intercepts 200 nW of RF power, radiated by the transmit antenna.
2) The aperture to realize a specific antenna gain is derived from

$$A = \frac{G}{\eta}\frac{\lambda^2}{4\pi}\ \text{m}^2$$

   The power received by the aperture depends on the RF even though the pfd remains unchanged. For a circular parabolic antenna with a diameter D, the aperture area A equals $(\pi/4) \times D^2$. By expressing the antenna gain in decibels, and the diameter in meters, we obtain

$$G = 20\log D + 20\log f_{\text{MHz}} + 10\log\eta - 39.6\ (\text{dB})$$

For $f = 4000$ MHz, $G = 30$ dB, and $\eta = 0.55$, D is calculated to be 1 m.

PART II Noise in a Communication Channel

1.5 Noise in Communication Systems

Noise in its broadest definition consists of any undesired signal in a communication circuit. Noise represents the fundamental constraint in the transmission capacity of a communication system. It is also of interest to the ITU and domestic radio regulation agencies. An essential element of radio regulations is the specification of allowable levels of radiation to other existing or proposed systems. Such a restriction to control the interference between communication systems is essential in a multisystem environment. Without the restriction, it is not possible to design a reliable communication system. Typically, the regulating agencies specify the allowable level of radiation outside the intended geographic area and allocated frequency spectrum. Also, the regulations include the interference guidelines of competing services and systems. The noise sources external to a communication system are as follows:

1) Cosmic radiation, including that from the sun
2) Anthropogenic (synthetic, caused by humans) noise such as power lines, electrical machinery, consumer electronics, and other terrestrial sources
3) Interference from other communications systems.

There is little to be done about the cosmic noise except to ensure that it is considered in the system design. Generally, anthropogenic noise occurs at low frequencies and is rarely a problem for communication systems operating above 1 GHz. Interference from other communication systems is strictly controlled by ITU and domestic agencies. Typically, this requires control of transmitter powers, antenna patterns, generation, and suppression of frequencies outside the

allocated spectrum. The regulations ensure that these sources of unwanted radiation are kept to a minimum and significantly below the levels of noise inherent in the design of a communication system. The dominant sources of noise in a communication system include the following:

- Interference from adjacent copolarized channels
- Interference from adjacent cross-polarized channels
- Multipath interference
- Thermal noise
- Intermodulation (IM) noise
- Noise due to channel imperfections.

These noise sources are now reviewed.

1.5.1 Adjacent Copolarized Channel Interference

One thing that is common to all communication systems is the channelization of the frequency spectrum. Channelization allows flexibility in meeting multidestination traffic requirements, as well as in maximizing the communication capacity of the system. In a subsequent section, modulation schemes employed for long-haul transmission are described. The most commonly used modulation scheme is frequency modulation (FM). One property of FM is that it produces sidebands of decreasing magnitudes extending to infinity. Consequently, there is no escape from some energy spilling over in the adjacent bands causing distortion. To a degree, this spillage can be controlled by the choice of channelization filters, highlighting the criticality of filter networks in a communication system. Figure 1.9 is a pictorial representation of adjacent channel interference.

1.5.2 Adjacent Cross-Polarized Channel Interference

The nature of EM radiation allows the polarization of energy in a given direction. This property is exploited in communication systems by employing orthogonal polarization of antenna beams. It allows frequency reuse that doubles the available bandwidth. Polarization can be linear or circular. The primary limitation is the level of cross-polarization isolation that is achievable in practical antenna networks. Therefore, this interference is totally dependent on and controlled by the antenna design. Typical cross-polarization isolation in practical systems is of the order 27–30 dB. It should be noted that polarization can be altered when EM radiation propagates through the atmosphere. This also needs to be accounted for in the design.

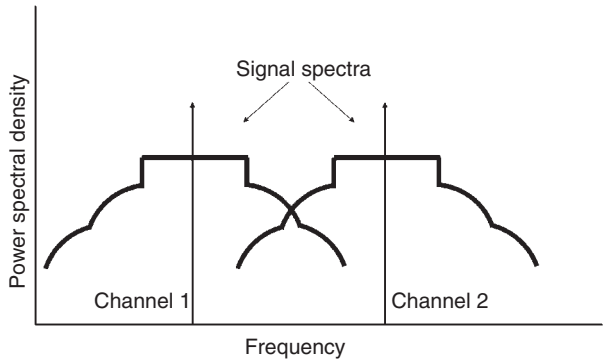


Figure 1.9 Adjacent channel interference.

1.5.3 Multipath Interference

This distortion mechanism is caused by obstacles in the spread energy field of the transmitter where the energy is reflected in the direction of the receiver. Such obstacles can be tall buildings or trees or foliage in an urban environment. Also, distortion can stem from refractive index gradients and other atmospheric effects. Interference occurs when the reflected rays from obstacles or ground are received by the receiver. These interfering signals are time-displaced echos of the original transmission. Figure 1.10 is a pictorial description of multipath interference. In a fixed LOS environment, the radio paths are optimized by taking such obstructions into account. This is not the case when it comes to mobile communication systems. The mobile terminal can be fixed or in motion. Even a handheld terminal might have micromotion. When a terminal is in motion, the path characteristics are constantly changing and multipath propagation tends to become the limiting factor. Mobile units experience multipath scattering, reflection, and diffraction by various obstructions and buildings in the vicinity. Although constructive and destructive fading can become quite complex, there are ways to deal with this problem, including frequency and space diversity and forward error correction. Regardless of these compensation techniques, multipath interference continues to be the principal cause of missed calls, fading, and disruptions in mobile communications. The limitation of the available bandwidth for this service further exacerbates the problem. This topic and other issues related to radio propagation are described by Freeman [7].

1.5.4 Thermal Noise

Thermal noise permeates all communication systems and is the ultimate limit to their performance. For this reason, thermal noise is addressed in more depth. Thermal noise is the name given to the electrical noise arising from the constant agitation of molecules and its constituents in a conductor. This agitation at the atomic level is a universal characteristic of all matter. Molecules consist of a nucleus and a cloud of electrons around it. The nucleus consists of neutrons and protons, and the number of protons is equal to the number of electrons. Electrons

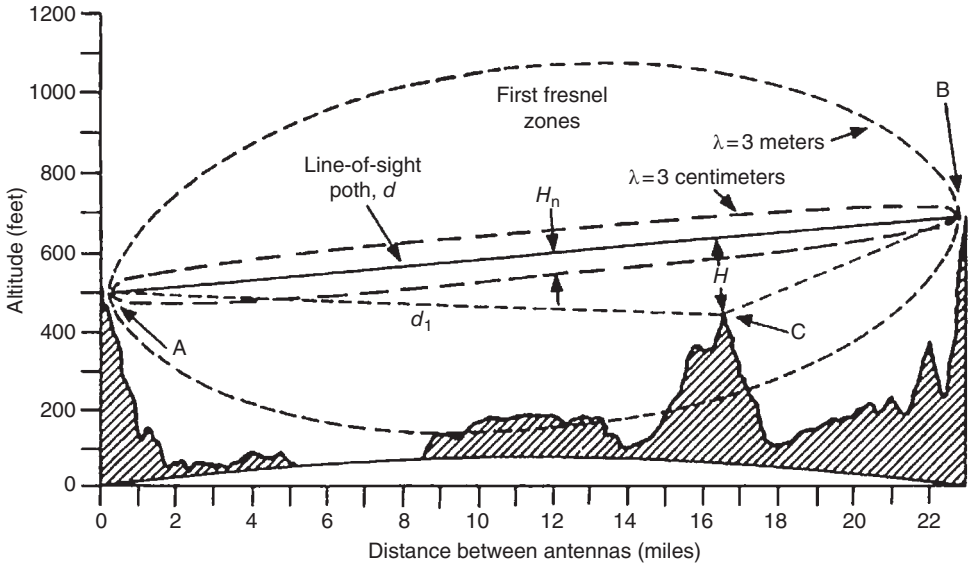


Figure 1.10 Multipath interference. (*Source*: Reproduced with the permission of Bell Telephone Laboratories, Inc.)

and positive ions (atoms with one or more electrons removed) are, on an average, distributed uniformly in a conductor, rendering the structure electrically neutral. The electrons in a conductor are in continual random motion, and in thermal equilibrium with the molecules. There is kinetic energy associated with this motion, and it increases with temperature. Since each electron carries a unit negative electric charge, each flight of an electron between collisions with molecules constitutes a short pulse of current. The random agitation of the large number of electrons creates statistical fluctuations from the neutrality that translates into electrical noise. The mean square velocity of electrons is directly proportional to the absolute temperature. The equipartition law due to Boltzmann and Maxwell (and the works of Johnson and Nyquist) states that for a thermal noise source, the available power in a 1 Hz bandwidth is given by

$$p_n(f) = kT \text{ W/Hz} \tag{1.13}$$

where k is the Boltzmann constant = 1.3805×10^{-23} J/K and T is the absolute temperature of the thermal noise source in degrees Kelvin. At a room temperature of 17°C or 290 K, the available power is $p_n(f) = 4.0(10^{-21})$ W/Hz or −174.0 dBm/Hz. The result given by the equipartition theory is one of a constant power density spectrum versus frequency. Because of this property, a thermal noise source is referred to as a *white noise source*, an analogy to white light that contains all the visible wavelengths of light. In all the reported measurements, the available power of a thermal noise source has been found to be proportional to the bandwidth over any range, from direct current to the highest microwave frequencies. If the bandwidths are unlimited, the results of the equipartition theory states that the available power of a thermal noise source should also be unlimited. Obviously, this is not possible. The reason is that the equipartition theory is based on classical mechanics, which tends to break down as the higher frequencies are approached. If the principles of quantum mechanics are applied to the problem, kT must be replaced by $hf/[\exp(hf/kT) - 1]$, where h is Planck's constant = 6.626×10^{-34} J s. By applying this result to the expression for available power of a thermal noise source [1], we obtain

$$p_n(f) = \frac{hf}{\exp(hf/kT) - 1} \text{ W/Hz} \tag{1.14}$$

This relationship demonstrates that at arbitrarily high frequencies, the thermal noise spectrum eventually drops to zero, but this does not mean that noiseless devices can be built at these frequencies. A quantum noise term equal to hf needs to be added to equation (1.14) in this case. Figure 1.11 offers a plot of the noise density as a function of frequency. The transitional region occurs at about 40 GHz for $T = 3$ K, at 400 GHz for $T = 30$ K, and at 4000 GHz at room temperature. For most practical purposes, the available noise power of a thermal noise source can be assumed to be directly proportional to the product of the bandwidth of the system or detector and the absolute temperature of the source.

Thus

$$p_a = kTB \text{ W} \tag{1.15a}$$

where B is the noise bandwidth of the system or detector in hertz and p_a is the available noise power in watts. By assuming an ambient temperature of 290 K and expressing the available noise power in dBm gives

$$p_a = -174 + 10 \log B \text{ dBm} \tag{1.15b}$$

This represents the minimum amount of noise power that must ultimately limit the SNR of a signal. This relationship represents an average value; it does not tell us anything about the statistical distribution. As stated earlier, thermal noise is attributed to the random motions of electrons in the conductors. Therefore, thermal noise might be regarded as the superposition

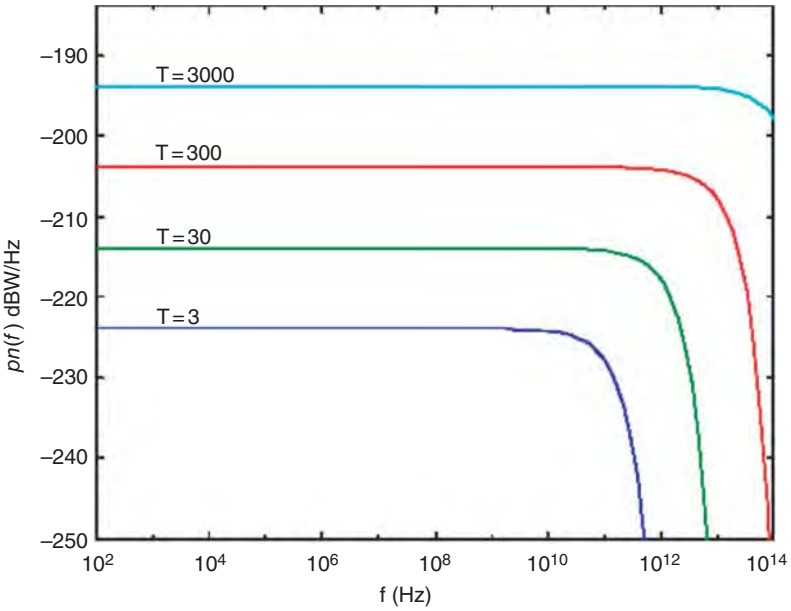


Figure 1.11 Thermal noise power densities as a function of frequency.

of an exceedingly large number of independent electronic contributions. It is well known in the field of statistics that the limiting form for the distribution function of the sum of a large number of independent quantities that can have various distributions is a Gaussian function. This result is known in statistics as the *central limit theorem*. Consequently, thermal noise satisfies the theoretical conditions for a Gaussian function and must follow this distribution. The Gaussian probability density function for the zero mean is reflected in Figure 1.12a and its equation is

$$p(V) = \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left(\frac{-V^2}{2\sigma_n^2}\right) \tag{1.16}$$

where V represents the instantaneous voltage and σ_n the standard deviation. The Gaussian distribution function is shown in Figure 1.12b and is given by the integral of equation (1.16)

$$P(V) = \frac{1}{\sigma_n \sqrt{2\pi}} \int_{-\infty}^{V} \exp\left(\frac{-x^2}{2\sigma_n^2}\right) dx \tag{1.17}$$

It can be readily shown that the mean square voltage (the expected value of V^2) is equal to the variance, σ_n^2. Thus, the rms (root mean square) voltage of the Gaussian distributed noise source is given by σ_n, the standard deviation.

Gaussian noise has a probability greater than zero that exceeds any finite magnitude no matter how large it is. As a result, the peak factor, given by the ratio of peak to rms voltage, does not exist for a thermal noise signal. For this particular case, it is convenient to modify the definition of the peak factor to be the ratio of the value, exceeded by the noise a certain percentage of the time to the rms noise value. This percentage of time is commonly chosen to be 0.01%. A table of the normal distribution indicates that signal magnitudes greater than $3.89\sigma_n$ (i.e., $|V| > 3.89\sigma_n$) occur less than 0.01% of the time.

Since σ_n is the rms value of the noise signal, the peak factor for a thermal noise signal is 3.89, or 11.80 dB. Inclusion of 0.001% peaks increases the peak factor by 1.1 dB, raising it to a value of 12.9 dB. The fact that thermal noise is white, as well as Gaussian, has led many engineers into

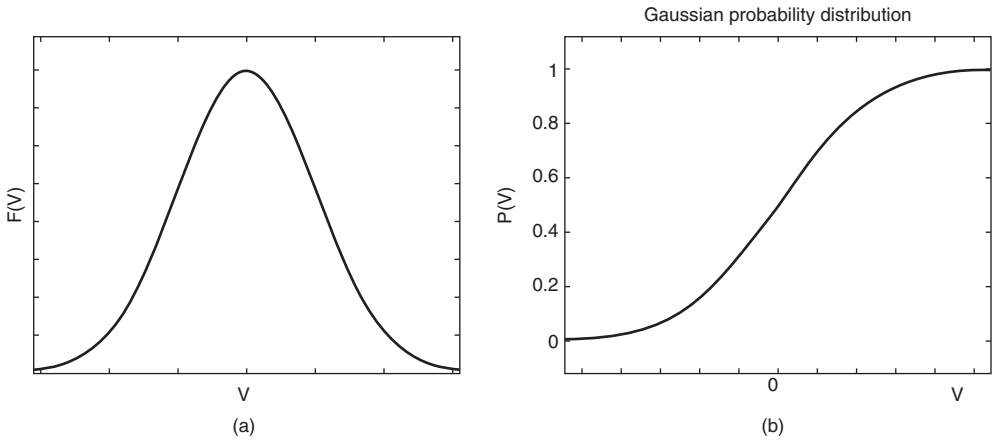


Figure 1.12 (a) Gaussian density and (b) distribution functions.

carelessly viewing white and Gaussian noise as synonymous, which is not always the case. For example, passing Gaussian noise through a linear network such as a filter will leave it Gaussian but can drastically change the frequency spectrum. A single impulse does not exhibit a Gaussian amplitude distribution but has a flat or white frequency spectrum.

1.5.4.1 Effective Input Noise Temperature

Since the available noise power of a thermal noise source is directly proportional to the absolute temperature of the source, an equivalent noise temperature can be attributed to it [1]. For resistive elements, the noise temperature is equal to the physical temperature of the resistor, that is, if a given noise source produces an available power of p_a watts in a small frequency interval of df hertz, the noise temperature of the noise source is given by $T = p_a/k\ df$. It should be emphasized that the concept of noise temperature does not have to be restricted to the noise sources alone and that the noise temperature does not have to equal the physical temperature of the source. For example, consider an antenna. The output noise is simply the noise collected by the aperture due to the radiating elements in the antenna's field of view. The physical temperature of the antenna has no bearing on this, and the noise temperature can still be used to define the noise power from the antenna.

Consider a two-port network with an available gain of $g_a(f)$. When it is connected to a noise source having a noise temperature of T, the available noise power in a small band df at the output of the network is $p_{\text{no}} = g_a(f)kT\ df + p_{\text{ne}}$. This power consists of two components: (1) power due to the external noise source, $g_a(f)kT\,df$ and (2) the power due to the internal noise sources of the network p_{ne}, which is the available noise power of the network when the input of the network is connected to a noise-free source. The effective noise temperature T_e of this equivalent representation of the internal noise source of the noisy network is then given by

$$T_e = \frac{p_{\text{ne}}}{g_a(f)k\ df} \tag{1.18}$$

The available noise power at the output of the network in terms of the effective input noise temperature now becomes

$$p_{\text{no}} = g_a(f)k(T + T_e)df \tag{1.19}$$

The effective input noise temperature T_e can vary as a function of the frequency, depending on how g_e and p_{ne} vary. The concept of the effective input noise temperature is useful when the source noise temperature differs from that of the standard temperature. It has a distinct advantage when the noise performance of a complete communication system is being evaluated. Another concept that is useful in analyzing active devices and communication systems is the noise figure.

1.5.4.2 Noise Figure

The IRE (Institute of Radio Engineers, precursor to IEEE) definition of the noise factor for a two-port network is as follows: "The noise figure (noise factor) at a specified input frequency is the ratio of the total noise power per unit bandwidth at a corresponding output frequency, available at the output when the noise temperature of the input source is standard (290 K); to that portion of this output power engendered at the input frequency by the input source." In terms of this definition

$$\text{Noise figure} = n_F = \frac{p_{\text{no}}}{g_a(f)kT_0df} \tag{1.20}$$

where $T_0 = 290\,\text{K}$ and is referred to as the standard temperature. A noise figure such as that described for a narrowband df is called a *spot noise figure*, which can vary as a function of frequency. The noise figure can also be related to the effective noise temperature. In equation (1.19), if T is replaced by T_0, as required in the definition of the noise figure, the output noise power p_{no} is then given by $g_a(f)k(T_0 + T_e)df$. The output noise in terms of the noise figure is given by equation (1.20). Equating these two expressions yields the relationship between the noise figure and the effective noise temperature as

$$n_F = 1 + \frac{T_e}{T_0} \tag{1.21}$$

and

$$T_e = T_0(n_F - 1) \tag{1.22}$$

The concept of the noise figure is most useful when the input source has a noise temperature approximately equal to the standard temperature. Consider the expression for the noise power at the output of the network as given by equation (1.20).

By rewriting this equation in terms of dBm (reference 1 mW), we have

$$P_{\text{no}} = N_F + G_a + 10\log df - 174\ \text{dBm} \tag{1.23}$$

The symbols are defined as follows:

$$P_{\text{no}} = \frac{10\log p_{\text{no}}}{10^{-3}}$$

$$N_F = 10\log n_F$$

$$G_a = 10\log g_a \tag{1.24}$$

Thus, the available noise power of the two-port network in dBm can be written as the sum of the noise power of the thermal noise source in dBm, the available gain of the network in dB, and the noise figure of the network in dB. As a result, the effects of internal noise sources of a two-port network are accounted for by adding the noise figure in dB to the available noise power of the source in dBm.

1.5.4.3 Noise of a Lossy Element

Any signal is attenuated by the lossy elements in its path. Lossy elements absorb energy, thereby raising the level of agitation of the molecules in the element resulting in additional noise. By adopting arguments similar to the preceding sections, the effective input noise temperature of a lossy element is derived as [1]

$$T_e = T(l_a - 1) \tag{1.25}$$

where l_a is the loss ratio of the element. The loss L_a of the element in dB is given by $L_a = 10 \log l_a$. The noise figure of the lossy element is then given by

$$n_F = 1 + \frac{T}{T_0}(l_a - 1) \tag{1.26}$$

If the lossy element is at the standard temperature T_0, then

$$n_F = l_a \quad \text{and} \quad N_F = L_a \text{ dB} \tag{1.27}$$

As an example, for a transmission line with a loss of 1 dB at room temperature, the effective input noise temperature is given by $T_e = 75$ K, and the noise figure is simply 1 dB.

1.5.4.4 Attenuators

Attenuators are used in communication systems to control power levels of different channels or transponders. Such networks can be constructed using lossy elements or reactive elements or a combination of both. If composed of ideal and entirely reactive elements (implies zero resistive component), an attenuator will not contribute any noise and will have an effective input noise temperature of zero. However, loss due to such attenuators must be included when dealing with a chain of network elements as described subsequently in equation (1.29). Practical reactive elements always have an associated resistive component, however small, and that contributes noise in the channel. The key point to remember is that it is the resistive loss associated with any device that contributes to system noise temperature.

Example 1.2

What is the amount of thermal noise power radiated by a source at a room temperature of 290 K in a bandwidth of 50 MHz? What are the peak values of this noise that do not exceed 0.01% and 0.001% of the time?

Solution:

Using equation (1.14), the thermal noise is given by

$$\begin{aligned} N_T &= kTB \\ &= -228.6 + 10 \log T + 10 \log B \quad \text{dBW} \end{aligned}$$

For $T = 290$ K and $B = 50 \times 10^6$ Hz, we have

$$\begin{aligned} N_T &= -127 \text{ dBW} \\ &= 0.2 \text{ pW} \end{aligned}$$

As a result, an antenna that is pointed at a 290 K source, where the source fills the antenna beamwidth, receives 0.2 pW of noise power in a bandwidth of 50 MHz. As described in Section 1.5.4, the peak factors for thermal noise not exceeding 0.01% and 0.001% of the time are 11.8 and 12.9 dB, respectively. As a consequence, the peak value of the noise can be as high as—115.2 dBW (3.0 pW) for 0.01% of the time and—114.1 dBW (3.9 pW) for 0.001% of the time.

Example 1.3
What is the noise power at the output of a LNA with a noise figure of 2 dB and a gain of 30 dB over a bandwidth of 500 MHz? What is the equivalent noise temperature of the LNA?

Solution:
Equations (1.20)–(1.22) lead to

$$\begin{aligned} N_{\text{LNA}} &= N_F + G_a + 10\log df - 174 \text{ dBm} \\ &= (2 \text{ dB}) + (30 \text{ dB}) + 10\log 500 \times 10^6 - 174 \\ &= -55 \text{ dBm} \end{aligned}$$

The equivalent noise temperature is given by

$$\begin{aligned} T_{\text{LNA}} &= T_0(n_F - 1) \\ &= 290(1.585 - 1) \\ &= 169.6 \text{ K} \end{aligned}$$

1.5.5 Noise in Cascaded Networks

The two networks connected in tandem in Figure 1.13 have effective input temperatures T_{e1} and T_{e2} and available gains g_1 and g_2.

Suppose that these two tandem networks are connected to a noise source with a noise temperature of T. In a small frequency band df, the noise power at the output due to the noise source alone is $g_1 g_2 kT\,df$. The noise power due to noise sources in the first network is $g_1 g_2 kT_{e1}\,df$ and in the second network is $g_2 kT_{e2}\,df$. The total noise appearing at the output of the second network is $kg_2(g_1 T + g_1 T_{e1} + T_{e2})df$. Since the portion of this noise due to the noise sources internal to the two networks is $kg_2(g_1 T_{e1} + T_{e2})\,df$, the effective input temperature T_e of the two networks in tandem is therefore given by

$$\begin{aligned} T_e &= \frac{kg_2(g_1 T_{e1} + T_{e2})df}{g_1 g_2 k\,df} \\ &= T_{e1} + \frac{T_{e1}}{g_1} \end{aligned} \tag{1.28}$$

This result can be easily generalized to n networks in tandem. The resulting effective input noise temperature is

$$(T_e)_1 = T_{e1} + \frac{T_{e2}}{g_1} + \frac{T_{e3}}{g_1 g_2} + \cdots + \frac{T_{en}}{g_1 g_2 \cdots g_{n-1}} \tag{1.29a}$$

It should be noted in equation (1.29a) that the reference point for the effective noise temperature is the input terminal of the first element in the chain. This relationship can be readily modified to compute the effective noise temperature with reference to any terminal within the chain. For example, if the chosen reference is the input port of element 3(T_{e3}), then the effective

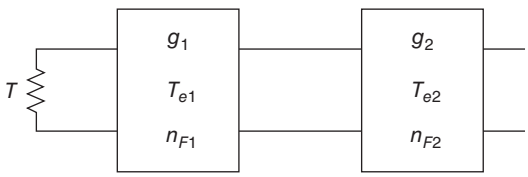


Figure 1.13 Noise in cascaded networks.

noise temperature is given by

$$(T_e)_3 = T_{e1}g_1g_2 + T_{e2}g_2 + T_{e3} + \frac{T_{e4}}{g_3} + \frac{T_{e5}}{g_3g_4} + \cdots \tag{1.29b}$$

Equation (1.29) is known as *Friis* formula, named in honor of H.T. Friis.

This relationship is advantageous in system calculations where the selected reference is typically the input terminal of the LNA. Such a reference provides the evaluation of the figure of merit defined by the ratio of the effective antenna gain to the noise temperature (G/T).

From the relationship between the noise figure and effective input noise temperature, it can be easily demonstrated that the resulting noise figure of n stages in tandem is

$$(n_F)_1 = n_{F1} + \frac{n_{F2} - 1}{g_1} + \cdots + \frac{n_{Fn} - 1}{g_1g_2 \cdots g_{n-1}} \tag{1.30}$$

The significance of these relationships lies in the fact that noise contribution after an amplifier in the chain is reduced by the gain of the amplifier. Typical amplifiers have gain in excess of 20 dB; this implies that the noise contribution by the elements after the amplifier in the chain will be reduced by a factor of 100. This is an important consideration in the design of a communication channel as well as multistage amplifiers. Example 1.4 illustrates the significance of these relationships.

Example 1.4

Calculate the noise temperature contributions in a 6 GHz, 500 MHz bandwidth receive section as denoted in Figure 1.14. Assume that the receive antenna has a noise temperature of 70 K.

Solution:

The noise temperatures and associated gain/loss ratios for the various elements are computed as follows:

Feed Network:

$$l_1 = 1.5849$$

$$g_1 = \frac{l}{l_1} = 0.631$$

$$T_{e1} = 290(l_1 - 1) = 169.6 \text{ K}$$

Bandpass Filter:

$$l_2 = 1.1885$$

$$g_2 = \frac{l}{l_2} = 0.8414$$

$$T_{e2} = 290(l_2 - 1) = 54.7 \text{ K}$$

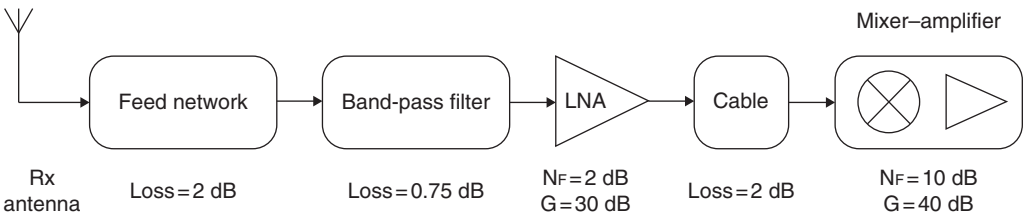


Figure 1.14 Example of noise calculations of a receive network.

LNA:

$$n_{F3} = 1.5849$$

$$g_3 = 1000$$

$$T_{e3} = 290(n_{F3} - 1) = 169.6 \text{ K}$$

Cable:

$$l_4 = 1.5849$$

$$g_4 = \frac{1}{14} = 0.631$$

$$T_{e4} = 290(l_4 - 1) = 169.6 \text{ K}$$

Mixer Amplifier:

$$n_{F5} = 10$$

$$g_5 = 10{,}000$$

$$T_{e5} = 290(n_{F5} - 1) = 2610 \text{ K}$$

The total system noise temperature referred to the input of the LNA is

$$\begin{aligned}(T_e)_{\text{sys}} &= (T_{\text{ant}} + T_{e1})g_1g_2 + T_{e2}g_2 + T_{e3} + \frac{T_{e4}}{g_3} + \frac{T_{e5}}{g_3g_4} \\ &= 127.2 + 46.0 + 169.6 + 0.17 + 4.14 \\ &= 347.1 \text{ K}\end{aligned}$$

It is interesting to note the relatively large contribution to the overall system noise temperature by the losses prior to the LNA and the relatively insignificant noise contributions by components after the LNA. These are important considerations in the design of communication systems, emphasizing the need to minimize losses prior to the LNAs in the receive section of a communication channel.

1.5.6 Intermodulation (IM) Noise

IM noise is generated by the nonlinearities present in communication systems. Similar to thermal noise, nonlinearities are present to some degree in all electrical networks. Nonlinearities can limit the useful signal levels in a system and thus become an important design consideration. The device that is the primary source of IM is the nonlinear HPA. It is an essential element of a communication system. The HPA's efficiency and the degree of nonlinearity are inversely related; thus, the HPA characteristics and operating power levels are important design parameters.

Consider the voltage transfer characteristics of a general two-port, which can be a device, network, or system, as depicted in Figure 1.15. For a two-port memory-less nonlinear network, the transfer function is described by the Taylor series expansion:

$$e_0 = a_1e_i + a_2e_i^2 + a_3e_i^3 + \cdots \tag{1.31}$$

For a single-frequency sinusoid input, $e_1 = A\cos(ax)$, we obtain

$$\begin{aligned}e_0 &= a_1A\cos ax + a_2A^2\cos^2 ax + a_3A^3\cos^3 ax + \cdots \\ &= K_0 + K_1\cos(ax) + K_2\cos(2ax) + K_3\cos(3ax) + \cdots\end{aligned} \tag{1.32}$$

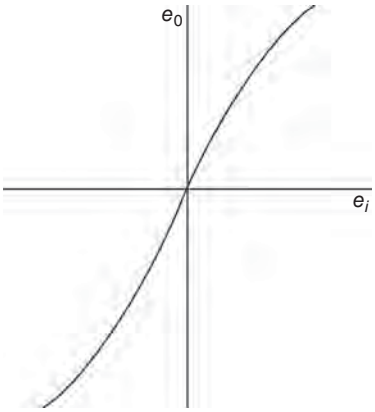


Figure 1.15 Transfer characteristics of a nonlinear two-port network.

where the K values are constants related to $a_1, a_2, a_3, \ldots$. Therefore, a single sinusoid input leads to an output containing the fundamental frequency and its harmonics. Similarly, if $e_i = A\cos\omega_{1i} + B\cos\omega_{2i} + \cos\omega_{3i} + \cdots$, then by using trigonometric identities, we obtain

$$e_0 = K_0 + K_{1i}f(\omega_i) + K_{2i}f(2\omega_i) + K_{3i}f(3\omega_i) + \cdots \tag{1.33}$$

where

$$\begin{aligned} f(\omega_i) &= \text{collection of first-order terms containing } \omega \\ f(2\omega_i) &= \text{second-order terms such as } 2\omega_1, \omega_1 + \omega_2, \omega_1 - \omega_2 \\ f(3\omega_i) &= \text{third-order terms such as } 3\omega_1, 2\omega_1 \pm \omega_2, 2\omega_2 \pm \omega_3, \omega_1 + \omega_2 + \omega_3, \\ &\qquad \omega_1 + \omega_2 - \omega_3, \ldots \\ K &= \text{a distinct constant associated with each term} \end{aligned}$$

Consequently, the output contains harmonics and all possible combinations of the sum and difference frequencies of the input signals. The dc (direct-current) term is of no interest and is filtered out. The desired output corresponds to the linear case, given by the first-order products K_{1i} due to term a_1. All the other outputs are spurious and contribute to the objectionable noise and interference. As the number of input signals increases, the number of IM products grows rapidly. These IM products fall either inside or outside of the RF channel, depending on the order of the product and the separations between the carrier frequencies. In the classic case of a single channel per carrier (SCPC), frequency-division multiplex (FDM-FM) system, IM products are so widespread that they resemble flat Gaussian or thermal noise. As a result, for RF channels employing a large number of carrier frequencies, there is a trade-off between thermal noise and IM noise. As the number of carriers increases, thermal noise per carrier is smaller, whereas the IM noise level is larger. There is some point of optimum loading of the wideband FM transmitter where the thermal noise and the IM noise in the channel when combined represent a minimum of the total system noise, as described in Figure 1.16. This provides guidance for the best operating power level for the amplifier. Typically, the third-order IM product is dominant and often forms part of the specification. Amplifiers are designed to have a carrier to third-order product values exceeding 20 dB. This topic is discussed further in Sections 1.8.3 and 1.8.4.

1.5.7 Distortion Due to Channel Imperfections

An ideal transmission channel transmits all signals without distortion over a certain bandwidth and completely attenuates all signals outside this bandwidth. Such a performance is characterized by a channel that provides a constant loss (ideally zero) and linear phase (i.e., constant

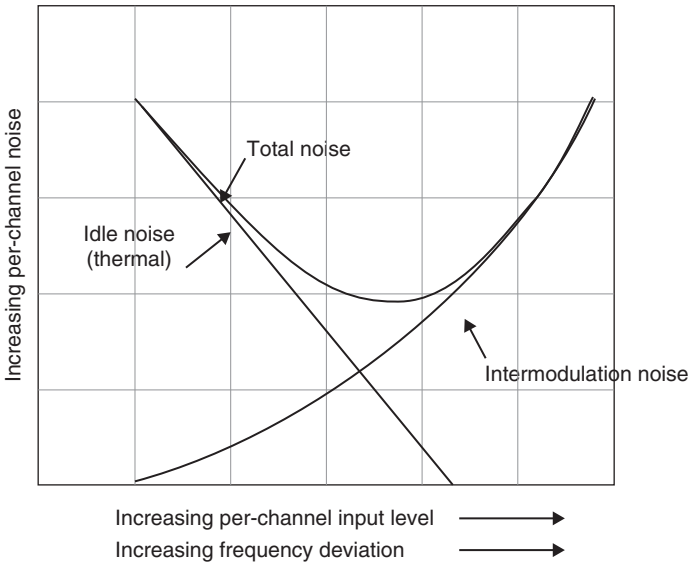
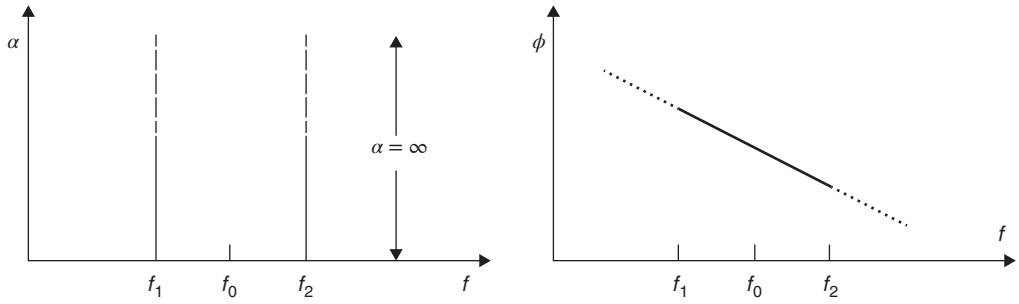


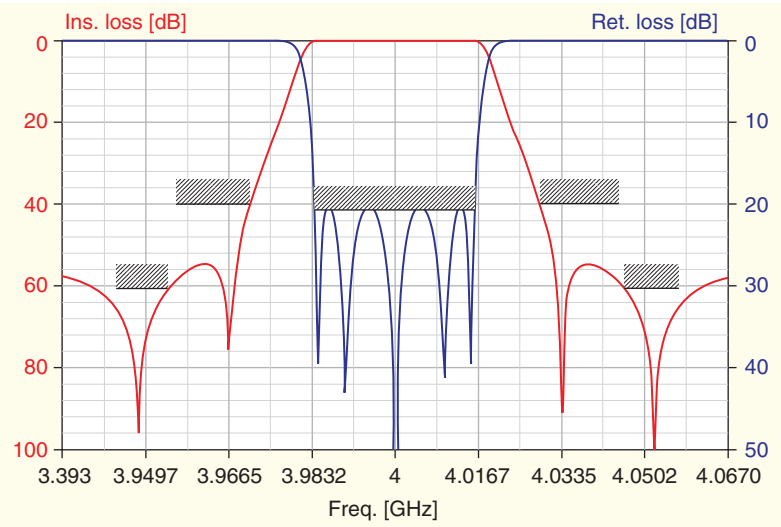
Figure 1.16 Effect of loading and thermal noise in a communication channel.

delay) over the passband and infinite attenuation outside it, as shown in Figure 1.17a and b. Such a filter performance is not feasible. The unit impulse response of such a filter exists for negative values of time, violating the condition of causality [11].

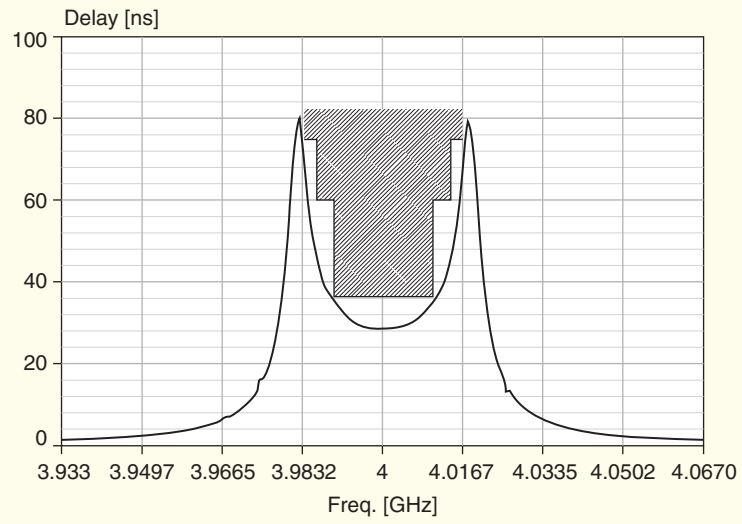
Although ideal filter characteristics are not feasible, it is possible to approach these characteristics as closely as desired. The characteristics of a practical filter, typical of narrowband channels in communication systems, are plotted in Figure 1.17c and d. There exists a trade-off between the design complexity and its departure from ideal filter characteristics. Other components in the communication system, including amplifiers, frequency converters, modulators, cables, and waveguides, are wideband devices, exhibiting a minimal deviation from the flat amplitude and group delays over the narrow channel bandwidths. In other words, filter networks represent the controlling features of amplitude and phase characteristics of a transmission channel. This has been a key driver for filter research and development for many decades.

All practical filters exhibit transmission deviations. Filters are passive and linear devices. Their response is time-invariant and has amplitude and phase shapes that are prescribed functions of frequency. Unlike nonlinear devices such as HPAs, no new frequencies are produced when an FM signal is passed through the filter. However, filters do change the relative amplitude and phase of the carrier and sidebands, which, in turn, is interpreted by the demodulator as additional modulation, causing distortion in the received signal. For insight into this process, consider an FM signal with many sidebands applied to a network with ideal transmission characteristics except at the frequency of one of the sideband components. The amplitude of this component gets altered and is equivalent to adding an extraneous signal to the applied FM signal. As a result, the demodulated output signal consists of the desired signal, which is proportional to the input modulating signal and the undesired signal. When many carriers are present, it can be demonstrated that the transmission deviations in an FM system can introduce baseband frequency components at the output that do not exist at the input. In this sense, the transmission deviations in an FM system have an effect similar to that of an amplifier nonlinearity. For that reason, the distortion introduced by transmission deviations is often called *IM noise*. There is no exact solution to the problem of determining IM noise produced by transmission deviations. An approximate approach for analog systems that has proved successful





(c)



(d)

Figure 1.17 Transmission channel characteristics: (a) an ideal amplitude response; (b) an ideal phase response; (c) a practical amplitude response; (d) a practical group delay response.

is described in Ref. [1] and is based on previous works [12–14]. The analysis assumes that the transmission characteristics of the channel are sufficiently smooth so that the gain and phase as a function of frequency can be represented by

$$Y_n(\omega) = [1 + g_1(\omega - \omega_c) + g_2(\omega - \omega_c)^2 + g_3(\omega - \omega_c)^3 + g_4(\omega - \omega_c)^4] \times e^{i[b_2(\omega-\omega_c)^2 + b_3(\omega-\omega_c)^3 + b_4(\omega-\omega_c)^4]} \tag{1.34}$$

where

ω_c = carrier frequency in radians per second (rad/s)

g_1, g_2, g_3, g_4 = linear, parabolic, cubic, and quartic gain coefficients, respectively

b_2, b_3, b_4 = parabolic, cubic, and quartic phase coefficients, respectively

This assumption is consistent with the frequency response that can be readily achieved by microwave filter networks. The second assumption is that the FM signal has a sufficiently low modulation index and that the bandwidth of the channel is much smaller than the carrier frequency. This is indeed the case for most communication systems. From these assumptions, it is possible to compute the distortion due to transmission deviations. Another related source of distortion is introduced by the HPAs that follow filter networks. The nonlinearity of amplifiers converts the amplitude variations introduced by filters to phase variations, causing distortion of the FM signal. If a limiter can successfully remove the amplitude variations before the signal reaches the device, it can be dismissed. A summary of these distortion terms and noise power (included in Appendix 1.A) is described in Ref. [15]. These values are applicable for analog FM transmission.

For digital systems, the impact of transmission deviations is relatively small, unless the data rate is very high. For advanced digital modulation schemes, sophisticated simulation tools are needed to compute distortion as a function of the amplitude and phase deviations in the RF channel. It leads to trade-offs between in-band response (transmission deviations) and out-of-band attenuation, consistent with microwave filter technology and other system design parameters [16]. Such trade-offs characterize the RF channels. In effect, RF channel filters control the amplitude and phase characteristics of a communication channel, that is, the channel filters define the effective usage of the available channel bandwidths.

It should be noted that for analog FM transmission, variations in the amplitude slope of a filter followed by a nonlinear amplifier cause AM-to-PM conversion, resulting in intelligible (coherent) cross talk between FM carriers. With digital modulation, cross talk is unintelligible (noncoherent), but there is still a modulation transfer, causing an increase in the required E_b/N_0 (ratio of energy per bit to noise density) in the link. Digital transmission is relatively insensitive to variations in group delay unless, of course, it is very large and the data rate is high (short symbol duration). In most cases, the effect of the group delay slope is small on the link E_b/N_0.

1.5.8 RF Link Design

A communication link is characterized by a number of RF links in tandem. In this section, we describe the carrier-to-noise ratio (CNR) for a single link and the impact of cascading a number of such links.

1.5.8.1 Carrier-to-Noise Ratio (*C*/*N*)

The carrier-to-noise ratio (C/N) of a signal is simply given by C/N, where C is the carrier power and N is the total noise present in a given bandwidth. If N_0 is the noise density as defined by the noise power in a bandwidth of 1 Hz, then, by virtue of equation (1.15), we have

$$N_0 = kT_s \quad \text{and} \quad N = N_0 B = kT_s B \tag{1.35}$$

where k is the Boltzmann constant, T_s is the total effective noise temperature, and B is the bandwidth of the carrier frequency. In a communication system, T_s is chosen as the total effective temperature referenced to the input of the LNA of the receive system. It includes the receive antenna noise temperature, lossy transmission lines, bandpass filter prior to the LNA, and the noise temperature of the LNA itself. If G_R is the receive antenna gain at the same reference (i.e., including the losses up to LNA), the carrier power at the reference point is [see equation (1.12b)]

$$C = \text{EIRP} + G_R - P_L \tag{1.36}$$

where $P_L = 20\log(4\pi r/\lambda)$ is referred to as the *path loss* between the transmit and receive antennas. Therefore,

$$\frac{C}{N} = \text{EIRP} + G_R - P_L + 228.6 - 10\log T_s - 10\log B$$

$$= \text{EIRP} + \frac{G}{T_s} - P_L + 228.6 - 10\log B \tag{1.37}$$

This represents the link equation for the thermal noise of the system. The factor $G_R - 10\log T_s$ or G/T_s represents the figure of merit of the receiving system. The link equation can also be expressed in the alternative form

$$\frac{C}{N_0} = \text{EIRP} + \frac{G}{T_s} - P_L + 228.6 \tag{1.38}$$

$$\frac{C}{T_s} = \text{EIRP} + \frac{G}{T_s} - P_L \tag{1.39}$$

All the terms in the link equations (1.37–1.39) are in decibels.

1.5.8.2 Multiple RF Links in Tandem

A communication channel, consisting of a number of RF links, is described in Figure 1.18.

Since the n links in tandem are physically different, the noise power generated in each of the links is not coherent with those generated in the other links. Thus, the noise power of all the links can be summed up as independent power sources to attain the total end-to-end noise power contribution and the overall C/N. The noise power of each link can be referenced to the carrier level, which can be normalized to unity. Therefore, by adding all the noise floors N_i, the overall carrier-to-noise ratio $(C/N)^*$ is given by

$$\left[\left(\frac{C}{N}\right)^*_T\right]^{-1} = \left[\left(\frac{C}{N}\right)^*_1\right]^{-1} + \left[\left(\frac{C}{N}\right)^*_2\right]^{-1} + \cdots + \left[\left(\frac{C}{N}\right)^*_n\right]^{-1} \tag{1.40}$$

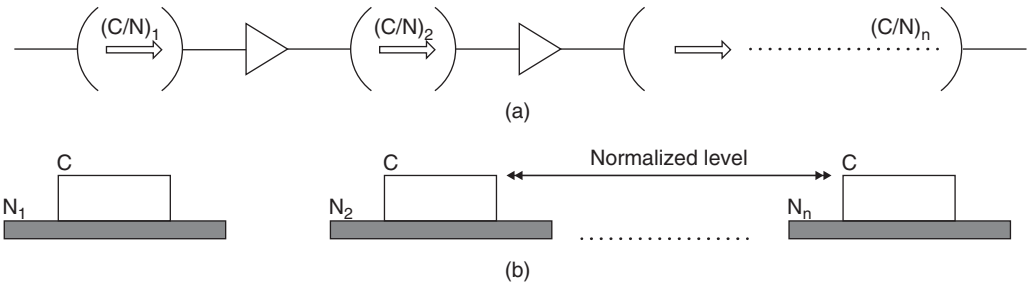


Figure 1.18 A communication channel depicting (a) RF Links in tandem and (b) normalized thermal noise of each link.

The terms $(C/N)_1^*$, $(C/N)_2^*$ signify that C/N is expressed in terms of ratios, as opposed to decibels.

Example 1.5

A transmit earth station has a 1 kW output power and an antenna with a gain of 55 dB. The transmission line and filter losses between the amplifier and antenna amount to 2 dB. Using Example 1.4 as reference for the receive network at the satellite

1) Compute the value of EIRP, the pfd, and the C/N_0 for the 6 GHz uplink for a satellite located at 40,000 km from the surface of the earth. Assume that the receive antenna at the satellite has a gain of 25 dB and include a margin of 3 dB due to losses through the atmosphere and antenna pointing accuracy.
2) Compute the value of the carrier to thermal noise at the satellite for a 36 MHz RF channel.
3) Compute the value of the combined uplink–downlink thermal noise, assuming that the power radiated by the satellite to the earth station in the 4 GHz downlink beam results in a CNR of 20 dB. What is the impact on the total noise if the uplink power is reduced by 10 dB?

Solution:

1) $$\begin{aligned}\text{Uplink EIRP} &= 30\ \text{dBW} + (-2\ \text{dB}) + 55\\ &= 83\ \text{dBW}\\ \text{PFD} &= \text{EIRP} - 10\log 4\pi r^2\\ &= 83 - 163\\ &= -80\ \text{dBW/m}^2\end{aligned}$$

This represents a power density of 10 nW per square meter at the satellite located in the geostationary orbit. From equation (1.38), we have

$$\begin{aligned}\frac{C}{N_0} &= \text{EIRP} - 3 - 20\log\frac{4\pi r}{\lambda} + G_R - L + 228.6 - 10\log T_s\\ &= 83 - 3 - 20\log\frac{4\pi\times 40\times 10^6}{(3\times 10^8)/(6\times 10^9)} + 25 - 2.75 + 228.6 - 10\log 347.1\\ &= 83 - 3 - 200 + 25 - 2.75 + 228.6 - 25.4\\ &= 105.45\ \text{dB Hz}\end{aligned}$$

2) $$\frac{C}{N} = \frac{C}{N_0} - 10\log 36\times 10^6 = 29.9\ \text{dB}$$

3) The combined value of the carrier to thermal noise, in terms of ratios, is given by

$$\left[\left(\frac{C}{N}\right)_T^*\right]^{-1} = \left[\left(\frac{C}{N}\right)_{\text{UL}}^*\right]^{-1} + \left[\left(\frac{C}{N}\right)_{\text{DL}}^*\right]^{-1}$$

$$\left(\frac{C}{N}\right)_{\text{UL}}^* = 10^{2.99} \quad \text{and} \quad \left(\frac{C}{N}\right)_{\text{DL}}^* = 10^2$$

The total thermal noise ratio is, therefore, given by

$$\begin{aligned}\left[\left(\frac{C}{N}\right)_T^*\right]^{-1} &= 10^{-2.99} + 10^{-2}\\ &= 0.001 + 0.01\\ &\approx 0.01\end{aligned}$$

For the ratio in decibels (dB), we obtain $(C/N)_T = 20$ dB.

If the uplink power is reduced by 10 dB, the $(C/N)_{\text{UL}}$ is reduced to 19.9 dB, and the carrier to total thermal noise power is

$$\begin{aligned}\left[\left(\frac{C}{N}\right)_T^*\right]^{-1} &= 10^{-1.99} + 10^{-2} \\ &= 0.102 + 0.01 \\ &= 0.0202\end{aligned}$$

Since this represents a CNR of 16.95 dB, the impact on the total noise is no longer negligible.

1.6 Modulation–Demodulation Schemes in a Communication System

In a communication system, the baseband signal, consisting of a large number of individual message signals, needs to be transmitted over the communication medium, separating the transmitter from the receiver. The efficiency of the transmission requires that this information be processed in some manner before it is transmitted. Modulation is the process whereby the baseband signal is suitably impressed on a carrier signal to increase its efficiency for transmission over the medium. Modulation can shift the signal frequencies to facilitate transmission or change the bandwidth occupancy, or it can alter the form of the signal to optimize the noise or distortion performance. At the receiver, a demodulation scheme reverses this process. Comprehensive coverage of this topic is described in references [2, 3, 6, 11].

Modulation techniques are categorized as linear or nonlinear, depending on whether the modulated signal varies linearly (i.e., the superposition holds) or varies nonlinearly with the message.

There are two forms of modulation: amplitude modulation and angle (phase or frequency) modulation. The process of modulation is represented by

$$M(t) = a(t)\cos[(\omega_c t + \phi(t)] \tag{1.41}$$

Here $a(t)$ represents the amplitude of the sinusoidal carrier and $\omega_c t + \phi(t)$ is the phase angle. Although both the amplitude and angle modulation can be present simultaneously, an amplitude-modulated system is one in which $\phi(t)$ is a constant and $a(t)$ is made proportional to the modulating signal. Similarly, in an angle-modulated system, $a(t)$ is held constant and $\phi(t)$ is rendered proportional to the modulating signal.

1.6.1 Amplitude Modulation

For the amplitude modulated wave, we have

$$M(t) = a(t)\cos\omega_c t \tag{1.42}$$

where the carrier is at frequency f_c and $a(t)$ is the modulating time function. If $a(t)$ is a single-frequency sinusoid of unit amplitude at frequency f_m, then $a(t) = \cos\omega_m t$, and the modulated wave is

$$M(t) = \cos\omega_m t \cos\omega_c t$$

This can be expanded to

$$M(t) = \frac{1}{2}\cos(\omega_c - \omega_m)t + \frac{1}{2}\cos(\omega_c + \omega_m)t \tag{1.43}$$

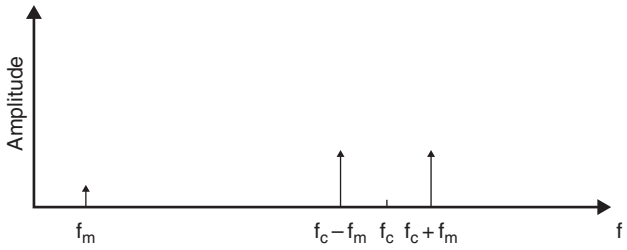


Figure 1.19 Amplitude-modulated carrier with a single sinusoid modulating frequency.

The modulated wave contains no component at the original carrier and only a sideband on either side of the carrier, spaced f_m hertz from the carrier, as reflected in Figure 1.19.

The effect of the product modulation is to translate $a(t)$ in the frequency domain so that it is reflected symmetrically about f_c. It can be shown that this is true for a complex waveform as well. If either sideband is rejected by a filter, the result is a single-sideband (SSB) signal, representing a pure frequency translation. A more general representation of the amplitude modulation is

$$M(t) = [1 + ma(t)] \cos \omega_c t \tag{1.44}$$

This expression is equivalent to adding a dc term of magnitude unity. Also, it is imperative that

$$|ma(t)| < 1 \tag{1.45}$$

so that the envelope of modulated wave remains undistorted, as denoted in Figure 1.20.

Here, m is defined as the modulation index with a maximum value of unity, representing 100% modulation. It represents the relative magnitude of the modulating wave for a carrier frequency of unity magnitude. The modulated wave is derived as

$$M(t) = \cos \omega_c t + \frac{m}{2} \cos(\omega_c - \omega_m)t + \frac{m}{2} \cos(\omega_c + \omega_m)t \tag{1.46}$$

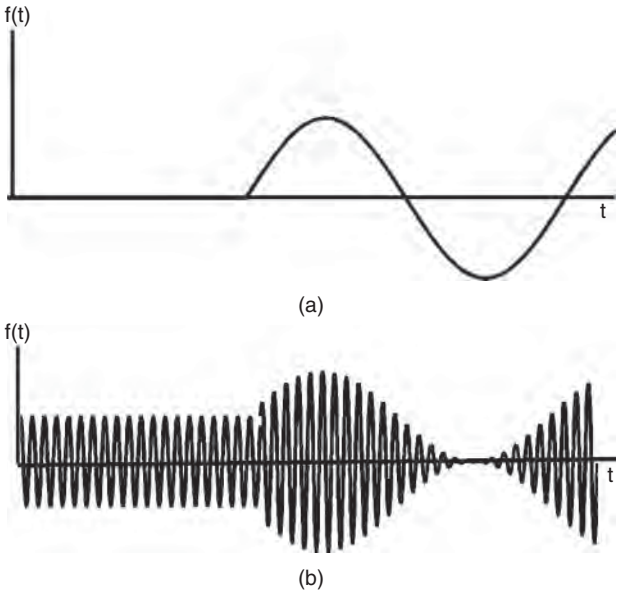


Figure 1.20 Amplitude modulation of a carrier: (a) modulating signal; (b) amplitude-modulated carrier.

The average power in each sideband is $m^2/4$ or a total sideband power of $m^2/2$ W. Thus, for 100% modulation, only one-third of the total power is in the information bearing sideband. For complex signals, the power in the sidebands is considerably less, amounting to a few percent of the total power. A second drawback of the AM wave is its sensitivity to amplitude deviations in the signal path. Any such deviation can result in distortion of the signal. Furthermore, for AM signals, linear HPAs are necessary for signal amplification to avoid excessive distortion of the signal. There are limits to obtaining practical linear amplifiers with an adequate gain and power, rendering AM impractical for long-haul transmissions. However, one substantial advantage of AM is the preservation of the bandwidth. As a consequence, it finds application for frequency translation and in multiplexing individual message channels by translating them to higher frequencies. A good example of amplitude modulation is in the formation of a baseband signal with a large number of voice channels.

1.6.2 Formation of a Baseband Signal

A message channel is composed of a large number of independent signals multiplexed to form a composite signal that occupies a continuous spectral range, referred to as the *bandwidth* of the signal. A hierarchy has been established for North American telephony communication systems to allow standardization and a high degree of commonality. The basic message channel, although originally intended for voice transmission, has also been adopted for data transmission. The basic group consists of 12 channels, each 4 kHz wide extending over the band 60–108 kHz. The equipment can be considered as a series of modulators with distinct carrier frequencies, followed by the appropriate bandpass filters and then multiplexed to form the composite signal as shown in Figure 1.21.

Here, modulation is amplitude modulation, and the single sideband is extracted via the bandpass filters. Amplitude modulation is a linear process, although the mixer to modulate the carrier is inherently a nonlinear device. Here, the mixer is used in a quasi-linear mode and simply translates the modulation frequency symmetrically about the carrier frequency. This group of 12 channels, occupying a 48 kHz bandwidth, is a basic building block in the Bell system. The next step in the FDM hierarchy is the combination of five groups in a 60-channel supergroup, for a bandwidth of 240 kHz, occupying a 312–552 kHz frequency band. Modern broadband transmission systems are capable of larger and larger systems, extending to groups of up to 3600 message channels. Although the grouping described here is generally used for long-haul Bell systems, it is not a universal standard. Along with telephony channels, television channels occupy bandwidths of 4–6 MHz or data channels ranging from kilobits to several megabits per second range, or Internet traffic. Therefore, a typical composite signal, referred to as a *baseband signal*, can vary from a simple basic group to a combination of groups of voice

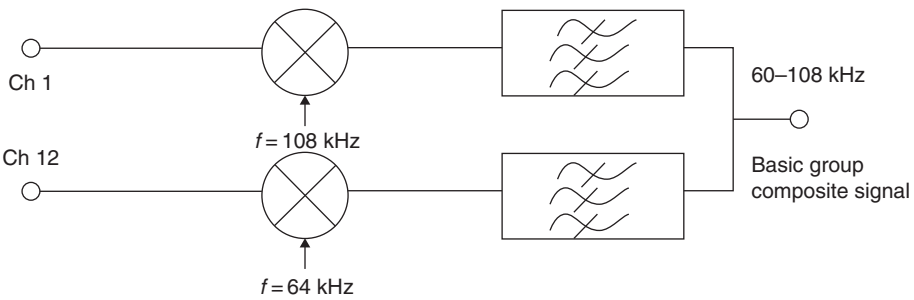


Figure 1.21 Formation of a message signal.

channels, television signals, and data channels. The signal composition is determined by the traffic requirements of the system. The hierarchy of data rates exists all the way up to optical fiber rates of 40 Gbps (gigabits per second).

1.6.3 Angle-Modulated Signals

The modulated wave for angle-modulated signals is represented by

$$M(t) = A\cos[\omega_c t + \phi(t)] \tag{1.47}$$

Phase modulation (PM) is defined as angle modulation in which the instantaneous phase deviation $\phi(t)$ is proportional to the modulating signal voltage. Frequency modulation is angle modulation (FM) in which the carrier varies with the integral of the modulating signal. The average power for a PM or an FM wave is proportional to the square of the voltage and results in

$$\begin{aligned} P(t) &= A_c^2\cos^2[\omega_c t + \phi(t)] \\ &= A_c^2\left[\frac{1}{2} + \frac{1}{2}\cos(2\omega_c t + 2\phi(t))\right] \end{aligned} \tag{1.48}$$

The second term is assumed to consist of a large number of sinusoids about the carrier frequency $2f_c$, with the average value of zero and

$$P_{\text{av}} = \frac{A_c^2}{2} \tag{1.49}$$

Thus, the average power of a FM wave is the same as the average power in the absence of modulation. This represents a major advantage, compared with the AM, where the average power of information bearing sidebands is a third or less than that of the average power of the carrier frequency. However, there are no free lunches, and this power advantage comes at the expense of increased bandwidths. FM frequency analysis is quite complicated and beyond the scope of this book. A brief account of the key results and underlying assumptions are now described.

1.6.3.1 Spectra of Analog Angle-Modulated Signals

Narrowband FM Phase and FM are special cases of angle modulation, and one form is easily derived from the other. The discussion is confined to FM, owing to its wider range of applications for both analog and digital communication systems. A frequency-modulated carrier is shown in Figure 1.22.

The modulating signal is assumed to be a repetitive sawtooth of period T, where $(2\pi/T) \ll \omega_c$. As the sawtooth modulating signal increases in magnitude, the FM signal oscillates more rapidly, resulting in the widening of the frequency spectrum. Note that its amplitude remains unchanged.

Analysis of the FM process is inherently much more complicated than that for AM, due to the nonlinearity of the FM process. Assume a sinusoidal modulating signal $v(t)$ at frequency f_m:

$$v(t) = a\cos\omega_m t \tag{1.50}$$

The instantaneous radian frequency ω_i is then

$$\omega_i = \omega_c + \Delta\omega\cos\omega_m t, \quad \Delta\omega \ll \omega_c \tag{1.51}$$

where $\Delta\omega$ is a constant, depending on the amplitude a. As a result, the instantaneous radian frequency varies about the unmodulated carrier frequency ω_c at the rate ω_m of the modulating signal and with a maximum deviation of $\Delta\omega$ radians. The phase variation $\theta(t)$ for this special

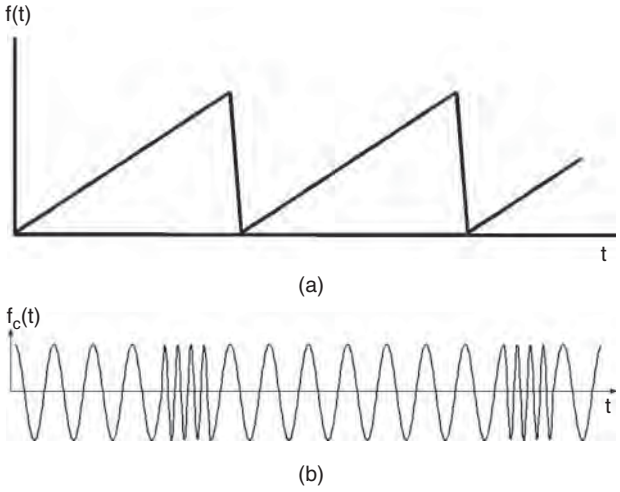


Figure 1.22 Frequency modulation: (a) modulating wave; (b) modulated FM carrier.

case is computed by

$$\theta(t) = \int \omega_i \, dt = \omega_c t + \frac{\Delta\omega}{\omega_m} \sin \omega_m t + \theta_0 \tag{1.52}$$

where θ_0 is a constant, providing a reference for the phase. By choosing it to be zero, we see that the frequency-modulated carrier is

$$M(t) = \cos(\omega_c t + m \sin \omega_m t) \tag{1.53}$$

with

$$m = \frac{\Delta\omega}{\omega_m} = \frac{\Delta f}{f_m}$$

where m is the modulation index and is given by the ratio of the frequency deviation to the baseband bandwidth. The modulated carrier frequency is written in its expanded form as $M(t) = \cos \omega_c t \cos(m \sin \omega_m t) - \sin \omega_c t \sin(m \sin \omega_m t)$. For $m \ll \pi/2$, we have

$$\begin{aligned} M(t) &\approx \cos \omega_c t - m \sin \omega_m t \sin \omega_c t \\ &\approx \cos \omega_c t - \frac{m}{2}[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t] \end{aligned} \tag{1.54}$$

Under this condition, the system is referred to as *narrowband FM* and has a form similar to that of AM carriers. It contains the original unmodulated carrier and two sideband frequencies, displaced $\pm\omega_m$ radians from ω_c. The bandwidth of the narrowband FM signal is thus $2f_m$, like that of an AM signal. Despite this similarity, there is a clear distinction between AM and the narrowband FM; in the FM case, the carrier amplitude is constant whereas for the AM, the amplitude varies in accordance with the modulating signal. The carrier and sideband terms are in phase for the AM, whereas the sidebands are in phase quadrature with the carrier for narrowband FM case.

Wideband FM The case where $m > (\pi/2)$ is referred to as *wideband FM*. Its analysis requires the expansion of $M(t)$ given by equation (1.53). In the equation, the terms $\cos(m \sin \omega_m t)$ and $\sin(m \sin \omega_m t)$ are periodic functions of ω_m and are expanded in a Fourier series of period

$2\pi/\omega_m$. This expansion can be expressed in terms of the Bessel functions [3]:

$$\begin{aligned} M(t) = {} & J_0(m)\cos\omega_c t - J_1(m)[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t] \\ & + J_2(m)[\cos(\omega_c - 2\omega_m)t + \cos(\omega_c + 2\omega_m)t] \\ & - J_3(m)[\cos(\omega_c - 3\omega_m)t - \cos(\omega_c + 3\omega_m)t] \\ & + \cdots \end{aligned} \tag{1.55}$$

Equation (1.55) represents a time function, consisting of a carrier and an infinite number of sidebands, spaced at frequencies, $\omega_c \pm \omega_m, \omega_c \pm 2\omega_m$, and so on.

These successive sets of sidebands are called *first-order sideband*, *second-order sideband*, and so on, the magnitudes of which are determined by the coefficients $J_1(m), J_2(m), \ldots$ respectively.

When two or more sinusoids are present in the modulating signal, the spectrum contains not only multiples of individual modulating frequencies but also all the possible sums and differences of multiples of the modulation frequencies. As more and more modulating signals are added, the complexity of the solution increases rapidly. Eventually, the baseband signal is represented by random noise, extending uniformly across the baseband from 0 to f_m hertz. The frequency spectrum of the corresponding FM signal appears as a continuum of sidebands. The magnitudes of the carrier and sideband terms depend on m, the modulation index, expressed by the appropriate Bessel function. From a theoretical viewpoint, the bandwidth required for 100% of the signal energy is infinite. For practical systems, only significant sidebands with a magnitude of at least 1% of the magnitude of the unmodulated carrier are considered. The number of significant sidebands varies with m and can be determined from the tabulated values of the Bessel function. For the large value of $m(> 10)$, the minimum bandwidth is given by $2\Delta f$, where Δf is the peak deviation. A general rule, postulated by J. R. Carson in 1939 for the minimum bandwidths of an FM signal, is

$$B_T \approx 2[f_m + \Delta f] \tag{1.56}$$

This is an approximate rule (referred to as *a rule of thumb*) and is appropriate for most applications. The actual bandwidth required is, to some extent, a function of the waveform of the modulating signal and the quality of transmission desired. It can be seen from this expression that for $m < 1$, the minimum bandwidth is given by $2f_m$, and for $m > 10$, the minimum bandwidth is $2\Delta f$. A more accurate assessment of bandwidth can be made by computing the coefficients of the Bessel function to any desired magnitude for a given modulation index. For modulation indices between 1 and 10, bandwidth given by equation (1.56) corresponds to the case of including sidebands with at least 10% of the magnitude of the unmodulated carrier frequency. For large values of m (>10), the number of significant sidebands in FM wave is equal to m. A key point is that wideband FM systems require greater bandwidth compared with that of AM systems, the extent of which is determined by the modulation index.

1.6.4 Comparison of FM and AM Systems

So far, only ideal AM and FM systems have been investigated. It was shown that an AM system is a linear process and conserves the bandwidth, but the modulation process transfers only one third of the input power to the information bearing sidebands, the remainder power remains at the carrier frequency. An FM system, however, is a nonlinear process. Modulation produces new frequencies, and the modulated signal requires much larger bandwidths. From a theoretical standpoint, the energy of an FM signal is dispersed over an infinite bandwidth. However, the bulk of the energy is contained in the first few sidebands.

For most FM signals, 99% of the energy is contained in a bandwidth of $2(f_m + \Delta f)$, where Δf is the peak deviation and f_m is the baseband frequency. The advantage of the FM system is that all

of the input power is transferred to the information-bearing sidebands. The average power at the carrier frequency, after modulation, is zero. Another advantage is that the amplitude of the FM envelope is nearly constant and, thus, the signals are amplified with a minimum of distortion by the practical nonlinear amplifiers. The critical parameter for any modulation scheme is the S/N ratio under different traffic conditions. For an AM system, this ratio is expressed by [3]

$$\frac{S}{N} = \frac{A_c^4}{8N^2 + 8NA_c^2} \tag{1.57}$$

where A_c is the voltage amplitude of the unmodulated carrier and N is the mean noise power. The CNR is

$$\frac{C}{N} = \frac{A_c^2}{2N} \tag{1.58}$$

Therefore

$$\frac{S}{N} = \frac{1}{2}\ \frac{(C/N)^2}{1 + 2(C/N)} \tag{1.59}$$

For $C/N \ll 1$, the output SNR drops as the square of the CNR. This is the suppression that is characteristic of envelope detection. For CNR $\gg 1$

$$\frac{S}{N} = \frac{1}{4}\frac{C}{N} \tag{1.60}$$

The output SNR is then linearly dependent on the C/N, another characteristic of envelope detectors. In addition, the relation shows that no SNR improvement is possible for AM systems. An increase in the transmission bandwidth $2f_m$, needed to pass the AM signals, serves only to increase the noise N, decreasing the output SNR.

For a FM system, the SNR is given by [3],

$$\frac{S}{N} = 3\left(\frac{\Delta f}{B}\right)^2 \frac{C}{2N_0 B} \tag{1.61}$$

Here, C is the average power of the FM carrier and $\Delta f/B$ is the modulation index m. Denoting $2N_0B = N$, the average noise power in the sidebands, we obtain

$$\frac{S}{N} = 3m^2\frac{C}{N} \tag{1.62}$$

Compare the FM and AM systems by assuming the same unmodulated carrier power and noise spectral density n_0 for both. For a 100% modulated FM signal, the C/N corresponds to the CNR of an AM system as described by equation (1.49), that is, $(S/N)_{\text{AM}} = C/N$. Equation (1.62) can be modified to read as

$$\left(\frac{S}{N}\right)_{\text{FM}} = 3m^2\left(\frac{S}{N}\right)_{\text{AM}} \tag{1.63}$$

For a large modulation index (corresponding to a wide transmission bandwidth, with $m \gg 1$), the SNR can be increased significantly over that of the AM case. For example, if $m = 5$, the FM output SNR is 75 times that of an equivalent AM system. Alternatively, for the same SNR at the output in both receivers, the power of the FM carrier can be reduced 75 times, but this requires an increase in the transmission bandwidth from $2B$ (AM case) to $16B$ (FM case). Note that this bandwidth is greater than that given by the rule of thumb criteria, as it is based on using sidebands with at least 1% magnitude of the carrier. Frequency modulation provides a substantial improvement in the SNR, but at the expense of increased bandwidth. This, of course, is a characteristic of all noise improvement systems.

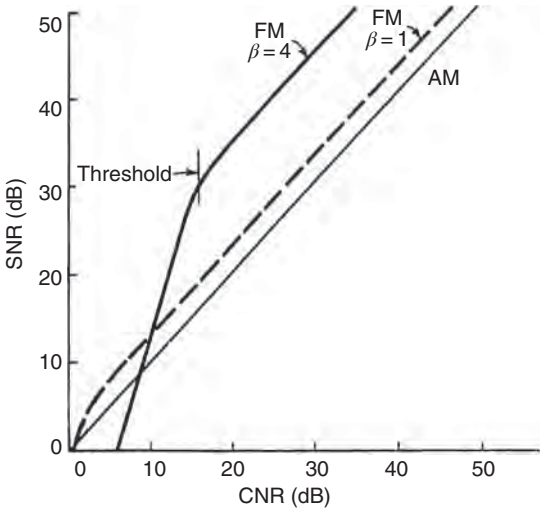


Figure 1.23 FM threshold effect. (*Source*: Schwartz 1980 [3]. Reproduced with the permission of McGraw-Hill Professional.)

Is it possible to continue increasing the output SNR indefinitely by increasing the frequency deviation, and the corresponding bandwidth? If the transmitter power is fixed, increasing the frequency deviation increases the required bandwidth with it, incurring more noise. Eventually, the noise power at the limiter becomes comparable with the signal power and the noise is found to "take over" the system. The output SNR falls off much more sharply than the input CNR. This effect is called the threshold effect, as described in Figure 1.23. For proper operation of FM systems, the CNR must be kept above the threshold value, typically >13 dB.

1.7 Digital Transmission

The widespread use of digital communication systems is the result of many factors. They include the relative simplicity of digital circuit design, the ease with which integrated circuit techniques can be applied to digital circuitry and the rapid advances in digital signal processing (DSP) techniques. The information content of a digital signal consists of discrete states, such as the presence or absence of a voltage, characterized as 1 or 0. This implies that a simple decision circuit can be employed as a regenerator, that is, a corrupted digital signal enters on one side and a clean perfect signal comes out the other side. Accumulated noise on the corrupted signal stops at the regenerator and does not accumulate as the signal passes through different stages of the communication channel. Using coding techniques (at the expense of small increase in bandwidth) to minimize the effect of noise in the detection of 1s and 0s at the regenerating stations (repeaters) at appropriate intervals allows almost error-free digital transmission over long distances. Digital techniques allow development of efficient compression techniques by eliminating statistical redundancies, removing unnecessary information such as pauses in voice communication, or removing small invisible parts of a picture or redundancies between images in video transmission. Compression techniques increase transmission capacity and more efficient utilization of the frequency spectrum (conservation of bandwidth). However, compression of information implies additional hardware and latency. Compression techniques have

matured to the point where digital systems now require less bandwidth than analog signals. In the early era of satellite communication, one 36 MHz transponder carried one analog TV channel; same transponder can now carry 10 digitally compressed channels. Not only that, the 10 digital channels can be combined into 1 signal to allow the transmitters to operate near saturation. It is apparent that digital communication systems require considerable amount of electronic circuitry. Such circuitry used to be expensive but nowadays electronics circuitry in the form of very-large-scale integrated (VLSI) circuits has become inexpensive. Although cost used to be a major factor in selecting analog communication over digital communications in the past that is no longer the case. Digital networks have therefore become the overwhelming choice for communication systems.

1.7.1 Sampling

The well-known Nyquist criterion states that "If a message that is magnitude–time function is sampled instantaneously at regular intervals, and at a rate at least twice the highest significant message frequency, then the samples contain all of the information of the original message."

This rather amazing result is fundamental to the ability to digitize analog signals without the loss of information. As an example, a message band limited to f_m hertz is completely specified by its amplitudes at any set of points in time spaced T seconds apart, where $T = \frac{1}{2}f_m$, as shown in Figure 1.24. The result implies that the bandwidth required to convert an analog signal into a digital representation is at least twice the bandwidth of the highest frequency in the signal. The amplitude-modulated pulse signal can then be transmitted to a receiver in any form that is suitable from a transmission standpoint. At the receiving end, a reverse process is followed to recreate the original pulse amplitude-modulated signal. To recover the original message, it is necessary to pass the impulses through an ideal low-pass filter with a cutoff of f_m. The output of this filter is a replica of the original message, delayed in time. Since information can be digitized without loss of accuracy, the challenge is then to exploit the potential for DSP for communications systems. The first step in this process is the quantization of the sampled signals.

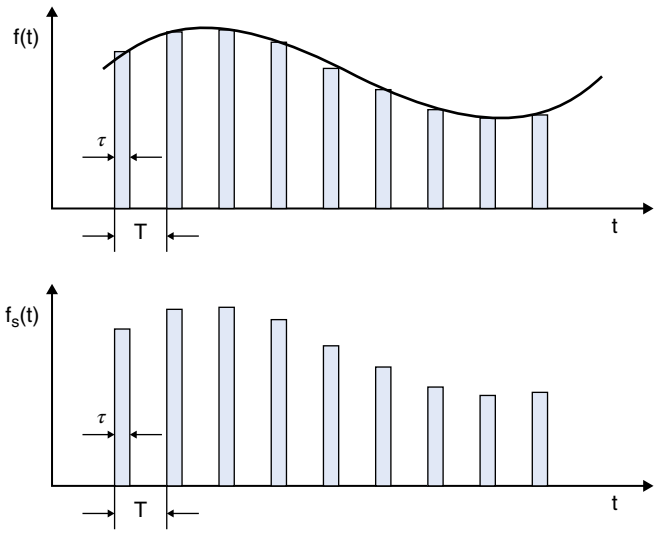


Figure 1.24 The sampling process (τ = sampling time; $T = 1/f_c$: sampling interval): (a) input function $f(t)$ and (b) sampled output $f_s(t)$.

1.7.2 Quantization

The process of quantization consists of breaking the amplitudes of the signals into a prescribed number of discrete amplitude levels. Then, when the message is sampled in a pulse amplitude-modulated (PAM) system, the discrete amplitude nearest the true amplitude is sent. Consequently, the process of quantization introduces an error in the representation of the amplitudes of the sampled signals. Moreover, this error is irretrievable. This seemingly purposeful distortion of the signal is kept below the noise introduced during transmission and at the receiver. In essence, the uncertainty introduced by the fundamental thermal noise, and noise due to imperfections in circuits and devices, limits the ability to distinguish between all the possible amplitude levels, thus making quantization possible. The advantage of quantization is that once the number of discrete amplitude levels is established, each level can then be coded in some arbitrary form before transmission. Thus, quantization makes it possible to deploy the full potential of DSP techniques to optimize the information flow in a communication system.

1.7.3 PCM Systems

Systems that embody the transmission of digitized and coded signals are commonly called *pulse-code-modulated* (PCM) systems. Binary digital systems constitute the most frequently encountered form of PCM systems. A quantized sample can be sent as a simple pulse with certain possible discrete amplitudes. However, if many discrete samples are required, design circuits become quite complicated and uneconomical. On the contrary, if several pulses are used as a code group to describe the amplitude of a sample, then each pulse can have only two states. In a binary system, a code group of m on/off pulses can be used to represent 2^m amplitudes. For example, eight pulses yield 2^8 or 256 amplitude levels. The m pulses must be transmitted in the general sampling interval allotted for the quantized sample. This constraint necessitates an increase in the bandwidth by a factor of m.

For the digital transmission of a 4 kHz voice channel, using 256 levels of quantization and binary code requires a bandwidth of $4 \times 2 \times 8$ or 64 kHz, a factor of 16 times greater than that of an analog system. The required bandwidth can be reduced by choosing a smaller number of amplitude levels for sampling or by using a code where the pulses can be represented by more than two amplitude levels. A quantized signal sample can be coded into a group of m pulses, each with n possible amplitude levels. Thus, if the signal is quantized into M possible amplitude levels, then, $M = n^m$. Each combination of n^m must correspond to one of the M levels. Consider an example with four amplitude levels ($n = 4$) to represent a pulse. The 256 amplitude levels of the sampled signal in the previous example can now be represented by four pulses per sample ($m = 4$). This requires a bandwidth of $2 \times n$ or eight times the bandwidth of the analog system. Similarly, if a smaller value is selected for M, the bandwidth is again reduced. For all such schemes, there is a trade-off involved in terms of system noise and signal power. This ability to trade the bandwidth and SNR, back and forth via coding and signal processing, is a characteristic of all PCM systems.

1.7.4 Quantization Noise in PCM Systems

The process of quantization introduces noise at the transmitting terminal. This noise depends on the number of amplitude intervals chosen to represent a signal. For a signal quantized with a uniform interval, the peak signal to rms noise ratio is given by [2, 3]

$$\left(\frac{S}{N}\right)^* = 3M^2$$
$$= 3n^{2m}$$

Table 1.1 SNR versus relative bandwidth for binary transmission.

Number of quantizing levels	Binary digits for coding	Relative bandwidth	Peak SNR (dB)
8	3	6	22.8
16	4	8	28.8
32	5	10	34.8
64	6	12	40.8
128	7	14	46.8
256	8	16	52.8
512	9	18	58.8
1024	10	20	64.8

For the binary case, $n = 2$ and the SNR in dB is

$$\left(\frac{S}{N}\right) = 4.8 + 6m \text{ dB} \tag{1.64}$$

This relation provides the trade-offs between the SNR and the relative bandwidth, summarized in Table 1.1.

The frequency spectrum of the quantizing noise with a uniform interval is essentially flat over the range of interest. Quantization does not need to be uniform. In fact, few message signals possess uniform amplitude distribution, and most have a large dynamic range. To overcome such constraints, quantization, typically, has nonuniform spacing and is optimized to achieve a relatively uniform signal-to-distortion ratio over a wide dynamic range. Such nonuniform quantization is referred to as *companding*. Quantization noise can be reduced to any desired degree by choosing increasing finer quantization levels. However, the larger the number of quantizing steps, the greater the required bandwidth is. It is, therefore, desirable to choose as few steps as necessary to meet the objectives of the transmission. Needless to say, many subjective tests have been carried out to determine the acceptable number of levels for voice, video, and data signals. High-quality speech transmission is readily achieved with 128 levels or a 7 bit PCM. Good-quality television requires 9 or 10 bit PCM.

1.7.5 Error Rates in Binary Transmission

The deliberate quantization error or noise imparted to the PCM signal is a major source of signal impairment and originates only at the transmitting (or coding) end of the system. This noise can be made arbitrarily small at the expense of increased bandwidth. The other type of noise that is always present is the thermal noise generated by dissipative elements in the path of the signals and the noise generated by active devices. These noises are random and follow a Gaussian distribution function. They are added to the incoming group of pulses at the receiver. The noise density and distribution function of such noise are the same as those given in equations (1.19) and (1.20).

To detect the presence or absence of a pulse in a binary system, a minimum SNR is required on the digital line. If the pulse power is too low compared with the noise power, the detector will make occasional errors, indicating a pulse where there is none or vice versa. However, if the signal power is increased, the error can be rendered arbitrarily small. To determine the probability of error quantitatively, assume the amplitude of the pulse to be V_p when it is present and zero when it is absent, designated by 1 and 0, respectively. The composite sequence of the

binary symbols and the received noise is sampled once every binary interval and decision must be made as to whether a 1 or 0 is present. A simple and perhaps obvious way to decide is to say that if the voltage plus noise sample exceeds $V_p/2$, it is a 1 and if it is less than $V_p/2$, it is a zero. Error then occurs if, with a pulse present, the composite voltage sample is less than $V_p/2$ or, with a pulse absent, if the noise alone exceeds $V_p/2$. To calculate the probability of error, let us assume that a 0 is sent. The probability of error is then just the probability that noise will exceed $V_p/2$ volts and be mistaken for a 1. The probability of error is therefore the voltage that appears somewhere between $V_p/2$ and ∞. Assuming noise to be Gaussian with an rms value of σ, the probability of error is given by [3]

$$P_{e0} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{V_{p/2}}^{\infty} \mathrm{e}^{-v^2/2\sigma^2}\, dv \tag{1.65}$$

In a similar manner, the probability of error when a pulse is sent and interpreted as zero:

$$P_{e1} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\infty}^{V_{p/2}} \mathrm{e}^{-(v-V_p)^2/2\sigma^2}\, dv \tag{1.66}$$

The two types of errors are mutually exclusive and are equivalent. If it is further assumed that the two binary signals are equally likely, then the system probability P_e is the same as P_{e0} or P_{e1}: $P_e = P_{e0} = P_{e1}$.

The probability function P_e is well known and available in various mathematical tables. It is plotted in decibels as a function of V_p/σ in Figure 1.25. It should be noted that P_e depends solely on V_p/σ, the ratio of peak signal to rms noise ratio. It is interesting that P_e has a maximum value of 1/2. Thus, even if the signal is entirely lost in the noise, the receiver cannot be wrong by more

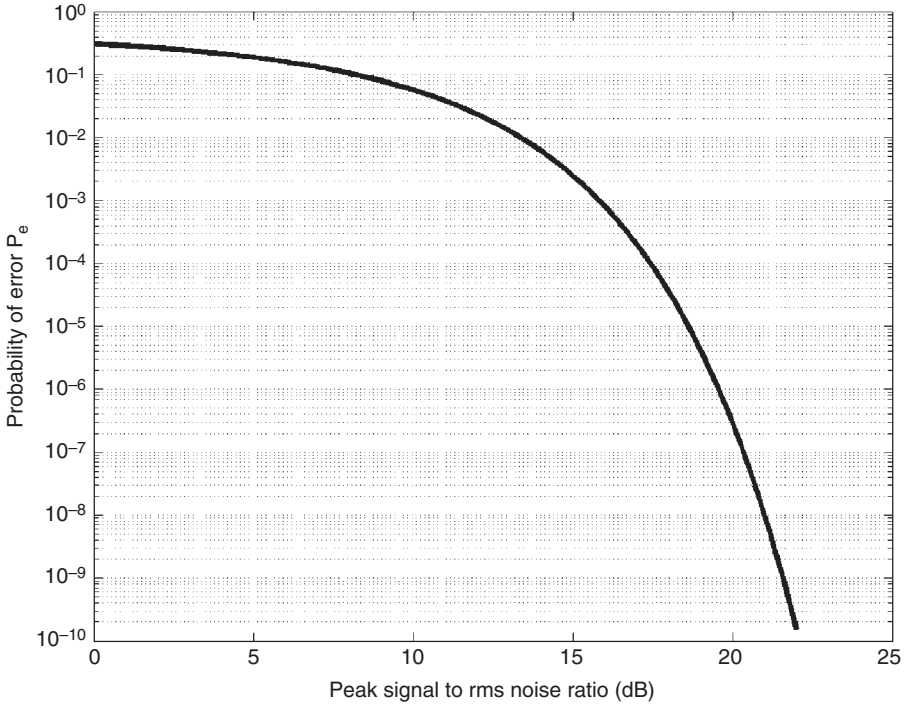


Figure 1.25 Probability of error versus peak signal to rms noise voltage ratio.

than half the time on the average. The probability curve indicates a sharp drop around the 16 dB level. Below this level, the error rate increases sharply and is called the *threshold effect*. For this reason, for the transmission of binary digits, the threshold level is chosen to be somewhere between 16 and 18 dB.

The implicit assumptions in this probability curve are (1) the statistics of the receive signal plus noise are Gaussian and (2) the transmission system is transparent, with no effect on the signal statistics or noise prior to detection. These assumptions allow the chosen decision level to be midway within a pulse amplitude.

1.7.6 Digital Modulation and Demodulation Schemes

Advances in data compression, digital modulation, and coding techniques coupled with the spectacular reduction in the costs of digital circuits are diverting more and more traffic to the digital domain. It is a matter of time before most, if not all, traffic is carried by using digital communication systems. A brief overview of digital modulation schemes, aspiring to achieve ever greater power and bandwidth efficiencies, approaching the Shannon limit, is outlined. The frequency spectrum of various modulation schemes has an impact on the desired filter characteristics, required to extract and process information in the communication channel.

Digital baseband signals can be modulated on a sinusoidal carrier by modulating one or more of its three basic parameters: amplitude, frequency, and phase. Accordingly, there are three basic modulation schemes: amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK).

1.7.6.1 Amplitude Shift Keying (ASK)

This type of modulation is characterized by the amplitude of the carrier wave switching between zero (off state) and some predetermined amplitude level (on state). The amplitude-modulated ASK signal is therefore given by

$$M(t) = Af(t)\cos\omega_c t \tag{1.67}$$

where $f(t) = 1$ or 0, over intervals T seconds long. This is analogous to the amplitude modulation in analog systems. The Fourier transform of the ASK signal is then given by

$$F(\omega) = \frac{A}{2}[F(\omega - \omega_c) + F(\omega + \omega_c)] \tag{1.68}$$

The binary signal simply shifts the frequency spectrum to the carrier frequency f_c with the signal energy distributed between the upper and lower sidebands. The required transmission bandwidth is twice the baseband bandwidth. For a pulse of amplitude A and width T (the binary interval), the spectrum is given by

$$A\frac{T}{2}\left[\frac{\sin(\omega - \omega_c)T/2}{(\omega - \omega_c)T/2} + \frac{\sin(\omega + \omega_c)T/2}{(\omega + \omega_c)T/2}\right] \tag{1.69}$$

This is the well-known $\sin(x)/x$ response of a pulse with a finite width, as illustrated in Figure 1.26.

1.7.6.2 Frequency Shift Keying (FSK)

The frequency-modulated signal by a binary pulse is represented by

$$M(t) = A\cos\omega_1 t \quad \text{or} \quad M(t) = A\cos\omega_2 t \tag{1.70}$$

where $-T/2 \le t \le T/2$.

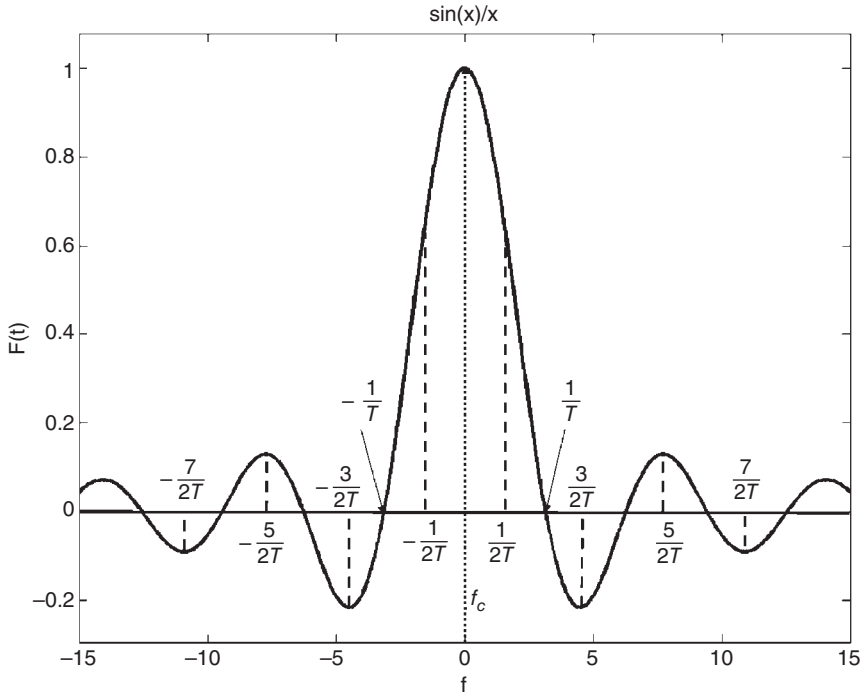


Figure 1.26 Spectrum of a periodic ASK signal.

In this scheme, one of the frequencies, say, f_1, can represent a one and the other frequency f_2, a zero. An alternative representation of FSK is, by letting $f_1 = f_c - \Delta f$ and $f_2 = f_c + \Delta f$:

$$M(t) = A\cos(\omega_c \pm \Delta\omega)t \tag{1.71}$$

The frequency deviates $\pm\Delta f$ about f_c, and Δf represents the frequency deviation. The frequency spectrum of the FSK is complex and has a form similar to that for analog FM.

1.7.6.3 Phase Shift Keying (PSK)

This form of modulation is characterized by a change of phase of the carrier frequency. Owing to the binary nature of the baseband signal, this simply implies a change of polarity. A PSK signal is represented in a form similar to ASK

$$M(t) = f(t)\cos\omega_c t, \quad \frac{-T}{2} < t < \frac{T}{2} \tag{1.72}$$

where $f(t) = \pm 1$. Here a 1 in the baseband binary stream corresponds to positive polarity and a 0 to negative polarity. The PSK signal has the same double-sideband characteristic as the ASK transmission. This result is similar to low index phase modulated analog systems. A comparison of the basic ASK, FSK, and PSK modulations is provided in Figure 1.27.

1.7.7 Advanced Modulation Schemes

As stated in Section 1.4, the two primary resources in a communication system are the signal power and the available transmission bandwidth. Advanced modulation schemes are geared to achieving higher efficiencies in either or both of these resources. Before proceeding further, let us examine the bandwidth and power efficiencies of a digital system.

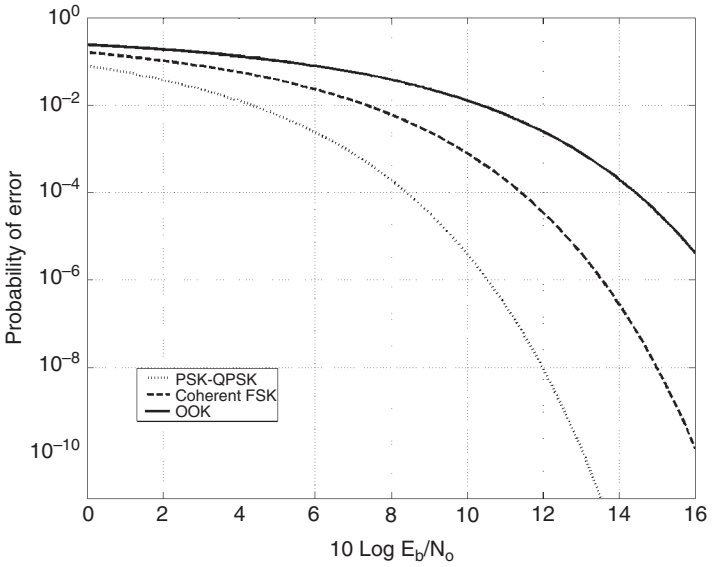


Figure 1.27 Comparison of ASK, FSK, and PSK modulation schemes. (*Source*: Schwartz 1980 [3]. Reproduced with the permission of McGraw-Hill Professional.)

1.7.7.1 Bandwidth Efficiency

As described in Section 1.3, Shannon's Theorem States

$$C = B\log_2\left(1 + \frac{S}{N}\right)$$

If the information rate R is equal to C, then

$$\frac{R}{B} = \log_2\left(1 + \frac{S}{N}\right) \tag{1.73}$$

defines the ultimate limit of the bandwidth efficiency, or the so-called Shannon limit. As described in Section 1.7.3, the quantized state of a signal can be represented by pulses of varying amplitudes or phases. Each state of a signal sample may be represented by [17]

$$M = 2^m$$
$$m = \log_2 M \tag{1.74}$$

Each state of M is referred to as a *symbol* and consists of m bits. Each symbol is transmitted as an electric voltage or current waveform. If the duration of the transmission of a symbol is T_s, then the data rate R is given by

$$R = \frac{m}{T_s} = \frac{\log_2 M}{T_s} \tag{1.75}$$

If T_b represents the duration of a bit ($= T_s/m$) and B is the allocated bandwidth, then the transmission bandwidth efficiency is expressed as

$$\frac{R}{B} = \frac{\log_2 M}{BT_s} = \frac{1}{BT_b} \tag{1.76}$$

The smaller the BT_b product, the higher is the bandwidth efficiency of the communication system. By assuming ideal Nyquist filtering, bandwidth B is simply given by $1/T_s$ and

$$\frac{R}{B} = \log_2 M \text{ bps/Hz} \tag{1.77}$$

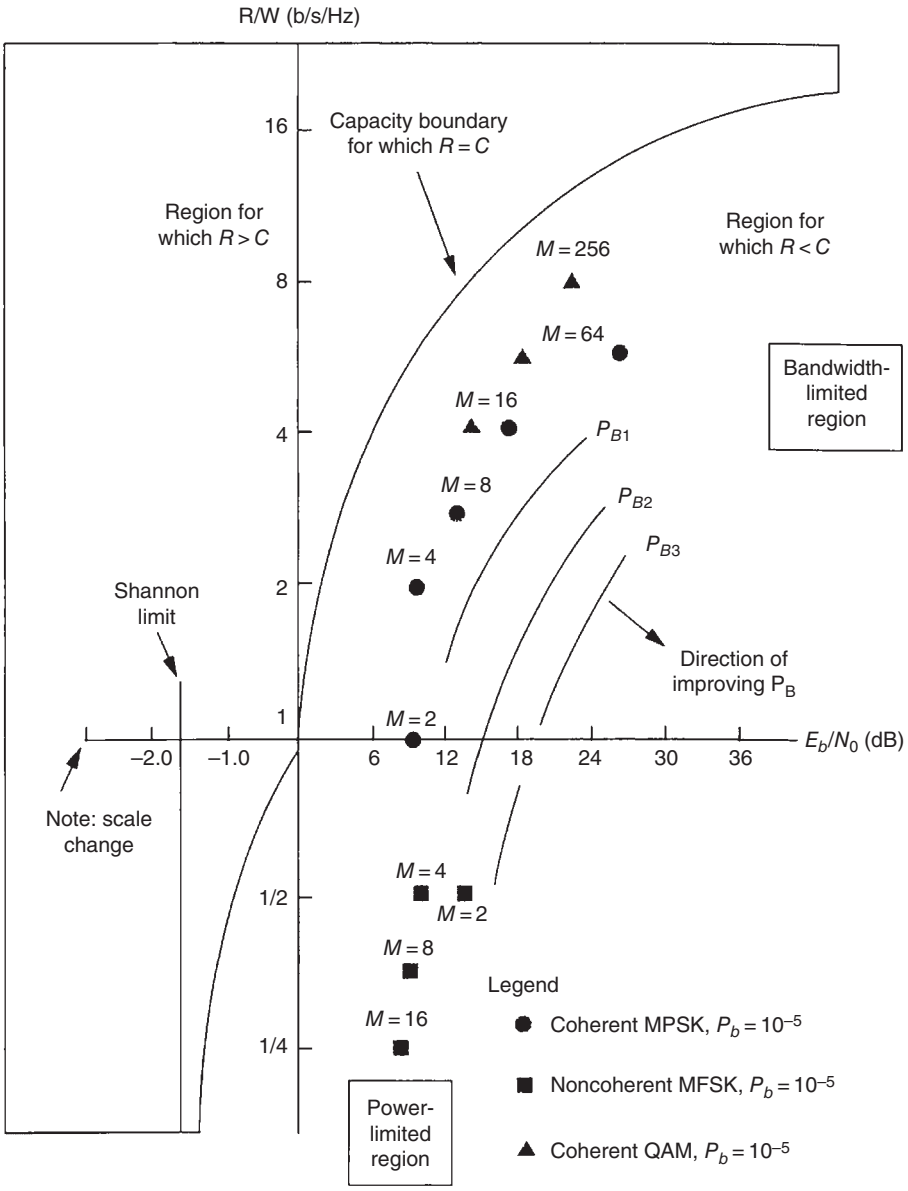


Figure 1.28 Bandwidth efficiency of digital modulation schemes. (*Source*: Xiong 1994 [18]. Reproduced with the permission of IEEE.)

where bps = bits per second. This represents the Shannon limit of bandwidth efficiency. As M increases, so does R/B. However, this comes at the cost of increased E_b/N_0. Figure 1.28 depicts the trade-off between the bandwidth and required E_b/N_0 for multiphase or M-ary (MPSK) modulation schemes. As expected, each modulation scheme leads to its own unique frequency spectrum. For example, Figure 1.29 traces the frequency spectrum for a binary (BPSK) and quadrature (QPSK), and offset QPSK modulation schemes. QPSK and offset QPSK are the most widely used modulation schemes in satellite communication systems.

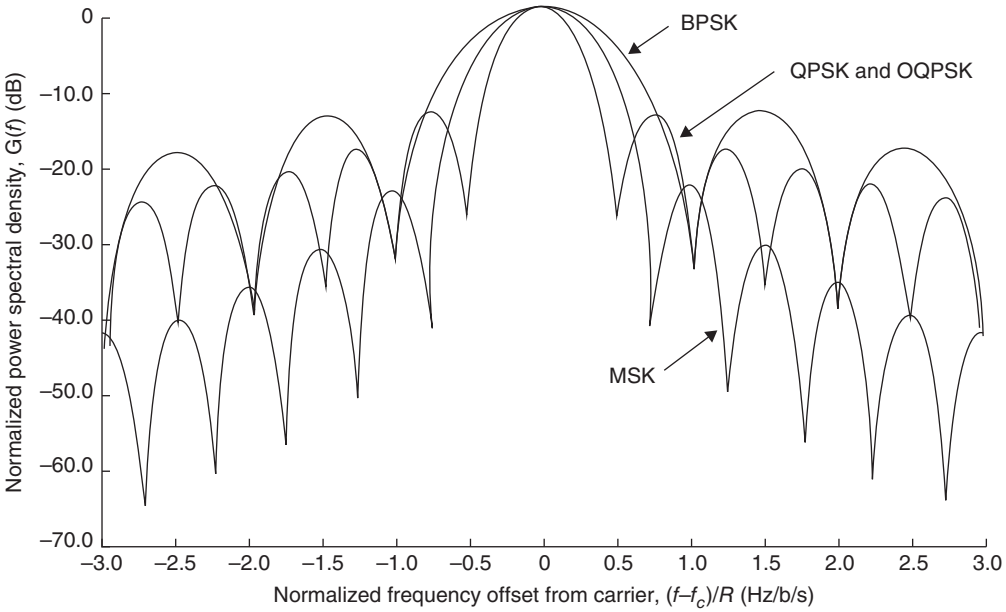


Figure 1.29 Normalized power spectral densities for PSK modulation schemes. (*Source*: Xiong 1994 [18]. Reproduced with the permission of IEEE.)

1.7.7.2 Power Efficiency

Power efficient modulation schemes are best suited for FSK modulation systems. In the binary FSK, the required bandwidth is twice the symbol rate and has a bandwidth efficiency of 0.5 bps/Hz. This is similar to the case of a narrowband analog FM that requires twice the bandwidth of the baseband signal. Again, analogous to the wideband analog FM, the power efficiency is improved by trading off the bandwidth. For MFSK scheme, the minimum bandwidth, required in accordance with the Nyquist criterion, is provided in [17]

$$B = \frac{M}{T_s} = MR_s \tag{1.78}$$

where $R_s(= 1/T_s)$ is the symbol rate.

By using M different orthogonal waveforms, each requiring a bandwidth of $1/T_s$, the bandwidth efficiency of incoherent orthogonal MFSK signals with Nyquist filtering is given by

$$\frac{R}{B} = \frac{\log_2 M}{M} \quad \text{bps/Hz} \tag{1.79}$$

Bandwidth–power trade-offs for MFSK modulation schemes are depicted in Figure 1.28, which shows how it is possible to achieve lower E_b/N_0 at the expense of increased bandwidth.

1.7.7.3 Bandwidth and Power Efficient Modulation Schemes

Table 1.2 describes a wide range of advanced digital modulation schemes [8]. They can be classified into two large categories: constant envelope and non-constant envelope. Generally, the constant envelope class is considered as the most suitable, where the effects of nonlinear amplification in HPAs are an important system consideration.

The PSK schemes (see Figure 1.29) have a constant envelope but discontinuous phase transitions from symbol to symbol. Classic PSK techniques include the BPSK and QPSK. More

Table 1.2 Advanced digital modulation schemes.

Abbreviation	Alternate abbreviation	Definition
ASK	—	Amplitude shift keying
FSK	—	Frequency shift keying (generic name)
BFSK	FSK	Binary frequency shift keying
MFSK	—	M-ary frequency shift keying
PSK	—	Phase shift keying (generic name)
BPSK	2PSK	Binary phase shift keying
DPSK	—	Differential BPSK
QPSK	4PSK	Quadrature phase shift keying
DPQSK		Differential QPSK (with differential demodulation)
DEQPSK		Differential QPSK (with coherent demodulation)
OQPSK	SQPSK	Offset QPSK, staggered QPSK
λ/4-QPSK	—	λ/4-quadrature phase shift keying
λ/4-DQPSK	—	λ/4-differential QPSK
CTPSK	—	λ/4-controlled transition PSK
MPSK	—	M-ary phase shift keying
CPM	—	Continuous phase modulation
SHPM	—	Single-h (modulation index) phase modulation
MHPM	—	Multi-h phase modulation
LREC	—	Rectangular pulse of length L
CPFSK	—	Continuous phase frequency shift keying
MSK	FFSK	Minimum shift keying, fast frequency shift keying
DMSK	—	Differential MSK
MSK	—	Gaussian MSK
SMSK	—	Serial MSK
TFM	—	Timed frequency shift keying
CORPSK	—	Correlative PSK
QAM	—	Quadrature amplitude modulation
SQAM	—	Superposed QAM
Q^2PSK	—	Quadrature phase shift keying
QPSK	—	Differential Q^2PSK
IJF OQPSK	—	Intersymbol-interference jitter-free OQPSK
SQORC		Staggered quadrature-overlapped raised-cosine modulation

Source: Reproduced from [8] with permission of John Wiley and Sons.

generally, modulation schemes with M-ary PSK (MPSK) and M-ary FSK (MFSK) signals can be used with a variety of trade-offs between power and bandwidth efficiencies.

The continuous phase modulation (CPM) schemes have not only a constant envelope but also continuous phase transitions from symbol to symbol. They have less side lobe energy in their spectra in comparison with the PSK schemes. A great variety of CPM schemes can be obtained by varying the modulation index and pulse frequency [8].

1.7.8 Quality of Service and *S/N* Ratio

The quality of a wireless communication link depends not only on its design but also on random effects of the propagation environment, such as rain attenuation, tropospheric and ionospheric scintillation, Faraday rotation, Doppler effect, and antenna pointing errors. As a consequence, transmission performance is defined probabilistically, in terms of the specific quality of the signal over a percentage of time. This has given rise to a variety of standards, some agreed on and others still open for discussion [8]. Conventionally, the quality of the signal is expressed in terms of SNR in the case of analog transmission and bit error rate (BER) in the case of digital transmission. For most applications, the specification calls for SNR to remain above a certain value or BER to remain below a certain value over 99% and 99.9% of the time, averaged over a year. Over these time periods, typical SNR for analog TV is specified to be 53 and 45 dB, respectively. At the present time, no consensus has emerged on the standards for digital transmission, although more and more traffic is moving toward it. With the widespread applications and advances in coding technology, it is becoming possible to have virtually error-free transmission for digital signals. For digital TV, satellite systems in the 14/11 GHz band are targeting a performance objective of one uncorrected error per transmission hour, specifically, BER $\leq 10^{-10}$ or 10^{-11}, depending on user bit rate [8].

It should be noted that SNR depends on two parameters, CNR and signal modulation. CNR is a measure of the efficiency of radio transmission at RF, whereas modulation provides the conversion of CNR into SNR. The variety of modulation and coding techniques provide the trade-offs between CNR and SNR. Systems are always designed to ensure that CNR levels are significantly larger than the threshold value of FM of digital demodulators.

PART III Impact of System Design on the Requirements of Filter Networks

1.8 Communication Channels in a Satellite System

Communication satellites (Figure 1.30) are radio relay stations in space. They serve much the same purpose as the microwave towers one sees spread over populated landmass areas. Satellite communication has evolved since the 1970s and represents a mature niche in the field of telecommunications [5–8, 19].

The satellites receive radio signals transmitted from the ground, amplify them, translate them in frequency, and retransmit them back to the ground. Since the satellites are at high altitude, they can see all the microwave transmitters and receivers (earth stations) on almost one-third of the earth. Thus, they can connect any pair of stations or provide point-to-multipoint services, such as television. The coverage can be extended to any part of the earth via intersatellite links or interconnection with long-distance fiber networks, giving satellite systems an inherent advantage of being insensitive to distance. Satellites are unique in their capability to provide global, seamless, and ubiquitous coverage, including mobile services via handheld units. The frequency plans for commercial satellite systems are listed in Tables 1.3 and 1.4.

It should be noted that frequency allocations are periodically addressed and revised by domestic and international regulatory agencies in order to accommodate new services. To determine the characteristics of communication channels, let us examine the typical block

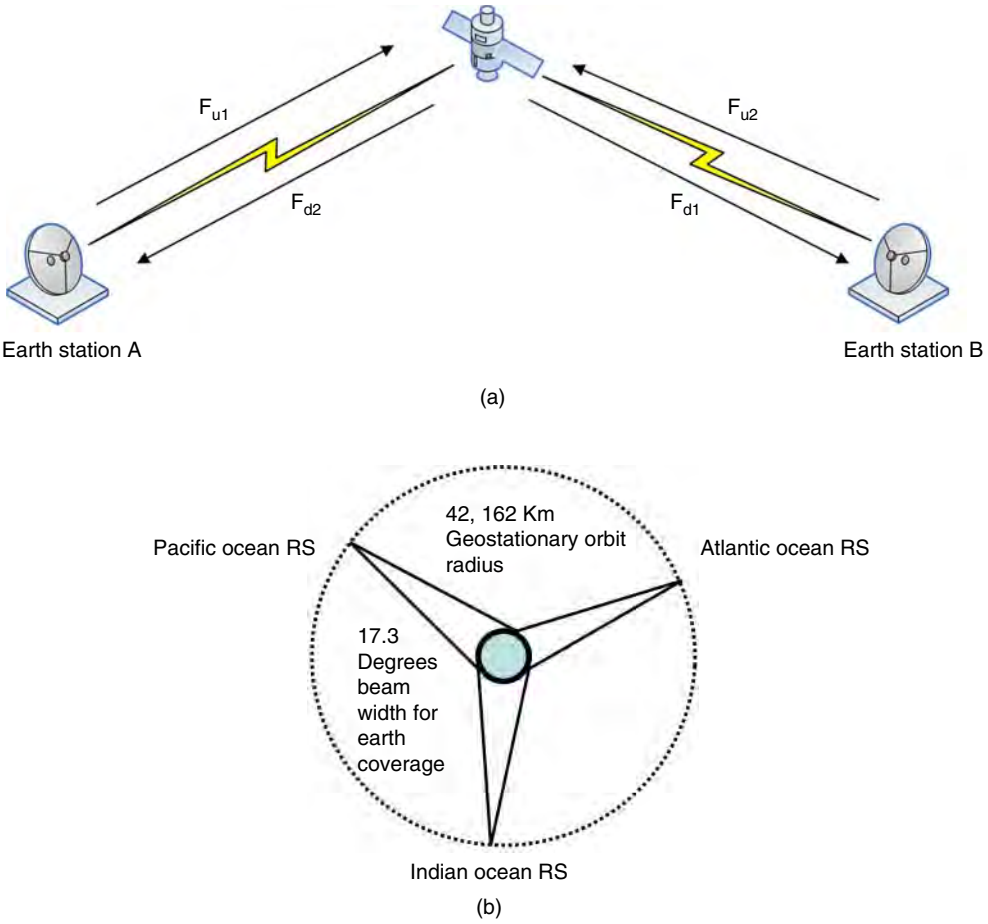


Figure 1.30 Satellite communications: (a) typical satellite link; (b) worldwide coverage using a three-geostationary satellite system.

Table 1.3 Frequency allocations for satellite systems.

Approximate frequency range (GHz)	Letter	Typical usage
1.5–1.6	L	Mobile satellite service (MSS)
2.0–2.7	S	Broadcasting satellite service (BSS)
3.7–7.25	C	Fixed satellite service (FSS)
7.25–8.4	X	Government satellites
10.7–18	Ku	Fixed satellite service (FSS)
18–31	Ka	Fixed satellite service (FSS)
44	Q	Government satellites

Table 1.4 Intersatellite frequency allocations.

Allocation frequency (GHz)	Total bandwidth (MHz)	Satellite services
22.55–23.55	1000	Fixed, mobile, broadcasting
59–64	5000	Fixed, mobile, radio location
126–34	8000	Fixed, mobile, radio location

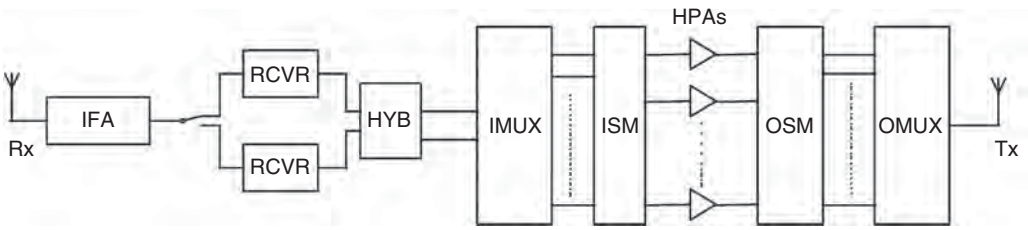


Figure 1.31 Communication subsystem block diagram.

diagram of a communication subsystem of a satellite repeater, as depicted in Figure 1.31. Most commercial satellite systems employ dual orthogonal polarization (linear or circular) providing twofold increase in the available bandwidth. Advance satellite systems employ multiple beams, which allow further reuse of the available bandwidth. However, such advance architectures come at the cost of higher complexity for the spacecraft.

Regardless of the satellite architectures, the block diagram of transponders on a given beam and polarization is essentially the same as the one in Figure 1.31. A receive antenna (Rx) is connected to a wideband filter (IFA - Input Filter Assembly), followed by a low-noise receiver (RCVR). After that, the signal is channelized into its various transponders via an input multiplexing network (IMUX). The allocated frequency band in a satellite system is divided into a number of RF channels, often referred to as *transponders*. Typical channel separation and bandwidths in Ku- and C-band satellite systems are listed in Table 1.5.

Each RF channel is amplified separately and recombined by an output multiplexer (OMUX) network into a composite wideband signal that feeds into the transmit antennas (Tx). Input and output switch matrices (ISM and OSM) are there to provide onboard reconfiguration of traffic flow from one transponder to another, from one beam to another, or various combinations thereof. Often, the ISM and OSM consist of mechanical switches. For ISM, loss is not a

Table 1.5 Typical channel separations and bandwidths in C and Ku bands for fixed satellite services.

Channel separation (MHz)	Desired usable bandwidth (MHz)
27	24
40	36
61	54
80	72

constraint, and, as a consequence, solid-state switches can be utilized to conserve mass and volume. For the OSM, loss is critical, and there is no substitute for mechanical switches by virtue of their very low loss. From the standpoint of the filtering requirements, the block diagram can be separated into three distinct groups: the front end, the channelizer section, and the high-power output circuits. Such a breakdown is typical of most communication repeaters. We now deal with each of the three subgroups separately.

1.8.1 Receive Section

The receive section consists of a wideband input receive filter, the LNA, frequency downconverter, and driver amplifier, as shown in Figure 1.32. For high reliability, satellites invariably employ a redundant receiver that requires a switch prior to the receiver. At the receive antenna, the signal strength is at its lowest level, and therefore, it is imperative to minimize the energy loss prior to the LNA. The wideband receive filter is required to ensure that only the signals in the *allocated frequency spectrum*, *typically comprising* 500 MHz bandwidth, are fed to the LNA and all other signals outside this range are attenuated. For the filters and transmission lines prior to the LNA, the designs are dictated by a need for low insertion loss in the passband. Typical requirement calls for an insertion loss of no more than a few tenths of a decibel.

The frequency downconverter consists of a mixer–local oscillator (LO) assembly. The separation between the transmit and receive frequencies is required to minimize the interference between them. A driver amplifier (DAMP) is used to attain the required power levels prior to channelization and final amplification of the signals. It also allows LNA to operate at low enough power level consistent with the lowest possible noise figure. The hybrid is a type of 3 dB power divider that provides two separate paths for the channelization of the transponders.

1.8.2 The Channelizer Section

A detailed block diagram of the channelizer or input multiplexer (IMUX) section is shown in Figure 1.33.

Once the signal has been amplified by the LNA, the loss in the subsequent equipment is no longer critical. This is due to the fact that the receive system noise temperature is reduced by the gain of the LNA, and the post-LNA losses have less impact (as described in Section 1.5.5). Instead, the design driver is the efficient channelization of the composite signal into its various RF channels or transponders. The criteria for channelization is that the signals incur a minimum

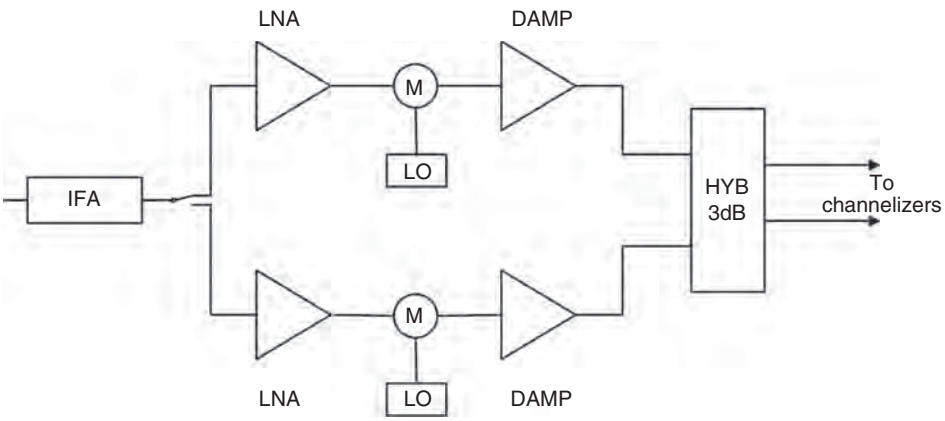


Figure 1.32 Satellite receiver block diagram.

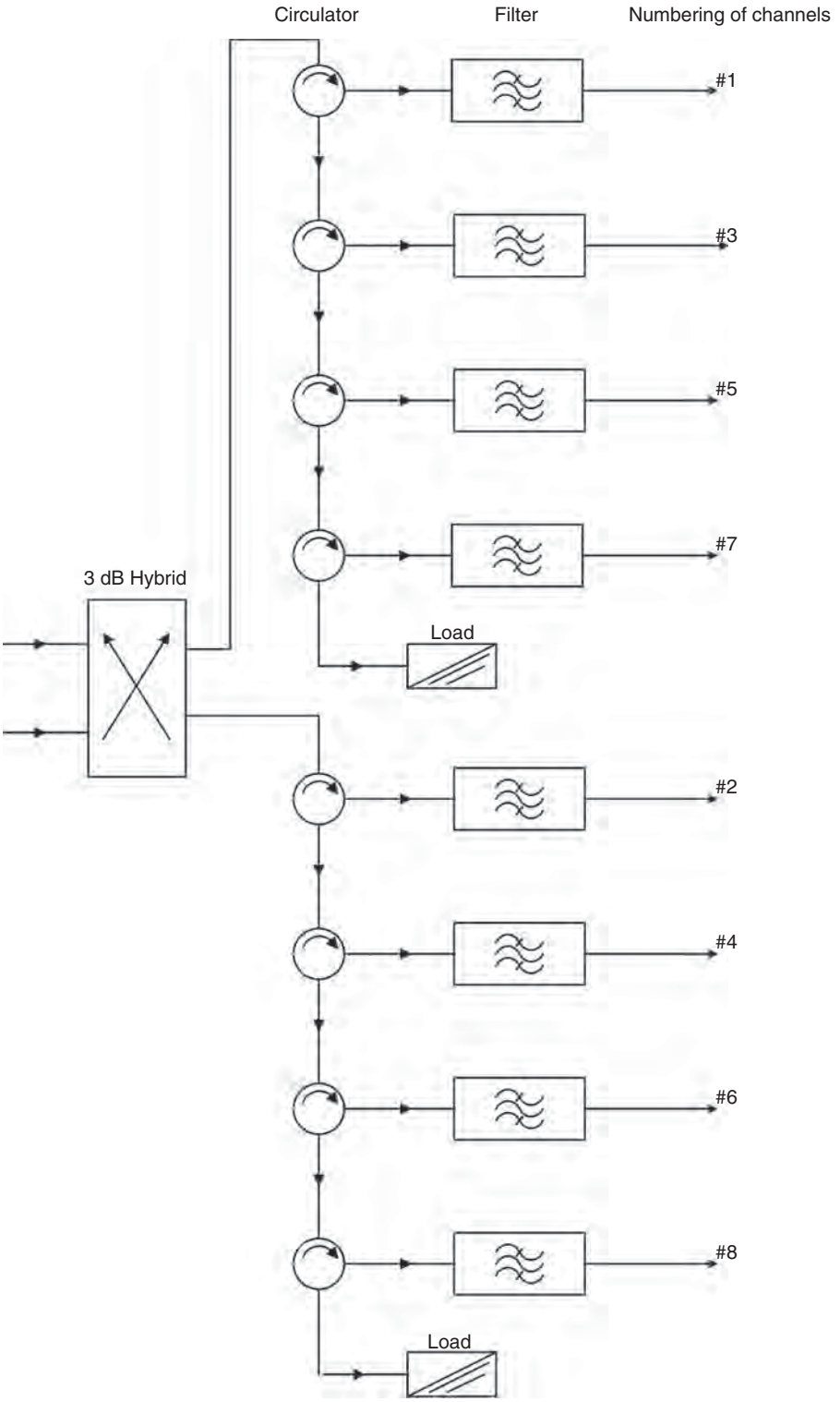


Figure 1.33 Input multiplexing network in a satellite.

of distortion in the passband and, at the same time, provide enough isolation (typically >20 dB) to control interference from other channels. This implies filter characteristics that emulate close to an ideal response as described in Figure 1.17a and b. Any departure from the ideal response means signals would incur distortion, especially at band edges where the amplitude and group delay tends to depart from its desired flat response. A guard band between channels (typically comprising 10% of channel bandwidth) is allocated to allow margin for the nonideal characteristics of practical filters. In essence, channelizing filters largely determine the effective usable bandwidth of each RF channel with a minimum of guard band. For communication channels, typical bandwidth requirements range from 0.3% to 2%. Such narrow band filters incur relatively large transmission deviations and often require a degree of phase and group delay equalization at the expense of additional hardware. As an example of a 6/4 GHz satellite system, the typical specification for a channelizing filter is described in Table 1.6.

The defining feature of the channelizing filter networks is the very stringent out-of-band rejection requirement. It necessitates the use of filters with transmission zeros just outside the passband to achieve a steep isolation response. Although the insertion loss at the band center is not critical, the amplitude variation across the passband is and can be minimized only by employing filters with high unloaded Q. An interesting feature of the channelization scheme is the use of channel dropping ferrite circulators and 3 dB power dividers with coaxial cables to ease the layout of the equipment. Circulators are three-port nonreciprocal devices, constructed from ferrite materials, biased with external magnets, to achieve the desired property of the unidirectional flow of energy. Such devices incur a few tenths of a decibel loss in coaxial or planar structures, and less than a tenth of a decibel loss in the waveguide realization. The reverse isolation is 30–40 dB. Circulators are inherently wideband devices, whose bandwidths range from 10% to 20%. It is possible to optimize the performance over narrower bandwidths. Circulators are widely used in radar and communication systems. As shown in Figure 1.33, use of 3 dB hybrid and channel dropping circulators enable design simplicity and flexibility for the layout of the multiplexing networks. However, this design simplicity comes at the expense of increased loss, which, as described earlier, is of little consequence for this portion of the network. Consequently, the narrowband channelizing filters need to be high-Q structures, whereas the wideband ancillary equipment, namely, circulators, isolators, and hybrid, use the conventional lower-Q but more compact coax designs.

Table 1.6 Typical specifications for channelizing filters in a 6/4 GHz satellite system.

Frequency band	3.7–4.2 GHz
Number of RF channels	12
Channel separation	40 MHz
Passband bandwidth	36 MHz
Rejection over narrowband	10–15 dB at band edge of adjacent copolarized channels, rising to 30–40 dB within 10–15% of channel bandwidth (2–3 MHz) beyond band edge over frequency band 3.7–4.2 GHz
Rejection over wideband	>40–45 dB over receive 5.925–6.425 GHz band
Insertion loss in passband	Not critical
Passband loss variation	<1 dB
Passband relative group delay	<1–2 ns over middle 70%, rising to 20–30 ns toward band edges
Operating temperature range	0–50°C

Narrowband filters incur relatively high group delay variation across its passband. Group delay varies inversely with the absolute bandwidth and is further constrained by the steepness of the required amplitude response of the filter. Group delay can be equalized using all-pass external equalizer network or by using higher-order self-equalized linear phase filters. Either way, group delay equalization increases the design complexity and cost. Most satellite systems employ some degree of equalization. An added benefit of group delay equalization is that it also reduces the amplitude variation across the passband at the expense of a small increase in mid-band loss. Trade-offs for the channelizing filter are critical in establishing the usable passband bandwidth and hence the efficiency in the utilization of the frequency spectrum.

1.8.3 High-Power Amplifiers (HPAs)

HPAs are required to raise the power levels of RF signals prior to their transmission back to earth. System level trade-offs determine the gain and maximum RF power of HPAs to ensure the desired communication capacity for a transponder. The traveling wave tube amplifier (TWTA) is the dominant power amplifier for communications satellites, although many satellites use solid-state power amplifiers (SSPA) as well. Over the years, the available power, reliability, and efficiency of TWTAs have significantly improved, allowing its continued dominance of the satellite market. However, SSPAs tend to exhibit better nonlinear characteristics for moderate power levels for C- and Ku-band satellite systems. As always, system-level trade-offs are required to select the appropriate HPA for a given system. HPAs are power-hungry, requiring its operation at a high efficiency. HPAs are inherently nonlinear devices and their characteristics involve trade-off between efficiency, output power level, and the amount of nonlinearity. Typical characteristics of a TWTA are depicted in Figure 1.34.

To obtain maximum RF power, the amplifier must be operated under the condition of saturation. However, at this power level, the amplifier is highly nonlinear and is suitable only for amplification of a single carrier. For multicarrier operation, it is essential to operate the TWTA in a backoff mode to keep the carrier-to-IM (C/I) ratio to an acceptable level. As described in Section 1.5.6, as the number of carriers in an RF channel increases, the number of IM products grows rapidly. These products fall inside as well as outside the RF channel. It is usual to specify the IM performance by the intercept point defined as the theoretical output power $\overline{IP}$, where

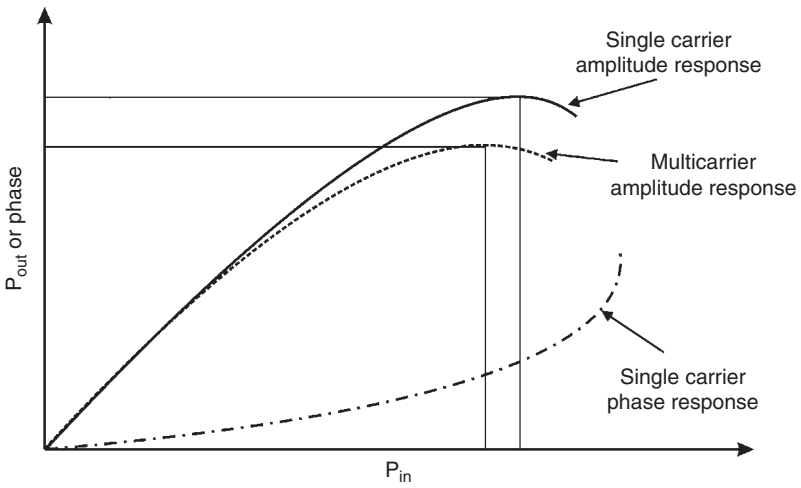


Figure 1.34 Typical characteristics of a high-power amplifier.

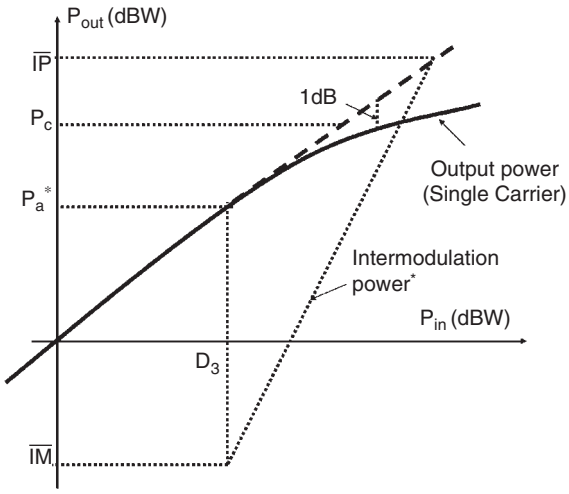


Figure 1.35 Third-order intermodulation (IM) intercept point in amplifiers. (*Source*: Reproduced with the permission of John Wiley and Sons.)

the extrapolated linear single carrier output power is equal to the extrapolated two-carrier IM power, as represented in Figure 1.35. The carrier to third order ($2f_1 \pm f_2$ or $2f_2 \pm f_1$) intermodulation (IM_3) is then given by [8, 19]

$$\frac{C}{IM_3} = 2(\overline{IP} - P_0) \text{ dB}$$

where P_0 is the single-carrier output power level. The value of the intercept point is specified by the tube manufacturer.

This relationship shows that carrier to third-order IM decreases 2 dB for every dB backoff of the output power of the HPA. Also, the greater the intercept point, the larger the value of C/IM. The intercept point is a characteristic of the amplifier and provides a measure of its linearity. There are good reasons for specifying HPA performance with respect to two carriers. It represents the minimum number of carriers to generate an IM product. In addition, third-order products have the highest power level and are critical in determining C/IM performance. The level of odd-order IM products tends to decrease between 10 and 20 dB with increasing order. IM products that fall inside the RF channel can be controlled only by the operating power level and linearity of the HPAs. There exists a trade-off between the efficiency and linearity of HPAs. Highly linear HPAs tend to have lower efficiency and vice versa.

The typical way to control IM is to operate HPA in the linear region of the power curve, which is usually 2–3 dB lower than its saturated power level. For two-carrier operation, HPAs are backed off by 2–3 dB below saturation power level. For multicarrier operation, the backoff can be as much as 10–12 dB to achieve acceptable performance. This represents a significant penalty in the output power to ensure acceptable IM performance. Obviously, linearity of amplifiers is a major issue. Often a "linearizer" is employed to improve the linearity of amplifiers, incurring extra hardware and cost. It should be noted that the prime reason for operating HPAs in a backoff mode for multicarrier operation is to control those IM products that fall inside the RF channel; those outside the RF channel are of no consequence, unless they lie within the transmit or receive frequency bands. IM products falling in the transmit band are attenuated by the high-power output filter or multiplexing networks that follow HPAs. IM products pertaining to the receive band are only possible if we consider products of much higher order. This is due to the fact that we are dealing with carriers in a single channel, not the whole transmit band. Such higher order products are not normally generated in the HPAs, even if they did, the power of such products would be negligible. However, high-power multiplexer and antenna equipment

handling all the channels can generate IM products, referred to as *passive inter-modulation* products or PIM, of lower order that can fall in the receive band. This topic is dealt with in a subsequent section on PIM. The other frequency band of significance corresponds to the second and third harmonics of carriers generated by HPAs. These harmonics, if not suppressed, can interfere with ground-based systems as well as with military- and science-based space systems. The harmonic suppression must be achieved prior to the transmit antenna. Harmonic suppression is part of the output multiplexing subsystem.

1.8.4 Transmitter Section Architecture

The transmit section of the satellite combines the outputs of various high-power channel amplifiers in an output multiplexing network for transmission via a common antenna. Once the signals have been amplified by the final amplifier, conservation of power becomes critical. Consequently, the challenge is how to achieve the lowest loss for each RF channel, preserve the useable bandwidth, and, at the same time, keep the design of multiplexer and antenna subsystems simple. This trade-off has given rise to two alternative multiplexing schemes for satellite systems.

Early satellite systems employed the noncontiguous multiplexing scheme. In such an architecture, alternate channels are combined in the so-called noncontiguous multiplexer, as depicted in Figure 1.36a. The output power of each noncontiguous multiplexer is then combined in a 3 dB hybrid as shown in Figure 1.36b. Power from each input port of the hybrid is divided equally between the two output ports with a phase difference of 90°. Thus, each output port of the hybrid contains all the RF channels at half the power. Such a scheme requires an antenna feed network with two input ports and is called *a dual-mode feed network*. This architecture was employed extensively for satellites in the 1970s and 1980s. Its advantage is the design simplicity of the multiplexer network. The disadvantage is that the architecture requires a more complex beam forming network, incurring a penalty in achievable antenna gain resulting in some loss of EIRP. The alternative is to combine the powers of all the transponders on a given polarization in a single device, the contiguous band multiplexer as shown in Figure 1.36c. Its main advantage is that it allows a simpler beamforming network and it is easier to optimize the antenna gain. Furthermore, owing to its inherent sharper amplitude characteristics, it reduces the effect of multipath in the satellite, resulting in improved RF channel characteristics [16]. Its disadvantage is that the design of such a multiplexer is more complex and it incurs slightly higher loss and group delay variation across the passband of RF channels. Over the years, technology advances have overcome the disadvantage of design complexity. Higher antenna gain achievable with this scheme more than compensates for the disadvantage of slightly higher loss in the multiplexer. As a result, most modern satellite systems employ contiguous multiplexing scheme since it yields better performance in terms of the overall transponder channel characteristics and EIRP. For some applications, especially for systems with narrower-band transponders (~20 MHz), a noncontiguous scheme offers a better design. Comparison of satellite architectures employing alternate multiplexing schemes is described in more depth in Ref. [16].

1.8.4.1 Output Multiplexer

The output multiplexer (OMUX) performs the function that is reverse to that of channelizing section. It combines the power of the individual RF channels to form a single composite signal for transmission back to earth via a transmit antenna. OMUX consists of a number of bandpass filters whose outputs are connected to a common manifold. Each filter corresponds to a particular transponder channel and is optimized to accept the amplified signal within the passband of that channel and reject frequencies corresponding to other transponders. In addition, the

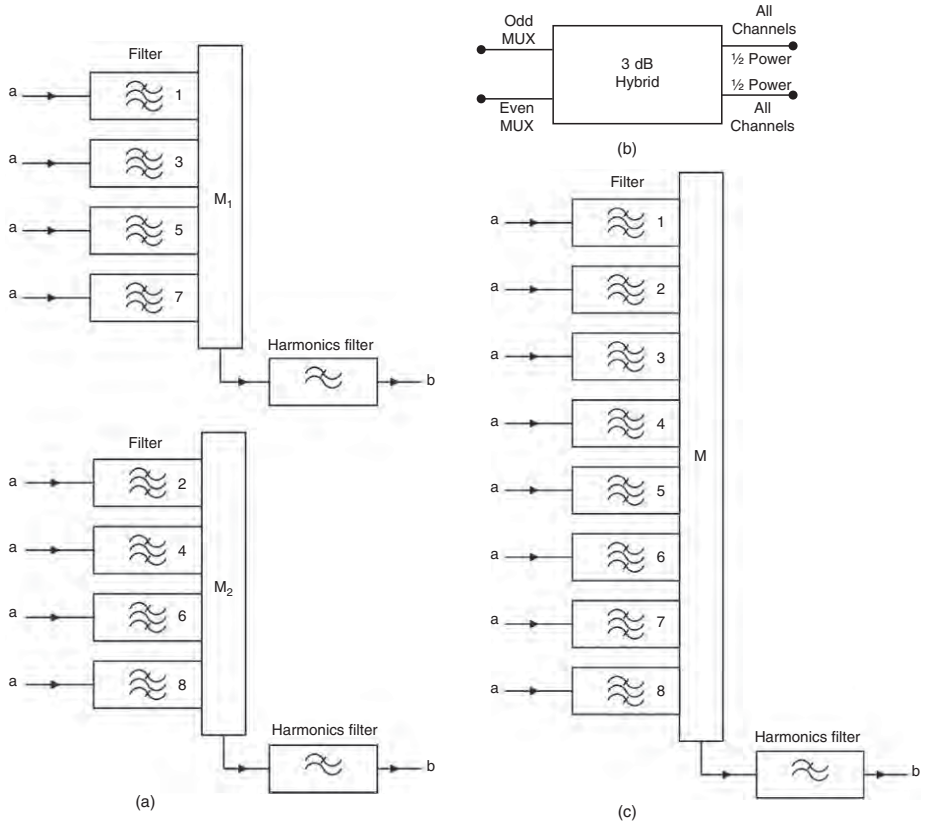


Figure 1.36 Alternative multiplexing schemes in a satellite system: (a) noncontiguous multiplexing scheme; (b) hybrid combiner; (c) contiguous multiplexing scheme.

Table 1.7 Typical specification of a channel in the high-power combining network in a 6/4 GHz satellite system.

Frequency band	3.7–4.2 GHz
Number of RF channels	12
Channel separation	40 MHz
Passband bandwidth	36 MHz
Passband insertion loss	Minimum (typically < 0.2–0.3 dB)
Rejection over narrowband[a)]	Shaped isolation response > 5–10 dB at band edges of adjacent copolarized channels, rising to 20–30 dB within 15–20% of channel bandwidth (3–4 MHz) beyond the band edge and over the frequency band 3.7–4.2 GHz
Rejection over wideband	>30–35 dB over the receive 5.925–6.425 GHz band
Passband relative group delay	<1–2 ns over middle 70%, rising to 10–20 ns toward band edges
Power handling	10–100 W per channel
Operating temperature range	0–50°C

a) The isolation response shape is dictated by the filter and multiplexer technology and the requirements of low-loss and high-power handling capacity.

performance of channel combining filters and overall multiplexer design is further optimized to achieve low loss and provide rejection over the receive band. TWTAs produce not only the desired amplified signals but also IM products and harmonics that must be suppressed. The OMUX provides part of that suppression. The typical requirements imposed for a channel in the OMUX of a 6/4 GHz satellite system are described in Table 1.7.

1.8.4.2 Harmonic Rejection Filters

The purpose of such filters is to provide a high rejection at the second and third harmonic while incurring a minimum of loss in the passband of the transponder. This is accomplished by a low-loss low-pass filter design. The power handling capability of such filters is a critical requirement. Two alternative designs are feasible to meet the power constraints as shown in Figure 1.37. One way is to use a single-harmonic filter connected after the multiplexer as shown in Figure 1.37a. The advantage is that it uses a single filter, incurring a lower mass and volume. However, it must be capable of handling the combined power of all the RF channels on the multiplexer, representing a major disadvantage. The alternative is to use a harmonic filter on a channel by channel basis, as depicted in Figure 1.37b. In this scheme, the harmonic filter is required to handle the power of a single channel and not the combined power of all the channels. Its disadvantage is the increase in the number of harmonic filters required.

Early satellite systems employed a single harmonic filter in OMUX. As the power of satellites has increased, there seems to be a shift toward employing harmonic filters for each individual channel. The particular choice is dictated by system level trade-offs.

1.8.4.3 PIM Requirements for High-Power Output Circuits

All active devices are inherently nonlinear and therefore generate IM products. What is not well understood is that all devices, including passive components such as filters, multiplexers, and antennas, have a degree of nonlinearity and are thus capable of generating IM products [20]. Such IM is known as passive inter-modulation or PIM. It arises owing to imperfections in the structure of materials. For most applications, the level of PIM is sufficiently low, and hence not an issue in system design. It can be an issue if a common antenna is used for the transmission and the reception of signals in a communication system. PIM is also problematic

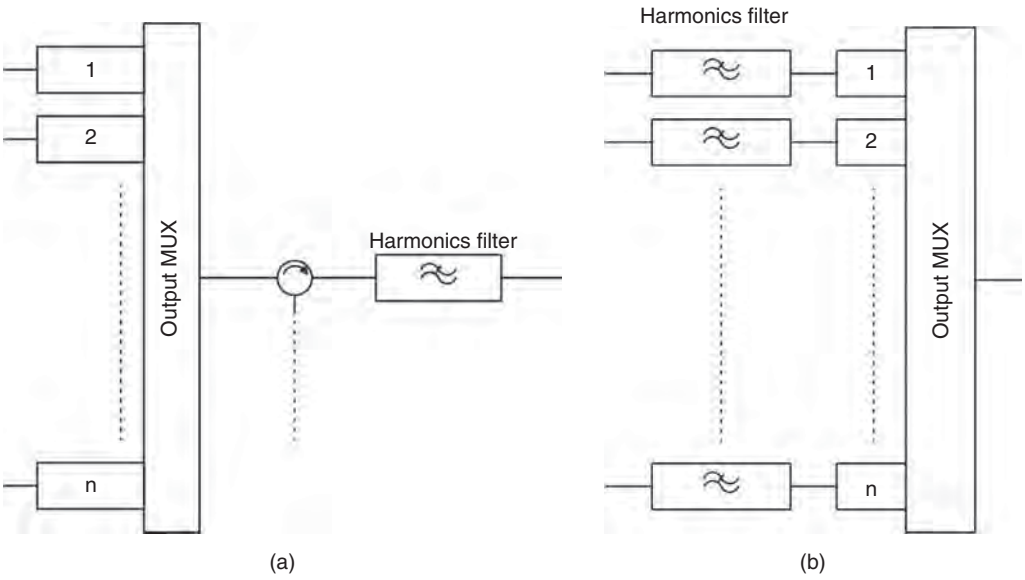


Figure 1.37 High-power output multiplexing network in a satellite, using (a) a single-harmonic filter and (b) individual channel harmonic filters.

if the high-power transmitter and low-level receivers are in close proximity, allowing coupling between the two. For example, the difference in the power levels between the transmit and receive section in a satellite repeater can be as high as 130–140 dB. This implies that the level of PIM should be between 160 and 170 dB or more below the power level of the transmitter to ensure that there is no interference with the low-level received signals. It has implications for the system design, as well as for the quality of materials and work standards for high-power equipment. One important impact on the system design pertains to the allocation of the transmit and receive frequency bands. As described in Section 1.8.3, the third-order IM is the dominant product. Consequently, the separation between the transmit and receive frequency bands should be such so as to avoid at least the third-order IM from falling into the receive band, preferably the fifth-order IM as well. If F_1 and F_2 represent the lower and upper band edge of the transmit frequency band, then

$$\text{Frequency location of IM products} = |m\ F_2 \pm nF_1|$$
$$\text{IM order} = m + n, m \text{ and } n \text{ are integers}$$

The frequency spectrum of IM products generated by two carriers is illustrated in Figure 1.38.

It is evident that even-order IM products are of no consequence since they lie well outside the bands of interest. Also, the odd-order IM products that lie below F_1 are of no consequence since the receive frequency band is invariably chosen to be higher than the transmit band. The reason for this choice is the higher efficiency of equipment at the lower transmit frequency band. The IM products that are of interest are the ones that are above F_2. The frequency locations of the upper odd-order IM products are given by $2F_2 - F_1$ for the third order, $3F_2 - 2F_1$ for the fifth order, and so on. As an example, transmit and receive bands for a 6/4 GHz satellite system are 3.7–4.2 and 5.925–6.425 GHz, respectively. In this case, the minimum odd-order IM product that falls in the receive band is the ninth-order product. This represents a relatively safe separation between the transmit and receive frequency bands. Some of the earlier military satellite systems had much closer separation that allowed third-order PIM to fall in the receive

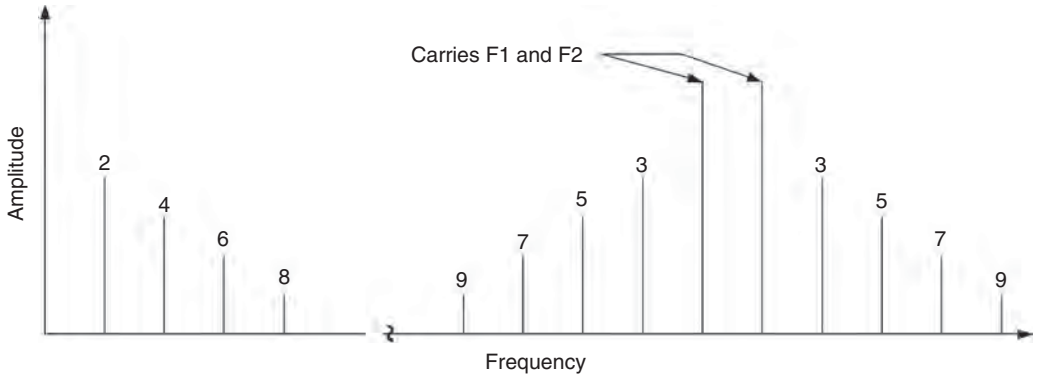


Figure 1.38 Frequency spectrum of IM products generated by two carriers.

band and that proved to be very problematic. It was only then that the system designers were made aware of the PIM phenomenon.

High-power output circuits are often subjected to a specification on PIM. It depends on the antenna design for the communication system. If a common antenna is used for transmission and reception, then any PIM products generated in the receive band can couple via the antenna feed networks into the receive section and interfere with the received signals. This can also occur for systems with independent transmit and receive antennas located in close proximity. A most conservative specification of PIM is to assume antenna Tx–Rx isolation to be zero and the PIM level to be equivalent to that of third-order IM, regardless of the order of the PIM that is permissible to fall into the receive band. Such a specification can put excessive constraints on hardware designers, and seldom can be met or demonstrated with a high degree of confidence. In practical systems, the Tx–Rx isolation is of the order of 20–30 dB for systems with a common antenna. It is much higher (60–100 dB) if separate transmit and receive antennas are used. If the permissible PIM falling into the receive band is of fifth order, its value is typically 10–20 dB lower than that of third-order PIM. Higher-order PIMs go down in level by 10–20 dB with each increasing order. This has been demonstrated by experiments for a number of satellite systems [21]. A typical PIM specification for the 6/4 GHz satellite system with a common antenna is −120 dBm for the third-order IM product.

1.9 RF Filters in Cellular Systems

Cellular radio evolved from mobile communication systems. The earliest application of mobile radio communications was to ocean-going vessels. This was followed by radio services to aircraft, and subsequently, on land for public safety services such as police, fire departments, and hospital ambulances. However, it was not until the early 1990s that mobile communications took hold as a means for personal communications, extending the reach of fixed telephony networks to people just about anywhere: in their backyards, in automobiles, urban centers, or remote regions. It is indeed possible to remain connected anywhere on the planet with access to open space. The conceptual layout of a cellular radio system is shown in Figure 1.39

The geographic area served by a cellular system is divided into small geographic cells, ideally in a hexagonal arrangement. Typically, the cell centers are spaced between 6 and 12 km, depending on the terrain and local climate. The system consists of cell sites, a switching center, and mobile units. Each cell represents a microwave radio repeater. The radio facility is housed

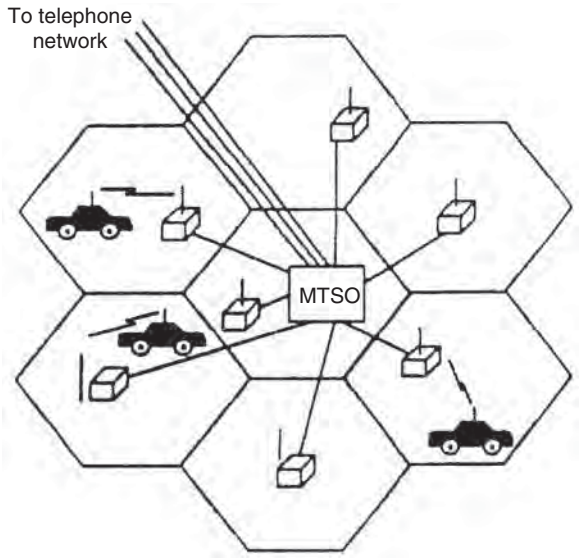


Figure 1.39 Conceptual layout of a cellular system. (*Source*: Freeman 1997 [7]. Reproduced with the permission of John Wiley and Sons.)

in a building or shelter and can connect and control any mobile unit within its own allocated area. The switching facility, referred to as the mobile telephone switching office (MTSO), provides switching and control functions for a group of cell sites. In addition, the MTSO provides connectivity with the public switched telecommunications network (PSTN), adding the distant reach of public telephony to mobile users.

Since the 1990s, cellular systems have seen explosive growth and this exponential growth is likely to continue in the foreseeable future. To meet this high demand, ITU, working with various domestic communication agencies (FCC, European Union, and others) has opened up increasing amount of frequency spectrum for mobile and fixed communication systems. The current state of frequency allocations is summarized in Table 1.8.

In addition, there is ongoing dialogue and meetings among ITU-R, mobile industry players, regulatory agencies, government, manufacturers, and incumbent service providers for further allocations of frequencies. Frequency allocation is a difficult process facing each nation, requiring coordination and harmonization with other systems and services not only within the national boundaries but with international use of the frequency spectrum. The availability of frequency spectrum does not imply that it is easy to get the license for its use. To obtain the license, the operator must successfully negotiate with any existing users of that spectrum, other applicants for the same spectrum, domestic regulatory agencies, and ITU-R. Sometimes it can take years to obtain licenses for the desired spectrum. Frequency allocations are reviewed on a regular basis. For mobile services, the biggest demand is for the frequency band 400–1000 MHz. This lower frequency band is most suitable for providing coverage with lower costs owing to propagation characteristics. L band is a good alternative if frequency spectrum is not available below 1 GHz. Wideband capacity is made available using higher frequency bands.

The architecture of cellular systems requires many trade-offs, especially with respect to the size (coverage area) of the cells, the number of channels allocated to each cell, the layout of the cell sites, and the traffic flow [7]. The total available (one-way) bandwidth is split up into N sets of channel groups. The channels are then allocated to cells, one channel set per cell in a regular pattern that repeats to fill the required number of cells. As N increases, the distance between channel sets (D) increases, reducing the level of interference. As the number of channels sets (N) increases, the number of channels per cell decreases; reducing the system capacity. Selecting

Table 1.8 Summary of frequency allocations from 300 MHz to 30 GHz for ITU region 2.

300–1000 MHz	1000–3000 MHz	3.0–6.0 GHz	6.0–30.0 GHz
300–328.6 (M,F,O)	1427–1525 (M,F,S)	3.3–3.4 (O,M,F)	6.0–7.075 (S,M,F)
335.4–399.9 (M,F)	1700–2690 (M,F,S)	3.4–4.2 (S,M,F)	7.075–7.25 (M,F)
406–430 (M,F,O)		4.4–4.5 (M,F)	7.25–7.85 (S,M,F)
440–470 (M,F,S)		4.5–5 (S,M,F)	7.85–7.9 (M,F)
470–608 (B,M,F)		5.15–5.35 (S,M)	7.9–8.5 (S,M,F)
614–790 (B,M,F)		5.47–5.725 (O,M)	10–10.6 (O,M,F)
790–960 (M,F,B)		5.85–6.0 (S,M,F)	10.7–13.25 (S,M,F)
			14.3–15.35 (S,M,F)
			17.3–19.7 (S, M,F)
			21.4–23.6 (S,M,F)
			24.25–24.64 (O,S,M)
			24.65–29.5 (S,M,F)

Note 1: Letters within brackets represent—M, mobile; F, fixed; S, satellite; B, broadcast; O, others.
The first letter indicates the incumbent service provider, followed by others in order of priority.
Others include radio location, radio astronomy, and radio navigation services
Note 2: Frequency allocations are shared with existing or proposed services requiring ongoing negotiations with ITU-R, (R refers to the radio communication sector) and existing service providers for specific frequency allocations.
Note 3: Spectrum allocations for ITU Regions 1 (Europe) and Region 3 (Asia) are very similar to Region 2 (Americas) allocations
Reference: ITU-R M.2024 (2000), "Summary of spectrum usage survey results."

the optimum number of channel sets is a compromise between capacity and quality. Note that only certain values of N lead to regular repeat patterns without gaps. These are $N = 3,4,7,9$, and 12, and then multiples thereof.

The use of directional antennas can improve performance, allowing smaller cells and higher operating capacity. Regardless of the architecture of the system, each cell site requires highly efficient RF filters to ensure the maximum use of the available frequency spectrum. Figure 1.40 shows the block diagram of a typical RF branching network in a base station using space diversity [22]. Only one transmitter is connected, the other is a redundant backup unit. In the receive direction, two antennas are used to provide space diversity. Use of space diversity is the norm in cellular systems. Thus, one transmit and two receive antennas are required in each direction, for a total of four antennas. Transmit and receive filters connected via a circulator can be replaced by a Diplexer as described in Section 18.5. The circulator coupled network represents the simplest structure but incurs added insertion loss of about 0.1 dB or less per path through the circulator. The diplexer represents the lowest loss structure at the expense of increased design complexity.

Cellular base stations combine the functions of a microwave repeater and a switching network. As a consequence, filtering requirements and constraints are very similar to those encountered in a satellite repeater. Receive and transmit filters must exhibit low loss; elsewhere, loss is not a constraint. Since the majority of cell sites are spread throughout the urban areas, real estate is expensive and, as a consequence, size of equipment is also a constraint. For extra high-reliability systems, both frequency diversity and space diversity are employed as described in Figure 1.41 [22]. It allows for frequency diversity and space diversity in each direction.

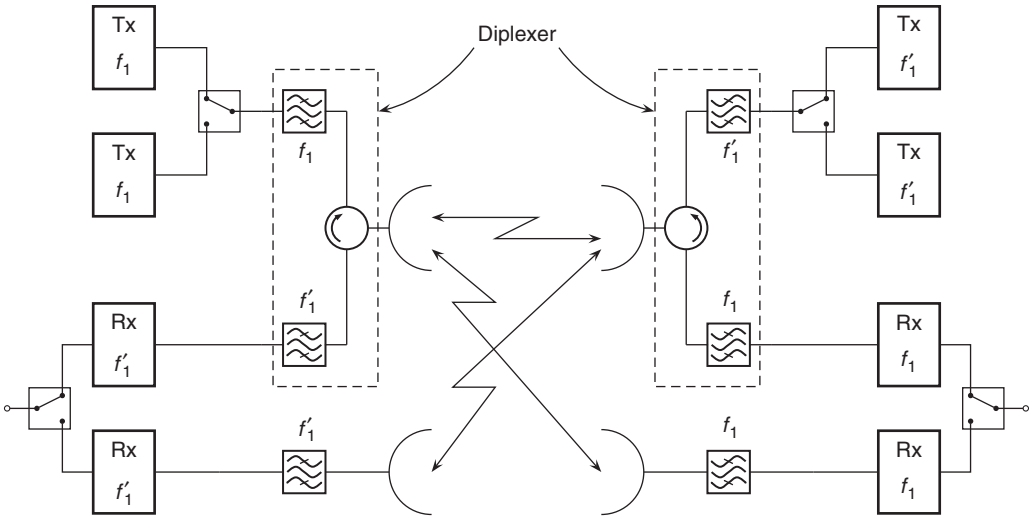


Figure 1.40 Block diagram of RF branching network with space diversity.

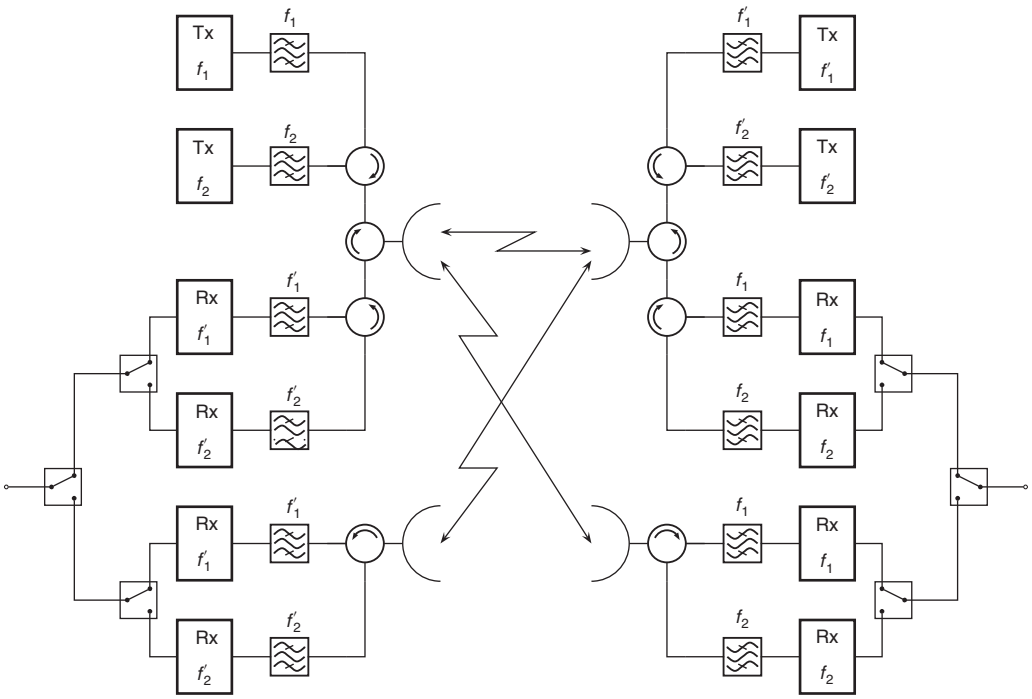


Figure 1.41 Block diagram of RF branching network with frequency and space diversity.

Cellular systems typically assign a single frequency band or channel to a base station to cover a defined geographical area requiring a single passband as shown in Figures 1.40 and 1.41. Owing to increasing demands for more capacity and services, additional frequency bands are being made available to meet this demand. Simple configurations of a two- and three-passband filter network using circulators are described in Figure 1.42. Such a configuration can be extended

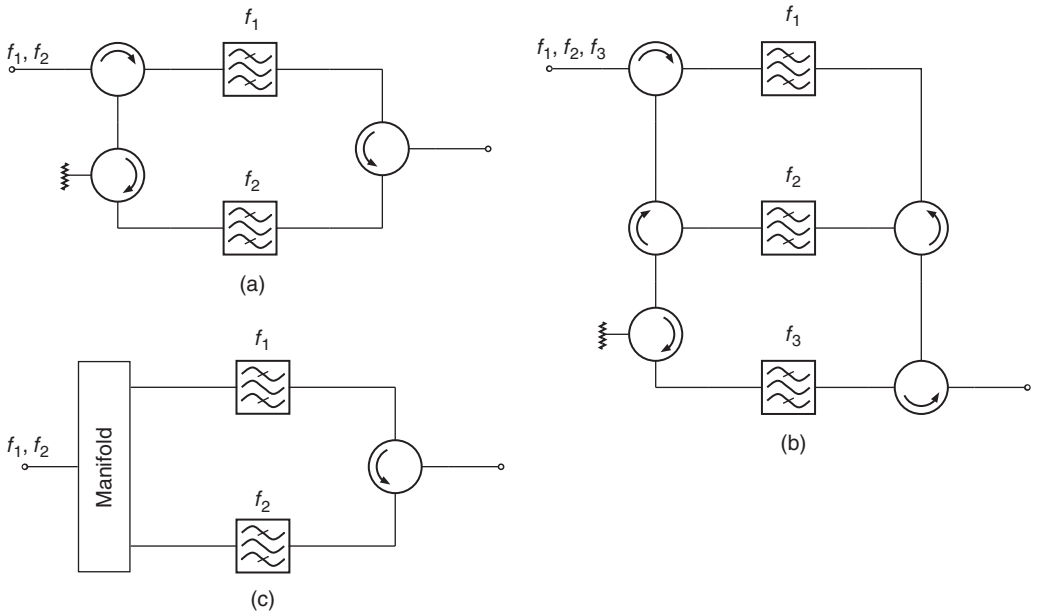


Figure 1.42 (a) Block diagram of circulator-coupled two-channel filter. (b) Three-channel filter. (c) Block diagram of a two-channel filter using manifold coupling at one end and circulator coupling at the other end.

readily to any number of passbands. Its key advantages are (1) simplicity, (2) modularity, and (3) ease of manufacture of single filters to desired performance. The disadvantages include additional hardware in terms of circulators (widely available for commercial applications), insertion loss penalty of approximately 0.3 dB (worst case) per additional passband, and slightly bigger footprint. This configuration is similar to the circulator-coupled multiplexers described in Chapter 18. The increased loss can be dealt with by increasing the power of transmitters by a corresponding amount; for example, to compensate for 0.3 dB loss, power needs to be increased by 7%. This is not an issue for the transmitters; however, it does imply increased usage of electricity that must be taken into account in the overall design of the base stations. The insertion loss penalty can be reduced by a factor of 2 by replacing the input circulators by a manifold-coupled multiplexer as described in Figure 1.42c at the expense of modularity.

Incorporation of additional bands in the base stations has spurred the development of multiband filters, the subject of the new Chapter 21 in this edition. Such filters are finding possible application in radio astronomy and military, suppressing narrowband jamming or interference signals that might occur within the wider usable passband width.

1.10 Ultra Wideband (UWB) Wireless Communication

Ultra Wideband (UWB) traces its origins in the evolution of spread spectrum techniques used for secure communications in military applications. UWB service occupies a very wide bandwidth and shares the frequency spectrum with many other existing services. UWB is a digital pulse wireless technology that can be used to transmit large amounts of digital data over a wide spectrum of frequency bands over short distances with very low power consumption [23–26]. In 1960s–1990s, UWB technology was restricted to defense applications under classified programs for highly secure communication. Since the year 2000, UWB spectrum has been

progressively released globally for short-range wireless communications. Major trends driving short-range wireless in general and UWB in particular are [24]:

- The growing demand for wireless data capability in portable devices
- Paucity of frequency spectrum, segmented and licensed by regulatory agencies
- The growth of high-speed wired access to the Internet in businesses, homes, and public places
- Shrinking costs of microprocessors and signal processing techniques

UWB enables high-speed data transfer over short-range wireless connection as well as applications in low data rate radar and imaging systems. Developments in high-speed microprocessors and fast switching techniques have made UWB technology commercially viable for short-range low-cost consumer communications. The key advantages of UWB can be summed up as [26]:

- Ability to share the frequency spectrum with many users
- Large communication capacity, excellent candidate for high data rate wireless applications
- Low-power density usually below environment noise ensures communication security
- Low sensitivity to multipath effects, ability to carry signals through doors and other obstacles that tend to reflect signals at more limited bandwidths
- Carrier-less UWB transmissions simplify transceiver architecture, thereby reducing cost and implementation time

In 2002, the FCC allowed UWB communication in the 3.1–10.6 GHz band for unlicensed operation having a −10 dB bandwidth greater than 500 MHz and a maximum EIRP spectral density of −41.3 dBm/MHz or 75 nW/MHz. The FCC defined UWB as any signal that occupies more than 500 MHz bandwidth in the 3.1–10.6 GHz band. In addition, FCC and subsequently ECC (Electronic Communications Committee in Europe) issued regulations for the radiated power for the UWB systems as described in Tables 1.9 and 1.10 [25, 26].

UWB is an overlay technology and is susceptible to in band interference from existing narrowband communication bands. Devices using UWB technology transmit ultra-low-power radio signals with short electrical pulses in the picosecond range across all frequencies simultaneously. UWB power levels are required to be below that of noise emissions allowed for electronic equipment. Due to low transmission power and large bandwidth, UWB systems are susceptible to interference from other existing systems. On the other hand, UWB systems are harmless to other communication systems because of their low transmit power.

A key component to remove the unwanted signals and noise from the UWB transmission systems is the bandpass filter. It must ensure that any interference from outside of UWB masks is kept well below the specified power levels for UWB transmissions. The existing communication

Table 1.9 FCC emission masks for indoor and outdoor UWB devices [26].

Frequency range (MHz)	Indoor EIRP (dBm/MHz)	Outdoor EIRP (dBm/MHz)
960–1610	−75.3	−75.3
1610–1990	−53.3	−63.3
1990–3100	−51.3	−61.3
3100–10600	−41.3	−41.3
Above 10600	−51.3	−51.3

Table 1.10 ECC emission masks for the UWB devices in Europe [26].

Frequency range (MHz)	EIRP (dBm/MHz)
Below 1600	−90
1600–2700	−85
2700–3400	−70
3400–3800	−80
3800–6000	−70
6000–8500	−41.3
8500–10600	−65
Above 10600	−85

systems employ narrowband channels (typically <2% bandwidth); for such channels, wide ranging microwave filter design and technologies have been developed over the last four decades. For UWB application, we require filters exceeding 100% bandwidth, a challenging proposition. Such wideband filters are realized using coupled transmission line structures. Many of the wideband filter design techniques are described in Ref. [26].

1.11 Impact of System Requirements on RF Filter Specifications

The emergence and widespread applications of wireless communication systems has stretched the required performance of all components and subsystems that constitute such systems. The scarcity of the available frequency spectrum for wireless communications has put new demands in signal processing and filtering in order to maximize the efficiency of the usage of the frequency spectrum. The factors affecting the specification of RF filters can be categorized as follows:

Frequency plan
Interference environment
Modulation–demodulation schemes
Operating environment
Location of filters in the communication link
HPA characteristics
Limitations of microwave filter technology
Each of these factors can be related to the appropriate area of system design.

1.11.1 Frequency Plan

As described in Section 1.2, the available frequency spectrum for commercially viable wireless communication is a limited natural resource. It is imperative that the communication systems make the most efficient use of the available bandwidth. RF filters perform the function of protecting the overall frequency band from outside interference, efficient channelization of the available bandwidth into various transponders to meet the traffic demands, and efficient power combining to form a common feed, minimizing the cost of the antennas. Consequently, the achievable performance of practical filters largely determines the effective usable bandwidth

of the allocated frequency spectrum. At higher frequencies, filter design is more complex and has higher sensitivity to manufacturing tolerances. Choice of narrower channel bandwidths implies higher loss, higher transmission deviations, and sensitivity. Typical allocation of guard band between RF channels is 10%. It implies an efficiency of 90% in the usage of the available frequency band. However, the allocation of 10% bandwidth to guard band is arbitrary; it merely provides a window to conduct trade-offs between passband performance within the constraint of providing a minimum of isolation at the band edges and beyond of the adjacent channels. For practical systems, performance can be optimized to within 1 dB of amplitude variation, and 2–3 ns of relative group delay over 70–80% of the passband while providing 20–30 dB (or greater) rejection over the adjacent channels. This implies minimal distortion of signals occupying the middle 80% of the passband, with distortion rising toward the band edges. System level trade-offs are critical in optimizing the filter design to maximize the usable bandwidth and hence the efficiency of the channel [16, 22]. The ultimate limit on the achievable bandwidth is imposed by the state of microwave filter technology and cost.

1.11.2 Interference Environment

The interference environment of an RF channel in a communication system is described in Section 1.3. The prime purpose of RF filters is to channelize (or combine) the allocated bandwidth with a minimum of guard band while maintaining a good passband performance. The near-out-of-band isolation requirements are dictated almost entirely by the interference from adjacent (copolarized) channels. The level of interference depends on the energy spectrum of signals in the adjacent channels. It is possible, therefore, to specify isolation in accordance with the shape of the interfering spectrum. This specification is the most severe constraint imposed on channel filters (see Sections 1.8.2 and 1.8.4). It drives the trade-off between the out-of-band amplitude response and the in-band amplitude and group delay variations. The level of such deviations depends on the filter technology, design complexity, and cost. Interference from other sources, such as cross-polarized channels, rain fades, multipath effects, and ionospheric scintillation, fall inside the RF channel and are therefore controlled by the system design. IM interference caused by nonlinear HPAs (Sections 1.5.6 and 1.8.3) is controlled by the operating power level of HPAs.

1.11.3 Modulation Schemes

Modulation schemes play a critical role in maximizing communication capacity of RF channels for different traffic requirements. They also provide a trade-off between power and bandwidth required for transmission of signals. Most systems require a high degree of flexibility in being able to handle different types of modulation schemes as well as a varying number of carriers in an RF channel. Such channel characteristics are best achieved by microwave filters exhibiting equiripple passbands and stopbands; such a response represents the optimal approach to a "brick wall" amplitude response. If necessary, a degree of phase and group delay equalization can be incorporated to achieve improved passband characteristics.

1.11.4 HPA Characteristics

HPAs are inherently wideband and nonlinear devices. In addition to amplifying the RF power, HPAs also generate harmonics of the fundamental frequency and wideband thermal noise. Another consequence of nonlinear characteristics is the generation of IM noise. As a result, HPAs control specification of harmonic suppression and wide out-of-band rejection response, as described in Sections 1.8.2 and 1.8.4.

1.11.5 Location of RF Filter in the Communication Link

A communication link consists of a cascade of transmitters and receivers. RF filters are required after HPAs, prior to LNAs and in between LNA and HPAs as described in Section 1.5. The impact on the specification of filters, based on their location in the communication link, is as follows:

Filter prior to LNA	Low loss
Filters after HPAs	Low loss and high power handling capability
Filters between LNA and HPAs	High selectivity

In addition to electrical requirements, there are always the constraints of size, mass, and cost. Low-loss and high-power handling capability imply larger size. However, if loss is not a constraint, then it is possible to trade-off smaller size at the expense of higher loss.

1.11.6 Operating Environment

For wireless communication, there are two operating environments: outer space or terrestrial space.

1.11.6.1 Space Environment

Key drivers for space applications are mass, size, reliability, and performance. In addition, the equipment must withstand the launch environment, radiation in outer space, and higher temperature range. These requirements have been drivers for the many innovations in filter technology over the last three decades. Typically, the microwave filters used in space are composed of air or dielectric-loaded metal structures and are therefore immune to the radiation environment. However, if high-temperature superconductors (HTS) are adopted to realize such filters, the radiation environment becomes a design constraint, which can be overcome by enclosing filter networks in a metal enclosure of appropriate thickness. Another phenomenon that is peculiar to this environment is multipactor breakdown, which can occur in the hard vacuum of outer space. It is dealt with in Chapter 20.

1.11.6.2 Terrestrial Environment

For terrestrial systems, cost is the main driver, along with the requirement of relatively large quantities of filters. This has spawned innovations in the area of volume production.

1.11.7 Limitations of Microwave Filter Technology

All electrical components dissipate energy; minimization of this energy loss is invariably a key design parameter. In communication systems, bulk of filter applications require narrow bandwidth filters. Energy loss in a filter depends upon two factors: (1) electrical conductivity of metal structures and loss tangent (reciprocal of unloaded Q) of dielectric materials used to realize the microwave filter and (2) the percentage bandwidth of the filter. The narrower the bandwidth of channels is, the higher are the losses incurred in the filter. Bandwidth requirements are a system constraint and is the raison d'etre for using filter networks. Silver has the highest electrical conductivity among metals and is employed as the plating material of microwave structures to achieve the lowest loss. Unplated aluminum or copper structures are often employed as well where specifications can be met without silver plating. For high-power applications, dissipated heat has to be conducted or radiated away to maintain the desired temperature environment, that implies that the materials must also have acceptable thermal

conductivity. All communication equipment has a typical specification of 0 to 50°C as the operating temperature environment. For the narrowband filters at microwave frequencies, this is critical since their performance is very sensitive with respect to any movement of the centre frequency. This implies that the materials must also exhibit excellent thermal stability, often in conflict with the requirements of high electrical and thermal conductivities. In other words, materials used for microwave filters must have low loss (high unloaded Q), good thermal conductivity, and excellent thermal stability.

For any microwave structure, especially narrow band devices such as filters, there is always a trade-off between size (and mass) and the dissipation of energy. The bigger the structure, the lower the loss and vice versa. Another factor that is crucial in determining the size of filters is the dielectric constant of materials. Advances in the quality of dielectric materials over the past few decades, namely, high dielectric constant, low loss (high Q), and excellent thermal stability has made dielectric loaded filters the preferred choice for many applications. Advances in materials are the key in realizing improved performance in ever more compact structures. Some of these trade-offs are discussed in more detail in Chapters 11 and 21.

Microwave filters are distributed structures. This allows realization of microwave filters in a large variety of topologies, as described in Chapters 9–14. Although most filters employ the dominant mode, it is possible to employ higher order propagation modes to realize high Qs at the cost of larger size; such features are well suited for high-power applications. Microwave filter design offers many choices and trade-offs and has given rise to a range of technologies in terms of its implementation. This is highlighted in Table 1.11.

Table 1.11 Microwave filter technologies for communications systems.

Application	Frequency band	Key requirements	Baseline filter technology	Competing technologies
IF Filters for Communication Systems at Large	100–600 MHz	Large number, small size, low power	Lumped Element Networks Microstrip	SAW, MMIC Active
Cellular Systems Handsets Infrastructure	800–900 MHz	Mass production/cost sensitive	Coax Dielectric Resonators Tunable Dielectric	SAW, MMIC, Active Tunable Filters, SIW
PCS/DCS Systems Handsets Infrastructure	2–5 GHz	Mass production/cost sensitive	Coax MIC	Dielectric, MMIC, SIW, Active Tunable Filters
Satellite Systems UHF	300 MHz	Low loss, high and low power, size and mass sensitive	Pressurized Coax	Dielectric
Mobile	1.5 GHz	Low loss, high and low power, size and mass sensitive	Coax Dielectric	Finline/SS/MIC SAW
FSS	4/6, 7/8, 12/14, 11/17 GHz	Low loss, high and low power, size and mass sensitive	Coax, WG, Dielectric	
Multimedia	20/30, 40/60 GHz	Low loss, high and low power, size and mass sensitive	WG, Dielectric	
LMDS	28, 38, 42 GHz	Large Volume, size and mass sensitive	Coax WG	Finline

1.12 Impact of Satellite and Cellular Communications on Filter Technology

In the 1970s and 1980s, the advent of satellite communications was the key R&D driver for microwave filter networks [27]. For space applications, the size and mass of onboard equipment have a critical impact on cost. In addition, power generation in space is costly, and thus, low loss for high-power equipment is equally critical. Lastly, the equipment must be ultra-reliable to operate in space, without failure, for the lifetime of the spacecraft. This led to many advances and innovations in the field of filters and multiplexing networks. These advances included the development of dual- and triple-mode waveguide filters with arbitrary amplitude and phase response, dielectric resonator filters, contiguous and non-contiguous multiplexing networks, surface acoustic wave (SAW) filters, HTS filters, and a variety of coaxial, finline, and microstrip filters. The evolution of widespread wireless cellular communication systems in the 1990s capitalized on this R&D and pushed the envelope further in developing materials and processes for large-scale and cost-effective production techniques for microwave filters for ground-based systems.

Computer-aided design and tuning by EM techniques have played a major role in this endeavor. In an interesting twist of circumstances, advances in production techniques are now being adopted for space microwave equipment to achieve lower costs. This is being driven by the new era of communication systems consisting of constellation of a large number (hundreds to thousands) of low earth orbiting satellites (LEOs). In addition, new frequency bands including millimeter waves are being made available to accommodate the ongoing demand for more and more services using wireless communication, be it satellite based or ground based cellular systems. This demand and new frequency bands will continue to fuel R & D for microwave filters. Some of the promising areas include continuing advances in dielectric materials, tunable filter technology and 3 D manufacturing techniques.

Summary

This chapter is devoted to an overview of communication systems, highlighting the relationship between the communication channel parameters and other elements of the system. The intent here is to provide the reader with sufficient background to be able to appreciate the critical role and requirements of RF filters in communication systems.

Starting with a description of a model of a communication system, the key questions addressed in this chapter include radio spectrum, concept of information, noise and the interference environment, and system considerations in the design of a communication channel.

The concept of information and its transmission over a medium is explored by reviewing the seminal work of Shannon on information theory, highlighting the fundamental trade-offs between signal power, noise in the system, and available bandwidth. This leads to a discussion of trade-offs between power, bandwidths, noise minimization, and antenna performance required to establish a radio link. The topic of noise is considered with emphasis on the thermal noise, and how it sets a fundamental noise floor for the communication channel. Brief overviews of analog and digital modulation are included. The chapter concludes with a discussion of system requirements and the specifications of microwave filter networks in satellite and cellular communication systems.

References

1 Members of Technical Staff (1971) *Transmission Systems for Communications*, 4th edn (revised), Bell Telephone Laboratories, Inc.

2 Taub, H. and Schilling, D.L. (1980) *Principles of Communications Systems*, 3rd edn, McGraw-Hill, New York.
3 Schwartz, M. (1980) *Information, Transmission, Modulation, and Noise*, 3rd edn, McGraw-Hill, New York.
4 Meyers, R.A. (ed.) (1989) *Encyclopedia of Telecommunications*, Academic Press, San Diego.
5 Gordon, G.D. and Morgan, W.L. (1993) *Principles of Communications Satellites*, Wiley, New York.
6 Haykin, S. (2001) *Communication Systems*, 4th edn, Wiley, New York.
7 Freeman, R.L. (1997) *Radio System Design for Telecommunications*, 2nd edn, Wiley, New York.
8 ITU (2002) *Handbook on Satellite Communications*, 3rd edn, Wiley, New York.
9 a Shannon, C.E. (1948) A mathematical theory of communications. *Bell System Technical Journal*, **27** (3), 379–623 and issue 4, 623–656, 1948; b Shannon, C.E. (1949) Communication in the presence of noise. *Proceedings of the IRE*, **37**, 10–21.
10 Krauss, J.N.D. (1950) *Antennas*, McGraw-Hill, New York.
11 Lathi, B.P. (1965) *Signals, Systems and Communication*, Wiley, New York.
12 Cross, T.G. (1966) Intermodulation noise in FM systems due to transmission deviations and AM/FM conversion. *Bell System Technical Journal*, **45**, 1749–1773.
13 Bennett, W.R., Curtis, H.E., and Rice, S.O. (1955) Interchannel interference in FM and PM systems under noise loading conditions. *Bell System Technical Journal*, **34**, 601–636.
14 Garrison, G.J. (1968) Intermodulation distortion in frequency-division multiplex FM systems—a tutorial summary. *IEEE Transactions on Communication Technology*, **16** (2), 289–303.
15 Kudsia, C.M. and O'Donovan, M.V. (1974) *Microwave Filters for Communications Systems*, Artech House, Norwood, MA.
16 Tong, R. and Kudsia, C. (1984) Enhanced performance and increased EIRP in communications satellites using contiguous multiplexers. Proceedings of the 10th AIAA Communication Satellite Systems Conference, Orlando, FL, March 19–22.
17 Sklar, B. (1993) Defining, designing, and evaluating digital communication systems. *IEEE Communications Magazine*, **11**, 91–101.
18 Xiong, F. (1994) Modem technologies in satellite communications. *IEEE Communications Magazine*, **8**, 84–98.
19 Maral, G. and Bousquet, M. (2002) *Satellite Communications Systems*, 4th edn, Wiley, New York.
20 Chapman, R.C., *et al.* (1976) Hidden threat: multicarrier passive component IM generation. Paper 76–296, AIAA/CASI 6th Communications Satellite Systems Conference, Montreal, April 5–8, 1976.
21 Kudsia, C. and Fiedzuisko, J. (1989) High power passive equipment for satellite applications. IEEE MTT-S Workshop Proceedings, Long Beach, CA, June 13–15, 1989.
22 Manning, T. (1999) *Microwave Radio Transmission Design Guide*, Artech House.
23 Roberto Aiello, G. and Rogerson, G.D. (2003) Ultra-wideband wireless systems. *IEEE Microwave Magazine*, **2**, 36–47.
24 Bedell, P. (2005) *Wireless Crash Course*, 2nd edn, McGraw Hill.
25 Wentzloff, D.D. *et al.* (2005) System design considerations for ultra-wideband communication. *IEEE Communication Magazine*, **8**, 114–121.
26 Zhu, L., Sun, S., and Li, R. (2012) *Microwave Bandpass Filters for Wideband Communications*, Wiley.
27 Kudsia, C., Cameron, R., and Tang, W.C. (1992) Innovations in microwave filters and multiplexing networks for communications satellite systems. *IEEE Transactions on Microwave Theory and Techniques*, **40**, 1133–1149.

Appendix 1A

Intermodulation Distortion Summary

Table 1A.1 Direct transmission deviations.

Transmission deviation	Order of distortion	NPR at top modulating frequency (without preemphasis)
Parabolic gain, A_2 (dB/MHz^2)	Third	$\frac{1.72 \times 10^4}{A_2^4 \sigma^4 f_m^4}$
Cubic gain, A_3 (dB/MHz^3)	Second	$\frac{33.6}{A_3^2 \sigma^2 f_m^4}$
Quartic gain, A_4 (dB/MHz^4)	Third	$\frac{6.32}{A_3^2 \sigma^2 f_m^4}$
Linear delay, B_1 (ns/MHz)	Second	$\frac{10^6}{\pi^2 B_1^2 \sigma^2 f_m^2}$
	Third	$\frac{7.5 \times 10^5}{\pi^4 B_1^4 \sigma^4 f_m^4}$
Parabolic delay, B_2 (ns/MHz^2)	Third	$\frac{7.5 \times 10^5}{\pi^2 B_2^2 \sigma^4 f_m^2}$
Cubic delay, B_3 (ns/MHz^3)	Second	$\frac{1.19 \times 10^6}{\pi^2 B_2^3 \sigma^2 f_m^6}$

Key: σ—multichannel RMS frequency deviation in MHz; f_m—top modulating frequency in MHz; A_n—is the nth-order amplitude coefficient in dB/(MHz)^n; B_n—is the nth-order group delay coefficient in ns/(MHz)^n; NPR—noise power ratio; a measure of intermodulation (IM) noise, given by ratio of white noise power spectral density to IM noise power spectral density in a communication channel.

Table 1A.2 Coupled transmission deviations—significant distortion terms only.

Transmission deviation	Order of distortion	NPR at top modulating frequency (without preemphasis)
Linear gain + linear AM/PM, A_1 (dB/MHz) + K_{P1}(•/dB/MHz)	Second	$\frac{3.28 \times 10^3}{K_{P1}^2 A_1^2 \sigma^2 f_m^2}$
Parabolic gain + constant AM/PM, A_2 (dB/MHz^2) + K_{P0}(•/dB)	Second	$\frac{3.28 \times 10^3}{K_{P0}^2 A_2^2 \sigma_m^2 f_m^2}$
Quartic gain + constant AM/PM, A_4 (dB/MHz^4) + K_{P0}(•/dB)	Second	$\frac{9.75 \times 10^2}{K_{P0}^2 A_4^2 \sigma^2 f_m^6}$
Linear delay + linear AM/PM, B_1 (ns/MHz) + K_{P1}(•/dB/MHz)	Second	$\frac{1.73 \times 10^3}{\pi^2 K_{P1}^2 B_1^2 \sigma^2 f_m^4}$
Parabolic delay + constant AM/PM, B_2 (ns/MHz^2) + K_{P0}(•/dB)	Second	$\frac{4.33 \times 10^7}{\pi^2 K_{P0}^2 B_2^2 \sigma^2 f_m^4}$

2

Fundamentals of Circuit Theory Approximation

Engineering problems represent exercises in approximation. The same can be said when we analyze any physical phenomena. At the most fundamental level, quantum mechanics can describe the properties of matter only in terms of probabilities. The point being made here is that with any engineering problem, there is always a tradeoff lurking somewhere, whether explicit or implicit. With that understanding, we can always optimize the parameters that are available within the constraints of a given system. The principles that unify communication theory and circuit theory approximations are explained in this chapter; it highlights the essential assumptions and the success of the frequency analysis approach that we take for granted in analyzing electrical networks.

2.1 Linear Systems

In nature, most systems, if not all, are nonlinear. Such systems are highly complex and not amenable to any general solution. Each problem and boundary condition requires an individual solution. However, most engineering problems can be approximated by linear systems over a limited range; herein lies our first assumption. It implies that we must ensure that all equipment used in the communication system satisfy the assumption of linearity over the operating range. Analysis of linear systems is well developed and covered in many textbooks [1].

2.1.1 Concept of Linearity

Any system can be described as a black box that receives an input signal (or driving function), processes this signal, and then produces an output signal or response function, as depicted in Figure 2.1.

A system whose output is proportional to its input constitutes a linear system. This does not imply that the response is linearly proportional to the driving signal, although this constitutes a special case of linearity. Thus, if $r(t)$ is the response to $f(t)$, then $kr(t)$ is the response to $kf(t)$, where k is an arbitrary constant. Symbolically, this implies

$$\begin{aligned} f(t) &\rightarrow r(t) \\ kf(t) &\rightarrow kr(t) \end{aligned} \tag{2.1}$$

However, linearity implies more than this. The defining property of a linear system is the law of superposition. This property states that if several causes are acting on a linear system, then the total effect on the system can be determined by considering each cause separately and assuming

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

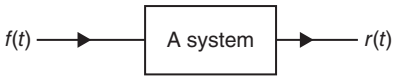


Figure 2.1 Block diagram of an electrical system.

all other causes to be zero. The total effect is then the sum of all the individual effects. Mathematically, this is expressed as follows:

$$\begin{aligned} f_1(t) &\rightarrow r_1(t) \\ f_2(t) &\rightarrow r_2(t) \\ f_1(t) + f_2(t) &\rightarrow r_1(t) + r_2(t) \end{aligned} \tag{2.2}$$

In more general terms, this can be stated as

$$a_1 f_1(t) + a_2 f_2(t) \rightarrow a_1 r_1(t) + a_2 r_2(t) \tag{2.3}$$

whereas this condition holds irrespective of the choice of a_1, f_1, a_2, f_2, and so on. Equation (2.3) represents the defining equation of a linear system.

We have restricted the discussion to a single-input/single-output system. This can be extended to multiple-input/multiple-output systems.

2.2 Classification of Systems

For linear and nonlinear systems alike, there can be a further classification:

- Time-invariant and time-variant systems
- Lumped-parameter and distributed-parameter systems
- Instantaneous (memoryless) and dynamic (with memory) systems
- Analog and digital systems

2.2.1 Time-Invariant and Time-Variant Systems

Systems whose parameters do not change with time are called *constant-parameter* or *time-invariant systems*. Linear time-invariant systems are characterized by linear equations with constant coefficients. Systems whose parameters change with time are called *variable-parameter* or *time-variant systems*. Linear time-variant systems are characterized by linear equations with time-dependent coefficients. Strictly speaking, all physical objects or systems change over time when measured in eons. However, when systems are considered over finite durations of time as measured in tens or even hundreds of years, most fall into the category of time-invariant systems. All our discussions are confined to such systems.

2.2.2 Lumped and Distributed Systems

Any system is a collection of individual elements that are interconnected in a specific way. A system is deemed as lumped if all the elements in the system are assumed to behave as lumped elements. In a lumped model, the energy in the system is considered to be stored in distinct isolated elements such as inductors and capacitors in an electrical system and springs in a mechanical system. Also, it is assumed that the disturbance initiated at any point is instantaneously propagated to any other point in the system. In electrical systems, this implies that the dimensions of the elements are very small compared with the wavelength of the signals to

be transmitted. For a lumped electrical component, the voltage across its terminals or current flowing through it is related by a lumped parameter.

In contrast to lumped-element systems, distributed systems are characterized by elements where the energy is distributed over the element's physical dimensions. For these systems, voltages and currents are distributed over the physical dimensions of the various elements. This implies that we need to deal with the variables of time as well as space. Moreover, in a distributed system, the propagation of energy takes a finite time to travel from a point in the system to any other point. The descriptive equations for distributed systems are partial differential equations, in contrast to the ordinary differential equations that describe lumped systems.

2.2.3 Instantaneous and Dynamic Systems

A system's output at any given instant depends on the past input. However, for many systems, the past history is irrelevant in determining the response. Such systems are said to be instantaneous or memoryless systems. More precisely, a system is said to be instantaneous if its output at any instant of time depends on its inputs at the same instant, and not on any past or future values of the inputs. A system whose output does depend on past inputs is said to be a system with memory or dynamic. In this book, we focus on instantaneous systems.

2.2.4 Analog and Digital Systems

A signal whose amplitude can take any value in a continuous range is an analog signal. Systems that deal with analog signals are referred to as *analog systems*. Signals, whether continuous or discrete, can also be represented by a finite number of values, referred to as *digital signals*. Systems that deal with digital signals are referred to as *digital systems*.

2.3 Evolution of Electrical Circuits: A Historical Perspective

In 1800, Volta published a paper that described, for the first time, the continuous flow of current by connecting a wire across a pair of dissimilar metals immersed in brine or a weak-acid electrolyte [2]. This important discovery represented the beginning of electrical circuits and circuit theory. Davy, Ampere, Coulomb, Ohm, and many others extended the concepts of electrical circuits to applications of direct current. Such circuits are characterized by currents, voltages, and electromotive force, which are all constants and not functions of time.

Two major discoveries on which the theory of time-varying currents and electromagnetic fields is based were made by Michael Faraday and James Clerk Maxwell. Faraday's discovery was experimental. He demonstrated that a changing magnetic field induces an electric field. Maxwell's discovery was intuitive and a leap in imagination. He postulated, by analogy, that a changing electric field would induce a magnetic field. In addition, he introduced the concept of displacement current so that time-varying electric fields play the same role as conduction currents and combined the two into a total current that he showed to be continuous. This generalization cannot be overstated. It allows propagation of electric disturbances through dielectric media, including air and free space. Without this generalization, electrical disturbance can only propagate through a conductive medium. Maxwell went one step further and calculated the velocity of such electromagnetic disturbance and found it to be very close to the velocity of light, as measured by Hippolyte Fizeau in 1849. He concluded that the two phenomena are

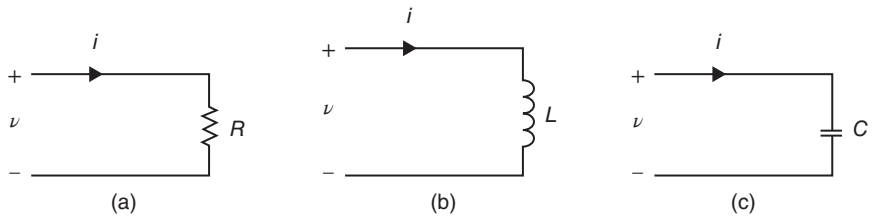


Figure 2.2 Relationships between the idealized elements in an electrical network: (a) a resistor; (b) an inductor; (c) a capacitor.

the same; specifically, light waves also represent an electromagnetic disturbance. The profound nature of this intuitive leap of imagination is evident in the following quotation [3]:

> It should be remembered that at this time no one had ever wittingly generated or detected electromagnetic waves. The concept was completely new, as was the notion of a displacement current. To link light to these hypothetical phenomena was a flash of brilliance seldom equalled in the history of science. It was not to be until eight years after Maxwell's death that these hypotheses would receive substantiation through the experiments of Hertz.

Maxwell combined all the fundamental relations of electrodynamics into four compact, beautiful equations, the Maxwell equations. This set of equations is the foundation of the whole field of electrical engineering.

2.3.1 Circuit Elements

Passive lumped electric circuits are characterized by three basic elements: capacitor (C), inductor (L), and resistor (R). The electromagnetic energy shuttles back and forth between the magnetic field of an inductor and the electric field of the capacitor while it is being gradually dissipated as thermal energy in the resistor. The electrical variables that occur in networks are voltages or potential v and currents i, where v and i are the instantaneous values of the voltage and current, respectively. The elements R, L, and C are defined in terms of two variables, v and i, as described in Figure 2.2. The elements are assumed to be linear, lumped, finite, passive, bilateral (LLFPB), and time invariant. Systems composed of such elements conform to linear equations (algebraic, differential, or difference equations) with constant coefficients.

When the values of variables v and i are related in terms of the element values, time becomes an important factor. Currents and voltages must therefore be expressed as functions of time $i(t)$ and $v(t)$, respectively. Therefore, the applied signals are also functions of time, usually voltages. The response of the network to these signals is a voltage or a current in some other part of the network. It should be noted that we are still dealing with time-invariant systems, that is, the values of R, L, and C remain independent of time.

2.4 Network Equation of Linear Systems in the Time Domain

To analyze electrical networks, we can write loop equations to obtain the relationships for the complete network. Figure 2.3 reflects a simple network with the applied signal $v(t)$, and the required response is the current flowing in the circuit.

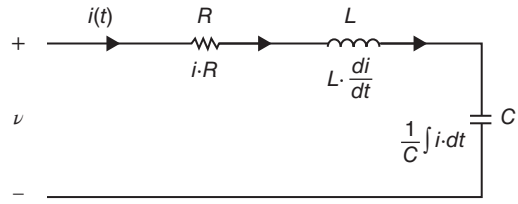


Figure 2.3 Time-domain analysis of a series *RLC* circuit.

The system equation is

$$iR + L\frac{di}{dt} + \frac{1}{C}\int i\,dt = v(t) \tag{2.4}$$

By differentiating this expression in a second-order differential equation, we obtain

$$L\frac{d^2i}{dt^2} + R\frac{di}{dt} + \frac{1}{C}i(t) = \frac{dv(t)}{dt} \tag{2.5}$$

Equation (2.5) indicates the nature of a differential equation characterizing lumped, linear, and time-invariant systems. For systems comprising of many more elements, branches, and loops, the general form of the differential equation is

$$\begin{aligned} &a_m\frac{d^m i}{dt^m} + a_{m-1}\frac{d^{m-1}}{dt^{m-1}} + \cdots + a_i\frac{di}{dt} + a_0 i(t) \\ &\quad = b_n\frac{d^n f}{dt^n} + b_{n-1}\frac{d^{n-1}}{dt^{n-1}} + \cdots + b_1\frac{df}{dt} + b_0 f(t) \end{aligned} \tag{2.6}$$

For convenience, we can replace differential operator d/dt by algebraic operator p. Consequently, the operation of integration with respect to t is replaced by $1/p$:

$$\begin{aligned} &p \equiv \frac{d}{dt}, \quad pf(t) = \frac{df(t)}{dt}, \ldots, p^m f(t) = \frac{d^m f}{dt^m} \\ &\frac{1}{p}f(t) = \int f(t)dt, \ldots \end{aligned}$$

With this notation, the system equation is

$$\begin{aligned} &(a_m p^m + a_{m-1}p^{m-1} + \cdots + a_1 p + a_0)i(t) \\ &\quad = (b_n p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + b_0)f(t) \end{aligned} \tag{2.7}$$

This is a linear differential equation with constant coefficients and is solved by classical methods. The solution consists of two components: a source-free component and the component due to source. For a stable system, the source-free component always decays with time. For this reason, the source-free component is designated as the *transient component*, while the component due to source is called the *steady-state component*. The source-free component is obtained by setting the driving function to zero. It is independent of the driving function and depends only on the nature of the system. This response is a characteristic of the system and is also known as the *natural response* of the system. We are concerned with circuits where the transient component has vanished, leaving the system response entirely due to the steady-state source. For most applications, the transient response vanishes quickly and is not a factor in analyzing circuits powered by steady-state sources.

Several methods are used to obtain the solution of a linear differential equation due to a source. Assuming that the driving function $f(t)$ yields only a finite number of derivatives, we can represent the component due to source to be a linear combination of $f(t)$ and all its higher derivatives [1]:

$$i(t) = h_1 f(t) + h_2\frac{df}{dt} + \cdots + h_{r+1}\frac{d^r f}{dt^r} \tag{2.8}$$

where only the first r derivatives of $f(t)$ are independent. Coefficients $h_1, h_2, \ldots, h_r$ are obtained by substituting equation (2.8) in equation (2.7) and equating the coefficients with similar terms on both sides. The solutions of such a set of simultaneous differential equations are very time-consuming and sometimes quite formidable. The problem of solving differential equations can be overcome by limiting the problem to a reduced class of functions of time for which simple differential relationships exist. This leads to the simpler world of frequency domain.

2.5 Network Equation of Linear Systems in the Frequency-Domain Exponential Driving Function

If the driving function $f(t)$ is an exponential function of time, it simplifies the analysis of electrical circuits. For an exponential function, all the derivatives are also exponential functions of the same form. None of the derivatives is independent. Thus, the component due to source must be the same form as the driving function itself. Thus, if driving function $f(t)$ is exponential, the response function $i(t)$ is given by

$$\begin{aligned} i(t) &= hf(t) \\ f(t) &= e^{st} \\ i(t) &= h\,e^{st} \end{aligned} \tag{2.9}$$

where s is the complex frequency variable $\sigma + j\omega$. The value of h is obtained by substituting equation (2.9) in equation (2.7) as follows:

$$\begin{aligned} &(a_m p^m + a_{m-1}p^{m-1} + \cdots + a_1 p + a_0)h\,e^{st} \\ &\quad = (b_n p^n + b_{n-1}p^{n-1} + \cdots + b_1 p + b_0)e^{st} \end{aligned} \tag{2.10}$$

Since

$$pe^{st} = \frac{d}{dt}e^{st} = se^{st} \quad \text{and} \quad p^r e^{st} = s^r e^{st}$$

equation (2.10) can be simplified to

$$\begin{aligned} &(a_m s^m + a_{m-1}s^{m-1} + \cdots + a_1 s + a_0)h \\ &\quad = (b_n s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0) \end{aligned}$$

therefore

$$h = \frac{b_n s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0}{a_m s^m + a_{m-1}s^{m-1} + \cdots + a_1 s + a_0} \tag{2.11}$$

The quantity h represents the system transfer function, usually denoted as $H(s)$, which characterizes the system in the frequency domain. This result forms the basis of frequency-domain analysis.

2.5.1 Complex Frequency Variable

At this point, we will describe the significance of the complex frequency variable in the analysis of electrical circuits. The complex frequency variable is given by $s = \sigma + j\omega$. A signal $f(t)$ may then be represented as

$$\begin{aligned} f(t) &= e^{st} \\ &= e^{\sigma t}(\cos \omega t + j\ \sin \omega t) \end{aligned}$$

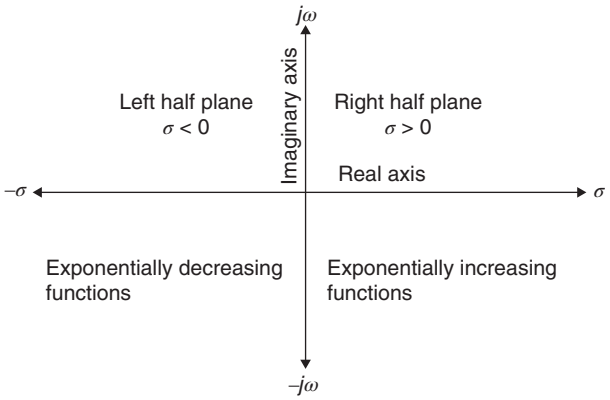


Figure 2.4 Complex frequency plane.

As a consequence, when s is complex, function e^{st} has real and imaginary parts. Furthermore, $e^{\sigma t}\cos\omega t$ and $e^{\sigma t}\sin\omega t$ represent functions oscillating at angular frequency ω, with the amplitude increasing or decreasing exponentially. This depends on whether σ is positive or negative. In the graphical representation of the complex frequency variable in Figure 2.4, the horizontal axis is the real axis represented by σ, and the vertical axis is the imaginary axis represented by ω.

Thus, the imaginary axis is associated with real frequencies. It should be mentioned that the left half of the plane represents exponentially decaying functions ($\sigma < 0$) and the right half plane represents exponentially growing functions ($\sigma > 0$). Each point on the complex frequency plane corresponds to a certain mode of the exponential function. This leads to the interesting question as to the significance of the frequencies that lie along the negative portion of the $j\omega$ axis. What is a negative frequency? By definition, frequency is an inherently positive quantity. The confusion arises because we are defining the frequency not as a number of cycles per second of a particular waveform, but as an index of the exponential function. As a result, the negative frequencies are associated with negative exponents. Signals of negative and positive frequencies can be combined to obtain real functions as follows: $e^{j\omega t} + e^{-j\omega t} = 2\cos\omega t$.

Similarly, the signals of two complex conjugate frequencies can form real signals. Any real function of time encountered in practice can be expressed as a continuous sum of the exponential functions that occur in pairs with complex conjugate frequencies. Thus, the use of a complex variable for the exponential functions represents a complete and powerful tool for analyzing electrical systems.

2.5.2 Transfer Function

A system with an exponential input is displayed in Figure 2.5.

In the frequency-domain analysis, the response of a system is characterized by its transfer function defined as

$$H(s) = \frac{\text{response function}}{\text{driving function}} \tag{2.12}$$

For an exponential driving function, the response of a linear time-invariant system is given by $H(s)e^{st}$. By definition, the transfer function of a system is defined with respect to an exponential

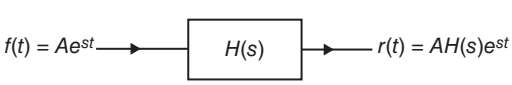


Figure 2.5 Electrical network with an exponential input function.

driving function. Equally, it should be remembered that the concept of transfer function is meaningful for linear systems alone. Furthermore, a system or electrical network can consist of a single-port, two-port, or multiport terminals. Thus there are a number of positions where a driving function can be applied and where a response can be observed. The transfer function is not a unique quantity unless the input and output terminals are specified. Filter networks are typically two-port networks, and $H(s)$ is usually specified in terms of the output to input voltage ratio. Using the more familiar notation,

$$H(s) = \frac{b_n s^n + b_{n-1} s^{n-1} + \cdots + b_1 s + b_0}{a_m s^m + a_{m-1} s^{m-1} + \cdots + a_1 s + a_0} = \frac{n(s)}{d(s)} \tag{2.13}$$

where $n(s)$ and $d(s)$ are the numerator and denominator polynomials of the transfer function, respectively. In addition, this establishes the link between the system's time-domain and frequency-domain representations. For the system in differential equation (2.10), $H(s)$ is a rational function in s. It is a quotient of polynomials with real coefficients. This is a very significant result and represents a huge simplification in the analysis of electrical circuits.

2.5.3 Signal Representation by Continuous Exponentials

In order to exploit the simplicity of equation (2.13) to analyze real-life systems, two conditions must be satisfied:

1) Any signal can be represented in terms of exponential waveforms.
2) The responses due to the individual exponential functions can be added up to obtain the overall response.

If these two conditions are satisfied, then the simple expression of the transfer function $H(s)$ is all that is needed to obtain the response of the system for any signal in the frequency domain.

The second condition represents the principle of superposition of linear systems. As stated earlier, a fundamental premise of network analysis is that it is confined to linear systems. We need to preserve the linearity of the systems and deal with the inevitable nonlinearities as distortions that must be controlled within specified bounds. As a consequence, for linear systems, the success of the frequency-domain analysis approach depends on our ability to represent a given function of time as a sum of various exponential functions to satisfy the first condition. Nearly all practical signals or waveforms are classified as either periodic or nonperiodic. A periodic function is expressed as a sum of discrete exponential functions by the well-known Fourier series. Similarly, a nonperiodic function is represented as a continuous sum of exponential functions by making use of the well-known Fourier and Laplace transforms. As a consequence, any function, periodic or nonperiodic, is amenable to representation as a sum of exponential functions. More detailed and rigorous analyses of signals and systems can be found in the text by Lathi [1]. Suffice it to say here that the basis of frequency-domain analysis rests in the assumptions of the system's linearity and the power of exponential functions to represent signals and waveforms encountered in practice.

2.5.4 Transfer Functions of Electrical Networks

Lumped electrical networks are composed of individual elements R, L, and C. An exponential voltage applied across each of these elements produces an exponential response (i.e., current) as a consequence of the properties of linearity and the time invariance of these elements. Thus, if current i through the element is given by $i(t) = e^{st}$, then the voltage v across the element is

Table 2.1 Properties of positive real functions.

1) An impedance or admittance function has the form Z(s) or $Y(s) = n(s)/d(s)$, where the coefficients in the numerator and denominator polynomials are real and positive; thus Z(s) is real when s is real and complex poles and zeros of Z(s) occur in conjugate pairs
2) The poles and zeros of Z(s) have either negative or zero real parts
3) The poles of Z(s) on the imaginary axis must be simple, and their residues must be real and positive
4) The degrees of the numerator and the denominator polynomials in Z(s) differ, at most, by 1; thus the number of finite poles and finite zeros of Z(s) differ, at most, by 1; Z(s) has neither multiple poles nor zeros at the origin

$v(t) = Z(s)e^{st}$, where $Z(s) = v(t)/i(t)$ is defined as the impedance of the element. If we consider the exponential voltage as the driving function and current as the resulting response through the element, then $v(t) = e^{st}$ and $i(t) = Y(s)e^{st}$, where $Y(s) = i(t)/v(t)$ is defined as the admittance of the element. It is evident that $Y(s) = 1/Z(s)$. Together, these functions are referred to as *immittance functions* and are easily evaluated as follows:

$$\begin{aligned} Z(s) = Y^{-1}(s) &= R \quad \text{for a resistor} \\ &= sL \quad \text{for an inductor} \\ &= \frac{1}{sC} \quad \text{for a capacitor} \end{aligned} \tag{2.14}$$

The transfer function of a system is the ratio of the response to an exponential excitation across a pair of specified terminals. It is composed of various combinations of the immittance functions of individual elements, R, L, and C, giving rise to a quotient of polynomial in s, the complex frequency variable. The coefficients of the polynomial are real, since R, L, and C, are physical elements and have real values. As a consequence, the transfer function of an electrical network has the same form as that in equation (2.11). Do all quotients of polynomials represent a physical network? Intuitively, the answer is "No." Physical networks must satisfy the law of the conservation of energy and must yield a real response to real stimuli. Such functions describe physical networks and are called *positive real functions*. The input impedance or driving-point impedance function of any network must represent a positive real function. The analysis of such functions has been described in many books [4]. Table 2.1 lists the properties of positive real functions.

2.6 Steady-State Response of Linear Systems to Sinusoidal Excitations

The steady-state response of a system to the exponential driving function e^{st} was found to be $H(s)e^{st}$. It follows that the steady-state response of the system to a function $e^{j\omega t}$ will be $H(j\omega)e^{j\omega t}$. Since $\cos \omega t + j \sin \omega t = e^{j\omega t}$ and the principle of superposition holds for linear systems, we can conclude that the steady-state response of the system to driving functions $\cos \omega t$ and $\sin \omega t$ is given by $\mathrm{Re}[H(j\omega)e^{j\omega t}]$ and $\mathrm{Im}[H(j\omega)e^{j\omega t}]$, respectively. $H(j\omega)$ is complex in general and is expressed in polar form as

$$H(j\omega) = |H(j\omega)|e^{j\theta(\omega)}$$

and

$$H(j\omega)e^{j\omega t} = |H(j\omega)|e^{j(\omega t+\theta)} \tag{2.15}$$

Therefore

$$\mathrm{Re}[H(j\omega)e^{j\omega t}] = |H(j\omega)|\cos(\omega t + \theta)$$

and

$$\mathrm{Im}[H(j\omega)e^{j\omega t}] = |H(j\omega)|\sin(\omega t + \theta) \tag{2.16}$$

The term $|H(j\omega)|$ signifies the magnitude of the response to a unit sinusoidal function. The response function is shifted in phase by angle θ with respect to the driving function. The angle θ is the phase angle of $H(j\omega)$, specifically, $\angle H(j\omega) = \theta$.

2.7 Circuit Theory Approximation

There are three basic assumptions implicit in circuit theory [5]:

1) The system is physically small enough that the propagation effects can be ignored; that is, the electrical effects occur instantaneously throughout the entire system. Ignoring the spatial dimensions results in what we refer to as lumped-parameter elements and systems.
2) The net charge on every component in the system is always zero. Thus, no component can collect a net excess of charge, although some components can hold equal but opposite charges.
3) There is no coupling between the components in a system.

Assumption 1 implies that the wavelengths of the signals are much larger than the physical dimensions of the electric components. At low frequencies, say, 1 MHz, the free-space wavelength is 300 m. Components measuring in the range of a few centimeter are well within the range of the assumption. At microwave frequencies, say, 1 GHz, the wavelength is 0.3 m. The dimensions of circuits in this range are comparable with the wavelength and, therefore, outside the range of the assumption. How can we then justify the use of circuit theory on the basis of the lumped-element representation at microwave frequencies? The answer lies in the modeling of microwave circuits such that their behavior can be predicted accurately using lumped-element representation. Such is indeed the case with distributed microwave structures over a limited frequency range. Consequently, circuit theory approximation is widely used in the analysis and synthesis for a large class of microwave filters.

Assumption 2 reflects the law of conservation of the charge or current in an electrical system. For time-varying signals, the concept of displacement current allows preservation of this principle when circuits include capacitors. Gustav Kirchhoff formalized these concepts, now referred to as *Kirchhoff's laws*:

1. The algebraic sum of all the currents at any node in a circuit equals zero.
2. The algebraic sum of all the voltages around any closed path in a circuit equals zero.

Assumption 3 implies that each element in circuit theory is independent of the other elements in the network. It allows for mutual coupling between two inductors that form a transformer. Such a mutual coupling is directly related to the inductances of the two separate elements forming the transformer.

In summary, the circuit theory is based on five ideal elements:

- An ideal voltage source is a circuit element that maintains a prescribed voltage across its terminals regardless of the device.
- An ideal current source is a circuit element that maintains a prescribed current within its terminals, regardless of the voltage across it.

- An ideal resistor is linear, time-invariant, bilateral, lumped element that follows Ohm's law.
- An ideal inductor is a linear, time-invariant, lumped circuit parameter that relates the voltage induced by a time-varying magnetic field to the current, producing the time-varying magnetic field.
- An ideal capacitance is a linear, time-invariant, lumped circuit parameter that relates the current induced by a time-varying electric field to the voltage, producing the time-varying electric field.

The application of the lumped-element circuit theory at microwave frequencies is an approximation, albeit a very good one over a limited frequency range. It is important to understand that lumped circuit techniques represent an approximation of the more encompassing EM theory. Consequently, it is possible to apply both the lumped circuit approximation and EM theory to a given problem. Circuit theory allows a good first approximation, and, if necessary, EM techniques can be applied to further refine the approximation, especially over wider frequency bands. Such is indeed the case for a large number of practical microwave engineering applications. The benefits of circuit theory are as follows:

1) It provides simple solutions (of significant accuracy) to problems that would otherwise become hopelessly complicated if we were to use electromagnetic field theory. We can analyze and build practical circuits with circuit theory.
2) The analysis and design of many useful electrical systems are less complicated if we divide them into subsystems, called *components*. We can then use the terminal behavior of each component to predict the behavior of the interconnection. The ability to derive the circuit models of physical devices makes the circuit theory an attractive approach.
3) Circuit analysis introduces a methodology for solving large networks of linked linear differential equations, prevalent in all fields of engineering and sciences. This allows access to a common pool of knowledge that exists in other engineering disciplines.
4) Circuit theory is an interesting area of study in its own right. Much of the remarkable development of electrical power systems can be attributed to the development of circuit theory as a separate discipline of study.

Summary

The principles that unify communication theory and circuit theory approximation are explained in this chapter. It highlights the necessity for communication systems to be linear over the specified range of bandwidth. When a signal, composed of a number of time-varying functions, is applied to a linear system, the overall response is given by the sum of the responses due to the individual functions. This follows from the principle of superposition, an inherent property of a linear system. As is well known, any function, periodic or nonperiodic, is amenable to representation by the sum of exponential functions via Fourier series or Laplace transforms. Thus, a communication channel can be characterized by an exponential driving function, and that represents a major simplification in the analysis of communication systems. This approach is called the *frequency analysis approach* or the *frequency-domain analysis*.

The next step in the simplification process for analyzing a communication channel is the circuit theory approximation. It assumes that the basic passive electrical components, namely, resistors, inductors, and capacitors, are discrete, lumped, linear elements. Such modeling, when combined with exponential driving functions, leads to the transfer functions of circuits in terms of a rational quotient of polynomials with real coefficients. This is a huge simplification in the analysis and synthesis of filter networks. Moreover, distributed elements can be represented by

lumped elements over limited bandwidths, and that allows the application of the circuit theory approximation as the first step in the design of microwave filters. EM techniques can then be applied to refine the approximation. This chapter highlights the fundamental assumptions and the success of the frequency analysis approach that we take for granted in analyzing electrical networks.

References

1 Lathi, B.P. (1965) *Signals, Systems and Communications*, Wiley, New York.
2 Volta, A. (1800) On the electricity excited by the mere contact of conducting substances of different kinds. *Philosophical Transactions of the Royal Society (London)*, **90**, 403–431.
3 Elliot, R.S. (1993) *Electromagnetics—History, Theory, and Applications*, IEEE Press.
4 Van Valkenburg, M.E. (1960) *Modern Network Synthesis*, Wiley, New York.
5 Nilsson, J.W. (1993) *Electric Circuits*, 4th edn, Addison-Wesley, Reading, MA.

3

Characterization of Lossless Lowpass Prototype Filter Functions

This chapter describes the synthesis process for the characteristic polynomials to realize the ideal, classical, prototype filters: the maximally flat, Chebyshev, and elliptic function filters. The chapter includes a discussion of filters that are not symmetric with respect to their center frequency. This leads to transfer function polynomials (with certain restrictions) with complex coefficients, a distinct departure from the more familiar characteristic polynomials with rational and real coefficients. This provides a basis for analysis of the most general class of filter functions in the lowpass prototype domain, minimum and nonminimum phase filters, exhibiting a symmetric or an asymmetric frequency response.

3.1 The Ideal Filter

In communication systems, filter networks are required to transmit and attenuate signals in specified frequency bands. Ideally, this must be accomplished with the minimum of distortion and loss of energy of the transmitted signal.

3.1.1 Distortionless Transmission

A signal is characterized by the amplitude and phase of its frequency components, referred to as the *waveform* of the signal. For distortionless transmission, the waveform must be preserved; that is, the output signal must be an exact replica of the input signal. This can be achieved only if the filter network has a constant amplitude and time delay for all the frequencies in the passband. A constant time delay implies that the phase shift is directly proportional to the frequency and is given by $-\omega t$. Outside the passband, the filter network is required to attenuate all frequencies, implying a brick-wall amplitude response, exhibiting infinite attenuation and a zero transition band. The unit impulse response of a brick-wall filter exists for the negative values of time, thus violating the causality condition [1]. A causal function starts at some finite time, say, $t = 0$ and has zero value for $t < 0$. The causality condition means that a physical system cannot anticipate the driving function. Causality condition gives rise to two fundamental constraints for the filter function: (1) the amplitude function can be zero at some discrete frequencies but cannot be zero over a finite band of frequencies and (2) the amplitude function cannot fall off to zero faster than a function of an exponential order. The implication is that the zero transition band is not allowed. However, it is possible to approach the ideal characteristic as closely as desired. Needless to say, it is the practical constraints that set the limit of the achievable performance. The ideal and practical response of a lowpass prototype filter is depicted in Figure 3.1.

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

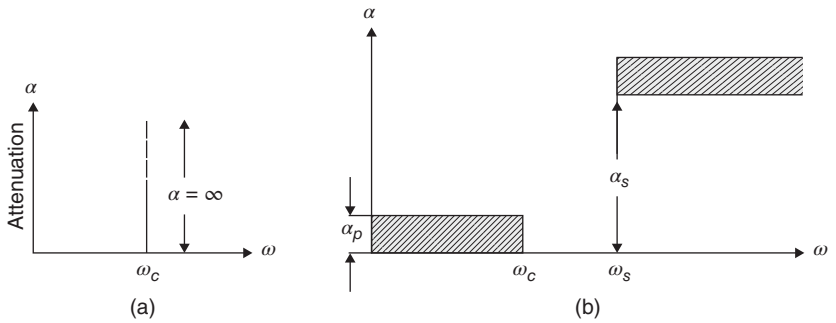


Figure 3.1 Lowpass filter amplitude response: (a) ideal filter and (b) practical filter. (Note that ω_c represents the passband cutoff frequency and ω_s represents the lower edge of stopband; α_p and α_s represent the maximum and minimum attenuation in the passband and stopband, respectively.)

3.1.2 Maximum Power Transfer in Two-Port Networks

Communication systems consist of a tandem connection of several two-port networks. Energy is transferred from the source to load via linear transducers such as amplifiers, modulators, filters, and cables. It is to be understood that although components such as modulators and amplifiers are inherently nonlinear, their operation is controlled to be linear with a minimal and acceptable amount of nonlinearity. For practical systems, it is imperative that all the available power is transferred at each stage so as to realize the highest levels of signal to noise ratios in the system. To achieve this condition, let us examine an ideal lossless network connected between a source and a load, as reflected in Figure 3.2.

The available power of a source is defined as the maximum power that can be drawn from the source. The maximum available power from a source of internal impendence Z_S is obtained when the load connected to its terminals is the complex conjugate of Z_S. Therefore, if $Z_S = R_S + jX_S$, then $Z_L = R_S - jX_S$ for a maximum power transfer.

Therefore, the available power from the source is $E^2/4R_S$. The power delivered to the load is maximized if the output impedance of the two-port network is the conjugate of Z_L. It should be noted that the available power depends only on the resistive component of the source. Similarly, the maximum power that can be delivered to the load depends solely on its resistive component. This leads us to the lossless two-port networks, terminated in resistors of equal value at each end, as the desired filter networks for communications systems. Such filter networks are referred to as the doubly terminated filters.

3.2 Characterization of Polynomial Functions for Doubly Terminated Lossless Lowpass Prototype Filter Networks

The synthesis of filter networks is carried out by developing lumped lossless lowpass filters that are normalized in terms of frequency and impedance and terminated in resistors of equal

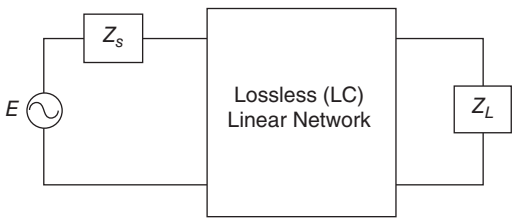


Figure 3.2 Power transfer in a two-port network.

value. By scaling in frequency and amplitude, it is then possible to derive filter networks over any desired frequency range and impedance levels. This simplifies the design of practical filters regardless of their frequency range and how they are realized physically. The lowpass prototype filter is normalized to a cutoff frequency of 1 rad/s and terminated in resistors of 1 Ω. The cutoff frequency of unity implies that the passband extends from $\omega = 0$ to $\omega = 1$.

As described in the previous section, even for an ideal filter, we cannot have a zero loss over the entire passband without violating the causality condition. However, it is possible to have a zero loss at a finite number of frequencies. Such frequencies are referred to as *reflection zeros*, for which no power is reflected, and thus the signal incurs zero loss. An obvious consequence of this is that all the reflection zeros of the filter function are confined to the passband. Furthermore, the magnitude of maximum reflected power over the passband serves as a design parameter of the filter prototype networks. The use of lossless elements makes the synthesis process much simpler. The effect of dissipation can be readily incorporated after the synthesis process with little loss of accuracy, provided the dissipation is small. Filter design based on lossless prototype networks is quite adequate for a variety of applications. This chapter deals with the circuit theory approximation for the design of lossless lowpass prototype filter functions.

Figure 3.3 illustrates a lossless two-port network terminated in resistors. This is a general representation of a doubly terminated filter network capable of maximum power transfer.

The maximum available power $P_{\max}$ from the ideal voltage source E is $E^2/4R_1$. Power P_2, delivered to the load R_2, is $|V_2|^2/R_2$. Therefore

$$\frac{P_{\max}}{P_2} = \left| \frac{1}{2} \sqrt{\frac{R_2}{R_1}} \frac{E}{V_2} \right|^2$$

$$= \frac{1}{4} \left| \frac{E}{V_2} \right|^2 \quad \text{for } R_1 = R_2 \tag{3.1}$$

For passive lossless two-port networks, P_2 can only be equal to or less than $P_{\max}$. For such networks, it is most convenient to define a characteristic function $K(s)$ by [2, 3]:

$$\frac{P_{\max}}{P_2} = 1 + |K(s)|^2_{s=j\omega} \tag{3.2}$$

For the lumped, linear, and time-invariant circuits considered here, $K(s)$ is a rational function in s with real coefficients. The ratio $P_{\max}/P$ is defined by the function $H(s)$, where

$$|H(s)|^2_{s=j\omega} = 1 + |K(s)|^2_{s=j\omega} \tag{3.3}$$

$H(s)$ is referred to as the *transducer function* [3] and can be written in the alternative form

$$H(s)H(-s) = 1 + K(s)K(-s) \tag{3.4}$$

since all network functions are rational in s with real coefficients. To characterize the polynomial $K(s)$, let us consider the reflection and transmission coefficients of the prototype network.

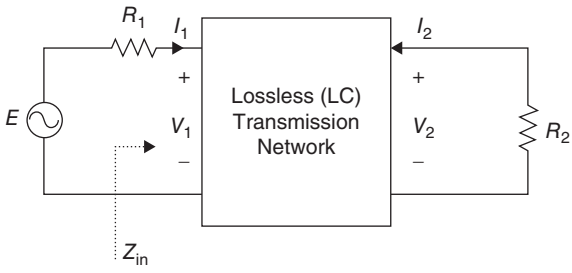


Figure 3.3 Doubly terminated lossless transmission network.

3.2.1 Reflection and Transmission Coefficients

In transmission-line theory, the reflection coefficient ρ is defined as

$$\rho(s) = \pm\frac{\text{reflected wave}}{\text{incident wave}} \tag{3.5}$$

The sign of ρ is positive when the ratio is defined in terms of voltages and negative when defined in terms of currents. The negative sign arises because the direction of current for the reflected wave is opposite that of incident wave. The use of symbol ρ for the reflection coefficient is quite common for analyzing two-port networks and their power relationships. The symbol Γ is used to denote the complex reflection coefficient and generally used when analyzing circuits in terms of scattering parameters. For the lossless network in Figure 3.3 and with $P_{\max}$ as the reference or maximum power available, the reflection coefficient in terms of power is given by

$$|\rho(j\omega)|^2 = \frac{\text{reflected power}}{\text{available power}} = \frac{P_r}{P_{\max}} \tag{3.6}$$

where P_r represents the reflected power. Since, the reflected power plus the transmitted power delivered to the load resistor R_2 must equal the available power:

$$\frac{\text{Reflected power}}{\text{Available power}} + \frac{\text{transmitted power}}{\text{available power}} = 1$$

or

$$|\rho(j\omega)|^2 + |t(j\omega)|^2 = 1 \tag{3.7}$$

where $t(s)$ is defined as the transmission coefficient and represents the ratio of the transmitted wave to the incident wave. Therefore

$$|t(j\omega)|^2 = \frac{\text{transmitted power}}{\text{available power}} = \frac{P_2}{P_{\max}} = 1 - |\rho(j\omega)|^2 \tag{3.8}$$

In terms of decibels, the transmission and reflection loss (often termed the *return loss*) are defined as

$$\begin{aligned} A &= 10\log\frac{1}{|t(j\omega)|^2} = -10\log|t(j\omega)|^2 \text{ dB} \\ R &= 10\log\frac{1}{|\rho(j\omega)|^2} = -10\log|\rho(j\omega)|^2 \text{ dB} \end{aligned} \tag{3.9}$$

Making use of equation (3.7), we can relate the transmission and return loss by

$$A = -10\log[1 - 10^{-R/10}] \text{ dB}$$

and

$$R = -10\log[1 - 10^{-A/10}] \text{ dB} \tag{3.10}$$

From the transmission-line theory [2], the reflection coefficient ρ for a two-port network is given by

$$\rho(s) = \frac{Z_{\text{in}}(s) - R_1}{Z_{\text{in}}(s) + R_1} = \frac{z_{\text{in}}(s) - 1}{z_{\text{in}}(s) + 1}, \quad \text{where} \quad z_{\text{in}}(s) = \frac{Z_{\text{in}}(s)}{R_1} \tag{3.11}$$

The input impedance of a terminated network is a positive real function (Section 2.5), and its normalized impedance z can be expressed as

$$z(s) = \frac{n(s)}{d(s)} \tag{3.12}$$

where $n(s)$ and $d(s)$ represent the numerator and denominator polynomials and $z(s)$ is a positive real function [2]. Thus

$$\rho(s) = \frac{z(s)-1}{z(s)+1} = \frac{n(s)-d(s)}{n(s)+d(s)} = \frac{F(s)}{E(s)} \tag{3.13}$$

Since $z(s)$ is positive real, we can conclude the following:

1) The denominator polynomial $n(s) + d(s) = E(s)$ must be the Hurwitz polynomial, with all its roots lying inside the left half of the s plane.
2) The numerator polynomial $n(s) - d(s) = F(s)$ may or may not be a Hurwitz polynomial. However, the polynomial coefficient must be real and, as a result, its roots must be real or at the origin or occur in conjugate complex pairs.

The magnitude of ρ, along the imaginary axis, is therefore given by

$$|\rho(j\omega)|^2 = \frac{F(s)F^*(s)}{E(s)E^*(s)} \tag{3.14}$$

where the asterisks indicate the complex conjugate of the respective functions. Furthermore, for $s = j\omega$, $F^*(s) = F(-s)$ and $E^*(s) = E(-s)$, and therefore

$$|\rho(j\omega)|^2 = \frac{F(s)F(-s)}{E(s)E(-s)}$$
$$|t(j\omega)|^2 = \frac{E(s)E(-s) - F(s)F(-s)}{E(s)E(-s)} = \frac{P(s)P(-s)}{E(s)E(-s)} \tag{3.15}$$

where $P(s)P(-s) = E(s)E(-s) - F(s)F(-s)$. Expressing the right-hand side of equation (3.15) by $P(s)P(-s)$ is justified only if it is a perfect square. This is indeed the case since the root pattern has a quadrantal symmetry and the roots on the imaginary axis have even multiplicity. Equation (3.15) is referred to as Feldtkeller's equation; it is for filters exhibiting symmetric characteristics. For a general form including fully canonical prototypes, asymmetric characteristics, or those where the reflection zeros do not lie on the imaginary axis, the more general paraconjugate form is described in Chapter 6.

The three polynomials are referred to as the *characteristic polynomials*. For lowpass prototype filter networks with symmetric characteristics, their properties are summarized as follows [3]:

1) $F(s)$ is a polynomial with real coefficients, and its roots lie along the imaginary axis as conjugate pairs. $F(s)$ can have multiple roots only at the origin. The roots represent frequencies at which no power is reflected, often termed *reflection zeros*. At these frequencies, the filter loss is zero. $F(s)$ is a pure odd or even polynomial with its highest coefficient as unity.
2) $P(s)$ is a pure even polynomial with real coefficients. Its roots lie on the imaginary axis in conjugate pairs. Such roots represent frequencies at which no power is transmitted, and the filter loss is infinite. These frequencies are often referred to as *transmission zeros* or *attenuation poles*. Its roots can also occur as conjugate pairs on the real axis or as a complex quad in the s plane. Such roots lead to linear phase (nonminimum phase) filters. The highest coefficient of $P(s)$ is unity.
3) $E(s)$ is a strict Hurwitz polynomial, as all its roots lie in the left half of the s plane.

In terms of the characteristic polynomials, we obtain

$$\rho(s) = \frac{F(s)}{E(s)} \quad \text{and} \quad t(s) = \frac{P(s)}{E(s)} \tag{3.16}$$

The reflected and transmitted powers are related by

$$|K(s)|^2_{s=j\omega} = \frac{|\rho(j\omega)|^2}{|t(j\omega)|^2} = \frac{P_r}{P_t} \tag{3.17}$$

where P_r is the reflected power and P_t is the transmitted power.

The transfer function and the characteristic function are then expressed as the ratio of the polynomials in the form

$$t(s) = \frac{P(s)}{E(s)} \tag{3.18}$$

and

$$K(s) = \frac{F(s)}{P(s)} \tag{3.19}$$

It should be noted that in terms of scattering parameters described in subsequent chapters, the reflection and transmission coefficients are synonymous with S_{11} and S_{21}, respectively, represented as

$$\rho(s) \equiv \Gamma(s) \equiv S_{11}(s)$$
$$t(s) \equiv S_{21}(s)$$

3.2.2 Normalization of the Transfer Function and Characteristic Polynomials

The transfer function of a lossless prototype filter, defined as the ratio of the transmitted power to the available power [equations (3.2) and (3.8)], can be represented as

$$|t(s)|^2_{s=j\omega} = \frac{1}{1 + |K(s)|^2_{s=j\omega}}$$

Without loss of generality, we can introduce a real and arbitrary constant $\boldsymbol{\varepsilon}$ and restate the relationship as

$$|t(s)|^2_{s=j\omega} = \frac{1}{1 + \boldsymbol{\varepsilon}^2|K(s)|^2_{s=j\omega}} \tag{3.20}$$

$\boldsymbol{\varepsilon}$ is often referred to as the ripple constant or factor. The ripple factor is used to normalize the transfer function of the filter so that the maximum permissible amplitude in the passband is restricted to a specified value. If ω_1 is the bandwidth corresponding to the maximum ripple in the passband, then

$$\boldsymbol{\varepsilon} = \sqrt{\frac{|t(j\omega_1)|^{-2} - 1}{|K(j\omega_1)|^2}} = \sqrt{\frac{1}{\left[|\rho(j\omega_1)|^{-2} - 1\right]|K(j\omega_1)|^2}} \tag{3.21}$$

For lowpass prototype networks, ω_1 is referred to as the cutoff frequency and is typically chosen to be unity. The value of $\boldsymbol{\varepsilon}$, in terms of the transmission and return loss in dB [see equation (3.9)] is given by

$$\boldsymbol{\varepsilon} = \sqrt{\frac{10^{-A_1/10} - 1}{|K(s)|^2_{s=j}}} = \sqrt{\frac{\left[10^{R_1/10} - 1\right]^{-1}}{|K(s)|^2_{s=j}}} \tag{3.22}$$

Since the polynomial $K(s)$ represents the ratio of polynomials $F(s)$ and $P(s)$, it is more appropriate to represent $K(s)$ in the form

$$K(s) = \varepsilon\frac{F(s)}{P(s)} = \frac{F(s)}{P(s)/\varepsilon}. \tag{3.23}$$

Polynomials $F(s)$ and $P(s)$ are formed by the zeros of reflection and transmission, respectively, and are assumed known with their highest coefficients as unity. Using equations (3.15) and (3.23), polynomial $E(s)$ is determined as

$$E(s)E(-s) = \frac{1}{\varepsilon^2}P(s)P(-s) + F(s)F(-s) \tag{3.24}$$

The leading coefficient of the polynomial $E(s)$ is therefore given by $\sqrt{\frac{1}{\varepsilon^2} + 1}$.

This represents the most general case of a fully canonical filter where the order of $P(s)$ is the same as $F(s)$ and $E(s)$, that is, N, the order of the filter. If the order of $P(s)$ is $<N$, the leading coefficient of $E(s)$ is the same as that of $F(s)$, namely, unity.

In the synthesis process, it is more convenient to have the leading coefficients of all the polynomials, including $E(s)$ to be unity. This can be achieved by expressing the ripple factor as a ratio of two constants, that is, $\boldsymbol{\varepsilon} = \frac{\varepsilon}{\varepsilon_R}$, and assigning ε to polynomial $P(s)$ and ε_R to $F(s)$. The leading coefficient of $E(s)$ is then evaluated as $\sqrt{\frac{1}{\varepsilon^2} + \frac{1}{\varepsilon_R^2}}$. For this coefficient to be unity, the two constants must be related as

$$\frac{1}{\varepsilon^2} + \frac{1}{\varepsilon_R^2} = 1 \quad \text{and} \quad \boldsymbol{\varepsilon} = \frac{\varepsilon}{\varepsilon_R} \tag{3.25}$$

The value of $\boldsymbol{\varepsilon}$ is determined by equation (3.22) as before. The values of ε and ε_R are then evaluated as

$$\varepsilon = \sqrt{1 + \boldsymbol{\varepsilon}^2} \quad \text{and} \quad \varepsilon_R = \frac{\sqrt{1 + \boldsymbol{\varepsilon}^2}}{\boldsymbol{\varepsilon}} \tag{3.26}$$

This relationship applies only for the fully canonical filters for which the order of $P(s)$ is the same as $F(s)$. Canonical filters tend to be highly sensitive (this topic is covered in more depth in later chapters) and their practical applications are therefore limited. For most practical filter networks, $P(s) < N$ (implies $\varepsilon_R = 1$), and one needs to deal with a single constant to normalize all the polynomials to have their highest coefficients as unity. This relationship is addressed more rigorously in Section 6.1.1. It should be noted that the enlarged epsilon represents the constant used to normalize the transfer function, whereas the normal epsilon is the constant used to normalize $P(s)$; the two are not the same. The notation used here is consistent with the notation in Chapter 6.

3.3 Characteristic Polynomials for Idealized Lowpass Prototype Networks

Filter requirements call for a low loss in the passband and a high loss in other frequency bands. Such a requirement can best be achieved by assigning all the zeros of $F(s)$ to the $j\omega$ axis in the passband region and all zeros of $P(s)$ to the $j\omega$ axis in the high-loss frequency bands. For some applications, the zeros of $P(s)$ have a non-$j\omega$-axis location. This results in an improved phase and group delay response in the passband at the expense of attenuation in the stopband. Such a trade-off is sometimes beneficial for the overall system requirements. In order to ensure that the characteristic polynomials conform to be positive real functions (Section 2.5), the pole/zero locations of lossless lowpass prototype networks normalized to a unity cutoff frequency are restricted as follows:

1) All the zeros of $K(s)$ must lie symmetrically on the $j\omega$ axis in the passband region.
2) All the poles of $K(s)$ must lie symmetrically on the $j\omega$ axis or real axis, or be distributed symmetrically along the real and imaginary axes, forming a conjugate complex quad.

The distribution of zero locations for polynomials $F(s)$ and $P(s)$ are illustrated in Figure 3.4. The polynomial $F(s)$ containing the reflection zeros has the form

$$\begin{aligned} F(s) &= s(s^2 + a_1^2)(s^2 + a_2^2)\cdots \quad \text{for odd-order filters} \\ &= (s^2 + a_1^2)(s^2 + a_2^2)\cdots \quad \text{for even-order filters} \end{aligned} \tag{3.27}$$

The polynomial $P(s)$ can be formed by factors $(s^2 \pm b_i^2)$ or $(s^4 \pm c_i s^2 + d_i)$ or any combination of such factors. Therefore, the characteristic function $K(s)$ assumes the following forms:

$$K(s) = \varepsilon \frac{F(s)}{P(s)} = \varepsilon \frac{s(s^2 + a_1^2)(s^2 + a_2^2)\cdots}{(s^2 \pm b_1^2)(s^2 \pm b_2^2)\cdots(s^4 \pm c_1 s^2 + d_1)\cdots} \tag{3.28}$$

for odd-order networks and

$$K(s) = \varepsilon \frac{F(s)}{P(s)} = \varepsilon \frac{(s^2 + a_1^2)(s^2 + a_2^2)(s^2 + a_3^2)\cdots}{(s^2 \pm b_1^2)(s^2 \pm b_2^2)\cdots(s^4 \pm c_1 s^2 + d_1)\cdots} \tag{3.29}$$

for even-order networks. The constraints on the a and b terms are

$|a_1|, |a_2|, \ldots < 1$

$|b_1|, |b_2|, \ldots > 1$, for $j\omega$-axis zeros and no restriction for real-axis zeros

$c_1, c_2, \ldots, d_1, d_2, \ldots > 0$

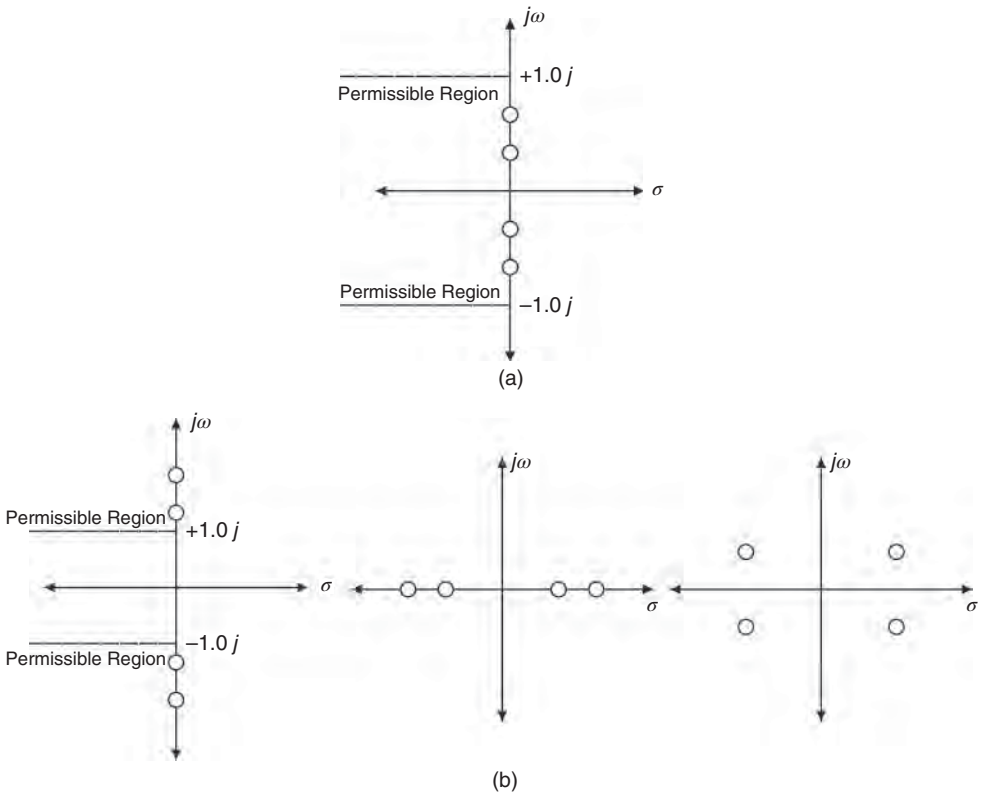


Figure 3.4 Permissible zero locations of lowpass prototype characteristic polynomials, (a) zero locations of $F(s)$, and (b) zero locations of $P(s)$ on the $j\omega$ axis, real axis, and as a complex quad.

These relationships clearly indicate that all the polynomials are rational with real coefficients and are either pure even or pure odd, reflecting the nature of positive real functions to realize lowpass prototype networks. From equations (3.28) and (3.29), it is evident that the frequencies $\pm ja_1, \pm ja_2, \ldots$ are zeros of reflection, that is, the frequencies at which all the power is transmitted and none is reflected. As a result, these frequencies are also called *attenuation zeros*. Similarly, frequencies $\pm jb_1, \pm jb_2, \ldots$ are zeros of transmission and no power is transmitted, and there is a total reflection of the power. For this reason, these frequencies are more commonly called *attenuation poles* or *transmission zeros*. Together, *a*'s and *b*'s are termed *critical frequencies*. The zeros that lie on the real axis or form a symmetric complex quad are used to improve the phase and group delay response at the expense of the amplitude response. Such filters are referred to as *linear phase filters*. The polynomial $E(s)$ is strictly Hurwitz, that is, all its zeros must lie in the left half of the s plane.

3.4 Lowpass Prototype Characteristics

Filters are specified in terms of their amplitude and phase or group delay response. Computation of these quantities is described in this section.

3.4.1 Amplitude Response

The insertion loss in dB is defined as

$$\begin{aligned} A(\text{dB}) &= 10 \log \frac{P_{\max}}{P_2} = 20 \log |H(j\omega)| \\ &= 20 \log \left| \varepsilon \frac{E(s)}{P(s)} \right| \text{ dB} \end{aligned} \tag{3.30}$$

If we denote $s_1, s_2, \ldots, s_n$ as the roots of the Hurwitz polynomial $E(s)$ and $p_1, p_2, \ldots, p_m$ as the roots of polynomial $P(s)$,

$$|H(j\omega)| = \varepsilon \frac{|(s - s_1)(s - s_2) \cdots (s - s_n)|}{|(s - p_1)(s - p_2) \cdots (s - p_m)|} \tag{3.31}$$

where ε is an arbitrary constant for normalizing the amplitude response and ε is computed by evaluating $|H|$ at $\omega = 0$. At zero frequency, the value of H is either unity or exceeds unity by a small amount determined by the maximum permissible ripple in the passband.

The insertion loss is then computed as

$$\begin{aligned} A(\text{dB}) = 20 \log e [&\ln |(s - s_1)| \\ &+ \ln |(s - s_2)| + \cdots - \ln |(s - p_1)| - \ln |(s - p_2)| - \cdots] \\ &+ 20 \log e \ln \varepsilon \end{aligned} \tag{3.32}$$

At real frequencies, $s = j\omega$, a typical term in the insertion loss is given by

$$\begin{aligned} |s - s_k|_{s=j\omega} &= |s - \sigma_k \mp j\omega_k|_{s=j\omega} \\ &= \sqrt{\sigma_k^2 + (\omega \mp \omega_k)^2} \end{aligned} \tag{3.33}$$

where $s_k = \sigma_k + j\omega_k$ represents the kth root of the respective polynomial.

3.4.2 Phase Response

The phase response for the filter network is computed by

$$\beta(\omega) = -\text{Arg } H(j\omega)$$

$$= -\tan^{-1}\frac{\mathrm{Im}\,H(j\omega)}{\mathrm{Re}\,H(j\omega)}$$
$$= \sum_{\text{poles}} \tan^{-1}\left(\frac{\omega-\omega_k}{\sigma_k}\right) - \sum_{\text{zeros}} \tan^{-1}\left(\frac{\omega-\omega_k}{\sigma_k}\right) \tag{3.34}$$

The group delay τ of the filter is calculated by

$$\tau = -\frac{d\beta}{d\omega} = \sum_{\text{zeros}} \left(\frac{\sigma_k}{\sigma_k^2 + (\omega-\omega_k)^2}\right) \tag{3.35}$$

The zeros of $P(s)$ are either located on the imaginary axis or distributed symmetrically about the imaginary axis. Consequently, they do not contribute to the phase variation (except for the integer multiples of π radians). Since the group delay is the derivative of the phase, it is entirely determined by the roots of $E(s)$. The slope of the group delay is given by

$$\frac{d^2\beta}{d\omega^2} = \sum_{\text{zeros}} 2\tau_k^2 \tan\beta_k - \sum_{\text{poles}} 2\tau_k^2 \tan\beta_k \tag{3.36}$$

The term τ_k is the group delay factor associated with the kth pole or zero such that

$$\tau_k = -\frac{d\beta_k}{d\omega} = \frac{\sigma_k}{\sigma_k^2 + (\omega-\omega_k)^2} \tag{3.37}$$

3.4.3 Phase Linearity

The phase linearity over a specified frequency band is defined as the deviation of the phase from the ideal linear phase response. Usually, the ideal phase curve is defined by the slope of the phase at zero frequency for the lowpass prototype filter. However, we can choose any reference frequency over the band of interest and define the phase linearity with respect to the slope at that reference frequency.

Let ω_{ref} be the reference frequency. The linear phase ϕ_L at a frequency ω is then expressed by

$$\phi_L(\omega) = \left(\frac{d\beta}{d\omega}\right)_{\omega=\omega_{\text{ref}}} (\omega-\omega_{\text{ref}})$$
$$= \tau_{\text{ref}}(\omega-\omega_{\text{ref}}) \tag{3.38}$$

where τ_{ref} is the absolute group delay at the reference frequency. The phase linearity $\Delta\phi$ at a given frequency ω is given by

$$\Delta\phi = \beta(\omega) - \phi_L \tag{3.39}$$

where $\beta(\omega)$ is the actual phase at ω. If the reference frequency is chosen to be zero, then

$$\Delta\phi = \beta(\omega) - \tau_0\omega, \quad \text{where} \quad \tau_0 = \left.\frac{d\beta}{d\omega}\right|_{\omega=0} \tag{3.40}$$

From a practical standpoint, it is desirable to compute the minimum phase deviation from a straight-line approximation of the phase over the band of interest. This implies curve fitting of a straight line over the phase response so as to obtain an equiripple phase about the straight line.

3.5 Characteristic Polynomials versus Response Shapes

In this section, we describe the possible forms of the characteristic function and the resulting response shapes for doubly terminated lowpass prototype filter networks.

3.5.1 All-Pole Prototype Filter Functions

Such filter functions are characterized by

$$t(s) = \frac{1}{E(s)} \tag{3.41}$$

with $P(s) \equiv 1$. There are no finite transmission zeros and the attenuation rises monotonically beyond the passband. All the attenuation poles are located at infinity. The response shape is determined by the form of the polynomial $F(s)$. The two basic forms of $F(s)$ are

$$F(s) \to s^n$$

or

$$F(s) \to s^m(s^2 + a_1^2)(s^2 + a_2^2)\cdots \tag{3.42}$$

The first form has all the zeros at the origin, and the transmission response is characterized by a maximally flat passband response, commonly known as the *Butterworth response.*

The more general second form of $F(s)$ is characterized by some of its zeros at the origin and others at the finite frequencies, $a_1, a_2, \ldots$. Figure 3.5 offers such an arbitrary response shape. By an appropriate choice of $a_1, a_2, \ldots$, we can obtain an equiripple response shape. A special case of the equiripple response exists when there are maximum permissible peaks in the passband. For such a case, $m = 1$ for odd-order filters and zero for even-order filters. The critical frequencies a_i are chosen via the Chebyshev polynomial to ensure the maximum number of equiripple peaks in the passband. For this reason, such filters are called *Chebyshev filters.*

3.5.2 Prototype Filter Functions with Finite Transmission Zeros

Such a filter function is characterized by

$$\begin{aligned} P(s) &= (s^2 + b_1^2)(s^2 + b_2^2)\cdots \\ F(s) &= s^m(s^2 + a_1^2)(s^2 + a_2^2)\cdots \end{aligned} \tag{3.43}$$

The difference between this and the all-pole filter function is the introduction of transmission zeros, which tends to increase the selectivity of the amplitude response. Figure 3.6 illustrates a comparison of the amplitude response of a fourth-order equiripple filter both with and without

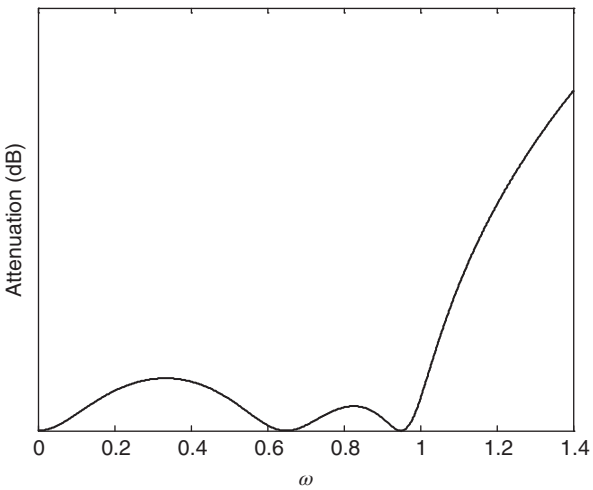


Figure 3.5 Lowpass amplitude response shape of an all-pole filter with an arbitrary distribution of reflection zeros.

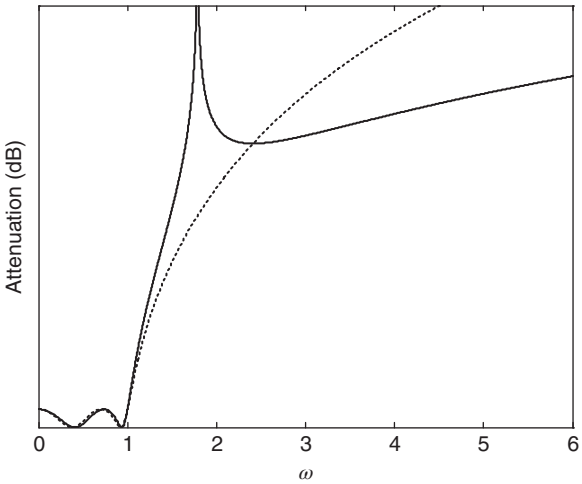


Figure 3.6 Lowpass amplitude response of a fourth-order filter with and without a finite transmission zero.

a finite transmission zero in the stopband. When the transmission zeros are introduced, the increased selectivity is obtained at the expense of a poorer wideband amplitude response in the stopband.

3.6 Classical Prototype Filters

This section describes the derivation of characteristic polynomials of the classical filters.

3.6.1 Maximally Flat Filters

Let us consider an nth-order polynomial $K_n(\omega)$ that satisfies the following conditions:

1) $K_n(\omega)$ is an nth-order polynomial.
2) $K_n(0) = 0$.
3) $K_n(\omega)$ is maximally flat at the origin.
4) $K_n(1) = \varepsilon$.

Condition 1 implies that

$$\varepsilon K_n(\omega) = c_0 + c_1\omega + c_2\omega^2 + \cdots + c_n\omega^n \tag{3.44}$$

Condition 2 requires that $c_0 = 0$. "Maximally flat at the origin" implies that as many derivatives as possible are zero at the origin. Thus

$$\varepsilon \frac{dK_n}{d\omega} = c_1 + 2c_2\omega + \cdots + nc_n\omega^{n-1} \tag{3.45}$$

For this polynomial to be zero at $\omega = 0$, it is required that $c_1 = 0$. Similarly, higher-order derivatives can be made to vanish by assigning zero to the higher-order coefficients. Thus, condition 3 requires

$$K_n(\omega) = c_n\omega^n \tag{3.46}$$

Finally, condition 4 yields $c_n = \varepsilon$. Summarizing, we can write

$$K_n(\omega) = \varepsilon\omega^n \tag{3.47}$$

where ε is the ripple factor, which defines the maximum amplitude response over the passband. The choice of ε as unity implies normalization with respect to the half-power point at the unity cutoff frequency. In terms of the characteristic polynomials, since all the poles of the filter are located at infinity, $P(s) = 1$. As a result, by virtue of equation (3.46), the polynomial $F(s)$ is given by

$$F(s) = s^n \tag{3.48}$$

The Hurwitz polynomial $E(s)$ is determined as follows. For the normalized (half-power) Butterworth lowpass prototype filter of degree n, we obtain

$$\begin{aligned} |K(j\omega)|^2 &= \omega^{2n} \\ |H(j\omega)|^2 &= 1 + \omega^{2n} = |E(j\omega)|^2 \end{aligned} \tag{3.49}$$

By analytic continuation, we can replace ω^2 with $-s^2$ to obtain

$$E(s)E(-s) = 1 + (-s^2)^n \tag{3.50}$$

The $2n$ zeros of this function are located on the unit circle. Their positions are given by [4]

$$s_k = \begin{cases} \exp\left[\dfrac{j\pi}{2n}(2k-1)\right] & \text{for } n \text{ even} \\ \exp\left(\dfrac{j\pi k}{n}\right) & \text{for } n \text{ odd} \end{cases} \tag{3.51}$$

where $k = 1,2,\ldots,2n$. Since $E(s)$ represents the Hurwitz polynomial, the zeros obtained from equation (3.46) that lie in the left half plane are assigned to $E(s)$. The amplitude response of a three-pole maximally flat filter is represented in Figure 3.7. The poles of the transfer function for this filter by using equation (3.51) are computed as $s_1 = -0.5 + 8660j$, $s_2 = -1.0$, and $s_3 = -0.5 - 0.8660j$.

3.6.2 Chebyshev Approximation

Let us define the characteristic function by

$$|K(j\omega)|^2 = \varepsilon^2 T_n^2\left(\frac{\omega}{\omega_c}\right) \tag{3.52}$$

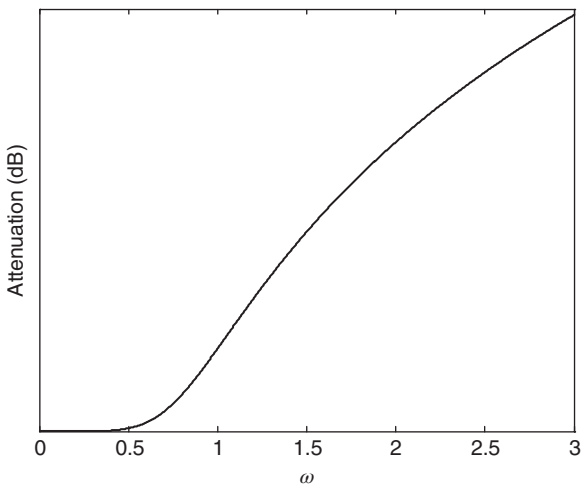


Figure 3.7 Lowpass amplitude response of three-pole maximally flat filter.

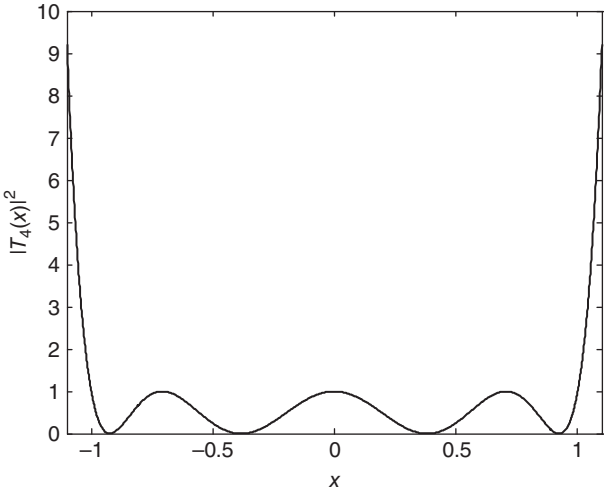


Figure 3.8 Plot of a fourth-order Chebyshev polynomial.

The polynomial $T_n(x)$ is the nth-degree polynomial that has the following properties:

1) T_n is even (odd) if n is even (odd).
2) T_n has all its zeros in the interval $-1 < x < 1$.
3) T_n oscillates between the values of ± 1 in the interval $-1 \leq x \leq 1$.
4) $T_n(1) = +1$.

The amplitude response, characterized by such a function, is shown in Figure 3.8. T_n can be derived as [5]

$$T_n(x) = \cos(n \cos^{-1} x) \tag{3.53}$$

This is a transcendental function and represents the well-known Chebyshev polynomial. Filters derived from this polynomial are referred to as *Chebyshev filters*. Equation (3.53) is represented by the recursion relationship

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{3.54}$$

Since $T_0(x) = 1$ and $T_1(x) = x$, we can generate the Chebyshev polynomial for any order n. For example, a third-order Chebyshev polynomial is computed by

$$T_3(x) = 4x^3 - 3x \tag{3.55}$$

In the complex frequency domain ($x \rightarrow s/j$), this equation would be represented by $T_3(s) = 4s^3 + 3s$. The roots of this polynomial are given by 0 and $\pm j0.8660$. In terms of the characteristic polynomials, $P(s) = 1$, and by virtue of equation (3.52), the polynomial $F(s)$ is given by

$$F(s) = T_n\left(\frac{s}{j}\right) \tag{3.56}$$

The Hurwitz polynomial is determined as follows:

$$H(s)H(-s) = 1 + \left[\varepsilon T_n\left(\frac{s}{j}\right)\right]^2 = E(s)E(-s) \tag{3.57}$$

The roots s_k are the solutions to

$$T_n\left(\frac{s_k}{j}\right) = \pm\frac{j}{\varepsilon} \tag{3.58}$$

This equation can be readily solved to yield [2]

$$\sigma_k = \pm\sinh\left(\frac{1}{n}\sinh^{-1}\frac{1}{\varepsilon}\right)\sin\frac{\pi}{2}\frac{2k-1}{n}$$
$$\omega_k = \cosh\left(\frac{1}{n}\sinh^{-1}\frac{1}{\varepsilon}\right)\cos\frac{\pi}{2}\frac{2k-1}{n} \tag{3.59}$$

where $k = 1,2,\ldots,2n$. The Hurwitz polynomial $E(s)$ is formed by the left half-plane roots obtained from equation (3.59). For example, for a 20 dB return loss, $\rho = 0.1$ and $\varepsilon = 0.1005$. This corresponds to an amplitude ripple of 0.0436 dB in the passband. By using equation (3.59), the roots of the Hurwitz polynomial are computed as -1.1714, -0.5857, and $\pm j1.3368$. It should be noted that in this formulation, polynomial $F(s)$ is being represented by the Chebyshev polynomial $T_n(s)$, with its highest coefficient as 4. The normalized $F(s)$, with its highest coefficient as unity is given by

$$F(s) = s\left(s^2 + \frac{3}{4}s\right)$$

Using the normalized $F(s)$, ε is computed as 0.4020. As expected, this value is four times the value using the unnormalized $F(s)$. The alternative representations of $F(s)$ have no impact on the location of poles and zeros of the various polynomials. Figure 3.9 exhibits the amplitude response of a three-pole Chebyshev filter.

3.6.3 Elliptic Function Filters

Let us examine a filter for which $|K(j\omega)| \le \varepsilon$ in the passband and $|K(j\omega)| \ge K_{\min}$ in the stopband. The maximum passband attenuation and minimum stopband attenuation are then related as

$$A_{\max} = 10\log(1+\varepsilon^2) \tag{3.60}$$
$$A_{\min} = 10\log(1+K_{\min}^2) \tag{3.61}$$

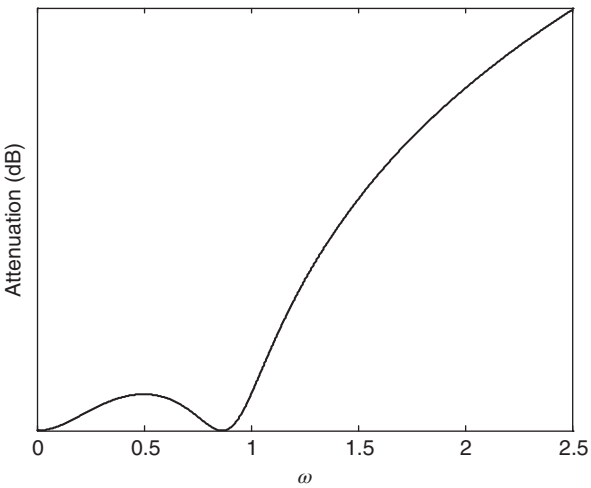


Figure 3.9 Lowpass amplitude response of a third-order Chebyshev filter.

There is an infinity of solutions to this approximation problem. However, the solution of most interest is when the response is equiripple in both passband and stopband. Such a response provides the sharpest rise in attenuation and is known as an *elliptic function response*, since its solution is dependent on the Jacobian elliptic functions. To examine this approximation problem, we express the characteristic function as [5]

$$|K(j\omega)|^2 = \left[\varepsilon R_n\left(\frac{\omega}{\omega_p}, L\right)\right]^2 \tag{3.62}$$

where $R_n((\omega/\omega_p), L)$ is a rational function. For filters having an equiripple passband and stopband, this function must have the following properties:

1) R_n is even (odd) if n is even (odd).
2) R_n has all its n zeros in the interval $-1 < (\omega/\omega_p) < 1$ and all its n poles outside that interval.
3) R_n oscillates between values of ± 1 in the interval $-1 < (\omega/\omega_p) < 1$.
4) $R_n(1, L) = +1$.
5) $1/R_n$ oscillates between values of $\pm 1/L$ in the interval $|\omega| > \omega_s$.

The problem now reduces to the determination of a rational function $R_n(x, L)$ that possesses the previous properties; in other words, all points in $|x| < 1$ and $|x| > x_L$, where $|R_n(x, L)| = 1$ and L, respectively, must represent maximum or minimum values of $R_n(x, L)$ at these points. Therefore

$$\left.\frac{dR_n(x,L)}{dx}\right|_{|R_n(x,L)|=1} = 0 \quad \text{except } |x| = 1$$
$$\left.\frac{dR_n(x,L)}{dx}\right|_{|R_n(x,L)|=L} = 0 \quad \text{except } |x| = x_L \tag{3.63}$$

where $x = (\omega/\omega_p)$ and $x_L = (\omega_s/\omega_p)$. As a result, $R_n(x, L)$ is defined through the differential equation

$$\left(\frac{dR_n}{dx}\right)^2 = M^2\frac{(R_n^2 - 1)(R_n^2 - L^2)}{(1 - x^2)(x^2 - x_L^2)}$$

or

$$\frac{CdR_n}{\sqrt{(1 - R_n^2)(L^2 - R_n^2)}} = \frac{Mdx}{\sqrt{(1 - x^2)({x_L}^2 - x^2)}} \tag{3.64}$$

where C and M are constants. This is the differential equation for the Chebyshev rational function, and its solution involves elliptic integrals. This rational function is derived and expressed as

$$R_n(x, L) = C_1 x \prod_{v=1}^{(n-1)/2} \frac{x^2 - sn^2(2vK/n)}{x^2 - [x_L/sn(2vK/n)]^2} \quad \text{for } n \text{ odd} \tag{3.65}$$

$$R_n(x, L) = C_2 x \prod_{v=1}^{(n/2)} \frac{x^2 - sn^2[(2v-1)K/n]}{x^2 - \{[x_L/sn(2v-1)K/n]\}^2} \quad \text{for } n \text{ even} \tag{3.66}$$

Here $sn[(vK/n)]$ is the Jacobian elliptic function, and K is the complete elliptic integral of the first kind (not to be confused with the characteristic function $K(j\omega)$) defined by [6]

$$K = \int_0^{\pi/2} \frac{d\xi}{\sqrt{1 - k^2\sin^2\xi}} \tag{3.67}$$

where k is the modulus of the elliptic integral. The modulus is less than unity for K to be real, permitting an alternative definition of k where

$$k = \sin\theta = \frac{\omega_p}{\omega_s} \tag{3.68}$$

and θ is the modular angle. By substituting for x, x_L, and $\sin\theta = \omega_p/\omega_s$, the characteristic function for n odd is expressed as [7]

$$|K(j\omega)| = \varepsilon C_1 \frac{\omega}{\omega_p} \prod_{v=1}^{(n-1)/2} sn^2\left(\frac{2vK}{n}\right) \frac{\omega^2 - sn^2\left(\frac{2vK}{n}\right)\omega_s^2 \sin^2\theta}{\omega^2 sn^2\left(\frac{2vK}{n}\right) - \omega_s^2} \tag{3.69}$$

If we choose the geometric mean of ω_p and ω_s as the normalizing frequency of unity, then

$$\sqrt{\omega_p \cdot \omega_s} = 1 \quad \text{or} \quad \omega_p = \frac{1}{\omega_s} = \sqrt{\sin\theta} \tag{3.70}$$

The characteristic function $|K(j\omega)|$ is rendered symmetric, such that the zeros of the numerator and denominator are reciprocal of each other

$$|K(j\omega)| = \varepsilon\omega \prod_{v=1}^{(n-1)/2} \left[\frac{\omega^2 - \left\{ \sqrt{\sin\theta} sn\left(\frac{2vK}{n}\right)\right\}^2}{\omega^2 \left\{ \sqrt{\sin\theta} sn\left(\frac{2vK}{n}\right)\right\}^2 - 1} \right] \tag{3.71}$$

In a similar manner, the symmetric form is obtained when n is even. We now express the characteristic function in the familiar symmetric form [7]:

$$K(s) = \varepsilon s \prod_{v=1}^{(n-1)/2} \frac{(s^2 + a_{2v}^2)}{(s^2 a_{2v}^2 + 1)} \quad \text{for } n \text{ odd} \tag{3.72}$$

$$K(s) = \varepsilon \prod_{v=1}^{n/2} \frac{(s^2 + a_{2v-1}^2)}{(s^2 a_{2v-1}^2 + 1)} \quad \text{for } n \text{ even} \tag{3.73}$$

where $a_v = \sqrt{\sin\theta}[sn(vK)/n]$ and $v = 1,2,\ldots,n$. It is evident that this representation implies that the cutoff frequency ω_p is normalized to $\sqrt{\sin\theta}$. For prototype filters, ω_p is usually selected to be unity. Such a normalization is achieved by dividing all the critical frequencies by $\sqrt{\sin\theta}$. For a given n and θ, the critical frequencies are determined and thus the characteristic function $K_n(\omega)$. The zeros of the Hurwitz polynomial $E(s)$ are obtained by the analytic continuation

$$1 + \varepsilon^2 K_n^2(s) = 0$$

or

$$P(s)P(-s) + \varepsilon^2 F(s)F(-s) = 0 \tag{3.74}$$

For the case where n is odd, $F(s)$ is an odd polynomial, and

$$[P(s) + \varepsilon F(s)][P(s) - \varepsilon F(s)] = 0 \tag{3.75}$$

where n is even, both $F(s)$ and $P(s)$ are even polynomials, and

$$P^2(s) + \varepsilon^2 F^2(s) = 0 \tag{3.76}$$

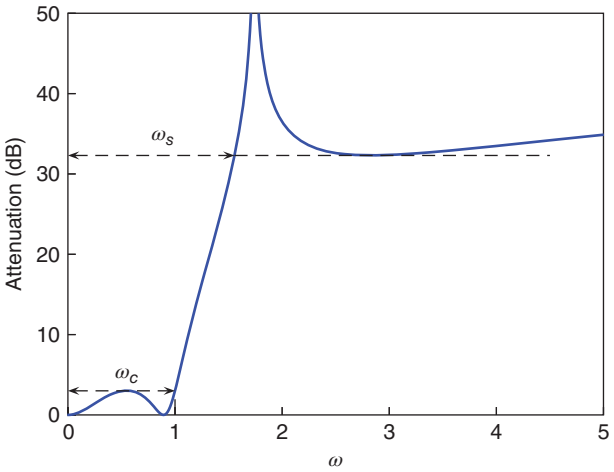


Figure 3.10 Amplitude response of a third-order lowpass prototype elliptic filter.

The roots of this equation determine all the poles in the s plane. Only the left half-plane poles are used to form the polynomial $E(s)$. This completes the derivation of the transfer function for elliptic function filters.

3.6.4 Odd-Order Elliptic Function Filters

The characteristic function of odd-order elliptic filters is represented by equation (3.72). The amplitude response of a third-order elliptic function filter is plotted in Figure 3.10. It has an attenuation zero at $\omega = 0$ and infinite attenuation at $\omega = \infty$. Here we consider an example of a third-order filter and compute its critical frequencies as follows.

Assuming $\theta = 40°$ as the modular angle, the frequencies corresponding to the equiripple passband and stopband by using equation (3.65) are calculated by

$$\omega_p = \sqrt{\sin\theta} = 0.801470$$
$$\omega_s = \frac{1}{\omega_p} = 1.247287$$

The next step in the computation process is the evaluation of K, the complete elliptic integral as 1.7865769, followed by the evaluation of the Jacobian elliptic function $sn(vK/3)$ [6]. This allows computation of critical frequencies via equation (3.72):

$$a_v = \sqrt{\sin\theta} sn\left(\frac{vK}{3}\right)$$

The critical frequencies are evaluated as $a_1 = 0.4407$, $a_2 = 0.7159$, and $a_3 = 0.8017$.

According to equation (3.72), a_2 represents the attenuation zero and $1/a_2$, the attenuation pole. The attenuation maxima in the passband is given by a_1, whereas a_3 represents the normalizing frequency ω_p, given by $\sqrt{\sin\theta}$. The reciprocal of a_3 represents ω_s, the frequency that corresponds to the attenuation minima in the stopband. It is more convenient to have prototype filters whose passband is normalized to the unity cutoff. This is obtained by dividing the critical frequencies by $\sqrt{\sin\theta}$. Thus, the attenuation zeros and poles normalized to a cutoff frequency of unity are as follows:

Normalized attenuation zero = 0.8929

Normalized attenuation pole = 1.7423
Normalized $\omega_s = 1.5557$

3.6.5 Even-Order Elliptic Function Filters

3.6.5.1 Class A Elliptic Filters

The characteristic function for the even-order class A elliptic function filter is represented by the symmetric equation (3.73). The attenuation response of such a four-pole filter is shown in Figure 3.11a. This filter has a finite attenuation at $\omega = 0$ and $\omega = \propto$ and has the sharpest rise in attenuation in the stopband. The critical frequencies are given by $a_v = \sqrt{\sin\theta}sn[(v/4)K]$ and are normalized with respect to $\omega_p = \sqrt{\sin\theta}$. Critical frequencies can be normalized to a unity cutoff by dividing them by $\sqrt{\sin\theta}$. By following the same procedure, described in the example for the three-pole filter, for a modular angle $\theta = 40°$, K, the complete elliptic integral is evaluated to be 1.786769. The computed critical frequencies with unity cutoff are summarized as follows:

Normalized attenuation zeros = 0.4267, 0.9405
Normalized attenuation poles = 1.6542, 3.6461
Normalized $\omega_s = 1.5557$

The physical realization of such a response represents a very sensitive circuit and, for that reason, has limited applications at microwave frequencies.

3.6.5.2 Class B Elliptic Filters

Even-order class B elliptic filters are derived in order to push the highest attenuation pole toward infinity, enabling a ladder realization without mutual inductances. The critical frequencies for this class of filters are derived by the frequency transformation as given in [7], where

$$b_\gamma = \sqrt{\frac{w}{a_\gamma^{-2} - a_1^2}}$$

$$b'_v = \sqrt{\frac{w}{a_\gamma^2 - a_1^2}}$$

$$w = \sqrt{(1 - a_1^2 a_n^2)\left(1 - \frac{a_1^2}{a_n^2}\right)} \tag{3.77}$$

where a_γ are derived by equation (3.73).

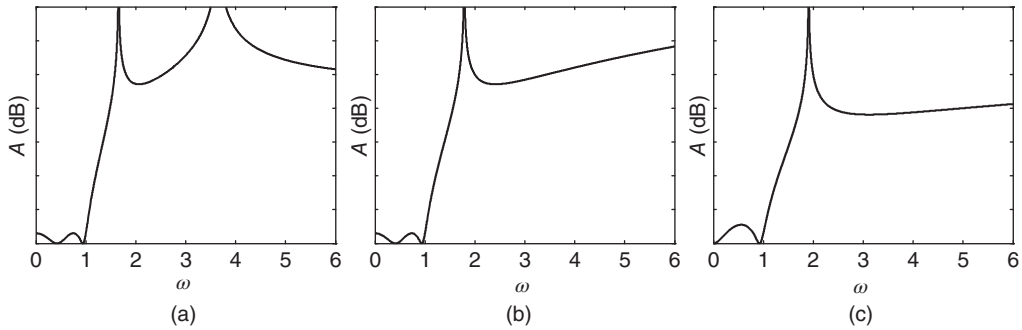


Figure 3.11 Lowpass amplitude response of four-pole elliptic filters: (a) class A, (b) class B, and (c) class C four elliptic filters.

The corresponding characteristic function for this class of filter is

$$|K(j\omega)| = \frac{(b_1^2 - \omega^2)(b_3^2 - \omega^2)\cdots(b_{n-1}^2 - \omega^2)}{(b_3'^2 - \omega^2)(b_5'^2 - \omega^2)\cdots(b_{n-1}'^2 - \omega^2)} \tag{3.78}$$

and the new cutoff frequency is given by

$$\omega_p = b_n = \sqrt{a_n a_{n-1}} < a_n \tag{3.79}$$

The response shape for this class of filter is plotted in Figure 3.11b. For such elliptic filters, the attenuation pole is shifted toward infinity. This type of filter provides the sharpest rise in attenuation, next only to the class A elliptic filters. Such a response can be realized without coupling between the input and output terminals, a major advantage in the implementation of practical filter networks at microwave frequencies. Consequently, the class B elliptic filter is the most widely used of even-order elliptic function filters.

By making use of the example of class A four-pole elliptic filter, the critical frequencies for the class B filter are computed by equations (3.77) as follows:

$$w = 0.8697$$
$$b_1 = 0.3212$$
$$b_3 = 0.7281$$
$$b_3' = 1.3879$$

The cutoff frequency is derived from $\omega_p = \sqrt{a_4 a_3} = 0.7775$, and the normalization with respect to a unity cutoff is obtained by dividing all critical frequencies by $\sqrt{a_4 a_3}$. Therefore

Normalized attenuation zeros = 0.4131, 0.9361
Normalized attenuation pole = 1.7851
Normalized $\omega_s = 1.6542$.

3.6.5.3 Class C Elliptic Filters

Even-order class C elliptic filters are characterized by a zero attenuation at $\omega = 0$ and an infinite attenuation at $\omega = \infty$. This class of filter function can be obtained through a frequency transformation [7]:

$$c_\gamma = \sqrt{\frac{a_\gamma^2 - a_1^2}{1 - a_\gamma^2 a_1^2}} = \sqrt{a_{\gamma-1} \cdot a_{\gamma+1}} \tag{3.80}$$

The corresponding characteristic function is

$$|K(j\omega)| = \frac{\omega^2(c_3^2 - \omega^2)(c_5^2 - \omega^2)\cdots(c_{n-1}^2 - \omega^2)}{(1 - c_3^2\omega^2)(1 - c_5^2\omega^2)\cdots(1 - c_{n-1}^2\omega^2)} \tag{3.81}$$

and the new cutoff frequency is given by $\omega_p = a_{n-1} < b_n < a_n$.

Using equations (3.68), (3.74), and (3.76), we can see that the cutoff frequency is progressively reduced from class A to class B and class C filters, implying a reduced sharpness in the rise of the attenuation in stopband. The response shape for a fourth-order filter using such a transformation is shown in Figure 3.11c. By following the same procedure as for classes A and B, the critical frequencies are computed using equation (3.75) and are summarized as follows:

Normalized attenuation zero = 0.9223
Normalized attenuation pole = 1.9069.

3.6.6 Filters with Transmission Zeros and a Maximally Flat Passband

These filters are characterized by all the attenuation zeros at the origin and the arbitrary distribution of the attenuation poles.

The characteristic function $K(s)$ has the form

$$K(s) = \varepsilon \frac{F(s)}{P(s)} = \varepsilon \frac{s^n}{(s^2 + b_1^2)(s^2 + b_2^2) \cdots (s^2 + b_m^2)} \tag{3.82}$$

where n is the order of the filter and $m(m \leq n)$ represents the number of transmission zeros. The conventional equiripple form of this class of filters is commonly referred to as *inverse Chebyshev filters.*

3.6.7 Linear Phase Filters

These filters are characterized by the presence of at least one pair of real-axis zeros or a complex quad of zeros or any combination thereof. Any type of minimum phase filter can be augmented by such zeros to improve the passband response. However, this improvement in group delay comes at the expense of a poorer amplitude response. The form of the characteristic function of a four-pole Chebyshev linear phase filter with a single pair of real-axis zeros is represented by

$$K(s) = \varepsilon \frac{(s^2 + a_1^2)(s^2 + a_2^2)}{(s^2 - b_1^2)} \tag{3.83}$$

where b_1 is the real-axis zero. Linear phase filters are described in more depth in Chapters 4 and 13.

3.6.8 Comparison of Maximally Flat, Chebyshev, and Elliptic Function Filters

The amplitude response of a four-pole lowpass prototype maximally flat, Chebyshev, and elliptic function (class B) filter is described in Figure 3.12. The response is normalized to a unity cutoff frequency. For the elliptic filter, the transmission zero is located at the normalized frequency of 1.8. A maximally flat filter exhibits monotonically rising attenuation. In the near-out-of-band

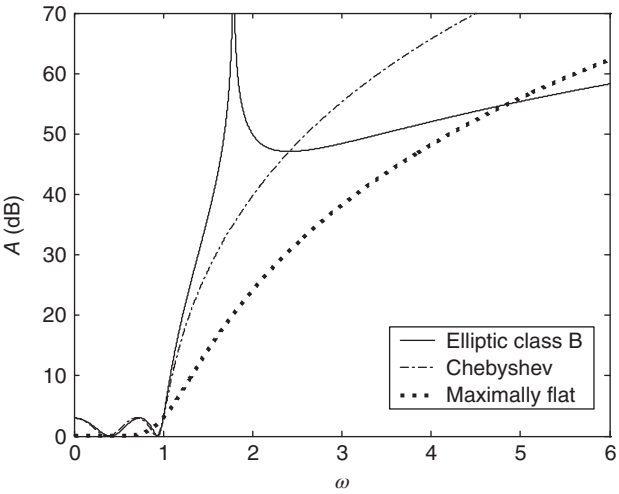


Figure 3.12 Comparison of four-pole maximally flat, Chebyshev, and elliptic function filters.

region, the filter provides less attenuation than an equivalent Chebyshev and elliptic filter. However, in the far-out-of-band region, the maximally flat filter provides higher attenuation and has the highest asymptotic slope.

The Chebyshev filter has an equiripple passband and a monotonically rising attenuation band. It provides a higher attenuation than does the maximally flat filter in the near-out-of-band region. The elliptic filter is characterized by an equiripple passband and stopband. This filter has the highest attenuation slope in the near-out-of-band region. However, in the far-out-of-band region, the filter provides less attenuation as compared with the maximally flat or Chebyshev filter.

This comparison highlights the trade-off between the near-out-of-band rejection and the far-out-of-band attenuation. For most applications, Chebyshev and elliptic function filters offer the best compromise between the near- and far-out-of-band attenuation responses. As always, one has to perform a detailed trade-off analysis of the amplitude and group delay response to determine the best design to meet a given specification.

3.7 Unified Design Chart (UDC) Relationships

The characteristic function $K(s)$ is given by

$$|K(s)|^2_{s=j\omega} = \frac{P_r}{P_t} = \frac{\text{reflected power}}{\text{transmitted power}} \tag{3.84}$$

If we choose ω_1 as the bandwidth, corresponding to the maximum ripple in the passband, and ω_3 as the bandwidth, corresponding to the minimum attenuation in the stopband, then

$$\left(\frac{P_r}{P_t}\right)_{s=j\omega_1}\left(\frac{P_t}{P_r}\right)_{s=j\omega_3} = \frac{|K(s)|^2_{s=j\omega_1}}{|K(s)|^2_{s=j\omega_3}} = FF(\omega_1, \omega_3) \tag{3.85}$$

where $FF(\omega_1, \omega_3)$ represents the ratio of the reflected and transmitted powers at a pair of frequencies ω_1 and ω_3. In terms of the insertion and reflection loss in decibels, we have

$$(R_1 - A_1) + (A_3 - R_3) = 10 \log FF(\omega_1, \omega_3)$$

or

$$(R_1 + A_3) - (A_1 + R_3) = F \tag{3.86}$$

where F simply represents the value $10 \log FF(\omega_1, \omega_3)$. Note that

R_1, A_1	Return loss and transmission loss in dB, corresponding to maximum ripple in the passband
R_3, A_3	Return loss and transmission loss in dB, corresponding to minimum attenuation in stopband
F	A dimensionless quantity with dB as the unit, referred to as the "characteristic factor"

Unified design chart (UDC) nomenclature for Butterworth, Chebyshev, and quasielliptic filters is illustrated in Figures 3A.1 and 3A.3. The value of the function F for the Butterworth filter is $20n \log(\omega_3/\omega_1)$, and for the Chebyshev filter it is $20 \log T_n(\omega_3/\omega_1)$, where T_n is the Chebyshev polynomial. For the elliptic filters, F is evaluated as $20 \log R_n(\omega/\omega_p, L)$. Plot of the

characteristic factor F versus (ω_3/ω_1) leads to the UDCs [8], which are generated as a function of the ratio of the characteristic polynomial $K(s)$ at a pair of frequencies, from equation (3.86). The ripple factor ε cancels out in this relationship, and thus, the UDCs are independent of it. On the other hand, the choice of ε governs the trade-off between the return loss in the passband and the attenuation in the stopband for a given F. The choice of ε alters the values of $R_1(A_1)$ and $A_3(R_3)$ within the constraint of constant F. In other words, for a given F, the dB for dB trade-off between the return loss in the passband and attenuation in the stopband is preserved and provided by the selection of the ripple factor. For most practical applications, the return loss in the passband is chosen to be >20 dB, implying $R_1 = 20$ dB and $A_1 = 0.0436$ dB. In the stopband, the attenuation required is typically >20 dB, implying $A_3 = 20$ dB and $R_3 = 0.0436$ dB. Thus, for practical applications, $F > 40$ dB, and A_1 and R_3 are negligible compared with F. Consequently

$$(R_1 + A_3) \approx F \tag{3.87}$$

Thus, a trade-off exists, dB for dB, between the return loss in the passband and transmission loss in the stopband. It should be noted that the relationship $F = (R_1 + A_3) - (A_1 + R_3)$ is an exact relationship. Also, there are only two independent variables, since the return loss and transmission loss in dBs are related by

$$A = -10\log[1 - 10^{-R/10}] \tag{3.88}$$

The UDCs provide, at first glance, the order of filter required for a given set of amplitude constraints. In addition, the charts provide the trade-offs between the return loss in the passband and attenuation in the stopband or, alternatively, between the passband and stopband bandwidths for a given order of filter and the amplitude constraints. Appendix 3A includes the UDCs for the maximally flat, Chebyshev, elliptic, and quasielliptic function filters. These charts provide a quick and accurate first-cut design of a filter to meet a given amplitude specification.

3.8 Lowpass Prototype Circuit Configurations

The characteristic polynomials of the lossless lowpass prototype filters are based on lumped circuit elements. The circuit configuration of such a lowpass prototype filter as a ladder network is depicted in Figure 3.13a and its dual, in Figure 3.13b. The circuit model of the prototype filter provides the link to the physical realization of the filter networks.

The next step in the design process is to determine the element values g_k from the characteristic polynomial of a desired lowpass prototype filter. In Chapters 6–8, the synthesis procedures are presented to determine the element values from characteristic polynomials for any class of filter function. In the case of the classic maximally flat and Chebyshev filters, explicit formulae are derived in [4]. The prototype ladder network of Figure 3.13 has all its transmission zeros located at infinity. As a result, only all-pole filters such as the maximally flat or the Chebyshev-type filter can be synthesized by such a configuration. For filters with finite transmission zeros, we need to introduce resonant circuits in the prototype network, as shown in Figure 3.14.

For such networks, some of the zeros of the polynomial $P(s)$ are located along the $j\omega$ axis and represent frequencies at which no power is transmitted. For the classic elliptic function filters, there is an analytic procedure to synthesize the lumped circuit elements [7]. The synthesis techniques for the general class of filters with arbitrary distribution of zeros of $P(s)$ are explained in Chapters 7 and 8.

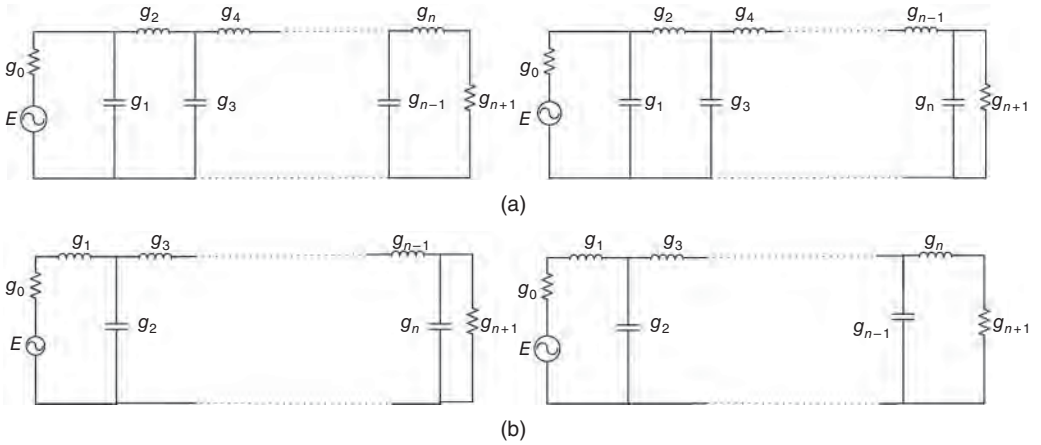


Figure 3.13 General forms of lowpass prototype lossless all-pole ladder filters: (a) shunt–series configuration and (b) series–shunt configuration.

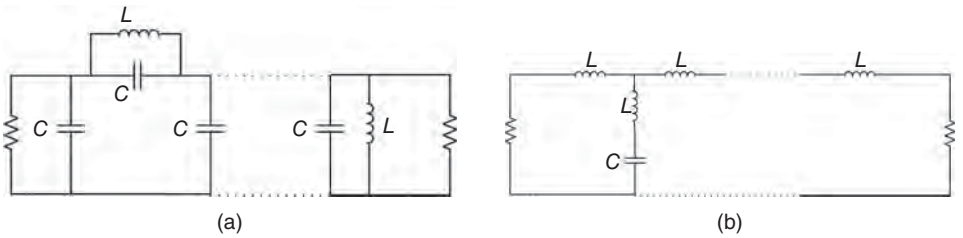


Figure 3.14 General forms of lowpass prototype lossless ladder filters with transmission zeros: (a) using parallel resonators and (b) using series resonators.

3.8.1 Scaling of Prototype Networks

The lowpass prototype filter serves as the basic model from which we can derive practical filters at any frequency with any bandwidth and impedance levels. This is accomplished by the use of appropriate frequency transformations to map the lowpass prototype filter response to the practical lowpass, highpass, bandpass, and bandstop filter networks. Let us consider two networks, the prototype network and the other, scaled network, as shown in Figure 3.15.

Suppose we desire to scale the driving point impedance $Z'(s)$ of the prototype network by

$$Z(s) = bZ'\left(\frac{s}{a}\right) \tag{3.89}$$

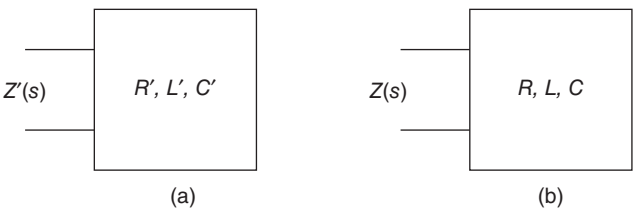


Figure 3.15 Scaling of prototype networks: (a) prototype network and (b) scaled network.

where $Z(s)$ is the equivalent impedance of the scaled network and a and b are dimensionless, real, and positive constants for scaling the frequency and impedance level, respectively. Both the impedance and frequency transformations are applied independently. If the two networks are identical in structure (topology), there is a simple relationship of the element values in the two networks [2]:

$$\begin{aligned} R &= bR' \\ L &= \frac{b}{a}L' \\ C &= \frac{1}{ab}C \end{aligned} \tag{3.90}$$

These relationships are restricted to transforming the lowpass prototype networks to practical lowpass filters. However, we must consider a more general frequency transformation to realize the highpass, bandpass, and bandstop filters.

Let us consider a general transformation $\phi(\omega)$, defined as a quotient of polynomials. Equation (3.89) represents the special case of $\phi(\omega)$, when the frequency is scaled by a constant factor. However, the most important case is one in which $\phi(\omega)$ is a reactance function [9]. In such a case, the inductors and capacitors are replaced by the networks of inductors and capacitors, but not resistors. This topic has been dealt with in many books [10, 11]. The reactance frequency transformations and equivalent element values are summarized in Tables 3.1 and 3.2.

3.8.2 Frequency Response of Scaled Networks

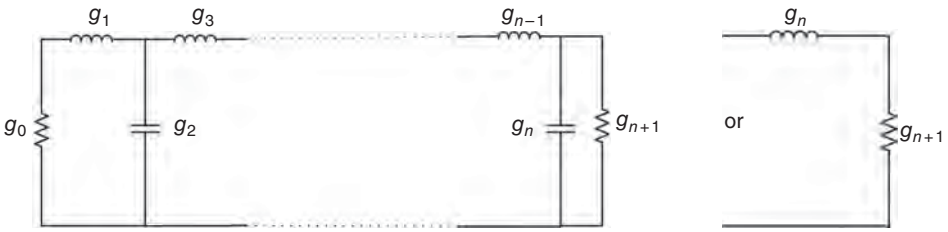
3.8.2.1 Amplitude Response

Under frequency transformation, the amplitude and phase response of the prototype network has a direct, one-to-one correspondence with the practical network derived from it. In other words, the amplitude and phase responses of the two networks at the same relative frequencies are identical. As an example, the amplitude response of the prototype filter at $\omega' = 0$ and 1 is the same as the one for a lowpass filter at $\omega = 0$ and ω_c. For a corresponding bandpass filter,

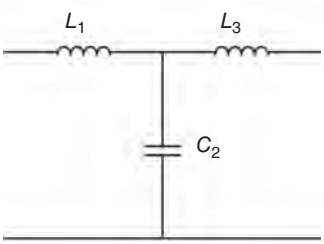
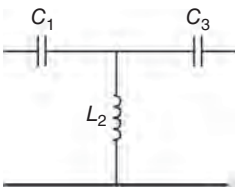
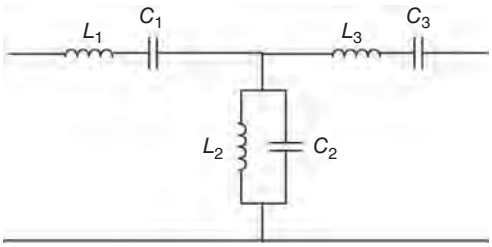
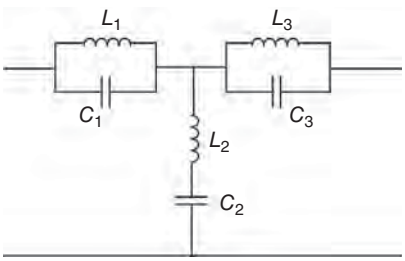
Table 3.1 Frequency transformations.

ω'	Normalized frequency variable of lowpass prototype filter ($\omega'_c = 1$)
ω	Unnormalized frequency variable
ω_c	Unnormalized cutoff frequency
ω_0	Band center frequency obtained as geometric mean of passband edges ω_2 and ω_1
$\omega_2 - \omega_1 = \Delta\omega$	Passband/stopband bandwidths
Transformation of lowpass prototype filter to	Equivalent frequency transformation variable
Lowpass filter	$\omega' = \dfrac{\omega}{\omega_c}$
Highpass filter	$\omega' = \dfrac{\omega_c}{\omega}$
Bandpass filter	$\omega' = \dfrac{\omega_0}{\Delta\omega}\left(\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}\right)$
Bandstop filter	$\omega' = \dfrac{1}{\dfrac{\omega_0}{\Delta\omega}\left(\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}\right)}$

Table 3.2 Lumped circuit transformations.



g_k: Element values of the lowpass prototype circuit

Circuit transformations lowpass prototype to	Transformed circuit configuration and equivalent element values	
Lowpass filter		$L_k = \frac{g_k}{\omega_c}$ $C_k = \frac{g_k}{\omega_c}$
Highpass filter		$L_k = \frac{1}{g_k \omega_c}$ $C_k = \frac{1}{g_k \omega_c}$
Bandpass filter		$L_k = \frac{g_k}{\Delta\omega}$ $C_k = \frac{\Delta\omega}{g_k \omega_0^2}$ $k = 1, 3, 5, \ldots$ $C_k = \frac{g_k}{\Delta\omega}$ $L_k = \frac{\Delta\omega}{g_k \omega_0^2}$ $k = 2, 4, \ldots$
Bandstop filter		$C_k = \frac{1}{g_k \Delta\omega}$ $L_k = \frac{g_k \Delta\omega}{\omega_0^2}$ $k = 1, 3, 5, \ldots$ $C_k = \frac{g_k \Delta\omega}{\omega_0^2}$ $L_k = \frac{1}{g_k \Delta\omega}$ $k = 2, 4, \ldots$

Note: The dual of these circuits has the same equivalent element values.

the amplitude response at the band center and band edge is the same as the prototype filter response at $\omega = 0$ and 1, respectively.

3.8.2.2 Group Delay Response

The group delay response is represented by the derivative of the phase response. Consequently, the group delay response of the transformed network depends on the absolute bandwidths of the practical networks. Let us consider the group delay response for the lowpass and bandpass filters.

3.8.2.3 Lowpass Filter Group Delay

The lowpass prototype to lowpass filter transformation is given by $\omega' = \omega/\omega_c$, and the group delay τ of the lowpass filter is given by

$$\tau = -\frac{d\beta}{d\omega} = -\frac{d\beta}{d\omega'}\frac{d\omega'}{d\omega} = \frac{1}{\omega_c}\tau' \tag{3.91}$$

where τ' is the group delay of the prototype network. Thus, for a lowpass filter with a cutoff frequency of 1 GHz, the group delay of the lowpass filter is simply given by $0.16\tau'$ ns. The group delay is, therefore, scaled by ω_c, the cutoff frequency of the lowpass filter.

3.8.2.4 Bandpass Filter Group Delay

By following the same procedure as for the lowpass filter, the group delay τ_{BPF} of the bandpass filter is readily derived as

$$\omega' = \frac{\omega_0}{\Delta\omega}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)$$

and

$$\begin{aligned}\tau_{\text{BPF}} &= -\frac{d\beta}{d\omega} = \frac{d\beta}{d\omega'}\frac{d\omega'}{d\omega} \\ &= \frac{\omega_0}{\Delta\omega}\left(\frac{1}{\omega_0} + \frac{\omega_0}{\omega^2}\right)\tau' \\ &\approx \frac{2}{\Delta\omega}\tau'\end{aligned} \tag{3.92}$$

At the band center, $\omega = \omega_0$, and

$$\tau_0 = \frac{2}{\Delta\omega}\tau_0' \tag{3.93}$$

τ_0' is the delay of the prototype filter at $\omega' = 0$. These relationships demonstrate that the shape and the relative values of the group delay are determined by the choice of the prototype filter and the absolute bandwidth of the practical filter. It is interesting to note that the group delay is independent of the choice of the center frequency of the bandpass filter, a point that often goes unnoticed.

3.9 Effect of Dissipation

In the physical realization of filter networks, the finite conductivities of the materials involved introduce dissipation of energy. As a consequence, the expected response of the transfer functions based on the lossless elements requires adjustments. The approximate effect of incidental dissipation in an inductance is to associate with it a series resistance; in a capacitance element,

the dissipation effectively produces a parallel conductance. To consider this effect on the transfer function, let us analyze the effect of changing the complex frequency variable s to $s+\delta$, where δ is a positive quantity, that is,

$$\begin{aligned} s &\rightarrow s+\delta \\ sL_k &\rightarrow sL_k+\delta L_k \\ sC_k &\rightarrow sC_k+\delta C_k \end{aligned} \tag{3.94}$$

where the k subscript denotes the kth element. This conversion then implies that each inductance is associated with a proportionate amount of series resistance δL_k and that each capacitance has a parallel conductance of the value δC_k. If r and r' are the associated series resistance and shunt conductance, then

$$\delta L_k = r \quad \text{or} \quad \delta = \frac{r}{L_k}$$

Similarly

$$\delta = \frac{r'}{C_k}$$

This common ratio δ is referred to as the *dissipation ratio* or factor. Since $s+\delta = s_k$ for the kth pole or zero, s assumes the value

$$s = s_k - \delta \tag{3.95}$$

Thus, all the zeros and poles of the transfer function are displaced along the real axis by an amount δ. It is assumed that the dissipation ratios of the inductances and capacitances are identical. Thus, by subtracting δ from the real part of the poles and zeros of the transfer function, we obtain the network response with finite unloaded Qs. Incorporating the displacement δ into the pole/zero configurations in equation (3.32) and making use of equation (3.37), value of insertion loss A_0 at $\omega = 0$ is evaluated as

$$A_0 \approx 20 \log e\, \delta \tau_0 \tag{3.96}$$

τ_0 is the absolute group delay at $\omega = 0$. The approximation assumes small amount of dissipation, that is, $\delta \ll 1$ and $\sigma \gg \delta$. This demonstrates that the insertion loss at $\omega = 0$ (or the band center for bandpass filters) is directly proportional to the absolute time delay at that frequency for small values of dissipation.

3.9.1 Relationship of Dissipation Factor δ and Quality Factor Q_0

The dissipation factor δ represents the displacement of the poles and zeros along the real (σ) axis in the complex frequency plane. δ has the dimensions of radian frequency.

Unloaded $Q(Q_0)$ represents the radian frequency times the energy stored, divided by the energy dissipated per cycle, or

$$\begin{aligned} Q_0 &= \omega \left(\frac{\text{energy stored}}{\text{average power loss}} \right) \\ &= 2\pi \left(\frac{\text{energy stored}}{\text{energy dissipated per cycle}} \right) \end{aligned} \tag{3.97}$$

Q_0 is a dimensionless quantity, and it provides an indication of the losses in the circuit. Assuming uniform dissipation for all inductances and capacitances comprising the filter network, the dissipation factors are evaluated as follows (refer to first edition of this book):

Equivalent δ for Lowpass and Highpass Filters

$$\delta = \frac{1}{Q_0} \tag{3.98}$$

Q_0 refers to the value of unloaded Q at the cutoff frequency of the practical lowpass and highpass filters.

Equivalent δ for Bandpass and Bandstop Filters

$$\delta = \frac{f_0}{\Delta f}\frac{1}{Q_0} \tag{3.99}$$

f_0 and Δf are the band center frequency and bandwidth of the filter, and Q_0 is the average unloaded Q of the resonant structures at band center.

Let us examine the example of an eight-pole 26-dB return loss Chebyshev filter. The group delay of the prototype network τ' is computed to be 5.73 s at $\omega = 0$. Let us now assume that the practical filters are derived and realized in a coaxial structure with an unloaded Q of 1000. The dissipation factor and loss for a lowpass and a bandpass filter are computed as follows:

Lowpass Filter. The dissipation factor is given by, $\delta = (1/1000) = 0.001$. Loss α at $\omega = \omega_c$ is therefore given by $\alpha = 20 \log e\, \delta\, \tau' = 8.686 \times 10^{-3} \times 5.73 = 0.05$ dB.

It should be noted that the loss is independent of the choice of the cutoff frequency of the lowpass filter. The loss depends solely on the choice of the prototype filter and the unloaded Q (defined at the cutoff frequency) of the practical structure.

Bandpass Filter. Let us assume a 1% bandwidth filter realized in a coax structure with an unloaded Q of 1000. The dissipation factor is evaluated as

$$\delta = \frac{f_0}{\Delta f}\frac{1}{Q_0} = \frac{100}{Q_0} = 0.1$$

By substituting a value of 0.1 for δ, the loss for the bandpass filter at $\omega = \omega_0$ is computed as 5.0 dB. This loss is 100 times the loss of the lowpass filter. The loss of the bandpass or bandstop filter depends on the percentage bandwidth, in addition to the choice of the prototype network and the quality factor of the practical structure. The loss does not depend on the choice of the center frequency *per se*, although Q_0 for a given structure varies with the choice of the center frequency. The dissipative loss for bandpass filters can be very high for narrow bandwidths and therefore require high-Q structures for most practical applications.

3.10 Asymmetric Response Filters

For some applications, it is desirable to have an asymmetric response in the amplitude or phase or both. This poses a challenge for the design of filter networks. The prototype filters are composed of lossless inductors and capacitors, terminated in a matching resistor. The result is a lowpass prototype network that is inherently symmetric with respect to the zero frequency. Also, the frequency transformation necessarily leads to networks that are symmetric with respect to the center of the passband. Consequently, as long as we confine a prototype network to real elements (L, C, and R), it is not feasible to realize an asymmetric response. However, in a bandpass circuit, it is possible to arrange the resonant circuits to yield an asymmetric frequency response. The challenge therefore lies in devising an equivalent lowpass prototype network, which, when transformed into bandpass filters via frequency transformation, yields the appropriate asymmetric response. Such a lowpass asymmetric prototype model would allow the use of numerous

and well-established general network synthesis techniques employed for symmetric lowpass prototype filters.

Baum [12] was the first to introduce a hypothetical frequency-invariant reactance (FIR) element in the design of filter networks. The element is used as a mathematical tool in the formulation of the lowpass prototype filter and becomes physically realizable only after a transformation from lowpass to bandpass or to bandstop filter. These hypothetical elements, introduced in the lowpass domain, render the filter response to be asymmetric with respect to zero frequency. Such a prototype circuit allows the realization of an asymmetric bandpass response via lowpass to bandpass frequency transformation. In the process, the fictitious FIR elements disappear as the frequency offsets to the resonant circuits, required in the realization of bandpass filters.

The challenge was how to incorporate this imaginary element in the synthesis procedure without violating the rules of circuit theory. Baum applied the concept of FIR elements to a few elementary synthesis procedures. By following Baum's introduction of the fictitious FIR elements, progress on the synthesis procedures incorporating such elements was slow, partly due to limited practical applications and equally due to the complexity of the problem. The textbooks [13, 14] that have addressed this problem are difficult to follow. Nevertheless, over time, this concept has been picked up by others and is now fully incorporated in the most general network synthesis techniques [15–18].

3.10.1 Positive Functions

The foundation of the classical theory is based on the concept of a positive real function. A network made up of L, C, and R elements can be described by the ratio of two polynomials in the complex frequency variable s with positive real constants. All the positive real functions exhibit this characteristic. The roots of such polynomials yield complex conjugate zeros instigating an amplitude and phase response that exhibits an even or odd symmetry about the zero frequency. The introduction of the fictitious constant reactances (FIR) elements in developing the lowpass prototype circuit models results in complex coefficients for the polynomials of the transfer function. This allows the unsymmetric distribution of the poles and zeros, giving rise to frequency response that is no longer symmetric with respect to the zero frequency. Such functions are called *positive functions* [14].

Ernst [19] has described the generalized two-port networks, including the hypothetical constant reactance element. The nomenclature for describing FIR elements is

$$\text{Frequency-invariant inductive element}: \quad X \geq 0, V(s) = jXI(s)$$
$$\text{Frequency-invariant capacitive element}: \quad B \geq 0, I(s) = jBV(s)$$

X and B represent the frequency-invariant inductance and capacitance, respectively. By applying the Laplace transform to the voltages and currents by assuming zero initial conditions, the complex impedance $Z(s)$ in the complex variable $s = \sigma + j\omega$ is defined as

$$Z(s) = \frac{V(s)}{I(s)}$$

Thus the impedances of FIR elements are

$$\text{FIR inductive element}: \quad X \geq 0, Z(s) = jX$$
$$\text{FIR capacitive element}: \quad B \geq 0, Z(s) = \frac{1}{jB} \tag{3.100}$$

Pictorially, FIR elements have been portrayed by a circle, a rectangle, or a set of parallel lines connected at each end. The alternative representations of this element are shown in Figure 3.16.

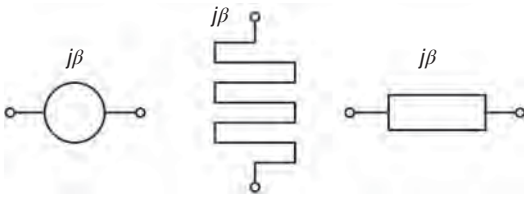


Figure 3.16 Different representations of the frequency-invariant reactance (FIR) element.

The impedance function of a loop or branch in a general network, consisting of R, L, C, and FIR elements, is

$$Z(s) = R + sL + \frac{1}{sC} + jX + \frac{1}{jB} \tag{3.101}$$

The driving point impedance of such a network is given by

$$\begin{aligned} Z(s) &= \frac{V(s)}{I(s)} \\ &= \frac{a_0 + a_1 s + a_2 s^2 + \cdots + a_n s^n}{b_0 + b_1 s^2 + b_2 s^2 \cdots + b_m s^m} \end{aligned} \tag{3.102}$$

where the coefficients a_i and b_i are now complex. Such functions are referred to as *positive functions*. If the FIR elements are not part of the network ($X = B = 0$), then all a_i and b_i are real, and $Z(s)$ becomes a positive real function. This begs the question: what constraints must be imposed on the positive functions to ensure physical realizability, when the lowpass prototype filter function is transformed to yield a real asymmetric bandpass filter? The imposed constraint is the same as in the case of the positive real functions, where the law of conservation of energy must hold. Invoking this constraint and working through the maze of complex power relationships [14], we find the following properties for the positive functions:

1) If $f(s)$ is a complex rational function in the complex variable $s = \sigma + j\omega$, then $f(s)$ is said to be a positive function if $\mathrm{Re}\{f(s)\} \geq 0$ for $\mathrm{Re}\{s\} \geq 0$.
   If, in addition, $f(s)$ is real for real s (i.e., when $j\omega = 0$), then $f(s)$ is said to be a positive real function.
2) There are no poles or zeros of $f(s)$ in the right half plane.
3) The poles and zeros of $f(s)$, along the $j\omega$ axis, are simple and have positive real residues.
4) The degrees of the numerator and denominator polynomials differ, at most, by 1. Thus, the number of finite poles and finite zeros differ, at most, by 1.
5) If $f(s)$ is a positive function, then $1/f(s)$ is also a positive function.
6) The linear combinations of the positive functions are a positive function.

These properties are very similar to those of positive real functions. The principal differences are as follows:

1) The coefficients of the numerator and denominator polynomials of $Z(s)$ or $Y(s)$ are real and positive for the positive real functions and complex for the positive functions.
2) The poles and zeros of $Z(s)$ or $Y(s)$ occur as complex conjugate pairs in the case of positive real functions, whereas there is no such restriction for positive functions.

This leads to the important question of how to represent the positive functions in terms of the lowpass prototype characteristic polynomials. Invoking the law of conservation of energy, it can be shown that the zeros of $P(s)$ must lie on the imaginary axis or appear as pairs, symmetrically located with respect to the imaginary axis. This derivation is detailed in Chapter 6. There is no requirement that such zeros have a corresponding complex conjugate pair as required

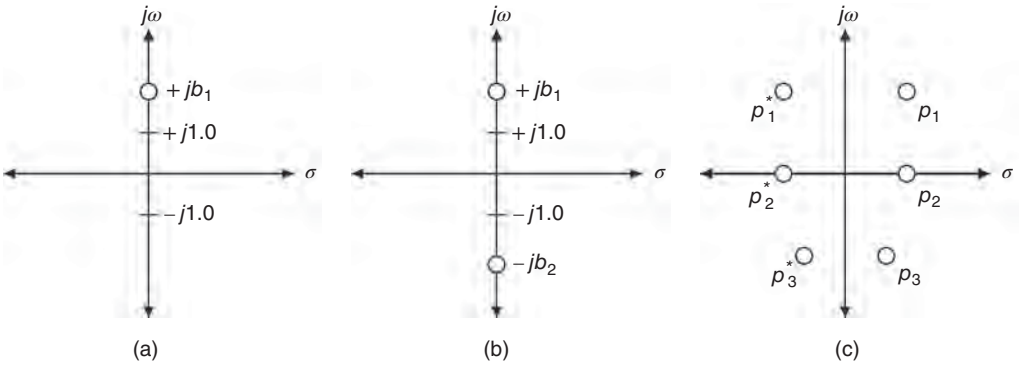


Figure 3.17 Permissible zero locations of $P(s)$ to realize asymmetric response filters: (a) asymmetric transmission zero location in the upper band, (b) asymmetric transmission zero locations in the upper and lower bands, and (c) asymmetric zero locations to realize linear phase response.

by positive real functions. Consequently, the polynomial $P(s)$ has complex coefficients. As for polynomial $F(s)$, all its zeros must lie on the imaginary axis, but they need not be distributed symmetrically around zero frequency, leaving $F(s)$ to possess complex coefficients as well. Of course, this results in complex coefficients for $E(s)$. Thus, for the general case, that includes the realization of symmetric or asymmetric lowpass prototype filters; the characteristic polynomials have the following properties:

1) The roots of $P(s)$ lie on the imaginary axis or appear as pairs of zeros located symmetrically with respect to the imaginary axis. The roots' degree is $\leq n$, where n is the order of the filter.
2) The roots of $F(s)$ lie on the imaginary axis where the degree is n.
3) $E(s)$ is a Hurwitz polynomial of degree n. All its roots lie in the left half plane of s.

As a result, $P(s)$ can be formed by any of the following factors or combinations of factors:

$$s \pm jb_i, \quad |b_i| > 1, \quad s^2 - \sigma_i^2, \quad (s - \sigma_i + j\omega_i)(s + \sigma_i + j\omega_i) \tag{3.103}$$

The polynomial $P(s)$ has the form

$$P(s) = s^m + jb_{m-1}s^{m-1} + b_{m-2}s^{m-2} + jb_{m-3}s^{m-3} + \cdots + b_0 \tag{3.104}$$

where m is the number of finite zeros of $P(s)$.

The polynomial $F(s)$ is formed by factors $(s + ja_i)$, where $|a_i| < 1$. It has the form

$$F(s) = s^n + ja_{n-1}s^{n-1} + a_{n-2}s^{n-2} + ja_{n-3}s^{n-3} + \cdots + a_0 \tag{3.105}$$

where n is the order of the filter. The possible locations of roots of $P(s)$ for lowpass prototype filters, normalized to unity cutoff frequency, are illustrated in Figure 3.17. A four-pole asymmetric filter with a Chebyshev passband and an equiripple stopband in the upper half of the frequency band is exhibited in Figure 3.18. Asymmetric classes of filters are dealt with extensively in Chapters 4, 6, and 7.

Summary

This chapter describes the synthesis process for the characteristic polynomials to realize the ideal classical prototype filters: maximally flat, Chebyshev, and elliptic function filters. The effect of uniform dissipation in filter elements is included by shifting the pole/zero pattern of

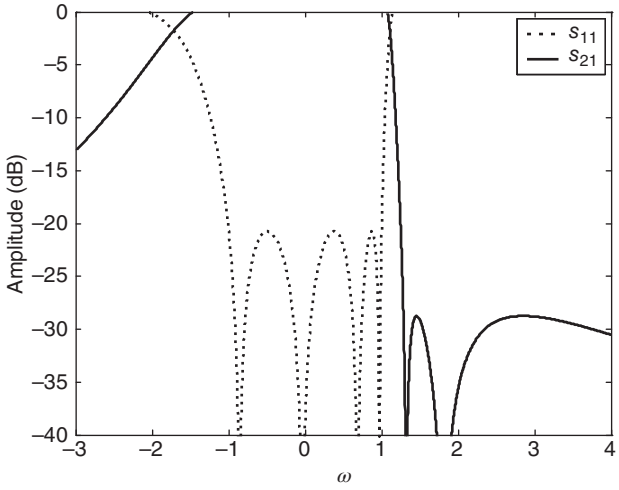


Figure 3.18 Lowpass amplitude response of a four-pole asymmetric filter.

the transfer function by the dissipation factor for the practical filter. The dissipation factor is derived in terms of the quality factor (unloaded Q) of the proposed structure and the bandwidth requirements of the practical filter. A brief review of the scaling of frequency and impedance of the prototype networks to realize practical filter networks is also included.

The chapter concludes with a discussion of filters that are not symmetric with respect to their center frequency. Such bandpass or bandstop filters are referred to as *asymmetric filters.* The synthesis of such filters in the lowpass prototype domain is accomplished by introducing two fictitious network elements, namely, the constant positive or negative reactance, referred to as a *FIR* element. This leads to transfer function polynomials with complex coefficients, implying that the characteristic polynomials no longer require their roots to be complex conjugate pairs. However, the law of the conservation of energy requires that the roots exhibit symmetry with respect to the imaginary axis to ensure physical realizability. The existing synthesis procedures can then be applied to synthesize asymmetric lowpass prototype filters. In the process of frequency transformation, the FIR elements disappear as the frequency offsets to the resonant circuits of bandpass and bandstop filters. Consequently, transfer functions with complex coefficients, within the constraint of polynomial roots exhibiting symmetry with respect to the imaginary axis, represent the most general lowpass prototype networks.

From the relationships described in this chapter, we can perform trade-offs to meet the filtering requirements of a communication channel for a wide range of applications, independent of the physical realization of the filter networks. In other words, it provides a reliable tool for system engineers and filter designers alike, to assess and specify realistic requirements for filters.

References

1 Lathi, B.P. (1965) *Signals, Systems and Communications*, Wiley, New York.
2 Van Valkenburg, M.E. (1960) *Modern Network Synthesis*, Wiley, New York.
3 Temes, G.C. and Mitra, S. (1973) *Modern Filter Theory and Design*, Wiley, New York.
4 Weinberg, L. (1957) Explicit formulas for Tschebyshev and Butterworth ladder networks. *Journal of Applied Physics*, **28**, 1155–1160.

5 Daniels, R.W. (1974) *Approximation Methods for Electronic Filter Design*, McGraw-Hill, New York.
6 Abramowitz, M. and Stegun, I.A. (1972) *Handbook of Mathematical Functions*, Dover Publications, New York.
7 Saal, R. and Ulbrich, E. (1958) On the design of filters by synthesis. *IRE Transactions on Circuit Theory*, **1958**.
8 Kudsia, C.M. and O'Donovan, V. (1974) *Microwave Filters for Communications Systems*, Artech House, Norwood, MA.
9 Papoulis, A. (1956) Frequency transformations in filter design. *IRE Transactions on Circuit Theory*, **CT-3**, 140–144.
10 Matthaei, G., Young, L., and Jones, E.M.T. (1980) *Microwave Filters, Impedance Matching Networks and Coupling Structures*, Artech House, Norwood, MA.
11 Hunter, I. (2001) *Theory and Design of Microwave Filters*, IEE.
12 Baum, R.F. (1957) Design of unsymmetrical bandpass filters. *IRE Transactions on Circuit Theory*, **1957**, 33–40.
13 Rhodes, J.D. (1976) *Theory of Electrical Filters*, Wiley, New York.
14 Belevitch, V. (1968) *Classical Network Theory*, Holden-Day, San Francisco.
15 Cameron, R.J. (1982) Fast generation of Chebyshev filter prototypes with asymmetrically-prescribed zeros. *ESA Journal*, **6**, 83–95.
16 Cameron, R.J. (1982) General prototype network synthesis methods for microwave filters. *ESA Journal*, **6**, 193–207.
17 Cameron, R.J. and Rhodes, J.D. (1981) Asymmetrical realizations for dual-mode bandpass filters. *IEEE Transactions on Microwave Theory and Techniques*, **MTT-29**, 51–58.
18 Bell, H.C. (1982) Canonical asymmetric coupled-resonator filters. *IEEE Transactions on Microwave Theory and Techniques*, **MTT-30**, 1335–1340.
19 Ernst, C. (2000) Energy storage in microwave cavity filter networks, Ph.D. thesis, Univ. Leeds, Aug. 2000.

Appendix 3A

Unified Design Charts

This appendix contains (i) UDCs for the Butterworth, Chebyshev and, quasielliptic filters with a single pair of transmission zeros, (ii) UDC for asymmetric filters with a single zero in the upper half of the frequency band and (iii) critical frequencies for the quasielliptic and asymmetric filters. An example of a channelizing filter in a satellite transponder is included to illustrate the versatility of these charts in determining the filter tradeoffs for a given specification.

Example 3A.1
Channelizing filters in a 6/4-GHz communication satellite system are required to separate 36 MHz channels with a 40 MHz spacing between them. By using UDCs, determine the trade-offs with respect to the order of the filter, and available passband bandwidths using an isolation constraint of 30 dB, or greater at $f_0 \pm 25$ MHz.

Solution
For this example, we shall assume a minimum return loss of, say, 20 dB in the passband, which is typical for such applications. Allowing for the operating environment and manufacturing tolerances, it is reasonable to assume a margin of ±0.25 MHz in the alignment of center frequency, implying that the passband bandwidth is no more than 39.5 MHz. Using the nomenclature of UDC, the criteria for filters trade-offs is $F \geq 20 + 30$ or ≥ 50 dB and

$$\frac{\omega_3}{\omega_1} \leq \frac{50}{36} \quad \text{or} \leq 1.39$$

$$\text{and } \frac{\omega_3}{\omega_1} \geq \frac{50}{39.5} \text{ or } 1.265$$

With this, the filter trade-offs are derived for the Chebyshev and quasielliptic (single pair of transmission zeros) filters using the UDCs. Trade-offs are summarized in Table 3A.1.

Table 3A.1 Filter trade-offs using unified design charts (UDCs).

		Chebyshev response		Quasielliptic response (single pair of transmission zeros)		
n	ω_3/ω_1	**Bandwidth (MHz)**	***F* (dB)**	ω_3/ω_1	**Bandwidth (MHz)**	***F* (dB)**
6		Not applicable		1.39	36	58
				1.26	39.7	50
7		Not applicable		1.39	36	67
				1.18	42.4	50
8	1.39	36	54	1.39	36	77
	1.34	37.3	50	1.125	44.45	50
9	1.39	36	62	Not necessary		
	1.265	39.5	50			

Key: n—order of filter; F—characteristic factor in dB; ω_3/ω_1—ratio of isolation bandwidth to passband bandwidth; $F \geq 50$ dB; $\omega_3/\omega_1 \leq 1.39; \geq 1.265$.

For the Chebyshev response filters, the minimum order of filter to satisfy the requirements is eight (Tables 3A.2–3A.4). However, such a design provides very little margin to allow for the operating environment, primarily the operating temperature range. Therefore, the realistic design is a Chebyshev filter of the ninth order. For such a design, the available bandwidth can be chosen between 36 and 39.5 MHz with the corresponding values of F varying between 62 and 50 dB. Therein lies the trade-off between bandwidth and isolation, or the return loss or a combination thereof. Greater bandwidth results in a better passband response, whereas higher isolation minimizes interchannel interference. A choice of $F = 55$ dB provides a margin of 5 dB, which can be shared between the return loss and isolation, say, a nominal return loss of 22 dB and a nominal isolation of 33 dB. The corresponding ω_3/ω_1 is 1.32, which yields a passband bandwidth of nearly 38 MHz. This represents a reasonable compromise and first-cut design.

Similar reasoning for the quasielliptic case yields a minimum order of six to satisfy the requirements. Here again, we choose a value of $F = 55$ dB, and as in the previous case,

Table 3A.2 Filter trade-offs summary.

Filter design	Passband bandwidth (MHz)	Return loss in passband	Isolation at $f_0 \pm 25$ MHz (dB)
Ninth-order Chebyshev	38	22	33
Sixth-order quasielliptic	37.7	22	33
Seventh-order quasielliptic	39	23	37

Table 3A.3 Critical frequencies for quasielliptic filters with a single pair of transmission zeros.

		Critical frequencies				
N	F (dB)	Attenuation zeros				Attenuation poles
4	40	0.4285	0.9419			1.5169
	50	0.4085	0.9343			1.9139
	60	0.3973	0.9268			2.4692
4	40	0.6567	0.9672			1.2763
	50	0.6321	0.9615			1.4921
	60	0.6161	0.9577			1.7870
6	40	0.3054	0.7785	0.9794		1.1695
	50	0.2917	0.7568	0.9752		1.3042
	60	0.2818	0.7415	0.9723		1.4869
7	40	0.5037	0.8476	0.9860		1.1137
	50	0.4863	0.8301	0.9807		1.2052
	60	0.4728	0.8170	0.9807		1.3293
8	40	0.2299	0.6327	0.8896	0.9899	1.0812
	50	0.2227	0.6158	0.8757	0.9876	1.1472
	60	0.2168	0.6021	0.8649	0.9859	1.2367

Table 3A.4 Critical frequencies for asymmetric filters with a single zero in the upper half of the frequency band.

		Critical frequencies				
N	***F* (dB)**	**Attenuation zeros**				**Attenuation poles**
4	40	−0.8943	−0.1994			1.3378
		0.5829	0.9600			
	50	−0.9008	−0.2415			1.6321
		0.5336	0.9509			
	60	−0.9060	−0.2748			2.0587
		0.4961	0.9440			
5	40	−0.9340	−0.4631			1.1925
		0.2145	0.7488	0.9766		
	50	−0.9367	−0.4843			1.3578
		0.1743	0.7151	0.9709		
	60	−0.9392	−0.5028			1.5918
		0.1406	0.6881	0.9665		
6	40	−0.9554	−0.6246	−0.0871		1.124
		0.4593	0.8339	0.9847		
	50	−0.9568	−0.6354	−0.1115		1.2297
		0.4267	0.8101	0.9809		
	60	−0.9580	−0.6454	−0.1337		1.3776
		0.3985	0.7907	0.9779		
7	40	−0.9681	−0.7259	−0.3059		1.0864
		0.1809	0.6090	0.8825	0.9893	
	50	−0.9688	−0.7318	−0.3203		1.198
		0.1582	0.5834	0.8651	0.9866	
	60	−0.9695	−0.7374	−0.3339		1.2618
		0.1372	0.5608	0.8507	0.9844	
8	40	−0.9761	−0.7923	−0.4610	−0.474	1.0636
		0.3677	0.7056	0.9127	0.9921	
	50	−0.9765	−0.7958	−0.4967	−0.0622	1.1176
		0.3480	0.6853	0.8995	0.9900	
	60	−0.9769	−0.7991	−0.4782	−0.0764	1.1922
		0.3296	0.6672	0.8885	0.9884	

$\omega_3/\omega_1 = 1.325$. This yields a passband bandwidth of 37.7 MHz. Thus, a sixth-order quasielliptic filter with a single pair of transmission zeros satisfies the requirement. By going a step further, a seventh-order quasielliptic filter offers wider margins at the cost of an increase in the order of filter by one. For example, a choice of 39 MHz passband bandwidth implies ω_3/ω_1 of 1.38. The corresponding value of F is 60 dB. This represents a 10 dB margin, shared between return loss and isolation. A good choice is to use a return loss of 23 dB and isolation of 37 dB. The three choices are summarized in Table 3A.2.

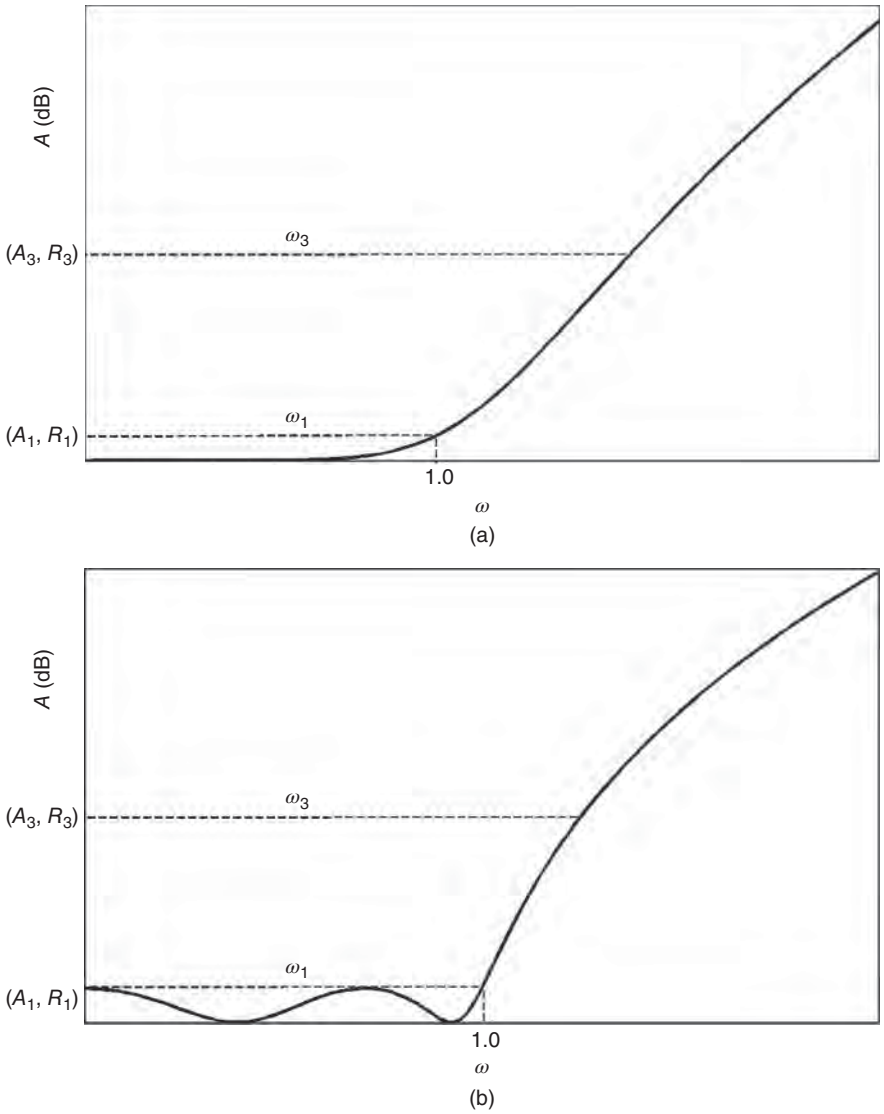
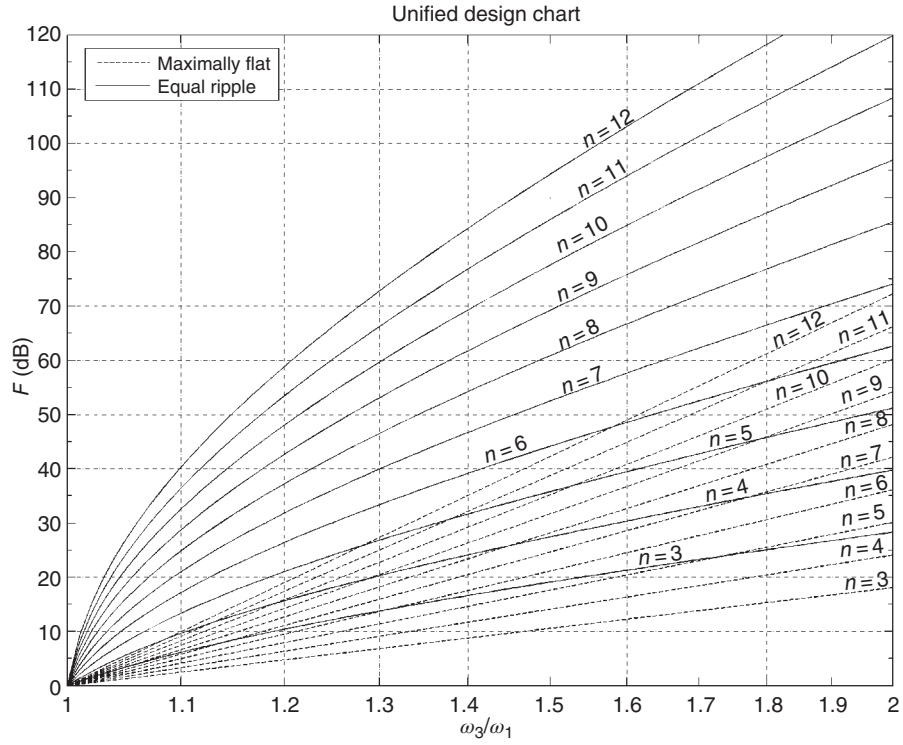


Figure 3A.1 UDC nomenclature for (a) Butterworth and (b) Chebyshev filters.

A further trade-off can be made, dB for dB, between the return loss and isolation. This example exhibits the simplicity of UDCs in developing filter trade-offs. Further optimization can be carried out by simulating these first-cut designs for both amplitude and phase/group delay response and then making minor adjustments. For many applications, use of UDC alone can be quite adequate for the design.



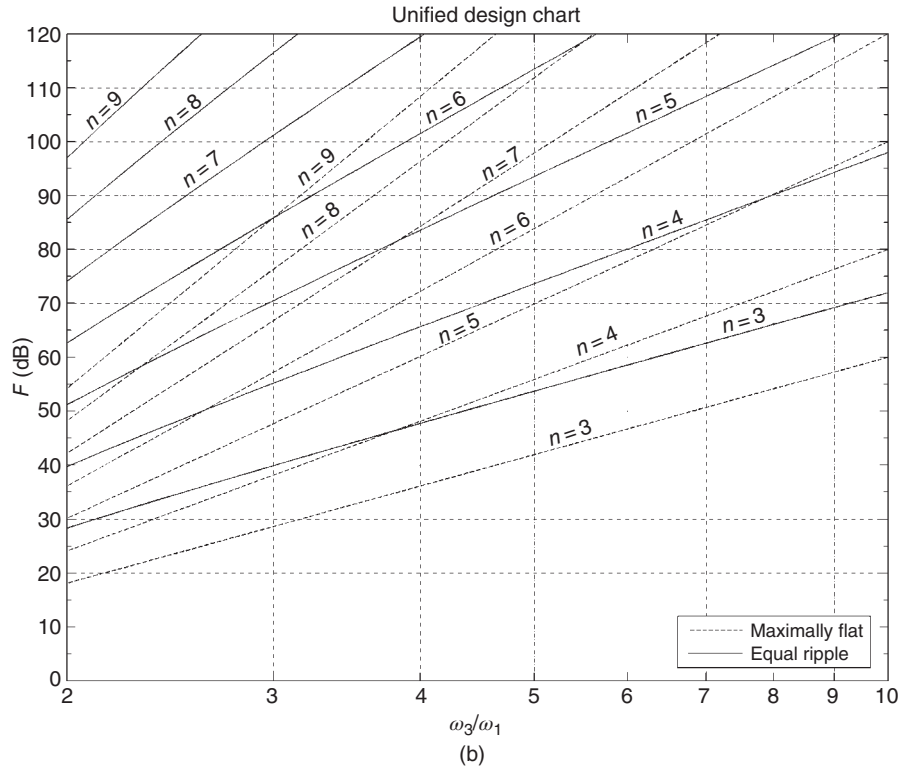


Figure 3A.2 Unified design chart for Butterworth and Chebyshev filters.

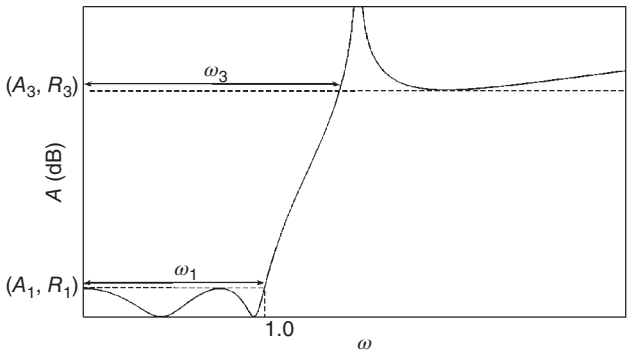


Figure 3A.3 UDC nomenclature for quasielliptic filter with a single pair of transmission zeros.

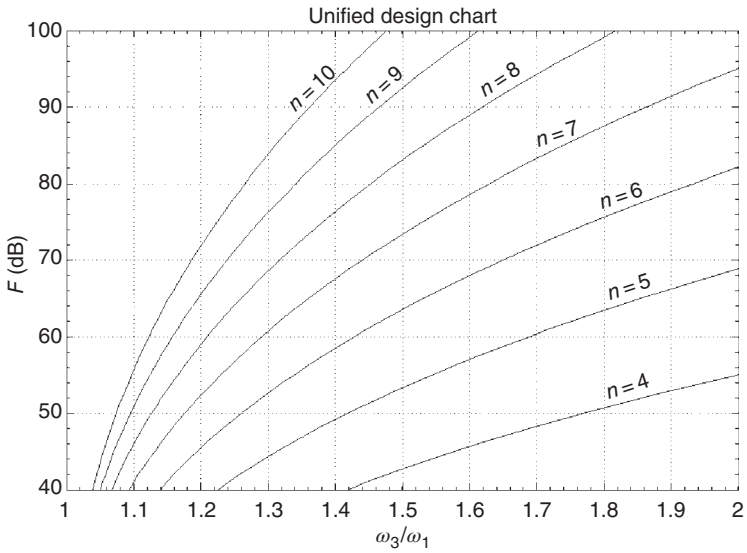


Figure 3A.4 Unified design chart for quasielliptic filters with a single pair of transmission zeros.

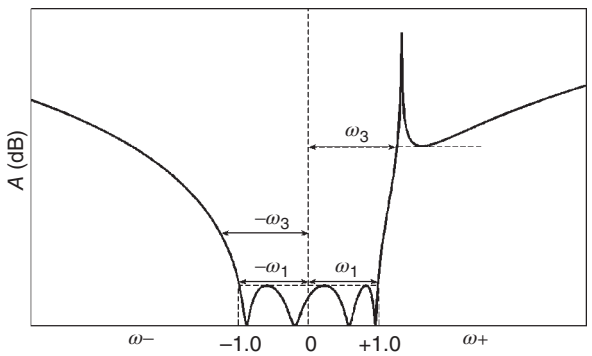


Figure 3A.5 Asymmetric filter with a single transmission zero in the upper half of frequency band.

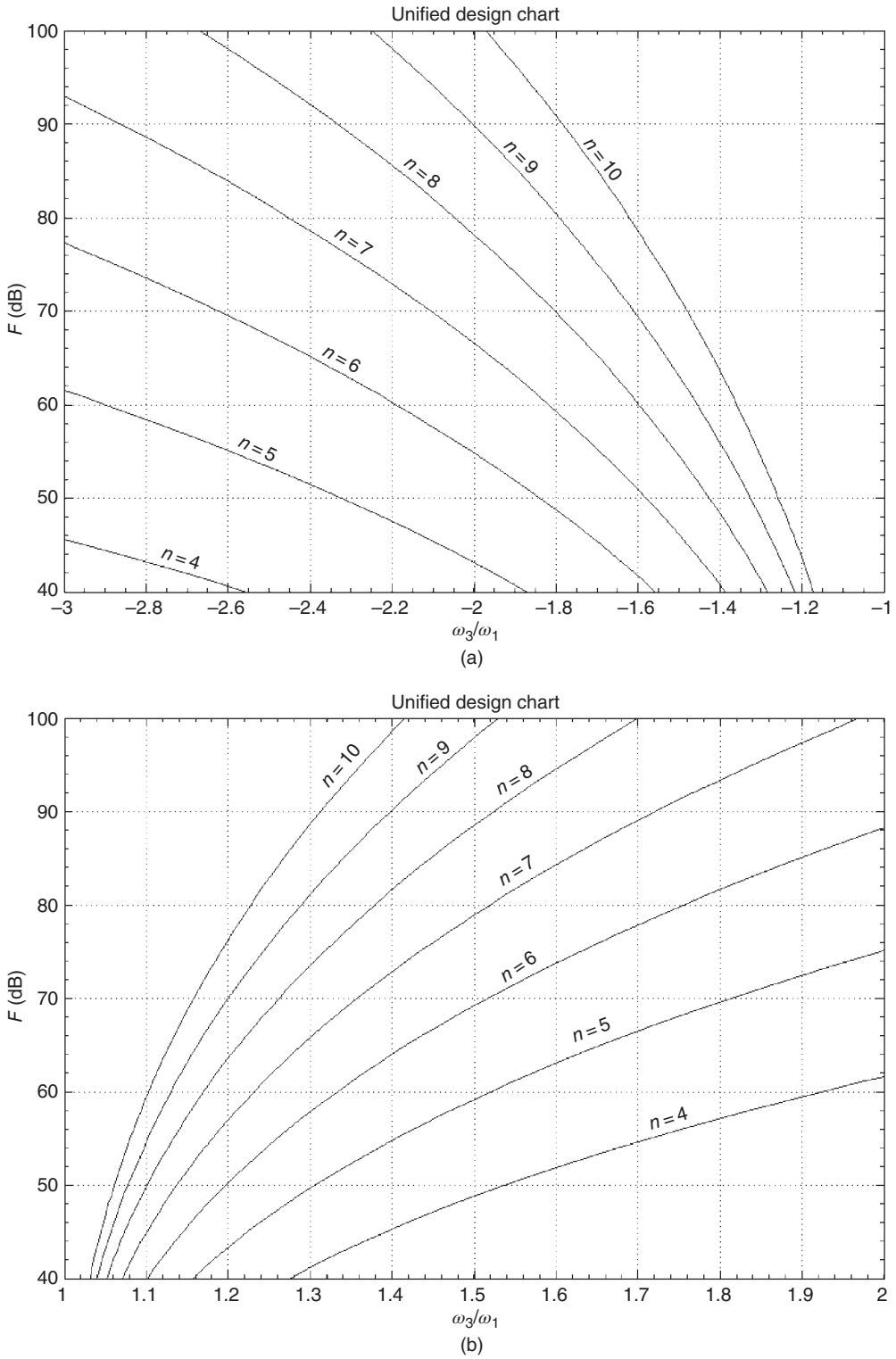


Figure 3A.6 Unified design chart for asymmetric filters with a single zero in the upper half of the frequency band: (a) lower half of frequency band and (b) upper half of frequency band.

4

Computer-Aided Synthesis of Characteristic Polynomials

A lowpass prototype filter can be completely characterized in terms of its critical frequencies, the poles, and zeros of the transfer function. This easily leads to the possibility of generating such critical frequencies for a desired response shape by adopting computer-aided optimization techniques. For most types of practical filters, analytic techniques already exist to compute the critical frequencies. However, such techniques are restricted to the well-defined filter functions such as the equiripple Chebyshev and elliptic functions or the monotonically increasing maximally flat function. Any departure from such filters necessitates a computer-aided optimization procedure.

This chapter is devoted to the synthesis of the characteristic polynomials of lowpass lossless prototype filters using an efficient computer-aided optimization technique. It includes minimum phase as well as linear phase filters exhibiting symmetric or asymmetric response. The technique is completely general and can be readily adapted to synthesize characteristic polynomials with an arbitrary response. Classical filters such as Chebyshev or elliptic function filters can be derived as special cases of the more general characteristic polynomials.

4.1 Objective Function and Constraints for Symmetric Lowpass Prototype Filter Networks

The transmission coefficient of a symmetric minimum phase lowpass prototype filter network is given by (Section 3.3)

$$|t(s)|^2_{s=j\omega} = \frac{1}{1+\varepsilon^2|K(s)|^2_{s=j\omega}}$$

where

$$K(s) = \varepsilon \frac{s(s^2+a_1^2)(s^2+a_2^2)\cdots(s^2+a_i^2)\cdots}{(s^2+b_1^2)(s^2+b_2^2)\cdots(s^2+b_i^2)\cdots}$$

for odd-order filter functions and

$$K(s) = \varepsilon \frac{(s^2+a_1^2)(s^2+a_2^2)\cdots(s^2+a_i^2)\cdots}{(s^2+b_1^2)(s^2+b_2^2)\cdots(s^2+b_i^2)\cdots} \tag{4.1}$$

for even-order filters.

The a and b terms are the attenuation zeros and poles, respectively (referred to as the *critical frequencies*), and ε is the ripple factor. The number of independent variables is represented by the finite critical frequencies. Attenuation zeros at origin or poles at infinity represent fixed

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

frequencies and are, therefore, not independent. ε provides a tradeoff between the passband and the stopband responses for a given characteristic factor F (Section 3.7). It does not represent an independent variable. Choice of a specific value of F implies an equality constraint. Alternatively, critical frequencies can be generated with F as a parameter. This is more efficient since it dispenses with the equality constraint in the optimization process and leads to the generation of unified design charts (Section 3.7).

Next comes the question as to which filter performance parameters are essential to determine a practical design. All practical applications require a maximum permissible ripple in the passband and a minimum of isolation in the stopband. Phase or group delay response, in the passband and elsewhere, is uniquely related to the amplitude response for minimum phase filters by Hilbert's transforms. For such filters, it is sufficient to optimize the amplitude response and accept the resulting phase response. We are concerned primarily with minimum phase filters, although the subsequent analysis is general enough to analyze and optimize nonminimum phase types of filters as well.

To constrain the amplitude ripples in the passband or stopband, we must determine the frequencies at which these amplitude maxima and minima occur. This is accomplished by differentiating the equation for the characteristic function $K(s)$ and equating it to zero:

$$\frac{\partial}{\partial s}|K(s)|^2 = 0, \quad \frac{\partial}{\partial s}\left|\frac{F(s)}{P(s)}\right| = 0, \quad \text{or} \quad |\dot{F}(s)P(s)| - |F(s)\dot{P}(s)| = 0 \tag{4.2}$$

Here, $\dot{F}(s)$ and $\dot{P}(s)$ represent the differentials with respect to the complex variable s. This determines the frequencies corresponding to the amplitude maxima in the passband and minima in the stopband for an arbitrary choice of ε. The transmission and reflection coefficients, or the transmission and return loss in decibels (dB), are computed as described by the equations in Section 3.4. The cutoff frequency ω_c, corresponding to the maximum ripple in the passband, is determined by solving for the highest frequency at which $t(s)$ is equal to the maximum ripple in the passband. Alternatively, $\omega_c = 1$ can be included as a constraint in the optimization process. All the other frequencies are then normalized with respect to $\omega_c = 1$.

For practical applications, the most convenient parameters for specifying the amplitude response are the return loss in passband and the transmission loss in the stopband, both parameters having units of dB.

By denoting the return loss and transmission loss at the ith attenuation maxima or minima as $R(i)$ and $T(i)$, respectively, the objective function U has terms of the following three types:

$$U = U_1 + U_2 + U_3$$

where

$$\begin{aligned} U_1 &= ABS[|R(i) \sim R(j)| + A_{ij}] \\ U_2 &= ABS[|T(i) \sim T(j)| + B_{ij}] \\ U_3 &= ABS[|R(i) \sim T(j)| + C_{ij}] \end{aligned} \tag{4.3}$$

where A_{ij}, B_{ij}, and C_{ij} represent arbitrary constants that constrain the differences in the amplitudes at the attenuation maxima and minima to some arbitrary values. For the equiripple passband, $A_{ij} = 0$, and for the equiripple stopband, $B_{ij} = 0$.

The term U_3 represents the characteristic factor F and is best utilized as an equality constraint. Thus, in a generalized form, the objective function U is represented by

$$U = \sum_{i \neq j} ABS[|R(i) - R(j)| - A_{ij}] + \sum_{k \neq \ell} ABS[|T(k) - T(\ell)| - B_{k\ell}] + U_3 \tag{4.4}$$

where the summations extend over all the attenuation maxima in the passband for the first term and over all the attenuation minima in the stopband for the second term. The return loss and transmission loss are calculated by

$$\begin{aligned} T &= -10\log|t(s)|^2_{s=j\omega}\text{dB} \\ R &= -10\log\left[1-|t(s)|^2_{s=j\omega}\right]\ \text{dB} \end{aligned} \tag{4.5}$$

For the prototype filter normalized to the unity cutoff, the constraints on the independent variables (the critical frequencies) are

$$\begin{aligned} &0 \le a_1, a_2, \ldots, < 1 \\ &b_1, b_2, \ldots, > 1 \end{aligned} \tag{4.6}$$

These constraints simply imply that the attenuation zeros lie in the passband and the transmission zeros lie in the stopband.

4.2 Analytic Gradients of the Objective Function

For the optimization process, we need to define an unconstrained artificial function that includes the objective function, as well as the inequality and equality constraints. Such an artificial function U_{art} is given by [1, 2]

$$U_{\text{art}} = U + \frac{r}{\sum_i|\phi_i|} + \frac{\sum_k|\psi_k|^2}{\sqrt{r}} \tag{4.7}$$

where r = a variable of optimization to control the imposed constraints. (It is a positive number with a typical range of r between 10^{-4} and 1.)

U = the unconstrained objective function

ϕ_i = ith inequality constraint

ψ_k = kth equality constraint

At the optimum value, $U_{\text{art}} = 0$. This implies that each individual term comprising U_{art} must be zero. The unconstrained function U tends to zero at the optimum value. To ensure the convergence of the constraint terms to zero, r is made smaller such that

$$\frac{r}{\sum_i|\phi_i|} \to 0 \quad \text{and} \quad \frac{\sum_k|\psi_k|^2}{\sqrt{r}} \to 0 \quad \text{as} \quad r \to 0$$

This is done in steps to ensure the smooth convergence of U_{art} to the desired level of accuracy. The gradient of U_{art}, with respect to the attenuation zeros a_i, is computed as

$$\begin{aligned} g_i &= \frac{\partial U_{\text{art}}}{\partial a_i} = \frac{\partial U}{\partial a_i} + r\frac{\partial}{\partial a_i}\frac{1}{\sum_i|\phi_i|} + \frac{1}{\sqrt{r}}\frac{\partial}{\partial a_i}\sum_k|\psi_k|^2 \\ &= \frac{\partial U}{\partial a_i} + \Phi + \Psi \end{aligned} \tag{4.8}$$

The same holds true for the attenuation poles. Each term on the right-hand side is evaluated independently in the following sections.

4.2.1 Gradient of the Unconstrained Objective Function

A typical form of the unconstrained function U is

$$U = |R(1) - R(2)| + |R(2) - R(3)| + \cdots + |T(4) - T(5)| + |T(5) - T(6)| + \cdots \tag{4.9}$$

where $R(k)$ and $T(k)$ are the values of the return loss and the transmission loss in dB at the kth attenuation maxima in the passband or minima in the stopband. Considering a typical term, we write

$$\begin{aligned} U_{ik} &= |R(i) - R(k)| = R(i) - R(k) \quad \text{if } R(i) \geq R(k) \\ &= -(R(i) - R(k)) \quad \text{if } R(i) \leq R(k) \end{aligned} \tag{4.10}$$

$$\therefore \frac{\partial U_{ik}}{\partial a_i} = \pm \left[\frac{\partial R(i)}{\partial a_i} - \frac{\partial R(k)}{\partial a_i} \right] \tag{4.11}$$

The positive sign holds when $R(i) \geq R(k)$ and vice versa. The partial derivatives $\partial R/\partial a_i$ are determined by considering the transmission function

$$|t|^2 = \frac{1}{1 + |K(s)|^2} = 1 - |\rho|^2 \tag{4.12}$$

where ρ is the reflection coefficient. The critical frequencies are independent of the choice of the ripple factor. As a consequence, it is convenient to assume ε to be unity. By making use of equation (4.1), we obtain

$$\begin{aligned} \therefore \frac{\partial |t|^2}{\partial a_i} &= -\frac{1}{[1 + |K(s)|^2]^2} 2K(s) \frac{\partial |K(s)|}{\partial a_i} \\ &= -4 \frac{a_i}{(s^2 + a_1^2)} \frac{K^2(s)}{[1 + K^2(s)]^2} \end{aligned} \tag{4.13}$$

Making use of equation (4.5), the gradients in terms of transmission and reflection loss in dB are given by [3]

$$\begin{aligned} \therefore \frac{\partial T}{\partial a_i} &= 40 \log e \frac{K^2(s)}{1 + K^2(s)} \frac{a_i}{(s^2 + a_i^2)} \\ &= 40 \log e |\rho|^2 \frac{a_i}{(s^2 + a_i^2)} \end{aligned} \tag{4.14}$$

and

$$\begin{aligned} \frac{\partial R}{\partial a_i} &= -40 \log e \frac{1}{1 + K^2(s)} \frac{a_i}{(s^2 + a_i^2)} \\ &= -40 \log e |t|^2 \frac{a_i}{(s^2 + a_i^2)} \end{aligned} \tag{4.15}$$

where e is the natural base for logarithms. In a similar manner, the gradients with respect to the attenuation poles b_i are computed as

$$\frac{\partial T}{\partial b_i} = -40 \log e |\rho|^2 \frac{b_i}{(s^2 + b_i^2)} \tag{4.16}$$

and

$$\frac{\partial R}{\partial b_i} = 40 \log e |t|^2 \frac{b_i}{(s^2 + b_i^2)} \tag{4.17}$$

Thus, the partial derivatives of the transmission or return loss in dB, at any frequency with respect to the independent variables, are determined analytically and thus the gradient of the unconstrained function U.

4.2.2 Gradient of the Inequality Constraint

The inequality constraint Φ has the form

$$\Phi = \frac{r}{|\phi_1|} + \frac{r}{|\phi_2|} + \cdots \tag{4.18}$$

where r is a positive number, and the Φ terms have the following forms:

$$\phi_k = \begin{cases} a_i \\ 1 - a_i \\ b_i - 1 \\ N - b_i, \quad N \text{ is a positive number greater than 1} \end{cases} \tag{4.19}$$

$$\begin{aligned} \therefore \frac{\partial}{\partial a_i}\frac{1}{|\phi_k|} &= -\frac{1}{\phi_k^2}\frac{\partial \phi_k}{\partial a_i}, \quad \phi_k > 0 \\ &= +\frac{1}{\phi_k^2}\frac{\partial \phi_k}{\partial a_i}, \quad \phi_k < 0 \end{aligned} \tag{4.20}$$

where a_i represents the ith independent variable. The gradients for the various forms of Φ are described in Table 4.1 [3].

As a result, the gradient of Φ has the following form:

$$\begin{aligned} \frac{\partial \Phi}{\partial a_i} + \frac{\partial \Phi}{\partial b_i} = r \cdot \Bigg[&\sum \left(\mp \frac{1}{a_i^2} \right) + \sum \pm \left[\frac{1}{(1 - a_i^2)} \right] \\ &+ \sum \left(\mp \frac{1}{(b_i - 1)^2} \right) + \sum \left(\pm \frac{1}{(N - b_i)} \right) \Bigg] \end{aligned} \tag{4.21}$$

This includes the inequalities associated with the attenuation zeros, as well as the poles.

4.2.3 Gradient of the Equality Constraint

The gradient of the equality constraint is computed by

$$\Psi = \frac{1}{\sqrt{r}} \frac{\partial}{\partial a_i} \sum |\psi_k|^2$$

where ψ_k has the form

$$\psi_k = A + R(1) + T(2) - T(1) - R(2) + \cdots \tag{4.22}$$

Table 4.1 Gradients of the inequality constraint.

Form of Φ	Gradient of Φ	Constraints
a_i	$\mp \frac{1}{a_i^2}$	$a_i \gtrless 0$
$1 - a_i$	$\pm \frac{1}{(1.0 - a_i)^2}$	$(1.0 - a_i) \gtrless 0$
$b_i - 1$	$\mp \frac{1}{(b_i - 1)^2}$	$(b_i - 1) \gtrless 0$
$N - b_i$	$\pm \frac{1}{(N - b_i)^2}$	$(N - b_i) \gtrless 0$

where A is a constant. $R()$ and $T()$ represent the return loss and transmission loss in dB at a maxima in the passband and minima in the stopband, respectively. The gradient is evaluated as follows:

$$\frac{\partial}{\partial a_i}|\psi_k|^2 = 2\psi_k \cdot \frac{\partial \psi_k}{\partial a_i} \tag{4.23}$$

$$\Psi = \frac{2}{\sqrt{r}} \sum_k \psi_k \left[\frac{\partial R(1)}{\partial a_i} + \frac{\partial T(2)}{\partial a_i} + \frac{\partial T(1)}{\partial a_i} - \frac{\partial R(2)}{\partial a_i} + \cdots + \frac{\partial R(1)}{\partial b_i} + \frac{\partial T(1)}{\partial b_i} + \cdots \right] \tag{4.24}$$

The values of $R()$ or $T()$ and their partial derivatives are derived by equations (4.14), (4.15)–(4.17). Thus, the gradients of the unconstrained objective function U_{art} are determined analytically.

4.3 Optimization Criteria for Classical Filters

In this section, we develop the appropriate objective functions for optimization to generate the known classes of filter functions.

4.3.1 Chebyshev Function Filters

This function is characterized by an equiripple response in the passband and a monotonically rising stopband. The objective function and constraints can be represented as

$$U = |R(1) - R(2)| + |R(2) - R(3)| + \cdots$$
$$0 < a_1, a_2, \ldots, < 1 \tag{4.25}$$

where $R(k)$ is the return loss at the kth attenuation maxima and a_i are the critical frequencies. For example, a sixth-order filter has the objective function

$$U = |R(1) - R(2)| + |R(2) - R(3)| + |R(3) - R(4)| \tag{4.26}$$

where $R(4)$ is characterized by $s = j$, as depicted in Figure 4.1. This term is used to normalize the critical frequencies to a unity cutoff frequency in the passband. By adopting the analytic gradients described by equations (4.21) and (4.24) in the optimization process, the critical frequencies are determined to be 0.2588, 0.7071, and 0.9659.

4.3.2 Inverse Chebyshev Filters

These filters are characterized by a monotonically increasing passband (maximally flat) and an equiripple stopband:

$$K(s) = \frac{s^6}{(s^2 + b_1^2)(s^2 + a_2^2)}$$

The response of such a filter with arbitrary locations of attenuation poles is illustrated in Figure 4.2. The objective function for an equiripple stopband normalized to unity cutoff frequency is as follows:

$$U = |T(1) - T(2)| + |T(2) - T(3)|, \quad b_1, b_2 > 1 \tag{4.27}$$

The critical frequencies are readily evaluated as 1.0379 and 1.4679.

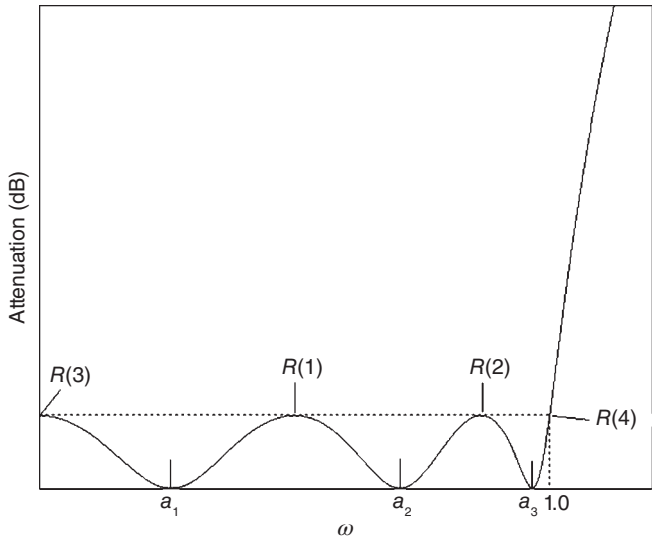


Figure 4.1 Attenuation zeros and maxima of a sixth-order Chebyshev filter.

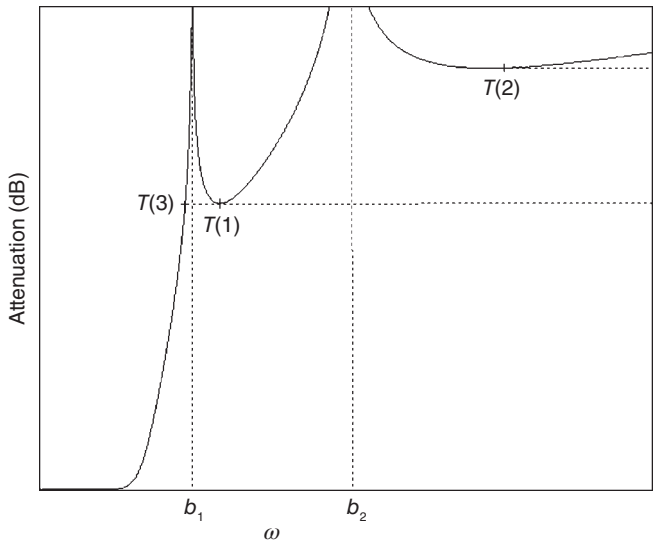


Figure 4.2 Attenuation poles and minima of a sixth-order inverse Chebyshev filter.

4.3.3 Elliptic Function Filters

These filters are characterized by an equiripple passband and equiripple stopband. For a six-pole class C elliptic function filter, displayed in Figure 4.3, the characteristic function has the form

$$K(s) = \frac{s^2(s^2 + a_1^2)(s^2 + a_2^2)}{(s^2 + b_1^2)(s^2 + b_2^2)}$$

and the objective function has the form

$$U = |R(1) - R(2)| + |R(2) - R(5)| + |T(3) - T(4)| \tag{4.28}$$

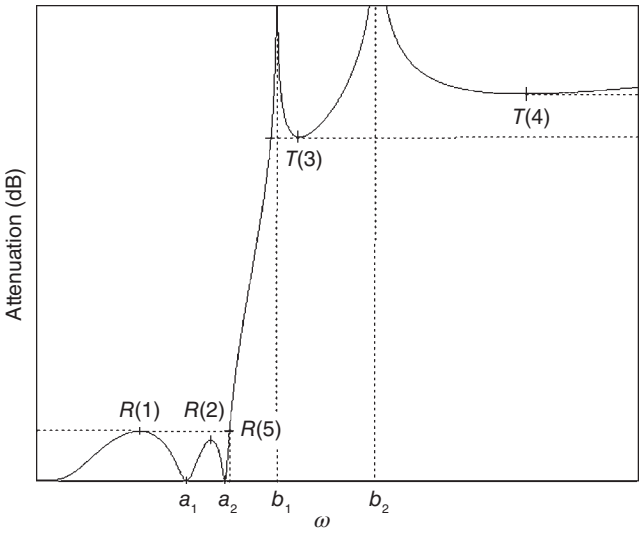


Figure 4.3 Attenuation maxima and minima of a sixth-order lowpass prototype class C elliptic function filter.

$R(5)$ is characterized by $s = j$ to incorporate the unity cutoff frequency. For filters with transmission zeros, it is necessary to impose some constraint on the characteristic factor F (Section 3.7), which, for the example under consideration, is expressed as

$$\begin{aligned} F &= |R(2) + T(3)| - |T(2) + R(3)| \\ &= A(\text{say}) \end{aligned} \tag{4.29}$$

As an equality constraint, this is written as follows:

$$\Psi = A - |R(2) + T(3)| + |T(2) + R(3)| \tag{4.30}$$

In general, equality constraints increase the computation time significantly. One way to bypass this equality constraint and at the same time decrease the number of independent variables by one is to fix one of the attenuation pole frequencies. This automatically constrains the characteristic factor F. From a practical standpoint, this is most useful since tradeoffs are generated by selecting the pole frequencies in the neighborhood where a high isolation is specified. With this approach, computations can be performed very efficiently. For example, we chose b_1 to be fixed at 1.2995. The remaining critical frequencies are then determined as $a_1 = 0.7583$, $a_2 = 0.9768$, and $b_2 = 1.6742$.

4.4 Generation of Novel Classes of Filter Functions

In order to demonstrate an unconventional filter, let us consider an eighth-order filter defined by the characteristic function

$$K(s) = \frac{s^4(s^2 + a_1^2)(s^2 + a_2^2)}{(s^2 + b_1^2)(s^2 + b_2^2)} \tag{4.31}$$

This function is characterized by a pair of attenuation poles, a double zero at the origin, and two finite reflection zeros as described in Figure 4.4. Here we introduce the following three cases for optimization.

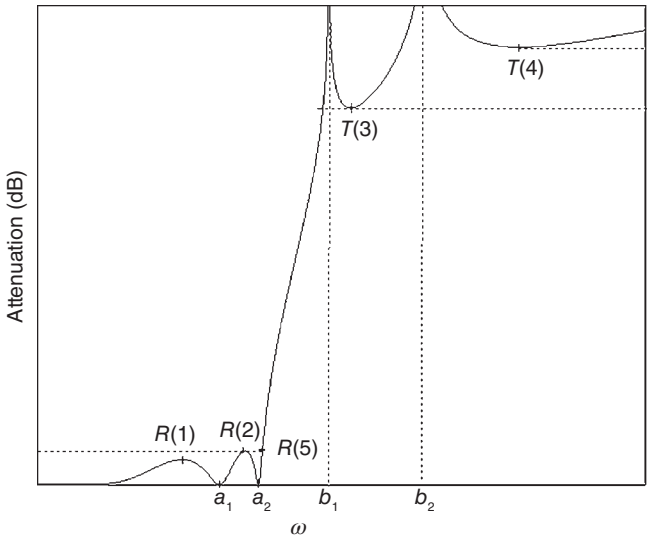


Figure 4.4 Lowpass prototype response of an eighth-order filter with two attenuation poles and a double zero at the origin.

4.4.1 Equiripple Passbands and Stopbands

Assuming a fixed attenuation pole $b_1 = 1.25$, we express the objective function and constraints as

$$U = |R(1) - R(2)| + |R(2) - R(5)| + |T(3) - T(4)|$$
$$0 < a_1, a_2, \ldots < 1, \quad b_2 > 1 \tag{4.32}$$

$R(5)$ is used to constrain the cutoff frequency to unity.

The computed critical frequencies are summarized in Table 4.2.

4.4.2 Nonequiripple Stopband with an Equiripple Passband

Here, we consider an equiripple passband and a nonequiripple stopband. As before, we fix the first attenuation pole at $b_1 = 1.25$, and the objective function and constraints are then calculated by

$$U = |R(1) - R(2)| + |R(2) - R(5)| + |T(3) - T(4) - B_{34}|$$
$$0 < a_1, a_2, \ldots, < 1, \quad b_2 > 1 \tag{4.33}$$

Table 4.2 Critical frequencies for an optimized eight-pole filter with two attenuation poles and a double attenuation zero at the origin and fixed attenuation pole at $b_1 = 1.25$.

	Computed critical frequencies			Characteristic factor
B_{34}(dB)	a_1	a_2	b_2	F (dB)
0	0.8388	0.9845	1.4671	67.4
10	0.8340	0.9839	1.6211	63.1
−10	0.8636	0.9878	1.1541	63.25

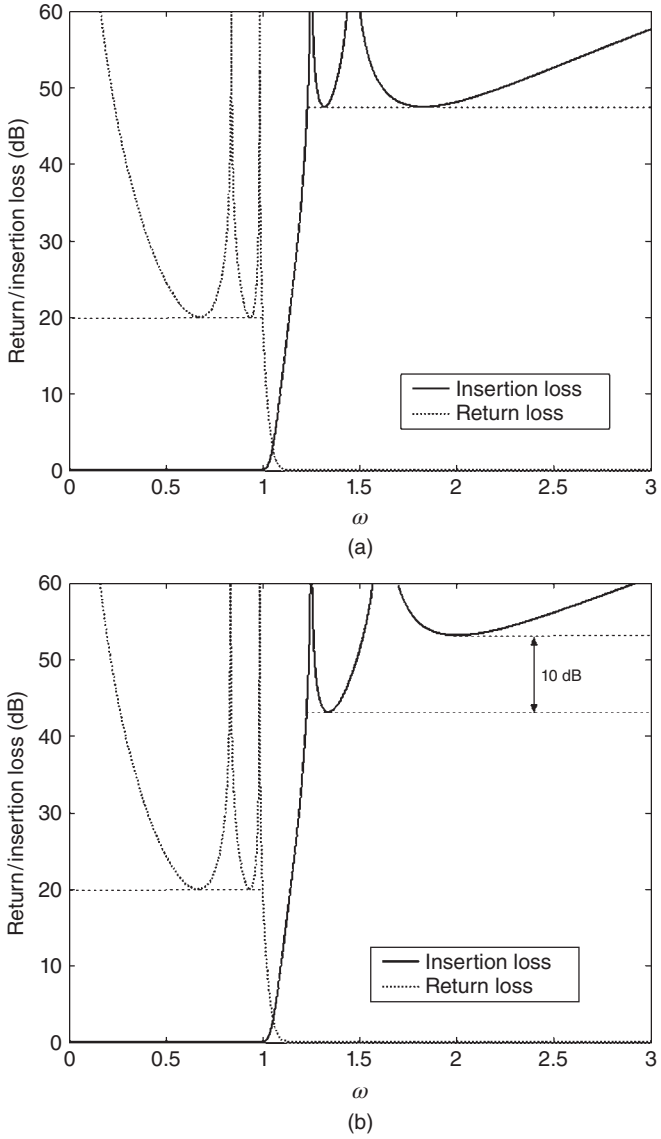


Figure 4.5 Frequency response of an optimized lowpass prototype eight-pole filter with two attenuation poles and a double zero at the origin: (a) an equiripple stopband; (b) a 10 dB differential in stopband minima.

The term B_{34} is an arbitrary constant. It represents the difference between the attenuation minima at $T(3)$ and $T(4)$ in dB. The computed critical frequencies for $B_{34} = \pm 10$ dB are listed in Table 4.2. $B_{34} = 10$ dB implies that the second minima in the stopband is greater than the first one by 10 dB, and $B_{34} = -10$ dB implies the reverse.

For the case $B_{34} = 0$ and 10 dB, the frequency response is exhibited in Figure 4.5.

4.5 Asymmetric Class of Filters

The characteristic polynomials for the asymmetric class of filters are discussed in Section 3.10. For the sake of simplicity, we first deal with the minimum phase filters for which the zeros of

$P(s)$ are confined to the imaginary axis. For such filters, the polynomial $K(s)$ is expressed as

$$K(s) = \frac{(s+ja_1)(s+ja_2)\cdots(s+ja_n)}{(s+jb_1)(s+jb_2)\cdots(s+jb_m)} \tag{4.34}$$

where $|a_i| < 1$, $|b_i| > 1$, $m \le n$, and n is the order of the filter network. Since the ripple factor ϵ does not alter the critical frequencies, it is assumed to be unity in the formulation of $K(s)$. The gradient of the transmission coefficient with respect to the reflection or transmission zeros is computed by equation (4.13). As an illustration, the gradient with respect to b_i is given by

$$|K(s)| = \frac{|s+ja_1||s+ja_2|\cdots}{|s+jb_1||s+jb_2|\cdots} \tag{4.35}$$

$$\ln|K(s)| = -\ln|s+jb_i| + \text{other terms}$$

$$\therefore \frac{1}{|K(s)|}\frac{\partial}{\partial b_i}|K(s)| = -\frac{1}{|s+jb_i|}\cdot\frac{\partial}{\partial b_i}|s+jb_i| + \cdots \tag{4.36}$$

Since $|s+jb_i| = (s^2+b_i^2)^{1/2}$, the gradient is computed to be

$$\frac{\partial}{\partial b_i}|K(s)| = -|K(s)|\frac{b_i}{s^2+b_i^2} \tag{4.37}$$

In a similar way

$$\frac{\partial}{\partial a_i}|K(s)| = |K(s)|\frac{a_i}{s^2+a_i^2} \tag{4.38}$$

By substituting the values of

$$\frac{\partial}{\partial a_i}|K(s)| \quad \text{or} \quad \frac{\partial}{\partial b_i}|K(s)|$$

in equation (4.13), we obtain

$$\frac{\partial|t|^2}{\partial a_i} = -2\frac{a_i}{(s^2+a_i^2)}\frac{|K(s)|^2}{\left[1+|K(s)|^2\right]^2} \tag{4.39}$$

$$\frac{\partial|t|^2}{\partial b_i} = 2\frac{b_i}{(s^2+b_i^2)}\cdot\frac{|K(s)|^2}{\left\{1+|K(s)|^2\right\}^2} \tag{4.40}$$

These relationships can be expressed in a more compact form in terms of transmission and reflection loss in dB, described by equations (4.14) and (4.15):

$$\frac{\partial T}{\partial a_i} = 20\log e|\rho|^2\frac{a_i}{s^2+a_i^2} \tag{4.41}$$

$$\frac{\partial R}{\partial a_i} = -20\log e|t|^2\frac{a_i}{s^2+a_i^2} \tag{4.42}$$

In a similar manner

$$\frac{\partial T}{\partial b_i} = -20\log e\cdot|\rho|^2\frac{b_i}{s^2+b_i^2} \tag{4.43}$$

$$\frac{\partial R}{\partial b_i} = 20\log e|t|^2\frac{b_i}{s^2+b_i^2} \tag{4.44}$$

The gradients differ by a factor of 2 for the asymmetric case as compared with the symmetric case. This is expected since we are dealing with a single zero as opposed to a pair of zeros for the symmetric case.

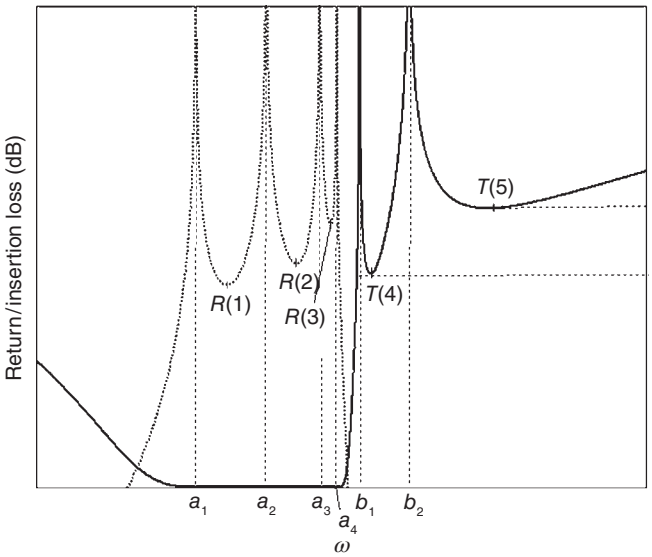


Figure 4.6 Lowpass prototype response of a four-pole asymmetric filter with two attenuation poles in the upper half of the stopband.

4.5.1 Asymmetric Filters with Chebyshev Passband

Let us consider a four-pole filter with four finite attenuation zeros in the passband and two attenuation poles in the upper half of the stopband, as shown in Figure 4.6.

For a Chebyshev passband and arbitrary stopband, the objective function is derived by

$$U = |R(1) - R(2)| + |R(2) - R(3)| + |T(4) - T(5) - B_{45}| \tag{4.45}$$

Here, B_{45} is an arbitrary constant and represents the difference between the attenuation minima at $T(4)$ and $T(5)$. For an equiripple stopband, $B_{45} = 0$. For filters requiring an arbitrary value of B_{45}, it is desirable to generate the unified design charts (Section 3.7) as a function of the first attenuation minima, often a critical specification requirement. This can be accomplished by choosing the attenuation pole frequency closest to the passband as a parameter to generate the unified design charts for an arbitrary value of B_{34}. This has the merit of reducing the number of variables for optimization by one. Sometimes, it is beneficial to fix both attenuation pole frequencies, in which case the objective function is simply given by

$$U = |R(1) - R(2)| + |R(2) - R(3)| \tag{4.46}$$

Such a function is much simpler to optimize. It provides a good indication for the starting values to optimize the objective function for specific values of B_{45}. For this example, b_1 and b_2 are chosen to be 1.25 and 1.5, respectively. Following the optimization procedure, the remaining critical frequencies for an equiripple Chebyshev passband are determined as −0.8466, 0.0255, 0.7270, and 0.9757, with a cutoff frequency of ±1.0. Figure 4.7 describes the filter response.

4.5.2 Asymmetrical Filters with Arbitrary Response

To demonstrate the flexibility of optimization approach, we consider a four-pole asymmetric filter with a double zero at the origin and two attenuation poles in the upper half of the stopband,

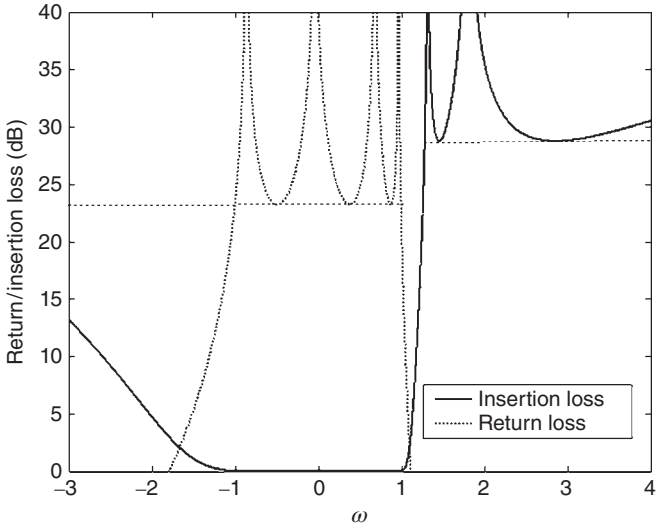


Figure 4.7 Lowpass prototype response of a fourth-order asymmetric filter with two attenuation poles in the upper half of the frequency band and optimized for an equiripple response.

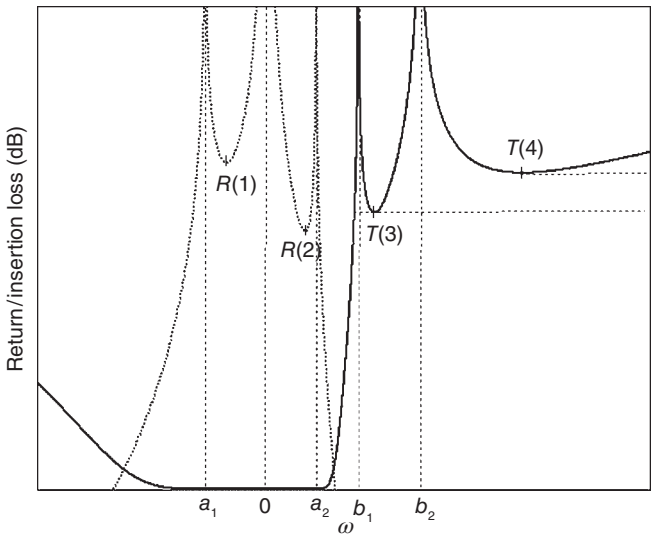


Figure 4.8 Lowpass prototype response of a fourth-order asymmetric filter with a double zero at the origin and two attenuation poles in the upper half of the frequency band.

as described in Figure 4.8. The characteristic polynomial for such a filter is

$$K(s) = \frac{s^2(s + ja_1)(s + ja_2)}{(s + jb_1)(s + jb_2)} \tag{4.47}$$

The objective function for this class of filter can be described as

$$U = |R(1) - R(2) + A_{12}| + |T(3) - T(4) + B_{34}| \tag{4.48}$$

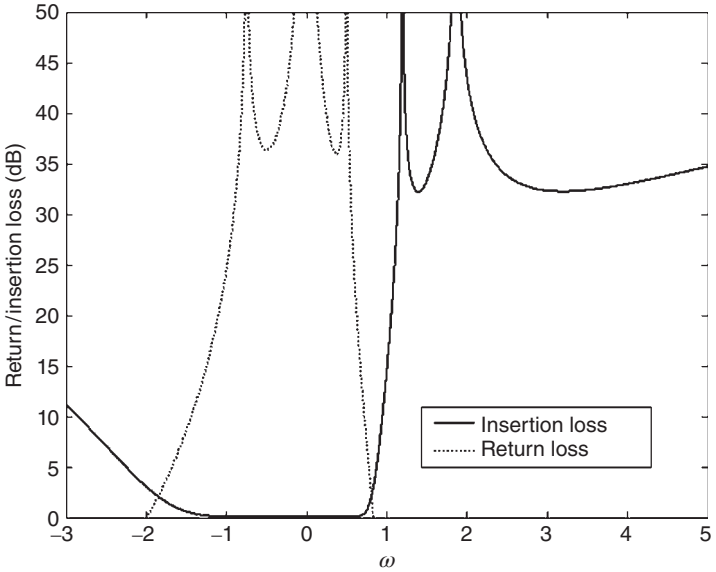


Figure 4.9 Lowpass prototype response of a four-pole asymmetric filter with a double zero at the origin, two attenuation poles in the upper half of the frequency band, and optimized for the equiripple response.

Table 4.3 Critical frequencies and polynomial characteristics for optimized four-pole asymmetric prototype filter with two equiripple peaks in passband and two transmission zeros in upper stopband.

a_i	b_i	Roots of $E(s)$	Frequency maxima in passband	Attenuation minimum in stopband
−0.745	1.2	$-0.8881 - 1.6837j$	—	—
0	1.865	$-1.4003 - 0.1067j$	−0.4963	1.3944
0	∞	$-0.6811 + 0.7139j$	0.3764	3.2005
0.5	∞	$-0.1689 + 0.8315j$	—	—

where A_{12} and B_{34} are constants that define the departure of the passband and stopband from an equiripple response, $A_{12} = 0$ represents an equiripple passband, and $B_{34} = 0$ represents an equiripple stopband. For this example, we optimize the objective function by choosing A_{12} and B_{34} to be zero. The optimized frequency response is plotted in Figure 4.9; the optimized characteristic polynomials and attenuation minima and maxima are listed in Table 4.3. An interesting aspect of such an asymmetric filter is that its cutoff frequency is no longer symmetric with respect to the zero frequency of the prototype filter. The band center must be calculated by computing the geometric mean of the equiripple passband. As a result, the filter response can be renormalized to the new band center frequency and passband bandwidth.

4.6 Linear Phase Filters

The technique described for the computer-aided optimization of characteristic polynomials for minimum phase filters is equally applicable for linear phase filters, both symmetric and

asymmetric. Linear phase filters, when the zeros of $P(s)$ are located symmetrically on the real axis, are readily derived from equation (4.17). The equations for the gradients are modified as

$$\frac{\partial T}{\partial b_i} = -40 \log e \cdot |\rho|^2 \cdot \frac{b_i}{(s^2 \pm b_i^2)} \tag{4.49}$$

and

$$\frac{\partial R}{\partial b_i} = 40 \log e \cdot |t|^2 \cdot \frac{b_i}{(s^2 \pm b_i^2)} \tag{4.50}$$

where the negative sign is applicable to the case of a linear phase filter with zero location at $\pm b_1$ along the real axis. The complex zeros of the transfer function are located symmetrically about the imaginary axis, as noted in Section 3.10. Let us define a complex zero by

$$\begin{aligned} Q(s) &= (s + \sigma_1 \pm j\omega_1)(s - \sigma_1 \pm j\omega_1) \\ &= s^2 \pm j2s\omega_1 - (\sigma_1^2 + \omega_1^2) \end{aligned} \tag{4.51}$$

A single pair of complex zeros gives rise to an asymmetric response, whereas a complex quad yields a symmetrical response. Following equation (4.51), we obtain

$$|Q(s)|^2 = \left\{(s^2 + \sigma_1^2 + \omega_1^2)^2 - 4s^2\sigma_1^2\right\} \tag{4.52}$$

$$\frac{\partial |Q(s)|}{\partial \omega_1} = -\frac{2\omega_1(3s^2 - \sigma_1^2 - \omega_1^2)}{|Q(s)|} \tag{4.53}$$

$$\frac{\partial |Q(s)|}{\partial \sigma_1} = -\frac{2\sigma_1(s^2 - \sigma_1^2 - \omega_1^2)}{|Q(s)|} \tag{4.54}$$

In terms of the characteristic function $K(s)$, we have

$$\frac{\partial}{\partial \sigma_1}|K(s)| = -\frac{|K(s)|}{|Q(s)|} \cdot \frac{\partial |Q(s)|}{\partial \sigma_1}$$

$$\frac{\partial}{\partial \omega_1}|K(s)| = -\frac{|K(s)|}{|Q(s)|} \cdot \frac{\partial |Q(s)|}{\partial \omega_1}$$

Finally, the gradient in terms of transmission and reflection loss in dB is

$$\frac{\partial T}{\partial \sigma_1} = -40 \log e \cdot |\rho|^2 \sigma_1 \frac{(s^2 - \sigma_1^2 - \omega_1^2)}{|Q(s)|^2} \tag{4.55}$$

$$\frac{\partial T}{\partial \omega_1} = -40 \log e |t|^2 \omega_1 \frac{(3s^2 - \sigma_1^2 - \omega_1^2)}{|Q(s)|^2} \tag{4.56}$$

Since linear phase filters require the simultaneous optimization of the amplitude and phase (or group delay), the problem becomes quite complex. For practical reasons, it is far more expeditious to assume specific values of a complex zero as well as $j\omega$-axis zeros [i.e., zeros of $P(s)$] and then optimize for an equiripple passband. By selecting zeros of $P(s)$ judiciously, and then tinkering with these values, one can optimize the characteristic polynomials for both amplitude and phase to achieve the desired response.

4.7 Critical Frequencies for Selected Filter Functions

With the optimization techniques presented in the preceding section, the critical frequencies are computed for an eight-pole lowpass prototype filter function, not amenable to analytical

techniques. Characteristic factor F is used as a parameter in these computations, and it includes the following filter functions:

1) An eight-pole filter with two transmission zeros having a 10 dB difference in the attenuation minima and an equiripple passband.
2) An eight-pole filter with a double zero at the origin, two transmission zeros, and an equiripple passband and stopband.

The computed critical frequencies are included in Appendix 4A. These data add to the range of unified design charts and also provide guidelines for developing software to synthesize characteristic polynomials for filters with arbitrary amplitude and phase response.

Summary

The transfer function of a lossless lumped-element lowpass prototype filter is completely characterized in terms of its poles and zeros. Such a characterization is well suited for computer-aided design techniques. The key lies in making sure that the optimization procedure is highly efficient. This is accomplished by determining the gradients of the objective function analytically and linking it directly to the desired amplitude response shape. In this chapter, we address this issue by deriving analytical gradients with respect to the critical frequencies for the most general lowpass prototype filter. This includes minimum phase and nonminimum phase filters exhibiting symmetric or asymmetric frequency response. The efficiency of this procedure is demonstrated by including examples of the classical Chebyshev and elliptic function filters as special cases of the design procedure. The technique is completely general and can be readily adapted to synthesize the characteristic polynomials for filters with arbitrary responses. To demonstrate the flexibility of this method, examples of some unconventional filters are included.

References

1 Kowalik, J. and Osbourne, M.R. (1968) *Methods for Unconstrained Optimization Problems*, Elsevier, New York.
2 Bandler, J.W. (1969) Optimization methods for computer-aided design. *IEEE Transactions on Microwave Theory and Techniques*, **17** (8), 533–552.
3 Kudsia, C. M. and Swamy, M. N. S. (1980) Computer-aided optimization of microwave filter networks for space applications. IEEE MTT-S International Microwave Symposium Digest, May 28–30, Washington, DC.

Appendix 4A

Table 4A.1 Critical frequencies for an eight-pole filter having two transmission zeros with 10 dB difference in attenuation minima and an equiripple passband.

$$K(s) = \frac{(s^2 + a_1^2)(s^2 + a_2^2)(s^2 + a_3^2)(s^2 + a_4^2)}{(s^2 + b_1^2)(s^2 + b_2^2)}$$

a_1	a_2	a_3	a_4	b_1	b_2	b_s	F a)
0.2702	0.7049	0.9255	0.9938	1.0460	1.2070	1.0384	40
0.2618	0.6885	0.9159	0.9925	1.0664	1.2514	1.0568	45
0.2539	0.6732	0.9067	0.9913	1.0914	1.3037	1.0802	50
0.2403	0.6464	0.8903	0.9891	1.1601	1.4321	1.1437	60
0.2295	0.6250	0.8768	0.9873	1.2520	1.5911	1.2306	70
0.2210	0.6081	0.8660	0.9857	1.3719	1.7876	1.3449	80
0.2147	0.5954	0.8577	0.9846	1.5192	2.0186	1.4860	90
0.2099	0.5857	0.8514	0.9837	1.6972	2.2901	1.6572	100

a) characteristic factor in dB, $F = (A_1 + R_3) - (A_3 + R_1) \approx (R_1 + A_3)$.

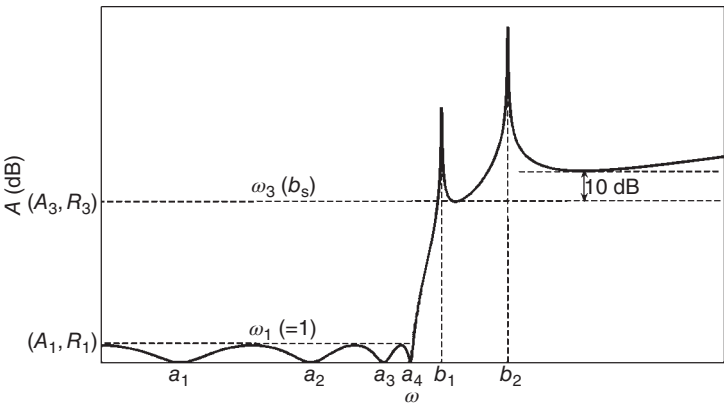
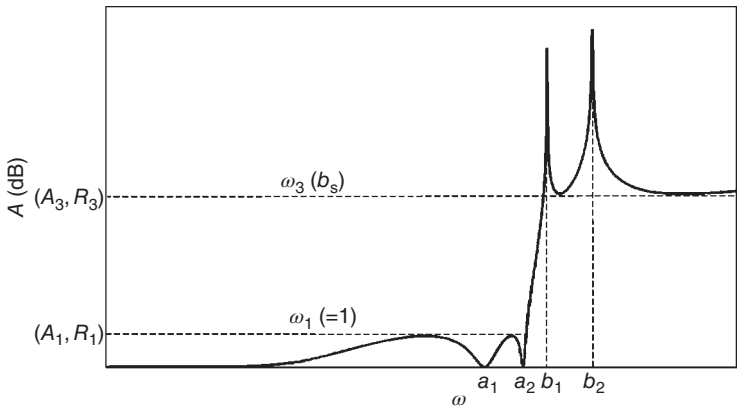


Table 4A.2 Critical frequencies for an eight-pole filter with a double zero at the origin, two transmission zeros, and equiripple passband and stopband.

$$K(s) = \frac{s^4(s^2 + a_1^2)(s^2 + a_2^2)}{(s^2 + b_1^2)(s^2 + b_2^2)}$$

a_1	a_2	b_1	b_2	F
0.9012	0.9926	1.0501	1.1566	40
0.8871	0.9909	1.0734	1.1990	45
0.8742	0.9893	1.1023	1.2473	50
0.8628	0.9878	1.1364	1.3011	55
0.8522	0.9863	1.1774	1.3626	60
0.8429	0.9851	1.2249	1.4315	65
0.8347	0.9839	1.2788	1.5074	70
0.8213	0.9820	1.4057	1.6807	80
0.8109	0.9804	1.5640	1.8911	90
0.8029	0.9793	1.7571	2.1430	100



5

Analysis of Multiport Microwave Networks

Any filter or multiplexer circuit can be viewed as a combination of multiport subnetworks. Figure 5.1 illustrates a circuit that consists of a five-port network connected to several two-port networks. The overall circuit in this case is a four-channel multiplexer. The five-port network is the combining manifold junction. Each filter could be viewed as a cascade of two-port or three-port subnetworks. In this chapter, we discuss several matrix representations of multiport microwave networks. Also, we present various techniques to analyze linear passive microwave circuits that are formed by connecting any number of multiport networks. As an example, we apply these techniques, at the end of this chapter, to demonstrate step by step the evaluation of the overall scattering matrix of a three-channel multiplexer.

The most commonly used matrices to describe a network are the $[Z]$, $[Y]$, $[ABCD]$, $[S]$, and $[T]$ matrices [1–3]. The $[Z]$, $[Y]$, and $[ABCD]$ matrices relate the voltage and current at the various ports by describing the network at the lumped-element level. The scattering matrix $[S]$ and the transmission matrix $[T]$ relate the incident and reflected normalized voltages at the various ports. While, at microwave frequencies, the voltage and currents cannot be measured, the $[Z]$, $[Y]$, and $[ABCD]$ matrices are often used to provide a physical insight into the equivalent circuit of the microwave network. In RF design, the most commonly quoted parameters are the S-parameters. They represent parameters that can be quantified and measured. In addition, the S matrix can be easily extended to represent the behavior of the network for the dominant mode, as well as the higher-order modes. It should be noted that the $[Z]$, $[Y]$, $[ABCD]$, $[S]$, and $[T]$ matrices are interchangeable and can fully describe the characteristics of any microwave network. Familiarity with the concepts of these matrices is crucial for understanding the materials presented in this book. These matrices are also essential to comprehend the various techniques presented in this chapter for the analysis of microwave filter networks.

5.1 Matrix Representation of Two-Port Networks

5.1.1 Impedance [*Z*] and Admittance [*Y*] Matrices

Let us consider the two-port network shown in Figure 5.2. The current and the voltages at the two ports are related by $[Z]$ and $[Y]$ matrices as follows:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{5.1}$$

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{5.2}$$

$$[V] = [Z][I] \tag{5.3}$$

$$[I] = [Y][V] \tag{5.4}$$

It is clear that the $[Z]$ and $[Y]$ matrices are related as

$$Z = [Y]^{-1} \tag{5.5}$$

It can be easily shown that Z_{11} is the input impedance at port 1 when port 2 is an open circuit: $I_2 = 0$. Y_{11} is the input admittance at port 1 when port 2 is a short circuit: $V_2 = 0$. Similarly, one can derive the physical meaning of the rest of the parameters from equations (5.1) and (5.2). For microwave networks that are lossless, the elements of the $[Z]$ and $[Y]$ matrices are pure imaginary. For networks that do not have ferrites, plasmas, or active devices, that is, networks that are reciprocal, one can show that $Z_{12} = Z_{21}$ and $Y_{12} = Y_{21}$.

In filter design, the $[Z]$ and $[Y]$ matrices are often used to describe the T or π lumped-element equivalent circuit of the microwave network as shown in Figure 5.3. For example, in dealing with a coupling iris in filters, the T or π equivalent circuit of the iris can easily identify whether the coupling is inductive or capacitive. As demonstrated in Chapter 14, the equivalent circuit can also identify the loading effects of the iris on the adjacent resonators.

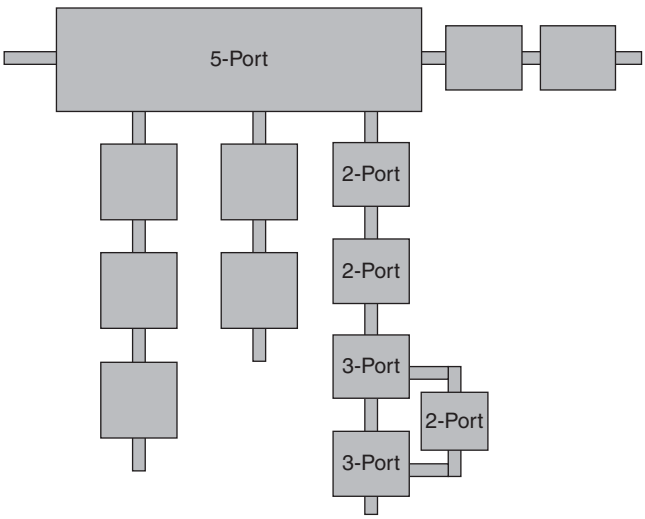


Figure 5.1 An example of a microwave circuit.

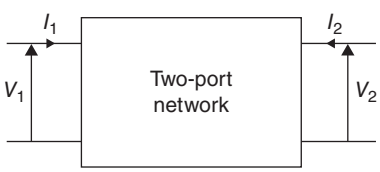


Figure 5.2 A two-port microwave network.

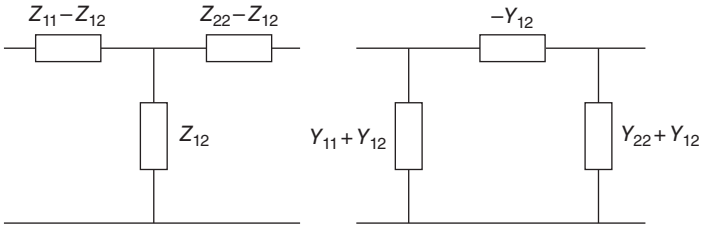


Figure 5.3 Equivalent *T* and π networks of a two-port microwave network.

5.1.2 The [*ABCD*] Matrix

The [*ABCD*] matrix representation of the two-port network depicted in Figure 5.4 is given in equation (5.6). Note that the current in port 2 is defined as $-I_2$; that is, the current flows out of port 2. This definition is adopted to ease the evaluation of the overall [*ABCD*] matrix in dealing with cascaded networks. For symmetrical networks $A = D$, and for reciprocal networks one can show that *ABCD* parameters satisfy equation (5.7). Figure 5.5 shows two networks connected in cascade. In this case, $I_3 = -I_2$, and the overall [*ABCD*] matrix is calculated by simple matrix multiplication of the two [*ABCD*] matrices:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix} \tag{5.6}$$

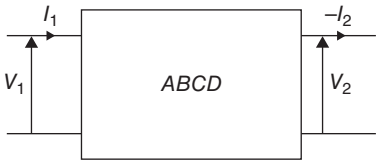
$$AD - CB = 1 \tag{5.7}$$

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix}, \quad \begin{bmatrix} V_3 \\ I_3 \end{bmatrix} = \begin{bmatrix} \overline{A} & \overline{B} \\ \overline{C} & \overline{D} \end{bmatrix} \begin{bmatrix} V_4 \\ -I_4 \end{bmatrix} \tag{5.8}$$

But V_2 and I_2 are related to V_3 and I_3 as

$$\begin{bmatrix} V_2 \\ -I_2 \end{bmatrix} = \begin{bmatrix} V_3 \\ I_3 \end{bmatrix} \tag{5.9}$$

Figure 5.4 The *ABCD* representation of a two-port network.



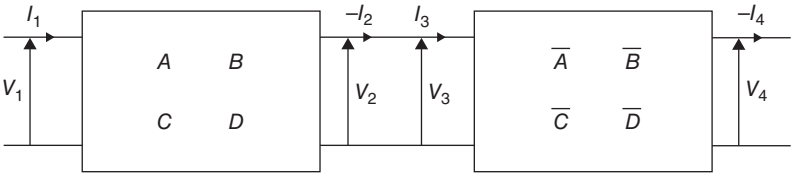


Figure 5.5 Two networks cascaded in series.

Substituting equations (5.8) and (5.9) in equation (5.6), we get

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \overline{A} & \overline{B} \\ \overline{C} & \overline{D} \end{bmatrix} \begin{bmatrix} V_4 \\ -I_4 \end{bmatrix} \tag{5.10}$$

The relation between the [Z] matrix and [$ABCD$] matrix can be obtained by setting $I_2 = 0$ and $V_2 = 0$ in both matrices as shown as follows:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{5.11}$$

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix} \tag{5.12}$$

Setting $I_2 = 0$ in equations (5.11) and (5.12) gives

$$V_1 = Z_{11} I_1, \quad V_2 = Z_{21} I_1 \tag{5.13}$$

$$V_1 = A V_2, \quad I_1 = C V_2 \tag{5.14}$$

C and D can then be related to the elements of the [Z] matrix as follows:

$$C = \frac{I_1}{V_2} = \frac{1}{Z_{21}}, \quad A = \frac{V_1}{V_2} = \frac{Z_{11}}{Z_{21}} \tag{5.15}$$

Similarly, by setting $V_2 = 0$ in both equations, one can easily show that

$$B = \frac{Z_{11}Z_{22} - Z_{12}Z_{21}}{Z_{21}}, \quad D = \frac{Z_{22}}{Z_{21}} \tag{5.16}$$

Table 5.1 depicts the [$ABCD$] matrix of several circuit components that are often used in the design of microwave circuits [3].

5.1.3 The Scattering [*S*] Matrix

S-parameters are important in microwave design because they are easier to measure and work with at high frequencies [4–6]. They are conceptually simple, analytically convenient, and capable of providing a great insight into the transmission and reflection of RF energy in microwave circuits.

S-parameters of the two-port network shown in Figure 5.6 relate the incident and the reflected normalized voltage waves at the two ports as described by equation (5.17). The relationships between the incident and reflected voltages V^+ and V^- and port currents and voltages I and V are explained in details in other publications [1, 4–6]:

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} \tag{5.17}$$

$$V_1^- = S_{11} V_1^+ + S_{12} V_2^+ \tag{5.18}$$

$$V_2^- = S_{21} V_1^+ + S_{22} V_2^+ \tag{5.19}$$

Terminating port 2 with a matched load, that is, $V_2^+ = 0$, results in

$$V_1^- = S_{11} V_1^+, \quad V_2^- = S_{21} V_1^+ \tag{5.20}$$

Table 5.1 [*ABCD*] parameters of the commonly used circuit configurations.

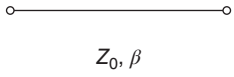
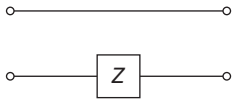
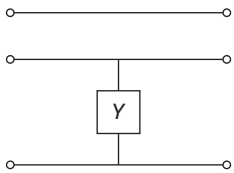
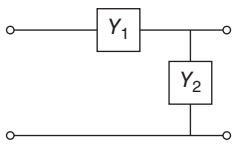
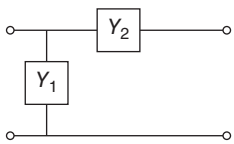
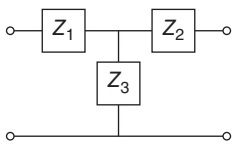
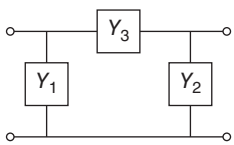
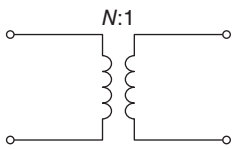
Circuit	*ABCD* parameters	
	$A = \cos \beta l$ $C = jY_0 \sin \beta l$	$B = jZ_0 \sin \beta l$ $D = \cos \beta l$
	$A = 1$ $C = 0$	$B = Z$ $D = 1$
	$A = 1$ $C = Y$	$B = 0$ $D = 1$
	$A = 1 + \frac{Y_2}{Y_1}$ $C = Y_2$	$B = \frac{1}{Y_1}$ $D = 1$
	$A = 1$ $C = Y_1$	$B = \frac{1}{Y_2}$ $D = 1 + \frac{Y_1}{Y_2}$
	$A = 1 + \frac{Z_1}{Z_3}$ $C = \frac{1}{Z_3}$	$B = Z_1 + Z_2 + \frac{Z_1 Z_2}{Z_3}$ $D = 1 + \frac{Z_2}{Z_3}$
	$A = 1 + \frac{Y_2}{Y_3}$ $C = Y_1 + Y_2 + \frac{Y_1 Y_2}{Y_3}$	$B = \frac{1}{Y_3}$ $D = 1 + \frac{Y_1}{Y_3}$
	$A = N$ $C = 0$	$B = 0$ $D = \frac{1}{N}$

Figure 5.6 The *S* matrix of a two-port network.

It is clear from equation (5.20) that the parameter S_{11} is the reflection coefficient at port 1, when port 2 is terminated with a matched load, and the parameter S_{21} is the transmission coefficient from port 1 to port 2. Similarly, one can show that S_{22} is the reflection coefficient at port 2 when port 1 is terminated by a matched load. The *S*-parameters are characteristics of the network itself and are defined assuming that all ports are matched. Having a different termination will change the reflection seen at the ports.

In reciprocal networks, $S_{12} = S_{21}$. For lossless networks, the S matrix satisfies a condition known as the *unitary condition*. This condition can be easily derived by equating the average power dissipated inside the network to zero. In lossless networks the energy is conserved. Thus the energy incident to the network is equal to the energy propagating away from it, which can be written as

$$[V^+]_t[V^+]^* = [V^-]_t[V^-]^* \tag{5.21}$$

where $[V^-]$ and $[V^+]$ are column vectors, $[\,]^*$ denotes complex conjugate, and $[\,]_t$ denotes transpose:

$$[V^+]_t[V^+]^* = \{[S][V^+]\}_t \cdot \{[S][V^+]\}^* \tag{5.22}$$

$$[V^+]_t[V^+]^* = [V^+]_t[S]_t[S]^*[V^+]^* \tag{5.23}$$

$$0 = [V^+]_t \cdot \{[U] - [S]_t[S]^*\}[V^+]^* \tag{5.24}$$

Equation (5.24) can hold only if the following equation is satisfied:

$$[S]_t[S]^* = [U] \tag{5.25}$$

This equation, where $[U]$ is the unitary matrix, is known as the *unitary condition*. It states that multiplication of S transpose by S complex conjugate equals the unity matrix U. For a two-port network, it can be written in the following form:

$$\begin{bmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{bmatrix} \begin{bmatrix} S_{11}^* & S_{12}^* \\ S_{21}^* & S_{22}^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.26}$$

$$|S_{11}|^2 + |S_{21}|^2 = 1 \tag{5.27}$$

$$|S_{12}|^2 + |S_{22}|^2 = 1 \tag{5.28}$$

$$S_{11}S_{12}^* + S_{21}S_{22}^* = 0 \tag{5.29}$$

In defining the *S*-parameters by using equation (5.17) [1], it is assumed that the characteristic impedances of all ports are equal. This is the case in the great majority of circuits. However, for cases where the ports have different characteristic impedances, one must use the normalized incident and reflected voltages in defining the scattering matrix. The incident and reflected voltages are normalized as follows:

$$a_1 = \frac{V_1^+}{\sqrt{Z_{01}}}, \quad a_2 = \frac{V_2^+}{\sqrt{Z_{02}}} \tag{5.30}$$

$$b_1 = \frac{V_1^-}{\sqrt{Z_{01}}}, \quad b_2 = \frac{V_2^-}{\sqrt{Z_{02}}} \tag{5.31}$$

Thus, in the general case, the scattering matrix is defined as [1, 4–6]

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{5.32}$$

To illustrate the importance of the normalization when dealing with a network with different characteristic impedances, consider the problem of calculating the scattering matrix of a shunt admittance jB between two ports in two cases. In case 1, the ports have the same characteristic impedance, whereas in case 2 the two ports have different characteristic impedances.

Case 1. Refer to Figure 5.7.

When port 2 is terminated in the characteristic admittance Y_0, the load Y_L seen at the input is then $Y_L = jB + Y_0$. The reflection coefficient Γ seen at the input port equals S_{11} and is then given by

$$S_{11} = \Gamma = \frac{Y_0 - Y_L}{Y_0 + Y_L} \tag{5.33}$$

$$S_{11} = \frac{Y_0 - (Y_0 + jB)}{2Y_0 + jB} = \frac{-jB}{2Y_0 + jB} \tag{5.34}$$

Because of the shunt configuration, voltage V_1 at port 1 equals voltage V_2 at port 2. The port voltages V_1 and V_2 are related to the incident and reflected voltages as follows:

$$V_1 = V_1^+ + V_1^-, \quad V_2 = V_2^- \tag{5.35}$$

Note that V_2^+ is zero since port 2 is terminated in a matched load:

$$V_1^+ + V_1^- = V_2^- \tag{5.36}$$

$$[1 + S_{11}]V_1^+ = V_2^- \tag{5.37}$$

$$S_{21} = \frac{V_2^-}{V_1^+} = [1 + S_{11}] = \frac{2Y_0}{2Y_0 + jB} \tag{5.38}$$

Similarly, by assuming that port 1 is terminated in a matched load, one can show that

$$S_{22} = \frac{-jB}{2Y_0 + jB}, \quad S_{12} = \frac{2Y_0}{2Y_0 + jB} \tag{5.39}$$

Note that $S_{12} = S_{21}$ since the network is reciprocal and $S_{11} = S_{22}$ since the network is symmetric.

Case 2. Refer to Figure 5.8.

First we assume that port 2 is terminated in a matched load, that is, it is terminated in Y_{02}. S_{11} is then given by

$$S_{11} = \frac{V_1^-}{V_1^+} = \frac{Y_{01} - (jB + Y_{02})}{Y_{01} + (jB + Y_{02})} \tag{5.40}$$

$$S_{11} = \frac{Y_{01} - jB - Y_{02}}{Y_{01} + jB + Y_{02}} \tag{5.41}$$

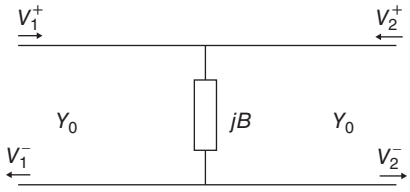


Figure 5.7 A shunt admittance connected between two ports having the same characteristic impedance.

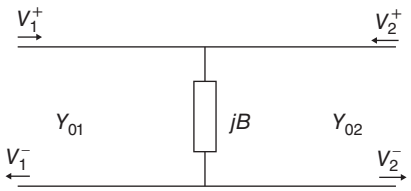


Figure 5.8 A shunt admittance connected between two ports having different characteristic impedances.

Following the same approach as in case 1, the S-parameters are derived by relating the voltages at the two ports, resulting in

$$V_1^+ + V_1^- = V_2^- \Rightarrow [1 + S_{11}] = S_{21} \tag{5.42}$$

$$S_{21} = \frac{V_2^-}{V_1^+} = 1 + \frac{Y_{01} - jB - Y_{02}}{Y_{01} + jB + Y_{02}} = \frac{2Y_{01}}{Y_{01} + jB + Y_{02}} \tag{5.43}$$

Similarly, if we terminate port 1 in a matched load, we can follow the same procedure to derive expressions for S_{22} and S_{12} as follows:

$$S_{22} = \frac{V_2^-}{V_2^+} = \frac{Y_{02} - jB - Y_{01}}{Y_{01} + jB + Y_{02}} \tag{5.44a}$$

$$S_{12} = \frac{V_1^-}{V_2^+} = \frac{2Y_{02}}{Y_{01} + jB + Y_{02}} \tag{5.44b}$$

Note that even though the circuit is reciprocal, the use of nonnormalized incident and reflected voltages leads to a solution that violates the reciprocity condition. However, if normalized voltages are used as shown below, a proper solution is obtained. When port 2 is terminated in a matched load, equation (5.43) is modified to reflect the normalization:

$$S_{21} = \frac{V_2^-/\sqrt{Z_{02}}}{V_1^+/\sqrt{Z_{01}}} = \frac{V_2^-}{V_1^+}\sqrt{\frac{Z_{01}}{Z_{02}}} \tag{5.45}$$

$$S_{21} = \frac{2\sqrt{Y_{01}Y_{02}}}{Y_{01} + jB + Y_{02}} \tag{5.46}$$

Similarly, when port 2 is terminated in a matched load, equation (5.45) is modified to read

$$S_{12} = \frac{V_1^-/\sqrt{Z_{01}}}{V_2^+/\sqrt{Z_{02}}} = \frac{V_1^-}{V_2^+}\sqrt{\frac{Z_{02}}{Z_{01}}} \tag{5.47}$$

$$S_{12} = \frac{2\sqrt{Y_{01}Y_{02}}}{Y_{01} + jB + Y_{02}} \tag{5.48}$$

It is noted that in this case the use of normalized incident and reflected voltage leads to a solution that satisfies the reciprocity condition. The preceding example highlights the importance of adopting normalized voltages when we deal with networks having ports with different characteristic impedances.

The relationship between the scattering matrix $[S]$ and the impedance matrix $[Z]$ can be derived by relating the voltage and current to incident and reflected voltages:

$$[V^-] = [S][V^+] \tag{5.49}$$

$$[V] = [Z][I] \tag{5.50}$$

By using transmission-line theory [1, 3] and assuming unity characteristic impedance for all ports, we get

$$[V] = [V^+] + [V^-] \tag{5.51}$$

$$[I] = ([V^+] - [V^-]) \tag{5.52}$$

$$[V^+] = \frac{1}{2}([V] + [I]) \Rightarrow [V^+] = \frac{1}{2}([Z] + [U])[I] \tag{5.53}$$

$$[V^-] = \frac{1}{2}([V] - [I]) \Rightarrow [V^-] = \frac{1}{2}([Z] - [U])[I] \tag{5.54}$$

$$[V^-] = ([Z] - [U])([Z] + [U])^{-1}[V^+] \tag{5.55}$$

$$[S] = ([Z] - [U])([Z] + [U])^{-1} \tag{5.56}$$

Similarly, the $[S]$ matrix is related to the $[Y]$ matrix as follows:

$$[S] = ([U] - [Y])([U] + [Y])^{-1} \tag{5.57}$$

Conversely, the $[Z]$ and $[Y]$ matrices are written in terms of the $[S]$ matrix as

$$[Z] = ([U] + [S])([U] - [S])^{-1} \tag{5.58}$$

$$[Y] = ([U] - [S])([U] + [S])^{-1} \tag{5.59}$$

These formulas are valid where all ports have the same characteristic impedance values. In the general case, where the ports have different impedances, the following relations are applied [2]:

$$[S] = [\sqrt{Y_0}]([Z] - [Z_0])([Z] + [Z_0])^{-1}[\sqrt{Z_0}] \tag{5.60}$$

$$[S] = [\sqrt{Z_0}]([Y_0] - [Y])([Y_0] + [Y])^{-1}[\sqrt{Y_0}] \tag{5.61}$$

$$[Z] = [\sqrt{Z_0}]([U] + [S])([U] - [S])^{-1}[\sqrt{Z_0}] \tag{5.62}$$

$$[Y] = [\sqrt{Y_0}]([U] - [S])([U] + [S])^{-1}[\sqrt{Y_0}] \tag{5.63}$$

where

$$[Z_0] = \begin{bmatrix} Z_{01} & 0 \\ 0 & Z_{02} \end{bmatrix} \tag{5.64}$$

$$[Y_0] = \begin{bmatrix} Y_{01} & 0 \\ 0 & Y_{02} \end{bmatrix} \tag{5.65}$$

$$[\sqrt{Z_0}] = \begin{bmatrix} \sqrt{Z_{01}} & 0 \\ 0 & \sqrt{Z_{02}} \end{bmatrix} \tag{5.66}$$

$$[\sqrt{Y_0}] = \begin{bmatrix} \sqrt{Y_{01}} & 0 \\ 0 & \sqrt{Y_{02}} \end{bmatrix} \tag{5.67}$$

5.1.4 The Transmission Matrix [*T*]

Another matrix that relates the incident and reflected voltages at the ports is known as the *transmission matrix* $[T]$ (see Figure 5.9). It is defined by equation (5.68). The $[T]$ matrix is similar to the $[ABCD]$ matrix in the sense that it facilitates calculation of the overall transmission matrix of two cascaded networks by a simple matrix multiplication. For reciprocal networks

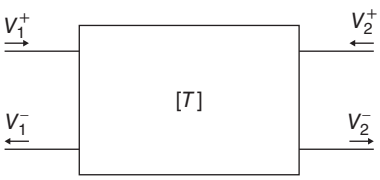


Figure 5.9 *T* matrix representation of a two-port network.

the determinant of the $[T]$ matrix equals unity as given in equation (5.69):

$$\begin{bmatrix} V_1^+ \\ V_1^- \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} V_2^- \\ V_2^+ \end{bmatrix} \tag{5.68}$$

$$T_{11}T_{22} - T_{12}T_{21} = 1 \tag{5.69}$$

Considering the two cascaded networks shown in Figure 5.10, the overall $[T]$ matrix of the cascaded network is given as

$$\begin{bmatrix} V_1^+ \\ V_1^- \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} V_2^- \\ V_2^+ \end{bmatrix}, \quad \begin{bmatrix} V_3^+ \\ V_3^- \end{bmatrix} = \begin{bmatrix} \overline{T}_{11} & \overline{T}_{12} \\ \overline{T}_{21} & \overline{T}_{22} \end{bmatrix} \begin{bmatrix} V_4^- \\ V_4^+ \end{bmatrix} \tag{5.70}$$

$$\begin{bmatrix} V_1^+ \\ V_1^- \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} \overline{T}_{11} & \overline{T}_{12} \\ \overline{T}_{21} & \overline{T}_{22} \end{bmatrix} \begin{bmatrix} V_4^- \\ V_4^+ \end{bmatrix} \tag{5.71}$$

The relation between $[S]$ and $[T]$ matrices is obtained by setting V_2^+ and V_2^- in both matrices such that

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} \tag{5.72a}$$

$$\begin{bmatrix} V_1^+ \\ V_1^- \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} V_2^- \\ V_2^+ \end{bmatrix} \tag{5.72b}$$

Letting $V_2^+ = 0$, from the $[S]$ matrix, we have

$$V_1^- = S_{11}V_1^+, \quad V_2^- = S_{21}V_1^+ \tag{5.73}$$

From the $[T]$ matrix we get

$$V_1^+ = T_{11}V_2^-, \quad V_1^- = T_{21}V_2^- \tag{5.74}$$

Using equations (5.73) and (5.74), we get

$$T_{11} = \frac{1}{S_{21}} \tag{5.75a}$$

$$T_{21} = \frac{S_{11}}{S_{21}} \tag{5.75b}$$

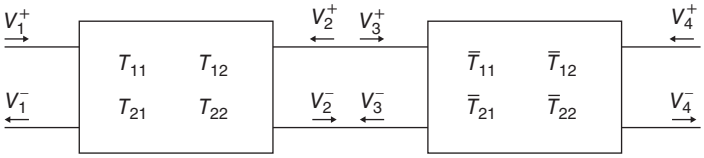


Figure 5.10 Two networks connected in cascade.

Letting $V_2^- = 0$ in equations (5.72a) and (5.72b), we then have

$$V_1^- = S_{11}V_1^+ + S_{12}V_2^+ \tag{5.76}$$

$$0 = S_{21}V_1^+ + S_{22}V_2^+ \tag{5.77}$$

$$V_1^+ = T_{12}V_2^+ \tag{5.78}$$

$$V_1^- = T_{22}V_2^+ \tag{5.79}$$

From equations (5.77) and (5.78), we get

$$T_{12} = -\frac{S_{22}}{S_{21}} \tag{5.80}$$

Substituting (5.80) in (5.78) and then using equation, (5.82) gives

$$V_1^- = -\frac{S_{11}S_{22}}{S_{21}}V_2^+ + S_{12}V_2^+ \tag{5.81}$$

$$T_{22} = \frac{V_1^-}{V_2^+} = S_{12} - \frac{S_{11}S_{22}}{S_{21}} \tag{5.82}$$

Similarly, one can show that

$$S_{11} = \frac{T_{21}}{T_{11}} \tag{5.83}$$

$$S_{12} = T_{22} - \frac{T_{21}T_{12}}{T_{11}} \tag{5.84}$$

$$S_{21} = \frac{1}{T_{11}} \tag{5.85}$$

$$S_{22} = -\frac{T_{12}}{T_{11}} \tag{5.86}$$

Table 5.2 summarizes the relationships between the $[Z]$, $[Y]$, $[ABCD]$, and $[S]$ matrices [3].

5.1.5 Analysis of Two-Port Networks

Consider the two-port network shown in Figure 5.11. If port 2 is terminated by a load, the network is reduced to a one-port network. The input impedance or the reflection coefficient at this port can be calculated from knowledge of the scattering matrix of the two-port network and the reflection coefficient of the terminating load:

$$V_1^- = S_{11}V_1^+ + S_{12}V_2^+ \tag{5.87}$$

$$V_2^- = S_{21}V_1^+ + S_{22}V_2^+ \tag{5.88}$$

The reflection coefficient seen at the output port is given by

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{V_2^+}{V_2^-} \tag{5.89}$$

Substituting equation (5.89) in equation (5.99) gives V_2^+ in terms of V_1^+:

$$V_2^+ = \Gamma V_2^- = \Gamma S_{21}V_1^+ + \Gamma S_{22}V_2^+ \tag{5.90}$$

$$(1 - \Gamma S_{22})V_2^+ = \Gamma S_{21}V_1^+ \tag{5.91}$$

$$V_2^+ = \frac{\Gamma S_{21}}{1 - \Gamma S_{22}}V_1^+ \tag{5.92}$$

$$V_1^- = S_{11}V_1^+ + \frac{S_{12}\Gamma S_{21}}{1 - \Gamma S_{22}}V_1^+ \tag{5.93}$$

Table 5.2 Relationship between [*S*], [*Z*], [*Y*], and [*ABCD*] matrices.

S		***Z***	***Y***	***ABCD***
S_{11}	S_{11}	$\frac{(Z_{11}-Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}}{(Z_{11}+Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}}$	$-\frac{(Y_{11}-Y_0)(Y_{22}+Y_0)-Y_{12}Y_{21}}{(Y_{11}+Y_0)(Y_{22}+Y_0)-Y_{12}Y_{21}}$	$\frac{A+B/Z_0-CZ_0-D}{A+B/Z_0+CZ_0+D}$
S_{12}	S_{12}	$\frac{2Z_{12}Z_0}{(Z_{11}+Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}}$	$-\frac{2Y_{12}Y_0}{(Y_{11}+Y_0)(Y_{22}+Y_0)-Y_{12}Y_{21}}$	$\frac{2(AD-BC)}{A+B/Z_0+CZ_0+D}$
S_{21}	S_{21}	$\frac{2Z_{21}Z_0}{(Z_{11}+Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}}$	$-\frac{2Y_{21}Y_0}{(Y_{11}+Y_0)(Y_{22}+Y_0)-Y_{12}Y_{21}}$	$\frac{2}{A+B/Z_0+CZ_0+D}$
S_{22}	S_{22}	$\frac{(Z_{11}+Z_0)(Z_{22}-Z_0)-Z_{12}Z_{21}}{(Z_{11}+Z_0)(Z_{22}+Z_0)-Z_{12}Z_{21}}$	$-\frac{(Y_{11}+Y_0)(Y_{22}-Y_0)-Y_{12}Y_{21}}{(Y_{11}+Y_0)(Y_{22}+Y_0)-Y_{12}Y_{21}}$	$\frac{-A+B/Z_0-CZ_0+D}{A+B/Z_0+CZ_0+D}$
Z_{11}	$Z_0\frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$	Z_{11}	$\frac{Y_{22}}{Y_{11}Y_{22}-Y_{12}Y_{21}}$	$\frac{A}{C}$
Z_{12}	$Z_0\frac{2S_{12}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$	Z_{12}	$\frac{-Y_{12}}{Y_{11}Y_{22}-Y_{12}Y_{21}}$	$\frac{AD-BC}{C}$
Z_{21}	$Z_0\frac{2S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$	Z_{21}	$\frac{-Y_{21}}{Y_{11}Y_{22}-Y_{12}Y_{21}}$	$\frac{1}{C}$
Z_{22}	$Z_0\frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$	Z_{22}	$\frac{Y_{11}}{Y_{11}Y_{22}-Y_{12}Y_{21}}$	$\frac{D}{C}$
Y_{11}	$Y_0\frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}$	$\frac{Z_{22}}{Z_{11}Z_{22}-Z_{12}Z_{21}}$	Y_{11}	$\frac{D}{B}$
Y_{12}	$Y_0\frac{-2S_{12}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}$	$\frac{-Z_{12}}{Z_{11}Z_{22}-Z_{12}Z_{21}}$	Y_{12}	$\frac{BC-AD}{B}$
Y_{21}	$Y_0\frac{-2S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}$	$\frac{-Z_{21}}{Z_{11}Z_{22}-Z_{12}Z_{21}}$	Y_{21}	$\frac{-1}{B}$
Y_{22}	$Y_0\frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}$	$\frac{Z_{11}}{Z_{11}Z_{22}-Z_{12}Z_{21}}$	Y_{22}	$\frac{A}{B}$

A	$\frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{2S_{21}}$	$\frac{Z_{11}}{Z_{21}}$	$-\frac{Y_{22}}{Y_{21}}$	A
B	$Z_0\frac{(1+S_{11})(1+S_{22})-S_{12}S_{21}}{2S_{21}}$	$\frac{Z_{11}Z_{22}-Z_{12}Z_{21}}{Z_{21}}$	$-\frac{1}{Y_{21}}$	B
C	$\frac{1}{Z_0}\frac{(1-S_{11})(1-S_{22})-S_{12}S_{21}}{2S_{21}}$	$\frac{1}{Z_{21}}$	$-\frac{Y_{11}Y_{22}-Y_{12}Y_{21}}{Y_{21}}$	C
D	$\frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{2S_{21}}$	$\frac{Z_{22}}{Z_{21}}$	$-\frac{Y_{11}}{Y_{21}}$	D

$T_{11}=1/S_{21}, T_{12}=-S_{22}/S_{21}, T_{21}=S_{11}/S_{21}, T_{22}=(S_{12}S_{21}-S_{11}S_{22})/S_{21}$

$S_{11}=T_{21}/T_{11}, S_{12}=(T_{11}T_{22}-T_{12}T_{21})/T_{11}, S_{21}=1/T_{11}, S_{22}=-T_{12}/T_{11}$

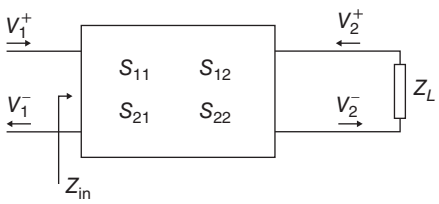


Figure 5.11 A two-port network terminated in load Z_L.

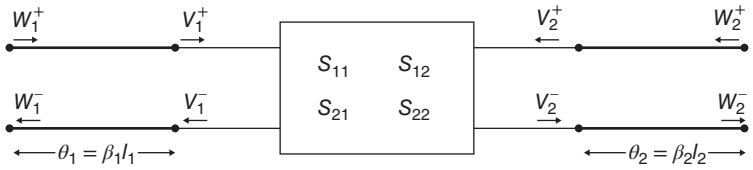


Figure 5.12 A two-port network with extended reference planes.

The reflection coefficient and input impedance at port 1 are then expressed as

$$\Gamma_{in} = \frac{V_1^-}{V_1^+} \tag{5.94}$$

$$\Gamma_{in} = S_{11} + \frac{S_{12}\Gamma S_{21}}{1 - \Gamma S_{22}} \tag{5.95}$$

$$Z_{in} = \frac{1 + \Gamma_{in}}{1 - \Gamma_{in}} \tag{5.96}$$

The scattering matrix of the two-port network with the added lengths of transmission lines (TL) as shown in Figure 5.12 (i.e., extending the reference planes) can be calculated as follows:

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = [S] \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} \tag{5.97}$$

We also have

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} e^{j\theta_1} & 0 \\ 0 & e^{j\theta_2} \end{bmatrix} \begin{bmatrix} W_1^- \\ W_2^- \end{bmatrix} \tag{5.98}$$

$$\begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} = \begin{bmatrix} e^{-j\theta_1} & 0 \\ 0 & e^{-j\theta_2} \end{bmatrix} \begin{bmatrix} W_1^+ \\ W_2^+ \end{bmatrix} \tag{5.99}$$

Substituting equations (5.98) and (5.99) in equation (5.97) gives

$$\begin{bmatrix} W_1^- \\ W_2^- \end{bmatrix} = \begin{bmatrix} e^{-j\theta_1} & 0 \\ 0 & e^{-j\theta_2} \end{bmatrix} [S] \begin{bmatrix} e^{-j\theta_1} & 0 \\ 0 & e^{-j\theta_2} \end{bmatrix} \begin{bmatrix} W_1^+ \\ W_2^+ \end{bmatrix} \tag{5.100}$$

The resultant $[S]$ matrix of the network shown in Figure 5.13 is then given by

$$\text{Resultant matrix} = \begin{bmatrix} e^{-j\theta_1} & 0 \\ 0 & e^{-j\theta_2} \end{bmatrix} [S] \begin{bmatrix} e^{-j\theta_1} & 0 \\ 0 & e^{-j\theta_2} \end{bmatrix} \tag{5.101}$$

5.2 Cascade of Two Networks

Different methods are used to deal with cascaded networks. Let us consider first the case when using the S-parameters directly. The overall scattering matrix of the cascaded networks $[S^{\text{cascade}}]$

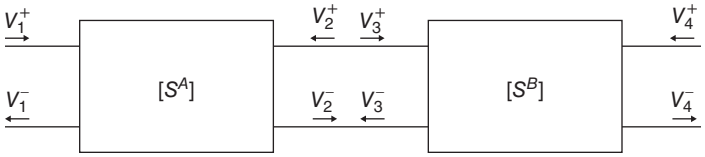


Figure 5.13 Two networks connected in cascade.

can be obtained from the scattering matrices of the two networks, $[S^A]$ and $[S^B]$, as follows:

$$V_1^- = S_{11}^A V_1^+ + S_{12}^A V_2^+ \tag{5.102}$$

$$V_2^- = S_{21}^A V_1^+ + S_{22}^A V_2^+ \tag{5.103}$$

$$V_3^- = S_{11}^B V_3^+ + S_{12}^B V_4^+ \tag{5.104}$$

$$V_4^- = S_{21}^B V_3^+ + S_{22}^B V_4^+ \tag{5.105}$$

In view of Figure 5.13, we also have

$$V_2^+ = V_3^- \tag{5.106}$$

$$V_2^- = V_3^+ \tag{5.107}$$

Solving equations (5.102)–(5.107) provides V_1^- and V_4^- in terms of V_1^+ and V_4^+ as

$$\begin{bmatrix} V_1^- \\ V_4^- \end{bmatrix} = [S^{\text{cascade}}] \begin{bmatrix} V_1^+ \\ V_4^+ \end{bmatrix} \tag{5.108}$$

where

$$S_{11}^{\text{cascade}} = S_{11}^A + \frac{S_{12}^A S_{11}^B S_{21}^A}{1 - S_{22}^A S_{11}^B} \tag{5.109}$$

$$S_{12}^{\text{cascade}} = \frac{S_{12}^A S_{12}^B}{1 - S_{22}^A S_{11}^B} \tag{5.110}$$

$$S_{21}^{\text{cascade}} = \frac{S_{21}^A S_{21}^B}{1 - S_{22}^A S_{11}^B} \tag{5.111}$$

$$S_{22}^{\text{cascade}} = S_{22}^B + \frac{S_{21}^B S_{22}^A S_{12}^B}{1 - S_{22}^A S_{11}^B} \tag{5.112}$$

In analyzing symmetric networks, the problem can be simplified by making use of symmetry. Consider the symmetric network shown in Figure 5.14. The network can be divided into two identical halves. The problem reduces to the analysis of one half of the circuit with an electric wall termination (i.e., a short circuit) and magnetic wall termination (i.e., an open circuit).

By evaluating the reflection coefficients S_e and S_m of the reduced circuits, the scattering matrix of the overall circuit can be written as

$$S_{11} = \frac{1}{2}[S_m + S_e] = S_{22} \tag{5.113}$$

$$S_{12} = \frac{1}{2}[S_m - S_e] = S_{21} \tag{5.114}$$

Equations (5.113) and (5.114) are derived by considering Figure 5.15. The network is decomposed into the superposition of two networks, one with odd-mode excitation, $V_1^+ = V/2$,

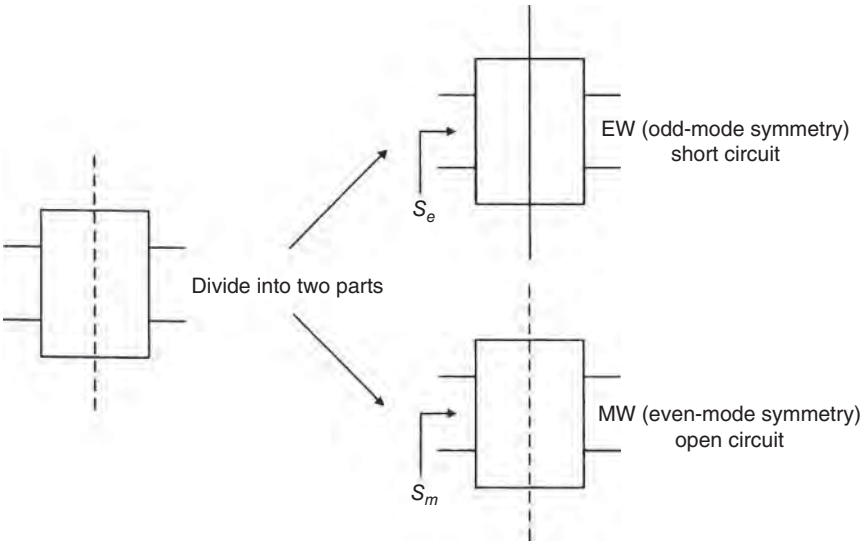


Figure 5.14 A two-port symmetric network divided into two identical halves (EW, MW = electric, magnetic walls).

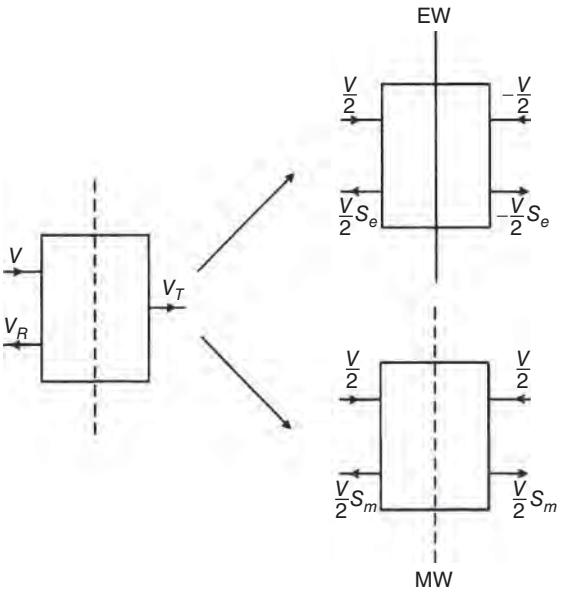


Figure 5.15 Applying superposition principle to a two-port symmetric network.

$V_2^+ = -V/2$, and another with even-mode excitation, $V_1^+ = V/2$, $V_2^+ = V/2$. These two excitations will provide the electric wall and magnetic wall, respectively, at the plane of symmetry. Note also that the superposition of the two networks produces the original excitation, $V_1^+ = V$, $V_2^+ = 0$. Addition of the reflected signals and the transmitted signals from both excitations provides the reflected and transmitted signals from which S_{11} and S_{21} can be calculated:

$$V_R = \frac{V}{2}S_e + \frac{V}{2}S_m = \frac{V}{2}(S_e + S_m) \tag{5.115}$$

$$V_T = \frac{V}{2}(S_m - S_e) \tag{5.116}$$

$$S_{11} = \frac{V_R}{V} = \frac{1}{2}[S_e + S_m] \tag{5.117}$$

$$S_{21} = \frac{V_T}{V} = \frac{1}{2}[S_m - S_e] \tag{5.118}$$

Example 5.1
Consider the cascaded four networks shown in Figure 5.16, and calculate the overall scattering matrix of the network assuming that the scattering matrices of networks *SX* and *SY* are given by

$$SX = \begin{pmatrix} \frac{1}{3} + j\frac{2}{3} & j\frac{2}{3} \\ j\frac{2}{3} & \frac{1}{3} - j\frac{2}{3} \end{pmatrix}, \quad SY = \begin{pmatrix} \frac{1}{3} - j\frac{2}{3} & j\frac{2}{3} \\ j\frac{2}{3} & \frac{1}{3} + j\frac{2}{3} \end{pmatrix}$$

Solution
The *S* matrix of a lossless interconnecting *TL* is written as follows:

$$S = \begin{pmatrix} 0 & e^{-j\theta} \\ e^{-j\theta} & 0 \end{pmatrix} = \begin{pmatrix} 0 & e^{-j\beta\ell} \\ e^{-j\beta\ell} & 0 \end{pmatrix}$$

For the 0.3λ line, $\theta = (2\pi/\lambda) \cdot (0.3\lambda) = 0.6\pi$, while for the 0.4λ line,

$$\theta = \frac{2\pi}{\lambda} \cdot (0.4\lambda) = 0.8\pi$$

$$S_{0.3\lambda} = \begin{pmatrix} 0 & e^{-j(0.6\pi)} \\ e^{-j(0.6\pi)} & 0 \end{pmatrix}, \quad S_{0.4\lambda} = \begin{pmatrix} 0 & e^{-j(0.8\pi)} \\ e^{-j(0.8\pi)} & 0 \end{pmatrix}$$

The overall network can be analyzed by a variety of techniques; we will solve this example using four different methods.

Method 1. Using the [ABCD] matrix approach, the solution can be summarized as follows:

1) Convert each *S* matrix into its *ABCD* equivalent matrix.
2) Multiply all *ABCD* matrices to obtain the overall *ABCD* matrix.
3) Convert the *ABCD* matrix into an *S* matrix.

Step 1. Convert each *S* matrix into its *ABCD* equivalent matrix. In view of Table 5.2, the conversion can be as follows:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} \frac{(1+S_{11})(1-S_{22}) + S_{12}S_{21}}{2S_{21}} & \frac{(1+S_{11})(1+S_{22}) - S_{12}S_{21}}{2S_{21}} \\ \frac{(1-S_{11})(1-S_{22}) - S_{12}S_{21}}{2S_{21}} & \frac{(1-S_{11})(1+S_{22}) + S_{12}S_{21}}{2S_{21}} \end{pmatrix}$$

Thus

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_X = \begin{pmatrix} 1 & -j2 \\ -j & -1 \end{pmatrix} \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix}_Y = \begin{pmatrix} -1 & -j2 \\ -j & 1 \end{pmatrix}$$

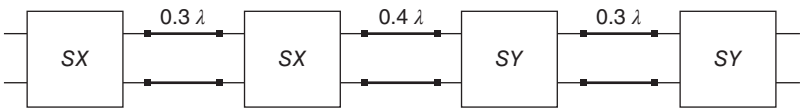


Figure 5.16 Four two-port networks connected in cascade.

and

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{0.3\lambda} = \begin{pmatrix} -0.309 & j0.951 \\ j0.951 & -0.309 \end{pmatrix}$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{0.4\lambda} = \begin{pmatrix} -0.809 & j0.588 \\ j0.588 & -0.809 \end{pmatrix}$$

Before we proceed, we should stop here and check for the *ABCD* matrices and apply the simple checks for reciprocity ($AD - BC = 1$). This can be a good guide for the accuracy of the calculations.

Step 2. Multiply all *ABCD* matrices to get the overall *ABCD* matrix:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{\text{overall}} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}_{X} \times \begin{pmatrix} A & B \\ C & D \end{pmatrix}_{0.3\lambda} \times \begin{pmatrix} A & B \\ C & D \end{pmatrix}_{X} \times \begin{pmatrix} A & B \\ C & D \end{pmatrix}_{0.4\lambda} \times \begin{pmatrix} A & B \\ C & D \end{pmatrix}_{Y} \times \begin{pmatrix} A & B \\ C & D \end{pmatrix}_{0.3\lambda} \times \begin{pmatrix} A & B \\ C & D \end{pmatrix}_{Y}$$

$$= \begin{pmatrix} 10.248 & j16.911 \\ -j6.151 & 10.248 \end{pmatrix}$$

Again, before proceeding further, check for reciprocity.

Step 3. Using Table 5.2, the *ABCD* matrix is converted into an *S* matrix as follows:

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} \dfrac{A+B-C-D}{A+B+C+D} & \dfrac{2(AD-BC)}{A+B+C+D} \\ \dfrac{2}{A+B+C+D} & \dfrac{-A+B-C+D}{A+B+C+D} \end{pmatrix}$$

Again note that the relation above holds when the characteristic impedances of both the input and the output lines are equal. Hence

$$(S)_{\text{overall}} = \begin{pmatrix} 0.996e^{j1.087} & 0.086e^{-j0.483} \\ 0.086e^{-j0.483} & 0.996e^{j1.087} \end{pmatrix}$$

The resultant *S* matrix exhibits the following:

a) $S_{12} = S_{21}$, or the matrix is in the form of $[S] = [S]^T$, which means that the circuit is reciprocal. This should be expected since each of the blocks forming the network was reciprocal.
b) $[S]_t[S]^* = [U]$ (the unitary condition). This is expected since each of the network blocks is lossless.

Method 2. Converting each *S* matrix to its *T* equivalent matrix. This can be of much interest if the modeling technique leads directly to the transmission matrix [7]. The solution can be summarized as follows:

1) Convert each *S* matrix into its *T* equivalent matrix.
2) Multiply all *T* matrices to get the overall *T* matrix.
3) Convert the *T* matrix into *S* matrix.

Step 1. Convert each *S* matrix into its *T* equivalent matrix.

Generally for normalized characteristic impedance at both the input and the output ports, the conversion can proceed as follows:

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{S_{21}} & \dfrac{-S_{22}}{S_{21}} \\ \dfrac{S_{11}}{S_{21}} & \dfrac{S_{12}S_{21} - S_{11}S_{22}}{S_{21}} \end{pmatrix}$$

Thus

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}_X = \begin{pmatrix} -j1.5 & 1+j0.5 \\ 1-j0.5 & j1.5 \end{pmatrix}$$

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}_Y = \begin{pmatrix} -j1.5 & -1+j0.5 \\ -1-j0.5 & j1.5 \end{pmatrix}$$

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}_{0.3\lambda} = \begin{pmatrix} e^{j0.6\pi} & 0 \\ 0 & e^{-j0.6\pi} \end{pmatrix}$$

$$\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}_{0.4\lambda} = \begin{pmatrix} e^{j0.8\pi} & 0 \\ 0 & e^{-j0.8\pi} \end{pmatrix}$$

Before we proceed, it is worthwhile to check the resultant T matrices for reciprocity (determinant of the matrix $T = 1$). This can be a good guide for the accuracy of our calculations.

Step 2. Multiply all T matrices to get the overall T matrix:

$$(T)_{\text{overall}} = (T)_X \times (T)_{0.3\lambda} \times (T)_X \times (T)_{0.4\lambda} \times (T)_Y \times (T)_{0.3\lambda} \times (T)_Y$$

$$= \begin{pmatrix} 11.577e^{j0.483499} & -j11.534 \\ j11.534 & 11.577e^{-j0.483499} \end{pmatrix}$$

Note that the determinant of this matrix is one.

Step 3. Convert the T matrix into an S matrix. This can be done using

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} \dfrac{T_{21}}{T_{11}} & \dfrac{T_{11}T_{22} - T_{12}T_{21}}{T_{11}} \\ \dfrac{1}{T_{11}} & \dfrac{-T_{12}}{T_{11}} \end{pmatrix}$$

$$(S)_{\text{overall}} = \begin{pmatrix} 0.996e^{j1.087} & 0.086e^{-j0.483} \\ 0.086e^{-j0.483} & 0.996e^{j1.087} \end{pmatrix}$$

It is clear that solutions of methods 1 and 2 are identical.

Method 3. Cascading directly using the scattering matrices as given in Section 5.3. From equations (5.109) and (5.112), we have

$$S_{11}^{\text{cascade}} = S_{11}^A + \frac{S_{12}^A S_{11}^B S_{21}^A}{1 - S_{22}^A S_{11}^B}$$

$$S_{21}^{\text{cascade}} = \frac{S_{12}^A S_{12}^B}{1 - S_{22}^A S_{11}^B}$$

$$S_{21}^{\text{cascade}} = \frac{S_{21}^A S_{21}^B}{1 - S_{22}^A S_{11}^B}$$

$$S_{22}^{\text{cascade}} = S_{22}^B + \frac{S_{21}^B S_{22}^A S_{12}^B}{1 - S_{22}^A S_{11}^B}$$

Applying this concept successively to the network given in Figure 5.16, starting first with the scattering matrix SY to represent output network B, and the scattering matrix of the 0.3λ section representing the input network A as shown in Figure 5.17, we get

$$S^{\text{cas1}} = \begin{pmatrix} 0.1222 + j0.7353 & 0.6340 - j0.2060 \\ 0.6340 - j0.2060 & 0.3333 + j0.6667 \end{pmatrix}$$

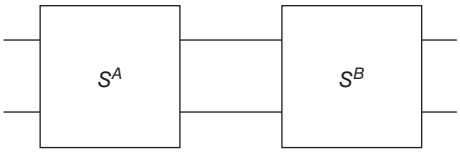


Figure 5.17 Two-port networks *A* and *B* connected in series.

Now S^{cas1} represents output network *B*, while network *SY* represents input network *A*:

$$S^{\text{cas2}} = \begin{pmatrix} 0.3460 - j0.8893 & 0.0277 + j0.2979 \\ 0.0277 + j0.2979 & 0.5038 + j0.8104 \end{pmatrix}$$

S^{cas2} represents the output network *B*, while the scattering matrix of the 0.4λ line represents network *A*:

$$S^{\text{cas3}} = \begin{pmatrix} 0.9527 + j0.0543 & 0.1527 - j0.2572 \\ 0.1527 - j0.2572 & 0.5038 + j0.8104 \end{pmatrix}$$

With S^{cas3} representing output network *B* and scattering matrix *SX* representing network *A*, we get

$$S^{\text{cas4}} = \begin{pmatrix} -0.0281 + j0.9744 & 0.2175 - j0.0501 \\ 0.2175 - j0.0501 & 0.4517 + j0.8638 \end{pmatrix}$$

Similarly, we get S^{cas5} as

$$S^{\text{cas5}} = \begin{pmatrix} -0.5500 - j0.8048 & -0.1149 - j0.1914 \\ -0.1149 - j0.1914 & 0.4517 + j0.8638 \end{pmatrix}$$

As a final step, with S^{cas5} representing output network *B* and scattering matrix *SX* representing network *A*, we get the overall *S* matrix as follows:

$$(S)_{\text{Overall}} = \begin{pmatrix} 0.4631 + j0.8821 & 0.0765 - j0.0402 \\ 0.0765 - j0.0402 & 0.4631 + j0.8821 \end{pmatrix}$$

$$= \begin{pmatrix} 0.996e^{j1.087} & 0.086e^{-j0.483} \\ 0.086e^{-j0.483} & 0.996e^{j1.087} \end{pmatrix}$$

The solution is identical to that given by methods 1 and 2. Method 3 can be easily programmed to deal with the cascade of a large number of two-port networks.

Method 4. In this method, we make use of network symmetry. In view of the scattering matrices *SX* and *SY*, it is clear that the network is symmetric, that is, it has a mirror-image symmetry at a vertical axis through the middle of the 0.4λ section (most Chebyshev filters are symmetric structures). The problem can be simplified by analyzing only half of the network terminated in magnetic and electric walls. It is worthy to note that the use of symmetry can considerably reduce the computation time. One needs first to calculate the scattering matrix of half of the original network, which is a cascade of the network *SX*, a 0.3λ transmission-line section, network *SX*, and a 0.2λ transmission-line section as shown in Figures 5.18 and 5.19. The *S* matrix of the reduced network S^{common} can be calculated using any of the above three methods and is given by

$$S^{\text{common}} = \begin{pmatrix} 0.5038 + j0.8104 & 0.2918 + j0.0657 \\ 0.2918 + j0.0657 & -0.8026 + j0.5160 \end{pmatrix}$$

The system is then reduced to the calculation of the reflection coefficient of the reduced two networks shown in Figures 5.18 and 5.19.

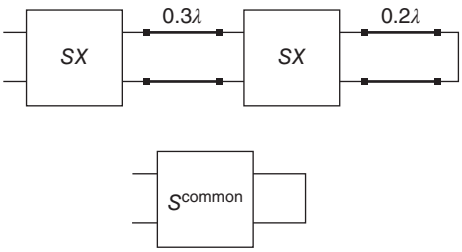


Figure 5.18 Half-network terminated by electric wall.

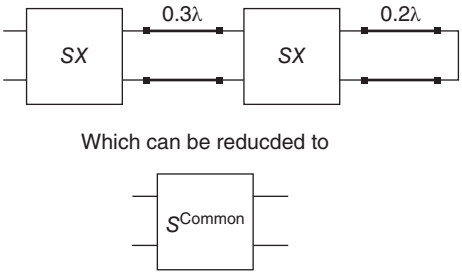


Figure 5.19 Half-network terminated by magnetic wall.

Electric wall case. With the output port of S^{common} terminated in the reflection coefficient Γ_L, in view of equation (5.98), the input reflection coefficient Γ_{in} can be written as follows:

$$\Gamma_{\text{in}} = S_{11}^{\text{common}} + \frac{S_{12}^{\text{common}} S_{21}^{\text{common}} \Gamma_L}{1 - S_{22}^{\text{common}} \Gamma_L}$$

For short circuit $\Gamma_L = -1$, then $S_e = \Gamma_{\text{in}} = 0.3867 + j0.9222$.

Magnetic wall case. For open circuit $\Gamma_L = 1$, then $S_m = \Gamma_{\text{in}} = 0.5396 + j0.8419$.

Hence, for the overall S matrix, we obtain

$$S_{11} = S_{22} = \frac{1}{2}(S_m + S_e) = 0.4631 + j0.8821$$

$$S_{12} = S_{21} = \frac{1}{2}(S_m - S_e) = 0.0765 + j0.0402$$

$$(S)_{\text{overall}} = \begin{pmatrix} 0.4631 + j0.8821 & 0.0765 - j0.0402 \\ 0.0765 - j0.0402 & 0.4631 + j0.8821 \end{pmatrix}$$

$$= \begin{pmatrix} 0.996e^{j1.087} & 0.086e^{-j0.483} \\ 0.086e^{-j0.483} & 0.996e^{j1.087} \end{pmatrix}$$

5.3 Multiport Networks

Consider the multiport microwave network shown in Figure 5.20. It consists of N ports having incident voltages $V_1^+, V_2^+, V_3^+, \ldots, V_N^+$ and reflected voltages $V_1^-, V_2^-, V_3^-, \ldots, V_N^-$. The scattering matrix is computed by

$$\begin{bmatrix} V_1^- \\ V_2^- \\ \cdot \\ \cdot \\ \cdot \\ V_N^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \cdot & \cdot & \cdot & S_{1N} \\ S_{21} & S_{22} & \cdot & \cdot & \cdot & S_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ S_{N1} & \cdot & \cdot & \cdot & \cdot & S_{NN} \end{bmatrix} \cdot \begin{bmatrix} V_1^+ \\ V_2^+ \\ \cdot \\ \cdot \\ \cdot \\ V_N^+ \end{bmatrix} \tag{5.119}$$

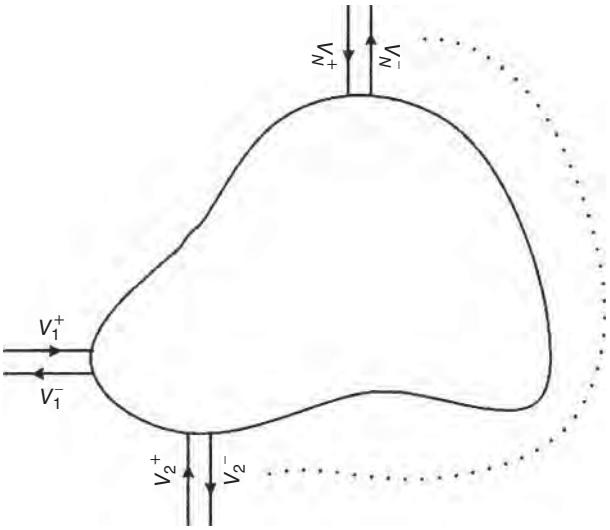


Figure 5.20 A multiport network of *N* ports.

The reciprocity and lossless unitary conditions can be extended to the N-port network. The reciprocity condition is $S_{ij} = S_{ji}$. The unitary condition is $[S]^t[S]^* = [U]$ and can be written as follows:

$$\begin{bmatrix} S_{11} & S_{21} & \cdot & \cdot & \cdot & S_{N1} \\ S_{12} & S_{22} & \cdot & \cdot & \cdot & S_{N2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ S_{1N} & \cdot & \cdot & \cdot & \cdot & S_{NN} \end{bmatrix} \cdot \begin{bmatrix} S_{11}^* & S_{12}^* & \cdot & \cdot & \cdot & S_{1N}^* \\ S_{21}^* & S_{22}^* & \cdot & \cdot & \cdot & S_{2N}^* \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ S_{N1}^* & \cdot & \cdot & \cdot & \cdot & S_{NN}^* \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} \tag{5.120}$$

If the terminal planes are shifted by adding lengths outward by $l_1, l_2, \ldots, l_N$, then the resultant scattering matrix $[S^T]$ is given by

$$\begin{bmatrix} e^{-j\beta_1 l_1} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & e^{-j\beta_2 l_2} & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & e^{-j\beta_N l_N} \end{bmatrix} [S] \begin{bmatrix} e^{-j\beta_1 l_1} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & e^{-j\beta_2 l_2} & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & e^{-j\beta_N l_N} \end{bmatrix} \tag{5.121}$$

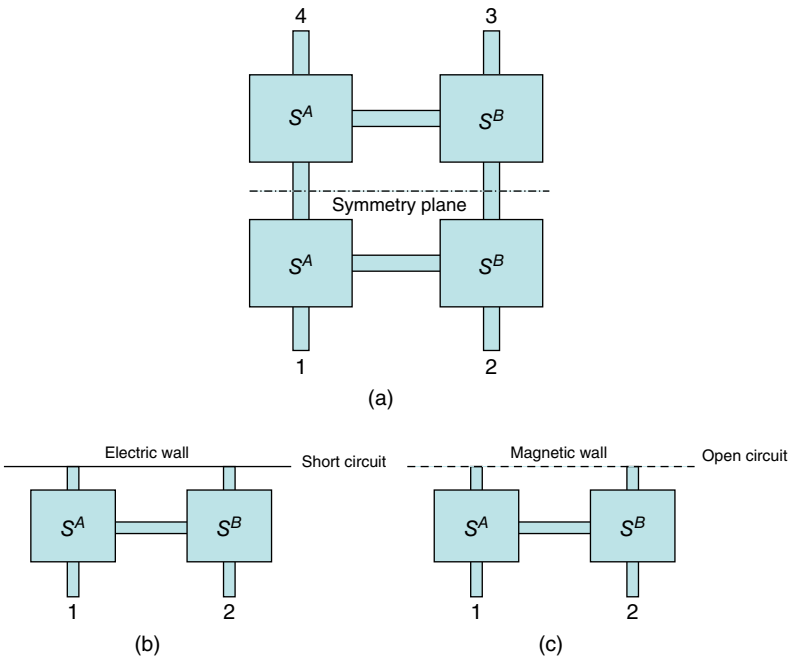


Figure 5.21 A four-port network (a) decomposed into two 2-port networks, (b) with electric wall, and (c) magnetic wall termination.

In general, an N-port network, with M-port terminated by loads, reduces to a network of $(N-M)$ ports, and the resultant matrix has a size of $(N-M)\times(N-M)$.

The same symmetry concept, applied to the two-port network, can be extended to multiport networks. Consider, for example, the four-port network shown in Figure 5.21a. An analysis of this network reduces to analysis of two networks, each having two ports with electric wall (EW) and magnetic wall (MW) terminations as shown in Figures 5.21b and 5.21c.

$$\begin{bmatrix} V_1^- \\ V_2^- \\ V_3^- \\ V_4^- \end{bmatrix} = \begin{bmatrix} [S_{11}^h] & [S_{12}^h] \\ [S_{21}^h] & [S_{22}^h] \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ V_3^+ \\ V_4^+ \end{bmatrix} \tag{5.122}$$

$$[S_{11}^h] = [S_{22}^h] = \frac{1}{2}[[S_m] + [S_e]] \tag{5.123}$$

$$[S_{12}^h] = [S_{21}^h] = \frac{1}{2}[[S_m] - [S_e]] \tag{5.124}$$

where $[S_m]$ and $[S_e]$ are (2×2) scattering matrices of the reduced two-port network with MW and EW terminations, respectively.

5.4 Analysis of Multiport Networks

Consider the generalized network shown in Figure 5.22. It consists of several multiport networks connected together. The overall circuit has P external ports and C internal ports. Some of

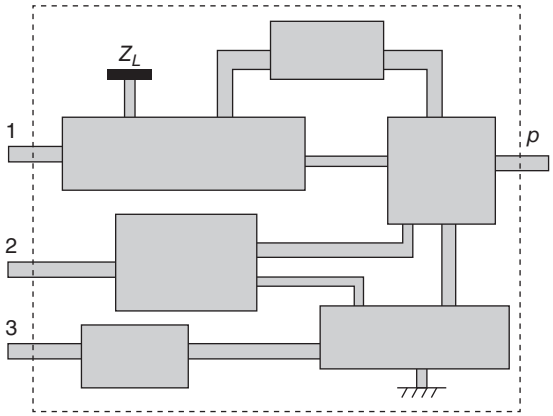


Figure 5.22 An arbitrary microwave network consisting of several multiport networks connected together.

the internal ports are terminated by loads. We will follow the analysis given in [2] to determine the scattering matrix of the resultant network, specifically to determine the scattering matrix that describes the relationships between external ports 1, 2, 3,…, *P*. The main steps are presented below:

Step 1. Identify the external ports *P* and the internal ports *C*. The internal ports include those ports that are terminated in loads and ports that are connected to each other. The external ports are the ports that define the scattering matrix of the resultant network.

Step 2. Rewrite the scattering matrices of the networks in the form of a large matrix of size $(P + C) \times (P + C)$. The incident and reflected voltages are arranged into two groups: one for the external *P* ports (V_P^+, V_P^-) and the other group for the internal C ports (V_C^+, V_C^-). The overall matrix can then be viewed as consisting of four submatrices: S_{PP} of size $(P \times P)$, S_{PC} of size $(P \times C)$, S_{CP} of size $(C \times P)$, and S_{CC} of size $(C \times C)$.

Step 3. Find the connection matrix Γ. This matrix relates the internal ports to each other. That is, it relates V_C^- to V_C^+ by the equation $V_C^- = \Gamma V_C^+$.

To calculate the scattering matrix of the overall network $[S^P]$, we need to relate $[V_P^-]$ to $[V_P^+]$. By dividing the ports into *P* ports and *C* ports, one can write the overall network scattering matrix as follows:

$$\begin{bmatrix} V_P^- \\ V_C^- \end{bmatrix} = \begin{bmatrix} S_{PP} & S_{PC} \\ S_{CP} & S_{CC} \end{bmatrix} \begin{bmatrix} V_P^+ \\ V_C^+ \end{bmatrix} \tag{5.125}$$

The vectors V_C^-, V_C^+ are related to each other by the connectivity matrix Γ as

$$V_C^- = \Gamma V_C^+ \tag{5.126}$$

Substituting equation (5.126) into V_P^- from equation (5.125) gives

$$\Gamma V_C^+ = S_{CP} V_P^+ + S_{CC} V_C^+ \tag{5.127}$$

$$[\Gamma - S_{CC}] V_C^+ = S_{CP} V_P^+ \tag{5.128}$$

$$V_C^+ = [\Gamma - S_{CC}]^{-1} S_{CP} V_P^+ \tag{5.129}$$

$$V_P^- = [S_{PP} V_P^+ + S_{PC}[\Gamma - S_{CC}]^{-1} S_{CP}] V_P^+ \tag{5.130}$$

The resultant *S* matrix $[S^P]$ that relates the external *P* ports can then be written as

$$[S^P] = [S_{PP} + S_{PC}[\Gamma - S_{CC}]^{-1} S_{CP}] \tag{5.131}$$

To illustrate this method, we consider the following two examples.

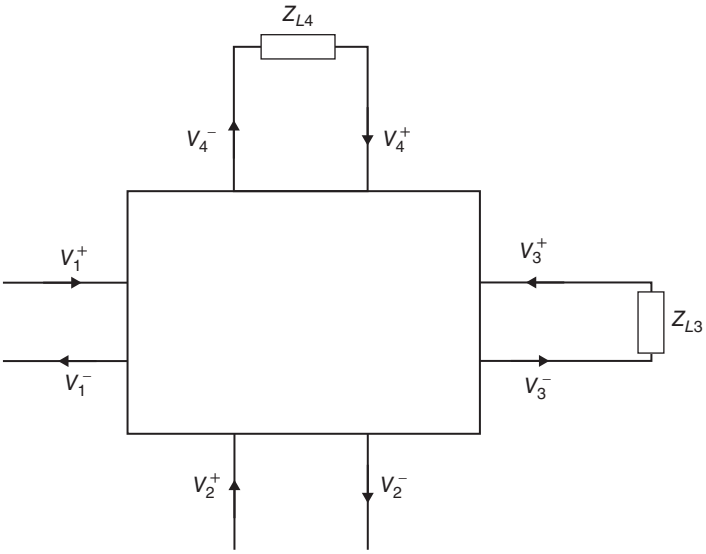


Figure 5.23 Schematic of the four-port network.

Example 5.2 Find the scattering matrix of the network shown in Figure 5.23.

Assume that the scattering matrix of the four-port network is given by

$$\begin{bmatrix} V_1^- \\ V_2^- \\ V_3^- \\ V_4^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ V_3^+ \\ V_4^+ \end{bmatrix}$$

In this problem, ports 1 and 2 are the external ports and are denoted by P, whereas ports 3 and 4 are internal ports and are denoted by C. The scattering matrix of the four-port network is then rewritten as

$$\begin{bmatrix} V_P^- \\ V_C^- \end{bmatrix} = \begin{bmatrix} S_{PP} & S_{PC} \\ S_{CP} & S_{CC} \end{bmatrix} \begin{bmatrix} V_P^+ \\ V_C^+ \end{bmatrix}$$

where

$$S_{PP} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{PC} = \begin{bmatrix} S_{13} & S_{14} \\ S_{23} & S_{24} \end{bmatrix}$$

$$S_{CP} = \begin{bmatrix} S_{31} & S_{32} \\ S_{42} & S_{42} \end{bmatrix}, \quad S_{CC} = \begin{bmatrix} S_{33} & S_{34} \\ S_{43} & S_{44} \end{bmatrix}$$

$$V_P^- = S_{PP} V_P^+ + S_{PC} V_C^+ \tag{5.132}$$

$$V_C^- = S_{CP} V_P^+ + S_{CC} V_C^+ \tag{5.133}$$

In view of Figure 5.23 we also have

$$V_3^+ = \Gamma_3 V_3^-, \quad \Gamma_3 = \frac{Z_{L3} - Z_{03}}{Z_{L3} + Z_{03}} \tag{5.134}$$

$$V_4^+ = \Gamma_4 V_4^-, \quad \Gamma_4 = \frac{Z_{L4} - Z_{04}}{Z_{L4} + Z_{04}} \tag{5.135}$$

$$[V_C^-] = \begin{bmatrix} \frac{1}{\Gamma_3} & 0 \\ 0 & \frac{1}{\Gamma_4} \end{bmatrix} [V_C^+] \tag{5.136}$$

$$V_C^- = \Gamma V_C^+ \tag{5.137}$$

Substituting equation (5.137) into equation (5.133) yields

$$\Gamma_i V_C^+ = S_{CP} V_P^+ + S_{CC} V_C^+ \Rightarrow V_C^+ = [\Gamma - S_{CC}]^{-1} S_{CP} V_P^+ \tag{5.138}$$

By substituting equation (5.138) into equation (5.132), the resultant scattering matrix between the external ports 1 and 2 is given as

$$V_P^- = S_{PP} V_P^+ + S_{PC}[\Gamma - S_{CC}]^{-1} S_{CP} V_P^+ \tag{5.139}$$

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = [S_{PP} + S_{PC}[\Gamma - S_{CC}]^{-1} S_{CP}] \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} \tag{5.140}$$

Example 5.3 Evaluate the *S*-parameters of the three-channel multiplexer shown in Figure 5.24.

The multiplexer consists of a combining junction (manifold) whose *S* matrix is $[S^M]$ and three-channel filter whose *S* matrices are $[S^A]$, $[S^B]$, and $[S^C]$. The objective is to evaluate the reflection coefficient at the manifold input, the reflection coefficients at the channel filter outputs, and the transmission coefficients between the input of the manifold and the channel filters; thus, we need to drive the scattering matrix among ports 1, 8, 9, and 10. These ports are the external ports and will be denoted by *P*. The internal ports in this problem are 2, 3, 4, 5, 6, and 7 and are denoted as ports *C*. Assume that $[S^M]$, $[S^A]$, $[S^B]$, and $[S^C]$ are given by

$$\begin{pmatrix} V_1^- \\ V_2^- \\ V_3^- \\ V_4^- \end{pmatrix} = \begin{pmatrix} S_{11}^M & S_{12}^M & S_{13}^M & S_{14}^M \\ S_{21}^M & S_{22}^M & S_{23}^M & S_{24}^M \\ S_{31}^M & S_{32}^M & S_{33}^M & S_{34}^M \\ S_{41}^M & S_{42}^M & S_{43}^M & S_{44}^M \end{pmatrix} \begin{pmatrix} V_1^+ \\ V_2^+ \\ V_3^+ \\ V_4^+ \end{pmatrix} \tag{5.141}$$

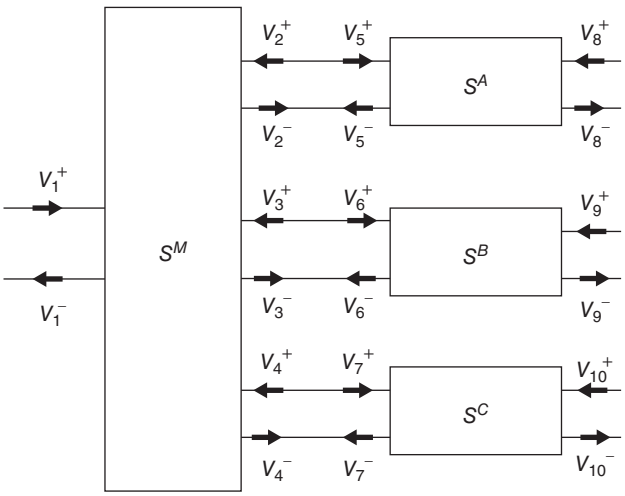


Figure 5.24 A three-channel multiplexer.

$$\begin{pmatrix} V_5^- \\ V_8^- \end{pmatrix} = \begin{pmatrix} S_{11}^A & S_{12}^A \\ S_{21}^A & S_{22}^A \end{pmatrix} \begin{pmatrix} V_5^+ \\ V_8^+ \end{pmatrix}, \quad \begin{pmatrix} V_6^- \\ V_9^- \end{pmatrix} = \begin{pmatrix} S_{11}^B & S_{12}^B \\ S_{21}^B & S_{22}^B \end{pmatrix} \begin{pmatrix} V_6^+ \\ V_9^+ \end{pmatrix} \tag{5.142}$$

$$\begin{pmatrix} V_7^- \\ V_{10}^- \end{pmatrix} = \begin{pmatrix} S_{11}^C & S_{12}^C \\ S_{21}^C & S_{22}^C \end{pmatrix} \begin{pmatrix} V_7^+ \\ V_{10}^+ \end{pmatrix} \tag{5.143}$$

We will rewrite equations (5.141)–(5.143) into one equation as

$$\begin{bmatrix} V_P^- \\ V_C^- \end{bmatrix} = \begin{bmatrix} S_{PP} & S_{PC} \\ S_{CP} & S_{CC} \end{bmatrix} \begin{bmatrix} V_P^+ \\ V_C^+ \end{bmatrix} \tag{5.144}$$

where

$$V_P^- = \begin{pmatrix} V_1^- \\ V_8^- \\ V_9^- \\ V_{10}^- \end{pmatrix}, \quad V_P^+ = \begin{pmatrix} V_1^+ \\ V_8^+ \\ V_9^+ \\ V_{10}^+ \end{pmatrix},$$

$$V_C^- = \begin{pmatrix} V_2^- \\ V_3^- \\ V_4^- \\ V_5^- \\ V_6^- \\ V_7^- \end{pmatrix}, \quad V_C^+ = \begin{pmatrix} V_2^+ \\ V_3^+ \\ V_4^+ \\ V_5^+ \\ V_6^+ \\ V_7^+ \end{pmatrix} \tag{5.145}$$

The matrices S_{PP}, S_{PC}, S_{CP}, and S_{CC} can be written as follows:

$$S_{PP} = \begin{pmatrix} S_{11}^M & 0 & 0 & 0 \\ 0 & S_{22}^A & 0 & 0 \\ 0 & 0 & S_{22}^B & 0 \\ 0 & 0 & 0 & S_{22}^C \end{pmatrix} \tag{5.146}$$

$$S_{PC} = \begin{pmatrix} S_{12}^M & S_{13}^M & S_{14}^M & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{21}^A & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{21}^B & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{21}^C \end{pmatrix} \tag{5.147}$$

$$S_{CP} = \begin{pmatrix} S_{21}^M & 0 & 0 & 0 \\ S_{31}^M & 0 & 0 & 0 \\ S_{41}^M & 0 & 0 & 0 \\ 0 & S_{12}^A & 0 & 0 \\ 0 & 0 & S_{12}^B & 0 \\ 0 & 0 & 0 & S_{12}^C \end{pmatrix} \tag{5.148}$$

$$S_{CC} = \begin{pmatrix} S_{22}^M & S_{23}^M & S_{24}^M & 0 & 0 & 0 \\ S_{32}^M & S_{33}^M & S_{34}^M & 0 & 0 & 0 \\ S_{42}^M & S_{43}^M & S_{44}^M & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{11}^A & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{11}^B & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{11}^C \end{pmatrix} \tag{5.149}$$

The internal ports are related as

$$V_2^- = V_5^+, \quad V_2^+ = V_5^- \tag{5.150}$$

$$V_3^- = V_6^+, \quad V_3^+ = V_6^- \tag{5.151}$$

$$V_4^- = V_7^+, \quad V_4^+ = V_7^- \tag{5.152}$$

The connection matrix Γ relates the internal ports to V_C^- to V_C^+ as $V_C^- = \Gamma V_C^+$, where Γ is given by

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \tag{5.153}$$

The scattering matrix of the three-channel multiplexer is expressed as

$$S^P = S_{PP} + S_{PC}[\Gamma - S_{CC}]^{-1} S_{CP} \tag{5.154}$$

It is noted that the connection matrix in this case consists of 1s and 0s. If there are TLs connecting the filters to the manifold, these TLs can be included in $[S^A]$, $[S^B]$, and $[S^C]$ by shifting the reference planes. Alternatively, one can take the TLs into consideration in writing the elements of the connection matrix Γ. In this case, the 1s in the connection matrix will be replaced by exponential terms of the form of $e^{\pm j\theta}$.

It should be noted that while we considered the case of only a three-channel multiplexer in Example 5.3, one can easily extend the procedure to formulate a computer program to calculate the scattering matrix of a multiplexer of N channels.

Summary

This chapter provides a review of the basic concepts used in the analysis of multiport microwave networks. These concepts are important for filter designers, since any filter or multiplexer circuit can be divided into smaller two-, three-, or N-port networks connected together. Five matrix representations of microwave networks are described, namely, $[Z]$, $[Y]$, $[ABCD]$, $[S]$, and $[T]$ matrices. These matrices are interchangeable, where the elements of any matrix can be written in terms of those of the other four matrices. A table is given summarizing the relationship between these matrices.

The $[Z]$, $[Y]$, and $[ABCD]$ matrices are useful in understanding and dealing with the lumped-element filter circuit models. The $[Z]$ and $[Y]$ matrices are useful in providing a physical insight of the circuit itself. For example, the measured S-parameters of an iris connected between two waveguides cannot directly identify the nature of the coupling provided by the iris. However, by converting the S-parameters to Z- or Y-parameters, the coupling can be readily identified, whether it is inductive (magnetic) or capacitive (electric) coupling. For filter or multiplexer circuits that have ports with unequal impedances, it is important to take into consideration the proper impedance normalization when switching from $[S]$ to $[Z]$ or $[Y]$ matrices. The concept of using symmetry to simplify the analysis of large symmetrical circuits is also described. This concept is particularly useful when using computation-intensive commercial software tools in designing filter circuits. An example is included to demonstrate

the flexibility in evaluating the overall S-parameters of a cascaded network four different ways by utilizing the equivalent [$ABCD$], [T], and [S] matrices, as well as by invoking the symmetry of the network.

An analysis is presented for evaluating the scattering parameters of microwave circuits formed by connecting several multiport networks together. The analysis is quite general and can be employed to analyze any filter or multiplexer circuit. The chapter concludes with an example showing the evaluation of the scattering parameters of a three-channel multiplexer. The reader can easily generalize the concept to evaluate the scattering matrix of multiplexers with any number of channels.

References

1 Collin, R.E. (1966) *Foundations for Microwave Engineering*, McGraw-Hill, New York.
2 Gupta, K.C. (1981) *Computer-Aided Design of Microwave Circuits*, Artech House, Dedham, MA.
3 Pozar, D. (1998) *Microwave Engineering*, 2nd edn, Wiley, New York.
4 Carlin, H.J. (1956) The scattering matrix in network theory. *IRE Transactions on Circuit Theory*, **3**, 88–96.
5 Youla, D.C. (1961) On scattering matrices normalized to complex port numbers. *Proceedings of IRE*, **49**, 1221.
6 Kurokawa, K. (March 1965) Power waves and the scattering matrix. *IEEE Transactions on Microwave Theory and Techniques*, **13**, 194–202.
7 Mansour, R.R. and MacPhie, R.H. (1986) An improved transmission matrix formulation of cascaded discontinuities and its application to E-plane circuits. *IEEE Transactions on Microwave Theory and Techniques*, **34**, 1490–1498.

6

Synthesis of a General Class of the Chebyshev Filter Function

In this chapter, some important scattering parameter relations that will be used in the polynomial representation of the transfer and reflection characteristics of prototype filtering functions are discussed, particularly the unitary conditions for a two-port network. This is followed by the derivation of the recursion technique that is now generally used to generate the transfer and reflection polynomials for Chebyshev filtering characteristics of the most general kind. Finally, the special cases of predistorted and dual-band filtering functions are discussed.

6.1 Polynomial Forms of the Transfer and Reflection Parameters $S_{21}(S)$ and $S_{11}(S)$ for a Two-port network

For the great majority of the filter circuits, we shall initially be considering two-port networks, that is, a "source port" and a "load port" (Figure 6.1).

For a two-port network, the scattering matrix may be represented by a 2×2 matrix:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{6.1}$$

where b_1 and b_2 are the voltage waves propagating away from ports 1 and 2, respectively, and a_1 and a_2 are the voltage waves incident at ports 1 and 2, respectively.

If the network is passive, lossless, and reciprocal, its 2×2 S-parameter matrix yields two conservation of energy equations:

$$S_{11}(s)S_{11}(s)^* + S_{21}(s)S_{21}(s)^* = 1 \tag{6.2}$$

$$S_{22}(s)S_{22}(s)^* + S_{12}(s)S_{12}(s)^* = 1 \tag{6.3}$$

and one unique orthogonality equation[1]:

$$S_{11}(s)S_{12}(s)^* + S_{21}(s)S_{22}(s)^* = 0 \tag{6.4}$$

where the S-parameters are now assumed to be functions of s $(= j\omega)$, the frequency variable.

1 For an Nth-degree polynomial $Q(s)$ with purely imaginary variable $s = j\omega$ and complex coefficients q_i, $i = 0, 1, 2, \ldots, N$, $Q(s)^*$ is the same as $Q^*(s^*)$ or $Q^*(-s)$, which might be a little more familiar to the reader.
If

$$Q(s) = q_0 + q_1 s + q_2 s^2 + \cdots + q_N s^N$$

then

$$Q(s)^* = Q^*(-s) = q_0^* - q_1^* s + q_2^* s^2 - \cdots + q_N^* s^N \quad (N \text{ even})$$

$$Q(s)^* = Q^*(-s) = q_0^* - q_1^* s + q_2^* s^2 - \cdots - q_N^* s^N \quad (N \text{ odd})$$

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

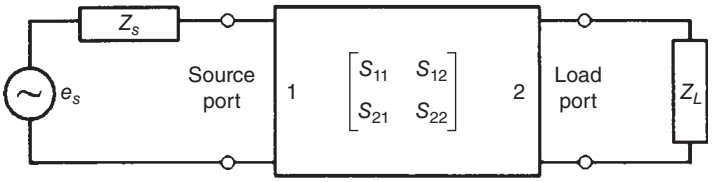


Figure 6.1 Two-port network.

The reflection parameter $S_{11}(s)$ at port 1 of the network is expressed as the ratio of two finite-degree polynomials $E(s)$ and $F(s)$ and the constant ε_R:

$$S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)} \tag{6.5}$$

where $E(s)$ is an Nth-degree polynomial with complex coefficients $e_0, e_1, e_2, \ldots, e_N$, N being the degree of the filter network under consideration. Also, $F(s)$ is an Nth-degree polynomial with complex coefficients $f_0, f_1, f_2, \ldots, f_N$. ε_R allows the normalization of the highest-degree coefficients of $E(s)$ and $F(s)$ to unity (i.e., e_N and $f_N = 1$). Because this is a lossless passive network, $E(s)$ is strictly Hurwitz, that is, all the roots of $E(s)$ (poles of $S_{11}(s)$) are in the left half of the complex plane. These poles need not be symmetric about the real axis. $F(s)$, the numerator polynomial of $S_{11}(s)$, for lowpass and bandpass filters is also of degree N. For bandstop filters the degree of $F(s)$ can be less than N. The roots of $F(s)$ (zeros of $S_{11}(s)$) are the points of zero reflected power ($b_i = 0$), or sometimes known as points of perfect transmission.

By reorganizing equation (6.2) and substituting for $S_{11}(s)$,

$$\begin{aligned} S_{21}(s)S_{21}(s)^* &= 1 - \frac{F(s)F(s)^*/\varepsilon_R^2}{E(s)E(s)^*} \\ &= \frac{P(s)P(s)^*/\varepsilon^2}{E(s)E(s)^*} \end{aligned}$$

Thus, the transfer parameter $S_{21}(s)$ can be expressed as the ratio of two polynomials:

$$S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)} \tag{6.6}$$

where $P(s)P(s)^*/\varepsilon^2 = E(s)E(s)^* - F(s)F(s)^*/\varepsilon_R^2$.

The effect of this conjugating operation ("paraconjugation") is to reflect the complex singularities of $Q(s)$ symmetrically about the imaginary axis (as opposed to reflecting about the real axis for the operation $Q(s) \rightarrow Q^*(s)$ ("conjugation")). If the N complex-plane singularities of $Q(s)$ are s_{0k}, $k = 1, 2, \ldots, N$, then the singularities of $Q(s)^*$ or $Q^*(-s)$ will be $-s_{0k}^*$. When generating the polynomial $Q(s)^*$ or $Q^*(-s)$ from the singularities of $Q(s)$ using the transformed variable $s_{0k} \rightarrow -s_{0k}^*$, the resultant polynomial must be multiplied by $(-1)^N$ to give the correct sign to the new leading coefficient.

$$Q(s)^* = Q^*(-s) = (-1)^N \prod_{k=1}^{N} (s + s_{0k}^*) = \prod_{k=1}^{N} (s - s_{0k})^*$$

Using this notation, the conservation of energy equation (6.2) (the Feldtkeller equation) becomes

$$E(s)E(s)^* = \frac{F(s)F(s)^*}{\varepsilon_R^2} + \frac{P(s)P(s)^*}{\varepsilon^2}$$

This general form of Feldtkeller's equation will deal with fully canonical characteristics (where the number of finite-position transmission zeros $n_{fz} = N$), asymmetric characteristics, and also those with off-axis reflection zeros (e.g., predistorted prototypes; see Section 6.4).

From equations (6.5) and (6.6), it is seen that $S_{11}(s)$ and $S_{21}(s)$ share a common denominator polynomial $E(s)$. The numerator of $S_{21}(s)$ is a polynomial $P(s)/\varepsilon$ whose zeros are the transmission zeros (Tx zeros or "TZ"s) of the filtering function. The degree n_{fz} of the polynomial $P(s)$ corresponds to the number of finite-position Tx zeros that the transfer function incorporates. This also implies that $n_{fz} \leq N$; otherwise, $S_{21}(s)$ exceeds unity as $s \to j\infty$, which is, of course, impossible for a passive network.

A TZ can be realized in two ways. The first case occurs when the degree n_{fz} of $P(s)$ is less than the degree N of the denominator polynomial $E(s)$ and $s \to j\infty$. At $s = j\infty$, $S_{21}(s) = 0$ and this is known as a "TZ at infinity." When there are no finite-position zeros ($n_{fz} = 0$), the filtering function is known as an "all-pole" response. When $0 < n_{fz} < N$, the number of Tx zeros at infinity is $N - n_{fz}$.

The second case occurs when the frequency variable s coincides with an imaginary-axis root of its numerator polynomial $P(s)$, that is, $s = s_{oi}$, where s_{oi} is a purely imaginary root of $P(s)$. The zeros are not necessarily on the imaginary axis, and, if there is one root s_{oi} that is complex, there must be a second root $-s_{0i}^{*}$ to make up a pair having symmetry about the imaginary axis. This ensures that polynomial $P(s)$ has coefficients that alternate between purely real and purely imaginary as the power of s increases. This is a condition that must hold if the filter is to be realized with purely reactive components (see Chapter 7).

Bringing into consideration the orthogonality unitary condition provides an important relation between the phases of the $S_{11}(s)$, $S_{22}(s)$, and $S_{21}(s)$ polynomials and between the zeros of $S_{11}(s)$ and $S_{22}(s)$ in the complex s-plane. If we apply the reciprocity condition $S_{12}(s) = S_{21}(s)$ to the unitary conditions described in equations (6.2)–(6.4) for a two-port network,

$$S_{11}(s)S_{11}(s)^{*} + S_{21}(s)S_{21}(s)^{*} = 1 \tag{6.7a}$$

$$S_{22}(s)S_{22}(s)^{*} + S_{21}(s)S_{21}(s)^{*} = 1 \tag{6.7b}$$

$$S_{11}(s)S_{21}(s)^{*} + S_{21}(s)S_{22}(s)^{*} = 0 \tag{6.7c}$$

If we write these vector quantities in polar coordinates and drop the "(s)" for the moment for presentational simplicity, $S_{11} = |S_{11}| \cdot e^{j\theta_{11}}$, $S_{22} = |S_{22}| \cdot e^{j\theta_{22}}$, and $S_{21} = |S_{21}| \cdot e^{j\theta_{21}}$ [1]. From equations (6.7a) and (6.7b), note that $|S_{11}| = |S_{22}|$, and so, from equation (6.7a),

$$|S_{21}|^2 = 1 - |S_{11}|^2$$

Rewriting equation (6.7c) in polar form,

$$\begin{aligned} |S_{11}|e^{j\theta_{11}} \cdot |S_{21}|e^{-j\theta_{21}} + |S_{21}|e^{j\theta_{21}} \cdot |S_{11}|e^{-j\theta_{22}} &= 0 \\ |S_{11}||S_{21}|(e^{j(\theta_{11}-\theta_{21})} + e^{j(\theta_{21}-\theta_{22})}) &= 0 \end{aligned} \tag{6.8}$$

This equation can be satisfied only if

$$e^{j(\theta_{11}-\theta_{21})} = -e^{j(\theta_{21}-\theta_{22})} \tag{6.9}$$

By replacing the minus sign in equation (6.9) with $e^{j(2k\pm1)\pi}$, where k is an integer,

$$e^{j(\theta_{11}-\theta_{21})} = e^{j((2k\pm1)\pi+\theta_{21}-\theta_{22})}$$

or

$$\theta_{21} - \frac{(\theta_{11} + \theta_{22})}{2} = \frac{\pi}{2}(2k \pm 1) \tag{6.10}$$

Since $S_{11}(s)$, $S_{22}(s)$, and $S_{21}(s)$ are rational polynomials in the variable s and share a common denominator polynomial $E(s)$, their phases may be written as

$$\theta_{21}(s) = \theta_{n21}(s) - \theta_{d}(s),$$

$$\theta_{11}(s) = \theta_{n11}(s) - \theta_d(s),$$
$$\text{and } \theta_{22}(s) = \theta_{n22}(s) - \theta_d(s) \tag{6.11}$$

where $\theta_d(s)$ is the angle of the common denominator polynomial $E(s)$ and $\theta_{n21}(s)$, $\theta_{n11}(s)$, and $\theta_{n22}(s)$ are the angles of the numerator polynomials of $S_{21}(s)$, $S_{11}(s)$, and $S_{22}(s)$, respectively.

By substituting these separated phases into equation (6.10), the $\theta_d(s)$ cancels and the following important relationship is obtained:

$$-\theta_{n21}(s) + \frac{(\theta_{n11}(s) + \theta_{n22}(s))}{2} = \frac{\pi}{2}(2k \pm 1) \tag{6.12}$$

This equation states that at any value of the frequency variable s, the difference between the angle of the $S_{21}(s)$ numerator vector and the average of phases of the $S_{11}(s)$ and $S_{22}(s)$ numerator vectors must be an odd multiple of $\pi/2$ radians, that is, orthogonal. The frequency independence of the right-hand side (RHS) of equation (6.12) allows us to observe two important properties of $\theta_{n21}(s)$, $\theta_{n11}(s)$, and $\theta_{n22}(s)$:

- The numerator polynomial of $S_{21}(s)$ ($P(s)/\varepsilon$) must have zeros that are either on the imaginary axis of the complex s-plane or in mirror-image pairs symmetrically arranged about the imaginary axis. In this way, the value of $\theta_{n21}(s)$ (the angle of $P(s)$) remains at an integral number of $\pi/2$ radians at any position of s, along the imaginary axis from $-j\infty$ to $+j\infty$.
- Similarly, the average of the phases of the $S_{11}(s)$ and $S_{22}(s)$ numerator polynomials, $(\theta_{n11}(s) + \theta_{n22}(s))/2$, must be evaluated to an integral number of $\pi/2$ radians at any value of s between $-j\infty$ and $+j\infty$. This means that the zeros of the numerator polynomial of $S_{11}(s)$ (i.e., the roots of the $F(s)$ polynomial[2]) and the zeros of the numerator polynomial of $S_{22}(s)$ (the roots of $F_{22}(s)$) must either be coincident upon the imaginary axis or arranged in mirror-image pairs about the imaginary axis (paraconjugate).

The definition of these angles is illustrated in Figure 6.3.

The second statement allows us to relate the numerator polynomials of $S_{11}(s)$ and $S_{22}(s)$, namely, $F(s)$ and $F_{22}(s)$, respectively. If their zeros (roots) are coincident upon the imaginary axis or are symmetrically arranged about the imaginary axis, then the ith zero of $F_{22}(s)$, s_{22i}, is related to the corresponding zero of $F(s)$, (s_{11i}), as follows:

$$s_{22i} = -s_{11i}^* \tag{6.13}$$

where $i = 1, 2, \ldots, N$.

By forming the $F_{22}(s)$ polynomial from the s_{22i} zeros,

$$F_{22}(s) = \prod_{i=1}^{N}(s - s_{22i})$$
$$= \prod_{i=1}^{N}(s + s_{11i}^*) \tag{6.14a}$$
$$= (-1)^N \prod_{i=1}^{N}(s - s_{11i})^* \tag{6.14b}$$

Since we already know that the polynomial $F(s) = \prod_{i=1}^{N}(s - s_{11i})$, it becomes clear that

$$F_{22}(s) = (-1)^N F(s)^* \tag{6.15}$$

The $(-1)^N$ term here effectively ensures that $F_{22}(s)$ is a monic polynomial irrespective of whether it is of even or odd degree. Because TZs are always either on the imaginary axis or are in pairs symmetrically positioned about the imaginary axis (Figures 6.2 and 6.3a), $P(s)^* = (-1)^{n_{fz}} P(s)$.

2 Formally, if the numerator polynomial of $S_{22}(s)$ is $F_{22}(s)$, then the numerator polynomial of $S_{11}(s)$ should be called $F_{11}(s)$. However, common usage in this book and elsewhere has meant that $F(s) \equiv F_{11}(s)$.

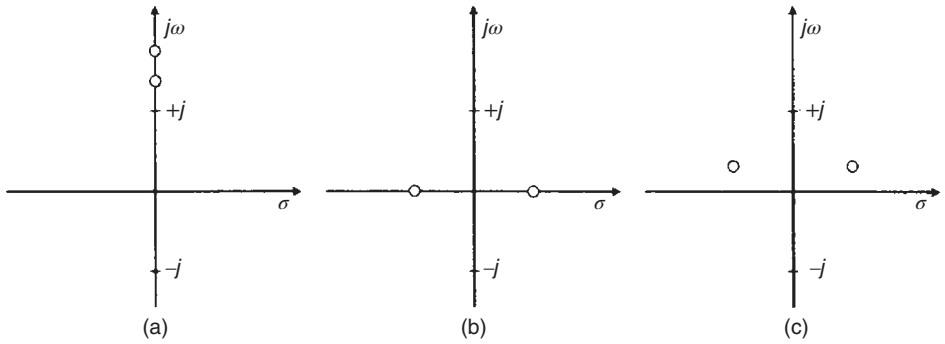


Figure 6.2 Possible positions for the finite-position roots of $P(s)$, the numerator of $S_{21}(s)$. (a) Two asymmetric imaginary-axis zeros, (b) real-axis pair, and (c) complex pair.

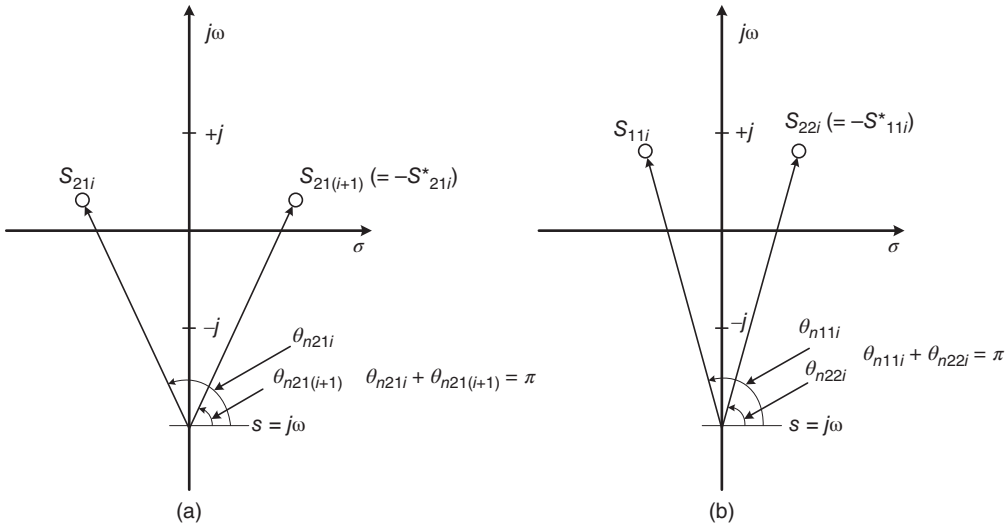


Figure 6.3 Angles of $S_{21}(s)$, $S_{11}(s)$, and $S_{22}(s)$ numerator polynomials for an arbitrary position of the frequency variable s on the imaginary axis: (a) for a transmission zero pair; (b) for an $S_{11}(s)$ zero and the complementary $S_{22}(s)$ zero.

Equation (6.15) leads to the first of two rules associated with the orthogonality condition:

- If two monic Nth-degree polynomials $Q_1(s)$ and $Q_2(s)$ have zeros that form paraconjugate pairs about the imaginary axis, or lie coincidentally on the imaginary axis itself, then one polynomial may be formed from the other through the relationship $Q_2(s) = (-1)^N Q_1(s)^*$. The multiplication by -1 when N is odd gives an expression for the angle of the numerator of $F_{22}(s)$ (equation (6.15)):

$$\theta_{n22}(s) = -\theta_{n11}(s) + N\pi \tag{6.16}$$

By substituting equation (6.16) into equation (6.12),

$$-\theta_{n21}(s) + \frac{N\pi}{2} = \frac{\pi}{2}(2k \pm 1) \tag{6.17}$$

Finally, turning to $\theta_{n21}(s)$, this is the angle of the numerator polynomial of $S_{21}(s)$, $P(s)$, which is of degree n_{fz}, the number of finite-position TZs in the transfer function. Because of the

symmetry of these TZs about the imaginary axis, $\theta_{n21}(s)$ is an integral number of $\pi/2$ radians regardless of the values of s and n_{fz} (Figure 6.3a), represented as

$$\theta_{n21}(s) = \frac{n_{fz}\pi}{2} + k_1\pi \tag{6.18}$$

where k_1 is an integer. Now by substituting into equation (6.17),

$$-\frac{n_{fz}\pi}{2} - k_1\pi + \frac{N\pi}{2} = \frac{\pi}{2}(2k \pm 1)$$
$$(N - n_{fz})\frac{\pi}{2} - k_1\pi = \frac{\pi}{2}(2k \pm 1) \tag{6.19}$$

Equation (6.19) implies that the integer quantity $(N - n_{fz})$ must be odd so that its RHS is satisfied. For the cases where $(N - n_{fz})$ is even, then an extra $\pi/2$ radians must be added to the left-hand side (LHS) of equation (6.19) to maintain the orthogonality condition, which is the same as adding $\pi/2$ to $\theta_{n21}(s)$ on the LHS of (6.18). Adding $\pi/2$ to $\theta_{n21}(s)$ is the equivalent of multiplying the polynomial $P(s)$ by j. Table 6.1 summarizes these conclusions.

This is the second of the two rules associated with the orthogonality condition:

- When the highest-degree coefficients of the Nth-degree polynomials $E(s)$ and $F(s)$ and the n_{fz} ($\leq N$) degree polynomial $P(s)$ are normalized to unity (i.e., monic polynomials), then the $P(s)$ polynomial must be multiplied by j when the integer quantity $(N - n_{fz})$ is even. Because the zeros of $P(s)$ lie on the imaginary axis or are positioned in paraconjugate pairs about the imaginary axis, $P(s)^* = (-1)^{n_{fz}}P(s)$, and so it may be reasoned that $P(s)^* = -(-1)^N P(s)$ for $(N - n_{fz})$ even or odd.

  In practice the polynomials $E(s)$, $F(s)$, and $P(s)$ are usually formed as products of their singularities, for example, $F(s) = \prod_{i=1}^{N}(s - s_{11i})$, so, their leading coefficients are automatically equal to unity in the majority of the cases.

  Knowing that $P(s)^* = -(-1)^N P(s)$ and $F_{22}(s) = (-1)^N F(s)^*$ and assuming that the network is reciprocal (i.e., $S_{21}(s) = S_{12}(s)$), then the S-parameters in the conservation of energy equations (6.2) and (6.3) may be written in terms of their rational polynomials (see equations (6.5) and (6.6)):

$$F(s)F_{22}(s)/\varepsilon_R^2 - P(s)^2/\varepsilon^2 = (-1)^N E(s)E(s)^* \tag{6.20}$$

for $(N - n_{fz})$ even or odd.

The orthogonality condition (equation (6.4)) may be rewritten as

$$\text{for } (N - n_{fz}) \text{ odd}, \quad F(s)P(s)^* + P(s)F_{22}(s)^* = 0 \tag{6.21a}$$
$$\text{for } (N - n_{fz}) \text{ even}, \quad F(s)[jP(s)]^* + [jP(s)]F_{22}(s)^* = 0$$
$$\text{or} \quad F(s)P(s)^* - P(s)F_{22}(s)^* = 0 \tag{6.21b}$$

Table 6.1 Multiplication of $P(s)$ by j to Satisfy the Orthogonality Condition.

N	n_{fz}	$(N - n_{fz})$	Multiply $P(s)$ by j
Odd	Odd	Even	Yes
Odd	Even	Odd	No
Even	Odd	Odd	No
Even	Even	Even	Yes

The unitary conditions may also be represented in S-matrix form [2]:

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \frac{1}{E(s)} \begin{bmatrix} F(s)/\varepsilon_R & jP(s)/\varepsilon \\ jP(s)/\varepsilon & (-1)^N F(s)^*/\varepsilon_R \end{bmatrix} \quad (N - n_{fz}) \text{ even} \tag{6.22a}$$

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \frac{1}{E(s)} \begin{bmatrix} F(s)/\varepsilon_R & P(s)/\varepsilon \\ P(s)/\varepsilon & (-1)^N F(s)^*/\varepsilon_R \end{bmatrix} \quad (N - n_{fz}) \text{ odd} \tag{6.22b}$$

For Chebyshev or Butterworth prototypes, where all the reflection zeros lie on the imaginary axis, $F_{22}(s) = (-1)^N F(s)^* = F(s)$ for N even or odd.

From the unitary conditions, three additional relationships may be derived:

- The relationship between the real constants ε and ε_R used to normalize the polynomials $P(s)$ and $F(s)$, respectively
- The relationships between the S-parameters and the polynomials of the $[ABCD]$ transfer matrix and the short-circuit admittance parameters
- The alternating pole principle that allows the accurate determination of either $E(s)$, $F(s)$, or $P(s)$, knowing the other two

6.1.1 Relationship Between ε and ε_R

The transfer and reflection functions $S_{21}(s)$ and $S_{11}(s)$ were defined earlier in terms of their rational polynomials,

$$S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)} \quad S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)} \tag{6.23}$$

respectively, where $E(s)$ and $F(s)$ are Nth degree, $P(s)$ is a polynomial of degree n_{fz}—the number of finite-position TZs—and ε and ε_R are real constants normalizing $P(s)$ and $F(s)$ such that $|S_{21}(s)|$ and $|S_{11}(s)|$ are ≤ 1 at any value of s, the frequency variable. It is assumed that the three polynomials have been normalized such that their highest-power coefficients are unity. Supplemental to this, for the cases where $(N - n_{fz})$ is an even integer, the polynomial $P(s)$ is multiplied by j to satisfy the unitary condition (6.7c) as mentioned above.

The real constant ε is determined by evaluating $P(s)/E(s)$ at a convenient value of s, where $|S_{21}(s)|$ or $|S_{11}(s)|$ are known, for example, at $s = \pm j$ where the equiripple return loss (RL) level for Chebyshev or the 3 dB (half-power point) for Butterworth filters is known. This sets the maximum level for $|S_{21}(s)|$ at 1, and if $n_{fz} < N$, $|S_{21}(s)| = 0$ at infinite frequency $s = \pm j\infty$. When $|S_{21}(s)| = 0$, the conservation of energy condition (6.7a) dictates that

$$S_{11}(j\infty) = \frac{1}{\varepsilon_R} \left| \frac{F(j\infty)}{E(j\infty)} \right| = 1 \tag{6.24}$$

Because the highest-degree coefficients (e_N and f_N) of $E(s)$ and $F(s)$, respectively, are unity, it is easily seen that $\varepsilon_R = 1$. When $n_{fz} = N$, that is, all N available TZs are at finite positions in the complex plane and therefore $P(s)$ is an Nth-degree polynomial (fully canonical function), then the attenuation at $s = \pm j\infty$ is finite and ε_R is derived from the conservation of energy condition equation (6.7a) by

$$\begin{aligned} &S_{11}(j\infty)S_{11}(j\infty)^* + S_{21}(j\infty)S_{21}(j\infty)^* = 1; \\ &\therefore \frac{F(j\infty)F(j\infty)^*}{\varepsilon_R^2 E(j\infty)E(j\infty)^*} + \frac{P(j\infty)P(j\infty)^*}{\varepsilon^2 E(j\infty)E(j\infty)^*} = 1 \end{aligned} \tag{6.25}$$

For the fully canonical case, $E(s)$, $F(s)$, and $P(s)$ are all Nth-degree polynomials with the highest-power coefficients = 1. Therefore, at $s = \pm j\infty$,

$$\frac{1}{\varepsilon_R^2} + \frac{1}{\varepsilon^2} = 1$$

$$\text{or}: \quad \varepsilon_R = \frac{\varepsilon}{\sqrt{\varepsilon^2 - 1}} \tag{6.26}$$

which is slightly greater than unity since ε is usually greater than 1.

Also, it is evident that for the fully canonical case, the insertion loss at $s = \pm j\infty$ is

$$S_{21}(\pm j\infty) = \frac{1}{\varepsilon} = 20\log_{10}\varepsilon \,\text{dB} \tag{6.27a}$$

and the RL is

$$S_{11}(\pm j\infty) = \frac{1}{\varepsilon_R} = 20\log_{10}\varepsilon_R \,\text{dB} \tag{6.27b}$$

Although the fully canonical characteristics are rarely used for the synthesis of bandpass filters, they are sometimes used for bandstop filters, which will be considered in a later chapter.

6.1.2 Relationship Between the Polynomials of the [*ABCD*] Transfer Matrix Parameters and the *S*-Parameters

From the parameter conversion formulae in Table 5.2, it may be seen that the first element $A(s)$ of the $[ABCD]$ transfer matrix may be expressed in terms of the equivalent S-parameters:

$$A(s) = \frac{(1 + S_{11}(s))(1 - S_{22}(s)) + S_{12}(s)S_{21}(s)}{2S_{21}(s)} \tag{6.28}$$

The S-parameters in turn may be expressed in terms of the $E(s)$, $F(s)$, and $P(s)$ polynomials (equations (6.5) and (6.6)):

$$A(s) = \frac{(1 + F_{11}/E)(1 - F_{22}/E) + (P/E)^2}{2P/E} \tag{6.29}$$

where F_{11}, F_{22}, P, and E are short-hand notations for $F(s)/\varepsilon_R$, $F_{22}(s)/\varepsilon_R$, $P(s)/\varepsilon$ and $E(s)$, and E^*, F^*, P^*, etc. imply the paraconjugate operations $E(s)^*$, $F(s)^*/\varepsilon_R$, and $P(s)^*/\varepsilon$, respectively. $P(s) \rightarrow jP(s)$ when $(N - n_{fz})$ is an even integer.

The unitary condition equations (6.7a) and (6.7c) may also be expressed in terms of the E(s), F(s), and P(s) polynomials:

$$F_{11}F_{11}^* + PP^* = EE^* \tag{6.30a}$$

$$F_{11}P^* + PF_{22}^* = 0 \tag{6.30b}$$

Substituting for F_{22} in equation (6.29) using equation (6.30b),

$$A(s) = \frac{1}{2P}\left[(E + F_{11}) + \frac{P}{P^*}F_{11}^* + \frac{P}{P^*E}(F_{11}F_{11}^* + PP^*)\right] \tag{6.31}$$

or, by using equation (6.30a),

$$A(s) = \frac{1}{2P}\left[(E + F_{11}) + \frac{P}{P^*}(E + F_{11})^*\right] \tag{6.32}$$

Since the n_{fz} finite-position TZs that make up the S_{21} numerator polynomial $P(s)$ can only exist on the imaginary axis of the complex plane or in pairs symmetrically positioned about the imaginary axis (see Figure 6.3a), the relationship $P(s)^* = (-1)^{n_{fz}}P(s)$ holds. Also it was shown

in Section 6.1 that when $(N - n_{fz})$ is even, $P(s) \to jP(s)$. Thus $P/P^* = -(-1)^N$ for $(N - n_{fz})$ even or odd.

Therefore equation (6.32) becomes (now writing out in full):

$$A(s) = \frac{\frac{1}{2}[(E(s) + F(s)/\varepsilon_R) - (-1)^N(E(s) + F(s)/\varepsilon_R)^*]}{P(s)/\varepsilon} = \frac{A_n(s)}{P(s)/\varepsilon} \tag{6.33a}$$

Similarly, the rest of the $[ABCD]$ matrix parameters may be expressed in terms of the coefficients of the transfer and reflection S-parameter polynomials:

$$B(s) = \frac{\frac{1}{2}[(E(s) + F(s)/\varepsilon_R) + (-1)^N(E(s) + F(s)/\varepsilon_R)^*]}{P(s)/\varepsilon} = \frac{B_n(s)}{P(s)/\varepsilon} \tag{6.33b}$$

$$C(s) = \frac{\frac{1}{2}[(E(s) - F(s)/\varepsilon_R) + (-1)^N(E(s) - F(s)/\varepsilon_R)^*]}{P(s)/\varepsilon} = \frac{C_n(s)}{P(s)/\varepsilon} \tag{6.33c}$$

$$D(s) = \frac{\frac{1}{2}[(E(s) - F(s)/\varepsilon_R) - (-1)^N(E(s) - F(s)/\varepsilon_R)^*]}{P(s)/\varepsilon} = \frac{D_n(s)}{P(s)/\varepsilon} \tag{6.33d}$$

An alternative formulation for the $[ABCD]$ polynomials may be found by using the relations $z + z^* = 2\mathrm{Re}(z)$, $z - z^* = 2j\mathrm{Im}(z)$ (z = complex number). For example, the numerator polynomial of the $A(s)$ polynomial, $A_n(s)$, may be expressed as follows for N even:

$$A_n(s) = j\,\mathrm{Im}(e_0 + f_0) + \mathrm{Re}(e_1 + f_1)s + j\,\mathrm{Im}(e_2 + f_2)s^2 + \cdots + j\,\mathrm{Im}(e_N + f_N)s^N$$

and for N odd,

$$A_n(s) = \mathrm{Re}(e_0 + f_0) + j\,\mathrm{Im}(e_1 + f_1)s + \mathrm{Re}(e_2 + f_2)s^2 + \cdots + j\,\mathrm{Im}(e_N + f_N)s^N$$

where e_i and f_i, $i = 0, 1, 2, \ldots, N$, are the complex coefficients of $E(s)$ and $F(s)/\varepsilon_R$, respectively. The full set of expressions is given in equation set (7.20).

Since the $[ABCD]$ parameters all have the common denominator $P(s)/\varepsilon$, they may be expressed in matrix form:

$$[ABCD] = \frac{1}{P(s)/\varepsilon}\begin{bmatrix} A_n(s) & B_n(s) \\ C_n(s) & D_n(s) \end{bmatrix} \tag{6.33e}$$

where $A_n(s)$, $B_n(s)$, $C_n(s)$, and $D_n(s)$ are the numerators in equation set (6.33) and $P(s) \to jP(s)$ when $(N - n_{fz})$ is even.[3]

---

3 If the network is to operate between complex source and load impedances Z_S and Z_L, equation (6.33) may be modified as follows (see Appendix 6A):

$$A_n(s) = \frac{1}{2}[(Z_S^*E(s) + Z_SF(s)/\varepsilon_R) - (-1)^N(Z_S^*E(s) + Z_SF(s)/\varepsilon_R)^*]$$

$$B_n(s) = \frac{1}{2}[Z_L^*(Z_S^*E(s) + Z_SF(s)/\varepsilon_R) + (-1)^NZ_L(Z_S^*E(s) + Z_SF(s)/\varepsilon_R)^*]$$

$C_n(s)$ as in equation (6.33c)

$$D_n(s) = \frac{1}{2}[Z_L^*(E(s) - F(s)/\varepsilon_R) - (-1)^NZ_L(E(s) - F(s)/\varepsilon_R)^*]$$

$$P(s) \to \sqrt{\mathrm{Re}(Z_S)\mathrm{Re}(Z_L)} \cdot P(s)$$

6.1.2.1 Short-Circuit Admittance Parameters

Having found the [*ABCD*] transfer matrix parameters from the transfer and reflection *S*-parameter polynomials, the rational polynomials for the elements of the short-circuit admittance parameter matrix [*Y*] may also be expressed in terms of the *E*(s), *F*(s), and *P*(s) polynomials:

$$[Y] = \begin{bmatrix} y_{11}(s) & y_{12}(s) \\ y_{21}(s) & y_{22}(s) \end{bmatrix} = \frac{1}{y_d(s)} \begin{bmatrix} y_{11n}(s) & y_{12n}(s) \\ y_{21n}(s) & y_{22n}(s) \end{bmatrix} \tag{6.34}$$

where $y_{11n}(s)$, $y_{12n}(s)$, $y_{21n}(s)$, and $y_{22n}(s)$ are the numerator polynomials of the rational polynomials $y_{11}(s)$, $y_{12}(s)$, $y_{21}(s)$, and $y_{22}(s)$, respectively, and $y_d(s)$ is their common denominator. The [*Y*] matrix parameters may be related to the [*ABCD*] matrix parameters through the conversion formulas (Table 5.2):

$$\begin{aligned} [Y] = \frac{1}{y_d(s)} \begin{bmatrix} y_{11n}(s) & y_{12n}(s) \\ y_{21n}(s) & y_{22n}(s) \end{bmatrix} &= \begin{bmatrix} D(s)/B(s) & -\Delta_{ABCD}/B(s) \\ -1/B(s) & A(s)/B(s) \end{bmatrix} \\ &= \frac{1}{B_n(s)} \begin{bmatrix} D_n(s) & -\Delta_{ABCD}P(s)/\varepsilon \\ -P(s)/\varepsilon & A_n(s) \end{bmatrix} \end{aligned} \tag{6.35}$$

where Δ_{ABCD} is the determinant of the [*ABCD*] matrix. For a reciprocal passive network, $\Delta_{ABCD} = 1$, and so

$$\begin{aligned} y_{11n}(s) &= D_n(s) \\ y_{22n}(s) &= A_n(s) \\ y_{12n}(s) &= y_{21n}(s) = -P(s)/\varepsilon \\ y_d(s) &= B_n(s) \end{aligned} \tag{6.36}$$

Now the relationship between the short-circuit admittance parameters and the $E(s)$, $F(s)/\varepsilon_R$, and $P(s)/\varepsilon$ polynomials of the transfer and reflection functions may be found by substituting for $A_n(s)$, $B_n(s)$, and $D_n(s)$ using equation set (6.33). These relationships will be used later for the synthesis of electrical networks in Chapter 7 and coupling matrices in Chapter 8.

It will be demonstrated in these chapters that to achieve a successful synthesis of electrical networks and coupling matrices representing filter responses, the $B_n(s)$ polynomial (and therefore the admittance matrix denominator polynomial $y_d(s)$; see equation (6.36)) must be of degree N, the same as the degree of the filtering function. The above procedure ensures that the polynomial $B_n(s)$ is always of degree N.

6.2 Alternating Pole Method for the Determination of the Denominator Polynomial *E*(*s*)

For the polynomial synthesis method to be described later, TZs are prescribed in the complex plane, which defines the $S_{21}(s)$ numerator polynomial $P(s)$. Then the coefficients of the $S_{11}(s)$ numerator polynomial $F(s)$ are found using an analytic or recursive method. It then remains to find the $S_{11}(s)$ and $S_{21}(s)$ denominator polynomial $E(s)$ to complete the design of the filtering function.

If two of the three polynomials are known, the third is derived from the conservation of energy equation (6.7a) as follows:

$$\begin{aligned} &S_{11}(s)S_{11}(s)^* + S_{21}(s)S_{21}(s)^* = 1 \\ &\text{or} \quad F(s)F(s)^*/\varepsilon_R^2 + P(s)P(s)^*/\varepsilon^2 = E(s)E(s)^* \end{aligned} \tag{6.37}$$

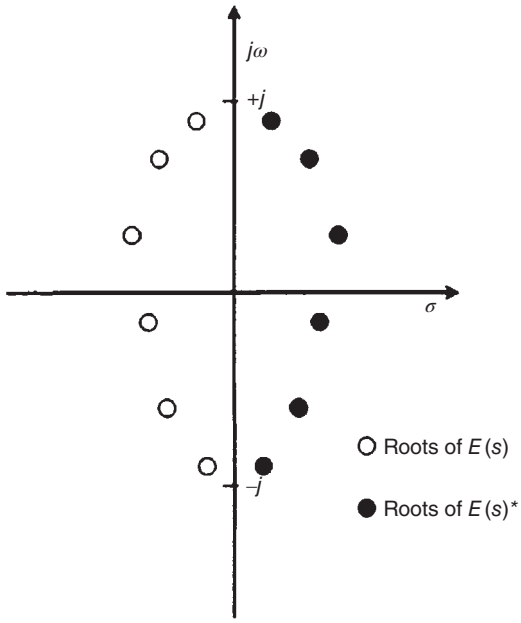


Figure 6.4 Pattern of the roots of $E(s)E(s)^*$ in the complex plane (symmetric about imaginary axis).

The LHS of equation (6.37) is constructed using polynomial multiplications to find $E(s)E(s)^*$, which must be a scalar quantity (Figure 6.4). This means that the $2N$ roots of $E(s)E(s)^*$ must form a symmetric pattern about the imaginary axis in the complex plane, so that, at any frequency s, the product $E(s)E(s)^*$ is scalar.

Since we know that the roots of $E(s)$ are strictly Hurwitz, those roots of $E(s)E(s)^*$ that are in the left-half plane must belong to $E(s)$, and those in the right-half plane, to $E(s)^*$. By choosing the N roots in the left-half plane, the polynomial $E(s)$ is formed.

Although this method is completely general, it does mean working with double-degree polynomials, and sometimes, with higher degree filter functions, the roots of $E(s)E(s)^*$ tend to cluster around $s = \pm j$, leading to inaccurate root finding. An alternative method due to Rhodes and Alseyab [3] is presented here, with which the roots of $E(s)$ can be found, without having to determine the roots of polynomials of degree $2N$.

Expanding equation (6.37) for the case $(N - n_{fz})$ odd gives

$$\varepsilon^2\varepsilon_R^2 E(s)E(s)^* = [\varepsilon_R P(s) + \varepsilon F(s)][\varepsilon_R P(s)^* + \varepsilon F(s)^*] - \varepsilon\varepsilon_R[F(s)P(s)^* + P(s)F(s)^*] \quad (6.38a)$$

The orthogonality unitary condition equation (6.21a) states that $F(s)P(s)^* + P(s)F_{22}(s)^* = 0$ for $(N - n_{fz})$ odd. Thus the term on the extreme right of equation (6.38a) will equal zero if $F(s) = F_{22}(s)$. This can only happen when all the zeros of $F(s)$ lie on the imaginary axis and are coincident with those of $F_{22}(s)$.

Similarly, expanding equation (6.37) for the case $(N - n_{fz})$ even gives

$$\begin{aligned}\varepsilon^2\varepsilon_R^2 E(s)E(s)^* &= [\varepsilon_R(jP(s)) + \varepsilon F(s)][\varepsilon_R(jP(s))^* + \varepsilon F(s)^*] \\ &\quad + j\varepsilon\varepsilon_R[F(s)P(s)^* - P(s)F(s)^*] \end{aligned} \quad (6.38b)$$

For the term on the extreme right of equation (6.38b) to equal zero, $F(s)P(s)^* - P(s)F(s)^* = 0$. Recalling the orthogonality unitary condition for $(N - n_{fz})$ even (equation (6.21b))

$$F(s)P(s)^* - P(s)F_{22}(s)^* = 0$$

it may be seen that the term on the extreme right of equation (6.38b) will equal zero if $F(s) = F_{22}(s)$. Once again this will only happen when all the zeros of $F(s)$ lie on the imaginary

axis and are coincident with those of $F_{22}(s)$. If the zeros of $F(s)$ and $F_{22}(s)$ do fulfill this condition, equation (6.38a) for $(N - n_{fz})$ odd reduces to

$$\varepsilon^2\varepsilon_R^2 E(s)E(s)^* = [\varepsilon_R P(s) + \varepsilon F(s)][\varepsilon_R P(s) + \varepsilon F(s)]^* \tag{6.39a}$$

and (6.38b) for $(N - n_{fz})$ even reduces to

$$\varepsilon^2\varepsilon_R^2 E(s)E(s)^* = [\varepsilon_R (jP(s)) + \varepsilon F(s)][\varepsilon_R (jP(s)) + \varepsilon F(s)]^* \tag{6.39b}$$

In the ω-plane, the polynomials $P(\omega)$ and $F(\omega)$ will have purely real coefficients. Because it is a condition that they have to be orthogonal (see equation (6.12)), equation (6.39b) may be modified as follows to find the ω-plane singularities for $(N - n_{fz})$ even or odd:

$$\varepsilon^2\varepsilon_R^2 E(\omega)E(\omega)^* = [\varepsilon_R P(\omega) - j\varepsilon F(\omega)][\varepsilon_R P(\omega) + j\varepsilon F(\omega)] \tag{6.40}$$

Rooting one of the two terms on the RHS of equation (6.39a) or (6.39b) results in a pattern of singularities alternating between the left-half and right-half planes, as shown in Figure 6.5.

Rooting the other term will give the complementary set of singularities, completing the symmetry of the pattern about the imaginary axis and ensuring that the RHS of equation (6.39) is properly scalar as the LHS demands.

As a result, it is only necessary to form one of the two terms in equation (6.39) from the known $P(s)$ and $F(s)$ polynomials and then root the resultant Nth-degree polynomial with complex coefficients to obtain the singularities. Knowing that the polynomial $E(s)$ must be Hurwitz, any singularity in the right-half plane can be reflected about the imaginary axis to lie in the image position in the left-half plane. Now, knowing the positions of the N singularities in the left-half plane, we can form the polynomial $E(s)$. In finding $E(s)$, we only have to deal with Nth-degree polynomials; also because the alternating singularities tend to be less clustered about $s = \pm j$, greater accuracy is guaranteed.

In practice, for most classes of filtering function, for example, Butterworth or Chebyshev, the reflection zeros lie on the imaginary axis, and the alternating singularity method can be applied to find $E(s)$. For certain specialized cases, for example, predistorted filters, where some or all of the reflection zeros lie in the complex plane and not on the imaginary axis, the conservation of energy method for finding $E(s)$ (equation (6.37)) has to be used.

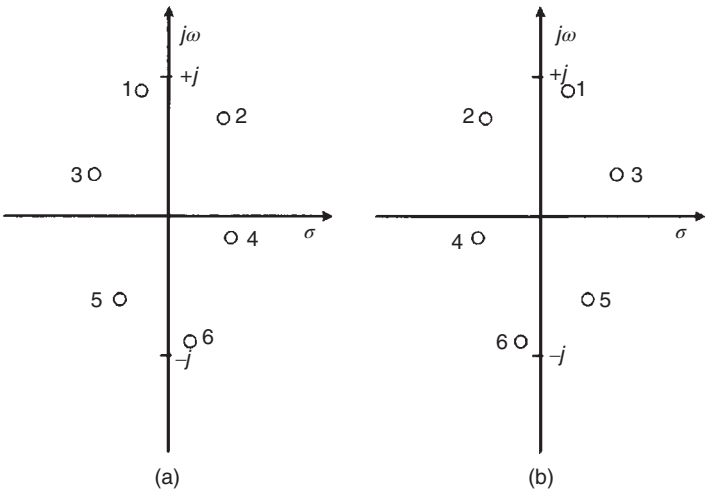


Figure 6.5 Zeros of the polynomials for a sixth-degree case where $(N - n_{fz})$ is even (equation (6.39b)): (a) $[\varepsilon_R(jP(s)) + \varepsilon F(s)]$ and (b) $[\varepsilon_R(jP(s)) + \varepsilon F(s)]^*$.

6.3 General Polynomial Synthesis Methods for Chebyshev Filter Functions

The methods presented in this section include the design of the transfer functions and the synthesis of the prototype filter networks with characteristics belonging to the general class of Chebyshev filter function:

- Even or odd degree
- Prescribed transmission and/or group delay equalization zeros
- Asymmetric or symmetric characteristics
- Single- or double-terminated networks

In this section, an efficient recursive technique for generating the Chebyshev transfer and reflection polynomials, given the arbitrary location of the TZs, is described [4, 5]. In the next chapter, the methods for generating the corresponding circuits and the "$N \times N$" and "$N + 2$" coupling matrices from these polynomials are described.

6.3.1 Polynomial Synthesis

To simplify the mathematics in this section, we will work in the ω-plane, where ω is the real frequency variable related to the more familiar complex frequency variable s by $s = j\omega$.

For any two-port lossless filter network composed of a series of N intercoupled resonators, the transfer and reflection functions can be expressed as a ratio of two Nth-degree polynomials:

$$S_{11}(\omega) = \frac{F(\omega)/\varepsilon_R}{E(\omega)} \quad S_{21}(\omega) = \frac{P(\omega)/\varepsilon}{E(\omega)} \tag{6.41}$$

$$\varepsilon = \frac{1}{\sqrt{10^{RL/10} - 1}} \left| \frac{P(\omega)}{F(\omega)/\varepsilon_R} \right|_{\omega \pm 1} \tag{6.42}$$

where RL is the prescribed RL level in dB. It is assumed that the polynomials $P(\omega)$, $F(\omega)$, and $E(\omega)$ are normalized such that their highest-degree coefficients are unity. $S_{11}(\omega)$ and $S_{21}(\omega)$ share a common denominator $E(\omega)$, and the polynomial $P(\omega) = \prod_{n=1}^{n_{fz}} (\omega - \omega_n)$ carries the n_{fz} transfer function finite-position TZs.[4] For a Chebyshev filter function, ε is a constant normalizing $S_{21}(\omega)$ to the equiripple level at $\omega = \pm 1$.

Since, by definition, a Chebyshev function has all its reflection zeros on the real axis of the ω-plane, the alternating pole formula for a lossless network, as given by equation (6.40), is as follows:

$$S_{21}(\omega) S_{21}(\omega)^* = \frac{P(\omega) P(\omega)^*}{\varepsilon^2 E(\omega) E(\omega)^*} = \frac{1}{\left[1 - j\frac{\varepsilon}{\varepsilon_R} kC_N(\omega)\right] \left[1 + j\frac{\varepsilon}{\varepsilon_R} kC_N(\omega)^*\right]}$$

where

$$kC_N(\omega) = \frac{F(\omega)}{P(\omega)} \tag{6.43}$$

and k is a constant.[5]

---

4 $P(\omega) = 1$ for $n_{fz} = 0$.

5 k is an unimportant normalizing constant inserted here to account for the fact that in general the polynomials of $C_N(\omega)$ have nonunity highest-degree coefficients, whereas the polynomials $P(\omega)$ and $F(\omega)$ in this text are assumed to be monic.

$C_N(\omega)$ is known as the filtering function of degree N, and its poles and zeros are the roots of $P(\omega)$ and $F(\omega)$, respectively. For the general Chebyshev characteristic, it has the form [4]

$$C_N(\omega) = \cosh\left[\sum_{n=1}^{N} \cosh^{-1}(x_n(\omega))\right] \tag{6.44a}$$

or by using the identity $\cosh\theta = \cos j\theta$, the alternative expression for $C_N(\omega)$ is given by

$$C_N(\omega) = \cos\left[\sum_{n=1}^{N} \cos^{-1}(x_n(\omega))\right] \tag{6.44b}$$

where $x_n(\omega)$ is a function of the frequency variable ω. For analyzing $C_N(\omega)$, equation (6.44a) can be adopted for $|\omega| \geq 1$ and equation (6.44b) for $|\omega| \leq 1$.

To properly represent a Chebyshev function, $x_n(\omega)$ requires the following properties:

- At $\omega = \omega_n$, where ω_n is a finite-position prescribed TZ or is at infinite frequency ($\omega_n = \pm\infty$), $x_n(\omega) = \pm\infty$
- At $\omega = \pm 1$, $x_n(\omega) = \pm 1$
- Between $\omega = -1$ and $\omega = 1$ (in-band), $1 \geq x_n(\omega) \geq -1$

The first condition is satisfied if $x_n(\omega)$ is a rational function with its denominator equal to $\omega - \omega_n$:

$$x_n(\omega) = \frac{f(\omega)}{\omega - \omega_n} \tag{6.45}$$

The second condition states that, at $\omega = \pm 1$,

$$x_n(\omega)|_{\omega=\pm 1} = \left.\frac{f(\omega)}{\omega - \omega_n}\right|_{\omega=\pm 1} = \pm 1 \tag{6.46}$$

This condition is satisfied if $f(1) = 1 - \omega_n$ and $f(-1) = 1 + \omega_n$, giving $f(\omega) = 1 - \omega\omega_n$. Therefore,

$$x_n(\omega) = \frac{1 - \omega\omega_n}{\omega - \omega_n} \tag{6.47}$$

Differentiating $x_n(\omega)$ with respect to ω shows that there are no turning points or inflection points between $\omega = -1$ and $\omega = 1$. If $x_n(\omega) = -1$ at $\omega = -1$ and $x_n(\omega) = +1$ at $\omega = +1$, then $|x_n(\omega)| \leq 1$, while $|\omega| \leq 1$, thus satisfying the third condition. Figure 6.6 illustrates the behavior of $x_n(\omega)$ for a case where the finite TZ $\omega_n = +1.3$.

By dividing by ω_n to deal with any poles at $\omega_n = \pm\infty$, the final form for the function $x_n(\omega)$ is

$$x_n(\omega) = \frac{\omega - 1/\omega_n}{1 - \omega/\omega_n} \tag{6.48}$$

In equation (6.48), $\omega_n = s_n/j$ is the position of the nth TZ in the complex frequency plane, and it may be verified from equations (6.44) and (6.48) that $C_N(\omega)|_{\omega=\pm 1} = 1$ when $|\omega| \leq 1$, $|C_N(\omega)| \leq 1$ and when $|\omega| > 1$, $|C_N(\omega)| > 1$, all of which are necessary conditions for a Chebyshev response. Also, as all N of the prescribed TZs ω_n approach infinity, $C_N(\omega)$ degenerates to the familiar pure Chebyshev function:

$$C_N(\omega)|_{\omega_n \to \infty} = \cosh[N\cosh^{-1}(\omega)] \tag{6.49}$$

The rule to be observed when prescribing the positions of the TZs is that symmetry must be preserved about the imaginary ($j\omega$) axis of the complex s-plane to ensure that the unitary conditions are preserved. Also, for the polynomial synthesis methods about to be described, the

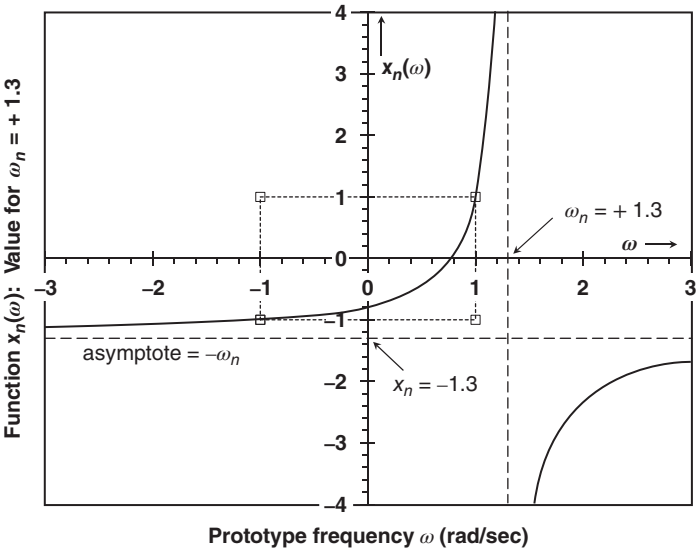


Figure 6.6 Function $x_n(\omega)$ in the ω-plane with a prescribed transmission zero at $\omega_n = +1.3$.

number of TZs with finite positions in the s-plane n_{fz} must be $\leq N$. If $n_{fz} < N$, those zeros without finite positions need to be placed at infinity. Although the "$N + 2$" coupling matrix (Chapter 8) can accommodate fully canonical filtering functions (i.e., the number of finite-position TZs n_{fz} equals the degree N of the filter), the $N \times N$ matrix can only accommodate a maximum of $N - 2$ finite-position zeros (minimum path rule[6]). When synthesizing the polynomials for the $N \times N$ coupling matrix, at least two of the TZs must be placed at infinity.

The next step is to solve equation (6.44a) to determine the coefficients of the numerator polynomial of $C_N(\omega)$. Normalizing this polynomial to set its highest-degree coefficient to unity will yield the monic polynomial $F(\omega)$. Now knowing ε and ε_R ($\varepsilon_R = 1$ for noncanonical cases) from equation (6.42), and the prescribed zero polynomial $P(\omega) = \prod_{n=1}^{n_{fz}} (\omega - \omega_n)$, the Hurwitz polynomial $E(\omega)$ common to the denominators of $S_{21}(\omega)$ and $S_{11}(\omega)$ can be evaluated from equation (6.43). It then becomes possible to proceed onward to prototype network synthesis, from which a real electrical network with the transfer and reflection characteristics $S_{21}(\omega)$ and $S_{11}(\omega)$ may be synthesized.

The first step in the polynomial synthesis procedure is to replace the $\cosh^{-1}$ term in equation (6.44a) with its identity:

$$C_N(\omega) = \cosh\left[\sum_{n=1}^{N} \ln(a_n + b_n)\right]$$

where

$$a_n = x_n(\omega) \quad \text{and} \quad b_n = (x_n^2(\omega) - 1)^{\frac{1}{2}} \tag{6.50}$$

6 The minimum path rule is a simple formula for calculating the number of finite-position transmission zeros (i.e., those that are not at $\omega = \pm\infty$) that an Nth-degree direct-coupled resonator network can be realized. If the number of resonators in the shortest path (along nonzero interresonator couplings) between the input (source) and output (load) terminals of a network equals $n_{\min}$, then the number of finite-position TZs (n_{fz}) that the network may realize is $n_{fz} = N - n_{\min}$.

Then

$$C_N(\omega) = \frac{1}{2}\left[e^{\sum \ln(a_n+b_n)} + e^{-\sum \ln(a_n+b_n)}\right]$$
$$= \frac{1}{2}\left[\prod_{n=1}^{N}(a_n + b_n) + \frac{1}{\prod_{n=1}^{N}(a_n + b_n)}\right] \tag{6.51}$$

Multiplying the numerator and denominator of the second term in equation (6.51) by $\prod_{n=1}^{N}(a_n - b_n)$ and recognizing that $\prod_{n=1}^{N}(a_n + b_n) \cdot \prod_{n=1}^{N}(a_n - b_n) = \prod_{n=1}^{N}(a_n^2 - b_n^2) = 1$,

$$C_N(\omega) = \frac{1}{2}\left[\prod_{n=1}^{N}(a_n + b_n) + \prod_{n=1}^{N}(a_n - b_n)\right] \tag{6.52}$$

Equation (6.52) is expressed in its final form by substituting for a_n, b_n, and $x_n(\omega)$ using equations (6.48) and (6.50):

$$a_n = \frac{\omega - 1/\omega_n}{1 - \omega/\omega_n}$$
$$b_n = \frac{\sqrt{(\omega - 1/\omega_n)^2 - (1 - \omega/\omega_n)^2}}{1 - \omega/\omega_n}$$
$$= \frac{\sqrt{(\omega^2 - 1)(1 - 1/\omega_n^2)}}{1 - \omega/\omega_n}$$
$$= \frac{\omega'\sqrt{(1 - 1/\omega_n^2)}}{1 - \omega/\omega_n} \tag{6.53}$$

where $\omega' = \sqrt{(\omega^2 - 1)}$, a transformed frequency variable.

Equation (6.52) then becomes

$$C_N(\omega) = \frac{1}{2}\left[\frac{\prod_{n=1}^{N}\left[(\omega - 1/\omega_n) + \omega'\sqrt{(1 - 1/\omega_n^2)}\right] + \prod_{n=1}^{N}\left[(\omega - 1/\omega_n) - \omega'\sqrt{(1 - 1/\omega_n^2)}\right]}{\prod_{n=1}^{N}(1 - \omega/\omega_n)}\right]$$
$$= \frac{1}{2}\left[\frac{\prod_{n=1}^{N}(c_n + d_n) + \prod_{n=1}^{N}(c_n - d_n)}{\prod_{n=1}^{N}(1 - \omega/\omega_n)}\right] \tag{6.54}$$

where

$$c_n = (\omega - 1/\omega_n) \text{ and } d_n = \omega'\sqrt{(1 - 1/\omega_n^2)}$$

Now comparing with equation (6.43), it may be seen that the denominator of $C_N(\omega)$ has the same zeros as $P(\omega)$, the numerator polynomial of $S_{21}(\omega)$ generated from the prescribed TZs ω_n. It may also be seen from equations (6.43) and (6.54) that the numerator of $C_N(\omega)$ has the same zeros as the numerator $F(\omega)$ of $S_{11}(\omega)$ and appears at first to be a mixture of two finite-degree polynomials, one in the variable ω purely and the other with each coefficient multiplied by the transformed variable ω'.

However the coefficients multiplied by ω' cancel each other when equation (6.54) is multiplied out. This is best proven by demonstration by multiplying out the left- and right-hand product terms in the numerator of $C_N(\omega)$ in equation (6.54) for a few low values of N:

$$\text{For } N = 1 \quad \text{Num}[C_1(\omega)] = \frac{1}{2}\left[\prod_{n=1}^{1}(c_n + d_n) + \prod_{n=1}^{1}(c_n - d_n)\right] = c_1$$

$$\text{Similarly for } N = 2 \quad \text{Num}[C_2(\omega)] = c_1c_2 + d_1d_2$$

$$\text{and for } N = 3 \quad \text{Num}[C_3(\omega)] = (c_1c_2 + d_1d_2)c_3 + (c_2d_1 + c_1d_2)d_3$$

$$\text{and so on}$$

At each stage, the expansions result in a sum of factors, where each factor is composed of multiples of c_n and d_n elements. Because of the positive sign in the LH product term in equation (6.54), the c_n, d_n factors resultant from multiplying out the LH term are always positive in sign. Multiplying out the RH product term in equation (6.54) produces the same c_n, d_n factors; however, the negative sign indicates that those factors containing an odd number of d_n elements are negative in sign and will cancel with the corresponding factors from the LH product term.

Now, the remaining factors contain only even numbers of d_n elements. Consequently, $\omega' = \sqrt{\omega^2 - 1}$, which is a common multiplier for all the d_n elements (equation (6.54)), is raised by even powers only, producing sub-polynomials in the variable ω. As a result, the numerator of $C_N(\omega)$ is a polynomial in variable ω only.

These relationships may be used to develop a simple algorithm to determine the coefficients of the numerator polynomial of $C_N(\omega)$. Normalizing the numerator of $C_N(\omega)$ to set its highest-degree coefficient to unity yields $F(\omega)$, the numerator of $S_{11}(\omega)$.

6.3.2 Recursive Technique

The numerator of equation (6.54) may be written as

$$\text{Num}[C_N(\omega)] = \frac{1}{2}[G_N(\omega) + G'_N(\omega)] \tag{6.55}$$

where

$$G_N(\omega) = \prod_{n=1}^{N}[c_n + d_n] = \prod_{n=1}^{N}\left[\left(\omega - \frac{1}{\omega_n}\right) + \omega'\sqrt{\left(1 - \frac{1}{\omega_n^2}\right)}\right] \tag{6.56a}$$

and

$$G'_N(\omega) = \prod_{n=1}^{N}[c_n - d_n] = \prod_{n=1}^{N}\left[\left(\omega - \frac{1}{\omega_n}\right) - \omega'\sqrt{\left(1 - \frac{1}{\omega_n^2}\right)}\right] \tag{6.56b}$$

The method for computing the coefficients of the numerator of $C_N(\omega)$ is a recursive technique, where the polynomials for the nth degree are built up from the results of the $(n-1)$th degree. First, let us consider the polynomial $G_N(\omega)$ (equation (6.56a)). This polynomial can be rearranged into two polynomials, $U_N(\omega)$ and $V_N(\omega)$, where the $U_N(\omega)$ polynomial contains the coefficients of the terms in the variable ω only, whereas each coefficient of the auxiliary polynomial $V_N(\omega)$ is multiplied by the transformed variable ω':

$$G_N(\omega) = U_N(\omega) + V_N(\omega)$$

$$\text{where} \quad U_N(\omega) = u_0 + u_1\omega + u_2\omega^2 + \cdots + u_N\omega^N$$

$$\text{and} \quad V_N(\omega) = \omega'(v_0 + v_1\omega + v_2\omega^2 + \cdots + v_N\omega^N) \tag{6.57}$$

The recursion cycle is initiated with the terms corresponding to the first prescribed TZ ω_1, that is, by putting $N = 1$ in equations (6.56a) and (6.57):

$$\begin{aligned} G_1(\omega) &= [c_1 + d_1] \\ &= \left(\omega - \frac{1}{\omega_1}\right) + \omega' \sqrt{\left(1 - \frac{1}{\omega_1^2}\right)} \\ &= U_1(\omega) + V_1(\omega) \end{aligned} \tag{6.58}$$

For the first cycle of the process, $G_1(\omega)$ needs to be multiplied by the terms corresponding to the second prescribed zero ω_2 (equation (6.56a)), given by

$$\begin{aligned} G_2(\omega) &= G_1(\omega) \cdot [c_2 + d_2] \\ &= [U_1(\omega) + V_1(\omega)] \cdot \left[\left(\omega - \frac{1}{\omega_2}\right) + \omega' \sqrt{\left(1 - \frac{1}{\omega_2^2}\right)}\right] \\ &= U_2(\omega) + V_2(\omega) \end{aligned} \tag{6.59}$$

Multiplying out this expression for $G_2(\omega)$, and, again, allocating terms purely in ω to $U_2(\omega)$, terms multiplied by ω' to $V_2(\omega)$, and recognizing that $\omega' V_n(\omega)$ results in $(\omega^2 - 1) \cdot (v_0 + v_1\omega + v_2\omega^2 + \cdots + v_n\omega^n)$ (equation (6.57)), a polynomial purely in ω and therefore to be allocated to $U_n(\omega)$:

$$U_2(\omega) = \omega U_1(\omega) - \frac{U_1(\omega)}{\omega_2} + \omega' \sqrt{\left(1 - \frac{1}{\omega_2^2}\right)} V_1(\omega) \tag{6.60a}$$

$$V_2(\omega) = \omega V_1(\omega) - \frac{V_1(\omega)}{\omega_2} + \omega' \sqrt{\left(1 - \frac{1}{\omega_2^2}\right)} U_1(\omega) \tag{6.60b}$$

Having obtained these new polynomials $U_2(\omega)$ and $V_2(\omega)$, the cycle is repeated with the third prescribed zero, and so on, until all N of the prescribed zeros (including those at $\omega_n = \infty$) are used, that is, $(N - 1)$ cycles.

This procedure may be very easily programmed—a compact Fortran subroutine is given below. The N ω-plane TZs (including those at infinity) are contained in the complex array XP:

```
                                          initialize with first prescribed zero ω1
      X=1.0/XP(1)
      Y=CDSQRT(1.0-X**2)
      U(1)=-X
      U(2)=1.0
      V(1)=Y
      V(2)=0.0
C
      DO 10  K=3, N+1                     multiply in second and subsequent
                                          prescribed zeros
      X=1.0/XP(K-1)
      Y=CDSQRT(1.0-X**2)
      U2(K)=0.0
      V2(K)=0.0
      DO  11  J=1, K-1                    multiply by constant terms
      U2(J)=-U(J)*X-Y*V(J)
```

```
11      V2(J)=-V(J)*X+Y*U(J)
        DO  12  J=2, K                    multiply by terms in ω
        U2(J)=U2(J)+U(J-1)
12      V2(J)=V2(J)+V(J-1)
        DO  13  J=3, K                    multiply by terms in ω²
13      U2(J)=U2(J)+Y*V(J-2)
        DO  14  J=1, K                    update U_n and V_n
        U(J)=U2(J)
14      V(J)=V2(J)
10      CONTINUE                          reiterate for 3rd , 4th … Nth
                                          prescribed zero.
```

If the same process is repeated for $G'_N(\omega) = U'_N(\omega) + V'_N(\omega)$ (equation (6.56b)), it will be found that $U'_N(\omega) = U_N(\omega)$ and $V'_N(\omega) = -V_N(\omega)$. Therefore, from equations (6.55) and (6.57),

$$\begin{aligned}\text{num}[C_N(\omega)] &= \frac{1}{2}[G_N(\omega) + G'_N(\omega)] \\ &= \frac{1}{2}((U_N(\omega) + V_N(\omega)) + (U'_N(\omega) + V'_N(\omega))) = U_N(\omega)\end{aligned} \tag{6.61}$$

Equation (6.61) demonstrates that the numerator of $C_N(\omega)$ (which has the same zeros as $F(\omega)$) is equal to $U_N(\omega)$ after $(N-1)$ cycles of this recursion method. Now the zeros of $F(\omega)$ are revealed by finding the roots of $U_N(\omega)$, and, along with the prescribed zero polynomial $P(\omega)/\varepsilon$, the denominator polynomial $E(\omega)$ is constructed by using the alternating singularity principle:

- Now knowing $F(\omega)$ and $P(\omega)$, construct the complex equation $P(\omega)/\varepsilon - jF(\omega)/\varepsilon_R = 0$ (equation (6.40)) and find its zeros (which will alternate between the lower and upper halves of the ω-plane). After prescribing the RL level, the ratio $\varepsilon/\varepsilon_R$ may be found from equation (6.42).
- Take the conjugate of any zero that is in the lower half of the ω-plane (this is equivalent to reflecting zeros in the right half of the s-plane about the imaginary axis to satisfy the Hurwitz condition).
- Reconstruct the polynomial to obtain $E(\omega)$.

To illustrate the procedure, the recursions are applied to a fourth-degree example with an equiripple RL level of 22 dB and prescribed zeros at $+j1.3217$ and $+j1.8082$, chosen to give two attenuation lobe levels of 30 dB each on the upper side of the passband.

Initializing (equation (6.58)), with $\omega_1 = +1.3217$,

$$U_1(\omega) = -0.7566 + \omega$$
$$V_1(\omega) = \omega'(0.6539)$$

After the first cycle, with $\omega_2 = +1.8082$,

$$U_2(\omega) = -0.1264 - 1.3096\omega + 1.5448\omega^2$$
$$V_2(\omega) = \omega'(-0.9920 + 1.4871\omega)$$

After the second cycle, with $\omega_3 = \infty$,

$$U_3(\omega) = 0.9920 - 1.6134\omega - 2.3016\omega^2 + 3.0319\omega^3$$
$$V_3(\omega) = \omega'(-0.1264 - 2.3016\omega + 3.0319\omega^2)$$

Table 6.2 4–2 Asymmetric Chebyshev Filter Function with Two Prescribed Transmission Zeros: (a) Normalized Transfer and Reflection Function Polynomials and (b) Corresponding Singularities.

(a)

s^i, i	$P(s)$	$F(s)\ (=U_4(s))$	$E_4(s)$
0	$-j2.3899$	$+0.0208$	$-0.1268 - j2.0658$
1	$+3.1299$	$-j0.5432$	$+2.4874 - j3.6255$
2	$+j1.0$	$+0.7869$	$+3.6706 - j2.1950$
3		$-j0.7591$	$+2.4015 - j0.7591$
4		$+1.0$	$+1.0$
	$\varepsilon = 1.1548$	$\varepsilon_R = 1.0$	

(b)

	Transmission Zeros (Prescribed)	Reflection Zeros (Roots of $U_4(s) = F_4(s)$)	Transmission/Reflection Poles (Roots of $E_4(s)$)	In-Band Reflection Maxima[a)] (Roots of $V_4(s)$)	Out-of-Band Rejection Lobe Positions[a)]
1	$+j1.3217$	$-j0.8593$	$-0.7437 - j1.4178$	$-j0.4936$	$+j1.4587$
2	$+j1.8082$	$-j0.0365$	$-1.1031 + j0.1267$	$+j0.3796$	$+j2.8566$
3	$j\infty$	$+j0.6845$	$-0.4571 + j0.9526$	$+j0.8732$	—
4	$j\infty$	$+j0.9705$	$-0.0977 + j1.0976$	—	—

a) See Section 6.3.4.

After the third and final cycle, with $\omega_4 = \infty$,

$$U_4(\omega) = 0.1264 + 3.2936\omega - 4.7717\omega^2 - 4.6032\omega^3 + 6.0637\omega^4$$
$$V_4(\omega) = \omega'(0.9920 - 1.7398\omega - 4.6032\omega^2 + 6.0637\omega^3)$$

At this stage, the polynomial $U_4(\omega)$ is the unnormalized numerator polynomial of the reflection function $S_{11}(\omega)$ ($F_4(\omega)$), and finding its roots yields the N in-band reflection zeros. The roots of $V_4(\omega)$ yield the $N-1$ in-band reflection maxima (see Section 6.3.4). The s-plane coordinates of the zeros together with the corresponding transmission poles are listed in Table 6.2b, and plots of the transfer and reflection characteristics are shown in Figure 6.7.

6.3.3 Polynomial Forms for Symmetric and Asymmetric Filtering Functions

Table 6.2a lists the coefficients of the $E(s)$, $F(s)$, and $P(s)$ polynomials formed from the singularities in Table 6.2b, normalized in each case to set the highest-degree coefficient to unity. In addition, the coefficients of $P(s)$ are multiplied by j to preserve the unitary conditions, because in this case, $N - n_{fz} = 2$ is an even integer (see Section 6.1). With these polynomials, we are now ready to proceed with the synthesis of a prototype electrical network that, under analysis, results in exactly the same performance characteristics as the original polynomials.

At this stage it is worth studying the form of the polynomials of the fourth-degree asymmetric filter used as an example above to help understand the distinctive characteristics of the electrical network that will be synthesized from them.

- Because the zeros of $E(s)$ (the poles of the filter function) are asymmetrically located on both the real and the imaginary axes, all except the leading coefficient of $E(s)$ are complex.

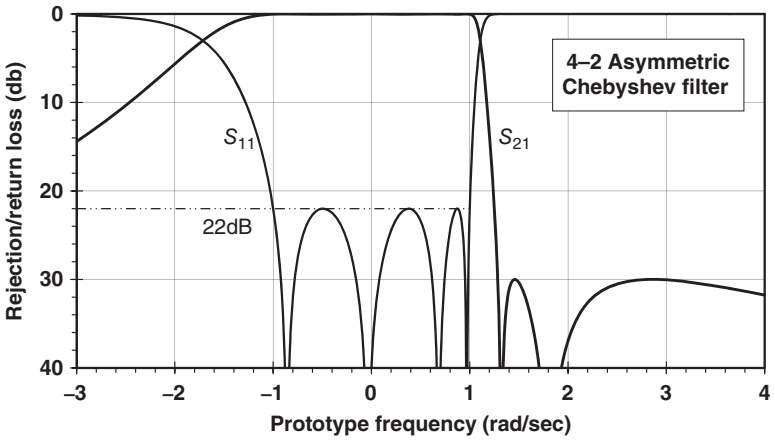


Figure 6.7 Lowpass prototype transfer and reflection characteristics of the 4–2 asymmetric Chebyshev filter with two prescribed transmission zeros at $s_1 = +j1.3217$ and $s_2 = +j1.8082$.

- Zeros of $F(s)$ (the zeros of the reflection function $S_{11}(s)$) are distributed asymmetrically along the imaginary axis. This means that the coefficients of $F(s)$ alternate between purely real and purely imaginary as the power of s increases.
- Similarly, because the zeros of $P(s)$ (the zeros of the transfer function $S_{21}(s)$) and all lie on the imaginary axis (or they can be distributed symmetrically about the imaginary axis), the coefficients of $P(s)$ also alternate between purely real and purely imaginary as the power of s increases.

By analyzing these polynomials with the frequency variable s ranging from $-j\infty$ to $+j\infty$, the asymmetric transfer and reflection characteristics are as shown in Figure 6.7.

In Chapter 7, we examine network synthesis techniques. After the polynomials of the $[ABCD]$ transfer matrix are formed from the coefficients of $E(s)$ and $F(s)$, an electrical network is synthesized from these matrix polynomials. As the network progressively builds up component by component, the polynomials are correspondingly reduced in degree one by one. This procedure is referred to as "extracting" the components from the polynomials. At the end of a successful synthesis cycle, when the electrical network has been fully synthesized, the coefficients of the polynomials are zero apart from one or two constants.

It can be demonstrated that if the coefficients of the polynomials are complex, it is necessary to extract components known as "frequency-invariant reactances" (FIRs) from the polynomials, in addition to the capacitors and inductors, in the course of building up the network and reducing the polynomial coefficients to zero. Eventually, the FIRs translate into frequency offsets from the bandpass filter's nominal center frequency ("asynchronously tuned"), giving the required asymmetric performance.

If the TZs of the filter function are prescribed symmetrically about the real axis, then the pattern of the singularities of the polynomials $E(s)$ and $F(s)$, which are generated from these TZs, is also symmetric about the real axis. Symmetry about the real and imaginary axes means that the polynomials are composed of purely real coefficients:

- $E(s)$ is an Nth-degree polynomial with real coefficients.
- $P(s)$ is an even polynomial of degree n_{fz} with real coefficients, where n_{fz} is the number of prescribed finite-position TZs ranging from zero ("all-pole" function) to N ("fully canonical").
- $F(s)$ is an Nth-degree even polynomial if N is even and odd if N is odd, again with purely real coefficients.

The purely real coefficients indicate that any FIR that is extracted during the synthesis process will be zero in value and that all the resonators of the filter are synchronously tuned at center frequency.

6.3.4 Finding the Positions of the In-Band Reflection Maxima and the Out-of-Band Transmission Maxima of the Generalized Chebyshev Prototype

During certain optimization processes, it is often useful to know where the prototype frequencies of the in-band RL maxima and out-of-band rejection lobe peaks (transmission maxima) of S_{21} are positioned. The positions of the in-band RL maxima are used during the optimization of manifold multiplexers as outlined in Chapter 18. For the design of individual filters, it is frequently the case that a certain minimum out-of-band rejection level is specified over certain bands, which may or may not be symmetric. Rejection levels may be tailored by optimizing the positions of TZs on the imaginary axis, and the positions of the minimum rejection levels that are produced between TZ positions ("rejection lobe peaks"), and the corresponding rejection levels, need to be known in order to ensure that the specification is being satisfied.

In the following section, an analytical method for computing the values of these critical frequencies for the generalized Chebyshev prototype is developed. The procedure starts with the differentiation of the generalized Chebyshev filtering function $C_N(\omega)$ as introduced earlier in this chapter in equation (6.44a) and equating it to zero:

$$\frac{d}{d\omega}[C_N(\omega)] = \frac{d}{d\omega}\cosh\left[\sum_{n=1}^{N}\cosh^{-1}x_n(\omega)\right]$$
$$= \sinh\left[\sum_{n=1}^{N}\cosh^{-1}x_n(\omega)\right]\cdot\sum_{n=1}^{N}\frac{x_n'}{\sqrt{x_n^2(\omega)-1}} = 0 \tag{6.62}$$

where $x_n(\omega) = \frac{\omega-1/\omega_n}{1-\omega/\omega_n}$ (equation (6.48)) and x_n' means $\frac{d}{d\omega}[x_n(\omega)]$.

The zeros of equation (6.62) will be the turning points of the filtering function. Rewriting equation (6.62) as the product of two expressions $T_1(\omega)$ and $T_2(\omega)$,

$$\frac{d}{d\omega}[C_N(\omega)] = T_1(\omega)\cdot T_2(\omega) = 0$$

where

$$T_1(\omega) = \sinh\left[\sum_{n=1}^{N}\cosh^{-1}x_n(\omega)\right] \tag{6.63}$$

and

$$T_2(\omega) = \sum_{n=1}^{N}\frac{x_n'}{\sqrt{x_n^2-1}} \tag{6.64}$$

Because $T_2(\omega)$ is only real for $|x_n| > 1$, its zeros represent the frequency positions of the out-of-band turning points and therefore the transmission maxima (see Figure 6.6). The frequency positions of the remaining $N-1$ in-band turning points of the Chebyshev filtering function $C_N(\omega)$ are then the zeros of $T_1(\omega)$, corresponding to the $N-1$ in-band RL maxima. These two expressions will now be used to compute the in-band reflection and out-of-band transmission maxima.

6.3.4.1 In-Band Reflection Maxima

Here it will be shown that the roots of the auxiliary polynomial $V_N(\omega)$, which has been developed in Section 6.3.2 by the recursion routine, are also the zeros of the expression $T_1(\omega)$ and are

therefore the turning points, or RL maxima, within the interval $\omega \pm 1$ of the filtering function. This can be achieved following the same procedure as for the development of $C_N(\omega)$, starting with equation (6.50) but now using the expression for $T_1(\omega)$ (equation (6.63)).

The only difference between equations (6.44a) and (6.63) is the "sinh" operator instead of the "cosh" operator. Because $\cosh(x) = {}^1\!/_2(e^x + e^{-x})$ and $\sinh(x) = {}^1\!/_2(e^x - e^{-x})$, there is the effect of introducing a minus sign into the equation sequence (6.51)–(6.61). For example, the equivalent to equation (6.52) is

$$T_1(\omega) = \frac{1}{2}\left[\prod_{n=1}^{N}(a_n + b_n) - \prod_{n=1}^{N}(a_n - b_n)\right]$$

where now it may be seen that there is a minus sign between the two product terms instead of a plus sign. This minus sign propagates through the procedure until finally the numerator of $T_1(\omega)$ is formed:

$$\begin{aligned}\text{Num}[T_1(\omega)] &= \frac{1}{2}[G_N(\omega) - G'_N(\omega)] = \frac{1}{2}((U_N(\omega) + V_N(\omega)) - (U'_N(\omega) + V'_N(\omega))) \\ &= V_N(\omega)\end{aligned} \tag{6.65}$$

Comparing with equation (6.61) it may be seen that the U_N and U'_N $(= U_N)$ polynomials now cancel instead of V_N and V'_N $(= -V_N)$. Thus the zeros of $V_N(\omega)$ will also be the zeros of $T_1(\omega)$, the turning points of the in-band RL. For the 4–2 example in Section 6.3.2, the zeros of the polynomial $V_N(\omega)$ are the positions of the $N-1$ in-band reflection maxima: $\omega_{r1} = -0.4936$, $\omega_{r2} = +0.3796$, and $\omega_{r3} = +0.8732$.

6.3.4.2 Out-of-Band Transmission Maxima

The positions of the out-of-band transmission maxima ("rejection lobes" or "attenuation minima") may be found by finding the zeros of the numerator of the polynomial $T_2(\omega)$. From equation (6.64),

$$T_2(\omega) = \sum_{n=1}^{N} \frac{x'_n}{\sqrt{x_n^2 - 1}} \tag{6.66}$$

By substituting for x_n and x'_n using

$$x_n(\omega) = \frac{\omega - 1/\omega_n}{1 - \omega/\omega_n}$$

and

$$x'_n(\omega) = \frac{dx_n(\omega)}{d\omega} = \frac{1 - 1/\omega_n^2}{(1 - \omega/\omega_n)^2}$$

$T_2(\omega)$ may be expressed in terms of the prescribed TZs ω_n (including those at infinity):

$$\begin{aligned}T_2(\omega) &= \sum_{n=1}^{N} \frac{x'_n}{\sqrt{x_n^2 - 1}} = \frac{1}{\omega'}\sum_{n=1}^{N} \frac{-\omega_n\sqrt{1 - 1/\omega_n^2}}{\omega - \omega_n} \\ &= \frac{1}{\omega'}\sum_{n=1}^{N} \frac{r_n}{\omega - \omega_n}\end{aligned} \tag{6.67}$$

where $\omega' = \sqrt{\omega^2 - 1}$, the transformed frequency variable, and $r_n = -\omega_n\sqrt{1 - 1/\omega_n^2}$, the residues of the partial fraction expansion. At the transmission maxima, $T_2(\omega) = 0$ and the ω' term disappears, and if there are n_{fz} finite-position TZs and therefore $N - n_{fz}$ zeros at infinite frequency, then the terms corresponding to $\omega_n = \infty$ in the summation will equal 1. Now the

expression for $T_2(\omega)$ may be written as

$$T_2(\omega) = (N - n_{fz}) + \sum_{n=1}^{n_{fz}} \frac{r_n}{\omega - \omega_n} = 0 \tag{6.68}$$

By using standard techniques, $T_2(\omega)$ may be reorganized from a partial fraction expansion into a quotient of two polynomials, the denominator of which will be the polynomial $P(\omega)$:

$$T_2(\omega) = (N - n_{fz}) + \sum_{n=1}^{n_{fz}} \frac{r_n}{\omega - \omega_n} = \frac{A(\omega)}{P(\omega)} = 0 \tag{6.69}$$

Rooting the numerator polynomial $A(\omega)$ will yield the positions of out-of-band transmission maxima ω_t. Some observations are as follows:

- The lobe positions corresponding to imaginary-axis TZs will occur at real frequencies, that is, ω_t is real. If there are complex zero pairs in the prototype characteristic (e.g., for group delay equalization), then the maxima ω_t will occur at complex frequencies and therefore may be easily identified and disregarded.
- If $N - n_{fz} > 0$, then $A(\omega)$ will be of degree n_{fz}, and, assuming that all the ω_t are real, there will be n_{fz} transmission maxima corresponding to the positions of the out-of-band rejection lobes.
- If $N - n_{fz} = 0$ (a fully canonical prototype), then $A(\omega)$ will be of degree $N - 1$. In this case, if all the ω_t are real, there will be $n_{fz} - 1$ rejection lobes.[7] If the fully canonical prototype is symmetric and even degree, one of the lobes will be at $\omega = \infty$.

This procedure for finding the positions of out-of-band rejection lobes may be used during optimization processes where prescribed minimum levels have been specified for the out-of-band rejections. Once the positions of the lobes have been determined, the corresponding rejection levels may be easily computed from the transfer function $S_{21}(\omega_t) = P(\omega_t)/\varepsilon E(\omega_t)$. For the 4–2 example in Section 6.3.2, application of the procedure to calculate the positions of the two out-of-band lobes results in $\omega_{t1} = +1.4587$ and $\omega_{t2} = +2.8566$.

This method has advantages in that only the positions ω_n of the n_{fz} prescribed TZs that make up the polynomial $P(\omega)$, and the degree N of the prototype, are required to determine the positions of the rejection lobes. Therefore only polynomials of up to degree n_{fz} are involved, ensuring greater accuracy. It also demonstrates that the value of the prescribed RL is not required, meaning that the same lobe positions (but not the rejection levels at those positions) will occur in a certain prototype even if the prescribed RL is changed. An approximation RL + Lobe Level (dB) = constant may be applied, that is, adding x dB to the design RL will reduce the rejection lobe levels by about x dB. This is an important consideration during filter tuning procedures, where tuning for RL levels above the values that the original prototype was designed to achieve will result in reduced out-of-band attenuation levels.

6.4 Predistorted Filter Characteristics

All the filter transfer and reflection polynomial synthesis techniques that have been considered so far have assumed that the filter networks that will eventually realize the filter response

7 In a fully canonical prototype, although there are N transmission zeros, there are only $N - 1$ rejection lobes. During an optimization process, when the levels of these lobes are being set to prescribed levels by optimizing the positions of the transmission zeros, only $N - 1$ of the zeros need to be used. This extra degree of freedom, within limits, can fix the position of one of the zeros during the optimization, which is sometimes useful if a high rejection over a narrow frequency band is required.

will comprise of purely reactive (nondissipative) components. If at a particular frequency, a two-port network has a transfer coefficient that is less than unity (i.e., $S_{21}(s) < 1$), then the power that does not transfer through to the output termination is reflected back to the input termination so that the conservation of energy condition for the network is satisfied. No power is absorbed internally by dissipative components.

In practice, the components of a filter are not perfectly lossless as the synthesis has assumed. Particularly in microwave filters, the conductivity of the material that the walls of the resonators are manufactured with (e.g., silver, copper) has a nonzero loss tangent, so some energy is absorbed within the resonator cavities, and the conservation of energy equations no longer hold true. The same holds for planar and distributed components whose critical elements are transmission lines.

Under analysis, the response of the filter with internal dissipative components will be less selective and have a more rounded appearance than the original (lossless) prototype circuit that was synthesized. Transmission and reflection zeros will be less distinct, and in bandpass filters the shoulders of the response near the band edges will be more rounded. Figure 6.8 shows the 4–2 asymmetric filter for which the lowpass prototype (LPP) polynomials were synthesized above and now transformed to a bandpass filter with a K_u-band center frequency of 12 GHz and design bandwidth 40 MHz. Firstly a lossless realization is assumed (Q_u = infinity), then with $Q_u = 10{,}000$ (typical of a dual-TE_{113}-mode cylindrical waveguide resonator at K_u-band) and then at $Q_u = 4000$ (TE_{101}-mode rectangular in rectangular waveguide). The effects of progressively worsening Q_u are clear.

To analyze the transfer and reflection responses of a filter with a finite Q_u, a positive real factor σ is added to the purely imaginary frequency variable $s = j\omega$ such that $s = \sigma + j\omega$. As described in Section 3.9, σ is calculated from

$$\sigma = \frac{f_0}{BW} \cdot \frac{1}{Q_u} \tag{6.70}$$

where Q_u is the unloaded Q-factor of the filter's resonators and f_0 and BW are, respectively, the center frequency and design bandwidth of the bandpass filter (Figure 6.9). The effect of adding σ to the frequency variable causes s to track to the right of the $j\omega$ axis as it moves from $-j\infty$ to $+j\infty$ in frequency.

The same response is obtained if all the transfer/reflection singularities, based on the lossless prototype filter case, are shifted to the right by the amount σ. By doing this, the magnitude and phase of the vectors drawn between the frequency variable position and the positions of the poles and zeros of the filter function remains unchanged. Now, we are effectively analyzing a lossless circuit, but the rightward shift of the singularities has resulted in an analyzed response equal to the lossy case.

Predistortion synthesis involves synthesizing an ideal transfer function (e.g., Chebyshev) and then moving all the poles of the filter function (the zeros of the polynomial $E(s)$) to the right by the amount σ, derived from the anticipated unloaded Q of the filter structure [6, 7]. With the poles in these new positions, a *lossless* analysis gives the predistorted response. In the presence of the finite Q_u that has been anticipated for the resonators of the real filter, the frequency variable $s = \sigma + j\omega$ tracks to the right of the imaginary axis by the amount σ. However, since the poles have also been pre-shifted to the right by this same amount, they are now in the correct relative positions in relation to s for the ideal response to be recovered. Figure 6.10 illustrates the procedure.

For filter functions with finite TZs, it is not possible to recover the ideal response with dissipative elements. In order to be able to synthesize a realizable network, there must be symmetry of the TZs about the imaginary axis; it is not possible to give them a one-way shift by the amount

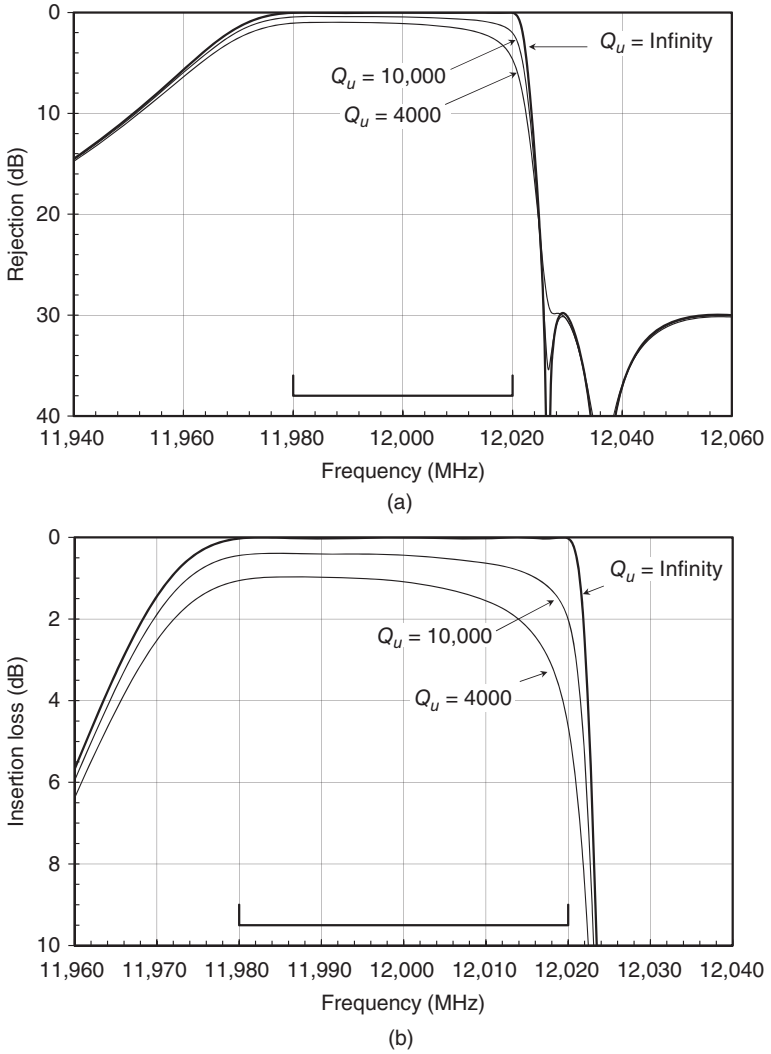


Figure 6.8 Transfer characteristics of a (4–2) bandpass filter with Q_u = infinity, 10,000, and 4000: (a) rejection, (b) in-band insertion loss.

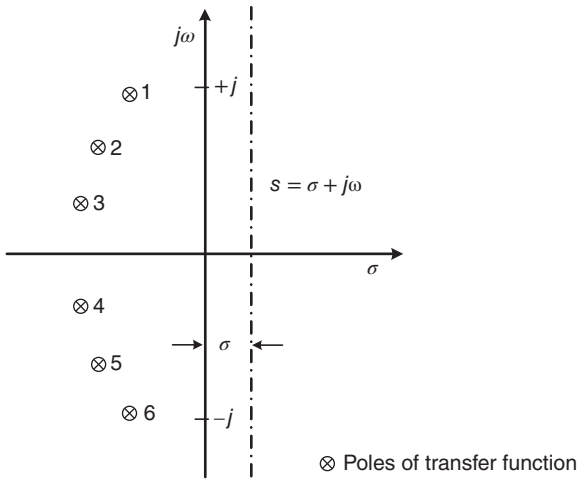


Figure 6.9 Track of the frequency variable *s* in the complex plane with a finite-value Q_u.

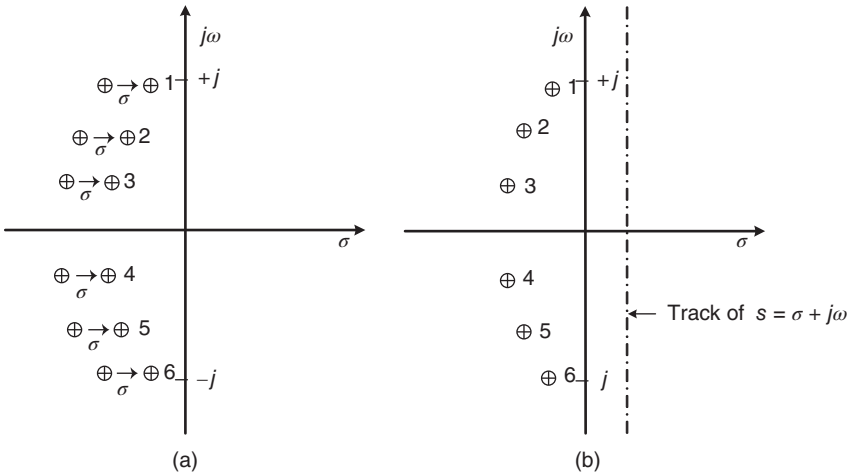


Figure 6.10 Predistortion synthesis: (a) poles shifted to right by amount σ before synthesis. (b) Poles are now in correct nominal positions relative to the track of the frequency variable with loss ($s = \sigma + j\omega$) to give equivalent lossless response under analysis.

σ. However the poles dominate the in-band response, and the deviation of the transfer characteristic $S_{21}(s)$ from the ideal response will not be significant if the TZs are left in their original positions.

So, the synthesis of a predistorted network might proceed as follows:

- Derive the $S_{21}(s)$ polynomials $E(s)$ and $P(s)$ for the ideal response.
- Compute the loss factor σ by using equation (6.70), knowing the center frequency and design bandwidth of the realized bandpass filter.
- Shift the positions of the poles by the real amount $+\sigma, p_k \rightarrow p_k + \sigma, k = 1, 2, \ldots, N$, that is, the poles are shifted to the right by an amount σ. The real part of p_k should always be negative to satisfy the Hurwitz condition.
- With the original $P(s)$ polynomial, recalculate ε. This is relatively arbitrary, provided $S_{21}(s) = (P(s)/\varepsilon)/E(s) \leq 1$ at any value of s, but is usually chosen so that $S_{21}(s) = 1$ at its maximum value.
- Recalculate the numerator polynomial $F(s)$ of the reflection function $S_{11}(s)$. This will have to be done using the conservation of energy equation (6.2) or (6.37) because in general the zeros of $S_{11}(s)$ and $S_{22}(s)$ are off-axis, and as mentioned in Section 6.3, the alternating singularity method cannot be used.

However, to fully compensate for a finite Q_u, effectively restoring it to infinity, usually incurs a rather large insertion loss penalty. In practice, it is usual to only partly compensate for the finite Q_u, for example, to compensate Q_u of 4000 to an effective Q_u (Q_{eff}) of 10,000. For partial compensation the equation for σ becomes (see equation (6.71))

$$\sigma = \frac{f_0}{BW}\left(\frac{1}{Q_{\text{act}}} - \frac{1}{Q_{\text{eff}}}\right) \tag{6.71}$$

Partial compensation tends to reduce insertion loss and improve RL without seriously degrading the in-band linearity performance from that of the ideal filter. Yu *et al.* [7] introduced a method whereby the amounts that the poles are shifted to the right are weighted, for example, in a sinusoidal fashion such that those poles closest to the imaginary axis (usually in the proximity

of the band edges) are moved less than those near the center of the band. This also helps to reduce insertion loss and improve RL (easier tuning).

Figure 6.11 demonstrates the effects on the 4–2 filter intended for realization in rectangular waveguide resonators with a Q_u of 4000. Figure 6.11a shows the response of the predistorted network analyzed with Q_u = infinity. This reveals the "cat's ear" predistortion of the characteristic. Figure 6.11b shows the filter response with the $Q_u = 4000$ that is expected from the real filter cavities. The in-band amplitude linearity is effectively improved to that of a filter with resonators of $Q_u = 10{,}000$, but with a standing insertion loss of about 5 dB. This loss is partially due to internal dissipative losses caused by finite Q_u values but mostly by reflected RF energy (RL), which is now tailored to compensate for the in-band roundings caused by the finite Q_u (Figure 6.11b).

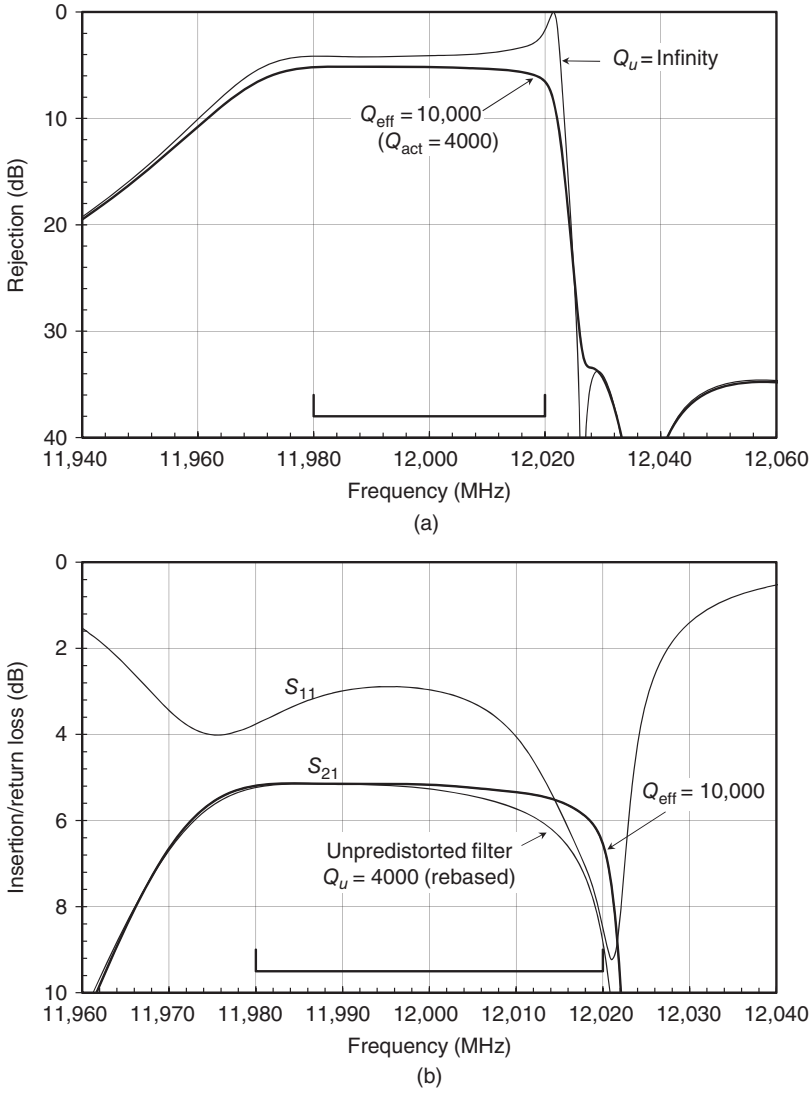


Figure 6.11 Predistorted 4–2 filter with $Q_{act} = 4000$ and $Q_{eff} = 10{,}000$. (a) Predistorted rejection characteristics with Q_u = infinity and $Q_{act} = 4000$ (effective Q of 10,000) and (b) in-band return loss and insertion loss compared with that of filter without predistortion.

The large in-band insertion loss implies that the predistortion technique can only be applied to filters in low power (input) subsystems where RF gain is readily available to compensate, and isolators can be installed at the input and output to capture the reflected energy.

Effectively improving the Q_u of a filter means that the same in-band performance can be achieved with a filter with smaller and lighter cavities than a high-Q_u equivalent. Figure 6.12 gives an example of two C-band (10–4–4) filters, one made from dielectric resonator cavities and the smaller filter with an equivalent in-band performance from smaller and less complex coaxial resonator cavities. The in-band amplitude and rejection performances of this predistorted filter are also given.

6.4.1 Synthesis of the Predistorted Filter Network

It was noted above that the zeros of reflection functions $S_{11}(s)$ and $S_{22}(s)$ (the roots of $F(s)$ and $F_{22}(s)$) of the predistorted filter will in general not be located on the imaginary axis. However, for a synthesizable network, the zeros from $F(s)$ and $F_{22}(s)$ must form mirror-image pairs about the imaginary axis to make up a symmetric overall pattern about the imaginary axis. Unlike the poles of the transfer and reflection function, there is no Hurwitz condition on the reflection zeros, and the individual zeros that make up the polynomial $F(s)$ can be arbitrarily chosen from the LHS or RHS of each pair. $F_{22}(s)$ is made up from the remaining zeros from each pair to form the complementary function to $F(s)$.

Thus, there are 2^N combinations of zeros that can be chosen to form $F(s)$ and $F_{22}(s)$, half of which are not simply exchanges of $F(s)$ and $F_{22}(s)$, that is, network reversals. Each combination leads to different values for the coefficients of $F(s)$ and, after the synthesis process, different values for the network elements.

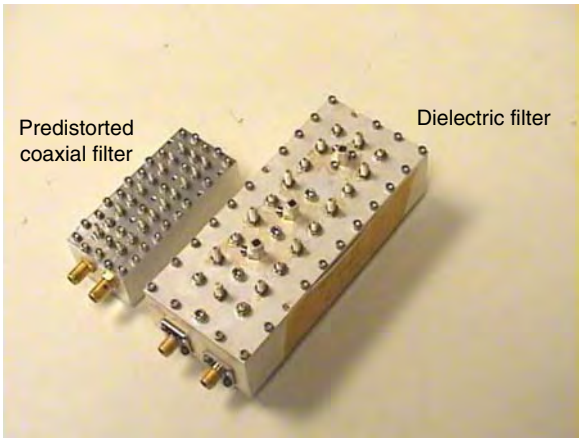
A parameter that may be used to broadly classify the solutions is μ, where

$$\mu = \left| \sum_{k=1}^{N} \mathrm{Re}(s_k) \right| \tag{6.72}$$

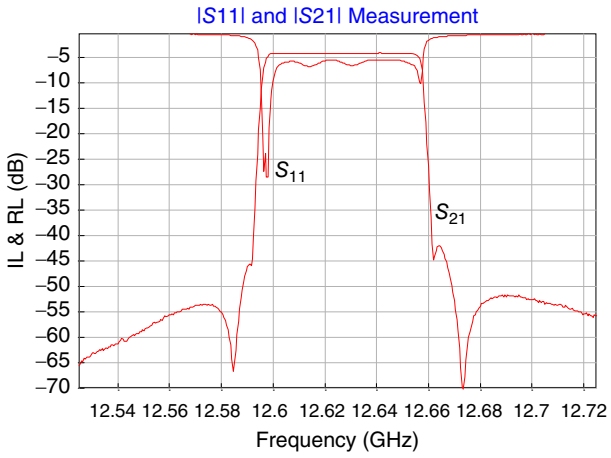
and s_k are the N zeros that have been chosen from the LHS or RHS of each pair to form $F(s)$. It can be demonstrated that if the zeros are chosen so as to minimize μ, then the values of the elements of the folded-array network are maximally symmetric about the physical center of the network; however, the network is maximally asynchronously tuned. The opposite holds true if μ is maximized, that is, all the zeros for $F(s)$ are chosen from the LHS or RHS of each mirror-image pair. Here, the network is as synchronously tuned as possible, but the network element values are maximally asymmetric about the center of the network.

Although this principle holds true for any filtering characteristic, symmetric or otherwise, the effect is most convincingly demonstrated with an even-degree symmetric characteristic. The example chosen here is a sixth-degree 23 dB Chebyshev filter. It has a pair of TZs at $\pm j1.5522$ to produce two rejection lobes on either side of the passband of 40 dB. The actual Q_u of the cavities is 4000, but predistortion has been applied to effectively raise the Q_u to 20,000.

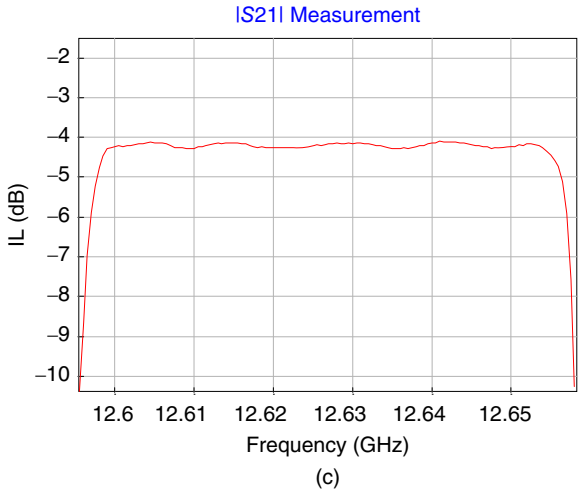
Three possible arrangements for the $S_{11}(s)$ zeros are considered, one where the value of μ is maximized and two for the case where $\mu = 0$. There is only one arrangement for $\mu = \max$ where the zeros of $F(s)$, s_k, are all in the left half (or right half) of the complex plane (Figure 6.13a). Two possible solutions for $\mu = 0$ are given in Figure 6.13b and c, where zeros s_k alternate between the left- and right-half planes as $j\omega$ increases. Another possibility occurs where the zeros with positive imaginary parts are all in the LH plane, and those with negative imaginary parts are in the RH plane. The characteristic has been normalized such that two of the zeros lie on the imaginary axis. Note that the zeros of $F_{22}(s)$ assume mirror-image positions about the imaginary axis to make up an overall symmetric pattern about the imaginary axis.



(a)



(b)



(c)

Figure 6.12 Predistorted C-band (10–4–4) coaxial filter: (a) Comparison with unpredistorted dielectric filter. (b) Rejection performance. (c) In-band insertion loss performance. (*Source*: Courtesy ComDev).

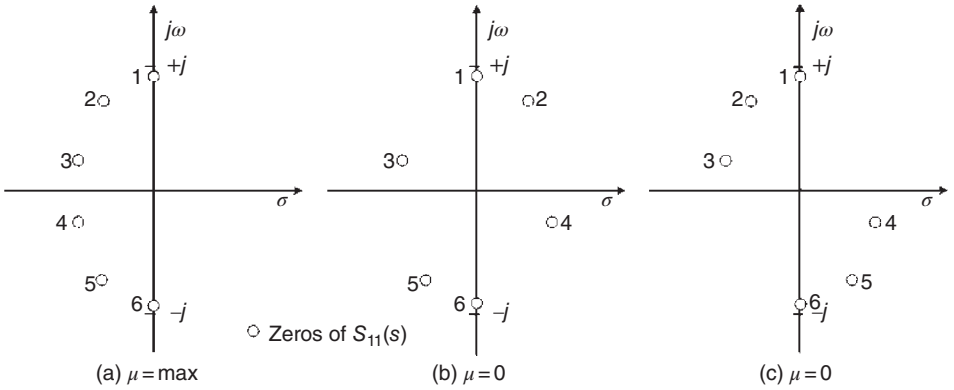


Figure 6.13 Possible arrangements for the zeros s_k of $F(s)$ for the symmetric 6–2 quasielliptic characteristic.

Let us apply the predistortion procedure to the filter function to increase its effective Q_u to 20,000. By using methods described in Chapter 8, the "$N + 2$" coupling matrices for the three cases are synthesized, yielding the values as shown in Table 6.3. It is evident that in the first case ($\mu = \max$) that the filter is synchronously tuned; that is, the diagonal entries of the coupling matrix, which represent the resonant frequency offsets at each resonator node, are all zero. However, for this case the coupling elements in the matrix are asymmetric in value about the center of the filter (i.e., asymmetric about the cross-diagonal). The two cases where $\mu = 0$ exhibit symmetry of coupling values about the cross-diagonal but are asynchronously tuned about the center of the filter.

Table 6.3 Folded Coupling Matrix Values for the Case Where $\mu = \max$ and Two Cases for $\mu = 0$.

Couplings	Case A ($\mu = \max$) (Figure 6.13a)	Case B ($\mu = 0$) (Figure 6.13b)	Case C ($\mu = 0$) (Figure 6.13c)
M_{S1}	0.3537	0.9678	0.9677
M_{12}	0.8417	0.7890	0.6457
M_{23}	0.5351	0.4567	0.7284
M_{34}	0.6671	0.7223	0.5299
M_{45}	0.6642	$=M_{23}$	$=M_{23}$
M_{56}	1.1532	$=M_{12}$	$=M_{12}$
M_{6L}	1.3221	$=M_{S1}$	$=M_{S1}$
M_{25}	−0.1207	−0.0940	−0.1403
Tuning offsets			
M_{11}	0.0	−0.0859	0.4616
M_{22}	0.0	−0.2869	0.1824
M_{33}	0.0	0.5334	−0.3541
M_{44}	0.0	$=-M_{33}$	$=-M_{33}$
M_{55}	0.0	$=-M_{22}$	$=-M_{22}$
M_{66}	0.0	$=-M_{11}$	$=-M_{11}$

As far as the electrical performance is concerned, the choice of which case to use for a practical application ($\mu = \max$ or $\mu = \min$) is arbitrary because $S_{21}(s)$ is the same for all cases. An interesting point is that for the $\mu = \min$ cases, $S_{11}(s)$ and $S_{22}(s)$ characteristics display conjugate symmetry about the band center of the filter in the presence of finite Q_u. The reason for this becomes clear when we remember that the loss is represented by an offset for $j\omega$ to the right of the imaginary axis as s tracks from $-j\infty$ to $+j\infty$. On the practical side, it may be easier to manufacture the physically symmetric $\mu = \min$ cases, for example, case B or case C in Table 6.3. Of these, case C is probably preferred because of the rather larger value for $M_{25.}$ However, in all cases, tuning is considerably more difficult than that for the regular Chebyshev filters because of the rather low value and amorphous shape of the RL characteristic. A "lossy synthesis" approach for lumped-element filters is presented in [8], which involves the inclusion of resistive elements among the reactive components. For these predistorted networks, the amplitude characteristic is shaped by absorption rather than reflection, and the network becomes easier to tune.

6.5 Transformation for Symmetric Dual-Passband Filters

Occasionally, applications arise for filters that exhibit two passbands within the confines of the passband of the original regular single-passband filter. The dual-band characteristic can be generated from a regular single-passband LPP filter characteristic by the application of a symmetric mapping formula. The mapping transforms the original equiripple behavior (if the original LP filtering characteristic is Chebyshev) of the single passband over two sub-bands. The lower sub-band ranges from the lower edge of the original passband at $s = -j$ to $s = -jx_1$, while the upper passband extends from $s = +jx_1$ to $s = +j$, where the fractional parameters $\pm jx_1$ are the prescribed frequency points within the main passband, defining the inner edges of the two sub-bands (Figure 6.14).

The effect of applying the transform is to double the degree of the filtering characteristic and the number of any TZs ($N \to 2N$, $n_{fz} \to 2n_{fz}$). Despite this, the original equiripple level and the levels of any rejection lobes (apart from the central lobe) are preserved. The dual-band characteristic is symmetric about zero frequency, even if the original LPP filtering function is asymmetric.

The dual-band characteristic is achieved by applying a symmetric frequency transform:

$$s = as'^2 + b \tag{6.73}$$

where s' is the frequency variable mapped from the prototype s-plane. Constants a and b are derived by considering the following boundary conditions:

$$\text{at } s' = \pm j, \quad s = -a + b = +j$$
$$\text{at } s' = \pm jx_1, \quad s = -ax_1^2 + b = -j$$

giving

$$a = \frac{-2j}{1 - x_1^2} \quad \text{and} \quad b = \frac{-j(1 + x_1^2)}{1 - x_1^2} \tag{6.74}$$

By substituting equation (6.74) back into equation (6.73), the transform formula is obtained:

$$s'_i = j\sqrt{\frac{1}{2}[(1 + x_1^2) - js_i(1 - x_1^2)]}$$

and

$$s'_{N+i} = s'^*_i \tag{6.75}$$

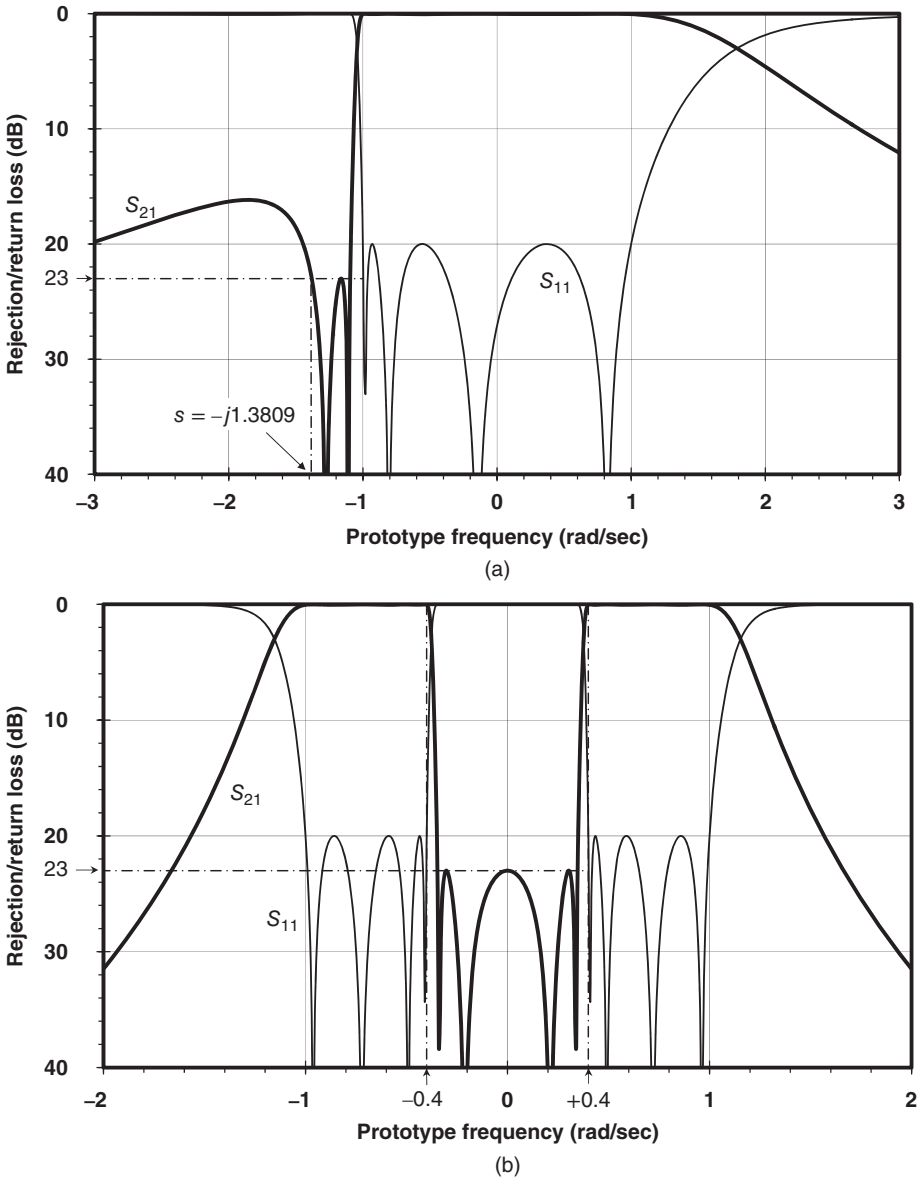


Figure 6.14 Dual-band filters: (a) (4–2) asymmetric lowpass prototype filter and (b) after transform to dual-band lowpass prototype with $x_1 = 0.4$.

for $i = 1, 2, \ldots, N$. With this transform, the N s-plane singularities of the $E(s)$, $F(s)$ polynomials and the n_{fz} singularities of the $P(s)$ polynomial of the LPP filter can be mapped to the $2N$ and $2n_{fz}$ singularities, respectively, of the dual-band characteristic. Network synthesis can then proceed as usual on the double-sized polynomials to realize the dual-band filter.

Any TZs below the main passband of the LPP will be mapped to frequencies *between* the two passbands of the dual-band characteristic. An inspection of equation (6.75) reveals that a TZ at $s = -j(1 + x_1^2)/(1 - x_1^2)$ maps to $s' = 0$, which imposes a limit to the position of the TZ on the lower side of the LPP, which can map into a pair of TZs in the dual-band characteristic. For

Table 6.4 Singularities of the Dual-Passband Transfer and Reflection Polynomials: (a) (4–2) Lowpass Prototype and (b) After Transform to (8–4) Dual-Band Characteristic with $x_1 = 0.4$.

	Reflection Zeros	Transmission Zeros	Transmission/Reflection Poles
	(a) 4–2 Asymmetric lowpass prototype—20 dB RL		
1	$-j0.9849$	$-j1.2743$	$-0.0405 - j1.0419$
2	$-j0.8056$	$-j1.1090$	$-0.2504 - j1.0228$
3	$-j0.1453$	$j\infty$	$-1.0858 - j0.4581$
4	$+j0.8218$	$j\infty$	$-0.8759 + j1.4087$
	$\varepsilon_R = 1.0$	$\varepsilon = 0.6590$	
	(b) 8–4 dual-band singularities		
	$\pm j0.9619$	$\pm j0.3380$	$-0.1679 \pm j1.0954$
	$\pm j0.7204$	$\pm j0.2117$	$-0.3247 \pm j0.7022$
	$\pm j0.4916$	$\pm j\infty$	$-0.1287 \pm j0.4086$
	$\pm j0.4078$	$\pm j\infty$	$-0.0225 \pm j0.3780$
	$\varepsilon_R = 1.0$	$\varepsilon = 3.7360$	

example, where $x_1 = 0.4$, $s_{\max} = -j1.3809$ and any TZ with a position more negative than this in the LPP will map into a pair of real-axis zeros in the dual-band response.

Figure 6.14b shows an eighth-degree dual-band filter with four TZs and $x_1 = \pm 0.4$. They are formed by the transformation of a (4–2) asymmetric filter with 20 dB RL level and optimized such that the rejection lobe level nearest to the passband and the point at $s = -j$ 1.3809 (which transforms to $s' = 0$) are both 23 dB (Figure 6.14a). After transformation, an equiripple rejection level of 23 dB in the interband region is obtained.

Table 6.4 gives the s-plane locations of the original (4–2) asymmetric filtering function, together with their transformed positions for the (8–4) symmetric dual-band filter. Figure 6.15 gives the coupling matrix (shown here in a folded-array configuration) synthesized from the transformed singularities by employing the methods described in Chapters 7 and 8. Note that the dual-band coupling matrix is synchronously tuned, even though the original prototype is asymmetric. Further similarity transforms (rotations) can be applied to the matrix to transform it to another configuration, for example, cascade quartet.

Further more advanced methods for the generation of multiband filters are presented in Chapter 21.

	S	1	2	3	4	5	6	7	8	*L*
S	0	0.8024	0	0	0	0	0	0	0	0
1	0.8024	0	0.8467	0	0	0	0	0	0	0
2	0	0.8467	0	0.4142	0	0	0	−0.2900	0	0
3	0	0	0.4142	0	0.2754	0	−0.4170		0	0
4	0	0	0	0.2754	0	0.2832		0	0	0
5	0	0	0	0	0.2832	0	0.2754	0	0	0
6	0	0	0	−0.4170	0	0.2754	0	0.4142	0	0
7	0	0	−0.2900	0	0	0	0.4142	0	0.8467	0
8	0	0	0	0	0	0	0	0.8467	0	0.8024
L	0	0	0	0	0	0	0	0	0.8024	0

Figure 6.15 Coupling matrix for (8–4) dual-band filtering characteristic (folded-array configuration).

Summary

In Chapter 3, we describe the classical filters, namely, the Butterworth, Chebyshev, and elliptic function filters. In Chapter 4, we describe a procedure for generating arbitrary filter functions using computer-aided design techniques. Although this technique is quite efficient, it does require use of optimization to generate the desired filter function. In this chapter, we describe an important class of equiripple filter function that can be derived by making use of a quasi-analytic technique.

The chapter begins with a review of some of the important scattering parameter relationships (particularly the unitary conditions) that are relevant to the synthesis procedures introduced in later chapters. This is followed by a discussion of the general kind of Chebyshev polynomial and its application in filter design. The classical Chebyshev filter, described in Chapter 3, is derived by making use of the Chebyshev polynomial of the first kind. Some books and many articles use the expression "Chebyshev polynomial" to refer exclusively to the Chebyshev polynomial $T_n(x)$ of the first kind. Filters, obtained by such polynomials, have their TZs located at infinity. The general Chebyshev polynomial described in this chapter can be utilized in generating filter functions with the maximum number of equiripple amplitude peaks in the passband for an arbitrary distribution of the zeros of the transfer function. To ensure physical realizability, the zeros of the transfer function must be located symmetrically about the imaginary axis. This constraint is related to satisfying the unitary condition. The derivation of this condition, along with a recursion technique to generate the transfer and reflection polynomials of this class of filters, is described. The procedure is completely general and is applicable for realizing symmetrical as well as asymmetrical response filters. Examples are included to illustrate the design procedure.

The final part of the chapter outlines the benefits of incorporating predistortion in the filter design for certain applications. The chapter concludes with a design example of a predistorted filter and a dual-band filter.

References

1 Collin, R.E. (1966) *Foundations for Microwave Engineering*, McGraw-Hill, New York.

2 Helszajn, J. (1990) *Synthesis of Lumped Element, Distributed and Planar Filters*, McGraw-Hill Book Company, United Kingdom.

3 J. D. Rhodes and S. A. Alseyab, "The generalized Chebyshev low-pass prototype filter," *IEEE Transactions on Circuit Theory*, vol. **8**, pp. 113-125, 1980.

4 R. J. Cameron, "Fast generation of Chebyshev filter prototypes with asymmetrically prescribed transmission zeros," *ESA Journal,* vol. **6**, pp. 83-95, 1982.

5 R. J. Cameron, "General coupling matrix synthesis methods for Chebyshev filtering functions," *IEEE Transactions on Microwave Theory and Techniques*, vol. **MTT-47**, pp. 433-442, 1999.

6 A. E. Williams, W. G. Bush and R. R. Bonetti, "Predistortion technique for multicoupled resonator filters methods for Chebyshev filtering functions," *IEEE Transactions on Microwave Theory and Techniques*, vol. **33**, pp. 402-407, 1985.

7 M. Yu, W.-C. Tang, A. Malarky, V. Dokas, R. J. Cameron and Y. Wang, "Predistortion technique for cross-coupled filters and its application to satellite communication systems," *IEEE Transactions and Microwave Theory and Techniques*, vol. **51**, pp. 2505-2515, 2003.

8 Hunter, I.C. (2001) *Theory and Design of Microwave Filters*, Electromagnetic Waves Series 48, IEE, London.

Appendix 6A

Complex Terminating Impedances in Multiport Networks

The foregoing discussions on S-parameters and their synthesis and analysis has assumed that (1) the network is a two-port with (2) unity source and load impedances. The analysis may be generalized to multiport networks with complex port impedances following the work of Kurokawa [A1]. The process starts with the classic power wave equations:

$$a_i = \frac{V_i + Z_i I_i}{2\sqrt{|\mathrm{Re}Z_i|}}, \quad b_i = \frac{V_i - Z_i^* I_i}{2\sqrt{|\mathrm{Re}Z_i|}} \tag{6A.1}$$

where a_i and b_i are the incident and reflected power waves at the ith port of an N-port network, V_i and I_i are the voltage at and the current flowing into the ith port of the network, and Z_i is the impedance of the circuit connected to the ith port.

Writing equations (6A.1) in matrix form for an N-port network,

$$\mathbf{a} = \mathbf{F}(\mathbf{v} + \mathbf{G}\mathbf{i}), \quad \mathbf{b} = \mathbf{F}(\mathbf{v} - \mathbf{G}^*\mathbf{i}) \tag{6A.2}$$

where **a**, **b**, **v**, and **i** are column vectors whose ith components are a_i, b_i, V_i, and I_i, respectively. **F** and **G** are $N \times N$ diagonal matrices whose ith diagonal elements are given by $f_i = 1/(2\sqrt{|\mathrm{Re}Z_i|})$ and $g_i = Z_i$, respectively, and Z_i are the complex impedances at the ports of the N-port network. Here * represents the complex conjugate transposed matrix.

The voltages and currents at each port are linearly related by the $N \times N$ impedance matrix **Z**, giving $\mathbf{v} = \mathbf{Z}\mathbf{i}$. Also **a** and **b** are related by the power wave scattering matrix **S** as $\mathbf{b} = \mathbf{S}\mathbf{a}$. Using these relations to eliminate **a**, **b,** and **v** from equation (6A.2),

$$\mathbf{F}(\mathbf{Z} - \mathbf{G}^*)\mathbf{i} = \mathbf{S}\mathbf{F}(\mathbf{Z} + \mathbf{G})\mathbf{i} \tag{6A.3}$$

from which the following expression for **S** may be obtained:

$$\mathbf{S} = \mathbf{F}(\mathbf{Z} - \mathbf{G}^*)(\mathbf{Z} + \mathbf{G})^{-1}\mathbf{F}^{-1} \tag{6A.4}$$

Equation (6A.4) may be rearranged to express the N-port admittance matrix $\mathbf{Y}$ ($= \mathbf{Z}^{-1}$) in terms of **S, F**, **G,** and **I** (the unit matrix):

$$\mathbf{Y} = \mathbf{F}^{-1}(\mathbf{S}\mathbf{G} + \mathbf{G}^*)^{-1}(\mathbf{I} - \mathbf{S})\mathbf{F} \tag{6A.5}$$

For a two-port network, these matrices may be represented as follows:

$$\mathbf{Y} = \begin{bmatrix} y_{11}(s) & y_{12}(s) \\ y_{21}(s) & y_{22}(s) \end{bmatrix} = \frac{1}{y_d}\begin{bmatrix} y_{11n}(s) & y_{12n}(s) \\ y_{21n}(s) & y_{22n}(s) \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} S_{11}(s) & S_{12}(s) \\ S_{21}(s) & S_{22}(s) \end{bmatrix}, \quad \mathbf{F} = \frac{1}{2}\begin{bmatrix} 1/\sqrt{|\mathrm{Re}Z_S|} & 0 \\ 0 & 1/\sqrt{|\mathrm{Re}Z_L|} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} Z_S & 0 \\ 0 & Z_L \end{bmatrix} \tag{6A.6}$$

where Z_S and Z_L are the source and load impedances, respectively; $y_{11n}(s)$, $y_{12n}(s)$ $(= y_{21n}(s))$, $y_{22n}(s)$ are the numerator polynomials of $y_{11}(s)$, $y_{12}(s)$ $(= y_{21}(s))$, and $y_{22}(s)$, respectively; and $y_d(s)$ is their common denominator.

In a later paper, Ge *et al.* [A2] showed how the elements of the two-port S-parameter matrix **S** in equation (6A.5) could be replaced by their rational polynomial equivalents:

$$S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)}, \quad S_{22}(s) = \frac{F_{22}(s)/\varepsilon_R}{E(s)}, \quad S_{12}(s) = S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)} \tag{6A.7}$$

and enabling, through equation (6A.6), the expression of $y_{11n}(s)$, $y_{12n}(s)$ $(= y_{21n}(s))$, $y_{22n}(s)$, and $y_d(s)$ in terms of the known termination impedances and coefficients of the $E(s)$, $F(s)$, and

$P(s)$ polynomials:

$$y_{11n}(s) = [Z_L^*(E(s) - F(s)/\varepsilon_R) - (-1)^N Z_L(E(s) - F(s)/\varepsilon_R)^*]$$
$$y_{22n}(s) = [(Z_S^*E(s) + Z_S F(s)/\varepsilon_R) - (-1)^N (Z_S^*E(s) + Z_S F(s)/\varepsilon_R)^*]$$
$$y_{12n}(s) = y_{21n}(s) = -2\sqrt{\mathrm{Re}(Z_S)\mathrm{Re}(Z_L)}P(s)/\varepsilon$$
$$y_d(s) = [Z_L^*(Z_S^*E(s) + Z_S F(s)/\varepsilon_R) + (-1)^N Z_L(Z_S^*E(s) + Z_S F(s)/\varepsilon_R)^*] \quad \text{(6A.8)}$$

These admittance parameters may then be used to begin the coupling matrix synthesis process, including complex termination impedances, as will be described later in Section 8.4.4. In Section 6.1.2 a derivation of the [*ABCD*] matrix polynomials for two-port networks with unit source and load impedances is presented, with a footnote noting the modifications required for complex terminations.

Two further useful equations may also be derived from reference [A1] for two-port networks with complex terminations; the impedances looking into the network from the driving point and load ports, Z_{in1} and Z_{in2} respectively:

$$\mathrm{Z}_{\text{in1}} = (Z_S^* + S_{11}Z_S)/(1 - S_{11}), \quad \mathrm{Z}_{\text{in2}} = (Z_L^* + S_{22}Z_L)/(1 - S_{22}) \quad \text{(6A.9)}$$

6A.1 Change of Termination Impedance

For an N—port network, the transfer coefficients S_{ij} and reflection coefficients S_{ii} at the ports are dependent upon the impedances that the ports are normalized to. In the great majority of cases, the terminating impedances are all unity, but for certain multiplexer problems it is useful to renormalize a set of S-parameters defined at one particular set of terminating impedances to another set of impedances, possibly including complex terminations. To convert the S-parameter matrix $\mathbf{S}$ defined at one set of termination impedances Z_i to a new S-parameter matrix $\mathbf{S}'$ defined with a new set of termination impedances Z_i', the following matrix equation may be used [A1,A3]:

$$\mathbf{S}' = \mathbf{A}^{-1}(\mathbf{S} - \mathbf{\Gamma}^*)(\mathbf{I} - \mathbf{\Gamma}\mathbf{S})^{-1}\mathbf{A}^* \quad \text{(6A.10)}$$

where $A_{ii} = \frac{(1-r_i^*)}{|1-r_i|}\sqrt{|1 - r_i r_i^*|}$ is the ith element in the diagonal matrix $\mathbf{A}$, $r_i = \frac{Z_i' - Z_i}{Z_i' + Z_i^*}$, and $\mathbf{\Gamma}$ is a diagonal matrix with $\Gamma_{ii} = r_i$. Z_i are the original impedances at the ith port of the network, Z_i' are the new impedances, and $\mathbf{A}^*, \mathbf{\Gamma}^*$ are the complex conjugate transposes of the diagonal matrices $\mathbf{A}, \mathbf{\Gamma}$, respectively.

Frickey [A4] gives conversions between various circuit parameters for two-port networks.

References

A1 Kurokawa, K. (1965) Power waves and the scattering matrix. *IEEE Transactions on Microwave Theory and Techniques*, 194–202, March 1965.

A2 Ge, C., Zhu, X.-W., Jiang, X., and Xu, X.-J. (2016) A general synthesis approach of coupling matrix with arbitrary reference impedances. *IEEE Microwave and Wireless Components Letters*, **25** (6), 349–351.

A3 Bodway, G.E. (1968) Circuit design and characterization of transistors by means of three-port scattering parameters. *Microwave Journal*, **11** (5), 7-1–7-11.

A4 Frickey, D.A. (1994) Conversions between S, Z, Y, H, ABCD, and T parameters which are valid for complex source and load impedances. *IEEE Transactions on Microwave Theory and Techniques*, **42** (2), 205–211.

7

Synthesis of Network-Circuit Approach

In the previous chapters, methods were established to derive the transfer and reflection polynomials for a broad class of lowpass prototype filter functions. The next step in the design process is to translate these polynomials into a prototype electrical circuit from which a real microwave filter may be developed. Two methods for doing this are available: the classical circuit synthesis method and the direct coupling matrix synthesis approach. In this chapter, the circuit synthesis approach, based on the [*ABCD*] transfer matrix, or the "chain" matrix, as it is sometimes known, is described.

There are many excellent and comprehensive contributions to the subject of circuit synthesis techniques in the literature [1–6]. It is not the intention to repeat the works here, but rather to utilize the theories to develop a general microwave filter synthesis technique. The techniques described here include both the symmetric and asymmetric lowpass prototype filters. As described in Sections 3.10 and 6.1, such prototype networks require the hypothetical frequency-invariant reactance elements, referred to by the acronym FIR [7]. Inclusion of such elements leads to transfer and reflection polynomials to have complex coefficients. In bandpass or bandstop filters, FIR elements appear as frequency offsets to the resonant circuits. The various components used in the microwave synthesis process include the following:

- Frequency-dependent reactive elements: lumped capacitors and inductors. Their number in the lowpass prototype network determines the "degree" or "order" of the filter network. Capacitors and inductors in a ladder network are interchangeable through the dual-network theorem.
- FIR[1] elements. The classical network synthesis theory is founded on the concept of the *positive real* condition for the driving-point immittance function $Z(s)$ or $Y(s)$, that is, for a realizable network, $\mathrm{Re}(Z(s))$ or $\mathrm{Re}(Y(s))$, when $\mathrm{Re}(s) > 0$ and $Z(s)$ or $Y(s)$ is real for s real [8]. With the introduction of FIRs into the network, the second condition (which implies symmetry of the frequency response) is no longer necessary and the driving-point function becomes a *positive* function [9]. The FIR represents an offset in the resonant frequency of a resonator from nominal (*nominal* here means zero for a lowpass resonator and f_0 for a bandpass resonator), as described in Figure 7.1. For the asymmetric class of filters, the driving-point immittance polynomials possess coefficients that cannot be extracted as capacitors or inductors. However, these coefficients can be extracted as FIRs that are realized eventually by a microwave resonant element (a waveguide cavity or a dielectric resonator) that is tuned to be offset from the *nominal* center frequency of the filter; thus, the filter is *asynchronously* tuned.
- Frequency-independent transmission-line phase lengths, particularly the quarter-wave (90°) impedance or admittance inverters (collectively known as *immittance inverters*).

1 In common usage, the acronym "FIR" is used for both frequency-independent reactances *and* susceptances.

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

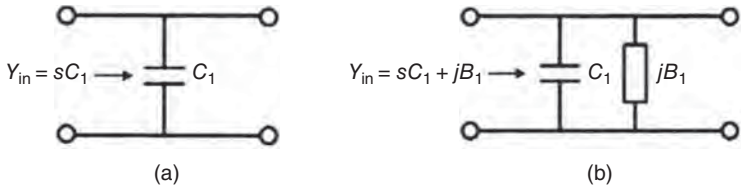


Figure 7.1 (a) $Y_{in} = 0$ when $s = s_0 = 0$; therefore, this lowpass resonator resonates at zero frequency; (b) $Y_{in} = sC_1 + jB_1 = 0$ when $s = s_0$. Thus the resonant frequency $s_0 = j\omega_0 = -jB_1/C_1$ is offset by $\omega_0 = -B_1/C_1$ from the nominal center frequency.

In a microwave circuit, these elements act as transformers with a frequency-invariant 90° phase shift and are approximately realized by a variety of microwave structures such as inductive irises and coupling probes. The use of inverters significantly simplifies the physical realization of filters at microwave frequencies.

- Coupling inverters act as coupling elements between the resonant nodes of a network, representing a microwave filter. Those that couple directly between sequentially numbered resonators are known as *mainline couplings*; those between nonsequential resonators, cross-couplings; and those between the source port and the load port and internal resonators of the filter, *input/output couplings*. There can be more than one coupling from the source or load ports to the internal resonators nodes in the filter, and there is a direct source–load coupling for canonical filter functions.

7.1 Circuit Synthesis Approach

For a two-port network operating between unity source and load terminations, the [*ABCD*] matrix representing the network has the following form [10]:

$$[ABCD] = \frac{1}{jP(s)/\varepsilon} \cdot \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} \tag{7.1a}$$

where

$$S_{12}(s) = S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)} = \frac{2P(s)/\varepsilon}{A(s) + B(s) + C(s) + D(s)} \tag{7.1b}$$

$$S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)} = \frac{A(s) + B(s) - C(s) - D(s)}{A(s) + B(s) + C(s) + D(s)} \tag{7.1c}$$

$$S_{22}(s) = \frac{(-1)^N F(s)^*/\varepsilon_R}{E(s)} = \frac{D(s) + B(s) - C(s) - A(s)}{A(s) + B(s) + C(s) + D(s)} \tag{7.1d}$$

The j that multiplies $P(s)$ in the denominator of equation (7.1) allows the cross-couplings to be extracted as inverters and, in addition to the j that multiplies $P(s)$ when $N - n_{fz}$ is even, preserves the orthogonality unitary condition (see Section 6.2). For nonunity source and load terminations R_S and R_L, the [*ABCD*] matrix is scaled as follows:

$$[ABCD] = \frac{1}{jP(s)/\varepsilon} \cdot \begin{bmatrix} \sqrt{\frac{R_L}{R_S}}A(s) & \frac{B(s)}{\sqrt{R_S R_L}} \\ \sqrt{R_S R_L}C(s) & \sqrt{\frac{R_S}{R_L}}D(s) \end{bmatrix} \tag{7.2}$$

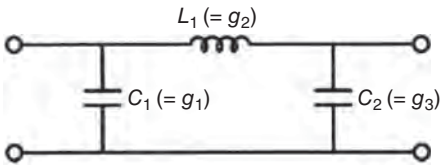


Figure 7.2 Basic lowpass prototype circuit (third-degree).

From equation (7.1), it is immediately clear that the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials share a common denominator polynomial $P(s)/\varepsilon$. It is demonstrated that these polynomials can be built up from the elements of the circuit that models the real microwave filter and then related to the coefficients of the $E(s)$ and $F(s)/\varepsilon_R$ polynomials, which, together with $P(s)/\varepsilon$, represent the desired transfer and reflection performance of the filter, respectively.

Our first task, therefore, is to relate the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials to the S parameters that represent the transfer and reflection characteristics of the filter function. This is accomplished by building the [*ABCD*] matrix for a simple three-element ladder network, relating it to a real filter circuit and examining the form of the resulting $A(s)$, $B(s)$, $C(s)$, $D(s)$, and $P(s)$ polynomials. These polynomials are then compared directly with the coefficients of the polynomials that constitute $S_{21}(s)$ and $S_{11}(s)$. We also study the form of the [*ABCD*] polynomials for the more advanced circuits, namely, the asymmetric, cross-coupled, and single-ended types, in order to generalize the procedure.

Figure 7.2 shows a simple lossless three-element prototype ladder circuit that is capable of realizing a lowpass filter. The three transmission zeros (TZs) for this circuit all lie at $s = j\infty$, which means that it is an all-pole filter. The network synthesis procedure is demonstrated by first building the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials, corresponding to a third-degree lowpass prototype ladder network with the elements C_1, L_1, and C_2 (sometimes known as the g values $g_1, g_2, g_3, \ldots$ of the prototype ladder network [11]). Then, the synthesis procedure is applied to demonstrate the recovery of the values of the original network elements C_1, L_1, and C_2 from the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials.

7.1.1 Buildup of the [*ABCD*] Matrix for the Third-Degree Network

Figure 7.2 represents a third-degree lowpass network.

Step A. Cascade Shunt C_1 with Series L_1

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ sC_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & sL_1 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & sL_1 \\ sC_1 & 1 + s^2 L_1 C_1 \end{bmatrix}$$

Second-degree network ($N = 2$)
Degree of polynomials:
$A(s) : N - 2$
$B(s) : N - 1$
$C(s) : N - 1$
$D(s) : N$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & sL_1 \\ sC_1 & 1 + s^2 C_1 L_1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ sC_2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 + s^2 L_1 C_2 & sL_1 \\ s(C_1 + C_2) + s^3 C_1 C_2 L_1 & 1 + s^2 C_1 L_1 \end{bmatrix}$$

Third-degree network ($N = 3$)
Degree of polynomials:
$A(s) : N - 1$
$B(s) : N - 2$
$C(s) : N$
$D(s) : N - 1$

Note the form of the polynomials $A(s)$, $B(s)$, $C(s)$, and $D(s)$. For a lowpass prototype network of degree N, where N is even, $B(s)$ and $C(s)$ are of degree $N-1$, $D(s)$ is of degree N, and $A(s)$ is of degree $N-2$. For N odd, $A(s)$ and $D(s)$ are of degree $N-1$, $C(s)$ is of degree N, and $B(s)$ is of degree $N-2$. Thus, $A(s)$ and $D(s)$ are always even, while $B(s)$ and $C(s)$ are always odd.

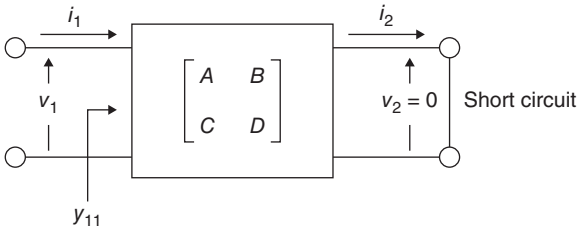
7.1.2 Network Synthesis

Now we reverse the process. The coefficients of the polynomials $A(s)$, $B(s)$, $C(s)$, and $D(s)$ are known, and we need to find the values of C_1, L_1, and C_2 and extract them from the polynomials in sequence.

The values of the components nearest to the ports of the network are found through the network's short-circuit admittance (y) parameters or the open-circuit (z) parameters and evaluating these parameters at $s = j\infty$. Knowing the value of the component, we can then extract it by premultiplying the $[ABCD]$ matrix with its inverse, leaving a unit matrix in cascade with a remainder matrix.

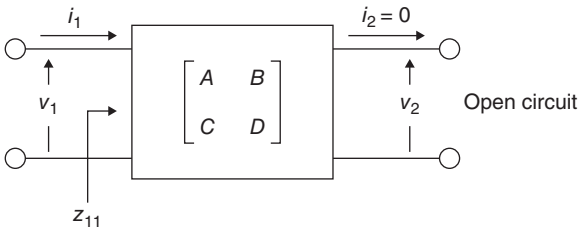
Step A.1. Evaluate C_1 We can find C_1 from the short-circuit admittance parameter y_{11} or the open-circuit impedance parameter z_{11}. Starting at the left-hand side (LHS) of the network, we first find y_{11} or z_{11} and evaluate C_1 at $s = j\infty$. First, we find y_{11} and z_{11}. For the general $[ABCD]$ matrix, we obtain

$$v_1 = A(s)v_2 + B(s)i_2$$
$$i_1 = C(s)v_2 + D(s)i_2$$



When

$$v_2 = 0, \quad y_{11} = \frac{i_1}{v_1} = \frac{D(s)}{B(s)}$$
$$i_2 = 0, \quad z_{11} = \frac{v_1}{i_1} = \frac{A(s)}{C(s)}$$



For our third-degree ladder network, the overall $[ABCD]$ matrix is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 + s^2 L_1 C_2 & sL_1 \\ s(C_1 + C_2) + s^3 C_1 C_2 L_1 & 1 + s^2 C_1 L_1 \end{bmatrix}$$

$$z_{11} = \frac{A(s)}{C(s)} = \frac{1 + s^2 L_1 C_2}{s(C_1 + C_2) + s^3 C_1 C_2 L_1}$$

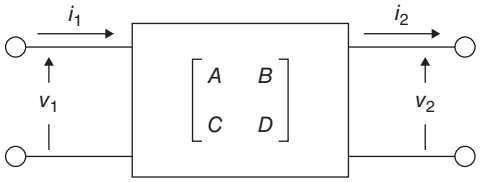
$$sz_{11}|_{s\to\infty} = \left.\frac{sA(s)}{C(s)}\right|_{s\to\infty} = \frac{1}{C_1}$$

Alternatively

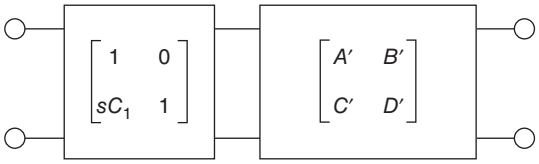
$$y_{11} = \frac{D(s)}{B(s)} = \frac{1 + s^2 L_1 C_1}{sL_1}$$

$$\left.\frac{y_{11}}{s}\right|_{s\to\infty} = \left|\frac{D(s)}{sB(s)}\right._{s\to\infty} = C_1$$

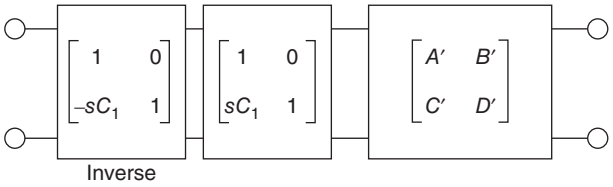
Step A.2. Extract C_1 The element extraction procedure is as follows. The $[ABCD]$ matrix



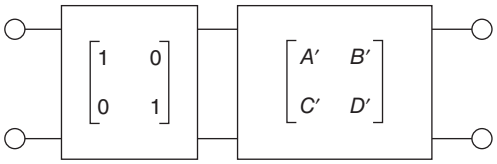
consists of a matrix for shunt C_1 in cascade with the remainder $[ABCD]$ matrix.



To extract C_1, premultiply by the inverse of the C_1 matrix.



This leaves the remainder matrix in cascade with a unitary matrix, which can be ignored.



The overall $[ABCD]$ matrix is given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 + s^2 L_1 C_2 & sL_1 \\ s(C_1 + C_2) + s^3 C_1 C_2 L_1 & 1 + s^2 C_1 L_1 \end{bmatrix} \tag{7.3}$$

Premultiply the matrix by the inverse matrix of shunt C_1 as follows:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -sC_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 + s^2 L_1 C_2 & sL_1 \\ s(C_1 + C_2) + s^3 C_1 C_2 L_1 & 1 + s^2 C_1 L_1 \end{bmatrix} = \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix}$$

$$\therefore A'(s) = A(s)$$

$$B'(s) = B(s)$$

$$C'(s) = C(s) - sC_1A(s)$$
$$D'(s) = D(s) - sC_1B(s)$$

The remainder matrix then becomes

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \begin{bmatrix} 1 + s^2L_1C_2 & sL_1 \\ sC_2 & 1 \end{bmatrix} \tag{7.4}$$

Step B. Evaluate the Series Element L_1 Using the Remainder Matrix from Step A.2.

$$z_{11} = \frac{A(s)}{C(s)} = \frac{1 + s^2L_1C_2}{sC_2}$$
$$\therefore L_1 = \left.\frac{z_{11}}{s}\right|_{s\to\infty} = \left.\frac{A(s)}{sC(s)}\right|_{s\to\infty} \tag{7.5}$$

Alternatively, we obtain

$$y_{11} = \frac{D(s)}{B(s)} = \frac{1}{sL_1}$$
$$\therefore \frac{1}{L_1} = \left.\frac{sD(s)}{B(s)}\right|_{s\to\infty} \tag{7.6}$$

Step B.1. Extract L_1 Premultiply the remainder matrix from the previous extraction (Step A.2), by the inverse of matrix of series L_1

$$\begin{bmatrix} 1 & -sL_1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 + s^2L_1C_2 & sL_1 \\ sC_2 & 1 \end{bmatrix} = \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} \tag{7.7}$$

$$\therefore A'(s) = A(s) - sL_1C(s)$$
$$B'(s) = B(s) - sL_1D(s)$$
$$C'(s) = C(s)$$
$$D'(s) = D(s)$$

The remainder matrix then becomes

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ sC_2 & 1 \end{bmatrix} \tag{7.8}$$

which is the matrix of C_2 in shunt. Premultiplication by the inverse matrix of C_2 leaves the unit matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Now we know the values of C_1, L_1, and C_2, and so the synthesis of the network is complete.

At each stage of the synthesis procedure, two values for each of the elements C_1, L_1, and C_2 are derived: one from the $C(s)$ and $A(s)$ polynomials and the other from the $B(s)$ and $D(s)$ polynomials. Nominally, they should be the same value, but with higher-degree networks, cumulative errors can cause the values to diverge quite significantly. Various techniques exist to mitigate the errors, such as taking advantage of any symmetry in the network or by using transformed variable (Z-plane) synthesis techniques [9, 12, 13]. With the numerical accuracies available from 32-bit computers, it is possible to attain six decimal places of accuracy for the values of components for moderate degrees of network, say, up to 12th degree, without using special techniques. However, it is important to use *only* the $A(s)$ and $C(s)$ polynomials or *only* the $B(s)$ and $D(s)$ polynomials to evaluate the series of components in the ladder network. Sometimes it is informative to evaluate the series first with the $A(s)$ and $C(s)$ polynomials and then with the $B(s)$ and $D(s)$ polynomials. When the values of the components, determined by the two methods, begin to

diverge in the sixth decimal place, then it will be known that the cumulating errors are starting to become significant.

7.2 Lowpass Prototype Circuits for Coupled-Resonator Microwave Bandpass Filters

The doubly terminated lowpass prototype circuits, connected to their terminating impedances or admittances [11], are shown in Figure 7.3. The element values in the network are termed the "g parameters" [11], which may be the value of a shunt capacitor or a series inductor. The source termination g_0 is resistive if g_1 is a capacitor, and conductive if g_1 is an inductor, and similarly for the load termination g_{N+1}. This gives an alternating impedance–admittance–impedance (or vice versa) reading left to right along the ladder network. The circuits are formulated in this way so that identical responses are obtained from whichever form is used.

At this stage, the dual-network theorem is applied to add unit inverters to the network to transform the series components (inductors) to shunt capacitors. Figure 7.4 shows the series

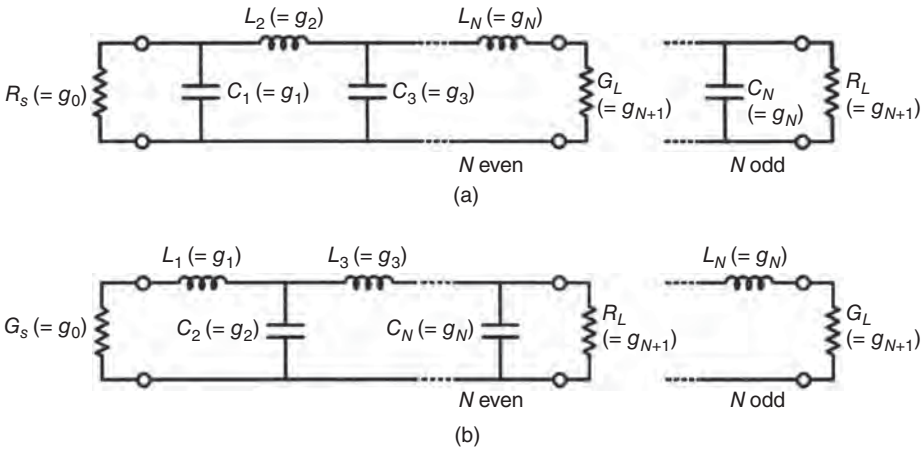


Figure 7.3 Lowpass prototype ladder networks: (a) where the leading component is a shunt capacitor; (b) the dual of the network.

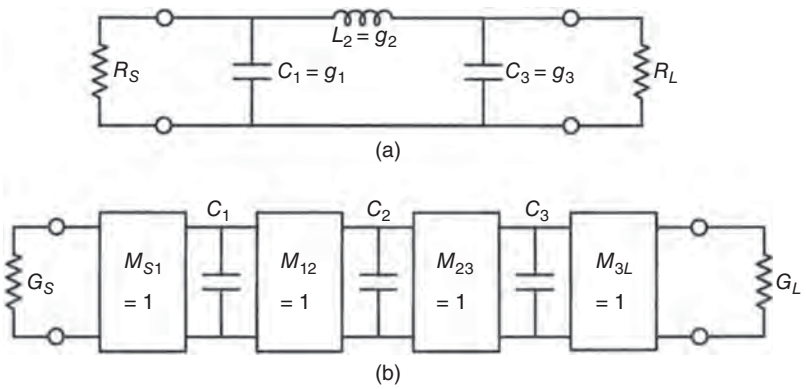


Figure 7.4 Third-degree lowpass prototype ladder network: the form after addition of the unit inverters.

inductor L_2 of a third-degree network that is transformed to a shunt capacitor (with the same value as that of the inductor, g_2). Also, two unit inverters M_{S1} and M_{3L} are added to transform the terminating impedances into conductances. Now the network is all admittance or conductances and is known as a *parallel resonator* lowpass prototype circuit. These inverters have a one-to-one relationship with the physical coupling elements in the finally realized filter structure.

A coupled-cavity microwave bandpass filter is realized directly from lowpass prototype circuits of the type shown in Figure 7.4. After the addition of the series and parallel resonating components [to form the *bandpass prototype* (BPP)], and the appropriate scaling to the center frequency and bandwidth of the RF bandpass filter, the design of the microwave structures or devices used to realize the elements of the bandpass circuit becomes apparent. These resonators are realized in the microwave structure by elements such as coaxial resonators. Figure 7.5 illustrates the procedure for a fourth-degree all-pole filter.

In this chapter, we concentrate on the synthesis of parallel-connected components in an open-wire transmission environment, because many waveguide and coaxial elements are modeled directly as shunt-connected elements. These include the inductive or capacitive apertures in rectangular waveguides [14]. If the series element is required, it can be obtained through the use of the dual-network theorem, where a shunt-connected capacitor becomes a series-connected inductor of the same value, or an admittance (J) inverter becomes an impedance (K) inverter [15–17].

7.2.1 Synthesis of [*ABCD*] Polynomials for Circuits with Inverters

We use the third-degree network in Figure 7.4 to derive the general form of the polynomials for an inverter-coupled parallel resonator network.

Cascade the first inverter M_{S1} and the first shunt capacitor C_1:

$$\begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ sC_1 & 1 \end{bmatrix} = j \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ sC_1 & 1 \end{bmatrix} = j \begin{bmatrix} sC_1 & 1 \\ 1 & 0 \end{bmatrix} \tag{7.9}$$

Cascade the second inverter M_{12}:

$$j \begin{bmatrix} sC_1 & 1 \\ 1 & 0 \end{bmatrix} \cdot j \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = - \begin{bmatrix} 1 & sC_1 \\ 0 & 1 \end{bmatrix} \tag{7.10}$$

Cascade the second shunt capacitor C_2:

$$- \begin{bmatrix} 1 & sC_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC_2 & 1 \end{bmatrix} = - \begin{bmatrix} 1 + s^2 C_1 C_2 & sC_1 \\ sC_2 & 1 \end{bmatrix} \tag{7.11}$$

Cascade the third inverter M_{23}:

$$- \begin{bmatrix} 1 + s^2 C_1 C_2 & sC_1 \\ sC_2 & 1 \end{bmatrix} \cdot j \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -j \begin{bmatrix} sC_1 & 1 + s^2 C_1 C_2 \\ 1 & sC_2 \end{bmatrix} \tag{7.12}$$

At this stage, we have included two frequency-variant elements C_1 and C_2 and have, in effect, a second-degree filter network ($N = 2$). Note the form of the polynomials:

$A(s)$ and $D(s)$	odd polynomial degree $N - 1$
$B(s)$	even polynomial degree N
$C(s)$	even polynomial degree $N - 2$

Then cascade the third shunt capacitor C_3:

$$-j \begin{bmatrix} sC_1 & 1 + s^2 C_1 C_2 \\ 1 & sC_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ sC_3 & 1 \end{bmatrix} = -j \begin{bmatrix} s(C_1 + C_3) + s^3 C_1 C_2 C_3 & 1 + s^2 C_1 C_2 \\ 1 + s^2 C_2 C_3 & sC_2 \end{bmatrix} \tag{7.13}$$

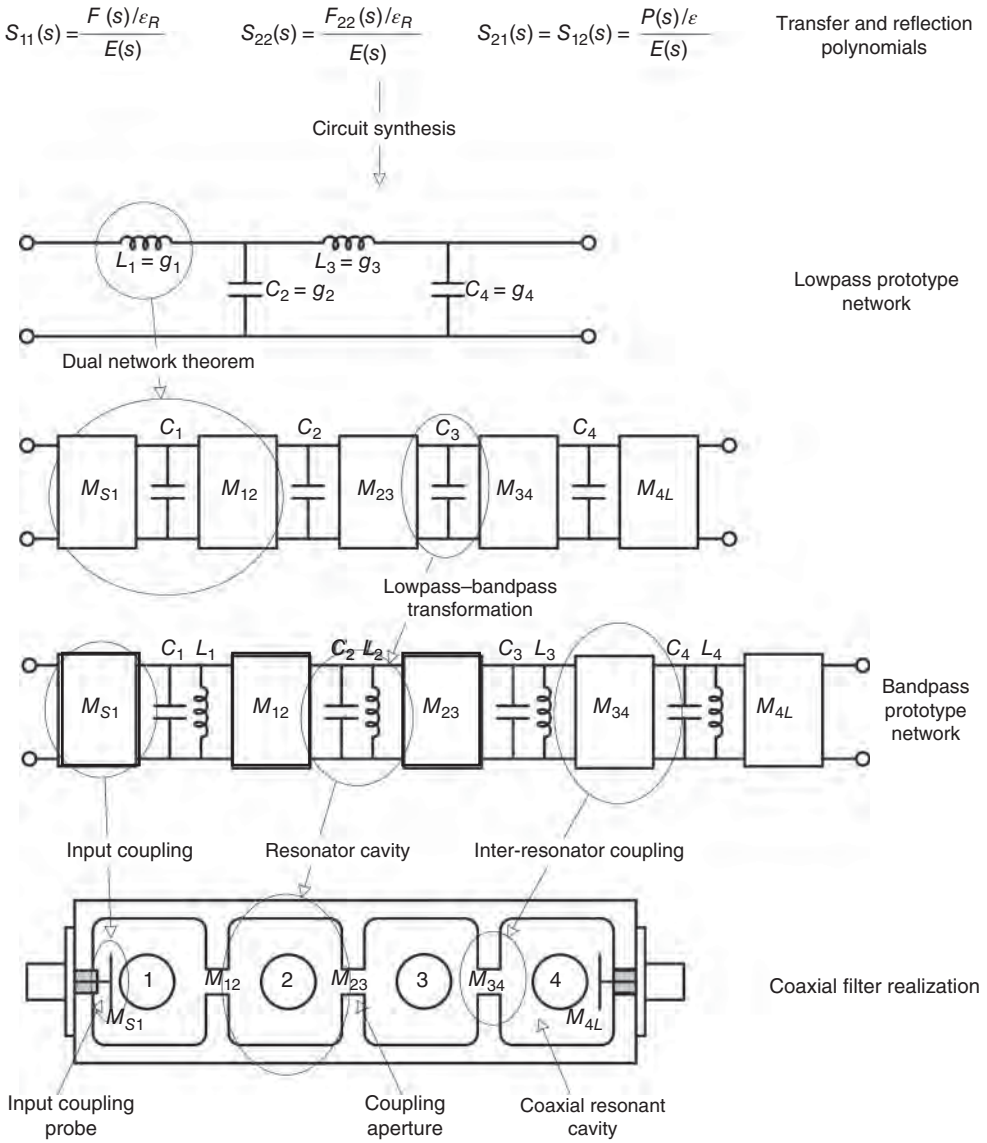


Figure 7.5 Steps in the synthesis process for a fourth-degree coaxial resonator bandpass filter.

Cascade the fourth inverter M_{3L}:

$$-j\begin{bmatrix} s(C_1+C_3)+s^3C_1C_2C_3 & 1+s^2C_1C_2 \\ 1+s^2C_2C_3 & sC_2 \end{bmatrix} \cdot j\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 1+s^2C_1C_2 & s(C_1+C_3)+s^3C_1C_2C_3 \\ sC_2 & 1+s^2C_2C_3 \end{bmatrix} \tag{7.14}$$

Note again the forms of the polynomials for this third-degree circuit ($N = 3$):

$A(s)$ and $D(s)$	Even polynomial degree $N-1$
$B(s)$	Odd polynomial degree N
$C(s)$	Odd polynomial degree $N-2$

These cascade operations reveal some important features of the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials that represent the filter ladder network. All are polynomials in variable s with real coefficients and are even or odd, depending on whether N is even or odd. Independently of whether N is even or odd, $B(s)$ is of degree N and is the highest degree. $A(s)$ and $D(s)$ are of degree $N-1$, and $C(s)$ is the lowest degree $N-2$.

Next, we relate the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials with the $S_{21}(s)$ and $S_{11}(s)$ transfer and reflection polynomials ($E(s), P(s)/\epsilon$, and $F(s)/\epsilon_R$). For the general case, the coefficients of $E(s)$ and $F(s)$ can have complex coefficients (see Section 3.10). To account for the general cases, it is necessary to add a frequency-independent reactance (FIR) in parallel with each shunt frequency-dependent element (the capacitors C_i) such that the value of the admittance at each node, instead of being simply sC_i, now becomes $sC_i + jB_i$ (see Figure 7.1).

The effect of adding the FIRs at each node is to convert the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials for the ladder network into complex-even and complex-odd forms; that is, their coefficients alternate between purely real and purely imaginary as the power of s increases. This is demonstrated by taking, for example, the $B(s)$ polynomial for the second-degree [$ABCD$] matrix (7.12) with no FIR added as follows:

$$B(s) = -j(1 + s^2 C_1 C_2)$$

Replacing sC_1 with $sC_1 + jB_1$ and sC_2 with $sC_2 + jB_2$, we obtain

$$\begin{aligned} B(s) &= -j[(1 - B_1 B_2) + j(B_1 C_2 + B_2 C_1)s + C_1 C_2 s^2] \\ &= j[b_0 + jb_1 s + b_2 s^2] \end{aligned} \tag{7.15}$$

where the coefficients b_i are purely real. Similarly for $D(s)$, we obtain

$$\begin{aligned} D(s) &= -j[sC_2] && \text{without a FIR} \\ D(s) &= -j[jB_2 + sC_2] && \text{with the FIR } jB_2 \\ &= j[jd_0 + d_1 s] \end{aligned} \tag{7.16}$$

where the coefficients d_i are the coefficients of the $D(s)$ polynomial and are purely real.

In a similar fashion, the coefficients of $A(s)$ and $C(s)$ are derived. In general, for the even-degree cases,

$$\begin{aligned} A(s) &= j[ja_0 + a_1 s + ja_2 s^2 + a_3 s^3 + \cdots + a_{N-1} s^{N-1}] \\ B(s) &= j[b_0 + jb_1 s + b_2 s^2 + jb_3 s^3 + \cdots + jb_{N-1} s^{N-1} + b_N s^N] \\ C(s) &= j[c_0 + jc_1 s + c_2 s^2 + jc_3 s^3 + \cdots + c_{N-2} s^{N-2}] \\ D(s) &= j[jd_0 + d_1 s + jd_2 s^2 + d_3 s^3 + \cdots + d_{N-1} s^{N-1}] \end{aligned} \tag{7.17}$$

Note that all the coefficients a_i, b_i, c_i, d_i, $i = 0, 1, 2, \ldots, N$ are real. In a similar way it can be shown that for the odd-degree case (see the third-degree matrix in equation (7.14)),

$$\begin{aligned} A(s) &= a_0 + ja_1 s + a_2 s^2 \\ B(s) &= jb_0 + b_1 s + jb_2 s^2 + b_3 s^3 \\ C(s) &= jc_0 + c_1 s \\ D(s) &= d_0 + jd_1 s + d_2 s^2 \end{aligned} \tag{7.18}$$

Typically, for N odd, we obtain

$$\begin{aligned} A(s) &= a_0 + ja_1 s + a_2 s^2 + ja_3 s^3 + \cdots + a_{N-1} s^{N-1} \\ B(s) &= jb_0 + b_1 s + jb_2 s^2 + b_3 s^3 + \cdots + jb_{N-1} s^{N-1} + b_N s^N \end{aligned}$$

$$C(s) = jc_0 + c_1 s + jc_2 s^2 + c_3 s^3 + \cdots + c_{N-2} s^{N-2}$$
$$D(s) = d_0 + jd_1 s + d_2 s^2 + jd_3 s^3 + \cdots + d_{N-1} s^{N-1} \tag{7.19}$$

Note that the j outside the brackets in equation (7.17) is a consequence of using inverters as coupling elements, each with its associated frequency-invariant 90° phase shift. For the even-degree case, there are an odd number of inverters along the main path of the signal and therefore an odd number of 90° phase shifts, thus the multiplication by j. For the purpose of the synthesis method described next, the j can be safely ignored.

The reason for adding the FIRs is to introduce the purely imaginary coefficients ja_i, jb_i into the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials in addition to the purely real coefficients. These purely imaginary coefficients need to be present in the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials to ensure that they are properly matched with the corresponding terms in the $E(s)$ and $F(s)$ polynomials when the latter are representing an asymmetric transfer function. In these cases, $E(s)$ and $F(s)$ have some complex coefficients, and with the inclusion of the FIRs, it is possible to relate network polynomials $A(s)$, $B(s)$, $C(s)$, and $D(s)$ with the transfer and reflection polynomials $E(s)$ $F(s)$ and $P(s)$ as follows.

Knowing now the form of the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials, of the polynomials constituting $S_{11}(s)$ and $S_{21}(s)$ ($E(s)$, $F(s)/\varepsilon_R$, and $P(s)/\varepsilon$), and also of the relationship between them [equation set (7.1)], the coefficients of $A(s)$, $B(s)$, $C(s)$, and $D(s)$ may be directly expressed in terms of the coefficients of $E(s)$ and $F(s)/\varepsilon_R$. These expressions are derived from equation set (6.33), using the relationships $z + z^* = 2\,\mathrm{Re}(z)$, $z - z^* = 2j\,\mathrm{Im}(z)$ (z = complex number).

From equation set (7.1),

$$[ABCD] = \frac{1}{jP(s)/\varepsilon} \cdot \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix}$$

where, for N even, we have

$$\begin{aligned}
A(s) &= j\,\mathrm{Im}(e_0 + f_0) + \mathrm{Re}(e_1 + f_1)s + j\,\mathrm{Im}(e_2 + f_2)s^2 + \cdots + j\,\mathrm{Im}(e_N + f_N)s^N \\
B(s) &= \mathrm{Re}(e_0 + f_0) + j\,\mathrm{Im}(e_1 + f_1)s + \mathrm{Re}(e_2 + f_2)s^2 + \cdots + \mathrm{Re}(e_N + f_N)s^N \\
C(s) &= \mathrm{Re}(e_0 - f_0) + j\,\mathrm{Im}(e_1 - f_1)s + \mathrm{Re}(e_2 - f_2)s^2 + \cdots + \mathrm{Re}(e_N - f_N)s^N \\
D(s) &= j\,\mathrm{Im}(e_0 - f_0) + \mathrm{Re}(e_1 - f_1)s + j\,\mathrm{Im}(e_2 - f_2)s^2 + \cdots + j\,\mathrm{Im}(e_N - f_N)s^N
\end{aligned} \tag{7.20a}$$

and for N odd

$$\begin{aligned}
A(s) &= \mathrm{Re}(e_0 + f_0) + j\,\mathrm{Im}(e_1 + f_1)s + \mathrm{Re}(e_2 + f_2)s^2 + \cdots + j\,\mathrm{Im}(e_N + f_N)s^N \\
B(s) &= j\,\mathrm{Im}(e_0 + f_0) + \mathrm{Re}(e_1 + f_1)s + j\,\mathrm{Im}(e_2 + f_2)s^2 + \cdots + \mathrm{Re}(e_N + f_N)s^N \\
C(s) &= j\,\mathrm{Im}(e_0 - f_0) + \mathrm{Re}(e_1 - f_1)s + j\,\mathrm{Im}(e_2 - f_2)s^2 + \cdots + \mathrm{Re}(e_N - f_N)s^N \\
D(s) &= \mathrm{Re}(e_0 - f_0) + j\,\mathrm{Im}(e_1 - f_1)s + \mathrm{Re}(e_2 - f_2)s^2 + \cdots + j\,\mathrm{Im}(e_N - f_N)s^N
\end{aligned} \tag{7.20b}$$

Here, e_i and f_i, $i = 0, 1, 2, \ldots, N$, are the complex coefficients of $E(s)$ and $F(s)/\varepsilon_R$, respectively.

It is easily demonstrated that these polynomials satisfy the equation set (7.1) and are the correct form for the representation of the electrical performance of a passive, reciprocal, and lossless microwave circuit.

7.2.1.1 Form

- The coefficients of $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials alternate between purely real and purely imaginary (or vice versa) as the power of s increases.
- For noncanonical networks, $\varepsilon_R = 1$ and so the highest-degree coefficients of $E(s)$ and $F(s)/\varepsilon_R$, e_N, and f_N, respectively, also equal unity. Therefore, the highest-degree coefficient

of $B(s)$, $b_N = \mathrm{Re}(e_N + f_N) = 2$, and the highest-degree coefficients of the $A(s)$, $C(s)$, and $D(s)$ polynomials, $a_N = j\,\mathrm{Im}(e_N + f_N)$, $c_N = \mathrm{Re}(e_N - f_N)$, and $d_N = j\,\mathrm{Im}(e_N - f_N)$, respectively, are all zero (e_N and f_N are real numbers equal to unity). In addition, the process for building the $E(s)$ polynomial from the $F(s)$ and $P(s)$ polynomials results in $\mathrm{Im}(e_{N-1}) = \mathrm{Im}(f_{N-1})$ for $n_{fz} < N$; therefore, $c_{N-1} = j\,\mathrm{Im}(e_{N-1} - f_{N-1}) = 0$ as well. Thus $B(s)$ is of degree N, $A(s)$ and $D(s)$ are of degree $N-1$, and $C(s)$ is of degree $N-2$, in accordance with equations (7.17) and (7.19). The exception to this is the canonical transfer functions, where $n_{fz} = N$, that is, where $\varepsilon_R \neq 1$. Here, the polynomial $C(s)$ is also of degree N.

7.2.1.2 Equations

- It is easily shown by simple addition and subtraction in equation set (7.20) that the following relationships hold:

$$\begin{aligned}
&A(s) + B(s) + C(s) + D(s) = 2E(s)\\
&A(s) + B(s) - C(s) - D(s) = 2F(s)/\varepsilon_R\\
&-A(s) + B(s) - C(s) + D(s) = 2F(s)^*/\varepsilon_R = 2F_{22}(s)/\varepsilon_R \quad (N \text{ even})\\
&-A(s) + B(s) - C(s) + D(s) = -2F(s)^*/\varepsilon_R = 2F_{22}(s)/\varepsilon_R \quad (N \text{ odd})
\end{aligned} \tag{7.21}$$

which clearly satisfy equation set (7.1), repeated here for convenience as follows:

$$[ABCD] = \frac{1}{jP(s)/\varepsilon}\begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix}$$

where

$$\begin{aligned}
S_{11}(s) &= \frac{F(s)/\varepsilon_R}{E(s)} = \frac{A(s) + B(s) - C(s) - D(s)}{A(s) + B(s) + C(s) + D(s)}\\
S_{22}(s) &= \frac{(-1)^N F(s)^*/\varepsilon_R}{E(s)} = \frac{D(s) + B(s) - C(s) - A(s)}{A(s) + B(s) + C(s) + D(s)}\\
S_{12}(s) &= S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)} = \frac{2P(s)/\varepsilon}{A(s) + B(s) + C(s) + D(s)}
\end{aligned}$$

If the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials are derived from equation (7.20), the constant ε is used directly, as determined previously. If the polynomials are normalized to b_N, the highest-degree coefficient of $B(s)$, ε is rederived by using the property of the $[ABCD]$ matrix that at any frequency, its determinant equals unity:

$$A(s)D(s) - B(s)C(s) = \left[\frac{P(s)}{\varepsilon}\right]^2 \tag{7.22}$$

If we evaluate equation (7.22) at zero frequency, we need to consider only the constant coefficients of the polynomials

$$a_0 d_0 - b_0 c_0 = \left[\frac{p_0}{\varepsilon}\right]^2$$

or

$$\varepsilon = \left|\frac{p_0}{\sqrt{a_0 d_0 - b_0 c_0}}\right| \tag{7.23}$$

As described in Chapter 8, short-circuit admittance parameters (y parameters) are used for the direct synthesis of coupling matrices for filter networks. At this stage, the y-parameter matrix $[y]$ is found by using the $[ABCD] \rightarrow [y]$ parameter conversion formula [11] as follows:

$$\frac{1}{P(s)/\varepsilon}\begin{bmatrix}A(s) & B(s)\\ C(s) & D(s)\end{bmatrix} \Rightarrow \begin{bmatrix}y_{11}(s) & y_{12}(s)\\ y_{21}(s) & y_{22}(s)\end{bmatrix}$$

$$\begin{bmatrix}y_{11}(s) & y_{12}(s)\\ y_{21}(s) & y_{22}(s)\end{bmatrix} = \frac{1}{y_d(s)}\begin{bmatrix}y_{11n}(s) & y_{12n}(s)\\ y_{21n}(s) & y_{22n}(s)\end{bmatrix}$$

$$= \frac{1}{B(s)}\begin{bmatrix}D(s) & \dfrac{-\Delta_{ABCD}P(s)}{\varepsilon}\\ \dfrac{-P(s)}{\varepsilon} & A(s)\end{bmatrix} \tag{7.24}$$

where $y_{ijn}(s)$, $i, j = 1,2$ are the numerator polynomials of $y_{ij}(s)$, $y_d(s)$ is their common denominator polynomial, and Δ_{ABCD} is the determinant of the $[ABCD]$ matrix, which for a reciprocal network, $\Delta_{ABCD} = 1$. From this, it becomes clear that

$$\begin{aligned} y_d(s) &= B(s)\\ y_{11n}(s) &= D(s)\\ y_{22n}(s) &= A(s)\\ y_{21n}(s) &= y_{12n}(s) = \frac{-P(s)}{\varepsilon} \end{aligned} \tag{7.25}$$

demonstrating that the short-circuit admittance parameters can be built up from the coefficients of $S_{11}(s)$ and $S_{21}(s)$, in the same way as $[ABCD]$ polynomials are by using equation set (7.20) (see Appendix 6A). The open-circuit impedance parameters Z_{ij} are built up in a similar way.

The $y(s)$ polynomials have the same form as $[ABCD]$ polynomials, namely, their coefficients alternate between purely real and purely imaginary as the power of s increases, $y_d(s)$ is of degree N, $y_{11n}(s)$ and $y_{22n}(s)$ are of degree $N-1$, and $y_{21n}(s)$ and $y_{12n}(s)$ are of degree n_{fz} (the number of finite-position TZs). The short-circuit admittance parameters are used for the direct coupling matrix synthesis method described in the next chapter.

7.2.2 Synthesis of [*ABCD*] Polynomials for the Singly Terminated Filter Prototype

Singly terminated filter networks are designed to operate from very high- or very low-impedance sources. They were once used to couple directly from the outputs of thermionic valve amplifiers and certain transistor amplifiers with very high internal impedances (effectively, current sources). Singly terminated filters are considered here because the admittance looking in at the input port has a characteristic that is particularly useful for the design of contiguous-channel manifold multiplexers, described in Chapter 18.

Figure 7.6 shows the development of a network with a zero-impedance source to the Thévenin equivalent network with a load impedance Z_L attached. The voltage v_L at the load can be written [8, 18] as follows:

$$v_L = \frac{-y_{12}v_S}{y_{22}} \cdot \frac{Z_L}{(Z_L + 1/y_{22})} \tag{7.26}$$

Thus, the voltage gain of this network is

$$S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)} = \frac{v_L}{v_S} = \frac{-y_{12}Z_L}{1 + Z_L y_{22}} \tag{7.27}$$

Setting the terminating impedance $Z_L = 1\ \Omega$ and replacing y_{12} and y_{22} with $y_{12n}(s)/y_d(s)$ and $y_{22n}(s)/y_d(s)$, respectively, we write

$$\frac{P(s)/\varepsilon}{E(s)} = \frac{-y_{12n}(s)}{y_d(s) + y_{22n}(s)} \tag{7.28}$$

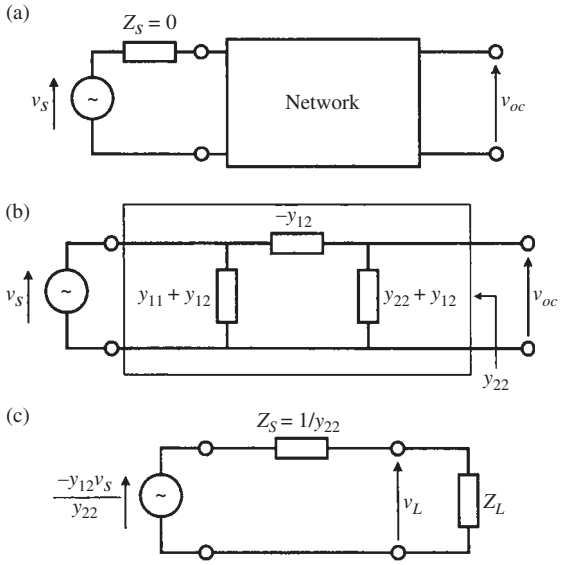


Figure 7.6 Singly terminated filter network: (a) network with zero-impedance voltage source; (b) network replaced with equivalent π network; (c) Thévenin equivalent circuit with load impedance Z_L attached.

By splitting $E(s)$ into its complex-even and complex-odd components, we obtain

$$E(s) = m_1 + n_1$$

where

$$\begin{aligned} m_1 &= \mathrm{Re}(e_0) + j\,\mathrm{Im}(e_1)s + \mathrm{Re}(e_2)s^2 + \cdots + \mathrm{Re}(e_N)s^N \\ n_1 &= j\,\mathrm{Im}(e_0) + \mathrm{Re}(e_1)s + j\,\mathrm{Im}(e_2)s^2 + \cdots + j\,\mathrm{Im}(e_N)s^N \end{aligned} \tag{7.29}$$

Substituting equation (7.29) into equation (7.28) yields

$$\frac{P(s)/\varepsilon}{m_1 + n_1} = \frac{-y_{12n}(s)}{y_d(s) + y_{22n}(s)} \tag{7.30}$$

$E(s)$ is normalized such that its highest-degree coefficient e_N equals unity. Also, we know that the coefficients of $y_{12n}(s)$, $y_{22n}(s)$, and $y_d(s)$ polynomials must alternate between purely real and purely imaginary as the power of s increases and that the $y_d(s)$ polynomial must be one degree greater than $y_{22n}(s)$. As a result,

$$\begin{aligned} &\text{For } N \text{ even:} && y_d(s) = m_1, && y_{22n}(s) = n_1 \\ &\text{For } N \text{ odd:} && y_d(s) = n_1, && y_{22n}(s) = m_1 \\ &\text{For } N \text{ even or odd:} && y_{12n}(s) = y_{21n}(s) = \frac{-P(s)}{\varepsilon} \end{aligned} \tag{7.31}$$

Now the polynomials $y_{12n}(s)$, $y_{22n}(s)$, and $y_d(s)$ for the singly ended filter network are known in terms of the desired transfer function. In addition, this gives the $A(s)$ and $B(s)$ polynomials of the $[ABCD]$ matrix, since $A(s) = y_{22n}(s)$ and $B(s) = y_d(s)$.

Thus, the $A(s)$ and $B(s)$ polynomials are easily found from the coefficients of the known $E(s)$ polynomial in a fashion similar to that used for the doubly terminated case. However, it remains to find the $C(s)$ and $D(s)$ polynomials to commence network synthesis. This is easily accomplished by a method due to Levy [13]. This method takes advantage of the property of an $[ABCD]$ matrix for a reciprocal network, for which the determinant of the matrix is unity at any value of the frequency variable s:

$$A(s)D(s) - B(s)C(s) = \left(\frac{P(s)}{\varepsilon}\right)^2 \tag{7.32}$$

Taking a fourth-degree example and this time writing the relationship (7.32) out in matrix form,

Fourth-degree case, $N = 4$

Degree of $P(s)/\varepsilon$ $= n_{fz}$ = number of finite-position transmission zeros

Degree of $A(s)$ and $D(s)$ $= N - 1 = 3$

Degree of $B(s)$ $= N = 4$

Degree of $C(s)$ $= N - 2 = 2$

$= N = 4$ for fully canonical cases ($n_{fz} = N$):

$$\begin{bmatrix} a_0 & 0 & 0 & 0 & b_0 & 0 & 0 & 0 & 0 \\ a_1 & a_0 & 0 & 0 & b_1 & b_0 & 0 & 0 & 0 \\ a_2 & a_1 & a_0 & 0 & b_2 & b_1 & b_0 & 0 & 0 \\ a_3 & a_2 & a_1 & a_0 & b_3 & b_2 & b_1 & b_0 & 0 \\ 0 & a_3 & a_2 & a_1 & b_4 & b_3 & b_2 & b_1 & b_0 \\ 0 & 0 & a_3 & a_2 & 0 & b_4 & b_3 & b_2 & b_1 \\ 0 & 0 & 0 & a_3 & 0 & 0 & b_4 & b_3 & b_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_4 & b_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_4 \end{bmatrix} \cdot \begin{bmatrix} d_0 \\ d_1 \\ d_2 \\ d_3 \\ -c_0 \\ -c_1 \\ -c_2 \\ -c_3 \\ -c_4 \end{bmatrix} = \frac{1}{\varepsilon^2} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \\ \cdots \\ \cdots \\ p_{2nfz} \\ 0 \\ 0 \end{bmatrix} \tag{7.33}$$

where p_i, $i = 0, 1, 2, \ldots, 2n_{fz}$ are $2n_{fz} + 1$ coefficients of $P(s)^2$. By inverting the left-hand square matrix (7.33) and postmultiplying it by the column matrix on the right-hand side (RHS), containing $2n_{fz} + 1$ coefficients of $P(s)^2$, the coefficients of the $C(s)$ and $D(s)$ polynomials are found. Because $y_{11n}(s) = D(s)$ and $y_{21n}(s)$, $y_{22n}(s)$, and $y_d(s)$ are already known from equation (7.25), the admittance matrix $[y]$ for the filter network is also fully defined.

In practice, it is very rare that a singly terminated fully canonical filter design is needed. For the noncanonical cases where $n_{fz} < N$, the last two rows and columns of the square matrix and the last two rows of the two column matrices can be omitted before matrix (7.33) is evaluated.

7.3 Ladder Network Synthesis

The preceding sections describe the evaluation of the overall $[ABCD]$ and the equivalent admittance $[y]$ matrix from the characteristic polynomials of the desired filter function. From this matrix, we can synthesize the network given its topology. The synthesis process is similar to that described in Section 7.1.

From equation (7.1), the $[ABCD]$ matrix for the transfer functions is represented by

$$[ABCD] = \frac{1}{jP(s)/\varepsilon} \cdot \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix}$$

where $A(s)$, $B(s)$, $C(s)$, $D(s)$, and $P(s)$ are polynomials in the frequency variable s with coefficients that alternate between purely real or purely imaginary (or vice versa), as the power of s increases. For an Nth-degree characteristic with n_{fz} finite-position TZs, $A(s)$ and $D(s)$ are of degree $N - 1$, $B(s)$ is of degree N, $C(s)$ is of degree $N - 2$, (except for fully canonical cases where $C(s)$ is of degree N as well as $B(s)$), and $P(s)$ is of degree n_{fz}. When $n_{fz} = 0$ (an all-pole network), then $P(s)/\varepsilon$ is of zero degree (i.e., a constant $= 1/\varepsilon$) and the network can be synthesized as a regular ladder network without cross-couplings. If $n_{fz} > 0$, indicating that there are finite-position TZs, then the network must be synthesized including the cross-couplings (couplings between nonadjacent resonator nodes), or by other means such as extracted pole.

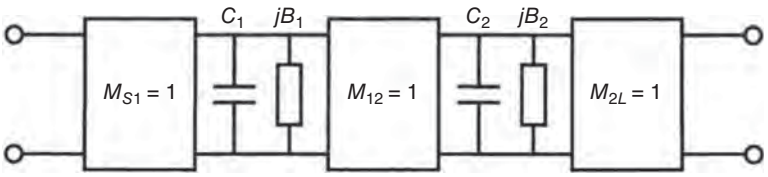


Figure 7.7 Second-degree ladder network with FIRs and inverters.

To illustrate the synthesis procedure, we begin with the synthesis of a simple second-degree network without cross-couplings, but including the FIRs (see Figure 7.7). Usually, FIRs are associated with cross-coupled networks that are realizing asymmetric filter characteristics. However, in this non-cross-coupled case, the network is of the asynchronously tuned or conjugately tuned symmetric type, introduced in Chapter 6 (predistorted filter with off-axis reflection zeros). The process for the extraction of parallel-coupled inverters is outlined, leading the way to the synthesis of more advanced networks, capable of realizing asymmetric and linear phase filters.

The [*ABCD*] matrix for the network in Figure 7.7 is constructed from equations (7.15) and (7.16):

$$[ABCD] = -j\begin{bmatrix}(jB_1 + sC_1) & (1 - B_1B_2) + j(B_1C_2 + B_2C_1)s + C_1C_2s^2\\ 1 & (jB_2 + sC_2)\end{bmatrix} \tag{7.34}$$

The procedure for extracting the elements is the reverse of building up the [*ABCD*] matrix from the individual elements in the ladder network. Knowledge of the type of element and the order in which it appears in the network is required beforehand. Each element is extracted by premultiplying the overall [*ABCD*] matrix with the inverse [*ABCD*] matrix of the elemental component to be extracted, leaving a remainder matrix in series with a unit matrix, which can be ignored. The procedure for the extraction of series- and shunt-connected frequency-variant components is given in Section 7.1.

In this case, the first extracted element is a frequency-invariant inverter of unity characteristic admittance $J = M_{S1} = 1$ (see Figure 7.8):

$$[ABCD] = \begin{bmatrix}0 & j/J\\ jJ & 0\end{bmatrix}$$

$$\text{Inverse} = \begin{bmatrix}0 & -j/J\\ -jJ & 0\end{bmatrix} = -j\begin{bmatrix}0 & 1/J\\ J & 0\end{bmatrix} \tag{7.35}$$

Putting $J = 1$ (unit inverter) and premultiplying the overall [*ABCD*] matrix (7.34) yields

$$\begin{bmatrix}A(s) & B(s)\\ C(s) & D(s)\end{bmatrix}_{\text{rem}} = -j\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\cdot(-j)\begin{bmatrix}A(s) & B(s)\\ C(s) & D(s)\end{bmatrix} = -\begin{bmatrix}C(s) & D(s)\\ A(s) & B(s)\end{bmatrix}$$

$$= -\begin{bmatrix}1 & (jB_2 + sC_2)\\ (jB_1 + sC_1) & (1 - B_1B_2) + j(B_1C_2 + B_2C_1)s + C_1C_2s^2\end{bmatrix} \tag{7.36}$$

The effect of extracting the unit inverter is to exchange the top and bottom rows of the [*ABCD*] matrix. Since the admittance, looking into the network, is given by (assuming a unity load termination)

$$Y_{\text{in}} = \frac{C(s) + D(s)}{A(s) + B(s)} \tag{7.37}$$

Figure 7.8 Extraction of the admittance inverter.

extracting the inverter, in effect, inverts the input admittance:

$$Y_{\text{in(rem)}} = \frac{1}{Y_{\text{in}}} \tag{7.38}$$

The next element in the chain to be extracted is the shunt capacitor C_1. We select the frequency-variant component first, instead of the FIR next to it. At this stage in the synthesis, the degree of the $D(s)$ polynomial is one degree greater than the $B(s)$ polynomial, and $C(s)$ is one degree greater than $A(s)$.

The short-circuit admittance parameter y_{11} is calculated by

$$y_{11} = \frac{D(s)}{B(s)} = \frac{(1 - B_1B_2) + j(B_1C_2 + B_2C_1)s + C_1C_2s^2}{(jB_2 + sC_2)} \tag{7.39}$$

The numerator polynomial is one degree higher than the denominator, indicating that the next extracted component must be frequency variant. Before the value of the component can be determined, we need to divide both sides of equation (7.39) by s and evaluate it as $s \to j\infty$:

$$\left.\frac{y_{11}}{s}\right|_{s\to j\infty} = \left.\frac{D(s)}{sB(s)}\right|_{s\to j\infty} = C_1 \tag{7.40a}$$

Therefore, $y_{11} = sC_1$, indicating that C_1 is the value of a frequency-variant admittance. Alternatively, the open-circuit admittance z_{11} may be used:

$$z_{11} = \frac{A(s)}{C(s)} = \frac{1}{(jB_1 + sC_1)}$$

Multiplying both sides by s and evaluating at $s = j\infty$ yields

$$sz_{11}|_{s\to j\infty} = \left.\frac{sA(s)}{C(s)}\right|_{s\to j\infty} = \frac{1}{C_1} \tag{7.40b}$$

or

$$z_{11} = 1/sC_1$$

Having determined the value of the shunt capacitor C_1, we now extract it from matrix (7.36) by

$$\begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix}_{\text{rem}} = -\begin{bmatrix} 1 & 0 \\ -sC_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & (jB_2 + sC_2) \\ (jB_1 + sC_1) & (1 - B_1B_2) + j(B_1C_2 + B_2C_1)s + C_1C_2s^2 \end{bmatrix}$$

$$\begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix}_{\text{rem}} = -\begin{bmatrix} 1 & (jB_2 + sC_2) \\ jB_1 & (1 - B_1B_2) + jsB_1C_2 \end{bmatrix} \tag{7.41}$$

Note that in the remainder matrix, both $C(s)$ and $D(s)$ decrease in degree by one and are now the same degree as $A(s)$ and $B(s)$, respectively. This indicates that the next component to be extracted should be frequency invariant. Also, note that there is no C_1 component in the remainder matrix, which indicates that the extraction of the parallel capacitor C_1 has been successful.

Again by evaluating at $s = j\infty$, but this time, without multiplying or dividing by s, we obtain

$$y_{11}|_{s\to j\infty} = \left.\frac{D(s)}{B(s)}\right|_{s\to j\infty} = \left.\frac{(1 - B_1B_2) + jsB_1C_2}{(jB_2 + sC_2)}\right|_{s\to j\infty} = jB_1$$
$$z_{11}|_{s\to j\infty} = \left.\frac{A(s)}{C(s)}\right|_{s\to j\infty} = \frac{1}{jB_1} \tag{7.42}$$

Now this shunt FIR can be extracted from the matrix in (7.40) by

$$\begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix}_{\text{rem}} = -\begin{bmatrix} 1 & 0 \\ -jB_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & (jB_2 + sC_2) \\ jB_1 & (1 - B_1B_2) + jsB_1C_2 \end{bmatrix}$$
$$= -\begin{bmatrix} 1 & (jB_2 + sC_2) \\ 0 & 1 \end{bmatrix} \tag{7.43}$$

Again, it is evident that both $C(s)$ and $D(s)$ have dropped by one degree and that jB_1 no longer appears in the remainder matrix, indicating a successful extraction.

The next component in the cascade is another unity admittance inverter that is extracted, as before, leaving a remainder matrix:

$$\begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix}_{\text{rem}} = j\begin{bmatrix} 0 & 1 \\ 1 & (jB_2 + sC_2) \end{bmatrix} \tag{7.44}$$

C_2 and jB_2 can now be extracted in the same way as for C_1 and jB_1. Because the $C(s)$ and $D(s)$ polynomials are one degree greater than $A(s)$ and $B(s)$, respectively, the next component to be extracted is required to be frequency-variant (C_2), followed by the FIR (jB_2). Following the same procedure as before, the remainder matrix after extracting C_2 and jB_2 is

$$\begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix}_{\text{rem}} = j\begin{bmatrix} 1 & 0 \\ -jB_2 - sC_2 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & (jB_2 + sC_2) \end{bmatrix}$$
$$= j\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{7.45}$$

which is simply the final inverter.

Now the circuit is fully synthesized and the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials are reduced to constants or zero. During the synthesis process, the polynomials are reduced degree by degree, while the ladder network is built up element by element.

However, all this has not yet affected the $P(s)$ polynomial, which forms the denominator of the $[ABCD]$ matrix. The $P(s)$ polynomial is simply a constant ($= 1/\varepsilon$) for all-pole transfer functions; that is, no finite-position TZs are present in the complex s plane. As a result, no cross-couplings (couplings between nonadjacent resonator nodes) are required. When TZs do appear in finite positions in the complex plane (on the imaginary axis or in pairs symmetric about the imaginary axis), then the $P(s)$ polynomial is finite in degree that corresponds to the number of finite-position zeros n_{fz}. To reduce the $P(s)$ polynomial to a constant, along with the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials, *parallel coupled inverters* (PCIs) need to be extracted at appropriate stages in the synthesis procedure. The theory behind the extraction of a PCI is presented. It is followed by an example of the synthesis process for the fourth-degree, asymmetric (4–2) filter, used as the example in Chapter 6 for the synthesis of $S_{21}(s)$ and $S_{11}(s)$ polynomials for the Chebyshev type of transfer function. The order in which the circuit elements are extracted from the polynomials is important, and so are the rules governing the extraction sequence.

7.3.1 Parallel-Coupled Inverters: Extraction Process

The inverter is extracted in parallel from between the input and output terminals of the two-port network, leaving a remainder matrix from which further parallel-coupled inverters can be extracted if necessary (see Figure 7.9).

The initial [*ABCD*] matrix is in the form, described by equation (7.1), where the $P(s)$ polynomial is finite in degree (i.e., where TZs are present). If $P(s)$ is less in degree than the initial or remainder network that it is being extracted from, then the value of the characteristic admittance, J, of the PCI, is zero. Note also that the polynomial $P(s)$ is multiplied by j; this allows the cross-couplings to be extracted as 90° inverters (or transformers). This is in addition to the multiplication of $P(s)$ by j, when $N - n_{fz}$ is even.

7.3.2 Extraction of Parallel Admittance Inverters

The overall [*ABCD*] matrix is

$$[ABCD] = \frac{1}{jP(s)/\varepsilon} \cdot \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} \tag{7.46}$$

The equivalent y matrix is

$$[y] = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \frac{1}{B(s)} \cdot \begin{bmatrix} D(s) & \dfrac{-jP(s)}{\varepsilon} \\ \dfrac{-jP(s)}{\varepsilon} & A(s) \end{bmatrix} \tag{7.47}$$

The [*ABCD*] matrix for the parallel admittance inverter J is

$$[ABCD]_{\text{inv}} = \begin{bmatrix} 0 & \dfrac{j}{J} \\ jJ & 0 \end{bmatrix} \tag{7.48}$$

and its equivalent y matrix is

$$[y]_{\text{inv}} = \begin{bmatrix} 0 & jJ \\ jJ & 0 \end{bmatrix} \tag{7.49}$$

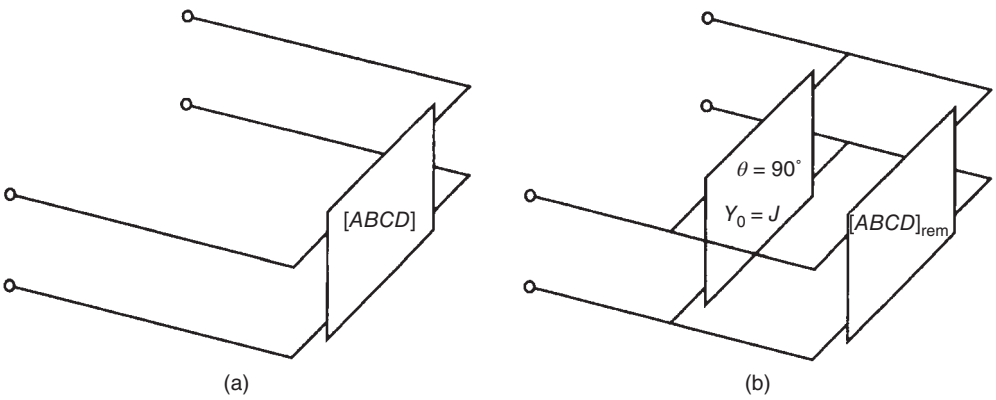


Figure 7.9 Extraction of a parallel-coupled inverter (PCI): (a) overall network as represented by [*ABCD*] matrix; (b) with an inverter of characteristic admittance J in parallel with the remainder matrix $[ABCD]_{\text{rem}}$.

The overall y matrix $[y]$ is the sum of the admittance matrix for the inverter and the admittance matrix $[y]_{\text{rem}}$; the remaining matrix of the network after the inverter is extracted:

$$[y] = [y]_{\text{inv}} + [y]_{\text{rem}} \tag{7.50}$$

Therefore

$$[y]_{\text{rem}} = [y] - [y]_{\text{inv}} = \frac{1}{B(s)} \cdot \begin{bmatrix} D(s) & \dfrac{-jP(s)}{\varepsilon} \\ \dfrac{-jP(s)}{\varepsilon} & A(s) \end{bmatrix} - \begin{bmatrix} 0 & jJ \\ jJ & 0 \end{bmatrix}$$

$$= \frac{1}{B(s)} \cdot \begin{bmatrix} D(s) & -j\left(\dfrac{P(s)}{\varepsilon} + JB(s)\right) \\ -j\left(\dfrac{P(s)}{\varepsilon} + JB(s)\right) & A(s) \end{bmatrix} \tag{7.51}$$

Reconverting to the $[ABCD]$ matrix yields

$$[ABCD]_{\text{rem}} = \frac{1}{jP_{\text{rem}}(s)} \cdot \begin{bmatrix} A_{\text{rem}}(s) & B_{\text{rem}}(s) \\ C_{\text{rem}}(s) & D_{\text{rem}}(s) \end{bmatrix} = \frac{-1}{y_{21\text{rem}}} \cdot \begin{bmatrix} y_{22\text{rem}} & 1 \\ \Delta_{y\text{rem}} & y_{11\text{rem}} \end{bmatrix} \tag{7.52}$$

$$\therefore A_{\text{rem}}(s) = \frac{-y_{22\text{rem}}}{y_{21\text{rem}}} = \frac{A(s)}{j(P(s)/\varepsilon + JB(s))} \tag{7.53a}$$

$$B_{\text{rem}}(s) = \frac{-1}{y_{21\text{rem}}} = \frac{B(s)}{j(P(s)/\varepsilon + JB(s))}$$

$$C_{\text{rem}}(s) = \frac{-\Delta_{y\text{rem}}}{y_{21\text{rem}}} = \frac{1}{B(s)} \cdot \frac{(A(s)D(s) + (P(s)/\varepsilon + JB(s))^2)}{j(P(s)/\varepsilon + JB(s))}$$

$$= \frac{A(s)D(s) + (P(s)/\varepsilon)^2 + 2JB(s)P(s)/\varepsilon + (JB(s))^2}{jB(s)(P(s)/\varepsilon + JB(s))} \tag{7.53b}$$

Since $A(s)D(s) - B(s)C(s) = -(P(s)/\varepsilon)^2$, it follows that

$$C_{\text{rem}}(s) = \frac{B(s)C(s) + 2JB(s)P(s)/\varepsilon + (JB(s))^2}{jB(s)(P(s)/\varepsilon + JB(s))} = \frac{C(s) + 2JP(s)/\varepsilon + J^2B(s)}{j(P(s)/\varepsilon + JB(s))} \tag{7.53c}$$

and finally

$$D_{\text{rem}}(s) = \frac{-y_{11\text{rem}}}{y_{21\text{rem}}} = \frac{D(s)}{j(P(s)/\varepsilon + JB(s))} \tag{7.53d}$$

Thus, the remainder $[ABCD]$ matrix has the following form:

$$[ABCD]_{\text{rem}} = \frac{1}{jP_{\text{rem}}(s)/\varepsilon} \cdot \begin{bmatrix} A_{\text{rem}}(s) & B_{\text{rem}}(s) \\ C_{\text{rem}}(s) & D_{\text{rem}}(s) \end{bmatrix}$$

$$= \frac{1}{j(P(s)/\varepsilon + JB(s))} \cdot \begin{bmatrix} A(s) & B(s) \\ C(s) + 2JP(s)/\varepsilon + J^2B(s) & D(s) \end{bmatrix} \tag{7.54}$$

Now the elements of the remainder $[ABCD]$ matrix, $[ABCD]_{\text{rem}}$, are known in terms of the elements of the original matrix and the value of parallel inverter J. The value J is found by knowing that after the extraction of the parallel inverter, the degree of $P_{\text{rem}}(s)$, the denominator polynomial of $[ABCD]_{\text{rem}}$ must be one degree less than that of $P(s)/\varepsilon$ polynomial of the original $[ABCD]$ matrix.

In order for the inverter to be extracted with a nonzero value, the $P(s)$ and $B(s)$ polynomials are both nth degree and are written out in terms of their coefficients as follows:

$$\frac{P(s)}{\varepsilon} = p_0 + p_1 s + p_2 s^2 + \cdots + p_n s^n \tag{7.55a}$$

and

$$B(s) = b_0 + b_1 s + b_2 s^2 + \cdots + b_n s^n \tag{7.55b}$$

Consequently, $jP_{\text{rem}}(s)/\varepsilon = j(P(s)/\varepsilon + JB(s))$, and therefore, the highest-degree coefficient of $P_{\text{rem}}(s)/\varepsilon$ is computed as follows:

$$\begin{aligned} jp_{n\text{rem}} &= j(p_n + Jb_n) \\ &= 0 \quad \text{if } (p_n + Jb_n) = 0, \text{ i.e., } J = \frac{-p_n}{b_n} \end{aligned} \tag{7.56}$$

Since the value of J is now known, the polynomials of the remainder matrix may be determined:

$$\begin{aligned} A_{\text{rem}}(s) &= A(s) & C_{\text{rem}}(s) &= C(s) + \frac{2JP(s)}{\varepsilon} + J^2 B(s) \\ B_{\text{rem}}(s) &= B(s) & D_{\text{rem}}(s) &= D(s) \\ \frac{P_{\text{rem}}(s)}{\varepsilon} &= \frac{P(s)}{\varepsilon} + JB(s) \end{aligned} \tag{7.57}$$

It is evident that the only polynomials affected by the extraction of a parallel-connected inverter are $P(s)$ (which drops by one degree) and $C(s)$. At the end of the synthesis procedure, the $P(s)$ polynomial will have reduced to zero degree, a constant.

7.3.3 Summary of the Principal Components of the Prototype Network

The main elements employed to synthesize the parallel-resonator cross-coupled network are summarized in Figure 7.10. Where it is appropriate, the formulas used to determine the elements' values from the $A(s)$, $B(s)$, $C(s)$, $D(s)$, and $P(s)$ polynomials, and the operations on the polynomials used to obtain the remainder polynomials, are included.

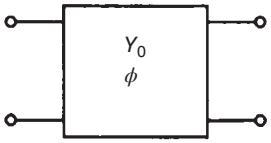
7.4 Synthesis Example of an Asymmetric (4–2) Filter Network

This example is intended to clarify the rules to be observed when synthesizing a network [19]:

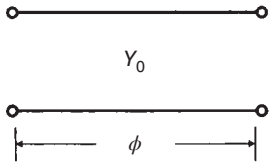
- Extractions are carried out at the source or load terminals only.
- Extractions are carried out on the $[ABCD]$ matrix remaining from the previous step, starting with the full matrix.
- When cross-couplings are present, use $D(s)/B(s)$ only to evaluate the values of $sC + jB$ pairs.
- For the shunt $sC + jB$ pairs, C is extracted first.
- For parallel (cross-coupling) inverters, $sC + jB$ are extracted at each end first.
- For fully canonical networks, a parallel inverter between the source and load terminals is extracted first.
- The last element to be extracted must be a parallel inverter.

At the end of extractions, $A(s)$, $B(s)$, $C(s)$, $D(s)$, and $P(s)$ polynomials should all be zeros, apart from the constant in $B(s)$.

(a)



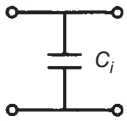
$$A_{rem}(s) = A(s)\cos\phi - jC(s)\sin\phi/Y_0$$
$$B_{rem}(s) = B(s)\cos\phi - jD(s)\sin\phi/Y_0$$
$$C_{rem}(s) = C(s)\cos\phi - jA(s)Y_0\sin\phi$$
$$D_{rem}(s) = D(s)\cos\phi - jB(s)Y_0\sin\phi$$
$$P_{rem}(s) = P(s)$$



When $\phi = \pi/2$ and $Y_0 = 1$ (a unit inverter), this simplifies to

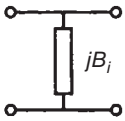
$$A_{rem}(s) = -jC(s)$$
$$B_{rem}(s) = -jD(s)$$
$$C_{rem}(s) = -jA(s)$$
$$D_{rem}(s) = -jB(s)$$
$$P_{rem}(s) = P(s)$$

(b)



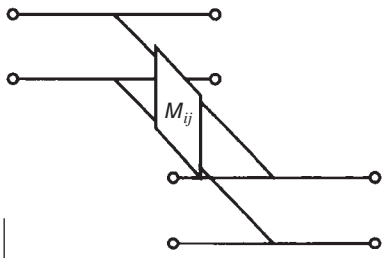
$$C_i = \left.\frac{D(s)}{sB(s)}\right|_{s=j\infty} = \left.\frac{C(s)}{sA(s)}\right|_{s=j\infty}$$
$$A_{rem}(s) = A(s)$$
$$B_{rem}(s) = B(s)$$
$$C_{rem}(s) = C(s) - sC_iA(s)$$
$$D_{rem}(s) = D(s) - sC_iB(s)$$
$$P_{rem}(s) = P(s)$$

(c)



$$B_i = \left.\frac{D(s)}{B(s)}\right|_{s=j\infty} \text{ or } \left.\frac{C(s)}{A(s)}\right|_{s=j\infty}$$
$$A_{rem}(s) = A(s)$$
$$B_{rem}(s) = B(s)$$
$$C_{rem}(s) = C(s) - B_iA(s)$$
$$D_{rem}(s) = D(s) - B_iB(s)$$
$$P_{rem}(s) = P(s)$$

(d)



$$M_{ij} = -\left.\frac{P(s)}{B(s)}\right|_{s=j\infty}$$
$$A_{rem}(s) = A(s)$$
$$B_{rem}(s) = B(s)$$
$$P_{rem}(s)/\varepsilon = P(s)/\varepsilon + M_{ij}B(s)$$
$$C_{rem}(s) = C(s) + 2M_{ij}P(s)/\varepsilon + M_{ij}^2B(s)$$
$$D_{rem}(s) = D(s)$$

Figure 7.10 Cross-coupled array components and their extraction formulas: (a) transmission line and special case where $\phi = \pi/2$ and $Y_0 = 1$ (unit inverter); (b) shunt-connected frequency-variant capacitor; (c) shunt-connected frequency-invariant reactance; (d) parallel cross-coupling inverter.

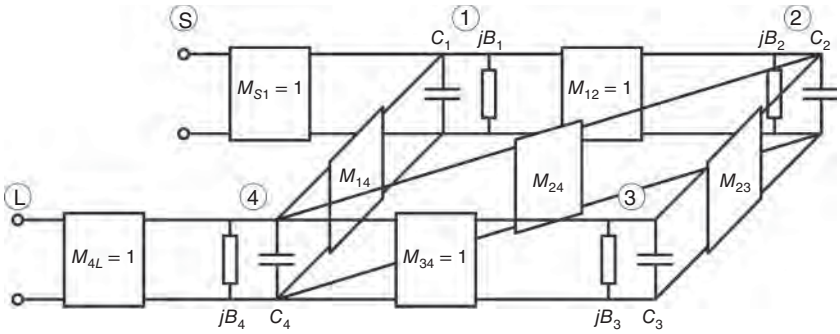


Figure 7.11 Folded cross-coupled network for the (4–2) asymmetric prototype.

To demonstrate the circuit synthesis approach, the (4–2) asymmetric prototype (see Figure 7.11), for which $S_{21}(s)$ and $S_{11}(s)$ functions are developed in Chapter 6, is chosen. The network is constructed as a folded array, with two cross-couplings, one diagonal and one straight. Because the original prototype is asymmetric, the shunt-connected FIRs in the network will be nonzero.

The polynomials corresponding to the zeros of the numerators and denominators of $S_{21}(s)$ and $S_{11}(s)$ rational functions, $E(s)$, $F(s)$, and $P(s)$, derived in Chapter 6 for the asymmetric (4–2) prototype, are summarized as follows:

$s^i, i =$	$E(s)$	$F(s)/\varepsilon_R$	$P(s)$
0	$-0.1268 - j2.0658$	$+0.0208$	$+2.3899$
1	$+2.4874 - j3.6255$	$-j0.5432$	$+j3.1299$
2	$+3.6706 - j2.1950$	$+0.7869$	-1.0
3	$+2.4015 - j0.7591$	$-j0.7591$	—
4	$+1.0$	$+1.0$	—
		$\varepsilon_R = 1.0$	$\varepsilon = 1.1548$

Note that $P(s)$ has been multiplied by j twice here, once because $N - n_{fz}$ is an even number and once to allow the cross-coupling elements to be extracted as inverters. With knowledge of the coefficients of $E(s)$, $F(s)$, and $P(s)$, we can build up $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials according to equations (7.20), (7.29), and (7.31), summarized as follows:

$s^i, i =$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\varepsilon$
0	$-j2.0658$	-0.1059	-0.1476	$-j2.0658$	$+2.0696$
1	$+2.4874$	$-j4.1687$	$-j3.0823$	$+2.4874$	$+j2.7104$
2	$-j2.1950$	$+4.4575$	$+2.8836$	$-j2.1950$	-0.8660
3	$+2.4015$	$-j1.5183$	—	$+2.4015$	—
4	—	$+2.0$	—	—	—

Network synthesis begins with the extraction of the input coupling inverter at the source end of the network:

Extraction number	Extracted element	Working node
1	Series unit inverter M_{S1}	S

The effect of extracting a unit inverter is to exchange $B(s)$ and $D(s)$ polynomials, and $A(s)$ and $C(s)$ polynomials, and to multiply them all by $-j$ (see Figure 7.10a):

$s^i, i=$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\varepsilon$
0	$+j0.1476$	-2.0658	-2.0658	$+j0.1059$	$+2.0696$
1	-3.0823	$-j2.4874$	$-j2.4874$	-4.1687	$+j2.7104$
2	$-j2.8836$	-2.1950	-2.1950	$-j4.4575$	-0.8660
3	—	$-j2.4015$	$-j2.4015$	-1.5183	—
4	—	—	—	$-j2.0$	—

Now we are at working node 1 and are able to extract a frequency-variant capacitor C_1 and an FIR jB_1, in that order:

Extraction number	Extracted element	Working node
2	Capacitor C_1	1
3	Reactance jB_1	1

Evaluate C_1 from

$$C_1 = \left.\frac{D(s)}{sB(s)}\right|_{s=j\infty} = 0.8328$$

Extract C_1:

$s^i, i=$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\varepsilon$
0	$+j0.1476$	-2.0658	-2.0658	$+j0.1059$	$+2.0696$
1	-3.0823	$-j2.4874$	$-j2.6103$	-2.4482	$+j2.7104$
2	$-j2.8836$	-2.1950	$+0.3720$	$-j2.3860$	-0.8660
3	—	$-j2.4015$	—	$+0.3098$	—
4	—	—	—	—	—

Evaluate jB_1 from

$$B_1 = \left.\frac{D(s)}{B(s)}\right|_{s=j\infty} = j0.1290$$

Extract jB_1:

$s^i, i =$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\epsilon$
0	$+j0.1476$	-2.0658	-2.0468	$+j0.3724$	$+2.0696$
1	-3.0823	$-j2.4874$	$-j2.2127$	-2.7691	$+j2.7104$
2	$-j2.8836$	-2.1950	—	$-j2.1028$	-0.8660
3	—	$-j2.4015$	—	—	—
4	—	—	—	—	—

After C_1 and jB_1 are extracted from working node 1 (see Figure 7.10), one of the terminals of the cross-coupling inverter M_{14} is encountered. To access the second of the two terminals, we need to turn the network (which is simply the exchange of the $A(s)$ and $D(s)$ polynomials) and extract the series unit inverter M_{4L}, followed by the capacitor–reactance pair $C_4 + jB_4$. This is accomplished according to steps 1–3 (in the preceding tabular list), yielding $C_4 = 0.8328$ and $B_4 = 0.1290$ (the same values as C_1 and B_1).

Extraction number	Extracted element	Working node
Turn network	—	—
4	Series unit inverter M_{4L}	L
5	Capacitor C_4	L
6	Reactance jB_4	L

Having extracted these components, the remainder polynomials are

$s^i, i =$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\epsilon$
0	$+j2.0468$	$+0.1476$	$+0.6364$	$+j2.0468$	$+2.0696$
1	-2.2127	$+j3.0823$	$+j1.3499$	-2.2127	$+j2.7104$
2	—	-2.8836	-0.2601	—	-0.8660
3	—	—	—	—	—
4	—	—	—	—	—

Now it may be seen that polynomial $P(s)$ is the same degree as $B(s)$, indicating that the network is ready for the extraction of the cross-coupling inverter M_{14}:

Extraction number	Extracted element	Working node
7	Parallel inverter M_{14}	1, 4

Evaluate

$$M_{14} = \left.\frac{-P(s)}{\epsilon B(s)}\right|_{s=j\infty} = -0.3003$$

from equation (7.56) and then extract by using equation (7.57). The remainder polynomials are given in the next tabular list. Note that only $C(s)$ and $P(s)$ have been affected by the extraction of the cross-coupling inverter:

$s^i, i =$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\varepsilon$
0	$+j2.0468$	$+0.1476$	-0.5933	$+j2.0468$	$+2.0253$
1	-2.2127	$+j3.0823$	—	-2.2127	$+j1.7848$
2	—	-2.8836	—	—	—
3	—	—	—	—	—
4	—	—	—	—	—

At this stage we have encountered one of the terminals of the diagonal cross-coupling inverter M_{24} and need to turn the network and extract the series unit inverter M_{12} and then the capacitor/reactance pair $C_2 + jB_2$ at node 2 before we can access its other terminal at node 2. Now M_{24} may be extracted in the same way as M_{14} above:

Extraction number	Extracted element	Working node
Turn network	—	—
8	Series unit inverter M_{12}	1
9	Capacitor C_2	2
10	Reactance jB_2	3
11	Parallel inverter M_{24}	2, 4

Thus, $C_2 = 1.3032$ and $B_2 = -0.1875$, followed by $M_{24} = -0.8066$. The polynomials, after they are extracted, are listed as follows:

$s^i, i =$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\varepsilon$
0	$+j0.5933$	$+2.0468$	—	$+j0.2362$	$+0.3744$
1	—	$+j2.2127$	—	—	—
2	—	—	—	—	—
3	—	—	—	—	—
4	—	—	—	—	—

Finally, one of the terminals of the last cross-coupling inverter M_{23} (which is actually a main-line inverter, but since it is the last inverter, it is treated as a cross-coupling inverter) is encountered at node 2. In a similar way to steps 8–11, the network is turned, and the series unit inverter M_{34} and the capacitor/reactance pair $C_3 + jB_3$ at node 3 are extracted. Only then can we access the other terminal at node 3 and extract it:

Extraction number	Extracted element	Working node
Turn network	—	—
12	Series unit inverter M_{34}	4
13	Capacitor C_3	3
14	Reactance jB_3	3
15	Parallel inverter M_{23}	2, 3

This gives $C_3 = 3.7296$ and $B_2 = -3.4499$, followed by $M_{23} = -0.6310$. After these elements are extracted, the remaining polynomials are all zeros except for the constant of $B(s) = 0.5933$. The synthesis is now complete. The values of the 15 elements that have been extracted are summarized as follows:

$$C_1 + jB_1 = 0.8328 + j0.1290 \quad M_{s1} = 1.0000 \quad M_{14} = -0.3003$$
$$C_2 + jB_2 = 1.3032 - j0.1875 \quad M_{12} = 1.0000 \quad M_{24} = -0.8066$$
$$C_3 + jB_3 = 3.7296 - j3.4499 \quad M_{34} = 1.0000 \quad M_{23} = -0.6310$$
$$C_4 + jB_4 = 0.8328 + j0.1290 \quad M_{4L} = 1.0000$$

7.4.1 Scaling at Resonator Nodes

The general canonical lowpass prototype network, synthesized using this circuit element extraction approach, consists of an array of N shunt-connected capacitors. Each is in parallel with an FIR, and the sequentially numbered capacitor/FIR nodes are coupled directly by 90° inverters (mainline couplings). If the filter transfer function contains finite-position TZs, the nonsequential nodes are coupled again by inverters (cross-couplings). In addition, there are two more inverters to couple the input (source) to the first node and the last node to the output terminal (load). In some cases, other internal nodes, apart from the first and last, are also connected to the source and load terminals. For fully canonical transfer functions, the source and load terminals are directly coupled with an inverter. In the real bandpass filter structure, the capacitor/FIR pairs become resonators, offset (or asynchronously tuned) with respect to the center frequency of the bandpass filter if the FIR is not zero in value. The inverters become input, output, or interresonator coupling devices, such as apertures, probes, and loops.

When each node is surrounded by inverters, it is possible to scale the impedance levels at each node to arbitrary values by adjusting the characteristic immittances of the inverters. Essentially, this is equivalent to adjusting the transformer ratios of the entry and exit coupling transformers such that the total energy transferring through the node remains the same.

For a network of nodes and couplings such as in Figure 7.10, the amount of energy passing through each inverter is represented by the scalar coupling coefficient k_{ij}, computed by

$$k_{ij} = \frac{M_{ij}}{\sqrt{C_i \cdot C_j}} \tag{7.58}$$

where M_{ij} is the characteristic immittance of the inverter coupling between the ith and jth nodes. For a given transfer function, all k_{ij} in the circuit must be kept constant. Therefore, M_{ij},

C_i, and C_j can be scaled to new values M'_{ij}, C'_i, and C'_j, provided that the following relationship is satisfied:

$$k_{ij} = \frac{M_{ij}}{\sqrt{C_i \cdot C_j}} = \frac{M'_{ij}}{\sqrt{C'_i \cdot C'_j}} \tag{7.59}$$

Thus, for a given network topology, there are an infinite number of combinations of M'_{ij}, C'_i, and C'_j that can be chosen. The different combinations result in the same electrical performance provided that the proper relationship with the starting values M_{ij}, C_i, and C_j is maintained by using equation (7.59).

In this chapter, we use two typical schemes.

- All mainline coupling inverters M_{ij} are scaled to unity value, giving the shunt capacitors C_i (and in most cases, the terminating admittance G_L) nonunity values. This form leads directly to the ladder network lowpass prototype forms often used for classical circuit-based filter design.
- All shunt capacitors at the resonator nodes are scaled to unity, which give the coupling inverters nonunity values. This form is adopted for the construction of the coupling matrix and is dealt with in the next chapter. Applying the scaling formulas to the asymmetric (4-2) example for which the element values were derived above, yields the following non-unity values for the inverters. Note that the values of the C'_i have been scaled to unity. Because the terminating admittances G'_S and G'_L have also been scaled to unity, these inverter values may be directly entered into a coupling matrix.

$$\begin{array}{llll} C'_1 = 1.0 & B'_1 = M'_{11} = B_1/C_1 = +0.1549 & M'_{12} = 0.9599 & G'_S = 1.0 \\ C'_2 = 1.0 & B'_2 = M'_{22} = B_2/C_2 = -0.1439 & M'_{23} = 0.2862 & G'_L = 1.0 \\ C'_3 = 1.0 & B'_3 = M'_{33} = B_3/C_3 = -0.9250 & M'_{34} = 0.5674 & \\ C'_4 = 1.0 & B'_4 = M'_{44} = B_4/C_4 = +0.1549 & M'_{14} = 0.3606 & M'_{S1} = 1.0958 \\ & & M'_{24} = 0.7742 & M'_{4L} = 1.0958 \end{array}$$

7.4.2 Scaling of Inverters to Unity

Figure 7.12 shows a typical lowpass prototype network (without cross-couplings) with source and load admittances, shunt capacitors C_i, and mainline coupling inverters M_{ij}. Typically, these components have nonunity values. To scale the inverters to unity, we progress through the network, starting at the source end. The first inverter at this end, M_{S1}, is scaled to unity by applying equation (7.59) with $M'_{S1} = 1$. Then

$$\frac{M_{S1}}{\sqrt{G_S C_1}} = \frac{1}{\sqrt{G'_S C'_1}} \tag{7.60}$$

Therefore, the new value for $C_1 (= C'_1)$ is

$$C'_1 = \frac{G_S C_1}{G'_S M^2_{S1}} \tag{7.61}$$

If we must retain the original terminating admittance value ($G'_S = G_S$), then

$$C'_1 = \frac{C_1}{M^2_{S1}} \tag{7.62}$$

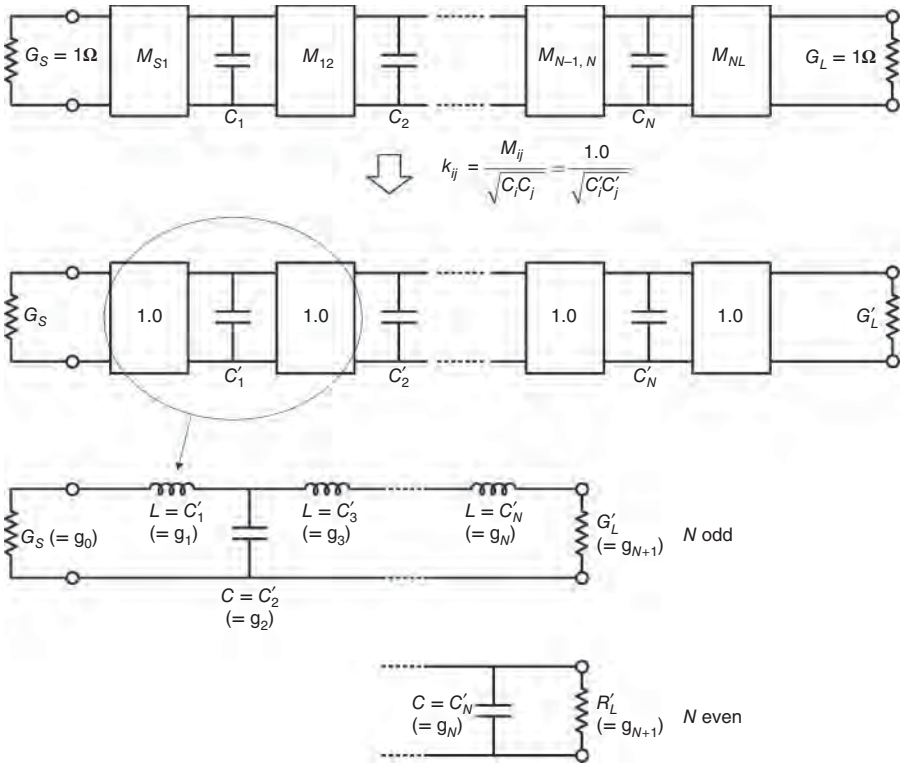


Figure 7.12 Lumped-element lowpass prototype ladder networks.

By applying (7.59) to the second inverter M_{12}, we obtain

$$\frac{M_{12}}{\sqrt{C_1 C_2}} = \frac{M'_{12}}{\sqrt{C'_1 C'_2}}$$

If $M'_{12} \to 1$, then

$$C'_2 = \frac{C_1 C_2}{C'_1 M_{12}^2} \tag{7.63}$$

This is repeated throughout the network until the last inverter M_{NL} is reached as follows:

$$\frac{M_{NL}}{\sqrt{C'_N G_L}} = \frac{1}{\sqrt{C'_N G'_L}} \text{ or } G'_L = \frac{G_L}{M_{NL}^2} \tag{7.64}$$

In general this means that the new terminating admittance G'_L is nonunity in value. However, for filter functions with a reflection zero (the point of perfect transmission) at zero frequency, G'_L is unity. The second network shown in Figure 7.12 shows the scaled network, with all inverters = 1.

Once the mainline couplings are scaled to unity, and values C'_i are known, any cross-coupling inverters M_{ik} must be scaled to account for the new capacitor values at each end as follows:

$$M'_{ik} = M_{ik}\sqrt{\frac{C'_i C'_k}{C_i C_k}} \tag{7.65}$$

If there are nonzero FIRs (jB_i) at the resonator nodes, they are scaled in the same proportion as the capacitor at that node by

$$B_i' = B_i \frac{C_i'}{C_i} \tag{7.66}$$

Finally, if it is required that both terminations are unity (which is the usual case), scaling can be applied from both ends of the network, progressing toward the center, which leaves an internal inverter with a nonunity value at the meeting point. This is the natural form of the networks synthesized by the element extraction procedure described earlier in this chapter. For the asymmetric (4–2) filter example in Section 7.3.1, all the mainline inverter and termination values are unity except for the central mainline inverter M_{23}, which has the value 0.6310.

Figure 7.12 further demonstrates how the dual-network theorem is applied to alternate C_i' and unit inverter pairs to create series-connected inductances of values C_i'. The network is now in the classic form [11], where the C_i' is equivalent to g_i, as well as $G_S = g_0$ and G_L (or $R_L = g_{N+1}$). Note that $G_L'(R_L')$ equals unity for symmetric odd-degree functions. For the general case, $G_L'(R_L') \neq 1$.

Summary

In the previous chapters, we introduced methods to generate the transfer and the reflection polynomials for a variety of filter functions. The next step in the design process is to translate these polynomials into a prototype electrical circuit from which a practical microwave filter can be developed. There are two ways to accomplish it: the classical circuit synthesis method and the direct-coupling matrix synthesis approach. In this chapter, the circuit synthesis approach, based on the [*ABCD*] transfer matrix or the chain matrix, as it is sometimes known, is presented.

Circuit elements used for the classical filter synthesis include the lossless inductors and capacitors and the terminating resistors. The lowpass prototype filters derived from these elements inherently lead to symmetric response filters. In Sections 3.10 and 6.1, such prototype networks are generalized to include asymmetric response filters by including the hypothetical FIR. The inclusion of such elements gives rise to transfer and reflection polynomials with complex coefficients, rendering the lowpass prototype filter response to be asymmetric. In practical bandpass or bandstop filters, FIR elements appear as frequency offsets to the resonant circuits. Another element that is required in the synthesis of filters at microwave frequencies is the frequency-invariant impedance or admittance inverter, commonly referred to as the *immittance inverter*. Such inverters significantly simplify the physical realization of the distributed microwave filter structure. In a microwave circuit, immittance inverters can be realized approximately over narrow bandwidths by a variety of microwave structures, such as inductive irises and coupling probes.

The chapter begins by establishing the relationship between *ABCD* parameters and the scattering parameters, related to the transfer and reflection polynomials of the filter function. It goes on to show the buildup of the overall *ABCD* matrix of a network by cascading *ABCD* matrices of the individual elements. The reverse process—derivation of the individual elements from the overall *ABCD* matrix—provides the basis of circuit synthesis technique. The procedure described in this chapter includes the synthesis of both the symmetric and asymmetric lowpass prototype filters.

The first step in the synthesis procedure is to compute the *ABCD* matrix from the optimized lowpass prototype filter polynomials. The next step is to follow the procedure reverse of that used for cascading a number of *ABCD* matrices corresponding to individual elements.

In other words, the individual elements in the filter are extracted from the overall *ABCD* matrix in the reverse order, from the highest to the lowest values. The synthesis procedure is broken down into two stages. The first stage of the synthesis process involves lumped-element lossless inductors, capacitors, and FIR elements. The second stage of the synthesis process includes the immittance inverters. Use of such inverters allows for the prototype electrical circuit in a form suitable for realization with intercoupled microwave resonators. The technique is applicable for synthesizing lowpass prototype filters with a symmetric or asymmetric response in ladder form, as well as cross-coupled topologies. A further generalization is introduced to allow the synthesis of singly terminated filters. Such filters are required in the design of contiguous multiplexers, described in Chapter 18. The well-known g_k parameters are readily derived by the synthesis procedure provided in this chapter. An example of an asymmetric filter is included to demonstrate the synthesis procedure.

References

1 Darlington, S. (1939) Synthesis of reactance 4-poles which produce insertion loss characteristics. *Journal of Mathematical Physics*, **18**, 257–353.
2 Cauer, W. (1958) *Synthesis of Linear Communication Networks*, McGraw-Hill, New York.
3 Guillemin, E.A. (1957) *Synthesis of Passive Networks*, Wiley, New York.
4 Bode, H.W. (1945) *Network Analysis and Feedback Amplifier Design*, Van Nostrand, Princeton, NJ.
5 van Valkenburg, M.E. (1955) *Network Analysis*, Prentice-Hall, Englewood Cliffs, NJ.
6 Rhodes, J.D. (1976) *Theory of Electrical Filters*, Wiley, New York.
7 Baum, R.F. (1957) Design of unsymmetrical band-pass filters. *IRE Transactions on Circuit Theory*, **4**, 33–40.
8 van Valkenburg, M.E. (1960) *Introduction to Modern Network Synthesis*, Wiley, New York.
9 Bell, H.C. (1979) Transformed-variable synthesis of narrow-bandpass filters. *IEEE Transactions on Circuits and Systems*, **26**, 389–394.
10 Carlin, H.J. (1956) The scattering matrix in network theory. *IRE Transactions on Circuit Theory*, **3**, 88–96.
11 Matthaei, G., Young, L., and Jones, E.M.T. (1980) *Microwave Filters, Impedance Matching Networks and Coupling Structures*, Artech House, Norwood, MA.
12 Orchard, H.J. and Temes, G.C. (1968) Filter design using transformed variables. *IEEE Transactions on Circuit Theory*, **15**, 385–408.
13 Levy, R. (1994) Synthesis of general asymmetric singly- and doubly-terminated cross-coupled filters. *IEEE Transactions on Microwave Theory and Techniques*, **42**, 2468–2471.
14 Marcuvitz, N. (1986) *Waveguide Handbook*, vol. Electromagnetic Waves Series 21, IEE, London.
15 Cohn, S.B. (1957) Direct coupled cavity filters. *Proceedings of IRE*, **45**, 187–196.
16 Young, L. (1963) Direct coupled cavity filters for wide and narrow bandwidths. *IEEE Transactions on Microwave Theory and Techniques*, **11**, 162–178.
17 Levy, R. (1963) Theory of direct coupled cavity filters. *IEEE Transactions on Microwave Theory and Techniques*, **11**, 162–178.
18 Chen, M.H. (1977) Singly terminated pseudo-elliptic function filter. *COMSAT Technical Reviews*, **7**, 527–541.
19 Cameron, R.J. (1982) General prototype network synthesis methods for microwave filters. *ESA Journal*, **6**, 193–206.

8

Synthesis of Networks: Direct Coupling Matrix Synthesis Methods

In this chapter, we examine the *coupling matrix* representation of microwave filter circuits. Modeling the circuit in matrix form is particularly useful because matrix operations can then be applied, such as inversion, similarity transformation, and partitioning. Such operations simplify the synthesis, the reconfiguration of the topology, and the performance simulation of complex circuits. Moreover, the coupling matrix includes the individual elements of the practical filter. As a result, each element in the matrix can be identified uniquely with an element in the finished microwave device. This enables the attribution of the individual electrical characteristics of each element to be taken into account, such as the Q_u values for each resonator cavity, different dispersion characteristics for the various types of mainline coupling and cross-coupling within the filter.

The basic circuit that the coupling matrix represents is reviewed, and the method to construct the matrix directly from a lowpass prototype (LPP) circuit, synthesized in Chapter 7, is outlined. This is followed by the presentation of three methods to synthesize the coupling matrix directly from the filter's transfer and reflection polynomials, the $N \times N$, $N + 2$, and the "folded lattice" method.

8.1 The Coupling Matrix

In the early 1970s Atia and Williams introduced the concept of the coupling matrix as applied to dual mode symmetric waveguide filters [1–4]. The circuit model that they used was a "bandpass prototype" (BPP), which has the form of Figure 8.1a. The circuit comprises a cascade of lumped-element series resonators intercoupled through transformers; each resonator comprising a capacitor of 1 F in series with the self-inductances of the mainline transformers, which total 1 H within each loop. This gives a center frequency of 1 rad/s, and the couplings will be normalized to give a bandwidth of 1 rad/s. In addition every loop is theoretically coupled to every other loop through cross-mutual couplings between the mainline transformers.

This circuit as it stands will only support symmetric filtering characteristics, but with the addition of a series connected frequency-invariant reactance (FIR) within each loop, the capability of the circuit will be extended to include the asymmetric cases as well. This is shown in Figure 8.1b where also the self-inductances of the mainline transformers have been separated out and represented as a separate inductor within each loop.

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

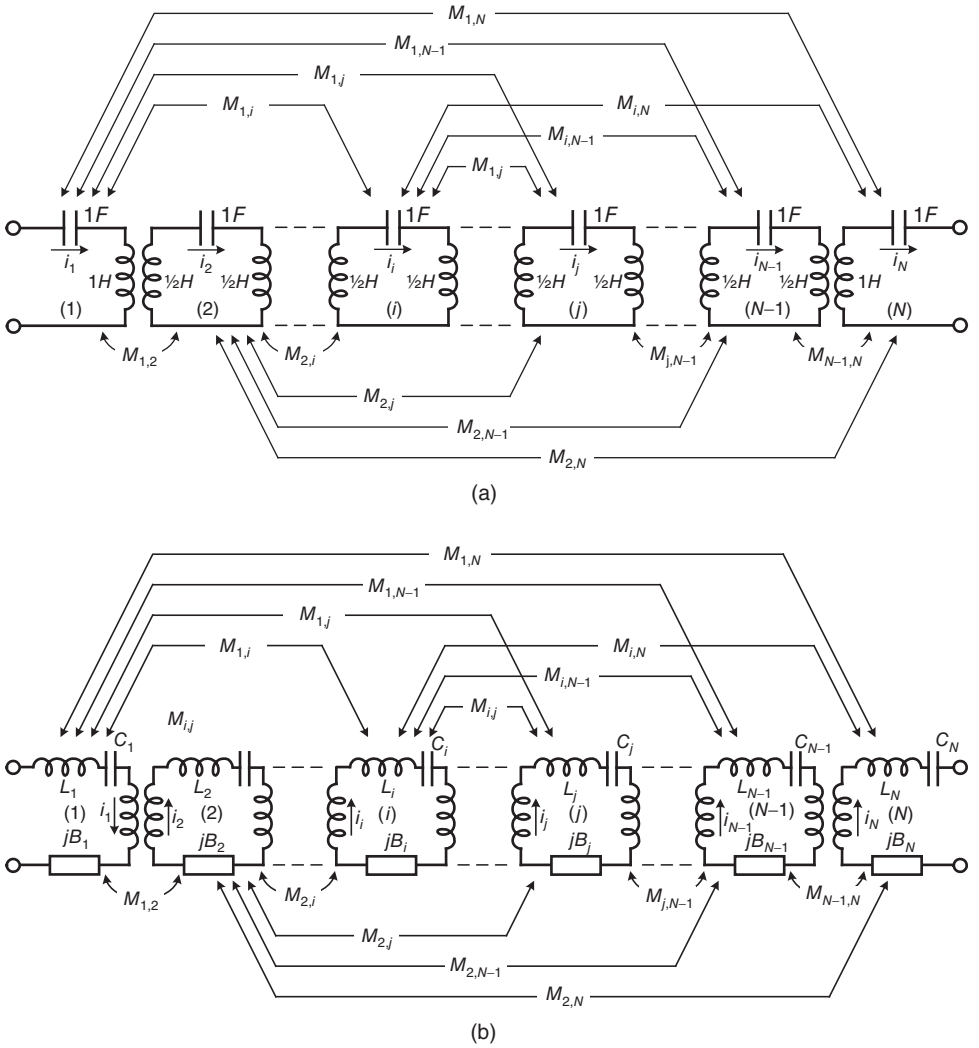


Figure 8.1 Multi-coupled series-resonator "bandpass prototype" network: (a) classical representation (*Source*: courtesy: A. E. Atia) (b) modified to include FIRs and separate self-inductors.

8.1.1 Bandpass and Lowpass Prototypes

The positive frequency characteristics of the BPP circuit may be related to the LPP characteristics that we have been considering in the earlier sections through the lumped element bandpass-to-lowpass mapping:

$$s = j\frac{\omega_0}{\omega_2 - \omega_1}\left[\frac{\omega_B}{\omega_0} - \frac{\omega_0}{\omega_B}\right] \tag{8.1}$$

where $\omega_0 = \sqrt{\omega_1\omega_2}$ is the center frequency of the BPP (=1 rad/s), ω_2 and ω_1 are the upper and lower band-edge frequencies, respectively (which can easily be shown to be $\omega_2 = 1.618$ and $\omega_1 = 0.618$ rad/s for a positive BPP, that is, a "bandwidth"[1] of $\omega_2 - \omega_1 = 1$ rad/s), and ω_B is the

1 The word "bandwidth" is between quotes here because, for the negative-frequency BPP, $\omega_2 - \omega_1$ becomes $(-\omega_2) - (-\omega_1) = -1$ rad/s, that is, its "bandwidth" is negative.

bandpass frequency variable. The mapping formula has been written in this way to ensure that the negative-frequency BPP characteristic correctly maps to the same point on the lowpass characteristic as the corresponding positive BPP frequency, even for asymmetric cases. For the negative-frequency BPP, all the frequency terms in (8.1) become negative so that an attenuation at $-\omega_B$ on the negative frequency BPP characteristic maps to the same point on the LPP characteristic in the lowpass domain as an attenuation at $+\omega_B$ on the positive BPP characteristic in the positive BP domain (Figure 8.2).

Because the coupling elements are frequency invariant, the series-resonator circuit itself may be transformed to the lowpass domain with the following steps:

1) Replace all mutual inductive couplings provided by transformers with inverters with the same values as the mutual couplings of the transformers. The inverters will provide the same amount of coupling energy between the resonator nodes as the transformers, with the same 90° phase change (Figure 8.3).
2) Transform the bandpass network to an LPP network with band edges at $\omega = \pm 1$ by letting the value of the series capacitance go to infinity (i.e., zero series impedance).

This is now in the form of the LPP circuits that were synthesized from the filter $S_{21}(s)$ and $S_{11}(s)$ polynomials in the last chapter. Because the coupling elements are assumed frequency invariant, synthesis in the lowpass or bandpass domains will yield the same values for the circuit

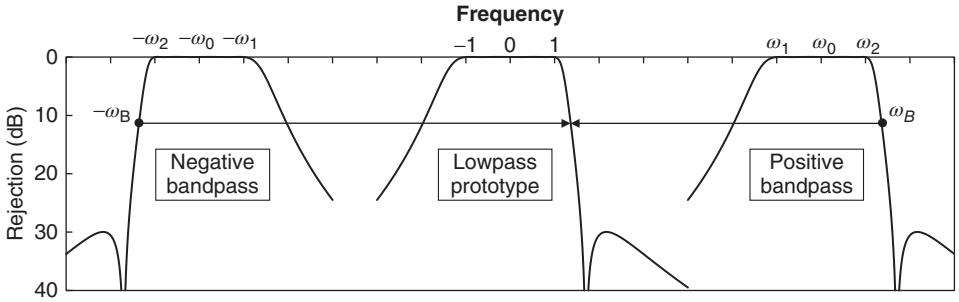


Figure 8.2 Bandpass to lowpass frequency mapping.

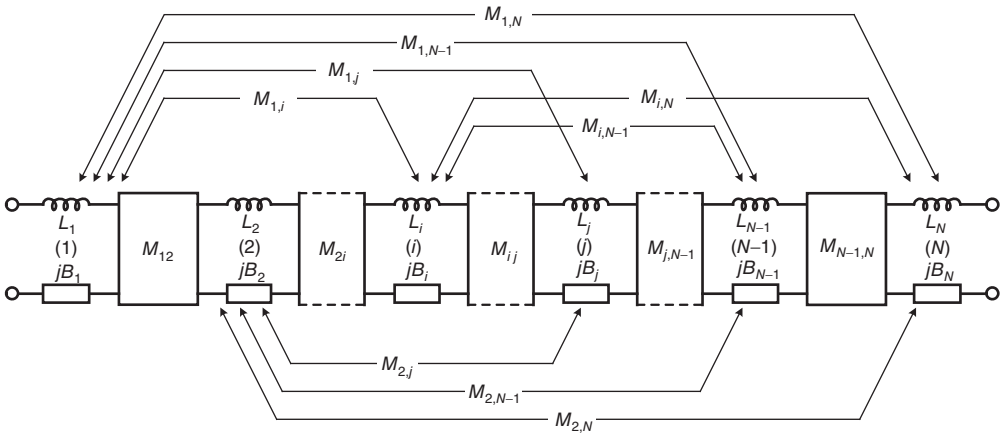


Figure 8.3 Lowpass prototype equivalent of the bandpass network as shown in Figure 8.1b with inverter coupling elements.

elements, and under analysis the insertion loss and rejection amplitudes will be the same; the frequency variables in the lowpass and bandpass domains were being connected by the lumped element frequency mapping formula (8.1).

8.1.2 Formation of the General $N \times N$ Coupling Matrix and Its Analysis

The two-port network of Figure 8.1b (in either its BPP or LPP form) operates between a voltage source generating e_g volts, internal impedance R_S ohms, and a load of R_L ohms. Being a series-resonator circuit with currents circulating in the loops, the overall circuit including the source and load terminations may be represented with an impedance matrix $[z']$ as shown in Figure 8.4.

Kirchhoff's nodal law (the vector sum of all currents entering a node is equal to zero) may be applied to the currents circulating in the series resonators of the circuit shown in Figure 8.1a, which gives rise to a series of equations that may be represented with the matrix equation [1–4]:

$$[e_g] = [z'][i] \tag{8.2}$$

where $[z']$ is the impedance matrix of the N—loop network plus its terminations. Equation (8.2) may be expanded as follows:

$$e_g[1, 0, 0, \ldots, 0]^t = [\mathbf{R} + s\mathbf{I} + j\mathbf{M}] \cdot [i_1, i_2, i_3, \ldots, i_N]^t \tag{8.3}$$

where $[\cdot]^t$ denotes matrix transpose, $\mathbf{I}$ is the unit matrix, e_g is the source voltage, and $i_1, i_2, \ldots, i_N$ are the currents in each of the N loops of the network.

It may be seen that the impedance matrix $[z']$ is itself the sum of three $N \times N$ matrices:

1) *The main coupling matrix $j\boldsymbol{M}$*
This is the $N \times N$ matrix containing the values of the mutual couplings between the nodes of the network (provided by the transformers in Figure 8.1 and equal in value to the immittance inverters in the LPP of Figure 8.3). If it is a coupling between sequentially numbered nodes, $M_{i,i+1}$, it is usually referred to as a "mainline coupling." The entries on the main diagonal $M_{i,i}$ ($\equiv B_i$, the FIR at each node) are the "self-couplings," while all other couplings between nonsequentially numbered nodes are known as "cross-couplings":

$$j\mathbf{M} = j\begin{bmatrix} B_1 & M_{12} & M_{13} & \cdots & M_{1N} \\ M_{12} & B_2 & & & \\ M_{13} & & \ddots & & \\ \vdots & & & \ddots & M_{N-1,N} \\ M_{1N} & & & M_{N-1,N} & B_N \end{bmatrix} \tag{8.4}$$

Because of the reciprocity of the passive network, $M_{ij} = M_{ji}$, and in the general case all entries are nonzero. For the purposes of analysis in the RF domain, any variation of the coupling values with frequency (dispersion) may be included at this stage.

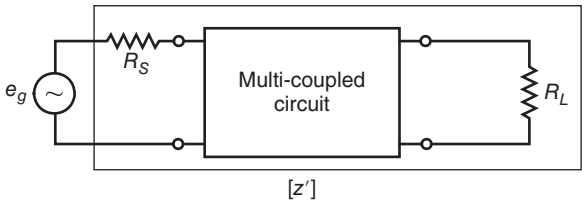


Figure 8.4 Overall impedance matrix $[z']$ of the series-resonator circuits of Figure 8.3 operating between a source impedance R_S and a load impedance R_L.

2) *The frequency variable matrix* $s\mathbf{I}$
This diagonal matrix contains the frequency-variable portion (either LPP or BPP) of the impedance in each loop, giving rise to an $N \times N$ matrix with all entries zero except for the diagonal filled with $s = j\omega$:

$$s\mathbf{I} = \begin{bmatrix} s & 0 & 0 & \cdots & 0 \\ 0 & s & & & \\ 0 & & \ddots & & \\ \vdots & & & \ddots & 0 \\ 0 & & & 0 & s \end{bmatrix} \tag{8.5}$$

For bandpass filters, the effects of a non-infinite resonator unloaded quality factor Q_u may be included by offsetting s by a positive real factor δ (i.e., $s \to \delta + s$), where $\delta = f_0/(BW \cdot Q_u)$, f_0 is the center frequency of the passband, and BW is its design bandwidth.

3) *The termination impedance matrix* $\boldsymbol{R}$
This $N \times N$ matrix contains the values of the source and load impedances in the R_{11} and R_{NN} positions. All other entries are zero:

$$\mathbf{R} = \begin{bmatrix} R_S & 0 & 0 & \cdots & 0 \\ 0 & 0 & & & \\ 0 & & \ddots & & \\ \vdots & & & \ddots & 0 \\ 0 & & & 0 & R_L \end{bmatrix} \tag{8.6}$$

8.1.2.1 "$N \times N$" and "$N + 2$" Coupling Matrices

The $N \times N$ impedance matrix $[z']$ (8.3) for the series-resonator network may be separated out into its purely resistive and purely reactive parts:

$$[z'] = \mathbf{R} + [s\mathbf{I} + j\mathbf{M}] = \mathbf{R} + [z] \tag{8.7}$$

The impedance matrix $[z]$ now represents the circuit as shown Figure 8.5a, a purely reactive network operating between a voltage source with internal impedance R_S and a load R_L.

The source and load terminations are in general nonzero and may be normalized to unity impedance by the insertion of impedance inverters M_{S1} and M_{NL} of impedance values $\sqrt{R_S}$ and $\sqrt{R_L}$, respectively, on the source and load side of the network (Figure 8.5b). In both the cases of Figure 8.5a and b, the impedance looking out from the network on the input side is R_S and on the output side is R_L.

The action of placing the two inverters either side of the $N \times N$ impedance matrix has two effects:

1) The terminating impedances become terminating conductances $G_S = 1/R_S$ and $G_L = 1/R_L$, respectively (and also the voltage source e_g is transformed into a current source $i_g = e_g/R_S$)[2].
2) The $[z]$ matrix surrounded by two inverters may be replaced by the dual network, which will be an admittance matrix $[y]$. The values of the input/output inverters themselves may be absorbed in the $[y]$ matrix by surrounding the $N \times N$ matrix by an extra row on top and bottom and an extra column on each side, so creating an $N + 2 \times N + 2$ matrix, commonly known as the "$N + 2$" matrix (Figure 8.5c). The dual of this network, which will have series

2 Terminating conductances G_S at the source end of the network are normally associated with zero conductance current sources and resistances R_S with zero impedance voltage sources. If it is convenient to do so, they may be interchanged by using the Norton or Thévenin equivalent circuits, as explained in the following.

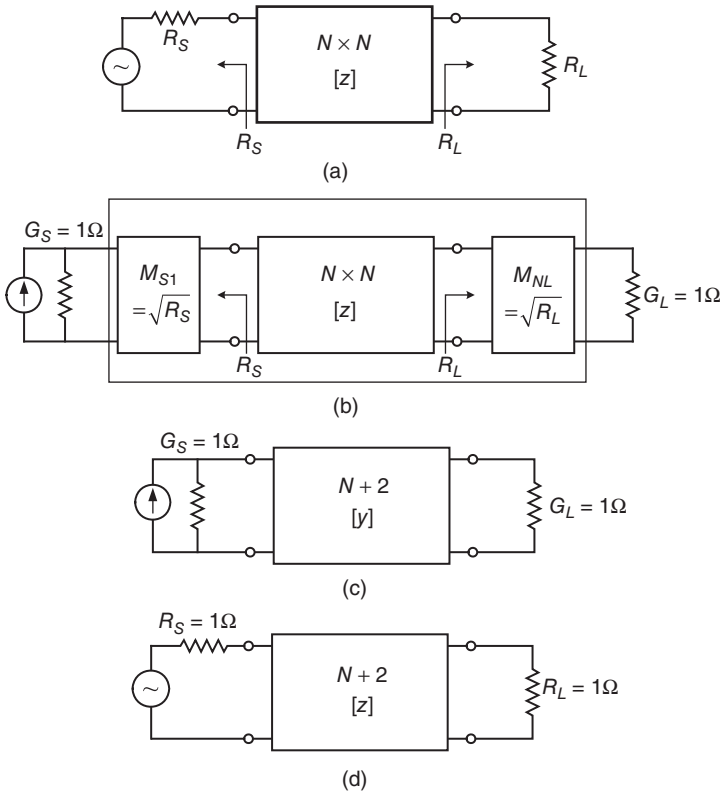
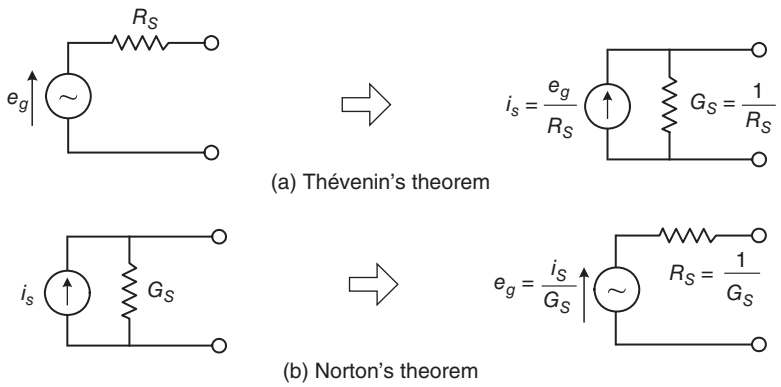


Figure 8.5 Configurations of the input and output circuits for the $N \times N$ and $N + 2$ coupling matrices. (a) The series-resonator circuit in Figure 8.4 represented as an ($N \times N$) impedance coupling matrix between terminations R_S and R_L. (b) The circuit in (a) with inverters to normalize the terminations to unity. (c) $N + 2$ matrix (parallel resonators) and normalized terminating conductances G_S and G_L. (d) $N + 2$ impedance matrix with series resonators and normalized terminating resistances R_S and R_L, the dual network of (c).

resonators and impedance inverter coupling elements, is shown in Figure 8.5d. Whether an impedance or admittance type, the $N + 2$ matrices will have the same values for their mainline and cross-coupling inverters.

The full $N + 2$ network and the corresponding coupling matrix is given in Figures 8.6 and 8.7, showing that in addition to the mainline input/output couplings M_{S1} and M_{NL}, it is also



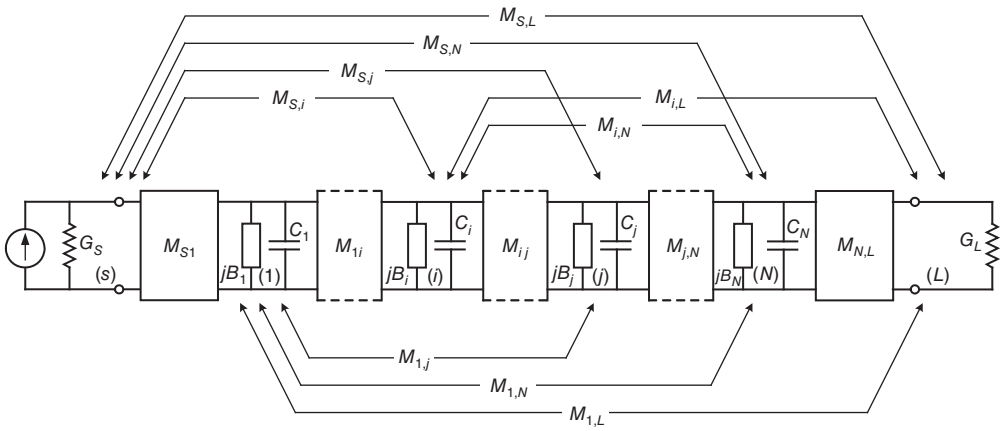


Figure 8.6 $N+2$ multi-coupled network with parallel "lowpass resonators."

Figure 8.7 Fourth-degree $N+2$ coupling matrix with all possible cross-couplings. The "core" $N \times N$ matrix is indicated within the double lines. The matrix is symmetric about the principal diagonal, that is, $M_{ij} = M_{ji}$.

	S	1	2	3	4	L
S	M_{SS}	M_{S1}	M_{S2}	M_{S3}	M_{S4}	M_{SL}
1	M_{S1}	M_{11}	M_{12}	M_{13}	M_{14}	M_{1L}
2	M_{S2}	M_{12}	M_{22}	M_{23}	M_{24}	M_{2L}
3	M_{S3}	M_{13}	M_{23}	M_{33}	M_{34}	M_{3L}
4	M_{S4}	M_{14}	M_{24}	M_{34}	M_{44}	M_{4L}
L	M_{SL}	M_{1L}	M_{2L}	M_{3L}	M_{4L}	M_{LL}

possible now to include other couplings between the source and/or load terminations and internal resonator nodes within the core $N \times N$ matrix. It is also possible to accommodate the direct source–load coupling M_{SL} that will be needed to realize fully canonical filter functions. Being an admittance matrix, it will have parallel resonators and admittance inverters for coupling elements.

8.1.3 Formation of the Coupling Matrix from the Lowpass Prototype Circuit Elements

In the previous chapter, the methods for building up the cross-coupled folded array network from the polynomials representing the desired filter response were demonstrated. The network that was developed comprised an array of parallel nodes (shunt-connected capacitors C_i and FIRs jB_i), interconnected by mainline and cross-coupling inverters. A coupling matrix may be built up directly from the values of these components, but first the frequency-variant

components in the circuit (the capacitors C_i at the nodes) need to be normalized to unity such that the resultant coupling matrix properly corresponds with the circuits in Figure 8.1.

The scaling may be done directly on the coupling matrix once its nonzero elements have been filled with the values of the components from the prototype circuit. Figure 8.7 shows a general "$N+2$" coupling matrix for a fourth-degree network that has all possible couplings present, symmetric about the principal diagonal. Because the circuit synthesis procedure has resulted in parallel "lowpass resonators" (the shunt $sC_i + jB_i$ at the nodes), the matrix is shown as an admittance matrix with conductances $G_S = 1/R_S$ and $G_L = 1/R_L$ at its source and load terminals, respectively. The standard scaling process may now be applied to the matrix, that is, to scale the capacitor C_i at the node $M_{i,i}$ to unity and multiply row i and column i by $[\sqrt{C_i}]^{-1}$ for $i = 1, 2, 3, \ldots, N$:

$$\begin{array}{l} \\ \times(\sqrt{G_S})^{-1} \rightarrow \\ \times(\sqrt{C_1})^{-1} \rightarrow \\ \times(\sqrt{C_2})^{-1} \rightarrow \\ \times(\sqrt{C_3})^{-1} \rightarrow \\ \times(\sqrt{C_4})^{-1} \rightarrow \\ \times(\sqrt{G_L})^{-1} \rightarrow \end{array} \begin{array}{c} \begin{array}{cccccc} \times(\sqrt{G_S})^{-1} & \times(\sqrt{C_1})^{-1} & \times(\sqrt{C_2})^{-1} & \times(\sqrt{C_3})^{-1} & \times(\sqrt{C_4})^{-1} & \times(\sqrt{G_L})^{-1} \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\ \left[\begin{array}{cccccc} G_S + jB_S & jM_{S1} & jM_{S2} & jM_{S3} & jM_{S4} & jM_{SL} \\ jM_{S1} & sC_1 + jB_1 & jM_{12} & jM_{13} & jM_{14} & jM_{1L} \\ jM_{S2} & jM_{12} & sC_2 + jB_2 & jM_{23} & jM_{24} & jM_{2L} \\ jM_{S3} & jM_{13} & jM_{23} & sC_3 + jB_3 & jM_{34} & jM_{3L} \\ jM_{S4} & jM_{14} & jM_{24} & jM_{34} & sC_4 + jB_4 & jM_{4L} \\ jM_{SL} & jM_{1L} & jM_{2L} & jM_{3L} & jM_{4L} & G_L + jB_L \end{array}\right] \end{array} \tag{8.8}$$

By applying the scaling operation, the elements on the diagonal are multiplied by C_i^{-1}, which scales the frequency dependent capacitors and the FIRs to values $sC_i \rightarrow s$ and $jB_i(\equiv M_{ii}) \rightarrow jB_i/C_i$, respectively. The off-diagonal coupling elements M_{ij} are also scaled by the row/column multiplications:

$$M_{ij} \rightarrow \frac{M_{ij}}{\sqrt{C_i \cdot C_j}} \qquad i,\ j = 1,2, \ldots, N, \quad i \neq j \tag{8.9a}$$

The input and output terminations may also be scaled to unity by multiplying the first row and column by $1/\sqrt{G_S}(= \sqrt{R_S})$. Then the M_{SS} element scales to $M_{SS} = G_S + jB_S \rightarrow 1 + jB_S/G_S$, and the mainline coupling from the source to the first resonator node M_{S1} scales

$$M_{S1} \rightarrow \frac{M_{S1}}{\sqrt{G_S C_1}} = M_{S1}\sqrt{\frac{R_S}{C_1}} \tag{8.9b}$$

Similarly, multiplying the last row and column by $1/\sqrt{G_L}(= \sqrt{R_L})$ scales the M_{LL} element to $M_{LL} = G_L + jB_L \rightarrow 1 + jB_L/G_L$, and the mainline coupling from the last resonator node to the load termination M_{NL} scales

$$M_{NL} \rightarrow \frac{M_{NL}}{\sqrt{C_N G_L}} = M_{NL}\sqrt{\frac{R_L}{C_N}} \tag{8.9c}$$

The 4-2 asymmetric filtering function that was used for the demonstration of the synthesis of the prototype electrical network is used again as an example of scaling for the construction of a coupling matrix. C_i, B_i, and M_{ij} that were extracted are firstly summarized and then equation set (8.9) is applied to normalize the C_i to unity and to derive the new values for B_i and M_{ij} (B_i' and M_{ij}', respectively).

$$C_1 = 0.8328 \quad B_1 = j0.1290 \quad M_{12} = 1.0 \quad G_S = 1.0$$
$$C_2 = 1.3032 \quad B_2 = -j0.1875 \quad M_{23} = 0.6310 \quad G_L = 1.0$$
$$C_3 = 3.7296 \quad B_3 = -j3.4499 \quad M_{34} = 1.0$$
$$C_4 = 0.8328 \quad B_4 = j0.1290 \quad M_{14} = 0.3003 \quad M_{S1} = 1.0$$
$$M_{24} = 0.8066 \quad M_{4L} = 1.0$$

Now applying equation set (8.9), a new set of C_i, B_i, and M_{ij} are obtained:

$$C'_1 = 1.0 \quad B'_1 = j0.1549 \quad M'_{12} = 0.9599 \quad G'_S = 1.0$$
$$C'_2 = 1.0 \quad B'_2 = -j0.1439 \quad M'_{23} = 0.2862 \quad G'_L = 1.0$$
$$C'_3 = 1.0 \quad B'_3 = -j0.9250 \quad M'_{34} = 0.5674$$
$$C'_4 = 1.0 \quad B'_4 = j0.1549 \quad M'_{14} = 0.3606 \quad M'_{S1} = 1.0958$$
$$M'_{24} = 0.7742 \quad M'_{4L} = 1.0958$$

and the coupling matrix for the 4-2 network may be constructed. The same scaling process, this time by −1.0, may then be used to change the signs of any negative mainline couplings to positive if required:

$$\mathbf{M} = \begin{array}{c} \\ S \\ 1 \\ 2 \\ 3 \\ 4 \\ L \end{array} \begin{array}{c} \begin{array}{cccccc} S & 1 & 2 & 3 & 4 & L \end{array} \\ \begin{bmatrix} 0.0 & 1.0958 & 0 & 0 & 0 & 0 \\ 1.0958 & 0.1549 & 0.9599 & 0 & 0.3606 & 0 \\ 0 & 0.9599 & -0.1439 & 0.2862 & 0.7742 & 0 \\ 0 & 0 & 0.2862 & -0.9250 & 0.5674 & 0 \\ 0 & 0.3606 & 0.7742 & 0.5674 & 0.1549 & 1.0958 \\ 0 & 0 & 0 & 0 & 1.0958 & 0.0 \end{bmatrix} \end{array} \tag{8.10}$$

Because all the capacitors C_i have been scaled to unity, the diagonal frequency matrix $s\mathbf{I}$ [see (8.5)] is now defined, and the termination matrix $\mathbf{R} = 0$ except for $R_{SS} = G_S = 1.0$ and $R_{LL} = G_L = 1.0$ [see (8.6)]. Now the $N + 2$ immittance matrix for this asymmetric 4-2 network may be constructed:

$$[y'] \text{ or } [z'] = \mathbf{R} + s\mathbf{I} + j\mathbf{M} \tag{8.11}$$

8.1.4 Analysis of the Network Represented by the Coupling Matrix

The network as represented by the coupling matrix may be analyzed in two ways:

1) As an [*ABCD*] matrix that can then be cascaded with the [*ABCD*] matrices of other components to form a compound network, for example, in a manifold multiplexer where the matrices of the different channel filters may be interconnected via lengths of transmission line, waveguide, coaxial, and so on
2) As a standalone network, incorporating the terminating impedances R_S and R_L

The network relevant to the first case is shown in Figure 8.8 where R_S and R_L have been separated out from the main network. Writing out (8.3) in full for this case ($N \times N$ matrix)[3]:

3 If working with the "$N + 2$" matrix, the frequency variable s is not entered into the top left and bottom right corners of the matrix, because these will be "nonresonant nodes" (NRNs).

$$\begin{bmatrix} e_g \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} R_S & 0 & 0 & \dots & 0 \\ 0 & 0 & & & \\ 0 & & \ddots & & \\ \vdots & & & \ddots & 0 \\ 0 & & & 0 & R_L \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ \\ i_N \end{bmatrix}$$

$$+ \begin{bmatrix} s + jM_{11} & jM_{12} & jM_{13} & \dots & jM_{1N} \\ jM_{12} & s + jM_{22} & & & \\ jM_{13} & & \ddots & & \\ \vdots & & & \ddots & jM_{N-1,N} \\ jM_{1N} & & & jM_{N-1,N} & s + jM_{NN} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ \\ i_N \end{bmatrix} \tag{8.12}$$

$$\begin{bmatrix} e_g - R_S i_1 \\ 0 \\ \vdots \\ \\ -R_L i_N \end{bmatrix} = \begin{bmatrix} s + jM_{11} & jM_{12} & jM_{13} & \dots & jM_{1N} \\ jM_{12} & s + jM_{22} & & & \\ jM_{13} & & \ddots & & \\ \vdots & & & \ddots & jM_{N-1,N} \\ jM_{1N} & & & jM_{N-1,N} & s + jM_{NN} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ \\ i_N \end{bmatrix} \tag{8.13}$$

Since $e_g - R_S i_1 = v_1$ and $-R_L i_N = v_N$ (see Figure 8.8), it may be seen that the coupling matrix here is in the form of an open-circuit impedance matrix $[z]$ [5]:

$$\begin{bmatrix} v_1 \\ 0 \\ \vdots \\ \\ v_N \end{bmatrix} = \begin{bmatrix} \\ [z] \\ \\ \\ \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ \\ i_N \end{bmatrix} \tag{8.14}$$

Inverting the $[z]$ matrix will give the currents in terms of the short-circuit admittance matrix $[y]$:

$$\begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ \vdots \\ i_N \end{bmatrix} = \begin{bmatrix} \\ [y] \\ (= [z]^{-1}) \\ \\ \end{bmatrix} \begin{bmatrix} v_1 \\ 0 \\ 0 \\ \vdots \\ v_N \end{bmatrix} \tag{8.15}$$

Since we are only interested at the moment in the voltages and currents at the terminals for this analysis, we may rewrite (8.15):

$$\begin{bmatrix} i_1 \\ i_N \end{bmatrix} = \begin{bmatrix} y_{11} & y_{1N} \\ y_{N1} & y_{NN} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_N \end{bmatrix} \tag{8.16}$$

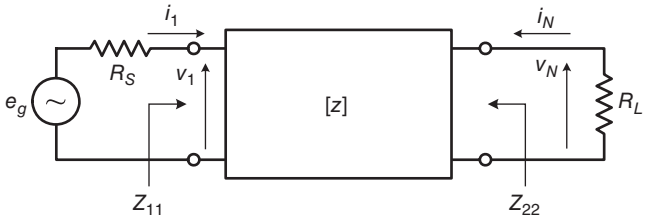


Figure 8.8 Network between source e_g of impedance R_S and load of impedance R_L.

where y_{11}, y_{1N}, and so on are the corner elements of the admittance matrix $[y]$. This may be converted into $[ABCD]$ parameters using the standard $[y] \rightarrow [ABCD]$ parameter transform [5], including the normalization to the source and load impedance values:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{-1}{y_{N1}} \begin{bmatrix} \sqrt{\frac{R_L}{R_S}} y_{NN} & \frac{1}{\sqrt{R_S R_L}} \\ \Delta_{[y]} \sqrt{R_S R_L} & \sqrt{\frac{R_S}{R_L}} y_{11} \end{bmatrix} \tag{8.17}$$

where $\Delta_{[y]}$ is the determinant of the submatrix in (8.16) given by $\Delta_{[y]} = y_{11}y_{NN} - y_{1N}y_{N1}$. If rapid analysis is required, for example, during a real-time parameter de-embedding process when a coupling matrix is being optimized to correspond to a measured characteristic, it may be more efficient to solve (8.14) using a Gaussian elimination procedure [6] rather than performing a full inversion on the $[z]$ matrix. In this case we shall be finding the short-circuit z-parameters z_{11}, z_{1N}, z_{N1}, and z_{NN}, and the equivalent $[ABCD]$ matrix is given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{z_{N1}} \begin{bmatrix} \sqrt{\frac{R_L}{R_S}} z_{11} & \frac{\Delta_{[z]}}{\sqrt{R_S R_L}} \\ \sqrt{R_S R_L} & \sqrt{\frac{R_S}{R_L}} z_{NN} \end{bmatrix} \tag{8.18}$$

where $\Delta_{[z]} = z_{11}z_{NN} - z_{1N}z_{N1}$. If the values of the input/output coupling inverters are used instead, ($M_{S1} = \sqrt{R_s}$, $M_{NL} = \sqrt{R_L}$).

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{z_{N1}} \begin{bmatrix} \frac{M_{NL}}{M_{S1}} z_{11} & \frac{\Delta_{[z]}}{M_{S1}M_{NL}} \\ M_{S1}M_{NL} & \frac{M_{S1}}{M_{NL}} z_{NN} \end{bmatrix} \tag{8.19}$$

Now the matrix may be cascaded with the $[ABCD]$ matrices of other networks, for example, transmission lines.

8.1.5 Direct Analysis

For the second analysis case, the full coupling matrix including the source–load terminations is used, as in (8.3):

$$[e_g] = [z'] \cdot [i] \quad \text{or} \quad [i] = [z']^{-1}[e_g] = [y'][e_g] \tag{8.20}$$

where $[z']$ and $[y']$ are the network open-circuit impedance and short-circuit admittance matrices, respectively, with the source and load impedances included. Referring to Figure 8.8 and (8.20), it may be easily seen that

$$i_1 = [y']_{11} e_g \tag{8.21a}$$

$$i_N = [y']_{N1} e_g = \frac{v_N}{R_L} \tag{8.21b}$$

Substituting (8.21b) into the "transmission coefficient" definition for S_{21} [5]:

$$\begin{aligned} S_{21} &= 2\sqrt{\frac{R_S}{R_L}} \cdot \frac{v_N}{e_g} = 2\sqrt{\frac{R_S}{R_L}} \cdot R_L [y']_{N1} \\ &= 2\sqrt{R_S R_L} \cdot [y']_{N1} \end{aligned} \tag{8.22}$$

The reflection coefficient at the input port is given by

$$\begin{aligned} S_{11} &= \frac{Z_{11} - R_S}{Z_{11} + R_S} \\ &= \frac{Z_{11} + R_S - 2R_S}{Z_{11} + R_S} \\ &= 1 - \frac{2R_S}{Z_{11} + R_S} \end{aligned} \tag{8.23}$$

where $Z_{11} = v_1/i_1$ is the impedance looking in at the input port (see Figure 8.8). The potential divider at the input port gives v_1 and (8.21a) gives i_1, and so Z_{11} may be expressed:

$$Z_{11} = \frac{v_1}{i_1} = \frac{e_g Z_{11}}{Z_{11} + R_S} \cdot \frac{1}{e_g [y']_{11}}$$

or

$$\frac{1}{Z_{11} + R_S} = [y']_{11} \tag{8.24}$$

Substituting into (8.23) finally gives the reflection coefficient at the input port of the network:

$$S_{11} = 1 - 2R_S [y']_{11} \tag{8.25a}$$

Similarly for the output port,

$$S_{22} = 1 - 2R_L [y']_{NN} \tag{8.25b}$$

The voltage v_1 incident at the network input (see Figure 8.8) may be related to the generator voltage e_g by substituting for $[y']_{11}$ in (8.25a) using (8.21a):

$$v_1 = (e_g - R_S i_1) = e_g (1 + S_{11})/2 \tag{8.26a}$$

This shows that at a point of perfect transmission ($S_{11} = 0$), $v_1 = e_g/2$, meaning that the impedance looking into the network at the input side $Z_{11} = R_S$. At this point the maximum power available from the generator is being transferred to the load, and if e_g is set to 2 V and the source impedance $R_S = 1\ \Omega$, the power P_i incident at the input will be 1 W. Also, if e_g in (8.26a) is substituted using (8.22), v_N may be directly related to v_1 through the network S-parameters:

$$\frac{v_N}{v_1} = \sqrt{\frac{R_L}{R_S}} \cdot \frac{S_{21}}{(1 + S_{11})} \tag{8.26b}$$

8.2 Direct Synthesis of the Coupling Matrix

In this section, two methods for the direct synthesis of the coupling matrix for a network with given transfer and reflection characteristics will be presented: the first for the $N \times N$ matrix and the second for the $N + 2$ matrix. In both cases the approach will be the same, namely, to formulate the two-port short-circuit admittance parameters in two ways: firstly from the coefficients of the polynomials $F(s)/\varepsilon_R$, $P(s)/\varepsilon$, and $E(s)$ that make up the desired transfer and reflection characteristics $S_{21}(s)$ and $S_{11}(s)$ and secondly from the elements of the coupling matrix itself. Equating the two formulations will then relate the coupling values of the matrix to the coefficients of the transfer and reflection polynomials.

Although the $N + 2$ matrix is more flexible than the $N \times N$ matrix and also happens to be a little easier to synthesize, it is informative to explore the principles behind the $N \times N$ matrix synthesis technique, and we shall go through this method before proceeding to the synthesis of the $N + 2$ matrix.

8.2.1 Direct Synthesis of the $N \times N$ Coupling Matrix

Recalling now the two-port short-circuit admittance matrix for the overall network that was defined in (8.16) [4]:

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \frac{1}{y_d} \begin{bmatrix} y_{11n} & y_{12n} \\ y_{21n} & y_{22n} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{8.27}$$

where y_{11}, y_{12}, and so on are the corner elements of the admittance matrix $[y]$, itself the inverse of the impedance matrix $[z] = s\mathbf{I} + j\mathbf{M}$ [equation (8.7)]. Referring to Figure 8.8 and taking the standard definitions for the $[y]$ matrix elements [5], the two-port y-parameters may be written in terms of the coupling matrix $\mathbf{M}$ and the frequency variable $s = j\omega$:

$$y_{11}(s) = \frac{y_{11n}(s)}{y_d(s)} = [z]_{11}^{-1} = \left.\frac{i_1}{v_1}\right|_{v_N=0} = [s\mathbf{I} + j\mathbf{M}]_{11}^{-1} = -j[\omega\mathbf{I} + \mathbf{M}]_{11}^{-1} \tag{8.28a}$$

$$y_{22}(s) = \frac{y_{22n}(s)}{y_d(s)} = [z]_{NN}^{-1} = \left.\frac{i_N}{v_N}\right|_{v_1=0} = [s\mathbf{I} + j\mathbf{M}]_{NN}^{-1} = -j[\omega\mathbf{I} + \mathbf{M}]_{NN}^{-1} \tag{8.28b}$$

$$y_{12}(s) = y_{21}(s) = \frac{y_{21n}(s)}{y_d(s)} = [z]_{N1}^{-1} = \left.\frac{i_N}{v_1}\right|_{v_N=0} = [s\mathbf{I} + j\mathbf{M}]_{N1}^{-1} = -j[\omega\mathbf{I} + \mathbf{M}]_{N1}^{-1} \tag{8.28c}$$

This is the essential step in the network synthesis procedure that relates the transfer function expressed in purely mathematical terms [i.e., $S_{11}(s)$, $y_{21}(s)$, etc. expressed as rational polynomials] to the real world of the coupling matrix, each element of which corresponds uniquely to a physical coupling element in the realized filter.

From equation (8.28), it may be seen that the characteristic polynomial corresponding to the matrix $-\mathbf{M}$ is the common denominator polynomial $y_d(s)$ of the elements of the admittance matrix $[y]$ [6]. Also, since $\mathbf{M}$ is real and symmetric about its principal diagonal, all of its eigenvalues are real [1]. Thus an $N \times N$ matrix $\mathbf{T}$ with rows of orthogonal vectors exists, which satisfies the equation

$$-\mathbf{M} = \mathbf{T} \cdot \mathbf{\Lambda} \cdot \mathbf{T}^t \tag{8.29}$$

where $\mathbf{\Lambda} = \text{diag}[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_N]$, λ_i are the eigenvalues of $-\mathbf{M}$ and $\mathbf{T}^t$ is the transpose of $\mathbf{T}$ such that $\mathbf{T} \cdot \mathbf{T}^t = \mathbf{I}$ [6]. The numerator and denominator polynomials of $y_{ij}(s)$ for the single- and double-terminated cases have already been derived from the coefficients of $E(s)$, $F(s)/\varepsilon_R$, and $P(s)/\varepsilon$ polynomials, which define the desired transfer and reflection characteristics of the network [equation sets (7.20), (7.25), and (7.31)]. In fact only two of the y-parameters are required, and we choose $y_{21}(s)$ and $y_{22}(s)$ because $y_{22}(s)$ results directly from the $A(s)$ and $B(s)$ polynomials in the case of the single-terminated network [the $C(s)$ and $D(s)$ polynomials are not needed]. Substituting (8.29) into (8.28b) and (8.28c) yields

$$y_{21}(s) = -j[\omega\mathbf{I} - \mathbf{T} \cdot \mathbf{\Lambda} \cdot \mathbf{T}^t]_{N1}^{-1} \tag{8.30a}$$

$$y_{22}(s) = -j[\omega\mathbf{I} - \mathbf{T} \cdot \mathbf{\Lambda} \cdot \mathbf{T}^t]_{NN}^{-1} \tag{8.30b}$$

Knowing that $\mathbf{T} \cdot \mathbf{T}^t = \mathbf{I}$, the right-hand sides (RHSs) of equation (8.30) may be developed as follows [7]:

$$\begin{aligned} [\omega\mathbf{I} - \mathbf{T} \cdot \mathbf{\Lambda} \cdot \mathbf{T}^t]^{-1} &= [\mathbf{T} \cdot (\mathbf{T}^t \cdot (\omega\mathbf{I}) \cdot \mathbf{T} - \mathbf{\Lambda}) \cdot \mathbf{T}^t]^{-1} \\ &= (\mathbf{T}^t)^{-1} \cdot (\mathbf{T}^t \cdot (\omega\mathbf{I}) \cdot \mathbf{T} - \mathbf{\Lambda})^{-1} \cdot (\mathbf{T})^{-1} \\ &= \mathbf{T} \cdot (\omega\mathbf{I} - \mathbf{\Lambda})^{-1} \cdot \mathbf{T}^t \\ &= \mathbf{T} \cdot \text{diag}\left(\frac{1}{\omega - \lambda_1}, \frac{1}{\omega - \lambda_2}, \ldots, \frac{1}{\omega - \lambda_N}\right) \cdot \mathbf{T}^t \end{aligned}$$

This leads to the general solution for an element i, j of an inverse eigenmatrix problem such as the RHSs of equation (8.30)

$$[\omega\mathbf{I} - \mathbf{T}\cdot\boldsymbol{\Lambda}\cdot\mathbf{T}^t]^{-1}_{ij} = \sum_{k=1}^{N}\frac{T_{ik}T_{jk}}{\omega-\lambda_k} \quad i,j = 1,2,\ldots,N \tag{8.31}$$

Therefore from equation (8.30),

$$y_{21}(s) = \frac{y_{21n}(s)}{y_d(s)} = -j\sum_{k=1}^{N}\frac{T_{Nk}T_{1k}}{\omega-\lambda_k} \tag{8.32a}$$

and

$$y_{22}(s) = \frac{y_{22n}(s)}{y_d(s)} = -j\sum_{k=1}^{N}\frac{T_{Nk}^2}{\omega-\lambda_k} \tag{8.32b}$$

Equations (8.32) show that the eigenvalues λ_k of $-\mathbf{M}$, multiplied by j, are also the roots of the denominator polynomial $y_d(s)$, which is common to the rational polynomial admittance functions $y_{21}(s)$ and $y_{22}(s)$. So now we can develop the first and last rows T_{1k} and T_{Nk} of the orthogonal matrix $\mathbf{T}$ by equating the residues of $y_{21}(s)$ and $y_{22}(s)$ with $T_{1k}T_{Nk}$ and T_{Nk}^2, respectively, at corresponding eigenvalue poles λ_k. Since the numerator and denominator polynomials of $y_{21}(s)$ and $y_{22}(s)$ have already been determined from the transfer and reflection polynomials of the desired filtering characteristic [equations (7.20), (7.25) for double-terminated networks, (7.29), (7.31) for single-terminated], the residues r_{21k} and r_{22k} may be calculated using partial fraction expansions[4]:

$$y_{21}(s) = \frac{y_{21n}(s)}{y_d(s)} = -j\sum_{k=1}^{N}\frac{r_{21k}}{\omega-\lambda_k}, \quad y_{22}(s) = \frac{y_{22n}(s)}{y_d(s)} = -j\sum_{k=1}^{N}\frac{r_{22k}}{\omega-\lambda_k}$$

Referring back to equation (8.32), it may now be seen that

$$\begin{aligned} T_{Nk} &= \sqrt{r_{22k}} \\ T_{1k} &= \frac{r_{21k}}{T_{Nk}} = \frac{r_{21k}}{\sqrt{r_{22k}}}, \quad k = 1,2,\ldots,N \end{aligned} \tag{8.33}$$

Starting with equation (8.28a) and following the same procedure, it may be shown that $T_{1k} = \sqrt{r_{11k}}$ where r_{11k} are the residues of $y_{11}(s)$. From this and equation (8.33), it may be seen that $T_{1k} = \sqrt{r_{11k}} = r_{21k}/\sqrt{r_{22k}}$, confirming the relationship $r_{21k}^2 = r_{11k}r_{22k}$ for a realizable network [8]. Having derived T_{1k} from $y_{21}(s)$ and $y_{22}(s)$ (equation (8.33)), the numerator polynomial of $y_{11}(s) = y_{11n}(s)/y_d(s)$ may be constructed, if required.

The network may be directly connected between the terminating resistances R_S and R_L, which in general will be non-unity in value. If it is desired to scale the terminating impedances to 1 Ω, input/output inverter values M_{S1} and M_{NL} may be found by scaling the magnitudes of the row vectors T_{1k} and T_{Nk} to unity for the "inner" network of Figure 8.5b:

$$M_{S1}^2 = R_S = \sum_{k=1}^{N}T_{1k}^2, \quad M_{NL}^2 = R_L = \sum_{k=1}^{N}T_{Nk}^2 \tag{8.34}$$

4 The residues of rational polynomials may be calculated as follows [6]:

$$r_{21k} = \left.\frac{y_{21n}(s)}{y_d'(s)}\right|_{s=j\lambda_k}, \; r_{22k} = \left.\frac{y_{22n}(s)}{y_d'(s)}\right|_{s=j\lambda_k}, \quad k = 1,2,\ldots,N$$

where $j\lambda_k$ are the roots of $y_d(s)$ and $y_d'(s)$ denotes differentiation of the polynomial $y_d(s)$ with respect to s.

Then

$$T_{1k} \rightarrow \frac{T_{1k}}{M_{S1}} \quad \text{and} \quad T_{Nk} \rightarrow \frac{T_{Nk}}{M_{NL}}$$

M_{S1} and M_{NL} are equivalent to the turns ratios n_1 and n_2 of two transformers at the source and load ends of the network, respectively, matching the terminating impedances with the internal network [4].

With the first and last rows of **T** now determined, the remaining orthogonal rows may be constructed with the Gram–Schmidt orthogonalization (GSO) process or similar [9], and finally the coupling matrix **M** synthesized using equation (8.29).

8.3 Coupling Matrix Reduction

The elements of the coupling matrix **M** that emerges from the synthesis procedure described in Section 8.2 will in general all have nonzero values (e.g., Figure 8.7). The nonzero values that will occur in the diagonal elements of the coupling matrices for electrically asymmetric networks represent the offsets from center frequency of each resonance (asynchronously tuned). Nonzero entries everywhere else means that in the network that **M** represents, couplings exist between every resonator or termination node and every other resonator or termination node. As this is clearly impractical, it is usual to annihilate couplings with a sequence of similarity transforms (sometimes called rotations) [10–12] until a more convenient form with a minimal number of couplings is obtained. The use of similarity transforms ensures that the eigenvalues and eigenvectors of the matrix **M** are preserved, such that under analysis the transformed matrix will yield exactly the same transfer and reflection characteristics as the original matrix.

There are several more practical canonical forms for the transformed coupling matrix **M**. Two of the better-known forms are the "arrow" form [13] and the more generally useful "folded" form [14, 15] (Figure 8.9). Either of these canonical forms may be used directly if it is convenient to realize the couplings or be used as a starting point for the application of further transforms to create an alternative resonator intercoupling topology, optimally adapted for the physical and electrical constraints of the technology with which the filter will eventually be realized, for example [16, 17]. The method for reduction of the coupling matrix to the folded form will be described here. The "arrow" form may be derived using a very similar method.

8.3.1 Similarity Transformation and Annihilation of Matrix Elements

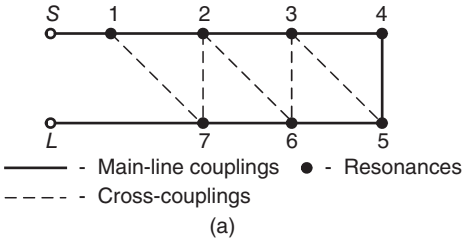
A similarity transform (or rotation) on an $N \times N$ coupling matrix $\mathbf{M}_0$ is carried out by pre- and post-multiplying $\mathbf{M}_0$ by an $N \times N$ rotation matrix **R** and its transpose $\mathbf{R}^t$ [10, 11]:

$$\mathbf{M}_1 = \mathbf{R}_1 \cdot \mathbf{M}_0 \cdot \mathbf{R}_1^t \tag{8.35}$$

where $\mathbf{M}_0$ is the original matrix, $\mathbf{M}_1$ is the matrix after the transform operation, and the rotation matrix **R** is defined as in Figure 8.10. The pivot $[i, j]$ $(i \neq j)$ of $\mathbf{R}_r$ means that elements $R_{ii} = R_{jj} = \cos\theta_r$, $R_{ji} = -R_{ij} = \sin\theta_r$ $(i, j \neq 1 \text{ or } N)$, and θ_r is the angle of the rotation. The other principal diagonal entries are 1 and all other off-diagonal entries are zero.

The eigenvalues of the matrix $\mathbf{M}_1$ after the transform are exactly the same as those of the original matrix $\mathbf{M}_0$, which means that an arbitrarily long series of transforms with arbitrarily defined pivots and angles may be applied, starting with $\mathbf{M}_0$. Each transform in the series takes the form

$$\mathbf{M}_r = \mathbf{R}_r \cdot \mathbf{M}_{r-1} \cdot \mathbf{R}_r^t \quad r = 1, 2, 3, \ldots, R \tag{8.36}$$



(a)

	1	2	3	4	5	6	7
1	s	m					xa
2	.	s	m			xa	xs
3	.	.	s	m	xa	xs	
4	.	.	.	s	m		
5	.	.	.	.	s	m	
6	.	.	.	.	.	s	m
7	.	.	.	.	.	.	s

Possible non-zero couplings:

s	–	Self coupling
m	–	Main-line coupling
xa	–	Asymmetric cross-coupling
xs	–	Symmetric cross-coupling
.	–	Couplings are equal-valued about the principal diagonal

(b)

Figure 8.9 $N \times N$ folded canonical network coupling matrix form—seventh-degree example. "s" and "xa" couplings are zero for symmetric characteristics.

	1	2	3	4	5	6	7
1	1						
2		1					
3			c_r		$-s_r$		
4				1			
5			s_r		c_r		
6						1	
7							1

$c_r \equiv \cos\theta_r$, $s_r \equiv \sin\theta_r$

Figure 8.10 Example of seventh-degree rotation matrix $\mathbf{R}_r$—pivot [3, 5], angle θ_r.

and under analysis the matrix $\mathbf{M}_R$ resultant at the end of the series of transforms will yield exactly same performance as the original matrix $\mathbf{M}_0$.

When a similarity transform of pivot $[i, j]$ and angle $\theta_r(\neq 0)$ is applied to a coupling matrix $\mathbf{M}_{r-1}$, the elements in rows i and j and columns i and j of the resultant matrix $\mathbf{M}_r$ change in value from the corresponding element values in $\mathbf{M}_{r-1}$. For the kth element in the row or column i or j of $\mathbf{M}_r$, and not on the cross-points of the pivot (i.e., $k \neq i, j$), the value will change according to the formulas:

$$\begin{aligned} M'_{ik} &= c_r M_{ik} - s_r M_{jk} \quad \text{For an element in row } i \\ M'_{jk} &= s_r M_{ik} + c_r M_{jk} \quad \text{For an element in row } j \\ M'_{ki} &= c_r M_{ki} - s_r M_{kj} \quad \text{For an element in column } i \\ M'_{kj} &= s_r M_{ki} + c_r M_{kj} \quad \text{For an element in column } j \end{aligned} \tag{8.37a}$$

	1	2	3	4	5	6	7
1	s	m	④	③	②	①	xa
2	.	s	m	⑨	⑧	xa	xs
3	.	.	s	m	xa	xs	⑤
4	.	.	.	s	m	⑩	⑥
5	.	.	.	.	s	m	⑦
6	.	.	.	.	.	s	m
7	.	.	.	.	.	.	s

③ etc – Coupling annihilated as a result of 3rd etc. transform in the sequence.

Figure 8.11 Seventh-degree coupling matrix: reduction sequence for folded canonical form. The shaded elements are those that may be affected by a similarity transform at pivot [3, 5], angle θ_r ($\neq 0$). All others remain unchanged.

where $k\,(\neq i, j) = 1, 2, 3, \ldots, N$, $c_r = \cos\theta_r, s_r = \sin\theta_r$, and the undashed matrix elements belong to the matrix $\mathbf{M}_{r-1}$, while the dashed to $\mathbf{M}_r$. For the elements on the cross-points of the pivot $[M_{ii}, M_{jj}, M_{ij}\,(=M_{ji})]$,

$$\begin{aligned} M'_{ii} &= c_r^2 M_{ii} - 2s_r c_r M_{ij} + s_r^2 M_{jj} \\ M'_{jj} &= s_r^2 M_{ii} + 2s_r c_r M_{ij} + c_r^2 M_{jj} \\ M'_{ij} &= M_{ij}(c_r^2 - s_r^2) + s_r c_r (M_{ii} - M_{jj}) \end{aligned} \tag{8.37b}$$

Two properties of a similarity transform that will be exploited for the matrix reduction process to be described in the following are noted here:

1) *Only* those elements in the rows and columns i and j of the pivot $[i, j]$ of a transform will be affected by the transform (provided its angle $\theta_r \neq 0$). All others will remain at their previous values.
2) If two elements facing each other across the rows and columns of the pivot of a transform are both zero before the application of the transform, they will still be zero after the transform. For example, if the elements M_{13} and M_{15} in Figure 8.11 happen to be both zero before the transform with pivot [3, 5], they will still be zero after the transform, regardless of the transform angle θ_r.

The equations in (8.37) may be used to annihilate (zero) specific elements in the coupling matrix. For example, to annihilate the nonzero element M_{15} (and simultaneously M_{51}) in the seventh-degree coupling matrix of Figure 8.11, a transform of pivot [3, 5] and angle $\theta_1 = -\tan^{-1}(M_{15}/M_{13})$ are applied to the coupling matrix (see last formula in (8.37a) with $k = 1$, $i = 3$, $j = 5$). In the transformed matrix, M'_{15} and M'_{51} will be zero and all values in rows and columns 3 and 5 (shaded in Figure 8.11) could be changed. Equation set (8.38) summarizes the angle formulas for annihilating specific elements in the coupling matrix with a rotation at pivot $[i, j]$:

$$\theta_r = \tan^{-1}(M_{ik}/M_{jk}) \quad \text{For the } k\text{th element in row } i\ (M_{ik}) \tag{8.38a}$$

$$\theta_r = -\tan^{-1}(M_{jk}/M_{ik}) \quad \text{For the } k\text{th element in row } j\ (M_{jk}) \tag{8.38b}$$

$$\theta_r = \tan^{-1}(M_{ki}/M_{kj}) \quad \text{For the } k\text{th element in column } i\ (M_{ki}) \tag{8.38c}$$

$$\theta_r = -\tan^{-1}(M_{kj}/M_{ki}) \quad \text{For the } k\text{th element in column } j\ (M_{kj}) \tag{8.38d}$$

$$\theta_r = \tan^{-1}\left(\frac{M_{ij} \pm \sqrt{M_{ij}^2 - M_{ii}M_{jj}}}{M_{jj}}\right) \qquad \text{For cross-pivot element } (M_{ii}) \tag{8.38e}$$

$$\theta_r = \tan^{-1}\left(\frac{-M_{ij} \pm \sqrt{M_{ij}^2 - M_{ii}M_{jj}}}{M_{ii}}\right) \qquad \text{For cross-pivot element } (M_{jj}) \tag{8.38f}$$

$$\theta_r = \frac{1}{2}\tan^{-1}\left(\frac{2M_{ij}}{(M_{jj} - M_{ii})}\right) \qquad \text{For cross-pivot element } (M_{ij}) \tag{8.38g}$$

The method for reducing the full coupling matrix $\mathbf{M}_0$ resultant from the synthesis procedure of Section 8.2 to the folded form of Figure 8.9 involves applying a series of similarity transforms to the matrix that progressively annihilates unrealizable or inconvenient elements one by one. The transforms are applied in a certain order and pattern that makes use of the two effects mentioned above, ensuring that once annihilated, an element is not regenerated by a subsequent transform in the sequence.

8.3.1.1 Reduction Procedure: Full Coupling Matrix to Folded Canonical Form

There are a number of transform sequences that will reduce the full coupling matrix to the folded form. The sequence used here involves alternately annihilating elements right to left along rows and top to bottom down columns as shown in the seventh-degree example in Figure 8.11, starting with the element in the first row and the $(N-1)$th column (M_{16}).

M_{16} is annihilated with a transform of pivot [5, 6] and angle $\theta_1 = -\tan^{-1}(M_{16}/M_{15})$. This is followed by a second transform, pivot [4, 5], angle $\theta_2 = -\tan^{-1}(M_{15}/M_{14})$, which will annihilate element M_{15}. The previously annihilated element M_{16} is unaffected by this transform because it is lying outside the rows and columns of the latest pivot and remains at zero. Now third and fourth transforms at pivots [3, 4] and [2, 3], and angles $\theta_3 = -\tan^{-1}(M_{14}/M_{13})$ and $\theta_4 = -\tan^{-1}(M_{13}/M_{12})$ will annihilate M_{14} and M_{13}, respectively, again without disturbing the previously annihilated elements.

After these four transforms, the elements in the first row of the matrix between the mainline coupling M_{12} and the element in the last (seventh) column will be zero. Because of the symmetry about the principal diagonal, the elements between M_{21} and M_{71} in the first column will also be zero.

Next, the three elements in column 7, M_{37}, M_{47}, and M_{57}, are annihilated with transforms at pivots [3, 4], [4, 5], and [5, 6] and angles $\tan^{-1}(M_{37}/M_{47})$, $\tan^{-1}(M_{47}/M_{57})$, and $\tan^{-1}(M_{57}/M_{67})$, respectively [see (8.37a)]. As with the rows, the columns are cleared down to the first mainline coupling encountered in that column. The couplings M_{13}, $M_{14,}$ M_{15}, and M_{16} annihilated in the first sweep will remain at zero, because they face each other across the pivot columns of the transforms of the second sweep and will therefore be unaffected.

Continuing on, a third sweep along row 2 annihilates M_{25} and M_{24} in that order, and the final sweep annihilates M_{46} in column 6. At this point it may be seen that the form of the folded canonical coupling matrix has been achieved (Figure 8.9), with its two cross-diagonals containing the symmetric and asymmetric cross-couplings. Table 8.1 summarizes the whole annihilation procedure.

The final values and positions of the elements in the cross-diagonals are automatically determined—no specific action to annihilate couplings within them needs to be taken. As the

Table 8.1 Seventh-degree example: similarity transform (rotation) sequence for the reduction of the full coupling matrix to the folded form.

Transform number	Element to be annihilated			$\theta_r = \tan^{-1}(cM_{kl}/M_{mn})$				
r			Pivot [*i*, *j*]	*k*	*l*	*m*	*n*	*c*
1	M_{16}	In first row	[5, 6]	1	6	1	5	−1
2	M_{15}	In first row	[4, 5]	1	5	1	4	−1
3	M_{14}	In first row	[3, 4]	1	4	1	3	−1
4	M_{13}	In first row	[2, 3]	1	3	1	2	−1
5	M_{37}	In seventh column	[3, 4]	3	7	4	7	+1
6	M_{47}	In seventh column	[4, 5]	4	7	5	7	+1
7	M_{57}	In seventh column	[5, 6]	5	7	6	7	+1
8	M_{25}	In second row	[4, 5]	2	5	2	4	−1
9	M_{24}	In second row	[3, 4]	2	4	2	3	−1
10	M_{46}	In sixth column	[4, 5]	4	6	5	6	+1

Total number of transforms $R = \sum_{n=1}^{N-3} n = 10$.

number of finite position prescribed transmission zeros (TZs) that the transfer function is realizing grows from 1 to the maximum permitted for the $N \times N$ matrix $(N-2)$, then the entries in the cross-diagonals will progressively become nonzero, starting with the asymmetric entry nearest to the principal diagonals (M_{35} in the seventh degree example). If the original filtering function that the matrix is realizing is symmetric, then the asymmetric cross-couplings M_{35}, M_{26}, and M_{17} will automatically be zero (and also, in general, the self-couplings in the principal diagonal M_{11} to M_{77}).

The regular pattern and order of the annihilation procedure makes it very amenable to computer programming, for any degree of coupling matrix.

8.3.1.2 Example of Usage

To illustrate the reduction procedure, an example is a seventh-degree 23 dB return loss single-terminated asymmetric filter, where a complex pair of TZs at $\pm 0.9218 - j0.1546$ in the s-plane are positioned to give group delay equalization over about 60% of the bandwidth, and a single zero is placed at $+j1.2576$ on the imaginary axis to give a rejection lobe level of 30 dB on the upper side of the passband.

With knowledge of the positions of the three Tx zeros, the numerator polynomial $F(s)$ of $S_{11}(s)$ may be constructed by the recursive technique of Section 6.3, and then by using the given return loss and the constant ε, the denominator polynomial $E(s)$ common to $S_{11}(s)$ and $S_{21}(s)$ is determined. The s-plane coefficients of these polynomials are given in Table 8.2. Note that the coefficients of the polynomial $P(s)$ have been multiplied by j because $N - n_{fz} = 4$ is an even number.

From these coefficients the numerators and denominators of y_{21} and y_{22} for this single-terminated network may be built up using equation sets (7.29) and (7.31). Now the residues resultant from the partial-fraction expansions of y_{21} and y_{22} will give the first and last rows of the orthogonal matrix **T** (8.33). After scaling (8.34), the remaining rows of **T** are

Table 8.2 7-1-2 Asymmetric single-terminated filter—coefficients of transfer and reflection polynomials.

s^n	Polynomial coefficients		
$n=$	$P(s)$	$F(s)$	$E(s)$
0	−1.0987	$-j0.0081$	$+0.1378 - j0.1197$
1	$-j0.4847$	+0.0793	$+0.8102 - j0.5922$
2	+0.9483	$-j0.1861$	$+2.2507 - j1.3346$
3	$+j1.0$	+0.7435	$+3.9742 - j1.7853$
4		$-j0.5566$	$+4.6752 - j1.6517$
5		+1.6401	$+4.1387 - j0.9326$
6		$-j0.3961$	$+2.2354 - j0.3961$
7		+1.0	+1.0
	$\varepsilon = 6.0251$	$\varepsilon_R = 1.0$	

	1	2	3	4	5	6	7
1	0.0586	−0.0147	−0.2374	−0.0578	0.4314	−0.4385	0
2	−0.0147	−0.0810	0.4825	0.3890	0.6585	0.0952	−1.3957
3	−0.2374	0.4825	0.2431	−0.0022	0.3243	−0.2075	0.1484
4	−0.0578	0.3890	−0.0022	−0.0584	−0.3047	0.4034	−0.0953
5	0.4314	0.6585	0.3243	−0.3047	0.0053	−0.5498	−0.1628
6	−0.4385	0.0952	−0.2075	0.4034	−0.5498	−0.5848	−0.1813
7	0	−1.3957	0.1484	−0.0953	−0.1628	−0.1813	0.0211

Figure 8.12 7-1-2 asymmetric single-terminated filter—$N \times N$ coupling matrix before application of reduction process. Element values are symmetric about the principal diagonal. $R_1 = 0.7220$, $R_N = 2.2354$.

found using an orthogonalization process (e.g., Gram–Schmidt [9]), and finally the coupling matrix **M** formed using equation (8.29). The element values of **M** are given in Figure 8.12.

To reduce this full coupling matrix to the folded form, a series of 10 similarity transforms may now be applied to **M** according to Table 8.2 and equation (8.35) in Section 8.3.1. Each transform is applied to the coupling matrix resultant from the previous transform, starting with **M** [=$\mathbf{M}_0$ in (8.36)]. After the last of the transforms in the series, the nonzero couplings in the matrix $\mathbf{M}_{10}$ will topologically correspond with couplings between filter resonators arranged in a folded pattern, ready for direct realization in a suitable technology (Figure 8.13). Note that the couplings M_{17} and M_{27} in the cross-diagonals, which are not needed to realize this particular transfer function, will automatically be zero. No specific action to annihilate them needs to be taken.

The results of analyzing this coupling matrix are given in Figure 8.14a (rejection/return loss) and Figure 8.14b (group delay). Return loss as shown here is indicative only, added to demonstrate that the in-band insertion loss for this single-terminated filter is equiripple. It may be seen that the 30 dB lobe level and equalized in-band group delay have not been affected by the transformation process.

	1	2	3	4	5	6	7
1	0.0586	0.6621	0	0	0	0	0
2	0.6621	0.0750	0.5977	0	0	0.1382	0
3	0	0.5977	0.0900	0.4890	0.2420	0.0866	0
4	0	0	0.4890	−0.6120	0.5038	0	0
5	0	0	0.2420	0.5038	−0.0518	0.7793	0
6	0	0.1382	0.0866	0	0.7793	0.0229	1.4278
7	0	0	0	0	0	1.4278	0.0211

Figure 8.13 7-1-2 asymmetric single-terminated filter—$N \times N$ coupling matrix after reduction to folded form ($\mathbf{M}_{10}$).

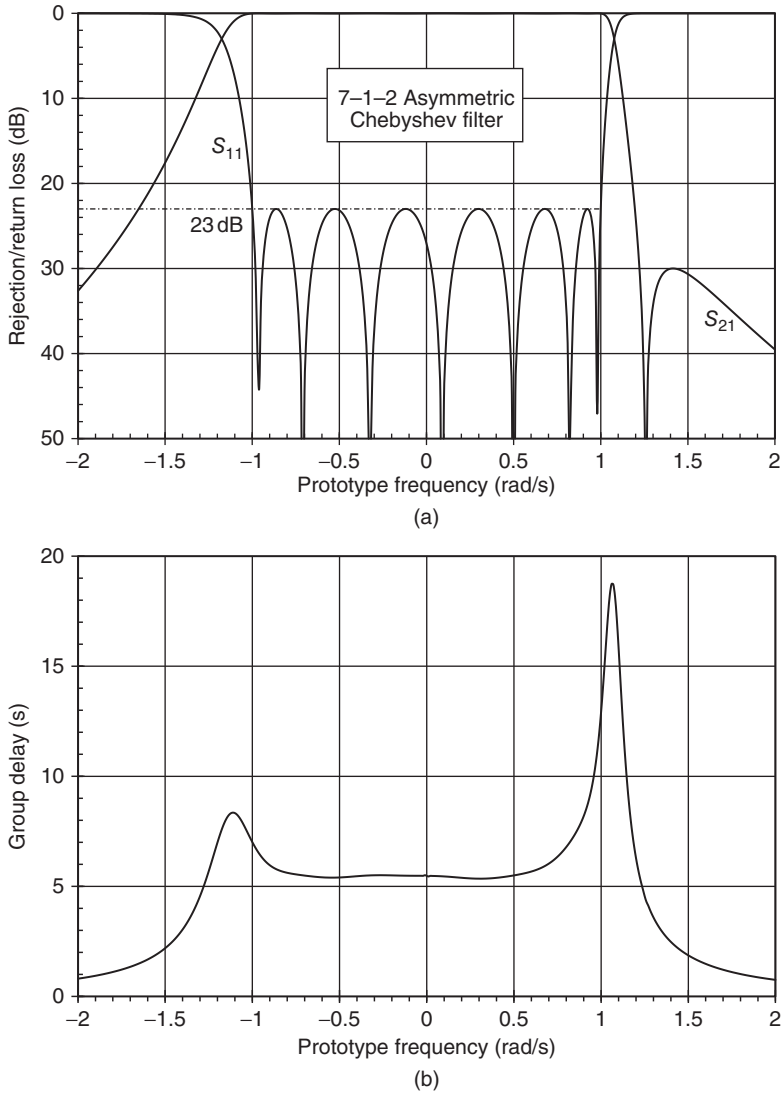


Figure 8.14 7-1-2 asymmetric single-terminated filter synthesis example—analysis of folded coupling matrix: (a) rejection and return loss and (b) group delay.

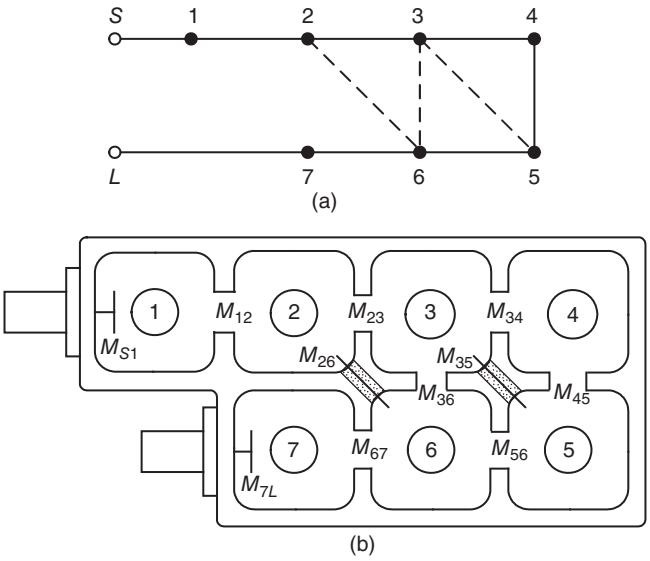


Figure 8.15 Realization in folded configuration: (a) folded network coupling and routing schematic and (b) corresponding realization in coaxial-resonator technology.

Figure 8.15a shows the topology of the folded network corresponding to the coupling matrix of Figures 8.13 and 8.15b a possible realization for the filter in coupled coaxial-resonator cavities. In this particular case, all the cross-couplings happen to be the same sign as the mainline couplings, but in general they will be mixed in sign.

8.4 Synthesis of the $N + 2$ Coupling Matrix

In this section a method is presented for the synthesis of the fully canonical or "$N + 2$" folded coupling matrix, which overcomes some of the shortcomings of the conventional $N \times N$ coupling matrix [18], and is actually easier to synthesize, not needing the GSO step. The $N + 2$ or "extended" coupling matrix has an extra pair of rows top and bottom and an extra pair of columns left and right surrounding the "core" $N \times N$ coupling matrix, which carry the input and output couplings from the source and load terminations to resonator nodes in the core matrix.

The $N + 2$ matrix has some advantages as compared with the conventional coupling matrix:

- Multiple input/output couplings may be accommodated; that is, couplings may be made directly from the source and/or to the load to internal resonators, in addition to the main input/output couplings to the first and last resonators in the filter circuit.
- Fully canonical filtering functions (i.e., Nth-degree characteristics with N finite-position TZs) may be synthesized.
- During certain synthesis procedures that employ a sequence of similarity transforms (rotations), it is sometimes convenient to temporarily "park" couplings in the outer rows or columns, while other rotations are carried out elsewhere in the matrix.

The $N + 2$ coupling matrix for a filter function may be directly created by firstly synthesizing the coupling matrix for an Nth-degree "transversal" circuit and then reducing this to the folded

form using exactly the same reduction technique as was used to reduce the "full" $N \times N$ coupling matrix to the folded form in Section 8.3.

8.4.1 Synthesis of the Transversal Coupling Matrix

The approach that will be employed to synthesize that the $N+2$ transversal coupling matrix will be to construct the two-port short-circuit admittance parameter matrix $[Y_N]$ for the overall network in two ways: the first from the coefficients of the rational polynomials of the transfer and reflection scattering parameters $S_{21}(s)$ and $S_{11}(s)$ that represent the characteristics of the filter to be realized and the second from the circuit elements of the transversal array network. By equating the $[Y_N]$ matrices as derived by these two methods, the elements of the coupling matrix associated with the transversal array network may be related to the coefficients of $S_{21}(s)$ and $S_{11}(s)$ polynomials.

8.4.1.1 Synthesis of Admittance Function $[Y_N]$ from the Transfer and Reflection Polynomials

The transfer and reflection polynomials that are generated in Section 6.3 for the general Chebyshev filtering function are in the form [see equations (6.5) and (6.6)]:

$$S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)}, \quad S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)} \tag{8.39}$$

where

$$\varepsilon = \frac{1}{\sqrt{10^{RL/10} - 1}} \cdot \left| \frac{P(s)}{F(s)/\varepsilon_R} \right|_{s=\pm j}$$

and RL is the prescribed return loss in dB; it is assumed that the polynomials $E(s)$, $F(s)$, and $P(s)$ have been normalized to their respective highest-degree coefficients. Both $E(s)$ and $F(s)$ are Nth-degree polynomials and N is the degree of the filtering function, while $P(s)$, which contains the finite-position prescribed TZs, is of degree n_{fz}, where n_{fz} is the number of finite-position TZs that have been prescribed. For a realizable network, n_{fz} must be $\leq N$.

ε_R is unity for all cases except for fully canonical filtering functions where all the TZs are prescribed at finite frequencies, that is, $n_{fz} = N$. In this case the value of $S_{21}(s)$ (in dB) is finite at infinite frequency and if the highest-degree coefficient of the polynomials $E(s)$, $F(s)$, and $P(s)$ are each normalized to unity, ε_R will have a value slightly greater than unity [see (6.26)]:

$$\varepsilon_R = \frac{\varepsilon}{\sqrt{\varepsilon^2 - 1}} \tag{8.40}$$

The numerator and denominator polynomials for the $y_{21}(s)$ and $y_{22}(s)$ elements of $[Y_N]$ may be built up directly from the transfer and reflection polynomials for $S_{21}(s)$ and $S_{11}(s)$—see equation sets (7.20) and (7.25). For a double-terminated network with source and load terminations of 1 Ω [19],

$$\begin{aligned}
\text{for } N \text{ even}, \quad & y_{21}(s) = y_{21n}(s)/y_d(s) = (P(s)/\varepsilon)/m_1(s) \\
& y_{22}(s) = y_{22n}(s)/y_d(s) = n_1(s)/m_1(s) \\
\text{and for } N \text{ odd}, \quad & y_{21}(s) = y_{21n}(s)/y_d(s) = (P(s)/\varepsilon)/n_1(s) \\
& y_{22}(s) = y_{22n}(s)/y_d(s) = m_1(s)/n_1(s)
\end{aligned}$$

where

$$\begin{aligned}
m_1(s) &= \mathrm{Re}(e_0 + f_0) + j\,\mathrm{Im}(e_1 + f_1)s + \mathrm{Re}(e_2 + f_2)s^2 + \cdots \\
n_1(s) &= j\,\mathrm{Im}(e_0 + f_0) + \mathrm{Re}(e_1 + f_1)s + j\,\mathrm{Im}(e_2 + f_2)s^2 + \cdots
\end{aligned} \tag{8.41}$$

and e_i and f_i, $i = 0, 1, 2, 3, \ldots, N$ are the complex coefficients of $E(s)$ and $F(s)/\varepsilon_R$, respectively. $y_{11}(s)$ may also be found here but, like the $N \times N$ matrix, is not actually needed for the synthesis of the $N+2$ matrix. In a similar way, $y_{21}(s)$ and $y_{22}(s)$ polynomials for single-terminated networks may be found from (7.29) and (7.31).

Knowing the denominator and numerator polynomials for $y_{21}(s)$ and $y_{22}(s)$, their residues r_{21k} and r_{22k}, $k = 1, 2, \ldots, N$ may be found with partial fraction expansions and the purely real eigenvalues λ_k of the network found by rooting the denominator polynomial $y_d(s)$, common to both $y_{21}(s)$ and $y_{22}(s)$. The Nth-degree polynomial $y_d(s)$ has purely imaginary roots $= j\lambda_k$ [see (8.32)]. Expressing the residues in matrix form yields the following equation for the admittance matrix $[Y_N]$ for the overall network:

$$[Y_N] = \begin{bmatrix} y_{11}(s) & y_{12}(s) \\ y_{21}(s) & y_{22}(s) \end{bmatrix} = \frac{1}{y_d(s)} \begin{bmatrix} y_{11n}(s) & y_{12n}(s) \\ y_{21n}(s) & y_{22n}(s) \end{bmatrix}$$

$$= j \begin{bmatrix} 0 & K_\infty \\ K_\infty & 0 \end{bmatrix} + \sum_{k=1}^{N} \frac{1}{(s - j\lambda_k)} \cdot \begin{bmatrix} r_{11k} & r_{12k} \\ r_{21k} & r_{22k} \end{bmatrix} \tag{8.42}$$

where the real constant $K_\infty = 0$ except for the fully canonical case where the number of finite-position TZs n_{fz} in the filtering function is equal to the filter degree N. In this case the degree of the numerator of $y_{21}(s)$ [$y_{21n}(s) = jP(s)/\varepsilon$] is equal to its denominator $y_d(s)$, and K_∞ has to be extracted from $y_{21}(s)$ first to reduce the degree of its numerator polynomial $y_{21n}(s)$ by one before its residues r_{21k} may be found. Note that in the fully canonical case, where the integer quantity $N - n_{fz} = 0$ is even, it is necessary to multiply $P(s)$ by j to ensure that the unitary conditions for the scattering matrix are satisfied.

Being independent of s, K_∞ may be evaluated at $s = j\infty$:

$$jK_\infty = \left.\frac{y_{21n}(s)}{y_d(s)}\right|_{s=j\infty} = \left.\frac{jP(s)/\varepsilon}{y_d(s)}\right|_{s=j\infty} \tag{8.43}$$

The process for building up y_d (8.41) results in its highest-degree coefficient having a value of $1 + 1/\varepsilon_R$, and since the highest-degree coefficient of $P(s) = 1$, the value of K_∞ may be found:

$$K_\infty = \frac{1}{\varepsilon} \cdot \frac{1}{(1 + 1/\varepsilon_R)} = \frac{\varepsilon_R}{\varepsilon} \frac{1}{(\varepsilon_R + 1)} \tag{8.44a}$$

and with the use of (8.40) an alternative form for K_∞ may be derived:

$$K_\infty = \frac{\varepsilon}{\varepsilon_R}(\varepsilon_R - 1) \tag{8.44b}$$

The new numerator polynomial $y'_{21n}(s)$ may now be determined:

$$y'_{21n}(s) = y_{21n}(s) - jK_\infty y_d(s) \tag{8.45}$$

which will be of degree $N - 1$, and the residues r_{21k} of $y'_{21}(s) = y'_{21n}(s)/y_d(s)$ may now be found as normal.

8.4.1.2 Synthesis of Admittance Function $[Y_N]$: Circuit Approach

The two-port short-circuit admittance parameter matrix $[Y_N]$ for the overall network may also be synthesized directly from the fully canonical transversal network, the general form of which is shown in Figure 8.16a. The circuit comprises a series of N individual first-degree lowpass sections, connected in parallel between the source and load terminations, but not to each other. The direct source–load coupling inverter M_{SL} is included to allow fully canonical transfer functions to be realized, according to the "minimum path" rule $n_{fz\max} = N - n_{\min}$, where $n_{fz\max}$ is the maximum number of finite-position TZs that may be realized by the network and $n_{\min}$ is the

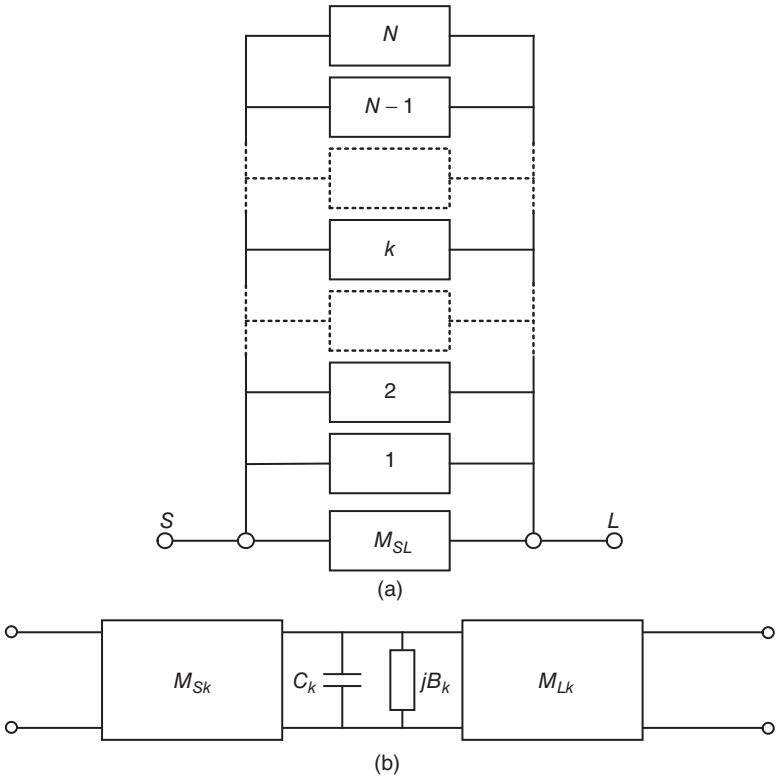


Figure 8.16 Canonical transversal array. (a) *N*—resonator transversal array including direct source–load coupling M_{SL}; (b) equivalent circuit of the *k*th "lowpass resonator" in the transversal array.

number of resonator nodes in the shortest route along nonzero couplings through the network between the source and load terminations. In fully canonical networks $n_{\min} = 0$ and so $n_{fz\max} = N$, the degree of the network.

Each of the N lowpass sections comprises one parallel-connected capacitor C_k and one frequency-invariant susceptance B_k, connected through admittance inverters of characteristic admittances M_{Sk} and M_{Lk} to the source and load terminations, respectively. The circuit of the *k*th lowpass section is shown in Figure 8.16b.

Fully Canonical Filtering Functions The direct source–load inverter M_{SL} in Figure 8.16a is zero except for fully canonical filtering functions, where the number of finite-position zeros equals the degree of the filter. At infinite frequency ($s = \pm j\infty$), all the capacitors C_k in the lowpass sections become parallel short circuits, which appear as open circuits at the source–load ports through the inverters M_{Sk} and M_{Lk}. So the only path between source and load is via the frequency-invariant admittance inverter M_{SL}.

If the load impedance is 1 Ω, the driving point admittance $Y_{11\infty}$ looking in at the input port will be (Figure 8.17)

$$Y_{11\infty} = M_{SL}^2 \tag{8.46}$$

Therefore the input reflection coefficient $S_{11}(s)$ at $s = j\infty$ is

$$S_{11}(s)|_{s=j\infty} \equiv |S_{11\infty}| = \left|\frac{(1 - Y_{11\infty})}{(1 + Y_{11\infty})}\right| \tag{8.47}$$

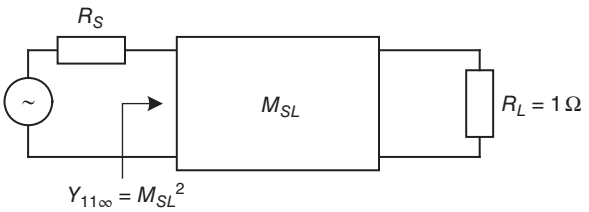


Figure 8.17 Equivalent circuit of transversal array at $s = \pm j\infty$.

Substituting for $|S_{11\infty}|$ in the conservation of energy equation using (8.47),

$$|S_{21\infty}| = \sqrt{1 - |S_{11\infty}|^2} = \frac{2\sqrt{Y_{11\infty}}}{(1 + Y_{11\infty})}$$
$$= \frac{2M_{SL}}{(1 + M_{SL}^2)} \tag{8.48}$$

Solving for M_{SL},

$$M_{SL} = \frac{1 \pm \sqrt{1 - |S_{21\infty}|^2}}{|S_{21\infty}|} = \frac{1 \pm |S_{11\infty}|}{|S_{21\infty}|} \tag{8.49}$$

At infinite frequency $|S_{21}(j\infty)| = |(P(j\infty)/\varepsilon)/E(j\infty)| = 1/\varepsilon$, because for a fully canonical filtering function, $P(s)$ and $E(s)$ will both be Nth-degree polynomials with their highest-degree coefficients normalized to unity. Similarly $|S_{11}(j\infty)| = |(F(j\infty)/\varepsilon_R)/E(j\infty)| = 1/\varepsilon_R$. Therefore,

$$M_{SL} = \frac{\varepsilon(\varepsilon_R \pm 1)}{\varepsilon_R} \tag{8.50}$$

Since ε_R is slightly greater than unity for a fully canonical network, choosing the negative sign will give a relatively small value for M_{SL}:

$$M_{SL} = \frac{\varepsilon(\varepsilon_R - 1)}{\varepsilon_R} \tag{8.51a}$$

and correctly gives $M_{SL} = 0$ for noncanonical filters where $\varepsilon_R = 1$. It can be shown that the positive sign will give a second solution $M'_{SL} = 1/M_{SL}$ [see (8.44)], but since this will be a large number, it is never used in practice [20]. An alternative form of (8.51a) involving ε_R only may be found with the use of (8.40):

$$M_{SL} = \sqrt{\frac{\varepsilon_R - 1}{\varepsilon_R + 1}} \tag{8.51b}$$

where again it may be seen that $M_{SL} = 0$ when $\varepsilon_R = 1$.

8.4.1.3 Synthesis of the Two-Port Admittance Matrix [Y_N]

Cascading the elements in Figure 8.16b gives an [*ABCD*] transfer matrix for the *k*th "low pass resonator":

$$[ABCD]_k = -\begin{bmatrix} \dfrac{M_{Lk}}{M_{Sk}} & \dfrac{(sC_k + jB_k)}{M_{Sk}M_{Lk}} \\ 0 & \dfrac{M_{Sk}}{M_{Lk}} \end{bmatrix} \tag{8.52}$$

which may then be directly converted into the equivalent short-circuit y-parameter matrix (Table 5.2):

$$[y_k] = \begin{bmatrix} y_{11k}(s) & y_{12k}(s) \\ y_{21k}(s) & y_{22k}(s) \end{bmatrix} = \frac{M_{Sk}M_{Lk}}{(sC_k + jB_k)} \cdot \begin{bmatrix} \dfrac{M_{Sk}}{M_{Lk}} & 1 \\ 1 & \dfrac{M_{Lk}}{M_{Sk}} \end{bmatrix}$$

$$= \frac{1}{(sC_k + jB_k)} \cdot \begin{bmatrix} M_{Sk}^2 & M_{Sk}M_{Lk} \\ M_{Sk}M_{Lk} & M_{Lk}^2 \end{bmatrix} \tag{8.53}$$

The two-port short-circuit admittance matrix $[Y_N]$ for the parallel-connected transversal array is the sum of the y-parameter matrices for the N individual sections, plus the y-parameter matrix $[y_{SL}]$ for the direct source–load coupling inverter M_{SL}:

$$[Y_N] = \begin{bmatrix} y_{11}(s) & y_{12}(s) \\ y_{21}(s) & y_{22}(s) \end{bmatrix} = [y_{SL}] + \sum_{k=1}^{N} \begin{bmatrix} y_{11k}(s) & y_{12k}(s) \\ y_{21k}(s) & y_{22k}(s) \end{bmatrix}$$

$$= j \begin{bmatrix} 0 & M_{SL} \\ M_{SL} & 0 \end{bmatrix} + \sum_{k=1}^{N} \frac{1}{(sC_k + jB_k)} \begin{bmatrix} M_{Sk}^2 & M_{Sk}M_{Lk} \\ M_{Sk}M_{Lk} & M_{Lk}^2 \end{bmatrix} \tag{8.54}$$

8.4.1.4 Synthesis of the *N*+2 Transversal Matrix

Now the two expressions for $[Y_N]$, the first in terms of the residues of the transfer function (8.42) and the second in terms of the circuit elements of the transversal array (8.54), may be equated. It may be seen immediately that $M_{SL} = K_\infty$, and for the elements with subscripts "21" and "22" in the matrices in the RHSs of equations (8.42) and (8.54),

$$\frac{r_{21k}}{(s - j\lambda_k)} = \frac{M_{Sk}M_{Lk}}{(sC_k + jB_k)} \tag{8.55a}$$

$$\frac{r_{22k}}{(s - j\lambda_k)} = \frac{M_{Lk}^2}{(sC_k + jB_k)} \tag{8.55b}$$

The residues r_{21k} and r_{22k} and the eigenvalues λ_k have already been derived from S_{21} and S_{22} polynomials of the desired filtering function [see (8.42)], and so by equating the real and imaginary parts in (8.55a) and (8.55b), it becomes possible to relate them directly to the circuit parameters:

$$\begin{aligned} & C_k = 1 \quad \text{and} \quad B_k (\equiv M_{kk}) = -\lambda_k \\ & M_{Lk}^2 = r_{22k} \quad \text{and} \quad M_{Sk}M_{Lk} = r_{21k} \\ \therefore \quad & M_{Lk} = \sqrt{r_{22k}} = T_{Nk} \\ \text{and} \quad & M_{Sk} = r_{21k}/\sqrt{r_{22k}} = T_{1k}, \quad k = 1, 2, \ldots, N \end{aligned} \tag{8.56}$$

It may be recognized at this stage that M_{Sk} and M_{Lk} constitute the unscaled row vectors T_{1k} and T_{Nk} of the orthogonal matrix $\mathbf{T}$ as defined in Section 8.2.1. Because the capacitors C_k of the parallel networks are all unity, and the frequency-invariant susceptances B_k ($=-\lambda_k$, representing the self-couplings $M_{11} \rightarrow M_{NN}$), the input couplings M_{Sk}, the output couplings M_{Lk}, and the direct source–load coupling M_{SL} are all now defined. As a result, the reciprocal $N + 2$ transversal coupling matrix $\mathbf{M}$ representing the network in Figure 8.16a may now be constructed. M_{Sk}

	S	1	2	3	...	k	...	$N-1$	N	L
S		M_{S1}	M_{S2}	M_{S3}	...	M_{Sk}	...	$M_{S,N-1}$	M_{SN}	M_{SL}
1	M_{1S}	M_{11}								M_{1L}
2	M_{2S}		M_{22}							M_{2L}
3	M_{3S}			M_{33}						M_{3L}
⋮	⋮				⋱					⋮
k	M_{kS}					M_{kk}				M_{kL}
⋮	⋮						⋱			⋮
$N-1$	$M_{N-1,S}$							$M_{N-1,N-1}$		$M_{N-1,L}$
N	M_{NS}								M_{NN}	M_{NL}
L	M_{LS}	M_{L1}	M_{L2}	M_{L3}	...	M_{Lk}	...	$M_{L,N-1}$	M_{LN}	

Figure 8.18 $N+2$ fully canonical coupling matrix **M** for the transversal array. The "core" $N \times N$ matrix is indicated within the double lines. The matrix is symmetric about the principal diagonal, that is, $M_{ij} = M_{ji}$.

$(=T_{1k})$ are the N input couplings and occupy the first row and column of the matrix from positions 1 to N; see Figure 8.18. Similarly M_{Lk} $(=T_{Nk})$ are the N output couplings and they occupy the last row and column of **M** from positions 1 to N. All other entries are zero. The terminating impedances R_S and R_L are proportional to $M_{SL}^2 + \sum_{k=1}^{N} M_{Sk}^2$ and $M_{SL}^2 + \sum_{k=1}^{N} M_{kL}^2$, respectively.

Certain networks, for example, where there are direct-coupled extracted poles at the source and/or load terminals as introduced in Chapter 10, require additional shunt FIRs B_S and B_L at the source and load terminations, respectively. Such FIRs arise as a result of synthesis procedures that produce prototype reflection characteristic polynomials $S_{11}(s) = (F(s)/\varepsilon_R)/E(s)$ and $S_{22}(s) = (F_{22}(s)/\varepsilon_R)/E(s)$ where $E(s)$, $F(s)$, and $F_{22}(s)$ have complex coefficients, including their highest-degree coefficients [21]. Thus the numerator admittance polynomials $y_{11n}(s)$ and $y_{22n}(s)$ of the short-circuit admittance matrix $[Y_N]$, as built up using equation (8.41), will also be of degree N, the same as its denominator polynomial $y_d(s)$.

Now before deriving the residues r_{11k} and r_{22k} through partial fraction techniques for this characteristic [equation (8.42)], a factor has first to be evaluated and extracted to reduce the degree of the numerator polynomials $y_{11n}(s)$ and $y_{22n}(s)$ by one, with the same technique used to reduce the degree of $P(s)$ polynomial by one for fully canonical transfer functions [see equations (8.43) and (8.45)]. The short-circuit admittance equation (8.42) in residue form becomes modified as follows, where K_S and K_L relate to the FIRs at the source and load terminals, respectively, and are evaluated at $s = j\infty$:

$$[Y_N] = j \begin{bmatrix} K_S & K_\infty \\ K_\infty & K_L \end{bmatrix} + \sum_{k=1}^{N} \frac{1}{(s - j\lambda_k)} \cdot \begin{bmatrix} r_{11k} & r_{12k} \\ r_{21k} & r_{22k} \end{bmatrix}$$

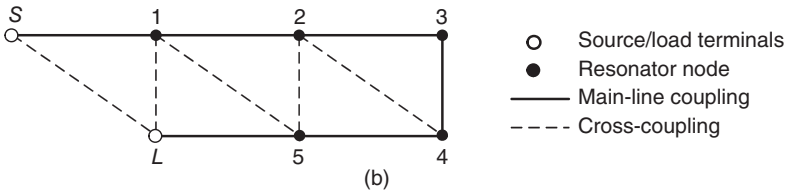
	S	1	2	3	4	5	L
S		m					xa
1	.	s	m			xa	xs
2	.	.	s	m	xa	xs	
3	.	.	.	s	m		
4	.	.	.	.	s	m	
5	.	.	.	.	.	s	m
L	.	.	.	.	.	.	

(a)

Possible nonzero couplings:

- s – Self coupling
- m – Main-line coupling
- xa – Asymmetric cross-coupling
- xs – Symmetric cross-coupling
- . – Coupling values are symmetric about the principal diagonal

All unspecified matrix entries are zero



(b)

Figure 8.19 Folded $N+2$ canonical network coupling matrix form—fifth-degree example: (a) folded coupling matrix form. "s" and "xa" couplings are in general zero for symmetric characteristics; (b) coupling and routing schematic.

With these FIRs B_S and B_L at the source and load terminals of the transversal circuit, the admittance equation (8.54) for the transversal network in circuit element form becomes modified as

$$[Y_N] = j\begin{bmatrix} B_S & M_{SL} \\ M_{SL} & B_L \end{bmatrix} + \sum_{k=1}^{N} \frac{1}{(sC_k + jB_k)} \begin{bmatrix} M_{Sk}^2 & M_{Sk}M_{Lk} \\ M_{Sk}M_{Lk} & M_{Lk}^2 \end{bmatrix}$$

Comparing these two equations, it may be seen that $B_S = K_S$ and $B_L = K_L$ and will appear in the transversal matrix as M_{SS} and M_{LL}, respectively.

8.4.2 Reduction of the $N+2$ Transversal Matrix to the Folded Canonical Form

With N input and output couplings, the transversal topology is clearly impractical to realize for most cases and must be transformed to a more suitable topology if it is to be of practical use. A more convenient form is the folded or "reflex" configuration [14], which may be realized directly or used as the starting point for further transformations to other topologies more suitable for the technology; it is intended to use for the construction of the filter.

To reduce the transversal matrix to the folded form, the formal procedure as described in Section 8.3 may be applied, working on the $N+2$ matrix instead of the $N \times N$ coupling matrix. This procedure involves applying a series of similarity transforms ("rotations"), which eliminate unwanted coupling matrix entries alternately right to left along rows and top to bottom down columns, starting with the outermost rows and columns and working inward toward the center of the matrix, until the only remaining couplings are those that can be realized by filter resonators in a folded structure (Figure 8.19).

As with the $N \times N$ matrix, no special action needs to be taken to eliminate unneeded "xa" and "xs" couplings in the cross-diagonals—they will automatically become zero if they are not necessary to realize the particular filter characteristic under consideration.

8.4.3 Illustrative Example

To illustrate the $N+2$ matrix synthesis procedure, an example is a fully canonical fourth-degree asymmetric filtering function with 22 dB return loss and 4 TZs: two at $-j3.7431$ and $-j1.8051$, which produce 2 attenuation lobes of 30 dB each on the lower side of the passband, and two at $+j1.5699$ and $+j6.1910$ producing a lobe of 20 dB on the upper side. These lobes occur at $-j22.9167$ and $-j2.2414$ on the lower side and $+j2.0561$ on the upper side, calculated using the methods of Section 6.3.4.

Applying the recursive technique of Section 6.3 yields the coefficients for the numerator and denominator polynomials of $S_{11}(s)$ and $S_{21}(s)$, repeated here for convenience:

$$S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)}, \quad S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)} \tag{8.57}$$

and these are shown in Table 8.3. Being fully canonical, $\varepsilon_R \neq 1$ and may be found from (8.40). Note that because $N - n_{fz} = 0$ and is therefore an even number, the coefficients of $P(s)$ have been multiplied by j in Table 8.3.

Now the numerator and denominator polynomials of $y_{21}(s)(= y_{21n}(s)/y_d(s))$ and $y_{22}(s)(= y_{22n}(s)/y_d(s))$ may be constructed using (8.41). The coefficients of $y_d(s)$, $y_{22n}(s)$, and $y_{21n}(s)$, normalized to the highest-degree coefficient of $y_d(s)$, are summarized in Table 8.4.

The next step is to find the residues of $y_{21}(s)$ and $y_{22}(s)$ with partial fraction expansions. Because the numerator of $y_{22}(s)$ ($y_{22n}(s)$) is one less in degree than its denominator $y_d(s)$, finding the associated residues r_{22k} is straightforward. However, the degree of the numerator of $y_{21}(s)$ ($y_{21n}(s)$) is the same as its denominator $y_d(s)$, and the factor K_∞ ($=M_{SL}$) has to be extracted first to reduce $y_{21n}(s)$ in degree by one.

This is easily accomplished by first finding M_{SL} by evaluating $y_{21}(s)$ at $s = j\infty$, that is, M_{SL} equals the ratio of the highest-degree coefficients in the numerator and denominator polynomials of $y_{21}(s)$ [see (8.43)]:

$$jM_{SL} = y_{21}(s)|_{s=j\infty} = \left.\frac{y_{21n}(s)}{y_d(s)}\right|_{s=j\infty} = j0.01509 \tag{8.58}$$

which may be seen is the highest-degree coefficient of $y_{21n}(s)$ in Table 8.4. Alternatively M_{SL} may be derived from (8.51).

M_{SL} may now be extracted from the numerator of $y_{21}(s)$ (8.45):

$$y'_{21n}(s) = y_{21n}(s) - jM_{SL}y_d(s) \tag{8.59}$$

Table 8.3 4-4 Filtering function—coefficients of $E(s)$, $F(s)$, and $P(s)$ polynomials.

s^i $i=$	Coefficients of S_{11} and S_{21} denominator polynomial $E(s)$ (e_i)	Coefficients of S_{11} numerator polynomial $F(s)$ (f_i)	Coefficients of S_{21} numerator polynomial $P(s)$ (p_i)
0	$1.9877 - j0.0025$	0.1580	$j65.6671$
1	$+3.2898 - j0.0489$	$-j0.0009$	$+1.4870$
2	$+3.6063 - j0.0031$	$+1.0615$	$+j26.5826$
3	$+2.2467 - j0.0047$	$-j0.0026$	$+2.2128$
4	$+1.0$	$+1.0$	$+j1.0$
		$\varepsilon_R = 1.000456$	$\varepsilon = 33.140652$

Table 8.4 4-4 Filtering function: coefficients of numerator and denominator polynomials of $y_{21}(s)$, $y_{22}(s)$, and $y'_{21n}(s)$.

s^i $i=$	Coefficients of denominator polynomial of $y_{22}(s)$ and $y_{21}(s)$ ($y_d(s)$)	Coefficients of numerator polynomial of $y_{22}(s)$ ($y_{22n}(s)$)	Coefficients of numerator polynomial of $y_{21}(s)$ ($y_{21n}(s)$)	Coefficients of numerator polynomial of $y_{21}(s)$ after extraction of M_{SL} ($y'_{21n}(s)$)
0	1.0730	$-j0.0012$	$j0.9910$	$j0.9748$
1	$-j0.0249$	+1.6453	+0.0224	+0.0221
2	+2.3342	$-j0.0016$	$+j0.4012$	$+j0.3659$
3	$-j0.0036$	+1.1236	+0.0334	+0.0333
4	+1.0		$+j0.0151$	

Table 8.5 4-4 Filtering function: residues, eigenvalues, and eigenvectors.

k	Eigenvalues λ_k	Residues		Eigenvectors	
		r_{22k}	r_{21k}	$T_{Nk} = \sqrt{r_{22k}}$	$T_{1k} = r_{21k}/\sqrt{r_{22k}}$
1	−1.3142	0.1326	0.1326	0.3641	0.3641
2	−0.7831	0.4273	−0.4273	0.6537	−0.6537
3	0.8041	0.4459	0.4459	0.6677	0.6677
4	1.2968	0.1178	−0.1178	0.3433	−0.3433

At this stage $y'_{21n}(s)$ will be one degree less than $y_d(s)$ (see Table 8.4) and the residues r_{21k} may be found as normal. The residues, the eigenvalues λ_k (where $j\lambda_k$ are the roots of $y_d(s)$), and the associated eigenvectors T_{1k} and T_{Nk} [see (8.56)] are listed in Table 8.5.

Note that for double-terminated lossless networks with equal source and load terminations, r_{22k} will be positive real for a realizable network, and $|r_{21k}| = |r_{22k}|$ [8].

Now knowing the values of the eigenvalues λ_k, the eigenvectors T_{1k}, T_{Nk}, and M_{SL}, the $N+2$ transversal coupling matrix (Figure 8.18) may be completed as shown in Figure 8.20.

Using the same reduction process as described in Section 8.3.1, but now operating upon the $N+2$ matrix, the transversal matrix may be reduced to the folded form with a series of six rotations, annihilating the elements M_{S4}, M_{S3}, M_{S2}, M_{2L}, M_{3L}, and finally M_{13} in sequence (see

	S	1	2	3	4	L
S	0	0.3641	−0.6537	0.6677	−0.3433	0.0151
1	0.3641	1.3142	0	0	0	0.3641
2	−0.6537	0	0.7831	0	0	0.6537
3	0.6677	0	0	−0.8041	0	0.6677
4	−0.3433	0	0	0	−1.2968	0.3433
L	0.0151	0.3641	0.6537	0.6677	0.3433	0

Figure 8.20 Transversal coupling matrix for a 4-4 fully canonical filtering function. The matrix is symmetric about the principal diagonal.

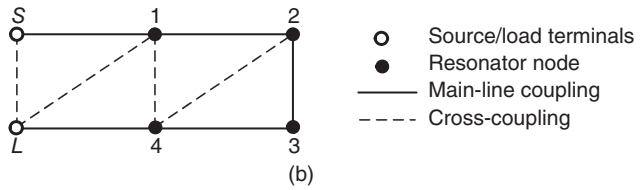
Table 8.6 Fourth-degree example: pivots and angles of the similarity transform sequence for the reduction of the transversal matrix to the folded configuration.

Transform number r	Pivot $[i, j]$	Element to be annihilated	$N+2$ matrix	$\theta_r = \tan^{-1}(cM_{kl}/M_{mn})$				
				k	l	m	n	c
1	[3, 4]	M_{S4}	In row "S"	S	4	S	3	−1
2	[2, 3]	M_{S3}	In row "S"	S	3	S	2	−1
3	[1, 2]	M_{S2}	In row "S"	S	2	S	1	−1
4	[2, 3]	M_{2L}	In column "L"	2	L	3	L	+1
5	[3, 4]	M_{3L}	In column "L"	3	L	4	L	+1
6	[2, 3]	M_{13}	In row 1	1	3	1	2	−1

Total number of transforms $R = \sum_{n=1}^{N-1} n = 6$.

	S	1	2	3	4	L
S	0	1.0600	0	0	0	0.0151
1	1.0600	–0.0023	0.8739	0	–0.3259	0.0315
2	0	0.8739	0.0483	0.8359	0.0342	0
3	0	0	0.8359	–0.0668	0.8723	0
4	0	–0.3259	0.0342	0.8723	0.0171	1.0595
L	0.0151	0.0315	0	0	1.0595	0

(a)



(b)

Figure 8.21 Fully canonical synthesis example. Folded coupling matrix for 4-4 filtering function. (a) Coupling matrix. Matrix is symmetric about the principal diagonal. (b) Coupling and routing schematic.

Table 8.6). The resulting folded configuration coupling matrix is shown in Figure 8.21a, and its corresponding coupling and routing schematic in Figure 8.21b.

The analysis of this coupling matrix is shown in Figure 8.22. It may be seen that the return loss and rejection characteristics are unchanged from those obtained from the analysis of the original S_{11} and S_{21} polynomials.

8.4.4 $N+2$ Coupling Matrix Synthesis for Networks with Complex Terminations

The transversal network technique for the synthesis of two-port $N+2$ coupling matrices may be extended to the synthesis of matrices designed to operate between frequency-invariant complex source and load terminations, which do not necessarily have to be equal. Instead of generating the short-circuit y-parameter polynomials through equation set (8.41), the y-parameter numerator and denominator polynomials y_{11n}, y_{22n}, y_{12n} $(= y_{21n})$ and y_d, modified for operation between complex source and load impedances Z_S and Z_L, are used. The derivations are

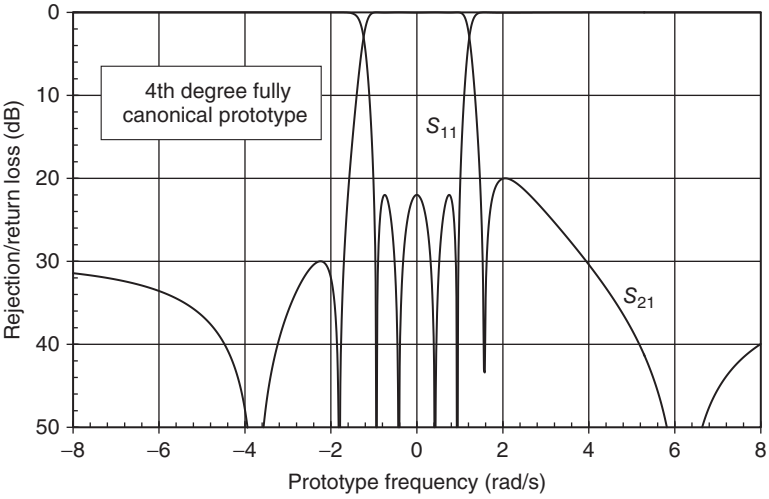


Figure 8.22 4-4 fully canonical synthesis example: analysis of folded coupling matrix. Rejection as $s \rightarrow \pm j\infty = 20\log_{10}(\varepsilon) = 30.407$ dB.

presented in Appendix 6A and are repeated below in equation set 8.60:

$$\begin{aligned}
y_{11n}(s) &= [Z_L^*(E(s) - F(s)/\varepsilon_R) - (-1)^N Z_L(E(s) - F(s)/\varepsilon_R)^*] \\
y_{22n}(s) &= [(Z_S^*E(s) + Z_S F(s)/\varepsilon_R) - (-1)^N (Z_S^*E(s) + Z_S F(s)/\varepsilon_R)^*] \\
y_{12n}(s) &= y_{21n}(s) = -2\sqrt{\mathrm{Re}(Z_S)\mathrm{Re}(Z_L)}P(s)/\varepsilon \\
y_d(s) &= [Z_L^*(Z_S^*E(s) + Z_S F(s)/\varepsilon_R) + (-1)^N Z_L(Z_S^*E(s) + Z_S F(s)/\varepsilon_R)^*]
\end{aligned} \tag{8.60}$$

Now the same procedure for generating the coupling matrix for the unit-termination transverse network may be followed, namely, finding the two-port admittance parameter matrix $[Y]$ in terms of the residues of the known rational transfer and reflection polynomials $E(s)$, $F(s)/\varepsilon_R$, and $P(s)/\varepsilon$, finding $[Y]$ in terms of the unknown circuit elements of the transversal network, equating, building up of the transversal network coupling matrix, and reconfiguring to a more convenient topology. Because the polynomials for y_{11n} and y_{22n} in the case of fully canonical transfer functions will in general have complex non-equal Nth-degree coefficients, a small modification to the procedure is necessary, which will be noted later and demonstrated through an example.

The first example is the noncanonical fourth-degree asymmetric Chebyshev prototype with 22 dB return loss level and two TZs at $s_{01} = +j1.3217$ and $s_{02} = +j1.8082$ that was used as the polynomial synthesis example in Chapter 6. In this example the terminating impedances at the source and load have been set at $Z_S = 0.5 + j0.6$ and $Z_L = 1.3 - j0.8$. Knowing the $E(s)$, $F(s)/\varepsilon_R$, and $P(s)/\varepsilon$ polynomials (Table 6.2), the straightforward application of equation set (8.60) yields the coefficients of the $y_{11}(s)$, $y_{22}(s)$, and $y_{21}(s)$ polynomials as given in Table 8.7.

Then the same procedure as outlined in Section 8.4 may be applied to find the residues and eigenvalues λ_k of the admittance function, from which, because the residue relationship $r_{11k}r_{22k} = r_{21k}^2$ still applies, the eigenvectors T_{1k} and T_{Nk} may be calculated and the transverse coupling matrix constructed (Table 8.8).

Table 8.7 4-2 Chebyshev filtering function: coefficients of the numerator and denominator polynomials of the $y_{11}(s)$, $y_{22}(s)$, and $y_{21}(s)$ polynomials.

s^i $i=$	Coefficients of the denominator polynomial of $y_{11}(s)$, $y_{22}(s)$, and $y_{21}(s)$ ($y_d(s)$)	Coefficients of the numerator polynomial of $y_{11}(s)$ ($y_{11n}(s)$)	Coefficients of the numerator polynomial of $y_{22}(s)$ ($y_{22n}(s)$)	Coefficients of the numerator polynomial of $y_{12}(s) = y_{21}(s)$ ($y_{12n}(s)$, $y_{21n}(s)$)
0	−0.7113	$-j2.1567$	$j0.7264$	$j1.2835$
1	$-j3.9495$	+4.3842	−0.4659	−1.6809
2	+2.6518	$-j0.4205$	$-j2.1752$	$-j0.5370$
3	$-j1.4611$	+2.4015	+0.9237	—
4	+1.0	—	—	—

Table 8.8 4-2 Filtering function with complex terminations: residues, eigenvalues, and eigenvectors.

k	Eigenvalues λ_k	Residues r_{21k}	Residues r_{22k}	Eigenvectors $T_{1k} = r_{21k}/\sqrt{r_{22k}}$	Eigenvectors $T_{Nk} = \sqrt{r_{22k}}$
1	−1.6766	0.3061	0.5740	0.4040	0.7576
2	0.2139	−0.3575	0.2772	−0.6789	0.5265
3	1.0697	0.0542	0.0724	0.2015	0.2691
4	1.8541	-0.2890×10^{-2}	0.4808×10^{-5}	−1.3179	0.0022

Applying a series of similarity transforms (Table 8.6) to reconfigure the transverse matrix into the folded form results in the coupling matrix as shown in Figure 8.23a. For the purposes of comparison, Figure 8.23b shows the equivalent coupling matrix with unity terminating impedances [equation (8.10)].

Comparing the coupling matrices in Figure 8.23, it may be seen that only the tuning state of the first and last resonators and the input and output couplings to them have been affected by the change in termination impedances.

8.4.4.1 Fully Canonical Characteristics with Complex Terminations

The same methods as above may be applied for the synthesis of coupling matrices for fully canonical filtering functions with complex termination impedances. However the application of the equation set (8.60) will result in nonzero Nth-degree coefficients for y_{11n} and y_{22n} polynomials. This will mean that these coefficients will have to be extracted first to allow the residue calculations to proceed, in the same way that the highest-degree coefficient was extracted from the $P(s)$ polynomial in the case of fully canonical characteristics [see equation (8.58)]. The extracted coefficients will occupy the M_{SS} and M_{LL} positions in the coupling matrix.

To demonstrate the procedure, the same fully canonical prototype as was used in Section 8.4.3 is taken, also with complex source and load terminations as above; $Z_S = 0.5 + j0.6$ and $Z_L = 1.3 - j0.8$. Applying the equation set (8.60) results in coefficients for the $y_{11}(s)$, $y_{22}(s)$, and $y_{21}(s)$ polynomials as shown in Table 8.9.

	S	1	2	3	4	L
S	0	1.5497	0	0	0	0
1	1.5497	−1.2860	0.9599	0	0.3606	0
2	0	0.9599	−0.1439	0.2862	0.7742	0
3	0	0	0.2862	−0.9250	0.5674	0
4	0	0.3606	0.7742	0.5674	0.8938	0.9611
L	0	0	0	0	0.9611	0

(a)

	S	1	2	3	4	L
S	0	1.0958	0	0	0	0
1	1.0958	0.1549	0.9599	0	0.3606	0
2	0	0.9599	−0.1439	0.2862	0.7742	0
3	0	0	0.2862	−0.9250	0.5674	0
4	0	0.3606	0.7742	0.5674	0.1549	1.0958
L	0	0	0	0	1.0958	0

(b)

Figure 8.23 4-2 asymmetric filtering function. (a) Coupling matrix with complex source and load terminations $Z_S = 0.5 + j0.6$ and $Z_L = 1.3 - j0.8$. (b) For comparison, the equivalent coupling matrix with source and load values $Z_S = Z_L = 1.0$.

Table 8.9 4-4 Fully canonical chebyshev filtering function: coefficients of the numerator and denominator polynomials of the $y_{11}(s)$, $y_{22}(s)$, and $y_{21}(s)$ polynomials.

s^i $i =$	Coefficients of denominator polynomial of $y_{11}(s)$, $y_{22}(s)$ and $y_{21}(s)$ ($y_d(s)$)	Coefficients of numerator polynomial of $y_{11}(s)$ ($y_{11n}(s)$)	Coefficients of numerator polynomial of $y_{22}(s)$ ($y_{22n}(s)$)	Coefficients of numerator polynomial of $y_{12}(s) = y_{21}(s)$ ($y_{12n}(s)$, $y_{21n}(s)$)
0	1.7478	j1.1236	j0.7264	$-j$1.2289
1	$-j$1.0043	+3.3196	−0.4659	−0.0278
2	+3.2728	+j1.5633	$-j$2.1752	$-j$0.4975
3	$-j$0.6612	+2.2481	+0.9237	−0.0414
4	+1.0	+j0.00028	$-j$0.00021	$-j$0.0187

The matrix values M_{SS} and M_{LL} may be evaluated:

$$y_{11}(s)|_{s=j\infty} = \left.\frac{y_{11n}(s)}{y_d(s)}\right|_{s=j\infty} = j0.00028 = jM_{SS}$$
$$y_{22}(s)|_{s=j\infty} = \left.\frac{y_{22n}(s)}{y_d(s)}\right|_{s=j\infty} = -j0.00021 = jM_{LL} \tag{8.61}$$

and then extracted from $y_{11}(s)$ and $y_{22}(s)$, so reducing their respective degrees by one:

$$y'_{11n}(s) = y_{11n}(s) - jM_{SS}y_d(s)$$
$$y'_{22n}(s) = y_{22n}(s) - jM_{LL}y_d(s) \tag{8.62}$$

Now the residues and eigenvectors of the rest of the network may be calculated, and the transverse coupling matrix constructed and transformed to the folded form. Table 8.10 gives

Table 8.10 4-4 Filtering function with complex terminations: residues, eigenvalues, and eigenvectors.

	Eigenvalues	Residues		Eigenvectors	
k	λ_k	r_{21k}	r_{22k}	$T_{1k} = r_{21k}/\sqrt{r_{22k}}$	$T_{Nk} = \sqrt{r_{22k}}$
1	−1.4748	−0.0504	0.4080	−0.0790	0.6387
2	−0.6719	0.3042	0.2697	0.5859	0.5193
3	0.9489	−0.2202	0.1833	−0.5143	0.4281
4	1.8590	−0.0626	0.0024	−1.2782	0.0490

	S	1	2	3	4	L
S	0.0003	1.4993	0	0	0	0.0187
1	1.4993	−1.3562	0.8743	0	−0.2987	0.0194
2	0	0.8743	0.0638	0.8347	0.0341	0
3	0	0	0.8347	−0.0823	0.8719	0
4	0	−0.2987	0.0341	0.8719	0.7135	0.9290
L	0.0187	0.0194	0	0	0.9290	−0.0002

Figure 8.24 4-4 fully canonical asymmetric coupling matrix with complex source and load terminations $Z_S = 0.5 + j0.6$ and $Z_L = 1.3 - j0.8$.

the eigenvalues, residues, and eigenvectors, and Figure 8.24 shows the folded coupling matrix for the network. Note that in comparison with the matrix with unit terminations (Figure 8.21), all the coupling values have changed.

8.5 Even- and Odd-Mode Coupling Matrix Synthesis Technique: the Folded Lattice Array

An alternative method is available for the synthesis of a coupling matrix using the folded lattice array [22–24]. The method is able to synthesize fully canonical matrices of even or odd degree, realizing asymmetric or symmetric characteristics. It has the advantages of simplicity—only the denominator polynomial $E(s)$ of $S_{21}(s)$ and $S_{11}(s)$ is required to commence the synthesis procedure (although the ratio $\varepsilon/\varepsilon_R$ is also needed for fully canonical characteristics), and only one-port networks of degree $N/2$ (N even) or $(N+1)/2$ (N odd) need to be synthesized. Residues and eigenvalues are not required, eliminating the need for polynomial rooting routines, which can sometimes cause inaccuracies when large-degree networks are concerned. However the method is restricted to networks with equal source and load terminations (thus excluding single-terminated networks) and characteristics where all the reflection zeros [zeros of the numerator polynomial $F(s)$ of $S_{11}(s)$] lie on the imaginary axis.

The method takes advantage of a property of the folded lattice array that all components, frequency-varying susceptances (capacitors), FIRs, and coupling inverters are equal-valued about the plane of symmetry that bisects the network. A network schematic for a sixth-degree example is shown in Figure 8.25a, and the corresponding coupling and routing diagram is given in Figure 8.25b with the bisecting plane of symmetry. In this case, node reactances

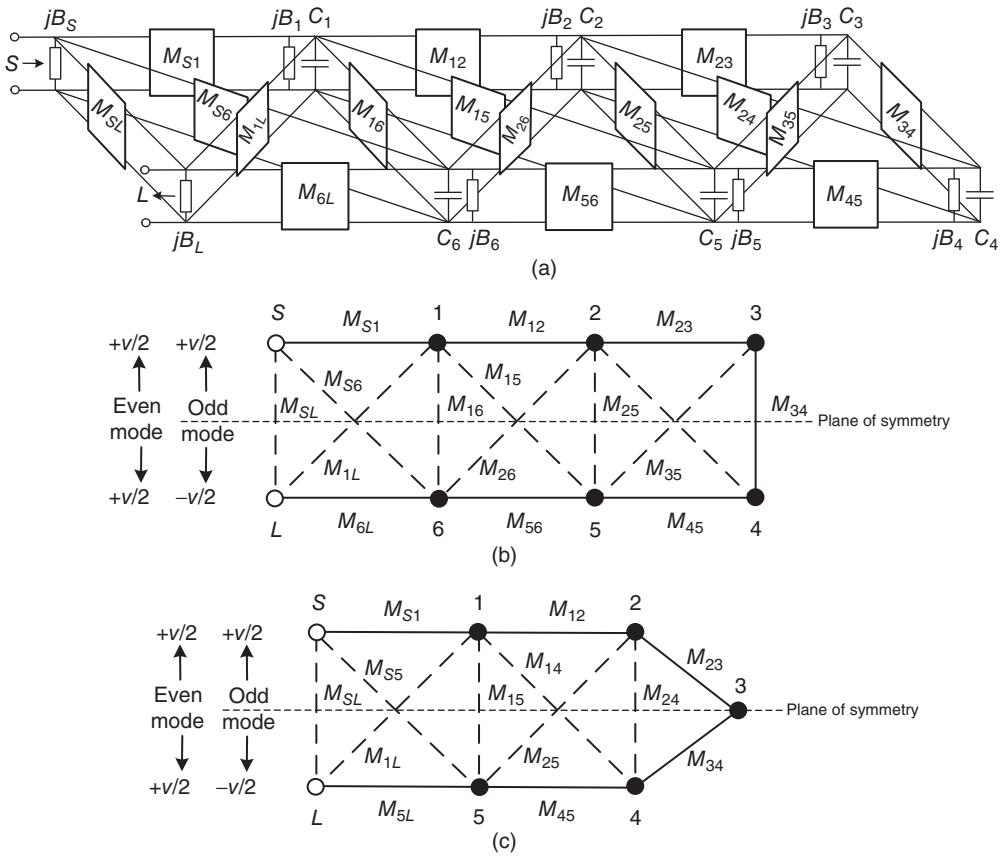


Figure 8.25 Folded lattice network: (a) circuit schematic—sixth-degree example; (b) corresponding coupling and routing diagram; (c) fifth (odd)-degree coupling and routing diagram. All components are equal-valued about the plane of symmetry.

$B_1 = B_6$ and $C_1 = C_6$, couplings $M_{S1} = M_{6L}$ and $M_{S6} = M_{1L}$, and so on. Figure 8.25c shows the equivalent for an odd (fifth)-degree example. Here the plane of symmetry bisects the middle resonant node (node 3), dividing the capacitor/FIR pair into two equal parts, each comprising a shunt-connected capacitor and FIR of values $C_3/2$ and $jB_3/2$, respectively. In both cases the source–load coupling M_{SL} that is necessary to realize fully canonical characteristics is included.

The values of the components of symmetric networks such as this may be synthesized by first bisecting the network into two identical one-port circuits using even- and odd-mode techniques. Each one port will have two different driving point admittances, Y_e or Y_o, depending on whether even-mode or odd-mode voltages, respectively, have been applied to the terminals of the original network. The even- and odd-mode admittances are also known from the desired transfer and reflection functions $S_{21}(s)$ and $S_{11}(s)$, which may be used to derive the values of the components of the even- and odd-mode one port. Then, through some simple formulation, the values of the components of the overall lattice network may be obtained, and the corresponding coupling matrix built up.

Firstly the network has to be separated into two identical one-port circuits along the plane of symmetry, which is done by applying even- and odd-mode voltages to the terminals of the symmetric lattice network as indicated in Figure 8.25.

8.5.1 "Straight" Cross-Couplings

For the "straight" cross-couplings (e.g., M_{16} in Figure 8.25a), the plane of symmetry divides the inverter into two 45° lengths of line, each of admittance M_{16}. The transfer [*ABCD*] matrix for this length of transmission line is [see equation (7.3)]

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & j\sin(\theta)/M_{ij} \\ jM_{ij}\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & j/M_{ij} \\ jM_{ij} & 1 \end{bmatrix} \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} \tag{8.63a}$$

The admittance Y_S looking into terminal 1 of this divided inverter with a load at terminal 2 of admittance Y_L is

$$Y_S = \frac{C + Y_L D}{A + Y_L B} \tag{8.63b}$$

Application of two equal voltages at the source and load terminals (even mode) will result in an open circuit at the bisection plane ($Y_L = 0$), giving $Y_S = jM_{16}$. Similarly, application of two equal voltages but opposite in phase at the source and load terminals (odd mode) will result in a short circuit at the bisection plane ($Y_L = \infty$), giving $Y_S = -jM_{16}$. Thus the bisected inverters will appear as shunt-connected FIRs $jK_1 = jK_6 = +jM_{16}$ (even mode) and $-jM_{16}$ (odd mode) at nodes 1 and 6—see Figure 8.26. jK_1 and jK_6 will add to the FIRs already present at those nodes, jB_1 and jB_6, respectively.

8.5.2 Diagonal Cross-Couplings

Due to the symmetry of values in the lattice network, including the diagonal couplings in each section, voltages at opposing resonator nodes (e.g., nodes 1 and 6, 2 and 5, etc. in the sixth-degree example) will be equal when equal voltages are applied at the network terminals (even mode) and equal but opposite in phase with odd-mode voltages at the terminals. Thus the diagonal couplings M_{15} and M_{26} ($M_{15} = M_{26}$) may be replaced by couplings of values $+M_{15}$ and $+M_{26}$ in parallel with the mainline couplings M_{12} and M_{56} for even-mode voltages and $-M_{15}$ and $-M_{26}$ for odd-mode voltages at the terminals (Figure 8.27).

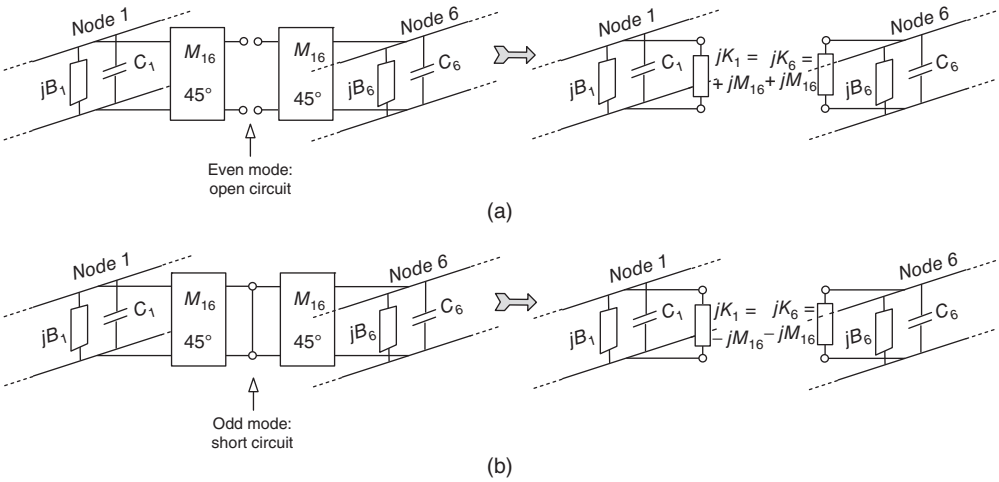


Figure 8.26 Bisection of "straight" cross-couplings through plane of symmetry: (a) even mode; (b) odd mode.

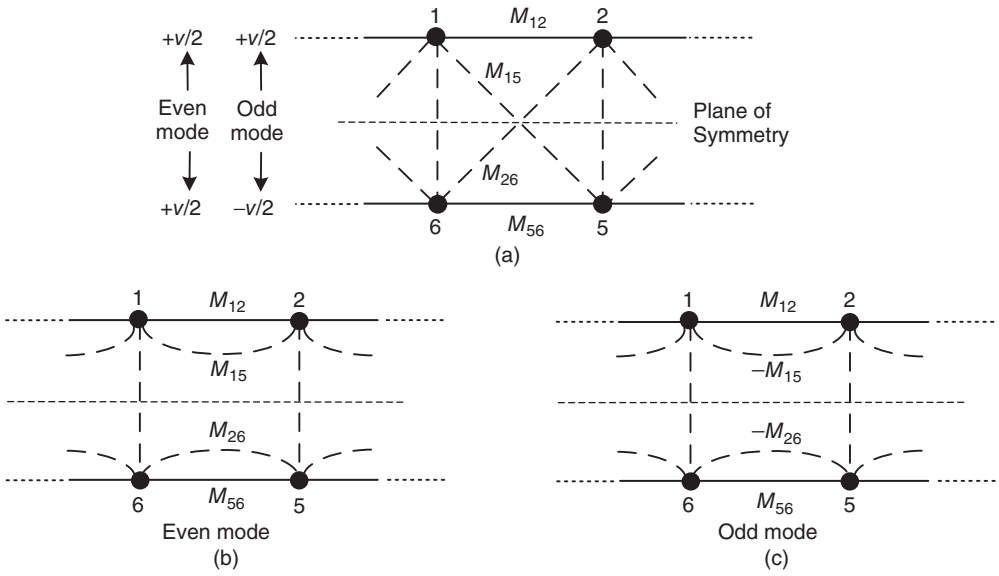


Figure 8.27 Bisection of diagonal cross-couplings through plane of symmetry: (a) first section of lattice network (sixth-degree example); (b) equivalent for even mode; (c) equivalent for odd mode.

8.5.3 Odd-Degree Network: Bisection of the Central Resonant Node

For odd-degree networks, the plane of symmetry bisects the central resonant node into two equal halves—see Figure 8.28. Here the plane is dividing the central node of the fifth-degree example in Figure 8.25c (node 3) into two parallel-connected halves 3a and 3b, each comprising a capacitor of value $C_3/2$ and an FIR of reactance $jB_3/2$. Note that the solid line between the two halves of node 3 in Figure 8.28 (nodes 3a and 3b) is a direct connection, not an inverter.

Under even-mode conditions, the voltages at nodes 3a and 3b will be equal and the 3a/3b connection will act as an open circuit. Under odd-mode conditions, the voltages at nodes 3a and 3b will be equal but opposite in phase, thus creating a short circuit that transforms back through the inverters M_{23} and M_{34} to create open circuits at nodes 2 and 4. Thus for odd-degree networks, the odd-mode one-port circuits will be one degree less than the even-mode circuits.

8.5.4 Even-Mode and Odd-Mode Circuits of the Symmetric Lattice Network

Now that the symmetric lattice network has been bisected into two identical networks along the plane of symmetry, the task is to synthesize the one-port even- and odd-mode circuits from the

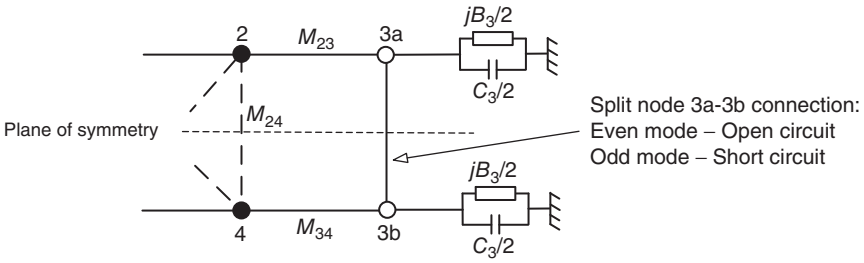


Figure 8.28 Bisection of central resonant node for an odd degree case—fifth-degree example. Because of symmetry, $M_{23} = M_{34}$.

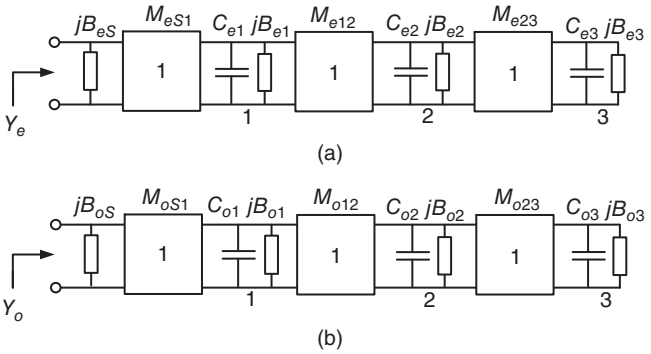


Figure 8.29 Even and odd-mode one-port networks; (a) even mode; (b) odd mode.

even- and odd-mode admittance functions Y_e and Y_o, the derivation of which will be detailed in the following. The synthesis procedure will result in two one-port networks, as shown in Figure 8.29 for a sixth-degree example. For even-degree filtering functions, the degree of the networks will be $N/2$, while for odd-degree functions the even-mode network will be of degree $(N+1)/2$ and the odd-mode network will be one degree less at $(N-1)/2$. In addition, for the even-mode circuit of odd-degree networks, the elements at the non-terminated end of the network will be half the values of the corresponding elements in the complete lattice network.

The synthesis of the one-port circuits proceeds in a conventional way as detailed in Section 7.1.2 [25]. Having derived the rational polynomial admittance functions Y_e and Y_o, elements are extracted in sequence starting with the elements at the ends of each network nearest to the terminals. If the function is fully canonical, the first elements will be the FIRs at the input terminals of each network, B_{eS} and B_{oS} in Figure 8.29. This is followed by the unity coupling inverters, which have the effect of inverting the admittance functions, and then the elements at the first resonant node of each network, first C_{e1} and then B_{e1} from the even-mode chain, C_{o1} and then B_{o1} from the odd-mode chain. The cycle is repeated until the final elements are extracted from each chain.

Finally the nodes are scaled to give unity values to the capacitors. For the first inverters M_{eS1} and M_{oS1}:

$$M_{eS1} = 1/\sqrt{C_{e1}}, \quad M_{oS1} = 1/\sqrt{C_{o1}} \tag{8.64a}$$

For the first nodes and the coupling inverters between nodes 1 and 2

$$\begin{aligned} &M_{e12} = 1/\sqrt{C_{e1}C_{e2}}, \quad M_{o12} = 1/\sqrt{C_{o1}C_{o2}} \\ &B_{e1} \to B_{e1}/C_{e1}, \quad B_{o1} \to B_{o1}/C_{o1} \\ &C_{e1} \to 1, \quad C_{o1} \to 1 \end{aligned} \tag{8.64b}$$

The cycle is repeated until all the capacitors are scaled to unity. To account for the final elements of the even-mode network of an odd-degree filtering function being extracted at half their true values, the scaling of the final inverter connecting to the central node (e.g., M_{23} in the fifth-degree example of Figure 8.25c) and the value of the central FIR are given by

$$M_{e(k-1,k)} = 1/\sqrt{2C_{ek}C_{e(k-1)}}, \quad B_{ek} \to B_{ek}/C_{ek}, \text{ and finally } C_{ek} \to 1 \tag{8.65}$$

where $k = (N+1)/2$.

Having scaled the even- and odd-mode networks and taking advantage of the symmetry about both diagonals, the coupling matrix for the symmetric lattice network may be completed. Taking the first section of the sixth-degree lattice network in Figure 8.25b as an example, it was shown that bisecting the "straight" cross-couplings of the lattice network gives $B_{e1} = (B_1 + M_{16})$ and $B_{o1} = (B_1 - M_{16})$. Adding and subtracting these equations yields the values of $B_1 = (B_{e1} + B_{o1})/2$, which is the self-coupling M_{11} (=M_{66}), and the straight cross-coupling $M_{16} = (B_{e1} - B_{o1})/2$. Similarly for the diagonal cross-couplings (Figure 8.27), $M_{e12} = (M_{12} + M_{15})$ and $M_{o12} = (M_{12} - M_{15})$. Again adding and subtracting gives the values of the mainline couplings $M_{12} = (M_{e12} + M_{o12})/2$ (= M_{56}) and the diagonal cross-couplings $M_{15} = (M_{e12} - M_{o12})/2$ (= M_{26}). For a fully canonical prototype $B_{eS} = (B_S + M_{SL})$ and $B_{oS} = (B_S - M_{SL})$, and this results in $B_S = (B_{eS} + B_{oS})/2$, which is the source self-coupling M_{SS} (=M_{LL}) and $M_{SL} = (B_{eS} - B_{oS})/2$, the direct source–load coupling.

These simple relationships may be applied to the remaining sections of the lattice network, and then symmetry used to complete the rest of the matrix elements. It now remains to derive the rational polynomials of the even- and odd-mode admittance functions $Y_e(s)$ and $Y_o(s)$ so that the even- and odd-mode one-port network synthesis may begin.

8.5.5 Development of the Even- and Odd-Mode Admittance Polynomials from the Transfer and Reflection Polynomials

The reflection and transfer functions $S_{11}(s)$ and $S_{21}(s)$ may be expressed in terms of the even- and odd-mode admittance functions $Y_e(s)$ and $Y_o(s)$ [26]:

$$S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)} = \frac{1 - Y_e Y_o}{(1 + Y_e)(1 + Y_o)} \tag{8.66a}$$

$$S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)} = \frac{Y_o - Y_e}{(1 + Y_e)(1 + Y_o)} \tag{8.66b}$$

From these equations it may be seen that the Hurwitz zeros of $E(s)$ are also the zeros of $(1 + Y_e)(1 + Y_o)$. This means that the numerators of the rational polynomials $(Y_e + 1)$ and $(Y_o + 1)$ are also Hurwitz, but at the moment it is not clear which of the zeros of $E(s)$ are to be allocated to $(Y_e + 1)$ and which to $(Y_o + 1)$. Because it is assumed that the zeros of $F(s)$ and $P(s)$ lie on the imaginary axis or in paraconjugate pairs about the imaginary axis, this may be resolved by considering the alternating pole definition of $S_{21}(s)$ (see Section 6.2).

$S_{21}(s)$ may be expressed in terms of its alternating poles as follows [see equations (6.39a) and (6.39b)]:

$$S_{21}(s)S_{21}(s)^* = |S_{21}(s)|^2 = \frac{1}{[1 + \varepsilon_1 C_N(s)][1 - \varepsilon_1 C_N(s)]} = \frac{1}{1 - \varepsilon_1^2 C_N^2(s)} \tag{8.67}$$

where the purely imaginary characteristic function $C_N(s) = F(s)/P(s)$ $(N - n_{fz}$ odd), $= F(s)/jP(s)$ $(N - n_{fz}$ even), and $\varepsilon_1 = \varepsilon/\varepsilon_R$.

Now $S_{11}(s)$ may be found from the conservation of energy principle:

$$|S_{11}(s)|^2 = 1 - |S_{21}(s)|^2 = \frac{-\varepsilon_1^2 C_N^2(s)}{1 - \varepsilon_1^2 C_N^2(s)} \tag{8.68}$$

The scalar ratio $|S_{11}(s)/S_{21}(s)|$ may then be formed from equations (8.67) and (8.68):

$$\left|\frac{S_{11}(s)}{S_{21}(s)}\right| = j\varepsilon_1 C_N(s) \tag{8.69}$$

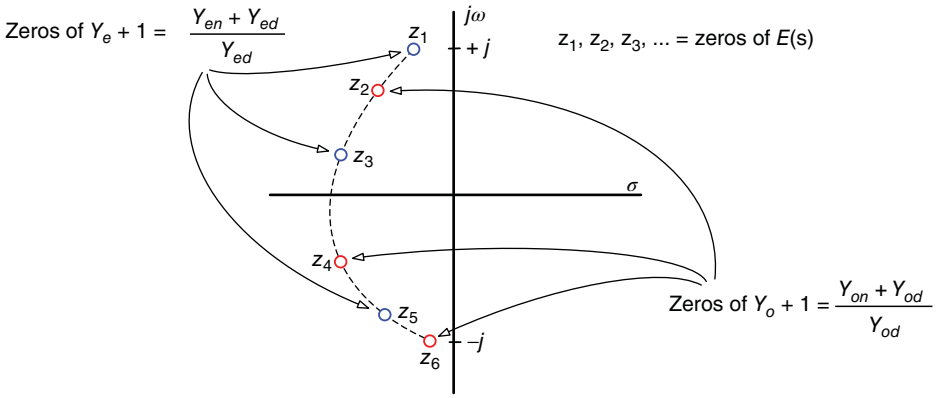


Figure 8.30 Synthesis of even- and odd-mode admittance functions from the Hurwitz zeros of $E(s)$ [poles of $S_{21}(s)$ and $S_{11}(s)$].

Knowing that there is a constant phase difference of $\pm\pi/2$ radians between $S_{11}(s)$ and $S_{21}(s)$ [equation (6.12)], this scalar quantity may be transformed into a vector by multiplication by $\pm j$. Then from equations (8.66) and (8.69):

$$\frac{S_{11}(s)}{S_{21}(s)} = \pm\varepsilon_1 C_N(s) = \frac{1 - Y_e Y_o}{Y_o - Y_e} \tag{8.70}$$

Considering the two parts of the RHS of equation (8.67) separately and substituting for $\varepsilon_1 C_N(s)$ using equation (8.70):

$$\frac{1}{1 + \varepsilon_1 C_N(s)} = \frac{Y_e - Y_o}{(Y_e - 1)(Y_o + 1)} \tag{8.71a}$$

$$\frac{1}{1 - \varepsilon_1 C_N(s)} = \frac{Y_o - Y_e}{(Y_e + 1)(Y_o - 1)} \tag{8.71b}$$

The poles of the left-hand sides of these two expressions alternate between the left and right halves of the s-plane, which, when superimposed, form a symmetrical pattern about the imaginary axis of the s-plane. Therefore it follows from equation (8.71a) that the zeros of $(Y_o + 1)$, which as explained above are Hurwitz, will also be the left half-plane (Hurwitz) zeros of $1 + \varepsilon_1 C_N(s)$. Similarly from equation (8.71b), the zeros of $(Y_e + 1)$ will be coincident with the left half-plane zeros of $1 - \varepsilon_1 C_N(s)$. Completing the symmetric pattern of zeros about the imaginary axis, $(Y_e - 1)$ and $(Y_o - 1)$ are the right half-plane zeros of $1 + \varepsilon_1 C_N(s)$ and $1 - \varepsilon_1 C_N(s)$, respectively. The situation is illustrated in Figure 8.30 for a sixth-degree example, where Y_{en} and Y_{on} are the numerator polynomials of Y_e and Y_o, respectively, while Y_{ed} and Y_{od} are their denominator polynomials.

Knowing the positions of the zeros of $E(s)$, the denominator polynomial of $S_{11}(s)$ and $S_{21}(s)$, the numerator and denominator polynomials of Y_e and Y_o may be constructed. The numerator polynomial of $(Y_e + 1)$, $Y_{en} + Y_{ed}$, is formed from alternately numbered zeros of $E(s)$, starting with the zero with the largest positive imaginary part, that is, z_1, z_3, and z_5 in the sixth-degree example of Figure 8.30. The numerator polynomial of $(Y_o + 1)$, $Y_{on} + Y_{od}$, is formed from the remaining zeros of $E(s)$, z_2, z_4, and z_6. For an odd-degree prototype, $Num(Y_o + 1)$ will be one degree less than $Num(Y_e + 1)$.

8.5.5.1 Canonical and Bandstop Cases

For noncanonical filter degrees where the number of finite zeros $n_{fz} < N$, forming the $(Y_e + 1)$ and $(Y_o + 1)$ numerator polynomials from the zeros of $E(s)$ will automatically set their

highest-degree coefficients at unity. However for fully canonical prototypes ($n_{fz} = N$), a vector multiplier $\overline{v}$ needs to be determined to account for the non-unity highest-degree coefficients that arise for these cases. The complex vector $\overline{v}$ may be determined from equation (8.67) as $\overline{v} = \sqrt{j + \varepsilon_1}$, and then $(Y_{en} + Y_{ed})$ is multiplied by $\overline{v}^*$ and $(Y_{on} + Y_{od})$ by $\overline{v}$.

To obtain bandstop characteristics, the vector $\overline{v}$ is multiplied by 45° (i.e., $\overline{v} \to \overline{v}\sqrt{j}$).

8.5.5.2 Construction of the Even- and Odd-Mode Admittance Polynomials

Since $Y_e(s)$ is a purely reactive function, its poles and zeros will interlace on the imaginary axis, meaning that the coefficients of the numerator and denominator polynomials will alternate between purely real and purely imaginary, that is, complex-even and complex-odd [8]. Forming the numerator polynomial $Num(Y_e + 1) = Y_{en} + Y_{ed}$ from alternate zeros of $E(s)$ will result in a polynomial with complex coefficients; $Y_{en} + Y_{ed} = e_0 + e_1 s + e_2 s^2 + e_3 s^3 + \cdots$. The numerator and denominator polynomials may be then be separated out and the even-mode admittance Y_e formed (ensuring that its numerator Y_{en} has the highest-degree real coefficient):

$$Y_e(s) = \frac{Y_{en}(s)}{Y_{ed}(s)} = \frac{j\mathrm{Im}(e_0) + \mathrm{Re}(e_1)s + j\mathrm{Im}(e_2)s^2 + \mathrm{Re}(e_3)s^3 + \cdots}{\mathrm{Re}(e_0) + j\mathrm{Im}(e_1)s + \mathrm{Re}(e_2)s^2 + j\mathrm{Im}(e_3)s^3 + \cdots} \tag{8.72}$$

Similarly the odd-mode admittance function $Y_o(s)$ may be built up from the remaining zeros of $E(s)$.

Having found the even- and odd-mode admittance functions from the zeros of $E(s)$, synthesis of the corresponding one-port circuits may begin, using the methods for extracting capacitors, FIRs, and inverters as detailed in Chapter 7. Then the simple formulas for deriving the coupling values to build up the coupling matrix for the symmetric lattice network may be applied. The whole synthesis procedure does not require zero-finding routines to calculate residues and is only dealing with polynomials of degree $N/2$ (even degree) or $(N+1)/2$ (odd degree), so ensuring greater accuracy.

8.5.6 Example of the Synthesis of the Coupling Matrix for a Symmetric Lattice Network

To illustrate the procedure, an example of the synthesis of a symmetric lattice coupling matrix for a sixth-degree asymmetric case will be taken. The example demonstrates that only the $E(s)$ zeros of the filter prototype are needed even if the prototype is asymmetric, an exception being that the constants ε and ε_R will also be needed for fully canonical characteristics.

The example to be used here is an asymmetric sixth-degree 22 dB return loss prototype, with 3 TZs on the upper side of the passband producing three rejection lobes of 40 dB each. The prototype singularities are shown in Table 8.11.

Choosing interleaved zeros 1, 3, and 5 of $E(s)$ from Table 8.11 and factoring them to form a third-degree polynomial,

$$Y_{en} + Y_{ed} = (-0.5464 - j0.6566) + (0.2404 - j1.6675)s + (1.1083 - j1.4277)s^2 + s^3$$

Separating out the real and imaginary parts of the coefficients of this polynomial into the complex-even and complex-odd polynomials Y_{en} and Y_{ed} to form $Y_e(s)$,

$$Y_e(s) = \frac{Y_{en}(s)}{Y_{ed}(s)} = \frac{-j0.6566 + 0.2404s - j1.4277s^2 + s^3}{-0.5464 - j1.6675s + 1.1083s^2}$$

Similarly, the polynomial factored from zeros 2, 4, and 6 of $E(s)$ in Table 8.11 will yield Y_{on} and Y_{od} and therefore $Y_o(s)$:

$$Y_o(s) = \frac{Y_{on}(s)}{Y_{od}(s)} = \frac{-j0.6926 + 1.6772s - j0.1720s^2 + s^3}{0.7121 - j0.2757s + 1.1083s^2}$$

Table 8.11 Transmission and reflection singularities: sixth-degree example.

i	Transmission/ reflection poles [roots of $F_N(s)$]	Reflection zeros [roots of $F_N(s)$]	Transmission zeros (prescribed)	In-band reflection maxima
1	$-0.0230 + j1.0223$	$+j0.9932$	$+j1.0900$	$+j0.9707$
2	$-0.1006 + j0.9913$	$+j0.9264$	$+j1.1691$	$+j0.8490$
3	$-0.2944 + j0.8646$	$+j0.7233$	$+j1.5057$	$+j0.5335$
4	$-0.6405 + j0.4248$	$+j0.2721$	$j\infty$	$-j0.0491$
5	$-0.7909 - j0.4592$	$-j0.3934$	$j\infty$	$-j0.7044$
6	$-0.3672 - j1.2441$	$-j0.9218$	$j\infty$	—
		$\varepsilon_R = 1.0$	$\varepsilon = 2.2683$	

Now the Darlington synthesis methods may be applied to extract elements from these even- and odd-mode polynomials to build up the corresponding one-port networks. Considering the even-mode network first (see Figure 8.29a), an input FIR B_{eS} (which will be zero in this non-canonical case) is extracted first, followed by a unit inverter M_{eS1}, then shunt capacitor C_{e1} followed by shunt FIR B_{e1}, then the next unit inverter M_{e12}, and so on until the Y_{en} and Y_{ed} polynomials are constants or zero and all the components have been extracted. Scaling at nodes may now be applied to render all the C_{ei} to unity, as summarized in Table 8.12a. The equivalent results for the odd-mode circuit are given in Table 8.12b.

Having extracted and scaled the even- and odd-mode element values, the bisecting formulas may be applied to find the elements in the upper triangle of the coupling matrix. For the input

Table 8.12 Synthesized capacitor and FIR values for even- and odd-mode one-port circuits: sixth-degree example.

(a) Even mode

	Element values: as extracted			Element values: scaled		
i	C_{ei}	B_{ei}	$M_{e(i,i+1)}$	C_{ei}	B_{ei}	$M_{e(i,i+1)}$
	—	$B_{eS} = 0.0$	$M_{e(S1)} = 1.0$	—	$B_{eS} = 0.0$	$M_{e(S1)} = 1.0528$
1	0.9023	0.0693	1.0	1.0	0.0768	0.7860
2	1.7939	−0.9026	1.0	1.0	−0.5031	0.1041
3	51.4005	−51.4712		1.0	−1.0014	

(b) Odd mode

	Element values: as extracted			Element values: scaled		
i	C_{oi}	B_{oi}	$M_{o(i,i+1)}$	C_{oi}	B_{oi}	$M_{o(i,i+1)}$
	—	$B_{oS} = 0.0$	$M_{o(S1)} = 1.0$	—	$B_{oS} = 0.0$	$M_{o(S1)} = 1.0528$
1	0.9023	0.0693	1.0	1.0	0.0768	1.0078
2	1.0913	0.5257	1.0	1.0	0.4818	0.5390
3	3.1541	−2.3042		1.0	−0.7305	

sections (refer to Figure 8.25),

$$M_{SS} = (B_{eS} + B_{oS})/2 \qquad \text{Input self-coupling}$$
$$M_{SL} = (B_{eS} - B_{oS})/2 \qquad \text{Canonical source-load cross coupling}$$
$$M_{S1} = (M_{e(S1)} + M_{o(S1)})/2 \qquad \text{Mainline coupling}$$
$$M_{SN} = (M_{e(S1)} - M_{o(S1)})/2 \qquad \text{Diagonal crosscoupling}$$
$$M_{11} = (B_{e1} + B_{o1})/2 \qquad \text{Self-coupling at first resonant node}$$
$$M_{1N} = (B_{e1} - B_{o1})/2 \qquad \text{First "straight" cross-coupling}$$

For subsequent lattice sections,

$$M_{i,i+1} = (M_{e(i,i+1)} + M_{o(i,i+1)})/2 \qquad \text{Mainline coupling}$$
$$M_{i,N-i} = (M_{e(i,i+1)} - M_{o(i,i+1)})/2 \qquad \text{Diagonal cross-coupling}$$
$$M_{i+1,i+1} = (B_{e(i+1)} + B_{o(i+1)})/2 \qquad \text{Self-coupling}$$
$$M_{i+1,N-i} = (B_{e(i+1)} - B_{o(i+1)})/2 \qquad \text{Straight cross-coupling}$$

where $i = 1, 2, \ldots, N/2 - 1$ (even degree), $(N-3)/2$ (odd degree). For odd degrees, the values of the couplings to the central resonator (e.g., M_{23} and M_{34} in the fifth-degree example of Figure 8.25c) are given by equation (8.65).

The inherent symmetry about both diagonals may then be used to complete the matrix entries. The full coupling matrix for the sixth-degree asymmetric example is given in Figure 8.31, showing the two diagonals about which the element values are symmetric.

	S	1	2	3	4	5	6	*L*
S	0	1.0528	0	0	0	0	0	0
1	1.0528	0.0768	0.8969	0	0	0.1109	0	0
2	0	0.8969	−0.0107	0.3216	0.2174	0.4924	0.1109	0
3	0	0	0.3216	−0.8660	0.1354	0.2174	0	0
4	0	0	0.2174	0.1354	−0.8660	0.3216	0	0
5	0	0.1109	0.4924	0.2174	0.3216	−0.0107	0.8969	0
6	0	0	0.1109	0	0	0.8969	0.0768	1.0528
L	0	0	0	0	0	0	1.0528	0

Figure 8.31 Coupling matrix for sixth-degree example synthesized as a symmetric lattice network, showing the two planes of symmetry.

	S	1	2	3	4	5	6	*L*
S	0	1.0528	0	0	0	0	0	0
1	1.0528	0.0768	0.9037	0	0	0	0	0
2	0	0.9037	0.1092	0.4298	0	0.4776	0.2201	0
3	0	0	0.4298	−0.7366	0.0399	0.3156	0	0
4	0	0	0	0.0399	−0.9954	0.1306	0	0
5	0	0	0.4776	0.3156	0.1306	−0.1306	0.8765	0
6	0	0	0.2201	0	0	0.8765	0.0768	1.0528
L	0	0	0	0	0	0	1.0528	0

Figure 8.32 Sixth-degree example—folded (reflex) coupling matrix topology after two similarity transforms (rotations).

Table 8.13 Rotation sequence to transform the sixth-degree symmetric lattice network to a folded (reflex) topology.

Transform number r	Pivot $[i, j]$	Element to be annihilated	Figure 8.31	$\theta_r = \tan^{-1}(cM_{kl}/M_{mn})$				
				k	l	m	n	c
1	[2, 5]	M_{15}	In row 1	1	5	1	2	−1
2	[3, 4]	M_{24}	In row 2	2	4	2	3	−1

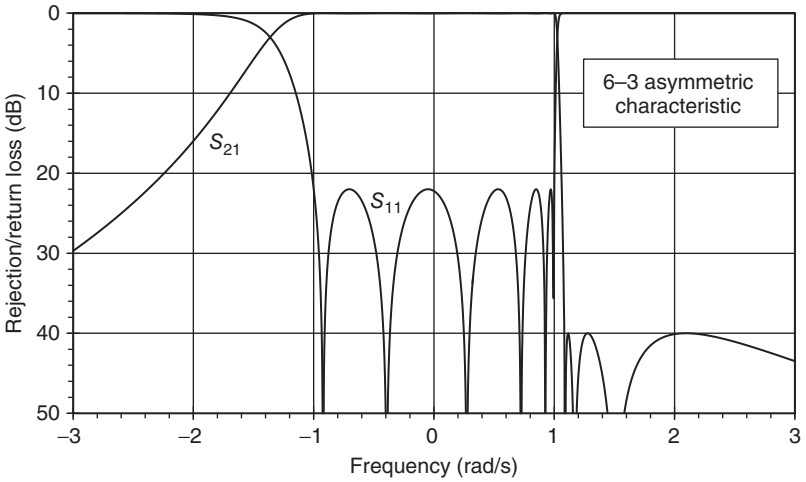


Figure 8.33 Analysis of sixth-degree symmetric lattice coupling matrix—transfer and reflection characteristics.

Rotations may now be applied to the coupling matrix to transform it to a more convenient topology if required—an example is given in Figure 8.32 where two rotations according to Table 8.13 have been applied to transform the lattice topology into a folded form.

Analysis of either of these coupling matrices will produce the transfer and reflection characteristics as shown in Figure 8.33. A simple series of rotations will transform the symmetric lattice coupling matrix to the cul-de-sac form, provided $N - n_{fz} \geq 3$. The pivots start at the center and work outward (see Table 9.7), each with an angle of 45°. For the sixth-degree case shown in Figure 8.33, the pivot sequence would be [3, 4], [2, 5], and then [1, 6].

Summary

In the 1970s, a synthesis technique, based on the coupling matrix representation of a filter network, was introduced for the design of dual-mode bandpass filters. The technique was based on a *bandpass-prototype* circuit, consisting of lossless lumped element resonators intercoupled via lossless transformers. Such a network is capable of realizing a symmetrical filter response. There are two principal advantages of this synthesis technique. Once the general coupling matrix with all the permissible couplings has been synthesized, it allows matrix operations on the coupling matrix to realize a variety of filter topologies. Another advantage is that the coupling matrix represents the practical bandpass filter topology. Therefore, it is possible to identify each element of the practical filter uniquely, including its Q value, dispersion characteristics, and sensitivity.

This permits a more accurate determination of the practical filter characteristics and an insight into ways to optimize filter performance.

At the beginning of this chapter, the concept of an $N \times N$ coupling matrix for synthesizing bandpass filters is introduced. The concept is then extended to include the hypothetical FIR element, as described in Sections 3.10 and 6.1. The FIRs make the synthesis process more general and allows the inclusion of asymmetric filter response in addition. Another generalization is introduced by establishing the equivalent LPP circuit of the coupling matrix. Such a prototype circuit allows the application of the synthesis techniques as discussed in Chapter 7.

The procedure to synthesize the $N \times N$ coupling matrix is described in depth. The concept is extended to $N + 2$ coupling matrix by separating out the purely resistive and purely reactive portions of the $N \times N$ matrix. The synthesis process yields the general coupling matrix with nonzero entries for all the couplings. To realize a filter that is based on such a coupling structure is clearly impractical, and so the next step in the process is to transform the matrix into a form that may be realized with the available couplings of the technology it is intended to employ. This is achieved by applying similarity transformations to the coupling matrix. Such transformations preserve the eigenvalues and eigenvectors of the matrix, thus ensuring that the desired filter response remains unaltered.

There are several practical canonical forms for the transformed coupling matrix **M**. Two of the better-known forms are the *arrow* form and the more generally useful *folded* form. Either of these canonical forms can be used directly if it is convenient to realize the couplings. Alternatively, it may be used as a starting point for the application of further transforms to create topologies better suited for the technology with which the filter is to be eventually realized. The method for the reduction of the general coupling matrix to the folded form is described in depth. The *arrow* form can be derived using a very similar method. Examples are included to illustrate the synthesis process for both the $N \times N$ and $N + 2$ coupling matrices.

References

1 Atia, A.E. and Williams, A.E. (1971) New types of bandpass filters for satellite transponders. *COMSAT Technical Review*, **1**, 21–43.

2 Atia, A.E. and Williams, A.E. (1972) Narrow-bandpass waveguide filters. *IEEE Transactions on Microwave Theory and Techniques*, **MTT-20**, 258–265.

3 Atia, A.E. and Williams, A.E. (1974) Nonminimum-phase optimum-amplitude bandpass waveguide filters. *IEEE Transactions on Microwave Theory and Techniques*, **MTT-22**, 425–431.

4 Atia, A.E., Williams, A.E., and Newcomb, R.W. (1974) Narrow-band multiple-coupled cavity synthesis. *IEEE Transactions on Circuits and Systems*, **CAS-21**, 649–655.

5 Matthaei, G., Young, L., and Jones, E.M.T. (1980) *Microwave Filters, Impedance Matching Networks and Coupling Structures*, Artech House, Norwood, MA.

6 Kreyszig, E. (1972) *Advanced Engineering Mathematics*, 3rd edn, John Wiley and Sons.

7 Frame, J.S. (1964) Matrix functions and applications, part IV: matrix functions and constituent matrices. *IEEE Spectrum*, **1**, Series of 5 articles.

8 van Valkenburg, M.E. (1955) *Network Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

9 Golub, G.H. and van Loan, C.F. (1989) *Matrix Computations*, 2nd edn, The John Hopkins University Press.

10 Gantmacher, F.R. (1959) *The Theory of Matrices*, vol. **1**, The Chelsea Publishing Co., New York.

11 Fröberg, C.E. (1965) *Introduction to Numerical Analysis*, Addison-Wesley, Reading, MA, Chapter 6.

12 Hamburger, H.L. and Grimshaw, M.E. (1951) *Linear Transformations in* n*-Dimensional Space*, Cambridge University Press, London.

13 Bell, H.C. (1982) Canonical asymmetric coupled-resonator filters. *IEEE Transactions on Microwave Theory and Techniques*, **MTT-30**, 1335–1340.

14 Rhodes, J.D. (1970) A lowpass prototype network for microwave linear phase filters. *IEEE Transactions on Microwave Theory and Techniques*, **MTT-18**, 290–300.

15 Rhodes, J.D. and Alseyab, A.S. (1980) The generalized Chebyshev low pass prototype filter. *International Journal of Circuit Theory Applications*, **8**, 113–125.

16 Cameron, R.J. and Rhodes, J.D. (1981) Asymmetric realizations for dual-mode bandpass filters. *IEEE Transactions on Microwave Theory and Techniques*, **MTT-29**, 51–58.

17 Cameron, R.J. (1979) A novel realization for microwave bandpass filters. *ESA Journal*, **3**, 281–287.

18 Cameron, R.J. (2003) Advanced coupling matrix synthesis techniques for microwave filters. *IEEE Transactions on Microwave Theory and Techniques*, **51**, 1–10.

19 Brune, O. (1931) Synthesis of a finite two-terminal network whose driving point impedance is a prescribed function of frequency. *Journal of Mathematical Physics*, **10** (3), 191–236.

20 Amari, S. (2001) Direct synthesis of folded symmetric resonator filters with source–load coupling. *IEEE Microwave and Wireless Components Letters*, **11**, 264–266.

21 He, Y., Wang, G., and Sun, L. (2016) Direct matrix synthesis approach for narrowband mixed topology filters. *IEEE Microwave and Wireless Components Letters*, **26** (5), 301–303.

22 Bell, H.C. (1974) Canonical lowpass prototype network for symmetric coupled-resonator bandpass filters. *Electronics Letters*, **10** (13), 265–266.

23 Bell, H.C. (1979) Transformed-variable synthesis of narrow-bandpass filters. *IEEE Transactions on Circuits and Systems*, **26** (6), 389–394.

24 Bell, H.C. (2007) The coupling matrix in lowpass prototype filters. *IEEE Microwave Magazine*, **8** (2), 70–76.

25 Darlington, S. (1939) Synthesis of reactance 4-poles which produce prescribed insertion loss characteristics. *Journal of Mathematical Physics*, **18**, 257–355.

26 Hunter, I.C. (2001) *Theory and Design of Microwave Filters*, Electromagnetic Waves Series 48, IEE, London.

9

Reconfiguration of the Folded Coupling Matrix

In Chapters 7 and 8, we described two different techniques for synthesizing the canonical folded array coupling matrix from the transfer and the reflection polynomials of the desired filter function. The folded array has several advantages for the filter designer:

- Layout possibilities are relatively simple.
- Positive and negative couplings can be implemented to realize advanced filter characteristics.
- The maximum $(N-2)$ number of finite-position transmission zeros can be realized if required.
- The diagonal couplings, necessary for asymmetric characteristics, can be implemented in some cases.

However, if the filter is to be realized in dual-mode technology (with two orthogonal resonant modes supported in the same physical resonator), a disadvantage becomes apparent. For a folded array realization, the input and output of the filter are in the same physical cavity. This imposes a limit on the input/output isolation that can be achieved with bandpass filters, typically 25 dB for the dual TE_{11n} (transverse electric) mode in cylindrical resonator cavities or dual TE_{10n} mode in square resonator cavities. Consequently, for dual-mode filters, the input and output couplings must occur in separate resonators. This chapter is devoted to the methods for the similarity transformations of the folded coupling array to realize a wide range of topologies suitable for dual-mode filter networks.

9.1 Symmetric Realizations for Dual-Mode Filters

In the early 1970s, the folded cross-coupled array was first introduced by Rhodes in a classical series of papers [1–3]. For symmetric even-degree characteristics, a method for synthesizing dual-mode longitudinal structure from the folded coupling matrix was also introduced [4]. The input to the filter is at one end of the structure and the output is at the other end, avoiding the low isolation problems associated with the folded realization.

Figure 9.1 illustrates the conversion of a (6–2) folded network to an inline or propagating topology. With the inline topology we can realize fewer finite-position transmission zeros than with the folded topology (minimum path rule). So the prototype function must be designed accordingly: two TZs for the 6th degree, four TZs for the 8th and 10th degrees, and six TZs for the 12th degree.

For the inline topology, Rhodes and Zabalawi adopted the even-mode coupling matrix as a base [4]. For even-degree symmetric networks, the even-mode network is formed by reflecting the quadrants of the $N \times N$ folded coupling matrix about the horizontal and vertical lines of symmetry, superimposed into the upper left quadrant. An alternative method for generating

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

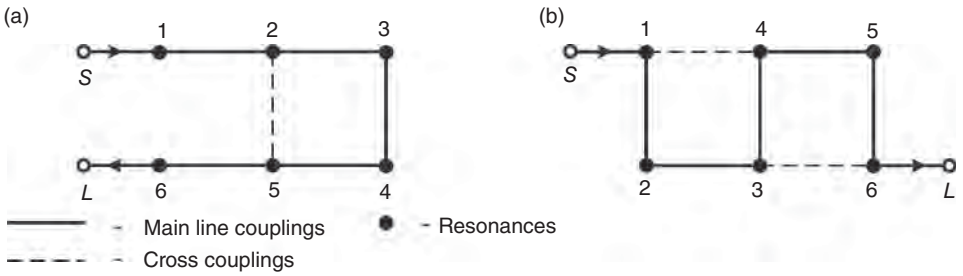


Figure 9.1 Sixth-degree network: (a) cross-coupled folded configuration; (b) after conversion to inline topology.

even-mode and odd-mode submatrices is outlined in Section 9.5.1. Equation (9.1) demonstrates the process for sixth-degree examples: (a) a 6×6 symmetric coupling matrix with the lines of symmetry; (b) the 3×3 even-mode matrix $\mathbf{M}_e$ and the associated K matrix, which is simply the $\mathbf{M}_e$ matrix with the subscripts to the elements along the diagonal adjusted to reflect their real position within the matrix, ($K_{22} = M_{25}$); (c) even-mode matrix after the rotation sequence; and (d) after unfolding to the full matrix:

$$\mathbf{M} = \left[\begin{array}{ccc|ccc} 0 & M_{12} & 0 & 0 & 0 & 0 \\ M_{12} & 0 & M_{23} & 0 & M_{25} & 0 \\ 0 & M_{23} & 0 & M_{34} & 0 & 0 \\ \hline 0 & 0 & M_{34} & 0 & M_{23} & 0 \\ 0 & M_{25} & 0 & M_{23} & 0 & M_{12} \\ 0 & 0 & 0 & 0 & M_{12} & 0 \end{array}\right] \tag{9.1a}$$

$$\mathbf{M}_e = \begin{bmatrix} 0 & M_{12} & 0 \\ M_{12} & M_{25} & M_{23} \\ 0 & M_{23} & M_{34} \end{bmatrix} = \begin{bmatrix} 0 & K_{12} & 0 \\ K_{12} & K_{22} & K_{23} \\ 0 & K_{23} & K_{33} \end{bmatrix} \tag{9.1b}$$

$$\mathbf{M}'_e = \begin{bmatrix} 0 & K'_{12} & K'_{13} \\ K'_{12} & 0 & K'_{23} \\ K'_{13} & K'_{23} & K'_{33} \end{bmatrix} \tag{9.1c}$$

$$\mathbf{M}' = \left[\begin{array}{ccc|ccc} 0 & K'_{12} & 0 & K'_{13} & 0 & 0 \\ K'_{12} & 0 & K'_{23} & 0 & 0 & 0 \\ 0 & K'_{23} & 0 & K'_{33} & 0 & K'_{13} \\ \hline K'_{13} & 0 & K'_{33} & 0 & K'_{23} & 0 \\ 0 & 0 & 0 & K'_{23} & 0 & K'_{12} \\ 0 & 0 & K'_{13} & 0 & K'_{12} & 0 \end{array}\right] \tag{9.1d}$$

Now the matrix is in the correct form for the sixth-degree inline filter (Figure 9.1b). Next, a series of rotations are applied to the even-mode matrix, with angles derived from the elements from the original folded coupling matrix, according to Table 9.1. However, the sequence of pivots and angles for each degree of the matrix does not seem to follow any particular pattern and needs to be determined individually for each degree. Table 9.1 lists the pivots and the degree that the rotations need to be applied to the even-mode matrix for even degrees 6–12 inclusive. The corresponding formulas for the angles for each rotation are given by equations (9.2)–(9.5).

Table 9.1 Symmetric even-degree prototypes: pivot and angle definitions for rotation sequences to reconfigure the folded coupling matrix to inline topology.

Degree *N*	Rotation no. *r*	Rotation pivot [*i, j*]	Angle θ_r	Angle equation	Annihilated elements in [M_e]
6	1	[2, 3]	θ_1	(9.2)	K_{22}
8	1	[3, 4]	θ_1	(9.3a)	—
	2	[2, 4]	θ_2	(9.3b)	$K_{22}\ K_{24}$
10	1	[4, 5]	θ_1	(9.4a)	—
	2	[3, 5]	θ_2	(9.4b)	—
	3	[2, 4]	θ_3	(9.4c)	$K_{44}\ K_{33}\ K_{25}$
12	1	[4, 5]	θ_1	(9.5a)	—
	2	[5, 6]	θ_2	(9.5b)	—
	3	[4, 6]	θ_3	$\theta_3 = \tan^{-1}(K_{46}/K_{66})$	—
	4	[3, 5]	θ_4	$\theta_4 = \tan^{-1}(K_{35}/K_{55})$	$K_{33}\ K_{35}\ K_{44}\ K_{46}$
	5	[2, 4]	θ_5	$\theta_5 = \tan^{-1}(K_{25}/K_{45})$	K_{25}

The coupling elements in the angle formulas are taken from the original $N \times N$ folded coupling matrix.

9.1.1 Sixth-Degree Filter

Only a single rotation is required to transform the folded sixth-degree network to the inline form in Figure 9.1, with pivot [2, 3] and angle θ_1 given by (see equation (8.37e))

$$\theta_1 = \tan^{-1}\left[\frac{M_{23} \pm \sqrt{M_{23}^2 - M_{25}M_{34}}}{M_{34}}\right] \tag{9.2}$$

A symmetric sixth-degree Chebyshev filter function with a 23 dB return loss and two TZs at $\pm j1.5522$ that produce two 40 dB rejection lobes is used as an example to show the realization. The $N \times N$ folded coupling matrix **M** for this filter is shown in Figure 9.2a and the corresponding

(a)
$$\mathbf{M} = \begin{bmatrix} 0 & 0.8867 & 0 & 0 & 0 & 0 \\ 0.8867 & 0 & 0.6050 & 0 & -0.1337 & 0 \\ 0 & 0.6050 & 0 & 0.7007 & 0 & 0 \\ 0 & 0 & 0.7007 & 0 & 0.6050 & 0 \\ 0 & -0.1337 & 0 & 0.6050 & 0 & 0.8867 \\ 0 & 0 & 0 & 0 & 0.8867 & 0 \end{bmatrix}$$

(b)
$$\mathbf{M}_e = \begin{bmatrix} 0 & 0.8867 & 0 \\ 0.8867 & -0.1337 & 0.6050 \\ 0 & 0.6050 & 0.7007 \end{bmatrix}$$

(c)
$$\mathbf{M}'_e = \begin{bmatrix} 0 & 0.8820 & -0.0919 \\ 0.8820 & 0 & 0.6780 \\ -0.0919 & 0.6780 & 0.5670 \end{bmatrix}$$

(d)
$$\mathbf{M}' = \begin{bmatrix} 0 & 0.8820 & 0 & -0.0919 & 0 & 0 \\ 0.8820 & 0 & 0.6780 & 0 & 0 & 0 \\ 0 & 0.6780 & 0 & 0.5670 & 0 & -0.0919 \\ -0.0919 & 0 & 0.5670 & 0 & 0.6780 & 0 \\ 0 & 0 & 0 & 0.6780 & 0 & 0.8820 \\ 0 & 0 & -0.0919 & 0 & 0.8820 & 0 \end{bmatrix}$$

Figure 9.2 Stages in the transformation of a sixth-degree folded coupling matrix to the symmetric inline form.

even-mode matrix $\mathbf{M}_e$, in Figure 9.2b. The application of equation (9.2) yields two solutions for $\theta_1 : -5.948°$ and $61.359°$. Taking the first of these solutions and applying a rotation to $\mathbf{M}_e$ at pivot [2, 3], we annihilate M_{22} to attain the modified even-mode matrix $\mathbf{M}'_e$ (Figure 9.2c). Finally, the unfolding yields the propagating form of the coupling matrix with symmetrically valued coupling elements, as presented in Figure 9.2d.

9.1.2 Eighth-Degree Filter

Refer to Figure 9.3.

The angles θ_1 and θ_2 required for the two rotations for the eighth-degree case are given by

$$\theta_1 = \tan^{-1}\left[\frac{M_{27}M_{34} \pm \sqrt{M_{27}^2 M_{34}^2 + M_{27}M_{45}(M_{23}^2 - M_{27}M_{36})}}{M_{23}^2 - M_{27}M_{36}}\right] \tag{9.3a}$$

and

$$\theta_2 = \tan^{-1}\left[\frac{M_{27}}{M_{23}\sin\theta_1}\right] \tag{9.3b}$$

9.1.3 10th-Degree Filter

Refer to Figure 9.4.

The 10th degree case needs three rotations, with angles resulting from

$$\theta_1 = \tan^{-1}\left[\frac{M_{45} \pm \sqrt{M_{45}^2 - M_{47}M_{56}}}{M_{56}}\right] \tag{9.4a}$$

$$\theta_2 = \tan^{-1}\left[\frac{s_1 M_{34} \pm \sqrt{s_1^2 M_{34}^2 - M_{38}(M_{47} + M_{56})}}{M_{47} + M_{56}}\right] \tag{9.4b}$$

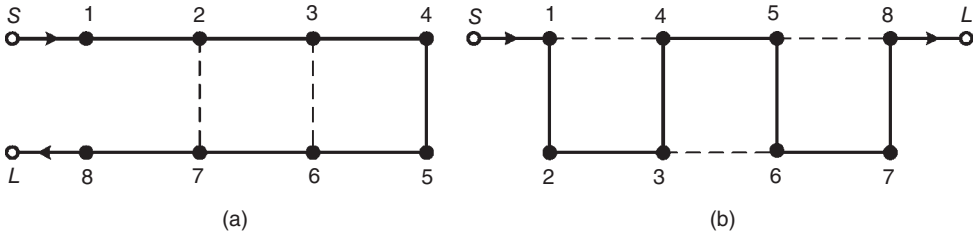


Figure 9.3 Eighth-degree network: (a) cross-coupled folded configuration; (b) after conversion to inline topology.

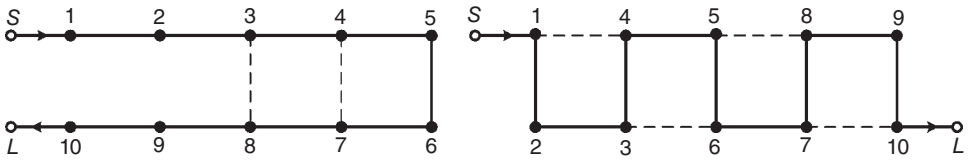


Figure 9.4 10th-degree network: (a) cross-coupled folded configuration; (b) after conversion to inline topology.

$$\theta_3 = \tan^{-1}\left[\frac{t_2 M_{23}}{M_{45} - t_1 M_{56} + c_1 t_2 M_{34}}\right] \tag{9.4c}$$

where $c_1 \equiv \cos\theta_1$, $t_2 \equiv \tan\theta_2$, etc.

9.1.4 12th-Degree Filter

Refer to Figure 9.5.

The 12th-degree filter case begins with the solution of a quartic equation, which is achieved analytically as follows [5]:

$$t_1^4 + d_3 t_1^3 + d_2 t_1^2 + d_1 t_1 + d_0 = 0 \tag{9.5a}$$

where $t_1 \equiv \tan\theta_1$ and

$$d = a_2 c_3^2 + a_3 c_2^2$$

$$d_0 = \frac{a_0 + a_3 c_0^2 + a_4 c_0}{d}$$

$$d_1 = \frac{a_1 + 2a_0 c_3 + 2a_3 c_0 c_1 + a_4(c_1 + c_0 c_3)}{d}$$

$$d_2 = \frac{a_2 + 2a_1 c_3 + a_0 c_3^2 + a_3(c_1^2 + 2c_0 c_2) + a_4(c_2 + c_1 c_3)}{d}$$

$$d_3 = \frac{2a_2 c_3 + a_1 c_3^2 + 2a_3 c_1 c_2 + a_4 c_2 c_3}{d}$$

$$a_0 = M_{58} \qquad b_0 = M_{49} M_{67} \qquad c = a_3 b_5 - a_4 b_4$$

$$a_1 = 2M_{45} \qquad b_1 = -2M_{45} M_{67} \qquad c_0 = \frac{a_0 b_4 - a_3 b_0}{c}$$

$$a_2 = M_{49} - \frac{M_{34}^2}{M_{3,10}} \qquad b_2 = M_{58} M_{67} - M_{56}^2 \qquad c_1 = \frac{a_1 b_4 - a_3 b_1}{c}$$

$$a_3 = M_{67} \qquad b_3 = -2M_{56} M_{45} \qquad c_2 = \frac{a_2 b_4 - a_3 b_2}{c}$$

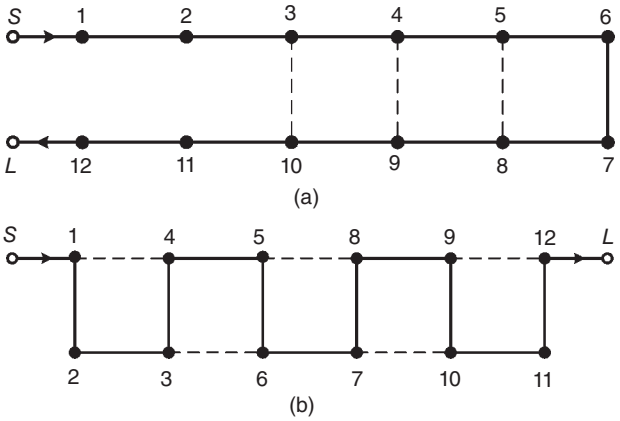


Figure 9.5 12th-degree network: (a) cross-coupled folded configuration; (b) after conversion to inline topology.

$$a_4 = -2M_{56} \qquad b_4 = M_{49}M_{58} - M_{45}^2 \qquad c_3 = \frac{a_3 b_3}{c}$$
$$b_5 = 2M_{56}M_{49}$$

Any one of the previous four solutions of t_1 from the quartic is then used to find θ_2 as follows:

$$\theta_2 = \tan^{-1}\left[\frac{c_0 + c_1 t_1 + c_2 t_1^2}{(1 + c_3 t_1)\sqrt{1 + t_1^2}}\right] \tag{9.5b}$$

Three further rotations are required, according to Table 9.1, by using the elements from the matrix resultant from the previous rotation in each case. After the final rotation and unfolding, the inline configuration in Figure 9.5b is obtained.

9.1.4.1 Conditions for Symmetric Realizations

The angle formulas for the rotations in reconfiguring the folded coupling matrix to the symmetric inline form involve the solutions of quadratic equations. Inevitably, there are certain patterns of transmission zeros in the original prototype filter functions that produce negative square roots and, therefore, cannot be realized with a symmetric structure. These conditions are discussed in more detail in the literature [4]. One useful pattern that does violate the "positive square-root condition" is the case of a real-axis TZ pair and an imaginary-axis TZ pair in an eighth-degree prototype (8–2–2). This particular problem is resolved by using the cascaded quartet (CQ) method, or the asymmetric realization method, discussed next.

9.2 Asymmetric Realizations for Symmetric Characteristics

The originating folded coupling matrix and the resultant inline topology of the asymmetric inline realization are exactly the same as those for the symmetric equivalent. However, the values of the inline coupling matrix and, thus, the dimensions of the physical elements that realize them are not equivalued about the physical center of the structure [6]. Although this means a greater design effort is needed to develop and manufacture a working filter, there are no restrictions on the pattern of TZs that the prototype can incorporate (apart from the minimum path rule, and the symmetry of the pattern of TZs about the imaginary axis (unitary condition), and about the real axis (symmetric characteristics)). Moreover, the computations required to produce the inline configuration are significantly less complex.

Like the symmetric configurations, there does not appear to be a rule governing the sequence of rotations or angles of rotation for the different degrees of filters. Each (even) degree has to be considered individually. For $N = 4$, the folded and inline matrices are of the same form and no further transformation is necessary. The first nontrivial case is $N = 6$, and solutions for $N = 6$, 8, 10, 12, 14 have been derived. Table 9.2 is a summary of the pivots and angles θ_r that are applied to obtain the inline configuration for these degrees. For the rth rotation, the elements M_{l_1,l_2} and M_{m_1,m_2} are again taken from the coupling matrix of the previous rotation and then applied to that matrix, beginning with the original folded matrix for $r = 1$. The resultant inline topologies are exactly the same as those for the symmetric inline structure except for the 6th-, 10th-, and 14th-degree cases. Here, the cross-coupling closest to the output of the filter is zero.

Table 9.2 Pivot positions and rotation angles for the general asymmetric inline realization for degrees 6, 8, 10, 12, and 14.

			$\theta_r = \tan^{-1}[cM_{l_1,l_2}/M_{m_1,m_2}]$				
Degree N	**Rotation number** r	**Pivot** $[i,j]$	l_1	l_2	m_1	m_2	c
6	1	[2, 4]	2	5	4	5	+1
8	1	[4, 6]	3	6	3	4	−1
	2	[2, 4]	2	7	4	7	+1
	3	[3, 5]	2	5	2	3	−1
	4	[5, 7]	4	7	4	5	−1
10	1	[4, 6]	4	7	6	7	+1
	2	[6, 8]	3	8	3	6	−1
	3	[7, 9]	6	9	6	7	−1
12	1	[5, 9]	4	9	4	5	−1
	2	[3, 5]	3	10	5	10	+1
	3	[2, 4]	2	5	4	5	+1
	4	[6, 8]	3	8	3	6	−1
	5	[7, 9]	6	9	6	7	−1
	6	[8, 10]	5	10	5	8	−1
	7	[9, 11]	8	11	8	9	−1
14	1	[6, 10]	5	10	5	6	−1
	2	[4, 6]	4	11	6	11	+1
	3	[7, 9]	4	9	4	7	−1
	4	[8, 10]	7	10	7	8	−1
	5	[9, 11]	6	11	6	9	−1
	6	[10, 12]	9	12	9	10	−1
	7	[5, 7]	4	7	4	5	−1
	8	[7, 9]	6	9	6	7	−1
	9	[9, 11]	8	11	8	9	−1
	10	[11, 13]	10	13	10	11	−1

9.3 "Pfitzenmaier" Configurations

In 1977, G. Pfitzenmaier introduced a configuration to avoid the input/output isolation problem associated with the folded configuration in a dual-mode structure for sixth-degree symmetric filter characteristics [7]. Pfitzenmaier demonstrated that the synthesized sixth-degree circuit can be transformed (not by using coupling matrix methods) to a topology where the input and output resonances (1 and 6) are in adjacent cavities of the dual-mode structure, avoiding the isolation problem. Furthermore, because it is possible to directly cross-couple resonances 1 and 6, the signal has only two resonances to pass through between the input and output. As a result,

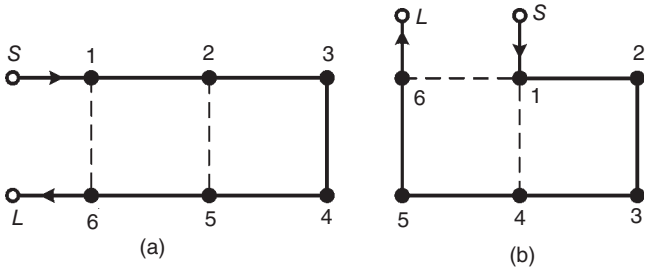


Figure 9.6 Pfitzenmaier configuration for (6–4) symmetric filtering characteristic: (a) original folded configuration; (b) after transformation to the Pfitzenmaier configuration.

by the minimum path rule, the Pfitzenmaier configuration is able to realize $N-2$ transmission zeros, the same as the folded structure. The coupling and routing diagram for a sixth-degree Pfitzenmaier configuration is illustrated in Figure 9.6.

The Pfitzenmaier configuration may be easily obtained for the sixth-degree filter, and any even-degree symmetric characteristic above 6, by using a sequence of coupling matrix rotations [8]. Unlike the asymmetric inline realization, the pivots and angles of the rotations in the sequence may be defined with simple equations. Starting with the folded matrix, a series of $R=(N-4)/2$ rotations is applied according to equation (9.6), after which the Pfitzenmaier configuration is obtained (Figure 9.7).

For the kth rotation, pivot $=[i,j]$ and the associated angle $=\theta_r$, where

$$\left.\begin{aligned} i &= r+1 \\ j &= N-i \\ \theta_r &= \tan^{-1}\frac{-M_{i,N-r}}{M_{j,N-r}} \end{aligned}\right\} r=1,2,3,\ldots,R \tag{9.6}$$

and N is the degree of the filter ($N=$ even integer ≥ 6).

The process is also demonstrated for an eighth-degree filter with six transmission zeros, where the total number of required rotations is $R=(N-4)/2=2$, tabulated as follows.

The transformation of the coupling matrix from folded to Pfitzenmaier topology is illustrated in Figure 9.7.

Filter degree N	Rotation number r	Pivot $[i,j]$	Element to be annihilated	$\theta_r=\tan^{-1}(cM_{i,N-r}/M_{j,N-r})$				
				i	$N-r$	j	$N-r$	c
8	1	[2, 6]	M_{27}	2	7	6	7	−1
	2	[3, 5]	M_{36}	3	6	5	6	−1

It is evident from Figure 9.8 that the Pfitzenmaier topology is easily accommodated by a dual-mode structure, for example, the cylindrical TE_{113}-mode waveguide or the dual-$TE_{01\delta}$-mode dielectric resonators. The M_{18} coupling is unaffected by this process and has the same (small) value as that in the original folded coupling matrix, benefiting the far-out-of-band rejection performance. Also, there are no restrictions on the pattern of the TZs, apart from those previously mentioned for the asymmetric inline configuration.

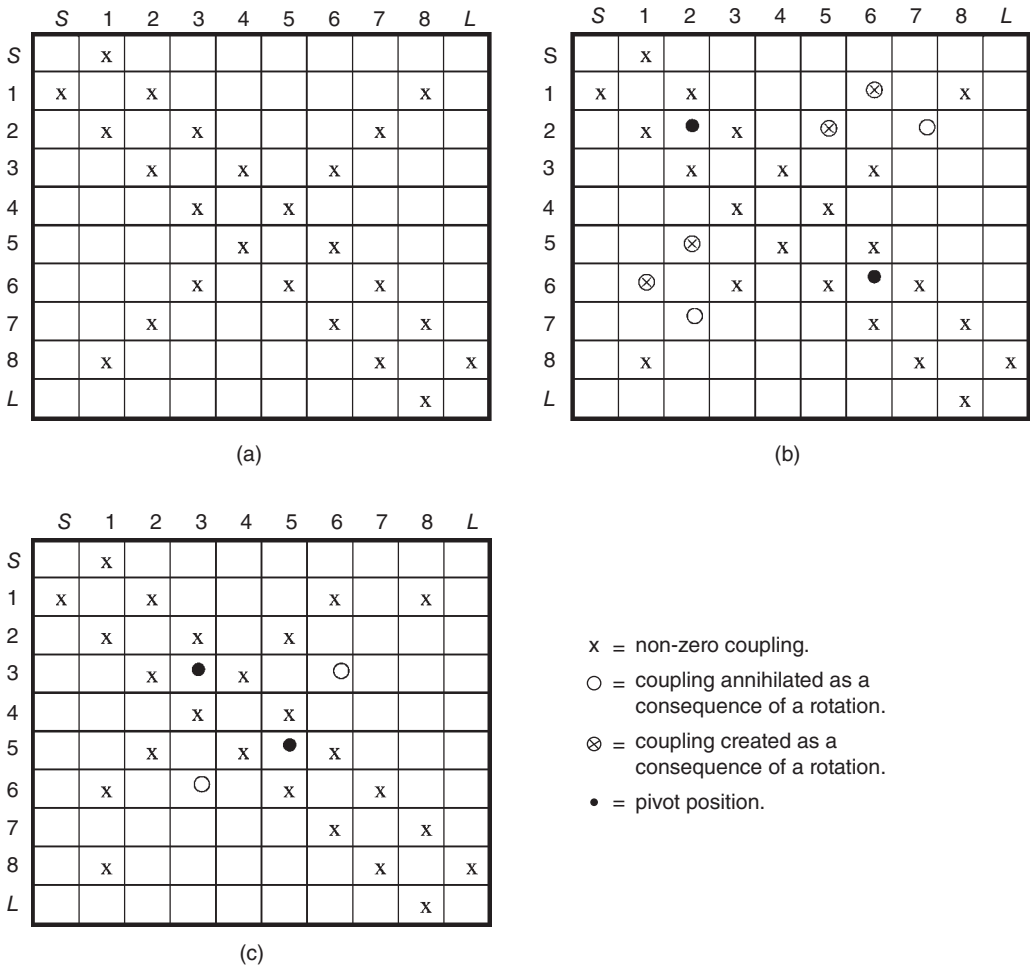


Figure 9.7 Formation of Pfitzenmaier topology: (a) original folded CM; (b) pivot at [2, 6] to annihilate M_{27}, creating M_{16} and M_{25}; (c) pivot at [3, 5] to annihilate M_{36}.

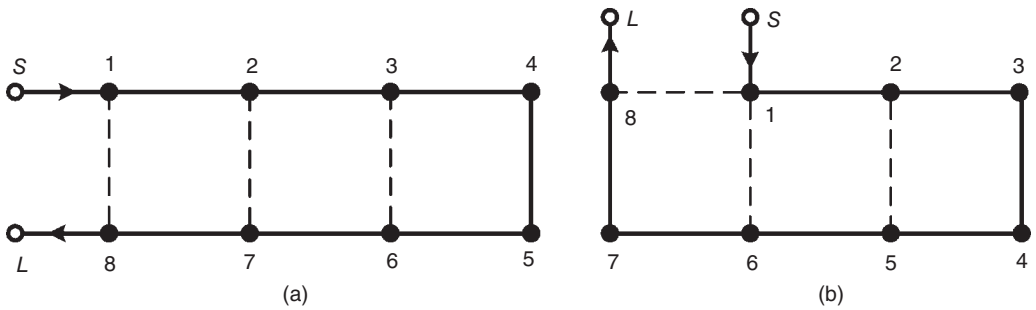


Figure 9.8 Pfitzenmaier configuration for 8th-degree symmetric filtering characteristic: (a) original folded configuration; (b) after transformation to Pfitzenmaier configuration.

9.4 Cascaded Quartets (CQs): Two Quartets in Cascade for Degrees Eight and Above

In the next chapter, a method is presented for creating a cascade of resonator node quartets by using trisections. However, a more direct method is available where the degree of the filter is eight or higher, and there are two pairs of transmission zeros (each pair realized by one of the two quartets in the cascade [6]). Because the angle of the first rotation is the solution of a quadratic equation, there are restrictions on the pattern of transmission zeros that can be realized; however, these restrictions are different from those for the symmetric inline configuration. For the CQ case, the two zero pairs can lie on the real axis or on the imaginary axis or one pair on each axis. Each pair must be symmetrically positioned with respect to both axes, which means that no off-axis pairs are allowed. The CQ configuration is able to realize the useful (8–2–2) case where there are two TZ pairs: one on the real axis and one on the imaginary axis. This pattern, though, contravenes the symmetric inline realizability conditions.

The procedure for synthesizing the pair of quartets necessitates a series of four rotations to an eighth-degree coupling matrix, starting with the folded coupling matrix. The first rotation angle is obtained by solving a quadratic equation (two solutions), expressed in terms of the coupling elements taken from the original folded coupling matrix

$$\begin{aligned} &t_1^2(M_{27}M_{34}M_{45} - M_{23}M_{56}M_{67} + M_{27}M_{36}M_{56}) \\ &\quad + t_1(M_{23}M_{36}M_{67} - M_{27}(M_{34}^2 - M_{45}^2 - M_{56}^2 + M_{36}^2)) \\ &\quad - M_{27}(M_{36}M_{56} + M_{34}M_{45}) = 0 \end{aligned} \tag{9.7}$$

where $t_1 \equiv \tan\theta_1$. Then, three further rotations are applied according to Table 9.3, after which the CQ topology for the eighth-degree filter is obtained as seen in Figure 9.9.

Table 9.3 Rotation sequence for CQ configuration.

Filter degree N	Rotation number r	Pivot $[i, j]$	Element to be annihilated	$\theta_r = \tan^{-1}[cM_{l_1,l_2}/M_{m_1,m_2}]$				
				l_1	l_2	m_1	m_2	c
8	1	[3, 5]	—					
	2	[4, 6]	M_{36}	3	6	3	4	−1
	3	[5, 7]	M_{27}, M_{47}	4	7	4	5	−1
	4	[2, 4]	M_{25}	2	5	4	5	+1

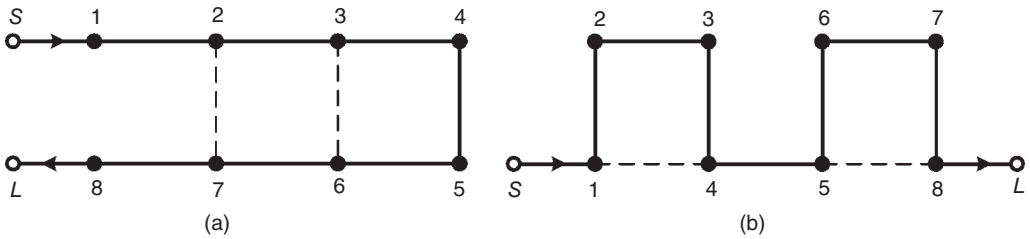


Figure 9.9 Cascade quartet (CQ) configuration with the eighth-degree symmetric filtering characteristic: (a) original folded configuration; (b) after the transformation to form two CQs.

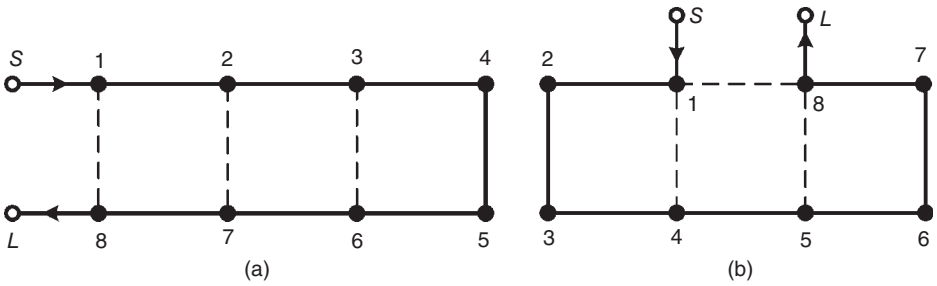


Figure 9.10 Quasi-Pfitzenmaier topology for canonical 8th-degree symmetric filtering characteristic: (a) original folded configuration; (b) after transformation to form two CQs and retaining the M_{18} coupling.

If the minimum path rule is applied to the topology in Figure 9.9b, it is obvious that the CQ solution realizes only four TZs in total. Accordingly, the original prototype characteristic should also incorporate only four TZs. Consequently, the M_{18} coupling in the folded coupling matrix, synthesized from the prototype polynomials, is zero. However, if there are six TZs in the prototype and the M_{18} coupling is present, it is not affected by the CQ rotations and needs to be realized within the CQ topology. This leads to the Pfitzenmaier-like topology, as shown in Figure 9.10, which may have advantages if unusual layout restrictions apply.

For higher-degree characteristics, the folded matrix is synthesized as normal; then the angle formulas and rotation pivots are applied to the central 8×8 submatrix with all the coupling value subscripts and pivot numbers increased by $(N-8)/2$. For example, for a 10–2–2 prototype, $(N-8)/2 = 1$. Coupling M_{67} in equation (9.7) becomes M_{78} and so on, and pivot [3, 5] becomes pivot [4, 6] and so forth. The quartets are shifted to different positions on the diagonal of the coupling matrix by the application of a single rotation according to Table 9.4 (where $i =$ the number of the first resonance node of the quartet). For example, to shift the first quartet ($i = 1$) in Figure 9.9b down the diagonal of the coupling matrix by one position (to become conjoined with the second quartet), we apply a rotation at pivot [2, 4] and angle $\theta = \tan^{-1}[-M_{14}/M_{12}]$ to annihilate coupling M_{14} and to create coupling M_{25}. Shifting the quartet up the diagonal creates a dual-input filter, with both couplings M_{S1} and M_{S3} present.

If the original prototype is asymmetric, each CQ section includes a diagonal cross-coupling, and this synthesis method is not applicable. For the asymmetric cases, the circuit must first be synthesized by using trisections, followed by rotations (hybrid synthesis), detailed in Chapter 10.

The CQ structure has some practical advantages. One is that the topology can be easily implemented in a dual-mode structure of some kind, and another is that each CQ is responsible for the production of a specific TZ pair, facilitating the development and tuning processes. The

Table 9.4 Rotations to shift a quartet up or down the diagonal of the coupling matrix by one position ($i =$ number of first resonance node of quartet).

Quartet shift	Pivot	Element to be annihilated	$\theta = \tan^{-1}[cM_{l_1,l_2}/M_{m_1,m_2}]$				
			l_1	l_2	m_1	m_2	c
Up diagonal	$[i, i+2]$	$M_{i,i+3}$	i	$i+3$	$i+2$	$i+3$	$+1$
Down diagonal	$[i+1, i+3]$	$M_{i,i+3}$	i	$i+3$	i	$i+1$	-1

cross-coupling in the CQ is positive if the TZ pair that it is realizing is on the real axis and negative if the pair is on the imaginary axis.

In the next two sections, two novel realizations are considered: the parallel-connected two-port networks and the cul-de-sac configuration [9]. The first is based on the $N+2$ transverse matrix and is derived by grouping residues and forming separate two-port subnetworks, which are then connected in parallel between the source and load terminations. The second is formed by a series of similarity transforms, operating on the folded coupling matrix.

9.5 Parallel-Connected Two-Port Networks

Closely related to the short-circuit admittance parameters, the eigenvalues and corresponding residues (see Chapter 8) of the filter function are separated into groups and subnetworks by the same matrix rotation procedures as those for the full matrix. Then, the subnetworks are connected in parallel between the source and load terminations to recover the original filtering characteristics. The transverse array itself is regarded as a parallel connection of N single-resonator groups.

Although the choice of residue groupings is arbitrary, difficult-to-realize couplings are created within the subnetworks, and between the internal nodes of the subnetworks and the source/load terminations, if the choice of the filter function and of the residue groupings is not restricted. The restrictions are summarized next:

- Filter functions may be fully canonical but must be symmetric and of even degree.
- There must be complementary pairs of eigenvalues and their associated residues within each group; that is, if an eigenvalue λ_i and its associated residues r_{21i} and r_{22i} are selected to be part of one group, there must also be the eigenvalue $\lambda_j = -\lambda_i$ and its associated residues $r_{21i}(= -r_{21i})$ and $r_{22j}(= r_{22i})$ in the same group. This implies that only doubly terminated between equal value source and load terminations can be synthesized with this simplified configuration.

If these restrictions are observed, the overall network consists of a number of two-port networks. The number corresponds to the number of groups that the residues are divided into, each connected in parallel between the source and the load terminals. If the filter function is fully canonical, the direct source–load coupling M_{SL} will also be present.

After the residues are divided into groups, the synthesis of the submatrices and their reduction to the folded form follows the same process as that for a single network, described for the transversal array synthesis in Chapter 8. To illustrate the process, we examine a 23 dB return loss sixth-degree filter characteristic, with two symmetrically placed transmission zeros at $\pm j1.3958$ and a pair of real-axis zeros at ± 1.0749. This filter is synthesized as two subnetworks, one of degree 2 and one of degree 4. The residues and eigenvalues are calculated based on the procedure outlined in Chapter 8, and the values are summarized in Table 9.5.

By grouping the eigenvalues and associated residues $k = 1$ and 6, we find the following folded matrix for the second-degree subnetwork, shown in Figure 9.11.

Now, the grouping of eigenvalues/residues $k = 2, 3, 4, 5$ yields the folded coupling matrix for the fourth-degree subnetwork of Figure 9.12.

Superimposing the two matrices yields the overall matrix, as shown in Figure 9.13.

The results of analyzing the overall coupling matrix are given in Figure 9.14a (rejection/return loss) and Figure 9.14b (group delay), indicating that the 25 dB lobe level and equalized in-band group delay are preserved.

Table 9.5 (6–2–2) symmetric filter function with residues, eigenvalues, and eigenvectors.

Eigenvalues		Residues		Eigenvectors	
k	λ_k	r_{22k}	r_{21k}	$T_{Nk} = \sqrt{r_{22k}}$	$T_{1k} = r_{21k}/\sqrt{r_{22k}}$
1	−1.2225	0.0975	−0.0975	0.3122	−0.3122
2	−1.0648	0.2365	0.2365	0.4863	0.4863
3	−0.3719	0.2262	−0.2262	0.4756	−0.4756
4	0.3719	0.2262	0.2262	0.4756	0.4756
5	1.0648	0.2365	−0.2365	0.4863	−0.4863
6	1.2225	0.0975	0.0975	0.3122	0.3122

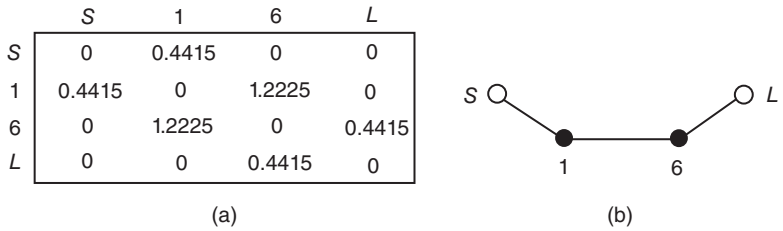


Figure 9.11 Coupling submatrix (a) and coupling–routing diagram (b) for residues $k = 1$ and 6.

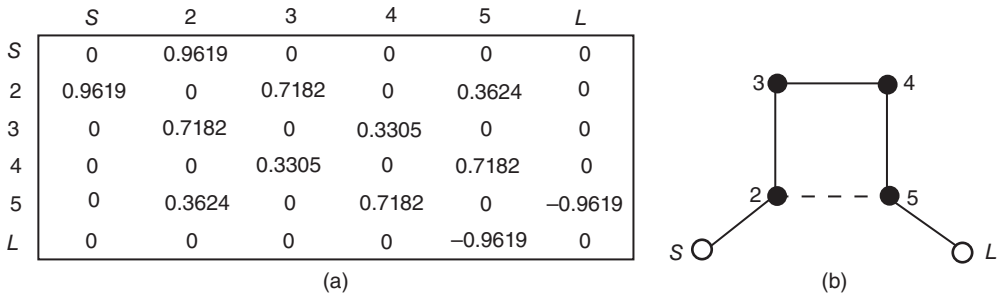


Figure 9.12 Coupling submatrix (a) and coupling–routing diagram (b) for residue group $k = 2, 3, 4, 5$.

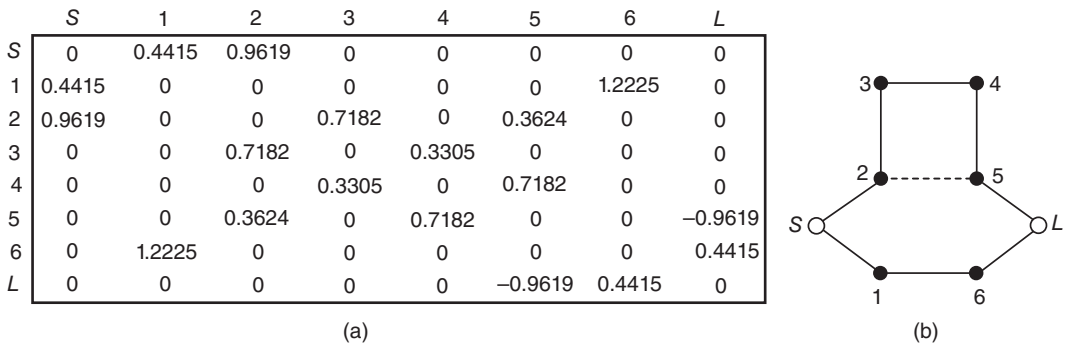


Figure 9.13 Superimposed 2nd- and 4th-degree submatrices: (a) coupling matrix; (b) coupling–routing diagram.

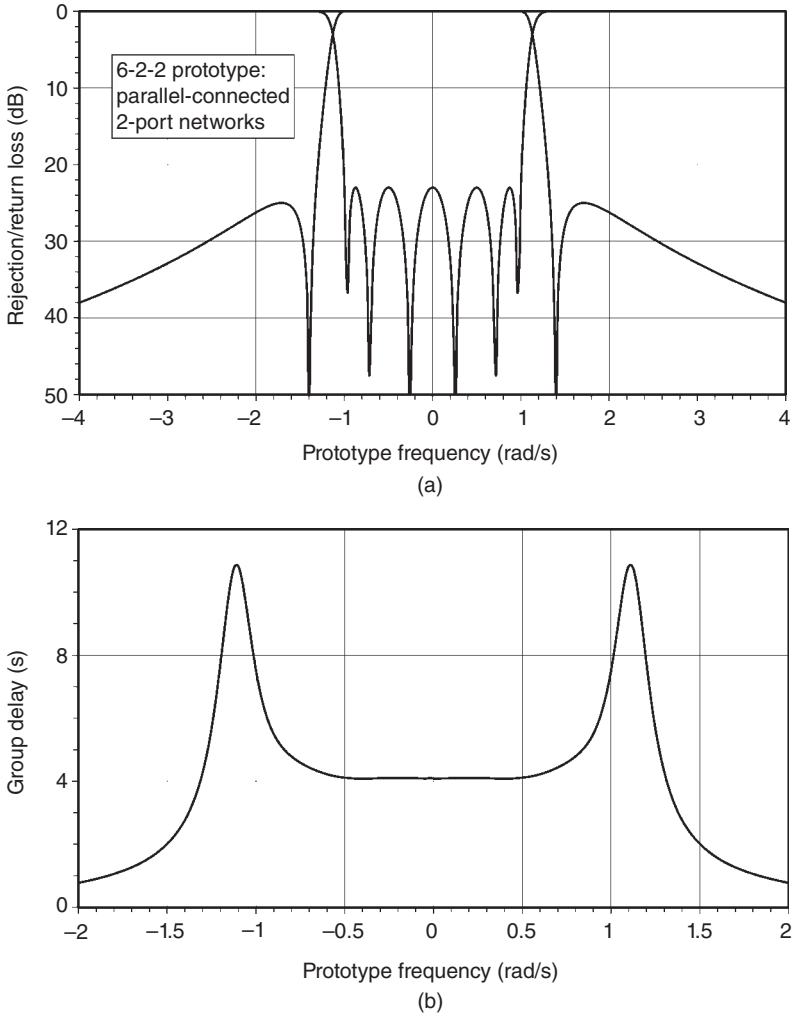


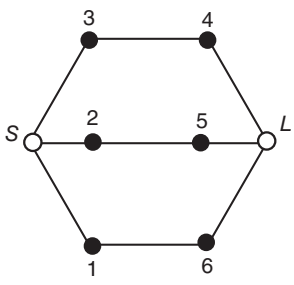
Figure 9.14 Analysis of parallel-connected two-port coupling matrix: (a) rejection and return loss; (b) group delay.

Other solutions for this topology are available, depending on the combinations of the residues that are chosen for the subnetworks. However, whatever combination is chosen, at least one of the input/output couplings is negative. Of course, the number of topology options increases as the degree of the filter function increases, for example, a 10th-degree filter can be realized as two parallel-connected two-port networks, one 4th-degree and one 6th-degree, or as three networks, one 2nd-degree and two 4th-degree, all connected in parallel between the source and load terminations. Also, each subnetwork itself can be reconfigured to other two-port topologies with further transformations.

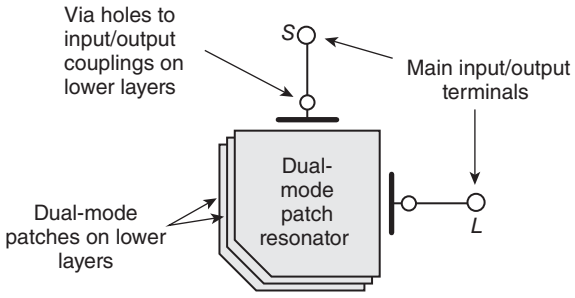
If the network is synthesized as $N/2$ parallel-coupled pairs, as shown in Figure 9.15b, a more direct synthesis route exists. Beginning with the transversal matrix, it is necessary to apply only a series of rotations to annihilate half the couplings in the top row from positions $M_{S,N}$ back to the midpoint $M_{S,N/2+1}$, implying $N/2$ rotations (see Figure 9.15a). Because of the symmetry of the values in the outer rows and columns of the transversal matrix, the corresponding entries M_{1L} to $M_{N/2,L}$ in the last column are annihilated simultaneously.

	S	1	2	3	4	5	6	*L*
S	0	0.4415	0.6877	0.6726	0	0	0	0
1	0.4415	0	0	0	0	0	1.2225	0
2	0.6877	0	0	0	0	1.0648	0	0
3	0.6726	0	0	0	0.3720	0	0	0
4	0	0	0	0.3720	0	0	0	0.6726
5	0	0	1.0648	0	0	0	0	–0.6877
6	0	1.2225	0	0	0	0	0	0.4415
L	0	0	0	0	0.6726	–0.6877	0.4415	0

(a)



(b)



(c)

Figure 9.15 Symmetric (6–2–2) filter configured as parallel-coupled pairs: (a) coupling matrix; (b) coupling–routing diagram; (c) multilayer dual-mode patch resonator realization.

The pivots of the rotations to annihilate these couplings start at position $[1, N]$ and progress along the cross-diagonal towards the center of the matrix to position $[N/2, N/2+1]$. For the sixth-degree example, this is a sequence of $N/2 = 3$ rotations, according to Table 9.6, and is applied to the transversal matrix.

After the series of rotations, the matrix as shown in Figure 9.15a is obtained, which corresponds to the coupling and routing diagram in Figure 9.15b. In each case, at least one of the input/output couplings is negative. An interesting example of a fourth-degree implementation of this topology in dielectric resonator technology is provided in Ref. [10].

A possible realization is with three planar dual-mode resonant patches, arranged in a triple-substrate stack, as shown in Figure 9.15c. Low-temperature co-fired ceramic (LTCC) multilayer substrates technology is well suited for building such filters.

Table 9.6 Symmetric (6–2–2) filter example: rotation sequence for reduction of the transversal matrix to the parallel-coupled pair form.

Transform number	Pivot [*i*, *j*]	Elements to be annihilated	$\theta_r = \tan^{-1}(cM_{kl}/M_{mn})$				
			k	***l***	***m***	***n***	***c***
1	[1, 6]	M_{S6} (and M_{1L})	*S*	6	*S*	1	−1
2	[2, 5]	M_{S5} (and M_{2L})	*S*	5	*S*	2	−1
3	[3, 4]	M_{S4} (and M_{3L})	*S*	4	*S*	3	−1

9.5.1 Even-Mode and Odd-Mode Coupling Submatrices

The even-mode and odd-mode submatrices for filter functions can be synthesized by the eigenvalue grouping process. If the eigenvalues of the function are arranged in the degree of descending value, then the odd-numbered group represents the odd-mode coupling matrix, and the even-numbered group, the even-mode matrix. The two coupling matrices are synthesized by the same methods as those for the full $N+2$ matrix by working with the two subgroups of eigenvalues and their associated residues. Once the two transversal submatrices are obtained, the rotation operations are carried out on the submatrices to annihilate the unwanted couplings.

The even-mode coupling matrices (Section 9.1), representing symmetric inline filter topology, can be generated by this method of residue grouping. For symmetric sixth-degree filter that was used to demonstrate the reconfiguration of the folded coupling matrix to the inline form, the eigenvalues are computed as follows: $\lambda_1 = -\lambda_6 = 1.2185$, $\lambda_2 = -\lambda_5 = 1.0729$, and $\lambda_3 = -\lambda_4 = 0.4214$. Using the even-numbered group of eigenvalues, and their associated residues and eigenvectors, we can synthesize the corresponding $N+2$ even-mode matrix $\mathbf{M}_e$, as denoted in matrix (9.8a). If the odd-numbered eigenvalue group is used instead, the odd-mode matrix $\mathbf{M}_o$ is produced [also shown in (9.8a)]:

$$\mathbf{M}_e = \begin{bmatrix} 0 & 0.7472 & 0 & 0 & 0 \\ 0.7472 & 0 & 0.7586 & -0.4592 & 0.7472 \\ 0 & 0.7586 & -0.4459 & 0.0891 & 0 \\ 0 & -0.4592 & 0.0891 & 1.0129 & 0 \\ 0 & 0.7472 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_o = \begin{bmatrix} 0 & 0.7472 & 0 & 0 & 0 \\ 0.7472 & 0 & 0.6151 & 0.6387 & -0.7472 \\ 0 & 0.6151 & 0.3053 & 0.4397 & 0 \\ 0 & 0.6387 & 0.4397 & -0.8723 & 0 \\ 0 & -0.7472 & 0 & 0 & 0 \end{bmatrix} \tag{9.8a}$$

In this case, one rotation is necessary to transform the $\mathbf{M}_e$ and $\mathbf{M}_o$ matrices to the folded form that is used in the reconfiguration example in Section 9.1. Rotation is carried out at pivot [2, 3] to annihilate M_{13} as shown in (9.8b). Now it can be seen that the core $N \times N$ matrix of $\mathbf{M}_e$ is the same as the even-mode matrix in Figure 9.2b, and as expected, $\mathbf{M}_o$ is its conjugate. These matrices can also be produced by applying the cul-de-sac synthesis process as outlined in Section 9.6.1:

$$\mathbf{M}_e = \begin{bmatrix} 0 & 0.7472 & 0 & 0 & 0 \\ 0.7472 & 0 & 0.8867 & 0 & 0.7472 \\ 0 & 0.8867 & -0.1337 & 0.6050 & 0 \\ 0 & 0 & 0.6050 & 0.7007 & 0 \\ 0 & 0.7472 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_o = \begin{bmatrix} 0 & 0.7472 & 0 & 0 & 0 \\ 0.7472 & 0 & 0.8867 & 0 & -0.7472 \\ 0 & 0.8867 & 0.1337 & 0.6050 & 0 \\ 0 & 0 & 0.6050 & -0.7007 & 0 \\ 0 & -0.7472 & 0 & 0 & 0 \end{bmatrix} \tag{9.8b}$$

9.6 Cul-de-Sac Configuration

The cul-de-sac configuration [9] in its basic form is restricted to doubly terminated networks and realizes a maximum of $N - 3$ transmission zeros. Otherwise, the configuration can accommodate any even- or odd-degree symmetric or asymmetric prototype. Moreover, such a form lends itself to a certain amount of flexibility in the physical layout of its resonators.

A typical cul-de-sac configuration is illustrated in Figure 9.16a for a 10th-degree prototype with the maximum allowable seven TZs (in this case three imaginary axes and two complex pairs). There is a central core of a quartet of resonators in a square formation (1, 2, 9, and 10 in Figure 9.16a), straight coupled to each other (no diagonal cross-couplings). One of these couplings is consistently negative, the choice of which one is arbitrary. The entry to and exit from the core quartet are from the opposite corners of square 1 and 10, respectively, in Figure 9.16a.

Some or all of the other resonators are strung out in cascade from the other two corners of the core quartet in equal numbers (even-degree prototypes) or one more than the other (odd-degree prototypes). The last resonator in each of the two chains has no output coupling, hence the nomenclature "cul-de-sac" for this configuration. Another possible configuration for an odd-degree (seventh-degree) characteristic is shown in Figure 9.16b. If it is practical to include a diagonal cross-coupling between the input and output of the core quartet, an extra transmission zero can be realized. This coupling will have the same value as in the original folded coupling matrix.

The procedure for the synthesis of the cul-de-sac network takes advantage from the symmetry of the ladder lattice network for both the symmetric and the asymmetric responses of a doubly terminated filter network [11]. This leads to a particularly simple and regular series of similarity rotations. Starting with the folded coupling matrix, elements are annihilated by a series of regular similarity transforms (for odd-degree filters), and cross-pivot transforms (for even-degree filters), beginning with a mainline coupling near the center of the matrix, and working outward along or parallel to the cross-diagonal.

For odd-degree filters, the angle formula takes the conventional form [see equation (8.38a)]:

$$\theta_r = \tan^{-1}\frac{M_{ik}}{M_{jk}} \tag{9.9}$$

where $[i,j]$ is the pivot coordinate and $k = j - 1$. For the even-degree characteristics, the cross-pivot transform is adopted [see equation (8.38g)]:

$$\theta_r = \frac{1}{2}\tan^{-1}\left(\frac{2M_{ij}}{(M_{jj} - M_{ii})}\right) \tag{9.10}$$

Table 9.7 lists the pivot coordinates and angle formula for the sequence of similarity transforms to be applied to the folded coupling matrix for degrees 4–9 and a general formula for the pivot coordinates for any degree ≥ 4.

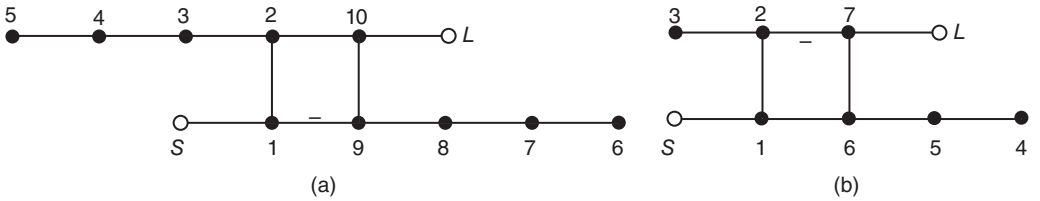


Figure 9.16 Cul-de-sac network configurations: (a) (10–3–4) filter network; (b) (7–1–2) filter network.

Table 9.7 Pivot coordinates for reduction of the folded matrix to the cul-de-sac configuration.

	Pivot position [*i*, *j*] and element to be annihilated				
	Similarity transform number *r*				
	r* = 1, 2, 3, … , *R		***R* = (*N* − 2)/2 (*N* even) = (*N* − 3)/2 (*N* odd)**		**Transform angle formula**
Degree *N*	***r* = 1**	**2**	**3**	***r***	θ_r
4	[2, 3] M_{23}				(9.10)
5	[2, 4] M_{23}				(9.9)
6	[3, 4] M_{34}	[2, 5] M_{25}			(9.10)
7	[3, 5] M_{34}	[2, 6] M_{25}			(9.9)
8	[4, 5] M_{45}	[3, 6] M_{36}	[2, 7] M_{27}		(9.10)
9	[4, 6] M_{45}	[3, 7] M_{36}	[2, 8] M_{27}		(9.9)
—	—	—	—	—	—
N (even)	$[i,j]$ $M_{i,j}$ $i = (N+2)/2 - 1$ $j = N/2 + 1$	—	—	$[i,j]$ $M_{i,j}$ $i = (N+2)/2 - r$ $j = N/2 + r$	(9.10)
N (odd)	$[i,j]$ $M_{i,j-1}$ $i = (N+1)/2 - 1$ $j = (N+1)/2 + 1$	—	—	$[i,j]$ $M_{i,j-1}$ $i = (N+1)/2 - r$ $j = (N+1)/2 + r$	(9.9)

The folded coupling matrix for the doubly terminated version of the example, which is employed to demonstrate the coupling matrix reduction process in Section 8.3, is shown in Figure 9.17a This is a seventh-degree, 23 dB return loss asymmetric filter, with a complex pair of transmission zeros to give group delay equalization over approximately 60% of the bandwidth and a single zero on the imaginary axis to give a rejection lobe level of 30 dB on the upper side of the passband.

To transform the folded network coupling matrix to the cul-de-sac configuration, two transformations, according to Table 9.7, are applied (pivots [3, 5] and [2, 6]) with the angles calculated from equation (9.9). Then the coupling matrix in Figure 9.17b is obtained, corresponding to the coupling–routing diagram of Figure 9.16b. The results of analyzing this coupling matrix are plotted in Figure 9.18a (rejection/return loss) and Figure 9.18b (group delay). It is evident that the 30 dB lobe level and equalized in-band group delay are not affected by the transformation process. Note that when the source and load are directly connected to the corners of the core quartet and the number of finite-position zeros of the prototype is less than $N - 3$, the values of the four couplings in the core quartet have the same absolute value.

The topologies of the filters in coaxial resonator technology, corresponding to the two coupling matrices of Figure 9.17, are shown in Figure 9.19a (folded) and Figure 9.19b (cul-de-sac). This demonstrates the rather simple construction of the cul-de-sac form, compared with that of the folded form. There are no diagonal cross-couplings and only one negative coupling in the cul-de-sac network.

9.6.1 Further Cul-de-Sac Forms

The cul-de-sac filter is defined in Figure 9.16 for an even-degree example and an odd-degree example. This basic form consists of a core quartet of resonator nodes with two opposite corners directly connected to the source and load terminations; the rest of the resonator nodes form

	S	1	2	3	4	5	6	7	*L*
S	0	1.0572	0	0	0	0	0	0	0
1	1.0572	0.0211	0.8884	0	0	0	0	0	0
2	0	0.8884	0.0258	0.6159	0	0	0.0941	0	0
3	0	0	0.6159	0.0193	0.5101	0.1878	0.0700	0	0
4	0	0	0	0.5101	–0.4856	0.4551	0	0	0
5	0	0	0	0.1878	0.4551	–0.0237	0.6119	0	0
6	0	0	0.0941	0.0700	0	0.6119	0.0258	0.8884	0
7	0	0	0	0	0	0	0.8884	0.0211	1.0572
L	0	0	0	0	0	0	0	1.0572	0

(a)

	S	1	2	3	4	5	6	7	*L*
S	0	1.0572	0	0	0	0	0	0	0
1	1.0572	0.0211	0.6282	0	0	0	0.6282	0	0
2	0	0.6282	–0.0683	0.5798	0	0	0	–0.6282	0
3	0	0	0.5798	–0.1912	0	0	0	0	0
4	0	0	0	0	–0.4856	0.6836	0	0	0
5	0	0	0	0	0.6836	0.1869	0.6499	0	0
6	0	0.6282	0	0	0	0.6499	0.1199	0.6282	0
7	0	0	–0.6282	0	0	0	0.6282	0.0211	1.0572
L	0	0	0	0	0	0	0	1.0572	0

(b)

Figure 9.17 Example of the seventh-degree filter in cul-de-sac configuration: (a) original folded coupling matrix; (b) after transformation to cul-de-sac configuration.

two chains connected to the other two corners. The form progressively evolves as a series of R rotations is applied to an originating folded coupling matrix, with pivots and angles as listed in Table 9.7. The number of rotations to be applied is given by $R = (N - 2)/2$ (N even) or $R = (N - 3)/2$ (N odd).

Two further forms can be obtained by stopping the series of rotations at the penultimate (i.e., $R - 1$ rotations), or continuing on with one extra rotation ($R + 1$ rotations). The first form leaves the core quartet not directly connected to the source and load terminations, as shown in Figure 9.20a for an eighth-degree example. The minimum path rule dictates that this form realizes fewer finite-position transmission zeros: $n_{fz} = N - 5$ for these cases, compared with $N - 3$ for the basic form. However, the advantage is that there is a greater isolation between input and output and the fact that the topology is suitable for realization in dual-mode cavities.

By continuing on with one extra rotation, the canonical cul-de-sac form is obtained, where the source and load terminations form part of the core quartet as in Figure 9.20b. This form realizes $N - 1$ transmission zeros, and, by including the direct source–load coupling M_{SL} (shown by the dashed line in Figure 9.20b), the form becomes fully canonical ($n_{fz} = N$). It can be demonstrated that the two branches form the even-mode and odd-mode subnetworks of the filtering function, and so, become very similar in principle of operation to the hybrid reflection mode bandpass and bandstop filters, described by Hunter and others [12, 13].

If extra unit coupling inverters are added to the network at the input and output ports so creating an $N + 4$ matrix (the additional inverters will have no effect on S_{11} and S_{21} except to change their phases by 180°), and then the rotation sequence is carried on by yet another rotation, a situation arises where all four nodes in the core quartet are non-resonant, as shown in Figure 9.20c. The values of the couplings in the core quartet will be $1/\sqrt{2}$, and together with the negative sign, it becomes evident that the core is a rat-race coupler. This is easily implemented in microstrip

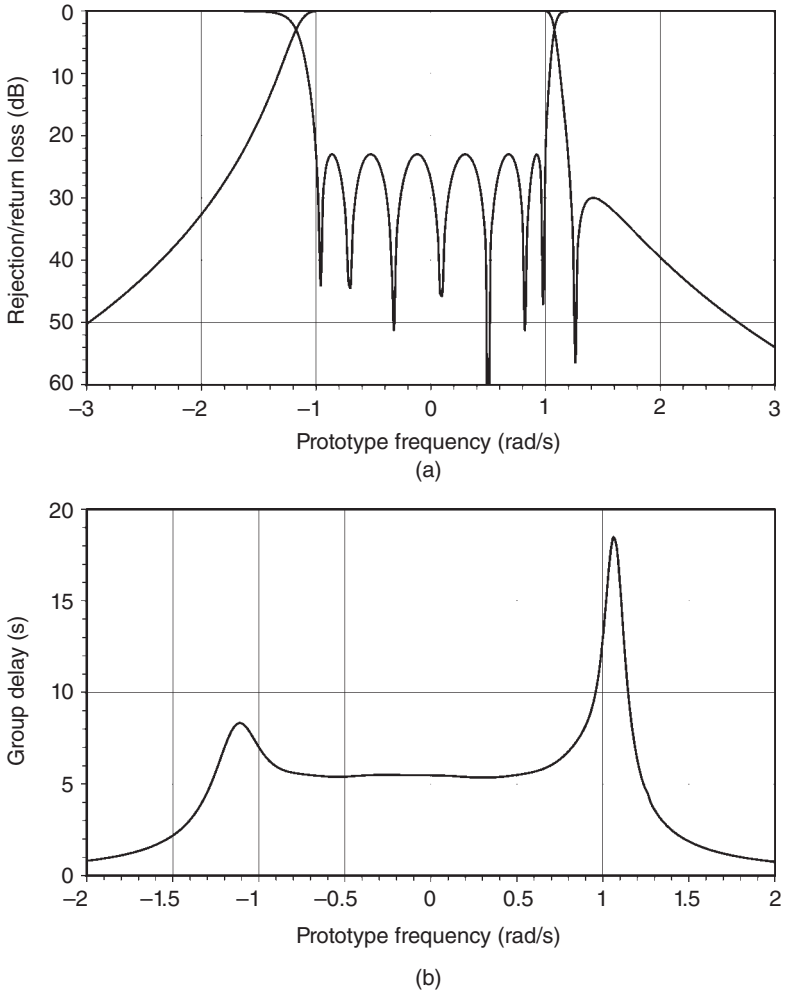


Figure 9.18 Example of the simulated performance of a (7–1–2) asymmetric filter: (a) rejection and return loss; (b) group delay.

or other planar technology, where the negative branch is realized with a 270° length instead of 90°. As with the network in Figure 9.20b, the two branches form the even-mode and odd-mode subnetworks of the filtering function, synthesized automatically by the cul-de-sac procedure.

To demonstrate the synthesis of the canonical form of the cul-de-sac network (Figure 9.20b), the sixth-degree characteristic that is an example for the symmetric inline synthesis (Section 9.1.1) is used again. We take the $N+2$ folded coupling matrix for this filter and apply three cross-pivot rotations at [3, 4], [2, 5], and [1, 6] to annihilate M_{34}, M_{25}, and M_{16}, respectively. Figure 9.21 shows the coupling matrix. It can be seen that the two branches of the canonical cul-de-sac network form the even-mode and odd-mode networks, as derived in Sections 9.1.1 and 9.5.1.

9.6.1.1 Proof-of-Concept Model

To illustrate the previous discussion, a proof-of-concept S-band model, realizing an asymmetric eighth-degree filter function, is synthesized in the indirect-coupled cul-de-sac

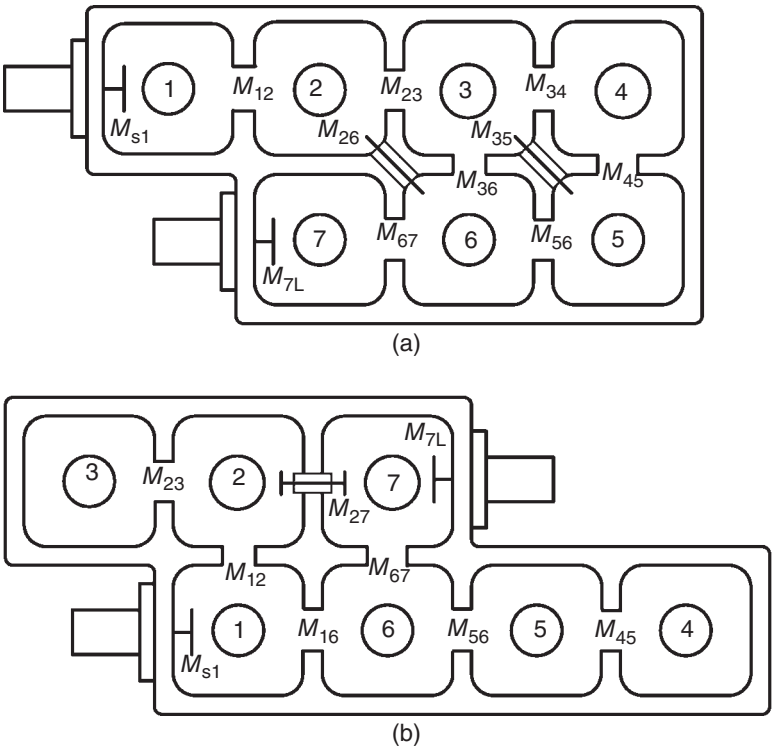


Figure 9.19 Example of a (7–1–2) asymmetric filter in coaxial cavity configurations: (a) folded network configuration; (b) cul-de-sac configuration.

configuration in Figure 9.20a, realized as a coaxial cavity filter, and measured. The prototype is an eighth-degree Chebyshev filter, with a 23 dB return loss and three prescribed transmission zeros at $s = -j1.3553$, $+j1.1093$, and $+j1.2180$, producing one rejection lobe level of 40 dB on the lower side of the passband and two of 40 dB on the upper side. Starting with the folded coupling matrix for this prototype (Figure 9.22a) and applying a series of $R - 1 = 2$ similarity transforms at pivots [4, 5] and [3, 6] of Table 9.7, we achieve the coupling matrix in Figure 9.22b, which corresponds to the configuration in Figure 9.20a.

The simulated and measured rejection and return loss performances of this filter are shown in Figure 9.23, which clearly shows the three transmission zeros. The somewhat large bandwidth of this filter (110 MHz) means that the return loss performance has degraded slightly because of the dispersion effects (the coupling values varying with frequency); however, these effects seem to be well predicted by the simulation model.

A sketch of the filter resonator layout is provided in Figure 9.24, showing the very uncomplicated coupling arrangement. The omission of the third rotation has left the core quartet (2, 3, 6, and 7) not directly connected to the input/output terminations, which should benefit far-out-of-band rejection. This same construction can support 8th-degree prototypes with the number of transmission zeros ranging from 0 to 3 and with symmetric or asymmetric characteristics.

As noted, all the couplings are positive except for one in the core quartet. It can be moved to any one of the four couplings for the greatest convenience and implemented as a capacitive probe, for example, if the filter is to be realized in coaxial resonator technology and the other

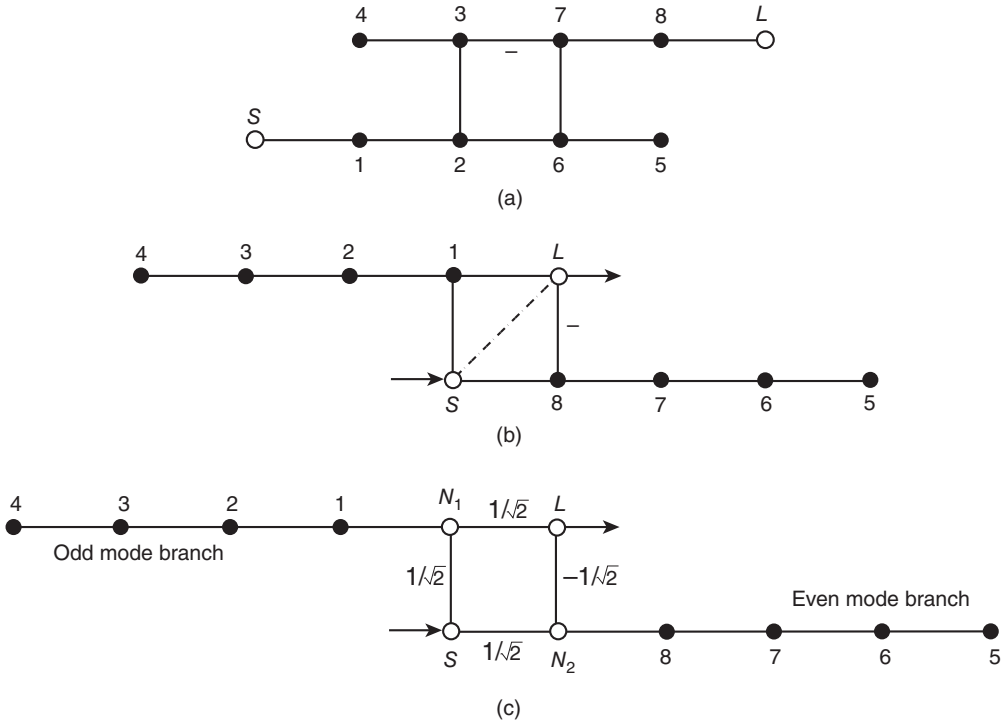


Figure 9.20 Three alternative forms for the cul-de-sac configuration: (a) the indirect-coupled, (b) fully canonical form, and (c) rat-race coupled even- and odd-mode one-port networks.

	S	1	2	3	4	5	6	*L*
S	0	0.7472	0	0	0	0	0.7472	0
1	0.7472	0	0.8867	0	0	0	0	0.7472
2	0	0.8867	–0.1337	0.6050	0	0	0	0
3	0	0	0.6050	0.7007	0	0	0	0
4	0	0	0	0	–0.7007	0.6050	0	0
5	0	0	0	0	0.6050	0.1337	0.8867	0
6	0.7472	0	0	0	0	0.8867	0	–0.7472
L	0	0.7472	0	0	0	0	–0.7472	0

Figure 9.21 Canonical cul-de-sac coupling matrix for symmetric (6–2) characteristic.

couplings as inductive apertures or inductive loops. Moreover, there are no diagonal couplings, even though the original prototype is asymmetric.

Some of the more important features of the cul-de-sac configuration are now summarized:

1) The cul-de-sac topology may be used to realize a completely general class of electrical filter function. It can be symmetric or asymmetric, even or odd degree, provided that the filter function is designed to operate between the equivalue source and load terminations (double terminated) and that the number of transmission zeros embodied by the filtering function is less than or equal to $N - 3$, where N is the degree of the function.
2) A filter function, incorporating transmission zeros, is realized with an absolute minimum number of coupling elements and their associated adjustment mechanisms. For a 10th-degree function with seven transmission zeros, 10 internal couplings are required.

	S	1	2	3	4	5	6	7	8	*L*
S	0	1.0428	0	0	0	0	0	0	0	0
1	1.0428	0.0107	0.8623	0	0	0	0	0	0	0
2	0	0.8623	0.0115	0.5994	0	0	0	0	0	0
3	0	0	0.5994	0.0133	0.5356	0	–0.0457	–0.1316	0	0
4	0	0	0	0.5356	0.0898	0.3361	0.5673	0	0	0
5	0	0	0	0	0.3361	–0.8513	0.3191	0	0	0
6	0	0	0	–0.0457	0.5673	0.3191	–0.0073	0.5848	0	0
7	0	0	0	–0.1316	0	0	0.5848	0.0115	0.8623	0
8	0	0	0	0	0	0	0	0.8623	0.0107	1.0428
L	0	0	0	0	0	0	0	0	1.0428	0

(a)

	S	1	2	3	4	5	6	7	8	*L*
S	0	1.0428	0	0	0	0	0	0	0	0
1	1.0428	0.0107	0.8623	0	0	0	0	0	0	0
2	0	0.8623	0.0115	0.3744	0	0	–0.4681	0	0	0
3	0	0	0.3744	–0.0439	0.8166	0	0	0.3744	0	0
4	0	0	0	0.8166	0.1976	0	0	0	0	0
5	0	0	0	0	0	–0.9590	0.2093	0	0	0
6	0	0	–0.4681	0	0	0.2093	0.0499	0.4681	0	0
7	0	0	0	0.3744	0	0	0.4681	0.0115	0.8623	0
8	0	0	0	0	0	0	0	0.8623	0.0107	1.0428
L	0	0	0	0	0	0	0	0	1.0428	0

(b)

Figure 9.22 (8–3) asymmetric filter example: (a) folded coupling matrix; (b) same matrix after transformation to the indirect cul-de-sac configuration.

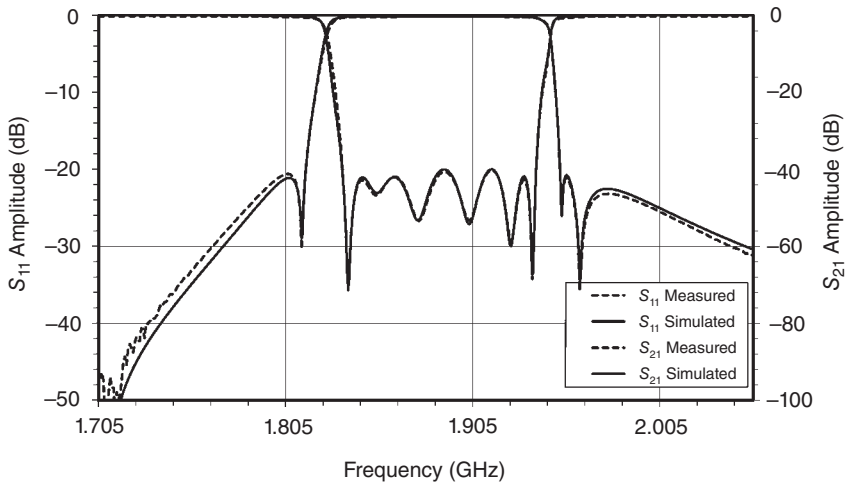


Figure 9.23 Simulated and measured performance of the (8–3) cul-de-sac filter.

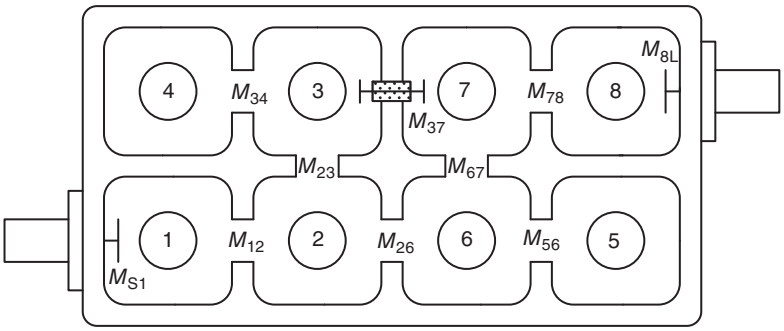


Figure 9.24 Indirect-coupled cul-de-sac configuration for the (8–3) asymmetric filter.

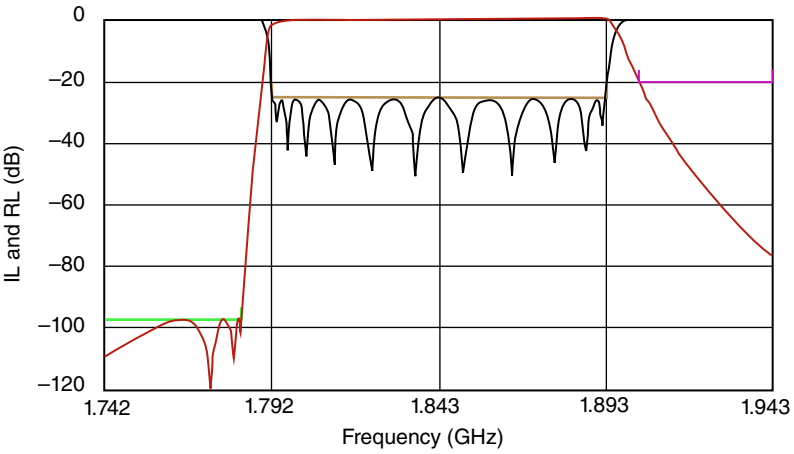
The equivalent folded structure requires 16 internal couplings, some of which are diagonal couplings.

3) All the couplings are straight, that is, they are either horizontal or vertical between the resonators, if these are arranged in a regular grid pattern. This significantly simplifies the design and eliminates the need for diagonal couplings, which are typically difficult to manufacture, assemble, and adjust and are quite sensitive to vibration and temperature variations. The absence of diagonal couplings enables the realization of asymmetric filter characteristics with dual-mode resonators, such as in dielectric or waveguide, and also usage of the high-Q_u TE_{011} cylindrical waveguide resonant mode.
4) The same structure can be used to embody any doubly terminated symmetric or asymmetric filter function of the same degree (provided that $n_{fz} \leq N - 3$). Therefore, the same basic housing, incorporating the coupling elements, can be used for a variety of filter characteristics. If the changes in the coupling values from the original design are small, for example, to correct for a dispersion distortion, the changes are achieved with tuning and iris adjustment screws without the need to alter the mechanical dimensions of the interresonator coupling apertures or to add further cross-couplings.
5) Regardless of the transfer characteristic, all the internal couplings have the same sign, with the exception of the one within the central quartet of resonators. The range over which the values of the internal interresonator couplings are spread is comparatively small, allowing the use of a common design for all the couplings with the same sign.
6) The values of the interresonator couplings are relatively large, compared with the levels of stray and spurious couplings typically found within complex filter structures. Not only is the design and tuning process eased, but the values of all interresonator couplings are also externally adjustable with screws, reducing the need for tight tolerances during manufacture and enabling optimal performance to be obtained during production.
7) The Q_u factor of a microwave resonant cavity decreases as the number of coupling elements is augmented, increasing the insertion loss of the filter of which the cavity forms a part. Owing to the minimal number of interresonator couplings in the cul-de-sac filter, with only one of those being a negative coupling (usually realized as a probe in coaxial and dielectric filters and can be relatively lossy, particularly if it is a diagonal coupling), the insertion loss is less than that of the equivalent filter with more coupling elements.
8) By short-circuiting certain cavities, the filter is split into smaller sections, easing the development and production tuning processes.
9) The cul-de-sac filter can be manufactured as a simple two-dimensional two-part housing and lid assembly with all the tuning screws on top. The result is a filter that is appropriate for mass production and tuning methods. The form is amenable to volume manufacturing processes

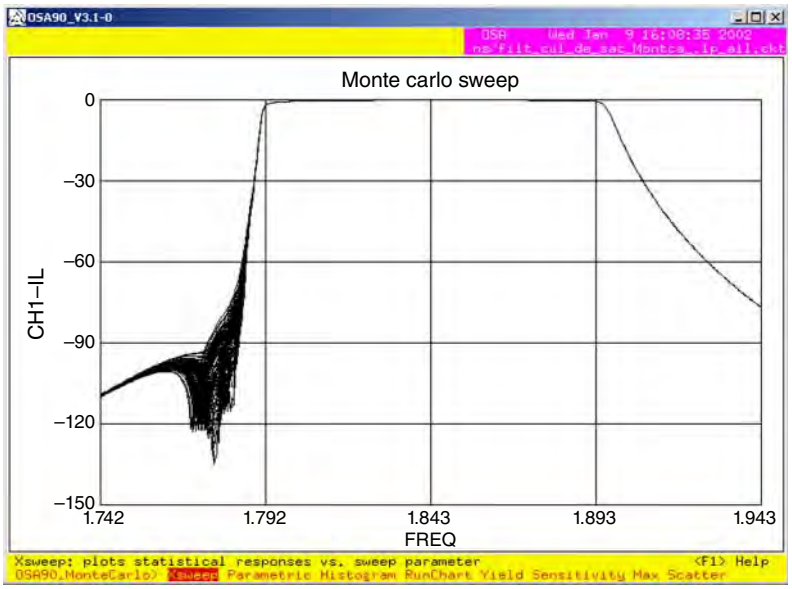
such as die casting. Moreover, a certain amount of flexibility is available in the layout of the filter cavities, enabling compact integration with other subsystems and components in the same system.

9.6.2 Sensitivity Considerations

Having the minimum number of internal couplings, it is not surprising that the cul-de-sac filter is more sensitive to variations in its coupling element values and tuning state than the equivalent filter realized in the folded configuration, or with trisections and so on. To assess the nature of this sensitivity and compare the alternative configurations for the same filter function, a brief qualitative simulation study was made.



(a)



(b)

Figure 9.25 Sensitivity to variations in the coupling value: (a) nominal 11–3 characteristic; (b) cul-de-sac—random variations to all the couplings; (c) cul-de-sac—0.2% increase to the self-couplings, 0.2% decrease to all others; (d) box filter—0.2% increase to self-couplings, 0.2% decrease to all others.

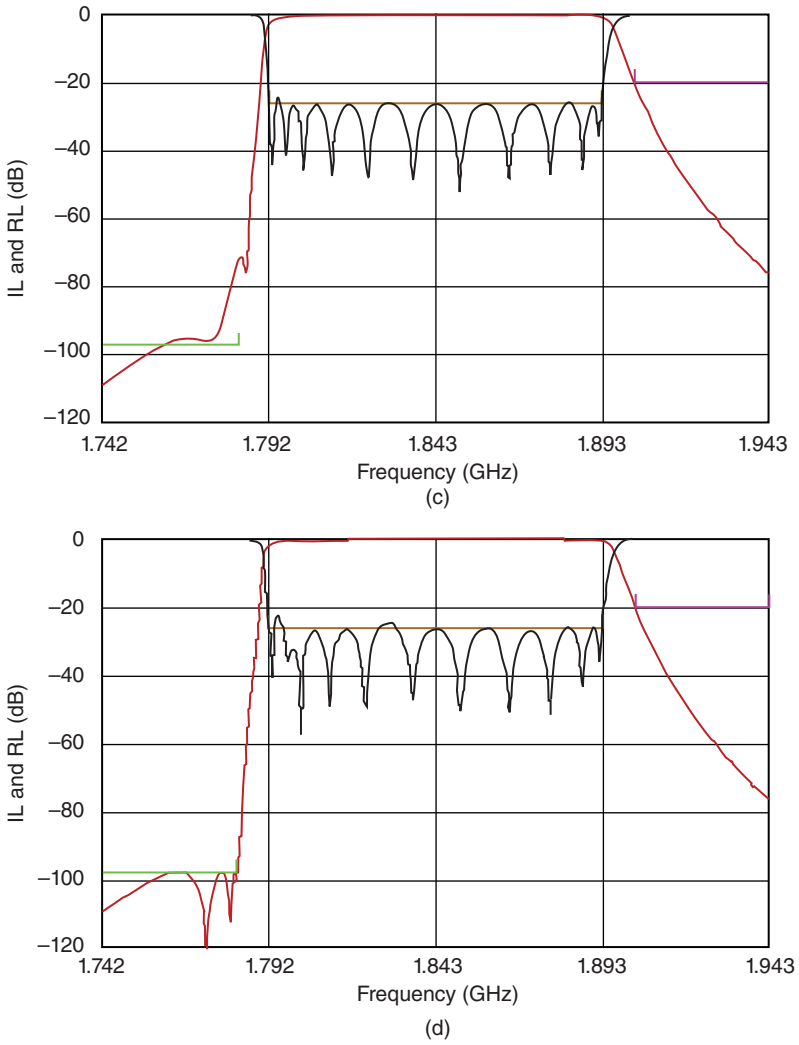


Figure 9.25 (*Continued*)

There are two principal types of variations in the coupling and resonator tuning states: *random*, due to component and manufacturing/assembly tolerances, and *uni-* or *bidirectional*, caused by temperature fluctuations during operation. The first category is assessed by a Monte Carlo sensitivity analysis package on an 11th-degree prototype, the nominal rejection and return loss characteristics of which are shown in Figure 9.25a. There are three transmission zeros on the lower side, providing 97 dB rejection lobe levels, realized in folded, box section, trisection (see Chapter 10), and cul-de-sac configurations. Here, the box and trisection were relatively little affected by 1% standard deviation variations in coupling values. However, the rejection performances of the folded and cul-de-sac configurations are substantially affected, as plotted in Figure 9.25b. Although the variations in the values in the coupling elements are adjustable with tuning screws during production, counteracting the effect of the manufacturing and assembly tolerances and restoring the rejection performance, there is an impact on volume production where more accurate tuning is required.

The second category of variation is investigated by applying a 0.2% increase to the self-couplings and a 0.2% decrease to the other coupling values. Here, the cul-de-sac fares the worst overall with the rejection performance suffering significantly as traced in Figure 9.25c. By comparison, the box filter is relatively unaffected (Figure 9.25d). This indicates that the thermal stability of the coupling values and resonant frequencies needs to be improved or that the mechanical design gives shifts in the same direction under the influence of temperature changes.

Summary

In Chapters 7 and 8, two different techniques for synthesizing the canonical folded array coupling matrix from the transfer and reflection polynomials of the desired filter function are described. For certain applications, this canonical form is used directly, if it is convenient to realize the couplings. Alternatively, the form can be used as a starting point for the application of further transforms to create topologies better suited for the technology with which the filter is eventually realized. One such technology that has had a major impact on filter design is the dual-mode filter realization, which utilizes two orthogonally polarized degenerate modes, supported in a single physical resonator, either a cavity, a dielectric disk, or a planar structure. The technology allows a significant reduction in the size of filters with a minor penalty in the unloaded Q and spurious response. Owing to the large reduction in size and mass, dual-mode filters are widely used in satellite and wireless communication systems. This chapter is devoted to the methods of similarity transformations to realize a wide range of topologies, appropriate for dual-mode filter networks. Besides longitudinal and folded structures, the chapter also includes structures referred to as *cascade quartets* and *cul-de-sac filters*. The chapter concluded with examples and a discussion of the sensitivity of some dual-mode filter topologies.

References

1 Rhodes, J.D. (1970) A lowpass prototype network for microwave linear phase filters. *IEEE Transactions on Microwave Theory and Techniques*, **18**, 290–301.

2 Rhodes, J.D. (1970) The generalized interdigital linear phase filter. *IEEE Transactions on Microwave Theory and Techniques*, **18**, 301–307.

3 Rhodes, J.D. (1970) The generalized direct-coupled cavity linear phase filter. *IEEE Transactions on Microwave Theory and Techniques*, **18**, 308–313.

4 Rhodes, J.D. and Zabalawi, I.H. (1980) Synthesis of symmetrical dual mode in-line prototype networks, in *Circuit Theory and Application*, vol. **8**, Wiley, New York, pp. 145–160.

5 Abramowitz, M. and Stegun, I.A. (eds) (1970) *Handbook of Mathematical Functions*, Dover Publications, New York.

6 Cameron, R.J. and Rhodes, J.D. (1981) Asymmetric realizations for dual-mode bandpass filters. *IEEE Transactions on Microwave Theory and Techniques*, **29**, 51–58.

7 Pfitzenmaier, G. (1977) An exact solution for a six-cavity dual-mode elliptic bandpass filter. IEEE MTT-S International Microwave Symposium Digest, San Diego, pp. 400–403.

8 Cameron, R.J. (1979) Novel realization of microwave bandpass filters. *ESA Journal*, **3**, 281–287.

9 Cameron, R.J., Harish, A.R., and Radcliffe, C.J. (2002) Synthesis of advanced microwave filters without diagonal cross-couplings. *IEEE Transactions on Microwave Theory and Techniques*, **50**, 2862–2872.

10 Pommier, V., Cros, D., Guillon, P., Carlier, A. and Rogeaux, E. (2000) Transversal filter using whispering gallery quarter-cut resonators, IEEE MTT-S International Symposium Digest, Boston, pp. 1779–1782.
11 Bell, H.C. (1982) Canonical asymmetric coupled-resonator filters. *IEEE Transactions on Microwave Theory and Techniques*, **30**, 1335–1340.
12 Hunter, I.C., Rhodes, J.D., and Dassonville, V. (1998) Triple mode dielectric resonator hybrid reflection filters. *IEE Proceedings of Microwaves, Antennas and Propagation*, **145**, 337–343.
13 Hunter, I.C. (2001) *Theory and Design of Microwave Filters* Electromagnetic Waves Series 48, IEE, London.

10

Synthesis and Application of Extracted Pole and Trisection Elements

Although the folded and transversal array coupling matrices form bases from which a wide variety of useful filter configurations may be derived, there are certain sections that are difficult or impossible to synthesize by coupling matrix operations alone. In this chapter, we shall follow the circuit-based method for synthesizing the "extracted pole" section and its direct application in bandpass and bandstop filter design. Then the rôle of the extracted pole section in the formation of another useful section able to realize a single transmission zero (TZ), the "trisection," will be examined. The trisection may be realized directly or used as part of a cascade with other sections. However it also finds application in the synthesis of more complex networks, where firstly a cascade of trisections is synthesized and then further reconfiguration is carried out using coupling matrix transforms. The techniques used to transform a cascade of trisections to form cascaded quartets, quintets, and sextets are outlined. Finally the synthesis of the "box section" from a trisection is described and its derivative, the "extended box" configuration.

10.1 Extracted Pole Filter Synthesis

The low-loss requirement for high-power filters usually means that the designs are low degree, with no in-band equalization provision (group delay equalization increases the insertion loss). Other requirements are a relatively steep out-of-band isolation to minimize the effect of spectrum spreading (due to the nonlinear nature of the high power amplifiers, HPAs) and to provide adequate channel-to-channel isolation.

Although it is simple to synthesize the prototype network for a low-degree filter with symmetric or asymmetric TZs, its realization can present some difficulties if negative cross-couplings are required.[1] In a low-power environment, the negative coupling is often realized as a capacitive probe; however, in high-power conditions, this probe can overheat (it is usually thermally decoupled from an effective heat sink) and cause arcing problems and an overall reduction in the Q_u of the cavity.

The need for negative couplings can be avoided by using the extracted pole principle. In effect, the TZs are removed from the network before the synthesis of the main body and realized as bandstop cavities. After the TZs are extracted, the remaining network does not require cross-couplings, either straight or diagonal, unless there are further TZs that have not been extracted. If the remaining zeros are a real-axis pair for the group delay equalization of a symmetric filter, the cross-couplings needed for their realization are the same sign as the mainline couplings.

1 "Negative" here means negative with respect to the other couplings in the network and inductive if the other couplings are capacitive, and vice versa.

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

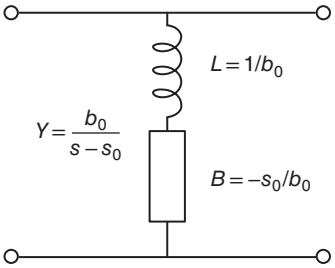


Figure 10.1 Shunt-connected series pair producing a short circuit at $s = s_0$.

10.1.1 Synthesis of the Extracted Pole Element

In the following section, the circuit-based procedure for extracting single (imaginary-axis) TZs is detailed. We follow the synthesis procedure based on the (4–2) asymmetric prototype filter example in Section 7.3. The procedure outlined here gives more flexibility than the method given in Ref. [1], which is for conjugate pole pairs in symmetric networks.

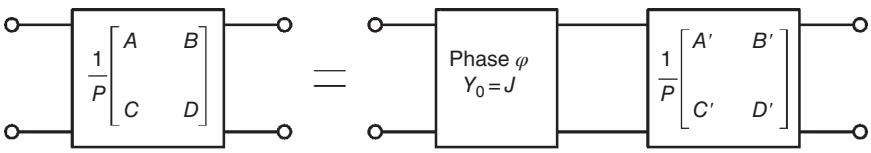
The prototype circuit that produces the TZ (or pole) is shown in Figure 10.1. It consists of an frequency-invariant reactance (FIR) of value $B = -s_0/b_0$ in series with an inductive element $L = 1/b_0$, shunt connected to the transmission line; b_0 is known as the *residue* of the pole. It is evident that when $s = s_0$, the impedance of the series pair is zero, short-circuiting the transmission line.

The pole is extracted from the [*ABCD*] matrix that represents the overall filter network in a three-stage operation:

1) A frequency-invariant phase shifter is first removed to prepare the circuit for extraction of the pole at $s = s_0$ (partial removal).
2) The shunt-connected resonant pair is removed at $s = s_0$.
3) A further phase length is extracted to prepare the remaining network for the extraction of another pole if required or for the extraction of a capacitor/FIR pair ($C + jB$) at the input to the main body of the filter.

10.1.1.1 Removal of Phase Length

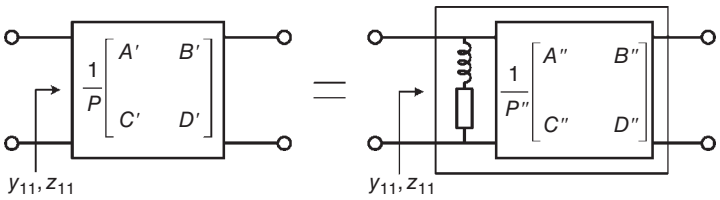
The [*ABCD*] matrix, representing the overall circuit, is considered as a cascade of a length of transmission line in cascade with a remainder matrix [$A'B'C'D'$] [the (s) has been omitted from the polynomials here for reasons of clarity, and so $A \equiv A(s)$, $B \equiv B(s)$, etc. and $P(\equiv P(s)/\varepsilon)$ is assumed to have absorbed the constant ε]:



$$\frac{1}{P}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cos\varphi & j\sin\varphi/J \\ jJ\sin\varphi & \cos\varphi \end{bmatrix} \cdot \frac{1}{P}\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix}$$

$$\therefore \frac{1}{P}\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \frac{1}{P}\begin{bmatrix} A\cos\varphi - jC\sin\varphi/J & B\cos\varphi - jD\sin\varphi/J \\ C\cos\varphi - jAJ\sin\varphi & D\cos\varphi - jBJ\sin\varphi \end{bmatrix} \tag{10.1}$$

The remainder matrix contains the shunt-connected resonant pair as the leading element:



The open-circuit admittance z_{11} at the input to the remainder matrix approaches zero and the short-circuit admittance y_{11} approaches infinity as s approaches the pole frequency s_0, expressed as

$$z_{11} = \frac{A'}{C'} = 0 \quad \text{and} \quad y_{11} = \frac{D'}{B'} = j\infty \quad \text{at } s = s_0 \tag{10.2}$$

Therefore, $A' = 0$ or $B' = 0$ at $s = s_0$. Taking the A' and B' elements from matrix (10.1), we obtain

$$A' = A\cos\varphi - \frac{jC\sin\varphi}{J} = 0,$$

$$\therefore \tan\varphi_0 = \left.\frac{AJ}{jC}\right|_{s=s_0} \quad \text{or} \quad J_0 = \left.\frac{jC\tan\varphi}{A}\right|_{s=s_0} \tag{10.3a}$$

$$B' = B\cos\varphi - \frac{jD\sin\varphi}{J} = 0,$$

$$\therefore \tan\varphi_0 = \left.\frac{BJ}{jD}\right|_{s=s_0} \quad \text{or} \quad J_0 = \left.\frac{jD\tan\varphi}{B}\right|_{s=s_0} \tag{10.3b}$$

Equations (10.3) show that either the admittance J of the inverter can be prescribed, giving a phase length φ_0 at which A' and B' are both equal zero, or the phase length φ can be prescribed ($\neq 90°$), which gives a characteristic admittance J_0 by which A' and B' go to zero. In practice, it is more common to prescribe $J = 1$ so that the interconnecting phase lengths are the same characteristic admittance as those of the interfacing transmission-line medium. If J is prescribed to be unity, the phase shift φ_0 is evaluated from equation (10.3), resulting in

$$\varphi_0 = \tan^{-1}\left.\frac{A}{jC}\right|_{s=s_0} \tag{10.4a}$$

$$\varphi_0 = \tan^{-1}\left.\frac{B}{jD}\right|_{s=s_0} \tag{10.4b}$$

Because the polynomials A' and B' go to zero when $s = s_0$, they are both divisible by the same factor $(s - s_0)$. This factor can be divided out to obtain the intermediate polynomials A^x and B^x:

$$A^x = \frac{A'}{(s - s_0)}, \quad B^x = \frac{B'}{(s - s_0)} \tag{10.5}$$

10.1.1.2 Removal of Resonant Pair

After extraction of the phase length, the residue b_0 of the resonant pair is evaluated. Referring to Figure 10.2, we observe that as s approaches s_0, the impedance or admittance of the network

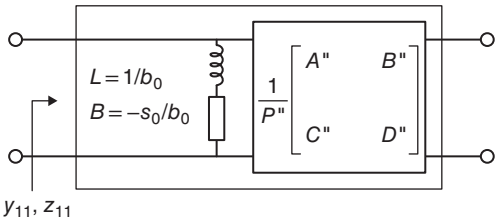


Figure 10.2 Input impedance/admittance as s approaches s_0.

is dominated by the shunt resonant pair:

At $s = s_0$

$$y_{11} = \frac{D'}{B'} = \left.\frac{b_0}{(s-s_0)}\right|_{s=s_0}, \quad \therefore\ b_0 = \left.\frac{(s-s_0)D'}{B'}\right|_{s=s_0} = \left.\frac{D'}{B^x}\right|_{s=s_0} \tag{10.6a}$$

$$z_{11} = \frac{A'}{C'} = \left.\frac{(s-s_0)}{b_0}\right|_{s=s_0}, \quad \therefore\ b_0 = \left.\frac{(s-s_0)C'}{A'}\right|_{s=s_0} = \left.\frac{C'}{A^x}\right|_{s=s_0} \tag{10.6b}$$

Now we extract the resonant pair

$$\begin{aligned}\frac{1}{P''}\begin{bmatrix}A'' & B''\\ C'' & D''\end{bmatrix} &= \begin{bmatrix}1 & 0\\ -b_0/(s-s_0) & 1\end{bmatrix}\cdot\frac{1}{P}\begin{bmatrix}A' & B'\\ C' & D'\end{bmatrix}\\ &= \frac{1}{P}\begin{bmatrix}A' & B'\\ C'-A'b_0/(s-s_0) & D'-B'b_0/(s-s_0)\end{bmatrix}\\ &= \frac{1}{P}\begin{bmatrix}A' & B'\\ C'-b_0A^x & D'-b_0B^x\end{bmatrix}\end{aligned} \tag{10.7a}$$

where $[A''B''C''D'']$ is the remainder matrix. By multiplying the right-hand matrix in equation (10.7a) top and bottom by the factor $(s-s_0)$, we obtain

$$\frac{1}{P''}\begin{bmatrix}A'' & B''\\ C'' & D''\end{bmatrix} = \frac{(s-s_0)}{P}\begin{bmatrix}\dfrac{A'}{(s-s_0)} & \dfrac{B'}{(s-s_0)}\\ \dfrac{C'-b_0A^x}{(s-s_0)} & \dfrac{D'-b_0B^x}{(s-s_0)}\end{bmatrix} \tag{10.7b}$$

Thus it is easily seen that $A'' = A'/(s-s_0) = A^x$ and $B'' = B'/(s-s_0) = B^x$ and that the numerators of C'' and D'' equal $C'-b_0A^x$ and $D'-b_0B^x$, respectively. Since $b_0 = C'/A^x = D'/B^x$ at $s=s_0$ [equation (10.6)], the numerators of C'' and D'' in equation (10.7b) go to zero at $s=s_0$. So, by extracting the resonant pair with residue b_0, numerators of C'' and D'' are conditioned to be divisible by the factor $(s-s_0)$, in the same way as the extraction of the phase length conditioned the numerators of A' and B' to be exactly divisible by $(s-s_0)$. By definition, the P polynomial of the original matrix is also divisible by $(s-s_0)$.

The process for extracting a single TZ at $s=s_0$ and obtaining the remainder matrix $[A''B''C''D'']$ is summarized as follows:

1) Calculate the phase length φ_0 or characteristic admittance J_0 of the length of transmission line to be extracted from (10.3).
2) Extract the transmission line to obtain the matrix $[A'B'C'D']$ [equation (10.1)].
3) Divide A' and B' by the factor $(s-s_0)$ to obtain the intermediate polynomials $A^x = A''$ and $B^x = B''$ from equation (10.5).
4) Calculate the value of the residue b_0 from equation (10.6).
5) Compute $C'-b_0A^x$ and $D'-b_0B^x$.

6) Divide these two polynomials by the factor $(s - s_0)$ to obtain C'' and D''.
7) Finally, divide the P polynomial by the factor $(s - s_0)$ to obtain P''.

Thus, the extraction of the pole has left all the polynomials in the remainder matrix one degree less than the corresponding polynomials in the original matrix. From the remainder matrix, a second pole may now be extracted by removing another length of transmission line at the new pole frequency s_{02}. If the next component to be extracted is an inverter or a shunt-connected $C + jB$ pair to start the synthesis of the main body of the filter, the phase length is removed at $s_0 = j\infty$. The main body of the filter itself is synthesized as a cross-coupled network to produce further TZs (usually real-axis pairs for group delay equalization), and if required, the filter's topology can be reconfigured with rotations (subject to the usual restrictions), regardless of whether extracted poles are present. Also, the procedure can be applied to create extracted poles within the main body of the filter, separating the mainline resonators into groups (see Figure 10.6b). This has certain advantages in tuning the filter but does tend to result in rather extreme values for the coupling elements.

10.1.2 Example of Synthesis of Extracted Pole Network

To illustrate the extracted pole procedure, the (4–2) filter for which we have already derived the prototype characteristics and $[ABCD]$ polynomials is used again to synthesize an extracted pole circuit. The method for single pole extraction described in the previous section allows for the extraction of a pole at any stage of the synthesis of the network, even between the main-body resonators. For a two-pole network, synthesizing the two poles to be physically on either side of the network gives the most practical coupling values; however, for the purposes of this demonstration, we choose to synthesize both poles to be physically on the input side of the filter. The procedure starts with the $[ABCD]$ polynomials for the (4–2) characteristic (see Section 7.3):

$s^i, i =$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\varepsilon$
0	$-j2.0658$	-0.1059	-0.1476	$-j2.0658$	$+2.0696$
1	$+2.4874$	$-j4.1687$	$-j3.0823$	$+2.4874$	$+j2.7104$
2	$-j2.1950$	$+4.4575$	$+2.8836$	$-j2.1950$	-0.8656
3	$+2.4015$	$-j1.5183$	—	$+2.4015$	—
4	—	$+2.0$	—	—	—

To prepare the network for the removal of the first pole, extract a transmission line of phase length θ_{S1} and unity characteristic admittance at $s_{01} = j1.8082$ as follows:

$$\theta_{S1} = \tan^{-1} \left. \frac{B(s)}{jD(s)} \right|_{s=s_{01}} = 48.9062^\circ$$

The polynomials after extracting this phase length are listed next:

$s^i, i =$	$A'(s)$	$B'(s)$	$C'(s)$	$D'(s)$	$P(s)/\varepsilon$
0	$-j1.2466$	-1.6265	-1.6539	$-j1.2780$	$+2.0696$
1	-0.6880	$-j4.6146$	$-j3.9006$	-1.5067	$+j2.7104$
2	$-j3.6160$	$+1.2755$	$+0.2411$	$-j4.8021$	-0.8656
3	$+1.5785$	$-j2.8078$	$-j1.8099$	$+0.4343$	—
4	—	$+1.3146$	—	$-j1.5073$	—

Divide the $A'(s)$ and $B'(s)$ polynomials by $(s - s_{01})$ to form the intermediate polynomials $A^x(s)$ and $B^x(s)$:

$s^i, i =$	$A^x(s)$	$B^x(s)$	$C'(s)$	$D'(s)$	$P(s)/\epsilon$
0	+0.6894	$-j$0.8995	−1.6539	$-j$1.2780	+2.0696
1	$-j$0.7618	+2.0546	$-j$3.9006	−1.5067	$+j$2.7104
2	+1.5785	$-j$0.4308	+0.2411	$-j$4.8021	−0.8656
3	—	+1.3146	$-j$1.8099	+0.4343	—
4	—	—	—	$-j$1.5073	—

Calculate the residue of the shunt-connected resonant pair b_{01} from

$$b_{01} = \left.\frac{D'(s)}{B^x(s)}\right|_{s=s_{01}} = 1.9680$$

Compute $C'(s) - b_{01}A^x(s)$ and $D'(s) - b_{01}B^x(s)$, and then divide each by $(s - s_{01})$ to obtain $C''(s)$ and $D''(s)$. Finally divide $P(s)$ by $(s - s_{01})$ to obtain $P''(s)$:

$s^i, i =$	$A''(s)$ $(= A^x(s))$	$B''(s)$ $(= B^x(s))$	$C''(s)$	$D''(s)$	$P''(s)/\epsilon$
0	+0.6894	$-j$0.8995	$-j$1.6650	−0.2722	$+j$1.1446
1	$-j$0.7618	+2.0546	+0.4073	$-j$2.9189	−0.8660
2	+1.5785	$-j$0.4308	$-j$1.8099	+0.5726	—
3	—	+1.3146	—	$-j$1.5073	—
4	—	—	—	—	—

Now the previous steps can be repeated to extract the second pole at $s_{02} = j1.3217$, giving $\theta_{12} = 27.5430°$ and $b_{02} = 3.6800$ and leaving the remainder polynomials:

$s^i, i =$	$A''(s)$	$B''(s)$	$C''(s)$	$D''(s)$	$P''(s)/\epsilon$
0	$-j$0.1200	+0.5082	+1.0240	$-j$1.9122	−0.8660
1	+0.5627	$-j$0.0274	$-j$2.3347	+1.1540	—
2	—	+0.4686	—	$-j$1.9443	—
3	—	—	—	—	—
4	—	—	—	—	—

Prepare the network to remove $C_3 + jB_3$, which is done by extracting a phase length at $s = j\infty$ such that

$$\theta_{23} = \tan^{-1}\left.\frac{B(s)}{jD(s)}\right|_{s=j\infty} = 13.5508°$$

Extracting this phase length leaves the following polynomials:

$s^i, i=$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\varepsilon$
0	$-j0.3566$	$+0.0460$	$+0.9673$	$-j1.9781$	-0.8660
1	—	$-j0.2970$	$-j2.4015$	$+1.1540$	—
2	—	—	—	$-j2.0000$	—
3	—	—	—	—	—
4	—	—	—	—	—

Synthesis of the main body can now proceed conventionally as in Section 7.4:

1) Extract $C_3 = 6.7343$ and $B_3 = 2.7126$.
2) Turn the network. If there is a pole present on the output side, a phase length θ_{5L} must be extracted now to prepare the network for the removal of the resonant pair, but in this case, both poles are on the input side, therefore, $\theta_{5L} = 0$.
3) Extract transmission line 90° and $Y_0 = M_{45} = 1$ (output coupling inverter).
4) Extract $C_4 = 0.8328$ and $B_4 = 0.1290$.
5) Finally extract a parallel-connected inverter $M_{34} = 2.4283$.

It is convenient to introduce two unity admittance line lengths of −90° and +90°, which have zero net effect on the transfer function, between the phase length θ_{23} and $C_3 + jB_3$. The first length is absorbed into $\theta_{23}(\theta_{23} \rightarrow \theta_{23} - 90° = -79.4492°)$, and the second forms the input coupling inverter to the main body of the filter M_{23}.

The extracted pole design in Figure 10.3 can be transformed through the dual-network theorem into a parallel resonant pair, connected through an inverter in parallel with the mainline (Figure 10.4a), or a series-connected parallel resonant pair between the two unit inverters in the mainline (Figure 10.4b). These circuits are now in the right form for realization in rectangular waveguide H-plane and E-plane open T junctions, respectively. Note Figure 10.4a, where the equivalent circuit for the H-plane junction includes an inverter [2].

By using the dual-network theorem, the shunt resonant pairs in the circuits in Figure 10.5 can also be developed into an inverter plus a length of transmission line that is 180° or an integral number of half-wavelengths at the resonant frequency of the TZ s_0 and short-circuited at one end. Eventually, this length is realized as a resonant cavity with the inverter becoming its input coupling. In practice, the inverter can be equated to its equivalent circuit, a shunt inductor at the center of a short length of negative transmission line. This is usually realized with an inductive aperture in waveguide.

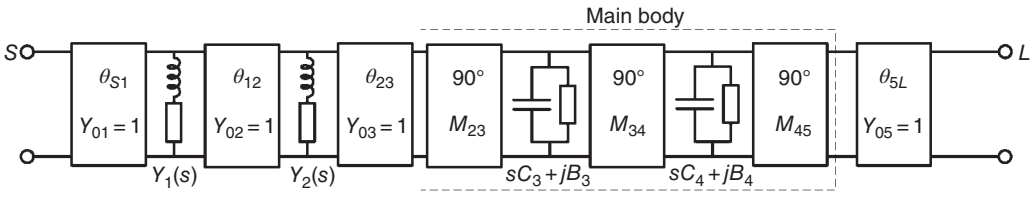


Figure 10.3 (4–2) extracted pole filter, synthesized network with both poles on input side.

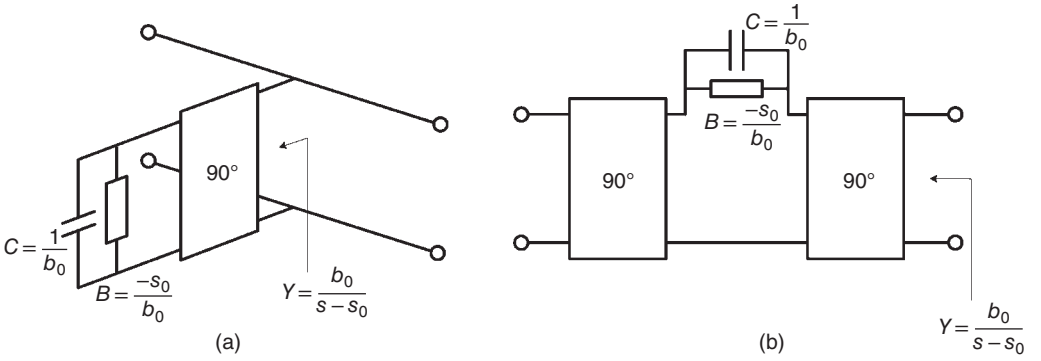


Figure 10.4 Possible realizations for the extracted pole: (a) parallel-connected (waveguide *H* plane); (b) series-connected (waveguide *E* plane).

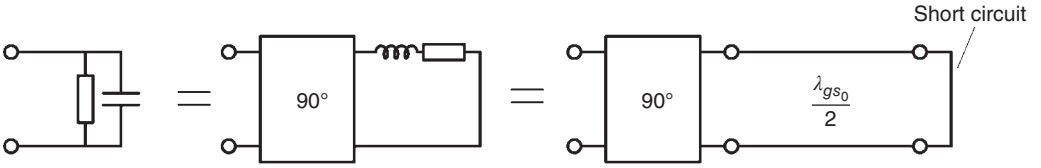


Figure 10.5 Development of the shunt resonant pair into an inverter and short-circuited half-wavelength transmission line.

Figure 10.6a depicts the extracted pole filter with TE_{011}-mode cylindrical resonant cavities with both poles on the input side of the main body. Although the structure incorporates two TZs, there are no negative couplings required (which would be difficult to implement with TE_{011} cavities). The couplings into and out of each mainline cavity are at right angles to each other to suppress the TM_{111} resonant mode, which is degenerate with the TE_{011} mode in cylindrical cavities. Figure 10.6b displays an alternative arrangement, where one of the extracted pole cavities is realized between the two mainline cavities.

10.1.3 Analysis of the Extracted Pole Filter Network

The extracted pole network can be analyzed by using the classical cascaded [*ABCD*] matrix method. A more useful method is based on the admittance matrix, which includes the coupling matrix [*M*] (see equation 8.7) and which may be analyzed in the same way as for a coupling matrix.

The overall admittance matrix [*Y*] is built up by the superimposition of the individual 2×2[*y*] submatrices. These submatrices themselves may make up more than one element and are formed by building the [*ABCD*] matrix of the element or "subcascade" and then converting it to [*y*] parameters. Figure 10.7 presents four such submatrices. The first is for a length of transmission line of phase θ_i and characteristic admittance Y_{0i}, followed by an extracted pole of admittance $Y_i(s)$; the second is for a length of transmission line in cascade with an inverter of characteristic admittance $Y_0 = M_{jk}$, representing the phase length between the last extracted pole and the input coupling inverter of the main body of the filter (M_{S1}). The third is for a shunt capacitor and FIR in cascade with a mainline coupling inverter, and the fourth is the same as the third but with an additional phase length. Pairing the components in this way reduces the number of nodes, reducing the size of the overall admittance matrix that eventually be inverted.

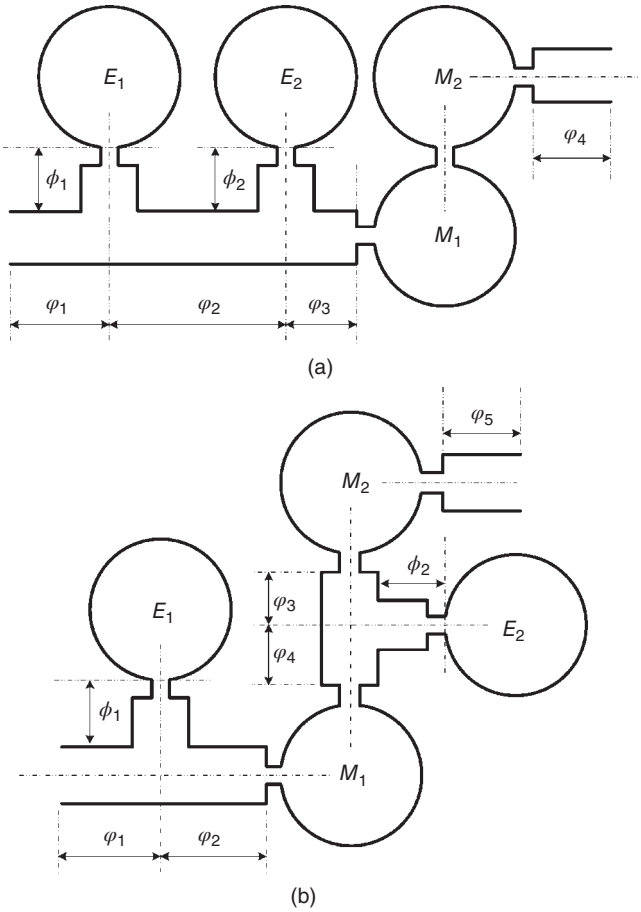


Figure 10.6 (4–2) extracted pole filter in cylindrical TE_{011}-mode resonator cavities: (a) both poles cavities on the input side; (b) one pole cavity synthesized within the main body.

Of course, the $[y]^{(i)}$ matrices can be reversed (for working at the output side of the filter) by simply exchanging the y_{11} and y_{22} elements in the submatrix.

The procedure for superimposing the individual 2×2 $[y]^{(i)}$ submatrices for a (4–2) filter with two extracted poles on the input side is shown in Figure 10.8, where the submatrix $[y]^{(1)}$ represents the input phase length + extracted pole 1, $[y]^{(2)}$ represents the interpole phase length + the second extracted pole, $[y]^{(3)}$ is the last phase length + input coupling inverter of the main body of the filter, and $[y]^{(4)}$ and $[y]^{(5)}$ represent the main body of the filter (which can be represented as an $N \times N$ cross-coupled matrix). The input phase length θ_{S1} and the final phase length θ_{5L} at the unity impedance source and load terminations define the input and output phase planes; they have no effect on the amplitudes of the transfer and reflection performance. It is in fact better to physically put the two extracted poles on either side of the main body of the filter if it is possible. This tends to result in more practical coupling values.

Adding the termination impedances ($1\,\Omega$) to the y_{11} and y_{66} positions of the matrix, inverting it, and applying the transfer and reflection formulas [(8.22) and (8.25)] yields the transfer and reflection performance of this network in the same way as for a regular coupling matrix.

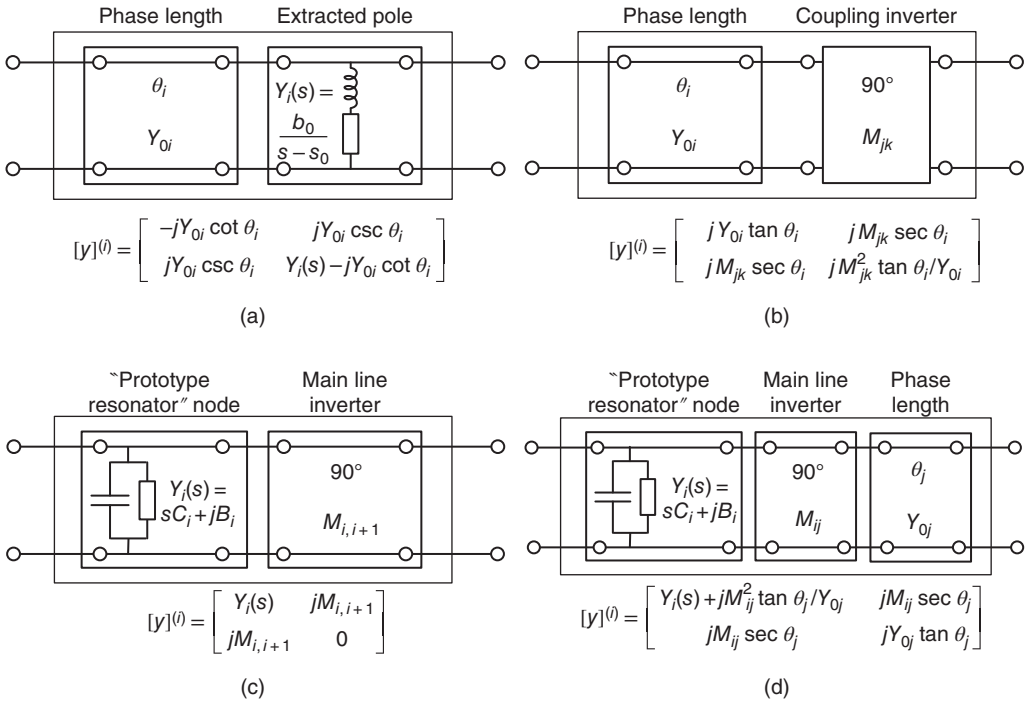


Figure 10.7 Admittance submatrices: (a) phase length + extracted pole; (b) phase length + coupling inverter; (c) prototype resonator node + inverter; (d) prototype resonator node + inverter + phase length.

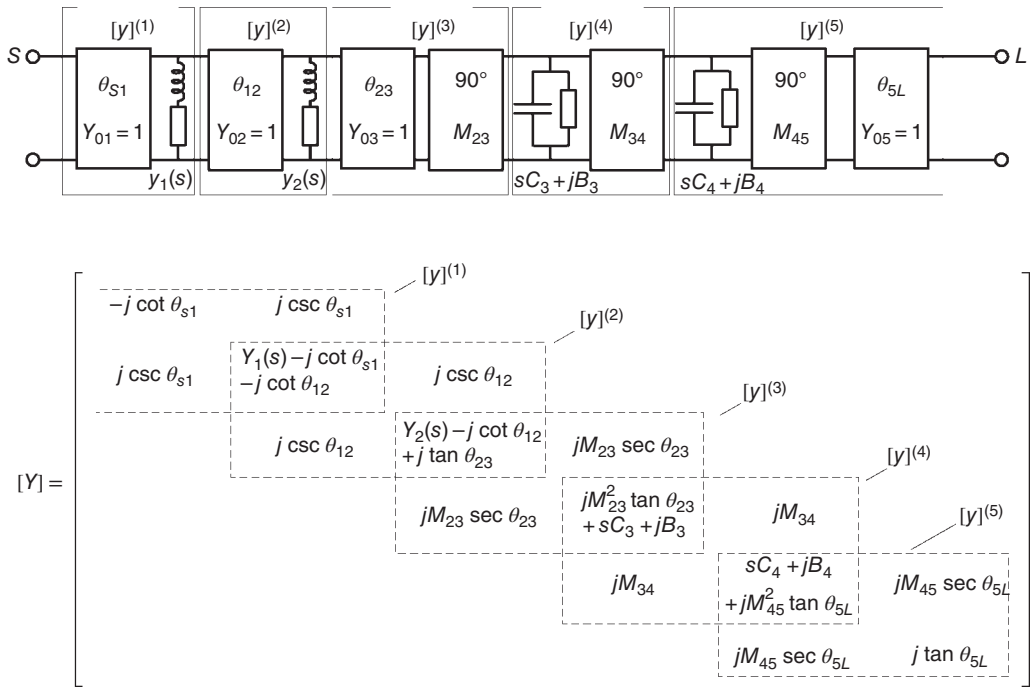


Figure 10.8 Superimposition of the admittance submatrices to form an overall admittance matrix.

10.1.4 Direct-Coupled Extracted Pole Filters

An alternative to the interpole phase lengths, and the phase lengths that couple between the last extracted pole and the main body of the filter, may be created through the use of the circuit equivalence as shown in Figure 10.9. This equivalence relates a length of transmission line of frequency-invariant phase length θ_i and characteristic admittance Y_0 to an inverter of admittance $Y_0 \text{cosec}\theta_i$ surrounded by 2 FIRs each of value $-jY_0 \cot\theta_i$.

The phase length between an extracted pole and the first $C + jB$ of the main body of the filter may be directly replaced with the inverter and the two FIRs as shown in Figure 10.9. The inverter itself forms a convenient input coupling inverter to the first resonator of the main body of the filter (avoiding the need to add the −90° and +90° line lengths as was done in the extracted pole synthesis example earlier), while the right-hand FIR may be added to the admittance ($sC + jB$) of the first resonator and realized as an additional frequency offset of that resonator. The FIR on the left-hand side (LHS) may be realized as a small capacitive or inductive obstacle or screw at the input port of the filter, or a non-resonant node (NRN) if convenient.

The process for developing a direct-coupled extracted pole filter from a phase-coupled type is most easily demonstrated through an example. Figure 10.10 shows the network for extracted pole synthesis of the (4–2) asymmetric filter, however, this time, with the two pole cavities realized on either side of the main body. The order of synthesis was to extract:

1) Phase length θ_{S1}
2) Resonant pair 1 at node 1
3) Phase length θ_{12}
4) $C_2 + jB_2$ and turn network
5) Phase length θ_{4L}
6) Resonant pair 2 at node 4
7) Phase length θ_{34}
8) $C_3 + jB_3$, and finally
9) Parallel-coupled inverter M_{23}

The values obtained for these components are summarized in Table 10.1.

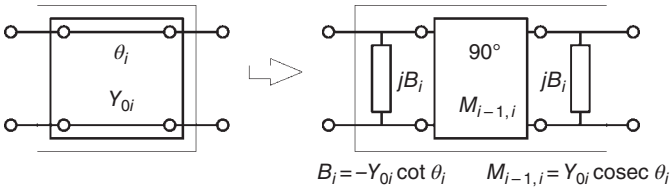


Figure 10.9 Equivalence of phase length with inverter surrounded by two FIRs.

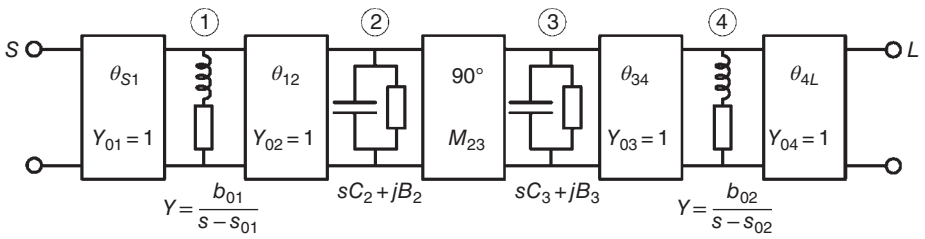


Figure 10.10 (4–2) asymmetric filter—phase-coupled extracted pole network.

Table 10.1 (4–2) extracted-pole filter—prototype network element values.

TZ no. i	TZ frequency s_{0i}	Extracted pole residue b_{0i}	$Y_i(s) = sC_i + jB_i$	Phase length $\theta_{i-1,i}$	Characteristic admittance Y_{0i}
1	$s_{01} = j1.3217$	$b_{01} = 0.5767$		$\theta_{s1} = +23.738°$	$Y_{01} = 1.0$
				$\theta_{12} = +66.262°$	$Y_{02} = 1.0$
			$C_2 = 1.3937$		
			$B_2 = 0.8101$		
				90.0°	$M_{23} = 1.7216$
			$C_3 = 2.8529$		
			$B_3 = 1.7180$		
				$\theta_{34} = +41.094°$	$Y_{03} = 1.0$
2	$s_{02} = j1.8082$	$b_{02} = 1.9680$		$\theta_{4L} = +48.906°$	$Y_{04} = 1.0$

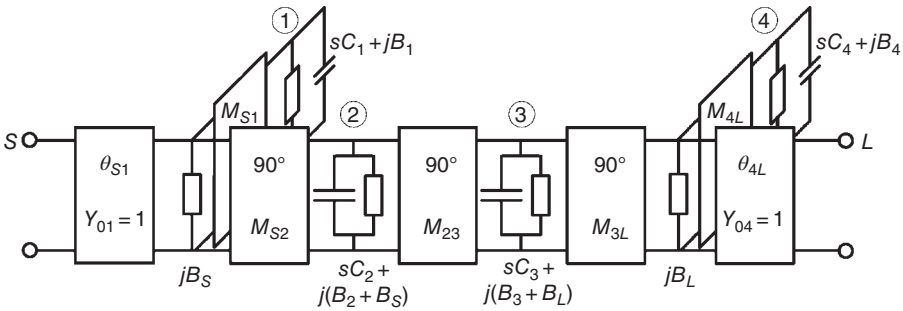


Figure 10.11 (4–2) asymmetric filter—direct-coupled extracted pole realization.

Figure 10.11 shows the configuration of the network after the phase lengths θ_{12} and θ_{34} have been replaced with the equivalent FIRs and inverters. Also the series resonant pairs Y_1 and Y_2 have transformed through the unit inverters M_{S1} and M_{4L} to form parallel resonators (see Figure 10.4). The values of the new components are

$$\begin{aligned}
M_{S2} &= \csc\theta_{12} = 1.0924 & M_{3L} &= \csc\theta_{34} = 1.5214 \\
B_S &= -\cot\theta_{12} = -0.4398 & B_L &= -\cot\theta_{34} = -1.1466 \\
C_1 &= \frac{1}{b_{01}} = 1.7340 & C_4 &= \frac{1}{b_{02}} = 0.5081 \\
B_1 &= -\frac{s_{01}}{b_{01}} = -2.2918 & B_4 &= -\frac{s_{02}}{b_{02}} = -0.9188 \\
B_2 &\to B_2 + B_S = 0.3703 & B_3 &\to B_3 + B_L = 0.5714
\end{aligned} \tag{10.8a}$$

Now scaling at the nodes to set all the C_i to unity,

$$\begin{aligned}
M_{S1} &= \frac{1}{\sqrt{C_1}} = 0.7594 & M_{4L} &= \frac{1}{\sqrt{C_4}} = 1.4029 \\
B_1 &\to \frac{B_1}{C_1} = -s_{01} = -1.3217 \equiv M_{11} & B_4 &\to \frac{B_4}{C_4} = -s_{02} = -1.8082 \equiv M_{44}
\end{aligned}$$

$$M_{S2} \rightarrow \frac{M_{S2}}{\sqrt{C_2}} = 0.9254 \qquad M_{3L} \rightarrow \frac{M_{3L}}{\sqrt{C_3}} = 0.9007$$

$$B_2 \rightarrow \frac{B_2}{C_2} = 0.2657 \equiv M_{22} \qquad B_3 \rightarrow \frac{B_3}{C_3} = 0.2003 \equiv M_{33}$$

$$M_{23} \rightarrow \frac{M_{23}}{\sqrt{C_2 C_3}} = 0.8634$$

$$M_{SS} \equiv B_S = -0.4398 \qquad M_{LL} \equiv B_L = -1.1466$$

$$C_1, C_2, C_3, C_4 \rightarrow 1 \tag{10.8b}$$

With all the C_i set to unity, the coupling matrix for the direct-coupled extracted pole filter may be constructed. Each FIR may be directly realized as a small obstacle as in Figure 10.11, for example, a screw near the input/output coupling devices, or as an NRN, in effect a detuned resonator, coupling the pole cavity and main-body cavity simultaneously [3]. In the coupling matrix and the corresponding coupling and routing diagram shown in Figures 10.12a and b, respectively, unity value inverters have been added to provide the input and output couplings to non-resonant cavities at the source and load ends of the network, respectively. Note that the values of the M_{11} and M_{NN} elements are the negatives of the prototype TZ positions (s_{01} and s_{02}, respectively). Also the values of B_S and B_L appear in the $M_{N1,N1}$ and $M_{N2,N2}$ positions, representing the parallel-connected FIRs at the input and output terminals. Figures 10.12c and 10.12d show two possible alternative waveguide configurations for the direct-coupled extracted pole network.

10.2 Synthesis of Bandstop Filters Using the Extracted Pole Technique

Bandstop filters are necessary when a specific narrow band of frequencies needs to be attenuated, for example, the second harmonic frequency at the output of a high-power saturated amplifier. Although the same job can be accomplished by a bandpass filter, it is often difficult to optimize the rejection performance over the stopband and at the same time maintain a low insertion loss and good return loss performance over the main signal band. However, the bandstop filter needs only to be optimized over its stopband and the bandstop cavities designed to be spurious-free over the passband.

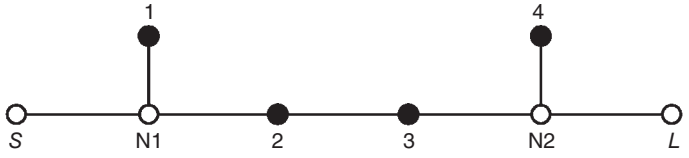
Bandstop filters are composed of a series of bandstop resonant cavities, placed at intervals along the main transmission-line medium carrying the signal, but not in the signal path, such that the power handling problems are mitigated. Figure 10.13 shows an example of a fourth-degree bandstop filter realized in a rectangular waveguide structure [4, 5].

Because the bandstop cavities of bandstop filters have the same equivalent circuit as the bandstop resonant cavities that realize the TZs of extracted pole bandpass filters, the same technique, used to synthesize the extracted poles, may be used for the bandstop filters. This approach was first noted [6] and then explained in greater detail in [7]. The difference now is that the $S_{21}(s)$ and $S_{11}(s)$ characteristics are exchanged, so that what used to be return loss becomes the transfer characteristic and vice versa. If the characteristic is a Chebyshev, the prescribed return loss level now becomes the equiripple stopband attenuation, whereas the former rejection characteristics become the out-of-stopband return loss performance.

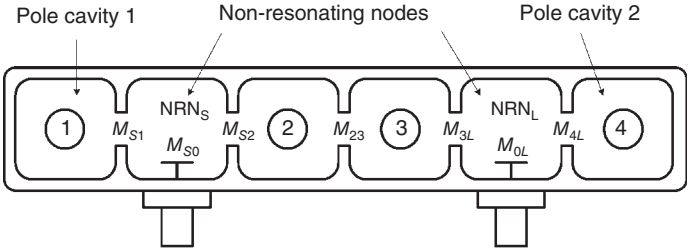
Thus, the numerator $F(s)/\varepsilon_R$ of the reflection function $S_{11}(s)$ becomes the numerator of the transfer function $S_{21}(s)$, and $P(s)/\varepsilon$ becomes the new numerator of $S_{11}(s)$. Therefore, for

$$[M] = \begin{array}{c} \\ S \\ 1 \\ 2 \\ 3 \\ 4 \\ L \end{array} \begin{array}{c} \begin{array}{cccccc} S & 1 & 2 & 3 & 4 & L \end{array} \\ \begin{bmatrix} -0.4398 & 0.7594 & 0.9254 & 0 & 0 & 0 \\ 0.7594 & -1.3217 & 0.0 & 0 & 0 & 0 \\ 0.9254 & 0.0 & 0.2657 & 0.8634 & 0 & 0 \\ 0 & 0 & 0.8634 & 0.2003 & 0.0 & 0.9007 \\ 0 & 0 & 0 & 0.0 & -1.8082 & 1.4029 \\ 0 & 0 & 0 & 0.9007 & 1.4029 & -1.1466 \end{bmatrix} \end{array}$$

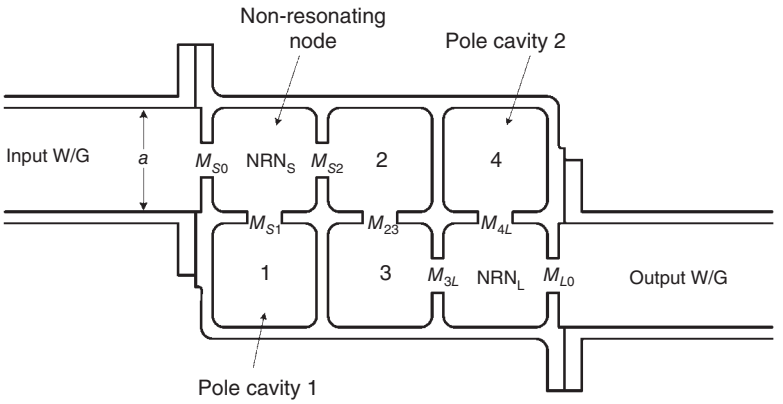
(a)



(b)



(c)



(d)

Figure 10.12 (4–2) asymmetric filter, the direct-coupled extracted pole realization: (a) coupling matrix. Nodes N1 and N2 (as well as the *S* and *L* nodes) are NRNs; (b) coupling and routing diagram; (c) realization in waveguide cavities; (d) alternative realization with waveguide resonant cavities.

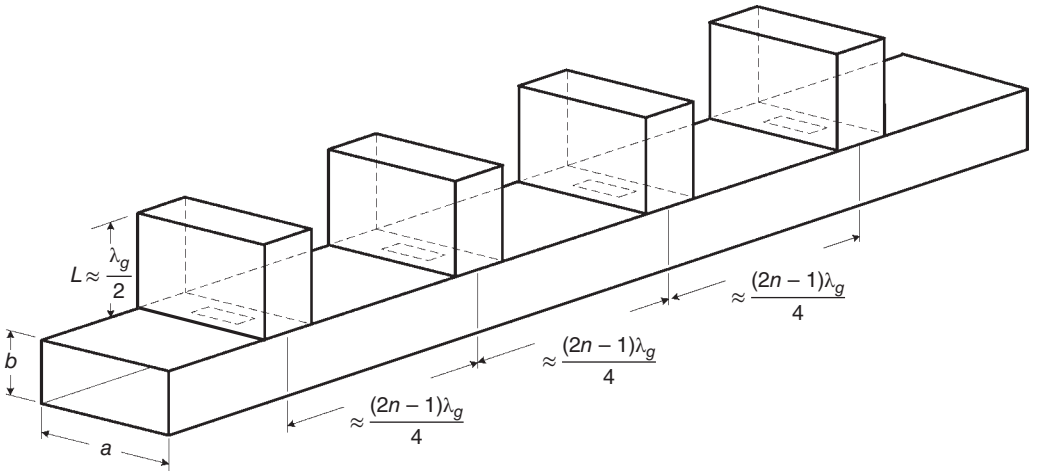


Figure 10.13 Fourth-degree bandstop filter realized in a rectangular waveguide.

an Nth-degree bandstop characteristic, there are N TZs to extract at the former reflection zero positions, in any order. Having extracted all the poles and their associated phase lengths, we need to extract a parallel-connected inverter. This can be located anywhere in the cascade but is equal to unity in all cases except for the canonical asymmetric and canonical even-degree symmetric filtering functions. This nonunity inverter represents a small reactive frequency-independent element, giving the finite return loss value at an infinite frequency that is inherent in the reflection polynomials for these cases.

The (4–4) asymmetric canonical prototype that is used for the $N + 2$ synthesis example in Section 8.4 is also adopted to demonstrate the bandstop synthesis process:

1) Exchange the $F(s)/\varepsilon_R$ and $P(s)/\varepsilon$ polynomials of the (4–4) lowpass prototype transfer and reflection functions $S_{21}(s)$ and $S_{11}(s)$. The j that multiplies $P(s)$ because $N - n_{fz}$ is even, and j that multiplies $P(s)$ to allow cross-couplings to be realized as inverters, should now multiply $F(s)/\varepsilon_R$, because this is now the new numerator of $S_{21}(s)$.
2) Extract the first phase length θ_{S1} at $s_{01} = -j0.9384$, the lowest of the $F(s)$ zeros, followed by the extraction of the resonant pair Y_1 at s_{01}.
3) Similarly, extract θ_{12} and Y_2 at $s_{02} = -j0.4228$.
4) Extract the phase length $\theta_{23}^{(1)}$ at $s_0 = j\infty$.
5) Turn the network and extract θ_{4L} and Y_4 at $s_{04} = j0.9405$ [the highest-frequency zero of $F(s)$].
6) Similarly extract θ_{34} and Y_3 at $s_{03} = j0.4234$.
7) Extract the phase length $\theta_{23}^{(3)}$ at $s_0 = j\infty$. Also, it is usually necessary to extract an extra 90° to attain the polynomials in the right form for the extraction of the final parallel-coupled inverter.
8) Extract the parallel-coupled inverter M_{23}. M_{23} is unity for all cases except when the original lowpass prototype is a symmetric even-degree canonical or any asymmetric canonical. In these cases, M_{23} has a value $\varepsilon_R \pm \sqrt{\varepsilon_R^2 - 1}$ and is approximately realized with a small reactive element, ideally an FIR, to provide the finite return loss that these prototypes have at infinite frequency.

The form of the synthesized (4–4) bandstop network is shown in Figure 10.14, and the corresponding element values are given in Table 10.2. Half-wavelengths (180°) can be added to the interresonator phase lengths, if these turn out to be negative or small for a more practical

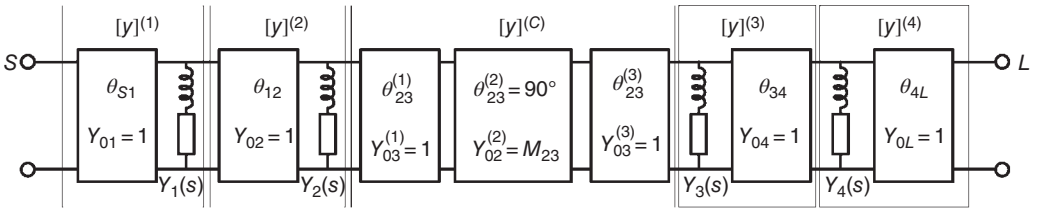


Figure 10.14 (4–4) bandstop filter network.

Table 10.2 Element values of the (4–4) asymmetric canonical bandstop filter.

TZ number i	TZ frequency s_{0i}	Phase length $\theta_{i-1,i}$	Extracted pole residue b_{0i}	Characteristic admittance Y_{0i}
1	$s_{01} = -j0.9384$	$\theta_{S1} = +34.859°$	$b_{01} = 0.6243$	$Y_{01} = 1.0$
2	$s_{02} = -j0.4228$	$\theta_{12} = +75.104°$	$b_{02} = 1.6994$	$Y_{02} = 1.0$
		$\theta_{23}^{(1)} = -19.962°$	—	$Y_{03}^{(1)} = 1.0$
		$\theta_{23}^{(2)} = +90.0°$	—	$Y_{03}^{(2)} = M_{23} = 1.0307$
3	$S_{03} = +j0.4234$	$\theta_{23}^{(3)} = 18.772°$	$b_{03} = 1.6024$	$Y_{03}^{(3)} = 1.0$
4	$s_{04} = +j0.9405$	$\theta_{34} = -72.334°$	$b_{04} = 0.6328$	$Y_{04} = 1.0$
		$\theta_{4L} = -36.438°$	—	$Y_{0L} = 1.0$
		$\sum\theta = k \cdot 90°$		

physical layout for the bandstop filter; however, the lengths should not be too long, because, ideally, they should be dispersionless (phase length frequency invariant), which in practice they are not.

For this example, the TZs are taken in ascending order, but there is no restriction on this, and they can be taken in any order. In addition, there is no restriction on where the final inverter is realized. It can be placed at the end of the network. In this case, the four poles are extracted in sequence, then a phase length at infinity, and finally the parallel inverter.

10.2.1 Direct-Coupled Bandstop Filters

The network synthesis methods presented in Chapter 7 can be used once the $S_{21}(s)$ and $S_{11}(s)$ functions are exchanged to create a bandpass-like filter configuration with bandstop filter characteristics. The resonant cavities are direct-coupled; therefore, the wideband performance is potentially better, and because the cavities are tuned to frequencies within the *stopband*, the main signal power routes mainly through the direct input/output coupling, bypassing the resonators and giving a minimal insertion loss and relatively high power handling.

As with the extracted pole types, to generate a bandstop characteristic from the regular lowpass prototype polynomials, it is necessary only to exchange the reflection and transfer functions (including the constants):

$$S_{11}(s) = \frac{P(s)/\varepsilon}{E(s)}, \quad S_{21}(s) = \frac{F(s)/\varepsilon_R}{E(s)} \tag{10.9}$$

Since $S_{21}(s)$ and $S_{11}(s)$ share a common denominator polynomial $E(s)$, the unitary conditions for a passive lossless network are preserved. If the characteristics are Chebyshev, then the original prescribed equiripple *return loss* characteristic becomes the transfer response, with a minimum reject level equal to the original prescribed return loss level. Because the degree of the new numerator polynomial for $S_{21}(s)\,(= F(s)/\varepsilon_R)$ is now the same as its denominator $E(s)$, the network that is synthesized is fully canonical. The new numerator of $S_{11}(s)$ is the original transfer function numerator polynomial $P(s)/\varepsilon$ and can have any number n_{fz} of prescribed TZs, provided $n_{fz} \leq N$, the degree of the characteristic. If $n_{fz} < N$, then the constant $\varepsilon_R = 1$.

The synthesis of the coupling matrix for the bandstop network is very similar to the synthesis of fully canonical lowpass prototypes for bandpass filters, as discussed in Section 8.4. First, we exchange the $P(s)/\varepsilon$ (degree n_{fz}) and $F(s)/\varepsilon_R$ (degree N) polynomials and then derive the polynomials for the rational short-circuit admittance parameters $y_{21}(s)$ and $y_{22}(s)$ of the network. For an even-degree doubly terminated network with source and load terminations of 1 Ω, we obtain

$$y_{21}(s) = \frac{y_{21n}(s)}{y_d(s)} = \frac{(F(s)/\varepsilon_R)}{m_1(s)}$$
$$y_{22}(s) = \frac{y_{22n}(s)}{y_d(s)} = \frac{n_1(s)}{m_1(s)} \tag{10.10a}$$

For an odd-degree network,

$$y_{21}(s) = \frac{y_{21n}(s)}{y_d(s)} = \frac{(F(s)/\varepsilon_R)}{n_1(s)}$$
$$y_{22}(s) = \frac{y_{22n}(s)}{y_d(s)} = \frac{m_1(s)}{n_1(s)} \tag{10.10b}$$

where

$$m_1(s) = \mathrm{Re}(e_0 + p_0) + j\mathrm{Im}(e_1 + p_1)s + \mathrm{Re}(e_2 + p_2)s^2 + \cdots$$
$$n_1(s) = j\mathrm{Im}(e_0 + p_0) + \mathrm{Re}(e_1 + p_1)s + j\mathrm{Im}(e_2 + p_2)s^2 + \cdots$$

and e_i and p_i, $i = 0,1,2,3,\ldots,N$, are the complex coefficients of $E(s)$ and $P(s)/\varepsilon$, respectively. If n_{fz} is less than N (which is the usual case), then the highest-degree coefficient of $P(s)/\varepsilon\ (= p_N)$ is zero.

Having built up the numerator polynomials $y_{21n}(s)$, $y_{22n}(s)$ and the denominator polynomial $y_d(s)$, the coupling matrix synthesis proceeds in exactly the same way as that in Section 8.4. Because the numerator of $S_{21}(s)$ is now the same degree as its denominator, the characteristic is now fully canonical and the coupling matrix representing it needs to incorporate a direct source–load coupling M_{SL}. M_{SL} is calculated as follows [see equation (8.43)]:

$$jM_{SL} = \left.\frac{y_{21}(s)}{y_d(s)}\right|_{s=j\infty} = \left.\frac{jF(s)/\varepsilon_R}{y_d(s)}\right|_{s=j\infty} \tag{10.11}$$

If the original bandpass characteristic is noncanonical, that is, the degree n_{fz} of $P(s)$ is less than N, and therefore $p_N = 0$, it can be seen from equation (10.10) that the leading coefficient of $y_d(s) = 1$. For a Chebyshev characteristic, the leading coefficient of $F(s)$ always equals unity, as does ε_R for noncanonical cases, and so from equation (10.11), it is evident that $M_{SL} = 1$. In other words, the direct source–load coupling inverter is the same characteristic impedance as the interfacing transmission lines from the source and to the load, and it can be simply constructed

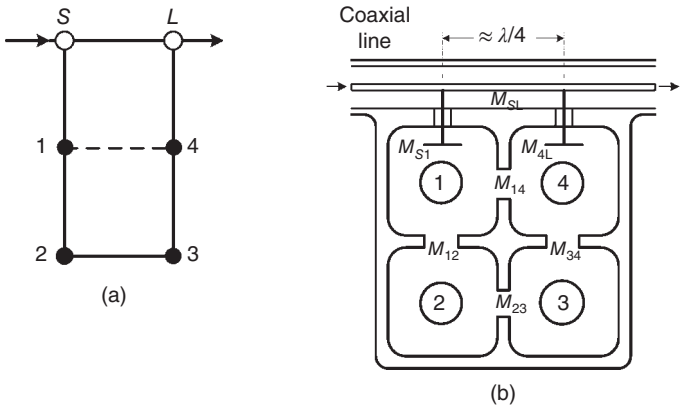


Figure 10.15 Direct-coupled (4–2) bandstop filter: (a) coupling–routing diagram; (b) possible realization with coaxial cavities.

from a 90° section of that line. For the fully canonical characteristics, M_{SL} is slightly less than unity and provides the finite return loss at infinite frequency that the fully canonical prototype requires.

The rest of the synthesis of the coupling matrix proceeds as for a folded lowpass prototype network for a bandpass filter. Diagonal cross-couplings appear if the characteristic is asymmetric. Figure 10.15 gives an example of a symmetric fourth-degree bandstop filter in folded configuration, showing that the main signal path is through the direct input/output coupling M_{SL}.

In general it is not desirable to include complex couplings in direct-coupled bandstop filters in folded form. It is best to restrict the applications to symmetric prototypes. Let us consider an example of a symmetric fourth-degree filter with 22 dB return loss (which becomes the stopband reject level) and two TZs (reflection zeros now) at $\pm j2.0107$ to give an out-of-band (return loss) lobe level of 30 dB. After synthesizing the folded ladder network and scaling all the C_i to unity, the following coupling matrix is obtained:

$$\boldsymbol{M} = \begin{array}{c} \\ S \\ 1 \\ 2 \\ 3 \\ 4 \\ L \end{array} \begin{array}{c} \begin{array}{cccccc} S & 1 & 2 & 3 & 4 & L \end{array} \\ \begin{bmatrix} 0.0 & 1.5109 & 0 & 0 & 0 & 1.0000 \\ 1.5109 & 0.0 & 0.9118 & 0 & 1.3363 & 0 \\ 0 & 0.9118 & 0.0 & -0.7985 & 0 & 0 \\ 0 & 0 & -0.7985 & 0.0 & 0.9118 & 0 \\ 0 & 1.3363 & 0 & 0.9118 & 0.0 & 1.5109 \\ 1.0000 & 0 & 0 & 0 & 1.5109 & 0.0 \end{bmatrix} \end{array} \tag{10.12a}$$

Note that the direct input/output coupling M_{SL} is unity for this noncanonical case, meaning that this coupling inverter is the same characteristic admittance as the input and output lines. It can be formed from a $90°(\approx \lambda/4)$ part of that line carrying the main signal power, between the input and output tap points to the main body of the filter. Figure 10.15b shows a possible realization with coaxial cavities; however, coupling M_{23} is negative in equation (10.12a) and needs to be realized as a probe.

It is interesting to note that with bandstop transfer and reflection characteristics, when the number of reflection zeros (formerly *transmission* zeros) is less than the degree of filter N (i.e., $n_{fz} < N$), a second solution is possible by working with the dual of the network. The dual network is obtained simply by multiplying the coefficients of the reflection numerator polynomials $F(s)$ and $P(s)$ by -1 and is equivalent to placing unit inverters at the input and output of the network.

If this is done for the (4–2) symmetric example and the bandstop matrix is resynthesized, the coupling matrix (10.12b) is obtained with all-positive couplings. In general, this uniformity of coupling sign does not occur:

$$\boldsymbol{M} = \begin{array}{c} \\ S \\ 1 \\ 2 \\ 3 \\ 4 \\ L \end{array} \begin{array}{c} \begin{array}{cccccc} S & 1 & 2 & 3 & 4 & L \end{array} \\ \begin{bmatrix} 0.0 & 1.5109 & 0 & 0 & 0 & 1.0000 \\ 1.5109 & 0.0 & 0.9118 & 0 & 0.9465 & 0 \\ 0 & 0.9118 & 0.0 & 0.7985 & 0 & 0 \\ 0 & 0 & 0.7985 & 0.0 & 0.9118 & 0 \\ 0 & 0.9465 & 0 & 0.9118 & 0.0 & 1.5109 \\ 1.0000 & 0 & 0 & 0 & 1.5109 & 0.0 \end{bmatrix} \end{array} \tag{10.12b}$$

10.2.1.1 Cul-de-Sac Forms for the Direct-Coupled Bandstop Matrix

If the number of reflection zeros of the bandstop characteristic is less than the degree of the network ($n_{fz} < N$), and the network is doubly terminated between equal source and load terminations, then a cul-de-sac form for the bandstop network, similar to that for bandpass filters (see Section 9.6), can be obtained by introducing two unity impedance 45° phase lengths at either end of the network. This is equivalent to multiplying the $F(s)$, $F_{22}(s)$, and $P(s)$ polynomials by j, which has no effect on the overall transfer and reflection responses of the network apart from the 90° phase changes.

Synthesizing the network by using either the circuit approach or the direct-coupling matrix approach yields networks as shown in Figure 10.16. These networks are characterized by the square-shaped core quartet of couplings, with the source and load terminals at adjacent corners at the input/output end; the other resonators are strung out in two chains from the other two corners in equal numbers if N is even and one more than the other if N is odd. There are no diagonal couplings even for asymmetric characteristics, and all the couplings are of the same sign. For these characteristics, where $n_{fz} < N$, the direct source–load coupling M_{SL} is always unity in value.

The bandstop filter may also be synthesized as the even- and odd-mode one-port networks of the lowpass prototype, attached to the branches of a coupler (see Figure 10.16c for a seventh degree example). If the coupler network is configured as a 3 dB hybrid coupler, instead of a rat-race coupler as for the bandpass filter (see Section 9.6.1), then the bandstop equivalent

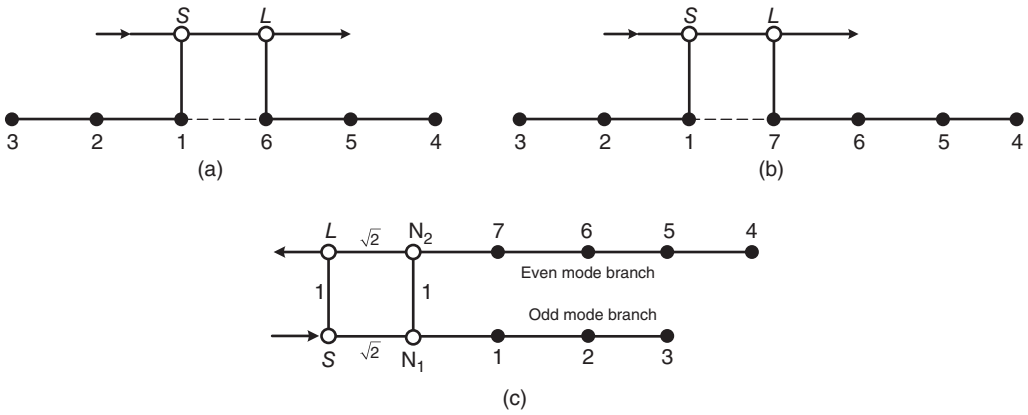


Figure 10.16 Cul-de-sac forms for direct-coupled bandstop filters: (a) sixth degree; (b) seventh degree. (c) seventh degree bandstop filter—hybrid-coupled even- and odd-mode networks.

response will be obtained, that is, the S_{21} and S_{11} responses exchange. The procedure is simply to first generate the $N+4$ coupling matrix with a series of cul-de-sac rotations as for the bandpass filter and then to replace the elements of the rat-race coupler with those of the 3 dB coupler, that is, $M_{S,L} = M_{N1,N2} = 1, M_{S,N1} = M_{N2,L} = \sqrt{2}$. There is no need to change the values of the even- and odd-mode networks, which are automatically generated by the cul-de-sac rotation sequence. Again this configuration is particularly suitable for realization in a planar technology, for example, microstrip.

Let us take the example of the (4–2) asymmetric characteristic that is used as the synthesis example in Chapters 7 and 8. Exchanging the $F(s)$ and $P(s)$ polynomials, multiplying them by j, and synthesizing the folded coupling matrix (Chapter 8) yields the coupling matrix (10.13). Figure 10.17 shows the corresponding coupling and routing diagram and a possible realization with rectangular waveguide resonators. The section of the waveguide transmission line between the input and output connections forms the direct input/output coupling M_{SL} and can be an odd number of quarter wavelengths. The less, the better:

$$\boldsymbol{M} = \begin{array}{c} \\ S \\ 1 \\ 2 \\ 3 \\ 4 \\ L \end{array} \begin{array}{c} \begin{array}{cccccc} S & 1 & 2 & 3 & 4 & L \end{array} \\ \begin{bmatrix} 0.0 & 1.5497 & 0 & 0 & 0 & 1.0000 \\ 1.5497 & 0.5155 & 1.2902 & 0 & 1.2008 & 0 \\ 0 & 1.2902 & -0.0503 & 0.0 & 0 & 0 \\ 0 & 0 & 0.0 & -1.0187 & 0.4222 & 0 \\ 0 & 1.2008 & 0 & 0.4222 & -0.2057 & 1.5497 \\ 1.0000 & 0 & 0 & 0 & 1.5497 & 0.0 \end{bmatrix} \end{array} \tag{10.13}$$

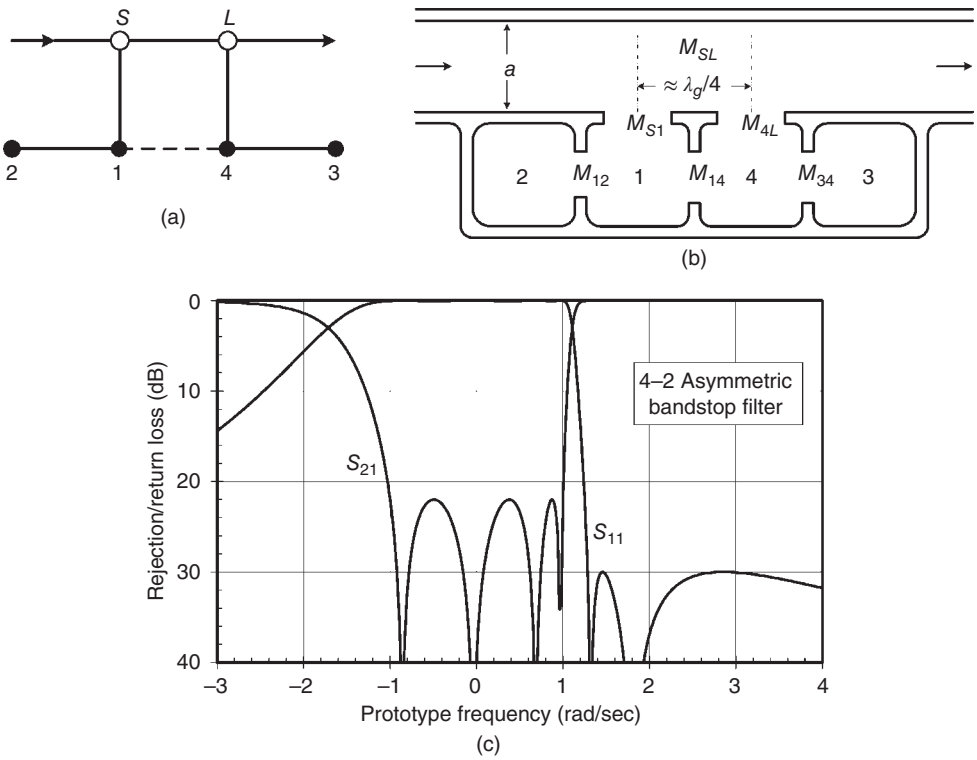


Figure 10.17 (4–2) Direct-coupled cul-de-sac bandstop filter: (a) coupling and routing diagram; (b) possible realization with waveguide cavities; (c) rejection and return loss performance.

The resonators of the direct-coupled bandstop filter can be realized with coaxial, dielectric, or planar resonators. A possible application is in high-power signal diplexing when ultralow loss is important but where out-of-band rejection is not a serious problem (or is provided by a low-loss wideband cover filter), and large levels of isolation are not required. For a diplexer application, each of the two bandstop filters is tuned to reject the frequency of the channel on the opposite arm of the combination point.

10.3 Trisections

Trisections are another section that, like the extracted pole section, are able to realize one TZ each. A trisection consists of three couplings between three nodes, the first and third of which can be source or load terminals or $C_i + jB_i$ *lowpass resonators*. The middle node is usually a lowpass resonator [2, 6, 8]. Figure 10.18 shows four possible configurations. Figure 10.18a is an internal trisection, and Figures 10.18b and c show the input and output trisections, respectively, where one node is the source or load termination. When the first and third nodes are the source and load terminations, respectively (Figure 10.18d), we have a canonical network of degree 1 with the direct source–load coupling M_{SL} providing the single TZ. Moreover, trisections can be cascaded with other trisections, either separately or conjoined (Figures 10.18e and f).

In general, if the TZ that the trisection is realizing is above the main passband, then the cross-coupling (e.g., M_{S2} in Figure 10.18b) is positive, and if the zero is below the passband, it is negative. The cross-coupling can be realized by a diagonal capacitive probe for a negative coupling and by an inductive loop for positive coupling.

Trisections have some properties that are very useful for the synthesis of advanced cross-coupled network topologies:

- A particular TZ may be associated with the cross-coupling in a particular trisection.
- The trisection may be represented in a coupling matrix.

The freedom to be able to synthesize the trisections in more or less unrestricted positions within the overall filter network, and then to apply rotations to the associated coupling matrix to obtain a final topology, opens up some configuration possibilities that cannot be easily

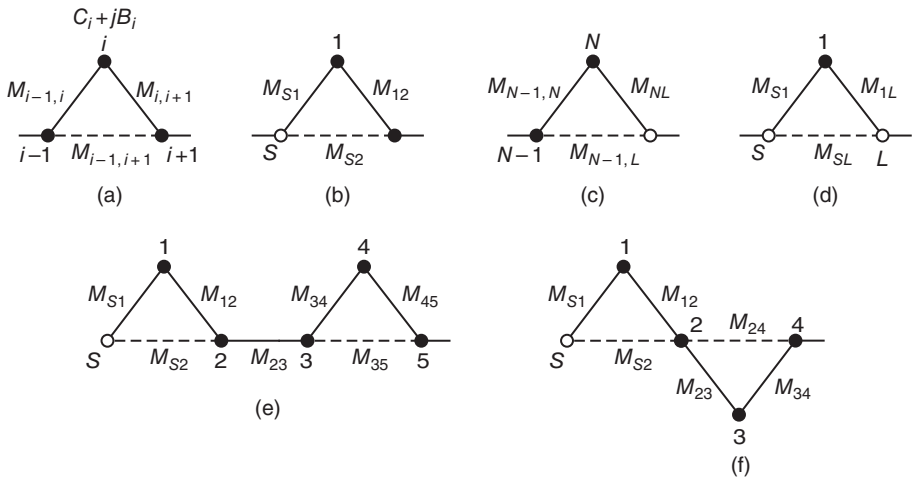


Figure 10.18 Coupling–routing diagram for trisections: (a) internal; (b) source connected; (c) load connected; (d) canonical; (e) nonconjoined cascaded; (f) conjoined cascaded.

obtained by using the folded or transversal matrix as the starting point. Two approaches for the synthesis of an individual trisection or a cascade of trisections are described: a hybrid method based on the extraction of circuit elements followed by conversion to a coupling matrix and an alternative method based purely on the coupling matrix. The methods used to transform the trisection-configured coupling matrix into cascaded sections (*n-tuplets* and *box* sections) are outlined.

10.3.1 Synthesis of the Trisection: Circuit Approach

The synthesis of the trisection follows a route similar to that for the synthesis of the extracted pole circuit, involving the extraction of only parallel-connected elements and lengths of transmission line [6]. For the extracted pole, a phase length is first extracted from the overall [*ABCD*] matrix that represents the filtering characteristic desired in order to prepare the network for the extraction of the parallel-connected series resonant pair to realize the TZ at $s = s_0$. In the case of a trisection, a shunt-connected FIR of admittance J_{13}, evaluated at $s = s_0$, is extracted first, followed by a unity admittance inverter to prepare the network for the extraction of the resonant pair by the same procedure as with the extracted pole. Then a further FIR of value $-1/J_{13}$ is extracted, another unity admittance inverter, and finally a parallel FIR of value J_{13}, the same value as the first FIR. The network is then transformed into the trisection network using the dual-network theorem and a circuit transform.

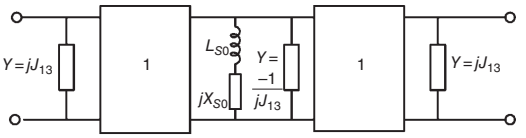
The procedure is illustrated in Table 10.3. Step 1 shows the initial network, consisting of two inverters, three parallel-connected FIRs, and the resonant pair, that is synthesized from the $A(s)$, $B(s)$, $C(s)$, $D(s)$, and $P(s)/\varepsilon$ polynomials. After these components have been extracted, all the remainder polynomials are one degree less than the originals, in the same way as with the extracted pole design. Steps 2 and 3 show the transformation of the network to the "π" network shown in step 3.

Now the well-known transform can be applied to the components in the series arm of the π network to form the new π network as shown in step 4. The transform is readily proved by equating the short-circuit input admittances y_{11} of the networks in the series arms of the π networks at $s = 0$ and $s = s_0$. At this stage, the π network formed by the two shunt admittances J_{13} at either side, and the admittance $-J_{13}$ that bridges them, are replaced by the equivalent admittance inverter of characteristic admittance $-J_{13}$ [4]. Lastly, the dual-network theorem is applied to the remaining series components in the series arm to form the shunt-connected pair $C_{S2} + B_{S2}$, the central lowpass resonator of the trisection.

In Table 10.3, the input and output ends of the trisection (nodes 1 and 3, respectively) are connected to nodes that are themselves lowpass resonators, but these nodes can be the source or load terminals. In this case, the filter is a dual-input type, with two couplings from the source terminal to internal resonators 1 and 2, and/or $N - 1$ and N to the load terminal. Of course, the trisections can be synthesized anywhere within larger-degree networks by simply indexing the subscripts of C_S, B_S, and J appropriately. If the central shunt-connected pair of the trisection is at node position k within the overall network, then k is called the *index* of the trisection, and the subscripts of the cross-coupling inverter become $-J_{k-1,k+1}$ ($k = 2$ for the trisection in Table 10.3). The value of this and the values of the trisection's central resonant pair $C_{Sk} + jB_{Sk}$ are entered into a coupling matrix along with the values of the other elements that have been extracted and scaled along with these other elements such that the values of all the capacitors are unity. Then, rotations may be applied to reconfigure the network.

Table 10.3 Trisection synthesis procedure.

1) Extract:
   a) Parallel frequency-invariant admittance jJ_{13}, to prepare network for extraction of pole at $s = s_0$
   b) Unit inverter
   c) Parallel-connected series inductor $L_{S0}(= 1/b_0)$+ frequency-invariant reactance $jX_{S0}(= -s_0/b_0)$ (extracted pole; see Section 10.1.1)
   d) Parallel frequency-invariant admittance $-1/jJ_{13}$
   e) Unit inverter
   f) Parallel frequency-invariant admittance jJ_{13}

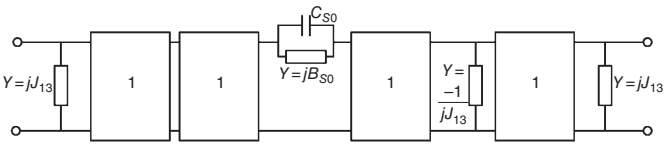


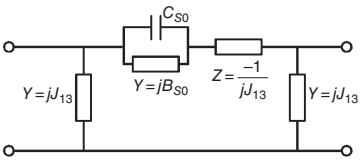
$$jJ_{13} = \left.\frac{D(s)}{B(s)}\right|_{s=s_0} \qquad L_{s0} = \frac{1}{b_0} \qquad jX_{s0} = \frac{-s_0}{b_0}$$

2) Dual-network theorem
   a) Convert parallel-connected $L_{S0} + jX_{S0}$ to series-connected parallel pair between two unit invertors:

   $B_{S0} = X_{S0}$
   $C_{S0} = L_{S0}$

   b) Convert parallel-connected admittance $-1/jJ_{13}$ between two unit inverters to series reactance $-1/jJ_{13}$





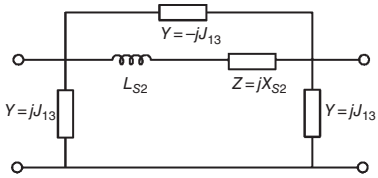
(Continued)

Table 10.3 (Continued)

3) Circuit transform:

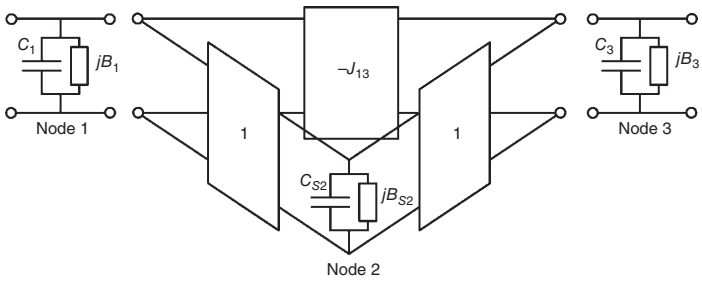
$$X_{S2} = \frac{J_{13} - B_{s0}}{-J_{13}^2} = \frac{J_{13} + \omega_0/b_0}{-J_{13}^2}$$

$$L_{S2} = \frac{-jB_{S0}}{s_0 J_{13}^2} = \frac{1}{b_0 J_{13}^2}$$



4) a) Inverter of characteristic admittance $-J_{13}$ formed by π network of J_{13} susceptances
   b) Dual-network theorem: Convert series-connected L_{S2} and jX_{S2} to parallel pair C_{S2} and jB_{S2} between two unit inverters:

   $C_{S2} = L_{S2}$
   $B_{S2} = X_{S2}$



For the elements of the trisection at index k within an Nth-degree network, the values of the scaled entries to the coupling matrix are

$$\begin{aligned} M_{k-1,k} &= \frac{1}{\sqrt{C_{k-1}C_{Sk}}} \\ M_{k,k+1} &= \frac{1}{\sqrt{C_{Sk}C_{k+1}}} \\ M_{k-1,k+1} &= \frac{-J_{k-1,k+1}}{\sqrt{C_{k-1}C_{k+1}}} \\ B_{Sk}(\equiv M_{k,k}) &\rightarrow \frac{B_{Sk}}{C_{Sk}} \\ C_{Sk} &\rightarrow 1 \end{aligned} \tag{10.14}$$

where C_{Sk} and B_{Sk} are the values of the central shunt elements of the trisection and C_{k-1} and C_{k+1} are the values of the shunt capacitors extracted immediately before and immediately after the trisection. If $k = 1$, $C_{k-1} = 1$ or if $k = N$, $C_{k+1} = 1$.

After the values of the trisection, scaled or unscaled, are obtained, the frequency of its TZ is checked by application of the following formulas:

$$\text{Unscaled values}: \omega_0 = -\frac{1}{C_{Sk}}\left[\frac{1}{J_{k-1,k+1}} + B_{Sk}\right] \tag{10.15a}$$

If the values are taken from the coupling matrix, then

$$\omega_0 = \frac{M_{k-1,k} \cdot M_{k,k+1}}{M_{k-1,k+1}} - M_{k,k} \tag{10.15b}$$

which is the solution to the trisection determinant evaluated at $s = j\omega_0$:

$$\det \begin{vmatrix} M_{k-1,k} & M_{k-1,k+1} \\ \omega_0 + M_{k,k} & M_{k,k+1} \end{vmatrix} = 0 \tag{10.15c}$$

In Figure 10.19, the (4–2) asymmetric prototype described in Section 10.1.2 is synthesized in a form similar to that in Figure 10.18e.

The $[ABCD]$ matrix polynomials that are synthesized for the (4–2) characteristic are provided in Section 7.4. The synthesis of the first trisection at index $k = 1$ with a TZ located at $s_0 = j1.8082$ proceeds as follows (see item 1 in Table 10.3):

1) Calculate the value of the parallel-connected FIR J_{S2} from

$$jJ_{S2} = \left.\frac{D(s)}{B(s)}\right|_{s=s_0}, \quad J_{S2} = -0.8722$$

2) Extract the FIR J_{S2} and the unit inverter from the $A(s)$, $B(s)$, $C(s)$, and $D(s)$ polynomials to obtain the remainder polynomials $A'(s)$, $B'(s)$, $C'(s)$, and $D'(s)$.
Now the procedure is similar to that used to extract the shunt-connected resonant pair (extracted pole), described in Section 10.1.

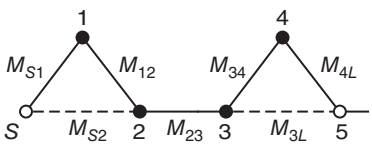


Figure 10.19 Trisection realization for the (4–2) asymmetric filter characteristic.

3) Divide $A'(s)$ and $B'(s)$ by the factor $(s - s_0)$ to obtain the polynomials A'' and B''.
4) Evaluate the value of the residue b_0 from

$$b_0 = \left.\frac{D'(s)}{B''(s)}\right|_{s=s_0}, \quad b_0 = 1.1177$$

Calculate B_{S1} and C_{S1} from

$$B_{S1} = -\frac{J_{S2} + \omega_0/b_0}{J_{S2}^2}, \quad C_{S1} = \frac{1}{b_0 J_{S2}^2}$$

$B_{S1} = -0.9801$ and $C_{S1} = 1.1761$.
5) Compute $C'(s) - b_0 A''(s)$ and $D'(s) - b_0 B''(s)$.
6) Divide these two polynomials by the factor $(s - s_0)$ to obtain $C''(s)$ and $D''(s)$.
7) Divide the $P(s)$ polynomial by the factor $(s - s_0)$ to obtain $P''(s)$. At this stage the coefficients of the polynomials have the values which are listed next:

$s^i, i =$	$A''(s)$	$B''(s)$	$C''(s)$	$D''(s)$	$P''(s)/\epsilon$
0	+0.9148	$-j1.1936$	$-j1.7080$	−0.7964	$+j1.1446$
1	$-j1.0108$	+2.7263	−0.1938	$-j3.5503$	−0.8660
2	+2.0945	$-j0.5716$	$-j2.4015$	+0.1484	—
3	—	+1.7443	—	$+j2.0000$	—
4	—	—	—	—	—

8) Extract a parallel-connected FIR of value $-1/J_{S2}$ from $[A''B''C''D'']$ matrix, followed by a unit inverter and finally an FIR of value J_{S2}.

This completes the extraction of the first trisection, leaving the coefficients of the $[ABCD]$ polynomials with the following values:

$s^i, i =$	$A(s)$	$B(s)$	$C(s)$	$D(s)$	$P(s)/\epsilon$
0	−0.6591	$-j0.5721$	$-j1.4896$	−0.6946	$+j1.1446$
1	$-j0.9652$	−0.4244	−0.1690	$-j3.0964$	−0.8660
2	—	$-j0.8038$	$-j2.0945$	+0.1294	—
3	—	—	—	$-j1.7443$	—
4	—	—	—	—	—

It is noted that at the end of this cycle, the $D(s)$ polynomial has the highest degree, instead of $B(s)$, which is typical. This is the consequence of not extracting an inverter at the beginning of the procedure, but has no effect on the rest of the synthesis. Since the trisection is removed, the shunt-connected pair $C_2 + jB_2$ may be extracted, where $C_2 = 2.1701$ and $B_2 = 1.3068$.

The second trisection at index $k = 4$ with the central shunt-connected pair $C_{S4} + jB_{S4}$, realizing the TZ $s_0 = j1.3217$, is now synthesized by turning the network [exchanging $A(s)$ and $D(s)$] and applying the same eight steps. The corresponding parameter values are $J_{3L} = -2.2739$, $b_0 = 0.0935$, $C_{S4} = 2.0694$, and $B_{S4} = -2.2954$. The following shunt-connected pair $C_3 + jB_3$ is

Table 10.4 Element values of the (4–2) asymmetric filter with two trisection.

Frequency-variant capacitors at nodes	Frequency-invariant reactances at node	Sequential coupling inverters	Nonsequential coupling inverters
$C_{S1} = 1.1761$	$B_{S1} = -0.9801$	$J_{S1} = 1.0$	$-J_{S2} = 0.8722$
$C_2 = 2.1701$	$B_2 = 1.3068$	$J_{12} = 1.0$	$-J_{3L} = 2.2739$
$C_3 = 7.2065$	$B_3 = 4.1886$	$J_{23} = 3.4143$	—
$C_{S4} = 2.0694$	$B_{S4} = -2.2954$	$J_{34} = 1.0$	—
—	—	$J_{4L} = 1.0$	—

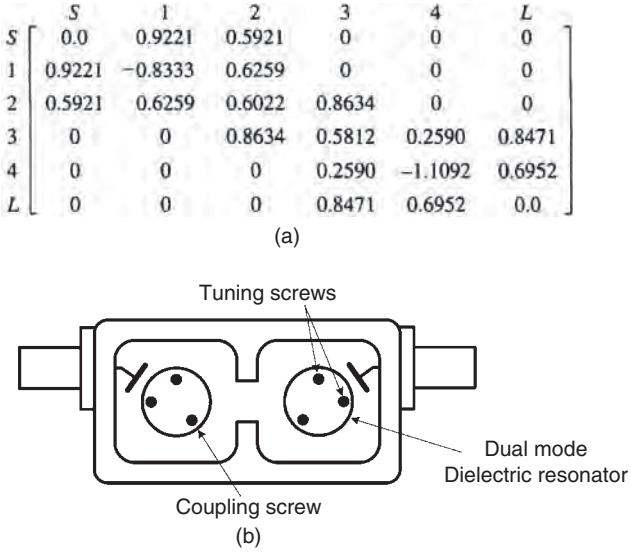


Figure 10.20 $N+2$ coupling matrix for the (4–2) asymmetric characteristic synthesized with trisections: (a) coupling matrix; (b) possible realization with two dual-mode dielectric resonator cavities.

extracted, where $C_3 = 7.2064$ and $B_3 = 4.1886$, and finally the parallel-connected inverter J_{23} is extracted as 3.4143 (see Table 10.4).

Scaling all the capacitors at the nodes to unity gives the $N + 2$ coupling matrix, corresponding to that in Figure 10.20a. Analyzing this matrix provides exactly the same prototype performance characteristics as those of the original folded configuration. Figure 10.20b shows a possible realization by employing two dual-mode dielectric resonator cavities. The angled input and output probes provide the dual-input/output couplings M_{S1} and M_{S2} and M_{4L} and M_{3L}, respectively.

10.3.2 Cascade Trisections: Coupling Matrix Approach

A method based purely on the coupling matrix for the generation of a cascade of trisections has been introduced by Tamiazzo and Macchiarella [9]. This method takes as its basis the *Bell's wheel* or *arrow* form for the canonical coupling matrix. This has the advantage that only the similarity transform operations are required, starting with any canonical form for the coupling matrix (e.g., full matrix, transverse matrix), which itself may be synthesized directly from the transfer and reflection polynomials of the filter function, discussed in Chapter 8. There is no need to first synthesize a circuit and then convert it to an equivalent coupling matrix. The pure

coupling matrix approach allows a seamless sequence of rotations to be applied to the starting coupling matrix: to generate the *arrow* configuration for the coupling matrix, cascading of trisections, and then to reconfigure to another form if it is required (e.g., cascade quartets).

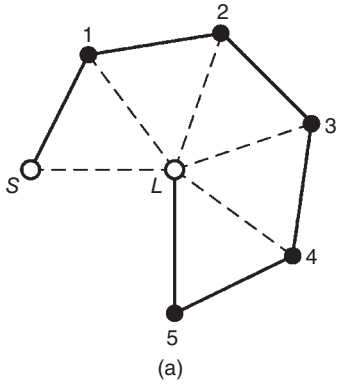
10.3.2.1 Synthesis of the Arrow Canonical Coupling Matrix

The folded cross-coupled circuit and its corresponding coupling matrix were introduced in Chapters 7 and 8 as one of the basic canonical forms of the coupling matrix, capable of realizing N TZs in an Nth-degree network. A second form was introduced by Bell [10] in 1982. The configuration later became known as the *wheel* or *arrow* form. As with the folded form, all the mainline couplings are present, and in addition, the source terminal and each resonator node are cross-coupled to the load terminal.

Figure 10.21a is a coupling–routing diagram for a fifth-degree fully canonical filter circuit, showing clearly why this configuration is referred to as the *wheel* canonical form with the mainline couplings forming the (partially incomplete) "rim" and the cross-couplings and input/output coupling forming the "spokes." Figure 10.21b displays the corresponding coupling matrix, where the cross-coupling elements are all in the last row and column and, along with the mainline and self-couplings on the main diagonals, give the matrix the appearance of an arrow pointing downward toward the lower right corner of the matrix. The network reversed, where all the cross-couplings are now in the first row and column of the matrix, giving the upward-pointing arrow form. If there are no TZs at all in the transfer function (all-pole function), the only entry in the last row and column of the arrow matrix is the output coupling M_{NL}. With one TZ, the entry $M_{N-1,L}$ appears in the last row/column, with two TZs, $M_{N-2,L}$ appears, and so on until finally there are N TZs (fully canonical) and the entry M_{SL} becomes nonzero (as in Figure 10.21b). Thus, as the number of TZs in the filter function increases from zero to N, the head of the arrow grows from the lower right corner of the matrix until the last row and column are completely filled.

10.3.2.2 Synthesis of the Arrow Coupling Matrix Form

The arrow matrix can be formed from the transversal or any other matrix form by a series of rotations very similar to those needed to form the folded coupling matrix (Section 8.3). The total number of rotations R in the sequence needed to generate an $N+2$ arrow matrix is given by $R = \sum_{r=1}^{N-1} r$, which equals 10 for the fifth-degree example in Figure 10.21. Table 10.3 gives



(a)

	S	1	2	3	4	5	L
S		X					X
1	X	X	X				X
2		X	X	X			X
3			X	X	X		X
4				X	X	X	X
5					X	X	X
L	X	X	X	X	X	X	

(b)

Figure 10.21 Fifth-degree wheel or arrow canonical circuit: (a) coupling–routing diagram (wheel); (b) $N+2$ coupling matrix (arrow).

Table 10.5 Fifth-degree example: similarity transform (rotation) sequence for reduction of the $N+2$ coupling matrix to the arrow form.

Transform number r	Pivot $[i, j]$	Element to Be annihilated		$\theta_r = -\tan^{-1}(M_{kl}/M_{mn})$			
				k	l	m	n
1	[1, 2]	M_{S2}	In row 1	S	2	S	1
2	[1, 3]	M_{S3}	—	S	3	S	1
3	[1, 4]	M_{S4}	—	S	4	S	1
4	[1, 5]	M_{S5}	—	S	5	S	1
5	[2, 3]	M_{13}	In row 2	1	3	1	2
6	[2, 4]	M_{14}	—	1	4	1	2
7	[2, 5]	M_{15}	—	1	5	1	2
8	[3, 4]	M_{24}	In row 3	2	4	2	3
9	[3, 5]	M_{25}	—	2	5	2	3
10	[4, 5]	M_{35}	In row 4	3	5	3	4

the sequence of pivots and angle formula for the fifth-degree case, reducing the initial $N+2$ coupling matrix (transversal, folded, or other) to the downward-pointing arrow form. As with the folded matrix, the final values and positions of the elements in the last row and column are automatically determined. No specific action to annihilate unneeded elements needs to be taken. The pattern of pivots and angle computations in the sequence is regular and is easily programmed (Table 10.5).

10.3.2.3 Creation and Repositioning of Trisections within the Arrow Coupling Matrix

The procedure followed to create a trisection, associated with the ith TZ $s_{0i} = j\omega_{0i}$ within the arrow-configured coupling matrix, involves applying a single rotation with pivot $[N-1, N]$ near the tip of the arrow at an angle calculated to satisfy the following determinant for the first trisection associated with the first TZ $s_{01} = j\omega_{01}$ embedded in the transformed matrix [see equation (10.15c)]:

$$\det \begin{vmatrix} M^{(1)}_{N-2,N-1} & M^{(1)}_{N-2,N} \\ \omega_{01} + M^{(1)}_{N-1,N-1} & M^{(1)}_{N-1,N} \end{vmatrix} = 0 \tag{10.16}$$

where $M^{(1)}_{ij}$ are elements in the transformed matrix $\mathbf{M}^{(1)}$. By following through the algebra associated with the application of the rotation at pivot $[N-1,N]$ to the original arrow matrix $\mathbf{M}^{(0)}$ [see equation set (8.35)], the angle θ_{01}, required to satisfy equation (10.16), is determined by

$$\theta_{01} = \tan^{-1}\left[\frac{M^{(0)}_{N-1,N}}{\omega_{01} + M^{(0)}_{N,N}}\right] \tag{10.17}$$

where $M^{(0)}_{N-1,N}$ and $M^{(0)}_{N,N}$ are taken from the original arrow matrix $\mathbf{M}^{(0)}$.

After application of this rotation at pivot $[N-1, N]$ and angle θ_{01}, coupling $M^{(1)}_{N-2,N}$ is created in the transformed matrix $\mathbf{M}^{(1)}$. A trisection is now embedded in the arrow structure at index position $k = N-1$. This trisection can be pulled up the diagonal to index position $k = N-2$

with the application of a rotation at pivot $[N-2, N-1]$ and angle θ_{12}, after which element $M^{(2)}_{N-2,N}$ is zero and $M^{(2)}_{N-3,N-1}$ is created in the new matrix $\mathbf{M}^{(2)}$ [see equation (8.35)]:

$$\theta_{12} = \tan^{-1}\frac{M^{(1)}_{N-2,N}}{M^{(1)}_{N-1,N}} \tag{10.18}$$

The trisection determinant (10.16) for the new position of the trisection yields $M^{(1)}_{N-2,N-1} \cdot M^{(1)}_{N-1,N} = M^{(1)}_{N-2,N} \cdot (\omega_{01} + M^{(1)}_{N-1,N-1})$, which is rearranged and substituted into equation (10.18) such that

$$\theta_{12} = \tan^{-1}\left[\frac{M^{(1)}_{N-2,N}}{M^{(1)}_{N-1,N}}\right] = \tan^{-1}\left[\frac{M^{(1)}_{N-2,N-1}}{\omega_{01} + M^{(1)}_{N-1,N-1}}\right] \tag{10.19}$$

Thus, from equations (10.17)–(10.19), it is evident that application of the rotation at pivot $[N-2, N-1]$ and angle θ_{12}, to pull the trisection up the diagonal by one position, automatically conditions it in its new position to again give a value of zero for its determinant. This applies for a shift of the trisection wherever it is on the diagonal of the arrow matrix.

The trisection is then pulled up the diagonal by a further index position ($k = N-3$) by another rotation at pivot $[N-3, N-2]$ and angle $\theta_{23} = \tan^{-1}(M^{(2)}_{N-3,N-1}/M^{(2)}_{N-2,N-1})$. The result is that element $M^{(3)}_{N-3,N-1}$ is zero, and $M^{(3)}_{N-4,N-2}$ is created in the new matrix $\mathbf{M}^{(3)}$, and so on until the trisection is in its desired position on the diagonal. Once the trisection has been pulled clear of the arrow structure of the matrix, the outermost cross-coupling elements in the last row and column disappear; in other words, the coupling diagram wheel (Figure 10.21a) "loses a spoke."

If the transfer function incorporates a second TZ at $s_{02} = j\omega_{02}$, this procedure is repeated to produce a second trisection. This time it is associated with s_{02}, and so on, until a cascade of trisections realizing $s_{01}, s_{02}, \ldots$ is formed. Then further rotations are applied to the cascade to form quartets and quintets, introduced in Section 10.3.3.

10.3.2.4 Illustrative Example

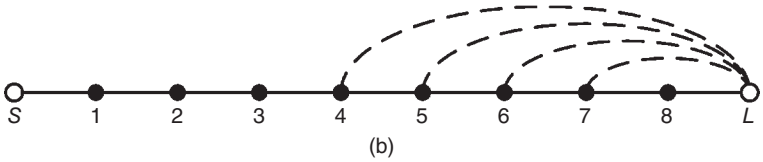
To demonstrate the application of the procedure, let us consider an example of a 23 dB return loss (8–2–2) asymmetric filter function with two TZs at $-j1.2520$ and $-j1.1243$ on the lower side of the passband producing two rejection lobes of 40 dB each and a complex zero pair at $\pm 0.8120 + j0.1969$ to equalize the group delay over approximately 50% of the passband. Two trisections realizing the two imaginary-axis zeros are synthesized and then merged to form a quartet.

Generating the $N+2$ coupling matrix for this transfer function and converting it to the arrow form using the reduction procedure outlined earlier (28 rotations) results in the matrix shown in Figure 10.22a and the corresponding coupling/routing diagram in Figure 10.22b. Note that because there are four TZs in the transfer function, there are four cross-coupling elements in the last row and column of the matrix, in addition to the mainline coupling M_{8L}.

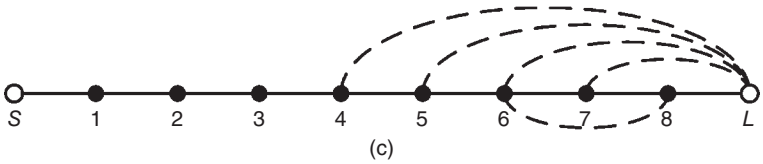
For the first imaginary-axis zero $s_{01} = -j1.2520$, the first rotation angle is evaluated from equation (10.17) as $\theta_{01} = -j1.2747^{\circ}$. Applying the first rotation at pivot [7, 8] to create the trisection (Figure 10.22c) and then five further rotations to pull the trisection up the diagonal from index position 7 (Figure 10.22d) to index position 2 (Figure 10.22e) completes the formation and positioning of the first trisection realizing the first TZ at s_{01}. The second cycle begins by evaluating θ_{67} with $s_{02} = -j1.1243$ in equation (10.17), taking the coupling values from the matrix resultant at the end of the first cycle. Applying the first rotation of the new cycle again at

	S	1	2	3	4	5	6	7	8	L
S	0	1.0516	0	0	0	0	0	0	0	0
1	1.0516	–0.0276	0.8784	0	0	0	0	0	0	0
2	0	0.8784	–0.0324	0.6147	0	0	0	0	0	0
3	0	0	0.6147	–0.0464	0.5813	0	0	0	0	0
4	0	0	0	0.5813	–0.1049	0.6073	0	0	0	–0.1789
5	0	0	0	0	0.6073	–0.2413	0.6171	0	0	0.2370
6	0	0	0	0	0	0.6171	–0.0477	0.5511	0	–0.3621
7	0	0	0	0	0	0	0.5511	0.3380	0.0027	0.8164
8	0	0	0	0	0	0	0	0.0027	1.1293	0.4691
L	0	0	0	0	–0.1789	0.2370	–0.3621	0.8164	0.4691	0

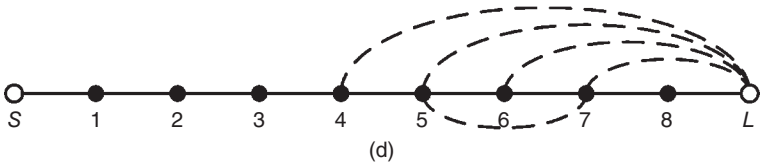
(a)



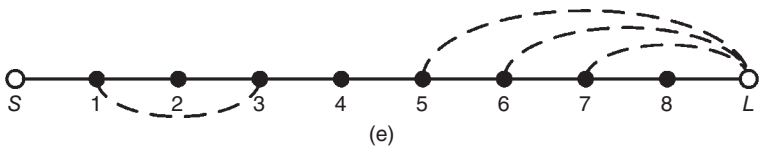
(b)



(c)



(d)



(e)

Figure 10.22 (8–2–2) filter synthesis example: (a) initial arrow coupling matrix; (b) corresponding coupling–routing diagram; (c) rotation 1 creates first trisection at index position 7; (d) rotation 2 pulls the trisection to index position 6; (e) rotation 6 positions the first trisection at index position 2.

pivot [7, 8] and then pulling the newly created trisection up the diagonal with three rotations to index position 4 completes the creation and positioning of the second trisection. The whole process is summarized in Table 10.6.

After the two trisections are created and repositioned, the coupling matrix and its corresponding coupling and routing diagram take the form as shown in Figures 10.23a and b. Note that because two TZs are extracted from the arrow structure and realized as trisections, the cross-couplings M_{4L} and M_{5L} at the extremes of the arrowhead in the last row and column now have zero values. This automatically occurs during the trisection shifting process. No specific action to annihilate them is needed.

Table 10.6 (8–2–2) filter example: rotation sequence for the creation and positioning of two trisections.

Rotation number r	Pivot $[i, j]$	Element to be annihilated	$\theta_{r-1,r} = \tan^{-1}(M_{kl}^{(r-1)}/M_{mn}^{(r-1)})$ [Equation (8.38a)]				Comments[a]
			k	l	m	n	
1	[7, 8]			$\theta_{01} = -1.2747°$			Create first trisection at IP7 with $\omega_{0i} = \omega_{01} = -1.2520$ [see equation (10.17)]; coupling M_{68} is created
2	[6, 7]	M_{68}	6	8	7	8	Rotations 2–6: to shift trisection 1 from IP7 to IP2
3	[5, 6]	M_{57}	5	7	6	7	
4	[4, 5]	M_{46}, M_{4L}	4	6	5	6	
5	[3, 4]	M_{35}	3	5	4	5	
6	[2, 3]	M_{24}	2	4	3	4	
7	[7, 8]			$\theta_{67} = 79.3478°$			Create second trisection at IP7 with $\omega_{0i} = \omega_{02} = -1.1243$ [see equation (10.17)]; coupling M_{68} is created again
8	[6, 7]	M_{68}	6	8	7	8	Rotations 8–10: to shift trisection 2 from IP7 to IP4
9	[5, 6]	M_{57}, M_{5L}	5	7	6	7	
10	[4, 5]	M_{46}	4	6	5	6	

a) IP = index position.

	S	1	2	3	4	5	6	7	8	L
S	0	1.0516	0	0	0	0	0	0	0	0
1	1.0516	–0.0276	0.6953	–0.5368	0	0	0	0	0	0
2	0	0.6953	–0.6920	0.4324	0	0	0	0	0	0
3	0	–0.5368	0.4324	–0.1272	0.4045	–0.4041	0	0	0	0
4	0	0	0	0.4045	0.7072	0.4167	0	0	0	0.0000
5	0	0	0	–0.4041	0.4167	–0.0435	0.6252	0	0	0.0000
6	0	0	0	0	0	0.6252	0.0058	0.7449	0	–0.4142
7	0	0	0	0	0	0	0.7449	–0.0436	0.5911	0.0858
8	0	0	0	0	0	0	0	0.5911	–0.1961	0.9628
L	0	0	0	0	0.0000	0.0000	0.4142	0.0858	0.9628	0

(a)

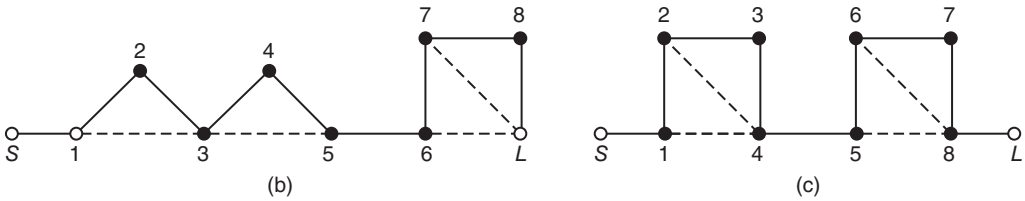


Figure 10.23 (8–2–2) asymmetric filter function: (a) coupling matrix after synthesis of two trisections; (b) coupling–routing diagram; (c) coupling–routing diagram after formation of two quartets.

The two trisections can now be a merged into a quartet by pulling the second trisection up the diagonal by one further position with a rotation at pivot [3, 4] annihilating M_{35} and then one rotation at [2, 3] annihilating either M_{13} or M_{24}. A second quartet (realizing the complex TZ pair) already exists as the remainder of the arrow formation (see Figure 10.23a); this can be pulled up the diagonal by one position by applying two rotations, one at pivot [7, 8] to annihilate M_{7L} and then one at [6, 8] to annihilate M_{6L}. The final form of the (8–2–2) cascade quartet coupling/routing diagram is shown in Figure 10.23c.

10.3.2.5 Trisections Realizing Complex Transmission Zeros

By applying the method described earlier, each trisection is created individually within the arrow structure and then moved to a new position on the diagonal. Each trisection has one TZ uniquely associated with it. By forming the trisection and moving it from the arrow formation, effectively that single TZ is removed from the arrow formation too. When the TZ s_{0i} is pure imaginary, equation (10.17) gives a real value for the rotation angle θ_{ti}, and the rotations may proceed conventionally.

However, only imaginary-axis TZs can exist individually. For off-axis TZs (to give group delay equalization), these must exist in mirror-image pairs about the imaginary axis for a realizable network. Because each trisection is capable of realizing only one of the zeros from the complex or real-axis pair, equation (10.17) yields a complex value $\theta_{ti} = a + jb$ for the initial rotation angle to be applied to the matrix (rotation 1 in Table 10.6). In such cases, the complex values for $\sin\theta_{ti}$ and $\cos\theta_{ti}$, which are needed for the rotation (see Section 8.3.1), are calculated [11] as follows:

$$\theta_{ti} = \tan^{-1}\left[\frac{M_{N-1,N}}{\omega_{0i} + M_{N,N}}\right] = \tan^{-1}(x + jy) = a + jb \tag{10.20}$$

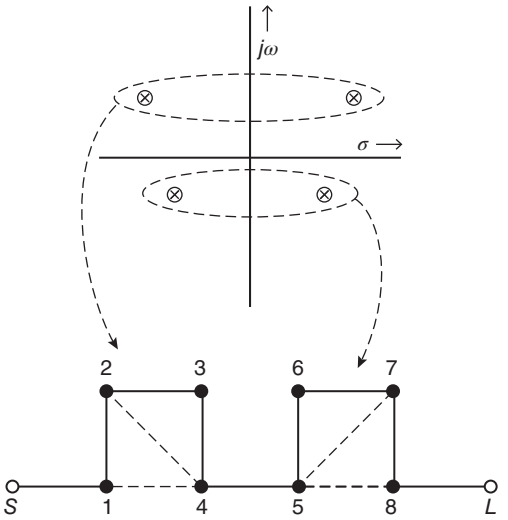


Figure 10.24 Complex zero pairs: realization with cascade quartets.

where

$$a = \frac{1}{2}\tan^{-1}\left[\frac{2x}{1-x^2-y^2}\right], \quad b = \frac{1}{4}\log_e\left[\frac{x^2+(y+1)^2}{x^2+(y-1)^2}\right]$$

Now knowing a and b, we can calculate $\sin\theta_{ti}$ and $\cos\theta_{ti}$ from

$$\begin{aligned}\sin\theta_{ti} &= \sin a\cosh b + j\cos a\sinh b = -j\sinh(j(a+jb))\\ \cos\theta_{ti} &= \cos a\cosh b - j\sin a\sinh b = \cosh(j(a+jb))\end{aligned} \tag{10.21}$$

The application of this first rotation by a complex angle gives complex values to some of the coupling elements of the matrix, and when the trisection is to be pulled up the diagonal, the rotation angles required to annihilate the necessary elements (e.g., rotations 2–6 in Table 10.6) also acquire complex values. For the rth rotation

$$\theta_r = \tan^{-1}\frac{M_{ij}}{M_{kl}} = \tan^{-1}(x_r + jy_r) = a_r + jb_r \tag{10.22}$$

and $\sin\theta_r$ and $\cos\theta_r$ are calculated in the same way as the initial rotation by using (10.20) and (10.21).

Consequently, after the first series of rotations has been applied to create and reposition the first complex TZ s_{01}, the coupling matrix has complex values, which of course are unrealizable. However, if the next trisection is created for the second TZ of the mirror-image pair ($s_{02} = -s_{01}^*$) and is pulled up the diagonal and merged with the first trisection (as in the eighth-degree example) to form a quartet (Figure 10.23c), then all the imaginary parts of the complex coupling values cancel out, leaving pure real coupling values. This procedure is necessary if cascade quartets are synthesized for more than one complex pair, as in Figure 10.24.

10.3.3 Techniques Based on the Trisection for Synthesis of Advanced Circuits

In this section, the synthesis of two useful network configurations is explained, the cascaded n-tuplet (such as quartets, quintets, sextuplets) and the *box* section, including the method for its development to *extended box* sections. Both are valuable for microwave filter designers, offering advantages in design and performance and in the manufacturing and tuning processes. Typically, the procedure for the synthesis of these networks starts with the generation of a cascade

of trisections, which is achieved by the circuit methods described earlier in this section or by using the arrow matrix method.

10.3.3.1 Synthesis Procedure for Cascade Quartets, Quintets, and Sextets

The method for synthesis of cascaded quartets is summarized first, followed by the extension of the method to cascaded quintets and sextuplets. It is possible to extend the method beyond sextuplets, but no practical application is foreseen for such configurations.

10.3.3.2 Cascaded Quartets

The procedure begins with the synthesis of a pair of conjoined trisections at s_{01} and s_{02}, as shown for the asymmetric (8–2–2) filter (two imaginary-axis TZs, one complex pair) example in Figure 10.23a. The rest of the network may then be synthesized conventionally as a folded network, if it is appropriate [automatically forming the second quartet of the (8–2–2) example]. The next step is to apply two rotations, the first at pivot [3, 4] to annihilate M_{35}. This rotation creates two new couplings, M_{14} and M_{24}. Now it can be seen that a quartet with both diagonal couplings M_{13} (which was originally present in the first trisection) and M_{24} has been formed. Either of these couplings can be eliminated with a second rotation pivot [2, 3] to annihilate either M_{13} or M_{24}.

10.3.3.3 Formal Procedure for Cascade Quartets, Quintets, and Sextets

When we follow the procedure for the synthesis of cascade quartets, it is noticeable that the sequence of rotations for the annihilation of the couplings to finally obtain the quartet has two stages:

1) Rotations are carried out to annihilate any trisection cross-coupling that is outside the 4×4 submatrix in the main matrix that the quartet eventually occupies [M_{35} in the (8–2–2) example]. Additional couplings are created during this process, but the new couplings are within the 4×4 submatrix. This stage in the process is known as "gathering in."
2) A sequence of rotations is applied within the 4×4 matrix to reduce the couplings (in general, they are nonzero) to the folded form. This sequence of rotations is the same as that used to reduce the $N \times N$ or $N + 2$ matrix in their raw forms to the folded form. The process is summarized in Table 10.7 and illustrated graphically in Figure 10.25. The (8–2–2) example is used for this illustration, where the first trisection is synthesized at index $k = 2$ and the second at $k = 4$, creating M_{13} and M_{35}. However, the process can be applied to the trisection group, wherever it occurs in the main matrix, indexing the subscripts of the couplings and the rotation coordinates appropriately.

Table 10.7 (8–2–2) example of the rotation procedure to transform the two trisections to the cascaded quartet form.

Rotation number r	Pivot $[i, j]$	Element to be annihilated		$\theta_r = \tan^{-1}(cM_{kl}/M_{mn})$				
				k	l	m	n	c
"Gather in" to 4×4 submatrix								
1	[3, 4]	M_{35}		3	5	4	5	+1
Reduction to folded form within 4×4 submatrix								
2	[2, 3]	M_{13}	In row 1	1	3	1	2	−1
or 2	[2, 3]	M_{24}	In row 1	2	4	3	4	+1

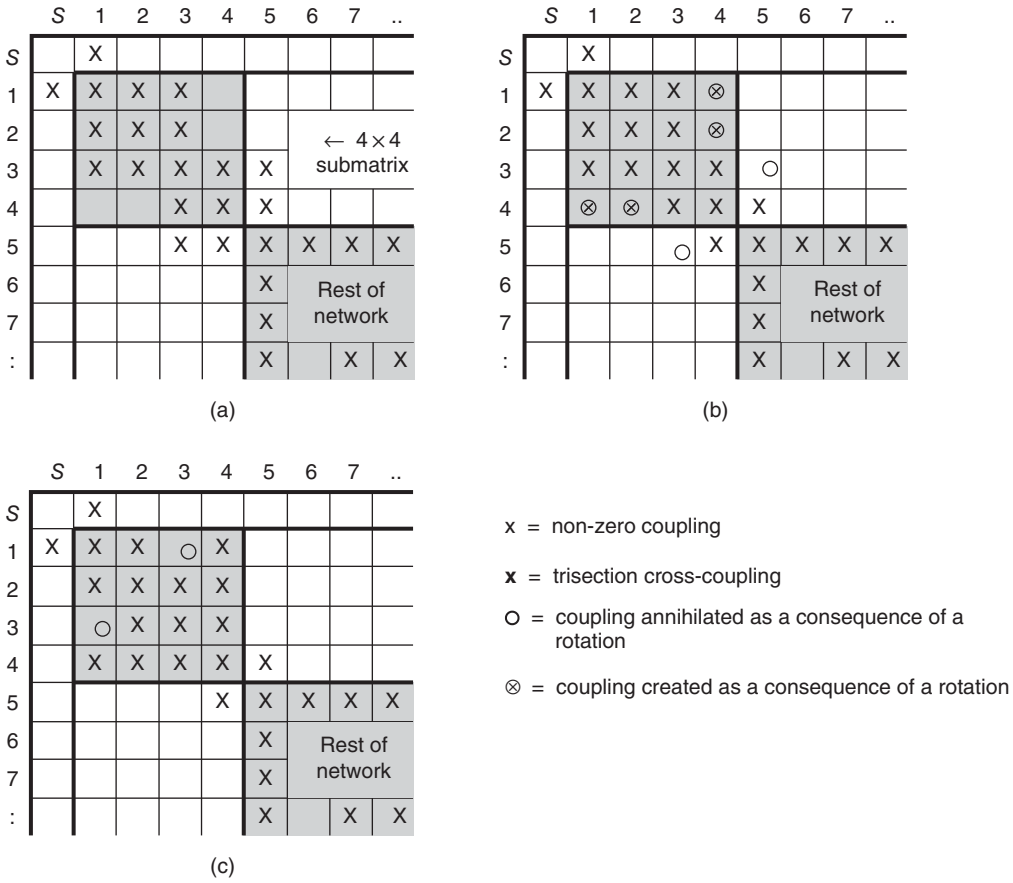


Figure 10.25 Formation of cascade quartet: (a) realization of TZs with two trisections (see Figure 10.26a) (trisection cross-couplings shown in bold); (b) annihilation of M_{35}, creating M_{14} and M_{24} (Figure 10.26b); (c) rotation at [2, 3] to annihilate M_{13} and thus create first quartet in Figure 10.26c.

Quartets are easily shifted by one index position within the network by applying a couple of rotations to the coupling matrix. For example, to move the first quartet in Figure 10.26b down the diagonal by one position (to become conjoined with the second quartet at node 5), two rotations are applied: the first at pivot [2, 4] to annihilate M_{14} (creating M_{25}) and the second at pivot [3, 4] to annihilate M_{24} (if required).

10.3.3.4 Cascaded Quintets

A quintet (see Figure 10.27 and Table 10.8) is capable of realizing three TZs and is created by first generating a cascade of three conjoined trisections from the three TZs (Figure 10.28).

10.3.3.5 Cascaded Sextets

A sextet (see Figure 10.29 and Table 10.9) is generated from a cascade of four conjoined trisections (Figure 10.30a) and is capable of realizing four TZs.

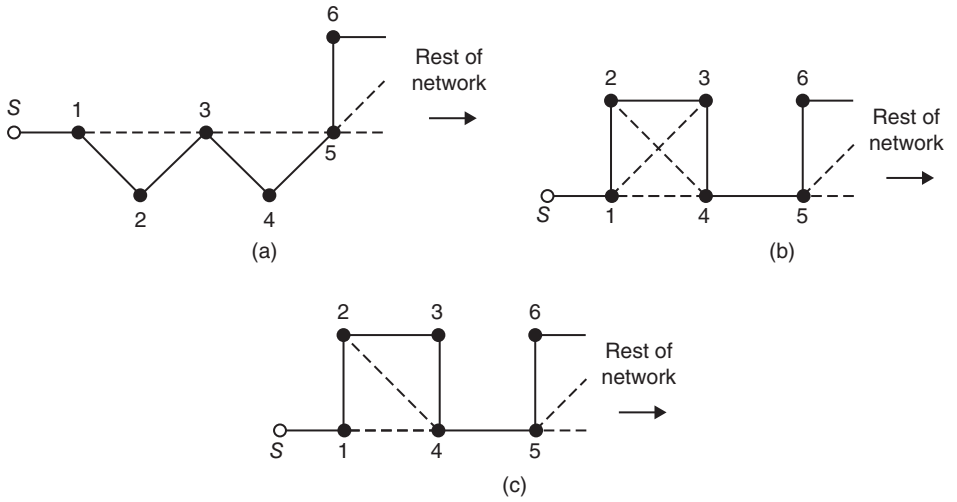


Figure 10.26 Transformation of two conjoined trisections to form a quartet section.

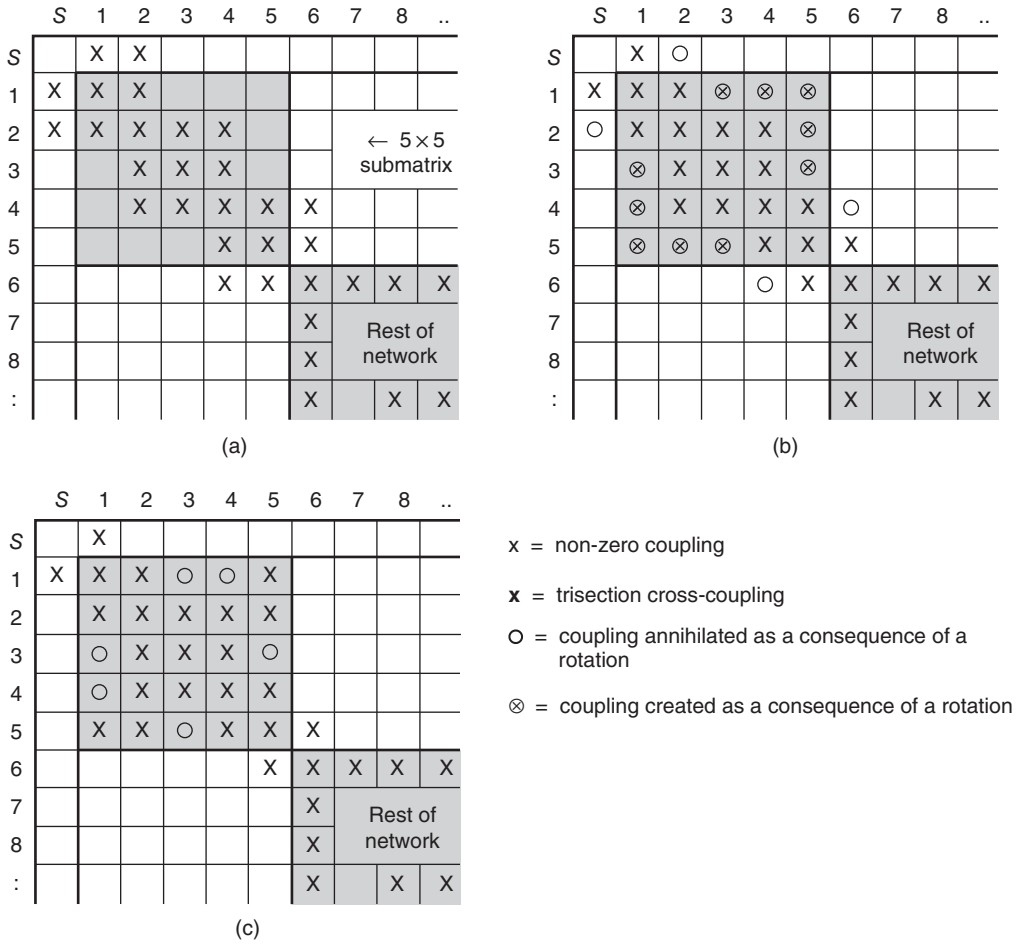


Figure 10.27 Formation of cascade quintet: (a) realization of TZs with three trisections (trisection cross-couplings shown in bold); (b) annihilation of M_{52} and M_{46}; (c) after series of three rotations to create the folded network in 5 × 5 submatrix.

Table 10.8 Rotation procedure to transform three conjoined trisections to the cascaded quintet form.

Rotation number r	Pivot $[i, j]$	Element to be annihilated		$\theta_r = \tan^{-1}(cM_{kl}/M_{mn})$				
				k	l	m	n	c
		"Gather in" to 5×5 submatrix						
1	[1, 2]	M_{S2}		S	2	S	1	−1
2	[4, 5]	M_{46}		4	6	5	6	+1
		Reduction to folded form within 5×5 submatrix						
3	[3, 4]	M_{14}	In row 1	1	4	1	3	−1
4	[2, 3]	M_{13}	—	1	3	1	2	−1
5	[3, 4]	M_{35}	In column 5	3	5	4	5	+1

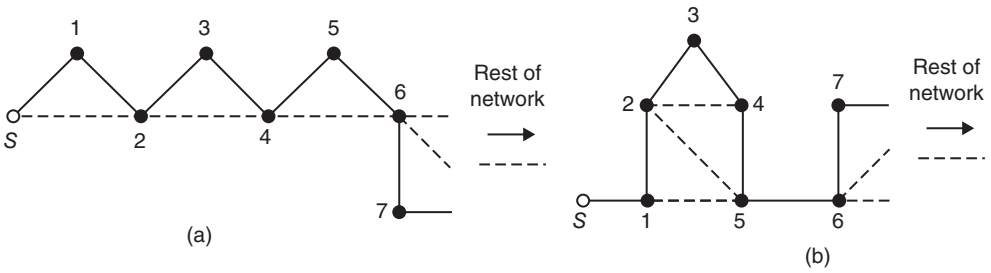


Figure 10.28 Transformation of three conjoined trisections to form a quintet section.

10.3.3.6 Multiple Cascaded Sections

Although this formal procedure is able to form cascaded n-tet sections relatively easily, it does actually need more "room" along the diagonal of the matrix as the number of sections or their order increases. This is because after the formation of the first trisection, which realizes the first TZ, each new conjoined trisection that is synthesized to realize one additional TZ takes up two additional circuit nodes. This may be seen in Figure 10.30a, for example, where four trisections take up the source termination plus eight resonator nodes for the production of four TZs, where the eventual sextet realizing those four zeros only occupies the first six nodes not including the source node (Figure 10.30b). This of course, leaves fewer nodes available in the coupling matrix for the synthesis of further trisections and n-tet sections.

The problem may be overcome if the direct-coupling matrix method to create the basic cascaded trisections is used (Section 10.3.2). Because we are working purely with rotations on coupling matrices, the n-tet sections may be individually created as the appropriate number of trisections is pulled out of the arrow formation. For example, two trisections may be first extracted from the arrow formation and then merged using the methods earlier to create a quartet section in its final position in the coupling matrix. Then further trisections may be extracted and merged to create a second n-tet section in cascade with the first. By forming the n-tet sections as the appropriate number of basic trisections are pulled out of the arrow formation, less "room" is needed along the diagonal of the coupling matrix than if all the trisections are produced first in a cascade, and then the n-tet sections formed as a second step in the overall process.

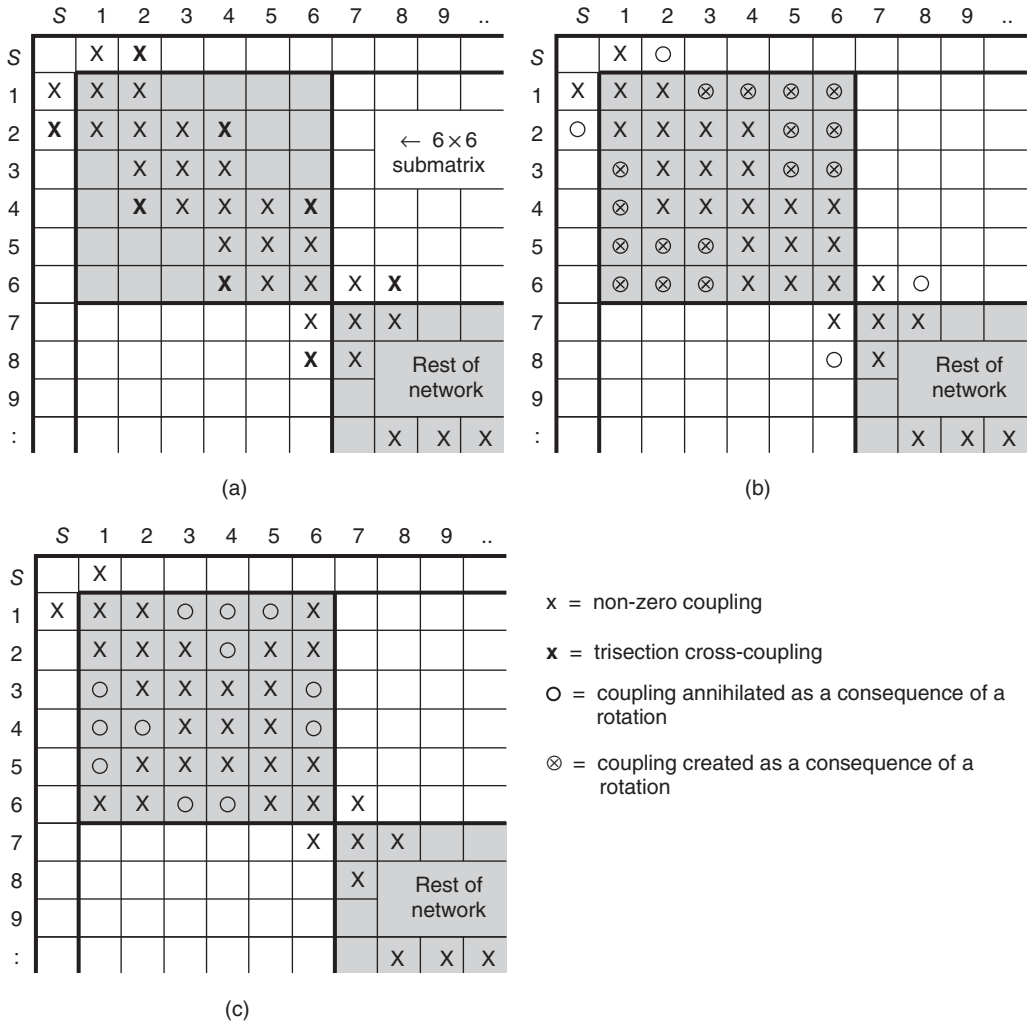


Figure 10.29 Formation of cascade sextet: (a) realization of TZs with four trisections (trisection cross-couplings are shown in bold); (b) annihilation of M_{52} and M_{68}; (c) after series of six rotations to create a folded network within the 6 × 6 submatrix.

Alternative circuit-based methods for the formation of trisections quartet and quintet sections may be found in Refs [12–15].

10.4 Box Section and Extended Box Configurations

For certain applications, specifications for channelizing filters require asymmetric rejection characteristics. This is particularly true for the front-end transmit/receive diplexers in the base stations of mobile communication systems and requires the use of filter characteristics optimally tailored to the rejection requirements; at the same time the maximum in-band amplitude and group delay linearity and lowest insertion loss must be maintained. In the course of synthesizing the networks for such characteristics and reconfiguring the interresonator mainline and cross-couplings, it is frequently found that diagonal couplings are required. In

Table 10.9 Rotation procedure to transform four trisections to cascaded sextet form.

Rotation number r	Pivot $[i, j]$	Element to be annihilated		$\theta_r = \tan^{-1}(cM_{kl}/M_{mn})$				
				k	l	m	n	c
		"Gather in" to 6×6 submatrix						
1	[1, 2]	M_{S2}		S	2	S	1	−1
2	[6, 7]	M_{68}		6	8	7	8	+1
3	[5, 6]	M_{57}		5	7	6	7	+1
4	[4, 6]	M_{47}		4	7	6	7	+1
		Reduction to folded form within 6×6 submatrix						
5	[4, 5]	M_{15}	In row 1	1	5	1	4	−1
6	[3, 4]	M_{14}	—	1	4	1	3	−1
7	[2, 3]	M_{13}	—	1	3	1	2	−1
8	[3, 4]	M_{36}	In column 6	3	6	4	6	+1
9	[4, 5]	M_{46}	—	4	6	5	6	+1
10	[3, 4]	M_{24}	In row 2	2	4	2	3	−1

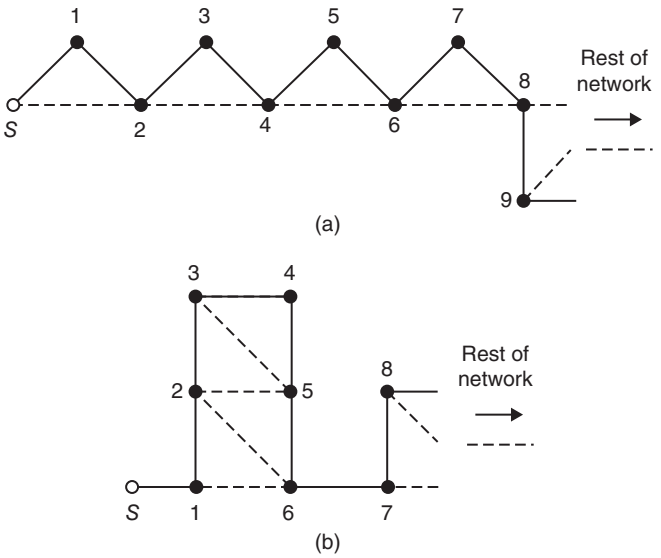


Figure 10.30 Transformation of four conjoined trisections to form a sextet section. Rotation 2 annihilates M_{68} and creates M_{57} and M_{47}, which are annihilated by rotations 3 and 4.

this section, a synthesis method is presented for the realization of symmetric or asymmetric filtering characteristics without the need for diagonal cross-couplings: the *box* configuration and a derivation known as the *extended box* configuration [16].

10.4.1 Box Sections

The box section is similar to the cascade quartet section, with four resonator nodes arranged in a square formation, but with the input and the output of the quartet at opposite corners of the square. Figure 10.31a illustrates the conventional quartet arrangement for a fourth-degree filter

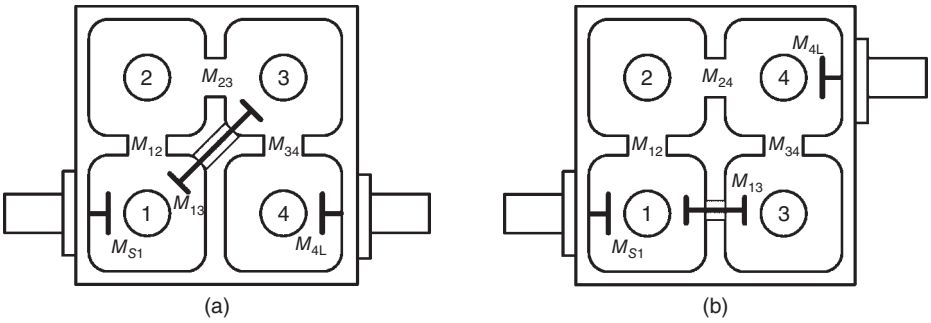


Figure 10.31 (4–1) asymmetric filter function: (a) realized with the conventional diagonal cross-coupling (M_{13}); and (b) realized with the box configuration.

with a single TZ, realized with a trisection. Figure 10.31b displays the equivalent box section realizing the same TZ, but without the need for the diagonal coupling. The application of the minimum path rule indicates that the box section can realize only a single TZ.

The box section is created by the application of a cross-pivot similarity transform [equation (8.38g)] to a trisection that has been synthesized in the coupling matrix for the filter. A cross-pivot similarity transform is the case where the coordinates of the coupling matrix element to be annihilated are the same as the pivot of the transform itself, that is, the element to be annihilated lies on the cross-points of the pivot [see equation (9.10)]. Arbitrary numbers of 90° are added to the cross-pivot rotation angle to create alternative solutions given by

$$\theta_r = \frac{1}{2}\tan^{-1}\left(\frac{2M_{ij}}{(M_{jj} - M_{ii})}\right) \pm k\frac{\pi}{2} \tag{10.23}$$

For the box section, the pivot is set to annihilate the second mainline coupling of the trisection in the coupling matrix; thus, pivot = [2, 3] annihilates element M_{23} (and M_{32}) in the fourth-degree example of Figure 10.31a. The equivalent coupling and routing schematic is given in Figure 10.32a. In the process of annihilating the mainline coupling M_{23}, the coupling M_{24} is created, as shown in Figure 10.32b. Then, by untwisting this section, the box section in Figure 10.32c is formed. In the resultant box section, one of the couplings is always negative, irrespective of the sign of the cross-coupling (M_{13}) in the original trisection.

To illustrate the procedure, we examine a fourth-degree 25 dB return loss Chebyshev filter with a single TZ at $s = +j2.3940$, providing a lobe level of 41 dB on the upper side of the stopband.

Figure 10.33a gives the $N+2$ coupling matrix for the (4–1) filter showing the M_{13} trisection cross-coupling, corresponding to the coupling diagram of Figure 10.32a. Figure 10.33b shows the coupling matrix after transformation to the box configuration. Figure 10.34a plots

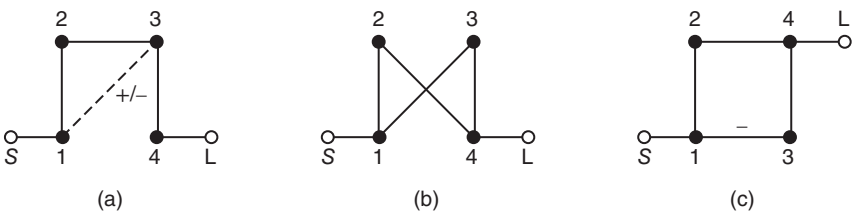


Figure 10.32 Formation of the box section for the (4–1) filter: (a) the trisection; (b) annihilation of M_{23} and creation of M_{24}; (c) untwisting to obtain box section.

	S	1	2	3	4	L
S	0	1.1506	0	0	0	0
1	1.1506	0.0530	0.9777	0.3530	0	0
2	0	0.9777	–0.4198	0.7128	0	0
3	0	0.3530	0.7128	0.0949	1.0394	0
4	0	0	0	1.0394	0.0530	1.1506
L	0	0	0	0	1.1506	0

(a)

	S	1	2	3	4	L
S	0	1.1506	0	0	0	0
1	1.1506	0.0530	0.5973	–0.8507	0	0
2	0	0.5973	–0.9203	0	0.5973	0
3	0	–0.8507	0	0.5954	0.8507	0
4	0	0	0.5973	0.8507	0.0530	1.1506
L	0	0	0	0	1.1506	0

(b)

	S	1	2	3	4	L
S	0	1.1506	0	0	0	0
1	1.1506	–0.0530	0.5973	–0.8507	0	0
2	0	0.5973	0.9203	0	0.5973	0
3	0	–0.8507	0	–0.5954	0.8507	0
4	0	0	0.5973	0.8507	–0.0530	1.1506
L	0	0	0	0	1.1506	0

(c)

Figure 10.33 (4–1) filter coupling matrices: (a) the trisection; (b) after transformation to box section (TZ on upper side of passband); (c) TZ on lower side of passband.

the measured results of a coaxial resonator (4–1) filter configured as in Figure 10.31b. Good correlation is obtained between the measured and EM simulated results.

If the TZ is placed at $-j2.3940$ below the passband instead of above, then the transformation to the box section results in the same values for the interresonator couplings, but complementary values for the self-couplings (M_{11}, M_{22}, … along the principal diagonal of the coupling matrix, in Figure 10.33c). Since the self-couplings represent offsets from the center frequency and are adjustable by tuning screws, the same filter structure can be used for the complementary filters of a transmit/receive diplexer, for example. Figure 10.34b shows the measured results of the same structure that gave the single TZ on the upper side of the passband (Figure 10.34a), now retuned to place the TZ below the passband.

In addition, box sections can be cascaded within higher-degree filters, indexing the coordinates of the cross-pivot rotation appropriately to correspond correctly with the position of each trisection. Figure 10.35 represents the coupling and routing diagrams for a tenth-degree filter with two TZs on the lower side of the passband. Figure 10.36 shows the measured and simulated return loss and rejection characteristics of a test filter.

It is evident in Figure 10.35b that we can realize this asymmetric characteristic in dual-mode cavities (there are no diagonal cross-couplings).

10.4.2 Extended Box Sections

A series of box sections may be formed in higher-degree networks by cascading trisections and applying the single cross-pivot rotation to eliminate the mainline coupling within each one and then untwisting. However the closest that box sections may be created by this method is where

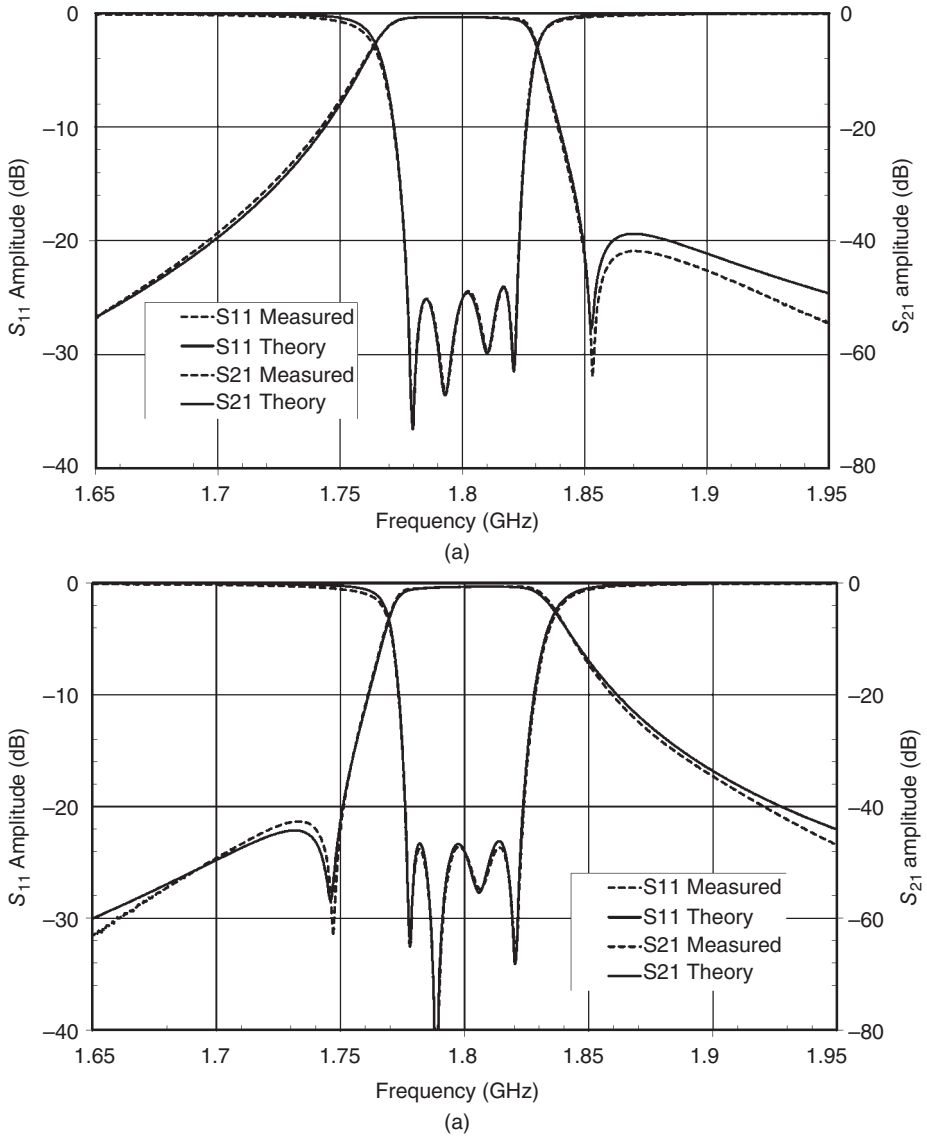


Figure 10.34 Measured results of the (4–1) filter: (a) transmission zero on upper side; (b) transmission zero on lower side.

two sections share a common resonator, for example, the 8th-degree network of Figure 10.37. This leads to a rather awkward and restricted layout and heavily loads the common resonator (resonator 4 of Figure 10.37) with four couplings. Moreover, only two TZs may be realized by this network.

A rather more convenient arrangement in terms of both physical layout and the number of TZs that may be realized is shown in Figure 10.38. Here the basic 4th-degree box section is shown, as well as the addition of pairs of resonators to form 6th-, 8th-, and 10th-degree networks. Application of the minimum path rule indicates that a maximum of $1,2,3,4,\ldots,(N-2)/2$ TZs may be realized by the 4th-, 6th-, 8th-, 10th-, …, Nth-degree networks, respectively. The resonators are arranged in two parallel rows with half the total number of resonators in

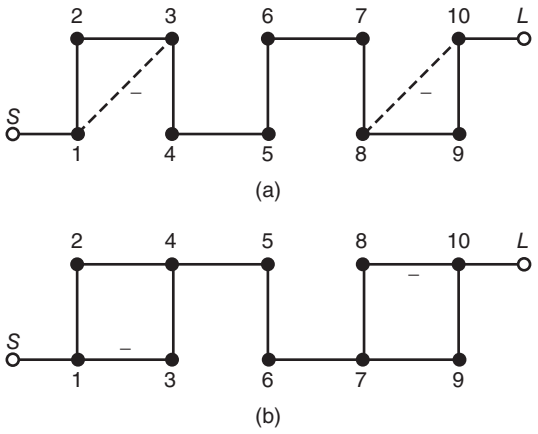


Figure 10.35 Coupling and routing diagrams of the (10–2) asymmetric filter: (a) synthesized with two trisections; (b) after transformation of trisections to two box sections by application of two cross-pivot rotations at pivots [2, 3] and [8, 9]. This form is suitable for realization in a dual-mode technology.

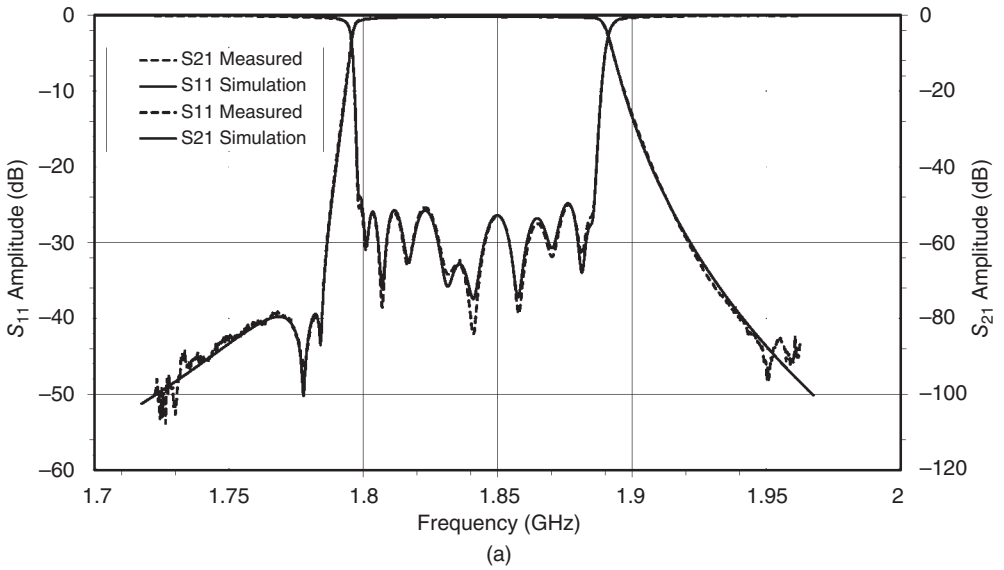


Figure 10.36 Simulated and measured return loss and rejection of the (10–2) asymmetric filter.

each row; input is at the corner at one end, and output is from the diagonally opposite corner at the other end. Even though asymmetric characteristics may be prescribed, there are no diagonal cross-couplings.

There appears to be no regular pattern for determining the sequence of rotations to synthesize the coupling matrix for the extended box sections from the folded network or any other canonical network. The sequences in Table 10.10 are derived by first determining the sequence of rotations to reduce the extended box section network to the corresponding folded network coupling matrix, as described in Section 8.3. This sequence of rotations is then reversed to transform the folded coupling matrix (which is readily derived from the S_{21} and S_{11} polynomials by using the methods of Chapters 7 and 8), back to the box section matrix.

Reversing the sequence means that some of the rotation angles θ_r are unknown *a priori* and can be determined only by relating the coupling elements of the final matrix with those of the initial folded matrix and then solving to find θ_r (similar to the procedure for synthesizing cascade quadruplets in an 8th-degree network [17]). Deriving the equations to solve for θ_r analytically involves a formidable amount of algebra, and the task is most easily performed by setting up a solver procedure.

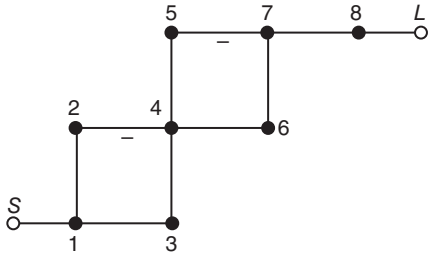


Figure 10.37 8th-degree network—two box sections conjoined at one corner.

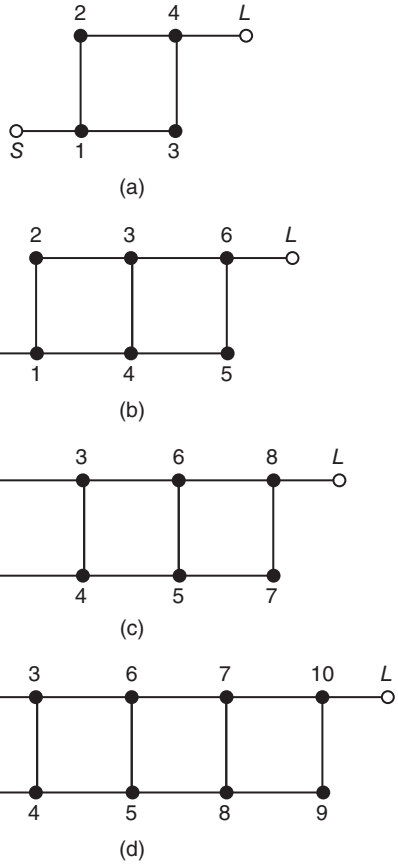


Figure 10.38 Coupling and routing diagrams for extended box section networks (a) 4th degree (basic box section); (b) 6th degree; (c) 8th degree; (d) 10th degree.

The procedure starts by setting the initial values for the unknown rotation angles (one for the 6th degree, two for the 8th degree, and five for the 10th degree) and then running through the sequence of rotations as shown in Table 10.10 to obtain the transformed coupling matrix. With these initial angles, some of the unwanted coupling matrix elements will have nonzero values. An error function can be evaluated by forming, for example, the root sum square of the values of all the elements in the coupling matrix that should be zero. The initial angles are then adjusted by the solver routine until the cost function is zero. By varying the value of the integer k in the cross-pivot angle formula [equation (10.23)] when it is used in the rotation sequences, several unique solutions can be found in most cases. This allows the choice of the coupling matrix that gives the most convenient coupling values for the technology to be used for the realization.

Let us consider an 8th-degree filter with 23 dB return loss and three prescribed TZs at $s = -j1.3553$, $+j1.1093$, and $+j1.2180$, which produce one rejection lobe level of 40 dB on the lower

Table 10.10 Pivot coordinates for reduction of folded matrix to extended box section configuration.

Degree *N*	Rotation number *r*	Rotation pivot [*i*, *j*]	Angle θ_r	Annihilating element(*s*)
6	1	[3, 4]	θ_1	—
	2	[2, 4]		M_{24}
	3	[3, 5]		M_{25}, M_{35}
8	1	[4, 5]	θ_1	—
	2	[5, 6]	θ_2	—
	3	[3, 5]		M_{37}
	4	[3, 4]		M_{46}
	5	[3, 5]		M_{35}
	6	[2, 4]		M_{24}, M_{25}
	7	[6, 7]		M_{67}, M_{37}
10	1	[5, 6]	θ_1	—
	2	[5, 7]	θ_2	—
	3	[6, 7]	θ_3	—
	4	[4, 7]	θ_4	—
	5	[4, 6]	θ_5	—
	6	[4, 5]		M_{47}
	7	[6, 8]		M_{48}
	8	[3, 5]		M_{37}, M_{38}
	9	[7, 8]		M_{57}, M_{68}
	10	[3, 4]		M_{46}, M_{35}
	11	[7, 9]		M_{79}, M_{69}
	12	[2, 4]		M_{24}, M_{25}

side of the passband and two at 40 dB on the upper side. The folded $N + 2$ coupling matrix is shown in Figure 10.39a. By finding the two unknown angles θ_1 and θ_2 of one of the possible solutions (+63.881° and +35.865°, respectively) and applying the sequence of seven rotations (Table 10.10), the coupling matrix for the extended box section is obtained (Figure 10.39b). It corresponds to the coupling and routing diagram in Figure 10.38c. The analyzed transfer and reflection characteristics for this coupling matrix are shown in Figure 10.40, demonstrating that the performance has been preserved intact.

Another optimization technique, due to Seyfert [18], is a synthesis method based on computer algebra, which captures a much larger number of solutions. The process actually finds all solutions, but some are unrealizable because of complex coupling values, and the number of real solutions depends on the pattern of TZs in the prototype. Table 10.11 summarizes the approximate number of real solutions that are expected from prototypes of degrees 6–12 with the maximum number of TZs allowed by the minimum path rule.

It can be seen from Figure 10.38 that because there are no diagonal cross-couplings, extended box sections are also suitable for realization with dual-mode resonator cavities. A practical

	S	1	2	3	4	5	6	7	8	*L*
S	0	1.0428	0	0	0	0	0	0	0	0
1	1.0428	0.0107	0.8623	0	0	0	0	0	0	0
2	0	0.8623	0.0115	0.5994	0	0	0	0	0	0
3	0	0	0.5994	0.0133	0.5356	0	–0.0457	–0.1316	0	0
4	0	0	0	0.5356	0.0898	0.3361	0.5673	0	0	0
5	0	0	0	0	0.3361	–0.8513	0.3191	0	0	0
6	0	0	0	–0.0457	0.5673	0.3191	–0.0073	0.5848	0	0
7	0	0	0	–0.1316	0	0	0.5848	0.0115	0.8623	0
8	0	0	0	0	0	0	0	0.8623	0.0107	1.0428
L	0	0	0	0	0	0	0	0	1.0428	0

(a)

	S	1	2	3	4	5	6	7	8	*L*
S	0	1.0428	0	0	0	0	0	0	0	0
1	1.0428	0.0107	0.2187	0	–0.8341	0	0	0	0	0
2	0	0.2187	–1.0053	0.0428	0	0	0	0	0	0
3	0	0	0.0428	–0.7873	0.2541	0	–0.2686	0	0	0
4	0	–0.8341	0	0.2541	0.0814	0.4991	0	0	0	0
5	0	0	0	0	0.4991	0.2955	0.4162	0.1937	0	0
6	0	0	0	–0.2686	0	0.4162	–0.2360	0	–0.7644	0
7	0	0	0	0	0	0.1937	0	0.9192	0.3991	0
8	0	0	0	0	0	0	–0.7644	0.3991	0.0107	1.0428
L	0	0	0	0	0	0	0	0	1.0428	0

(b)

Figure 10.39 Extended box section configuration of an 8th-degree example: (a) original folded coupling matrix; (b) after transformation to extended box section configuration.

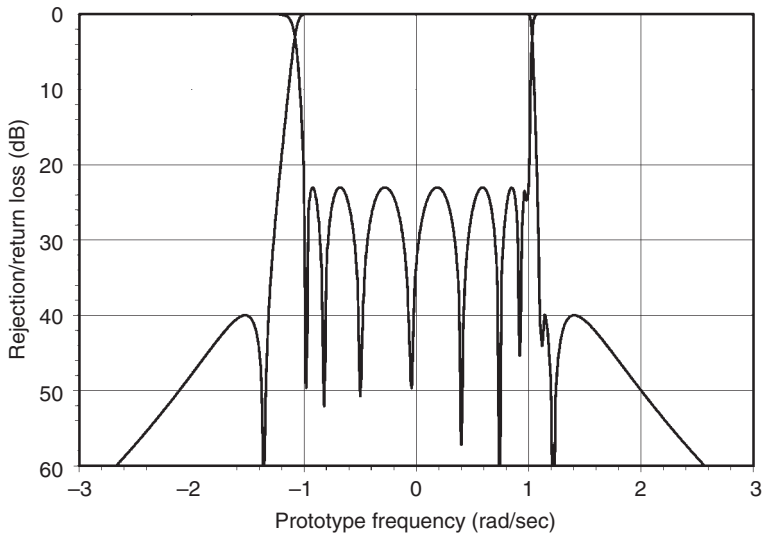


Figure 10.40 (8–3) asymmetric filter in an extended box section configuration—simulated rejection and return loss performance.

Table 10.11 Number of real solutions for the extended box configuration.

Extended box degree	Maximum number of finite-position transmission zeros n_{fz}	Number of real solutions (Approximate)
6	2	6
8	3	16
10	4	58
12	5	>2000

advantage arises from all the straight couplings being present, in that any spurious stray couplings become less significant. Also odd degree filters (7th, 9th, 11th, 13th, … etc.) can be synthesized by using the solutions for the 6th-, 8th-, 10th-, and 12th-degree filters, respectively.

By working on the subsections of the main coupling matrix, hybrid forms may also be created. An example is an 11th-degree filter with three TZs, initially synthesized with three trisections (Figure 10.41a). The first two trisections are then transformed into an asymmetric cascade quad section with two similarity transforms at pivots [4, 5] and [3, 4] to eliminate couplings M_{46} and M_{24}, respectively (Figure 10.41b). Then the iterative procedure described earlier for the 6th-degree case is applied to the upper left 6×6 submatrix of the main matrix to form the extended box section at the LHS of the network (Figure 10.41c). Finally, the basic box section at the right is formed with a cross-pivot transform at pivot [8, 9] to eliminate coupling M_{89}. Alternatively, applying the cross-pivot transform at pivot [9, 10] forms the basic box section at the extreme right of the network.

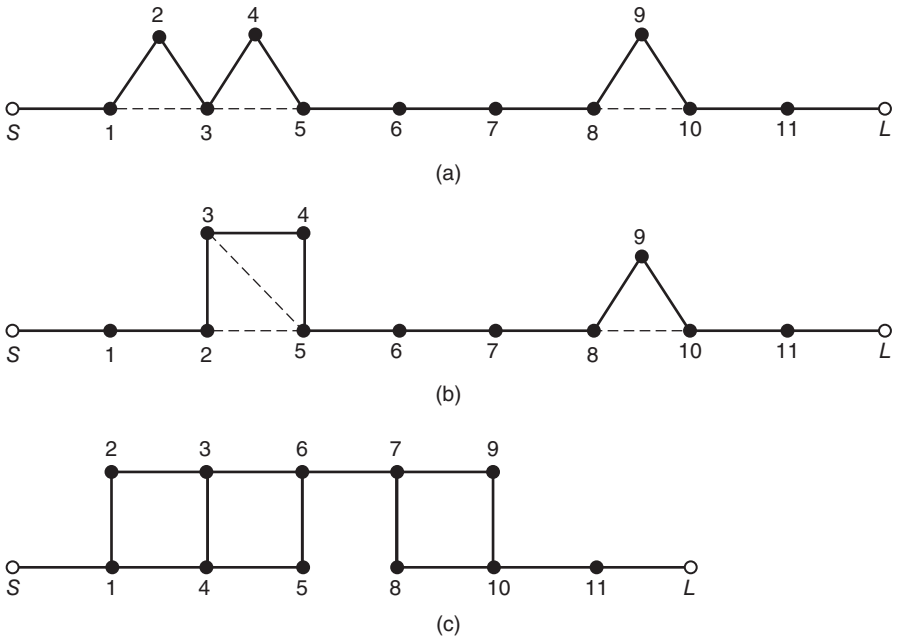


Figure 10.41 Stages in the synthesis of an (11–3) network with cascaded box sections: (a) initial synthesis with trisections; (b) formation of cascaded quad section; (c) formation of the 6th-degree extended box section in cascade with a basic box section.

Summary

Synthesis techniques described in Chapters 7, 8, and 9 are based on lossless lumped-element inductors, capacitors, FIR, and frequency-invariant K and J inverters. This leads to the folded and transversal array coupling matrices from which a wide variety of filter topologies are derived. In this chapter, two advanced circuit sections are introduced, the extracted pole section and the trisection. These sections are capable of realizing one TZ each. Also, they can be cascaded with other circuit elements in the filter network. Application of these sections extends the range of filter topologies for realizing microwave filters.

For high-power applications, it is preferable to avoid negative cross-couplings. Such couplings are often realized as capacitive probes, susceptible to overheating under high-power conditions. Negative couplings are effectively removed by using the extracted pole principle, presented in this chapter. It allows removal of TZs (realized as bandstop cavities) before the synthesis of the remainder network. Since the TZs are extracted, the remaining network does not require cross-couplings, either straight or diagonal, unless there are further TZs that have not been extracted. The synthesis procedure based on the extracted pole section is described. Many examples are included to illustrate the procedure, and the role of the extracted pole section in the formation of a trisection is explained in depth.

Trisections have some properties that are advantageous for the synthesis of more advanced cross-coupled network topologies. A particular TZ can be associated with the cross-coupling in a particular trisection, and a trisection can be represented in a coupling matrix. The freedom to be able to synthesize trisections in more or less unrestricted positions in the filter network, and apply rotations to the coupling matrix to obtain a final topology, opens up some configuration possibilities that cannot be easily obtained by using the folded or transversal matrix as the starting point. The trisection is realized directly or as part of a cascade with other sections. It allows the synthesis of more complex networks, where a cascade of trisections is synthesized first and then further reconfiguration is carried out by using coupling matrix transforms. The power of this technique is demonstrated by applying it to transform a cascade of trisections to form cascaded quartet, quintet, and sextet filter topologies. Finally, the synthesis of the box section and its derivative, the extended box configuration, is explained. Examples are included to illustrate the intricacies of this synthesis procedure.

References

1 Rhodes, J.D. and Cameron, R.J. (1980) General extracted pole synthesis technique with applications to low-loss TE_{011} mode filters. *IEEE Transactions on Microwave Theory and Techniques*, **28**, 1018–1028.

2 Levy, R. (1976) Filters with single transmission zeros at real or imaginary frequencies. *IEEE Transactions on Microwave Theory and Techniques*, **24**, 172–181.

3 Amari, S., Rosenberg, U., and Bornemann, J. (2004) Singlets, cascaded singlets, and the nonresonating node model for advanced modular design of elliptic filters. *IEEE Microwave Wireless Components Letters*, **14**, 237–239.

4 Matthaei, G., Young, L., and Jones, E.M.T. (1980) *Microwave Filters, Impedance Matching Networks and Coupling Structures*, Artech House, Norwood, MA.

5 Rhodes, J.D. (1972) Waveguide bandstop elliptic filters. *IEEE Transactions on Microwave Theory and Techniques*, **20**, 715–718.

6 Cameron, R.J. (1982) General prototype network synthesis methods for microwave filters. *ESA Journal*, **6**, 193–206.

7 Amari, S. and Rosenberg, U. (2004) Direct synthesis of a new class of bandstop filters. *IEEE Transactions on Microwave Theory and Techniques*, **52**, 607–616.

8 Levy, R. and Petre, P. (2001) Design of CT and CQ filters using approximation and optimization. *IEEE Transactions on Microwave Theory and Techniques*, **49**, 2350–2356.

9 Tamiazzo, S. and Macchiarella, G. (2005) An analytical technique for the synthesis of cascaded *N*-tuplets cross-coupled resonators microwave filters using matrix rotations. *IEEE Transactions on Microwave Theory and Techniques*, **53**, 1693–1698.

10 Bell, H.C. (1982) Canonical asymmetric coupled-resonator filters. *IEEE Transactions on Microwave Theory and Techniques*, **30**, 1335–1340.

11 Stegun, I.A. and Abramowitz, M. (eds) (1970) *Handbook of Mathematical Functions*, Dover Publications, New York.

12 Levy, R. (1995) Direct synthesis of cascaded quadruplet (CQ) filters. *IEEE Transactions on Microwave Theory and Techniques*, **43**, 2939–2944.

13 Yildirim, N., Sen, O.A., Sen, Y. *et al.* (2002) A revision of cascade synthesis theory covering cross-coupled filters. *IEEE Transactions on Microwave Theory and Techniques*, **50**, 1536–1543.

14 Reeves, T. Van Stigt, N. and Rossiter, C. (2001) *A Method for the Direct Synthesis of General Sections*. IEEE MTT-S International Microwave Symposium Digest, Phoenix, pp. 1471–1474.

15 Reeves, T. and Van Stigt, N. (2002) *A Method for the Direct Synthesis of Cascaded Quintuplets*. IEEE MTT-S International Microwave Symposium Digest, Seattle, pp. 1441–1444.

16 Cameron, R.J., Harish, A.R., and Radcliffe, C.J. (2002) Synthesis of advanced microwave filters without diagonal cross-couplings. *IEEE Transactions on Microwave Theory and Techniques*, **50**, 2862–2872.

17 Cameron, R.J. and Rhodes, J.D. (1981) Asymmetric realizations for dual-mode bandpass filters. *IEEE Transactions on Microwave Theory and Techniques*, **29**, 51–58.

18 Seyfert, F. *et al*. (2002) *Design of Microwave Filters: Extracting Low Pass Coupling Parameters from Measured Scattering Data*, International Workshop Microwave Filters, Toulouse, France, June 24–26.

11

Microwave Resonators

Resonators are the basic building blocks of any bandpass filter. A resonator is an element that is capable of storing both frequency-dependent electric energy and frequency-dependent magnetic energy. A simple example is an *LC* resonator, where the magnetic energy is stored in the inductance *L* and the electric energy is stored in the capacitance *C*. The resonant frequency of a resonator is the frequency at which the energy stored in electric field equals the energy stored in the magnetic field. At microwave frequencies, resonators can take various shapes and forms. The shape of the microwave structure affects the field distribution and hence the stored electric and magnetic energies. Any microwave structure has an inherent resonant frequency determined by its physical characteristics and dimensions. In this chapter, we examine various resonator configurations highlighting their main applications and limitations. We also review the available techniques to calculate and measure the resonant frequency and unloaded *Q* factors of microwave resonators.

11.1 Microwave Resonator Configurations

The main design considerations of microwave resonators are the resonator size, unloaded *Q*, spurious performance, and power handling capability. The unloaded *Q* represents the inherent losses in the resonator. The higher the losses are, the lower is the *Q* value. It is therefore desirable to use resonators with high *Q* values since this reduces the insertion loss of the filter and improves its selectivity performance. In contrast to *LC* resonators, which have only one resonant frequency, microwave resonators can support an infinite number of electromagnetic (EM) field configurations or resonant modes. The spurious performance of a resonator is determined by how close the neighboring resonant modes are to the operating mode. The neighboring resonant modes act as spurious modes interfering with the filter performance. It is therefore desirable to increase the spurious-free window of the resonator in order to improve the filter out-of-band rejection performance.

The resonant modes in microwave resonators exist in the form of a single mode representing one electric resonator or in the form of degenerate modes (i.e., modes having the same resonance frequency with different field distributions). These degenerate modes allow the realization of two electric resonators within the same physical resonator (dual-mode resonators) or three electric resonators within the same physical resonator (triple-mode resonators). Examples of dual modes are TE_{11} modes, which exist in circular waveguide cavities [1–4]; HE_{11} modes, which exist in dielectric resonators [5, 6]; or TM_{11} modes, which exist in circular or square patch microstrip resonators [7]. Cubic waveguide cavities and cubic dielectric resonators can support triples modes [8]. The key advantage of operating in dual-mode or triple-mode configurations is size reduction. However, these modes do have an impact on the

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

Table 11.1 Comparison between various modes of operation.

Parameter	Single mode	Dual mode	Triple mode
Size	Large	Medium	Small
Spurious response	Good	Fair	Fair
Unloaded Q	High	Medium	Medium
Power handling capability	High	Medium	Medium
Design complexity	Low	Medium	High

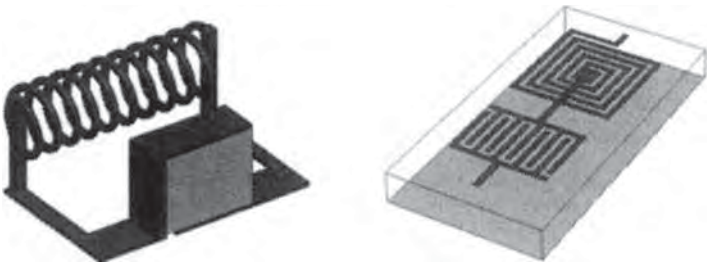


Figure 11.1 Lumped-element resonators realized by (a) a chip inductor and a chip capacitor and (b) spiral inductor and interdigital capacitor.

unloaded Q, spurious performance, and power handling capability of the cavity resonator. A summary of the features of each mode of operation is given in Table 11.1.

Microwave resonators are grouped into three categories: lumped-element LC resonators, planar resonators, and three-dimensional (3D) cavity-type resonators. Figure 11.1a shows a lumped-element resonator constructed by using a chip inductor and a chip capacitor. The lumped-element resonator can be printed on a dielectric substrate in the form of a spiral inductor and an interdigital capacitor, as shown in Figure 11.1b. They are very small in size and offer a wide spurious-free window; however, they have a relatively low Q value.

Examples of planar resonators are depicted in Figure 11.2. Planar resonators can take the form of a length of a microstrip transmission line, terminated in a short circuit or open circuit, or it can take the form of meander line, folded line, ring resonator, patch resonator, or any other configuration [9–11]. Any printed structure effectively acts as a resonator whose resonant frequency is determined by the resonator's dimensions, substrate dielectric constant, and

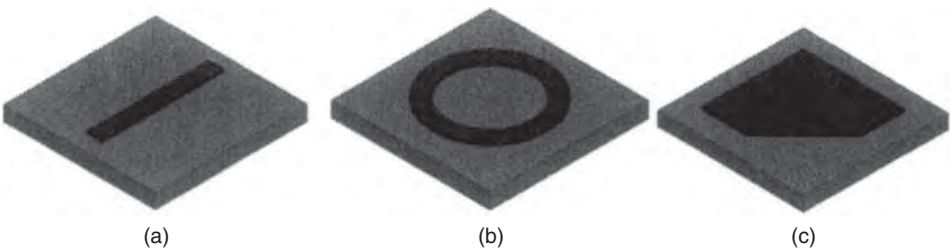


Figure 11.2 Examples of microstrip resonator configurations: (a) half-wavelength resonator; (b) ring resonator; (c) rectangular patch resonator.

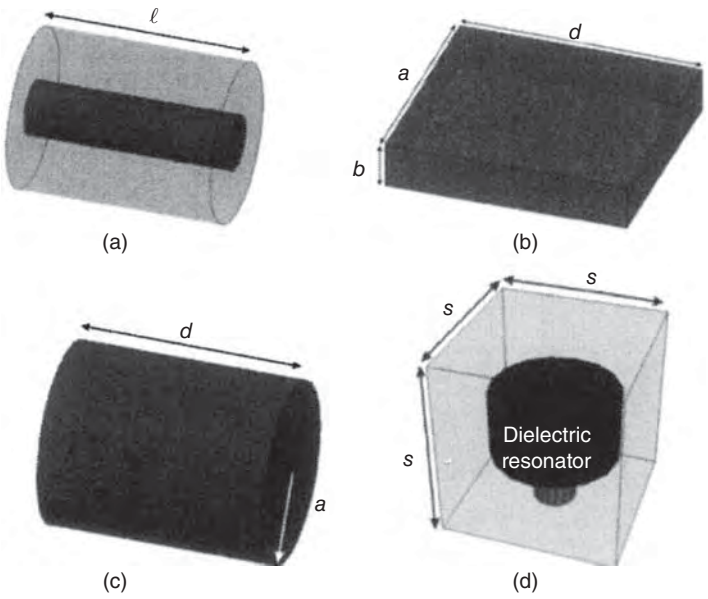


Figure 11.3 Examples of three-dimensional cavity resonators: (a) coaxial resonator; (b) rectangular waveguide resonator; (c) circular waveguide resonator; (d) dielectric resonator.

substrate height. Figure 11.3 illustrates examples of 3D coaxial, rectangular waveguide, circular waveguide, and dielectric resonators. The coaxial resonator consists of a length of a coaxial line, terminated in a short circuit at both ends. The waveguide resonators are rectangular and circular waveguides terminated in a short circuit at both ends. The dielectric resonator cavity consists of a dielectric resonator with a high dielectric constant mounted on a support structure with a low dielectric constant inside a below-cutoff metallic housing. 3D resonators are bulky in size; however, they offer very high Q values; in addition, they are capable of handling high RF power levels.

The selection of a resonator or filter configuration for a particular application involves several tradeoffs among filter insertion loss (i.e., resonator Q), filter size, cost, power handling requirements, etc. A comparison between these resonators is given in Figures 11.4 and 11.5. Typically, lumped-element filters are employed in low-frequency applications. The filters are suitable for integration in microwave monolithic integrated circuits (MMICs) or RFIC circuits. A typical Q value for lumped-element resonators is between 10 and 50 at 1 GHz. Planar resonators are usually employed in wideband, compact, and low-cost applications. The typical Q value for planar filters is in the range of 50–300 at 1 GHz. If planar filters are implemented in a superconductor [9], they can offer Q values ranging from 20,000 to 50,000 at 1 GHz. However, in this case the filters need to be cooled down to very low temperatures, below 90 K. Coaxial, waveguide, and dielectric resonators offer a Q value ranging from 3000 to 30,000 at 1 GHz. 3D resonators are widely employed to construct filters for low-loss wireless and space communication applications [10, 11]. The dielectric resonator technology, in particular, is emerging as the technology of choice for miniature high Q filters. Chapter 16 is dedicated to dielectric resonator filters.

Figures 11.1–11.3 illustrate only a few of the most common resonators; hundreds of other resonators with different configurations have been reported in the literature. The availability of commercial EM software simulators in the early 1990s has helped researchers in the field to devise and propose different planar and 3D resonator configurations. In particular, there have been tremendous innovations in planar resonator configurations. Simple and complex planar

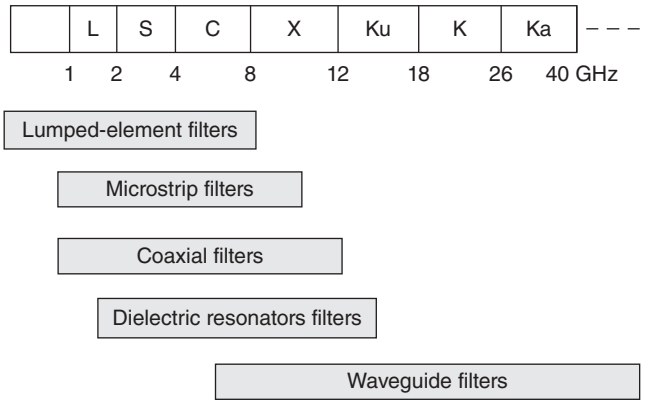


Figure 11.4 Application of the various resonator configurations.

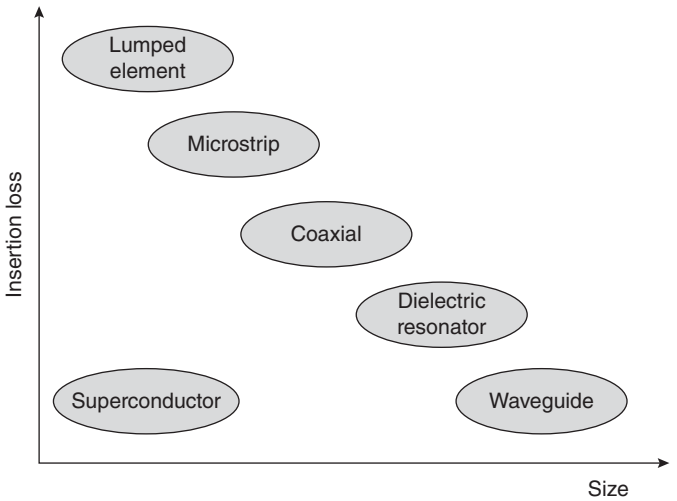


Figure 11.5 Typical relative size and insertion loss of various resonators.

resonator configurations can be handled by the same low-cost photolithographic fabrication process, enabling researchers to readily build prototype units to verify their designs.

11.2 Calculation of Resonant Frequency

11.2.1 Resonance Frequency of Conventional Transmission-Line Resonators

For resonators that are constructed from a section of waveguide, coaxial, or microstrip line, terminated in a short circuit or open circuit, the resonant frequency is calculated from the knowledge of the guide wavelength λ_g. Consider the transmission line shown in Figure 11.6, which is terminated in short circuit at both ends. By assuming a lossless structure, the resonant frequency is determined by standing at any point along the transmission line and calculating the impedance or the admittance seen from both sides. At resonance, we have

$$Z_1 + Z_2 = 0 \quad \text{or} \quad Y_1 + Y_2 = 0 \tag{11.1}$$

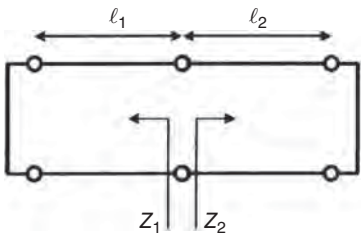


Figure 11.6 A transmission line terminated in short circuit at both ends.

Z_1 and Z_2 are the impedances seen at the input of transmission lines terminated in short circuit and are given by

$$Z_1 = j\tan\beta l_1 \quad \text{and} \quad Z_2 = j\tan\beta l_2 \tag{11.2}$$

Then, $Z_1 + Z_2 = 0$ gives $\tan\beta l_1 = -\tan\beta l_2$, which leads to

$$\beta l_1 = -\beta l_2 + q\pi, \quad \beta(l_1 + l_2) = \beta l = \frac{2\pi}{\lambda_g} l = q\pi, \quad q = 1, 2, 3, \ldots \tag{11.3}$$

$$l = q\frac{\lambda_g}{2}, \quad q = 1, 2, 3, \ldots \tag{11.4}$$

Using equation (11.4), we now consider calculating the resonant frequency of transmission-line coaxial, microstrip, and waveguide resonators:

1) *Coaxial line resonators* (Figure 11.3a). Coaxial lines are nondispersive media with λ_g given by

$$\lambda_g = \frac{c_0}{f\sqrt{\varepsilon_r}}, \quad f_{0q} = q\frac{c_0}{2l\sqrt{\varepsilon_r}} \tag{11.5}$$

where ε_r is the relative permittivity of the dielectric material that completely fills the space between the inner and outer connectors, c_0 is the free-space speed of light, $c_0 = 2.998 \times 10^8$ m/s.

2) *Microstrip line resonators* (Figure 11.2a). In microstrip lines, the EM fields are not completely confined in the dielectric. The propagation is described by defining an effective dielectric constant ε_{eff} [10]. The guide wavelength λ_g is then given by

$$\lambda_g = \frac{c_0}{f\sqrt{\varepsilon_{\text{eff}}}}, \quad f_{0q} = q\frac{c_0}{2l\sqrt{\varepsilon_{\text{eff}}}} \tag{11.6}$$

Microstrip lines are dispersive media, and ε_{eff} is frequency dependent. However, at low frequencies ε_{eff} can be approximated as [2]

$$\varepsilon_{\text{eff}} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\frac{1}{\sqrt{1 + 12h/W}} \tag{11.7}$$

where h is the substrate height and W is the microstrip width. More accurate expressions for ε_{eff} are found in [10, 11].

3) *Rectangular waveguide resonators* (Figure 11.3b). The propagation constant β in rectangular waveguides is given by [1]

$$\beta = \sqrt{\left(\frac{2\pi f}{c}\right)^2 - \left(\frac{n\pi}{a}\right)^2 - \left(\frac{m\pi}{b}\right)^2} \tag{11.8}$$

where $c = 1/\sqrt{\mu\varepsilon}$, applying equation (11.3), gives

$$f_{omnq} = \frac{c}{2\pi}\sqrt{\left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2 + \left(\frac{q\pi}{d}\right)^2} \quad \text{for } \text{TE}_{nmq} \text{ and } \text{TM}_{nmq} \text{ modes} \tag{11.9}$$

4) *Circular waveguide resonators* (Figure 11.3c). The propagation constants of TE and TM modes in circular waveguides are given by [1]

$$\beta = \sqrt{\left(\frac{2\pi f}{c}\right)^2 - \left(\frac{\rho'_{nm}}{a}\right)^2} \quad \text{for TE}_{nm} \text{ modes} \tag{11.10}$$

$$\beta = \sqrt{\left(\frac{2\pi f}{c}\right)^2 - \left(\frac{\rho_{nm}}{a}\right)^2} \quad \text{for TM}_{nm} \text{ modes} \tag{11.11}$$

where ρ_{nm} and ρ'_{nm} are mth roots of the Bessel functions $J_n(x)$ and $J'_n(x)$, respectively [1]. The resonant frequencies of TE_{nmq} and TM_{nmq} modes are given

$$f_{omnq} = \frac{c}{2\pi}\sqrt{\left(\frac{\rho'_{nm}}{a}\right)^2 + \left(\frac{q\pi}{d}\right)^2} \quad \text{for TE}_{nmq} \text{ modes} \tag{11.12}$$

$$f_{omnq} = \frac{c}{2\pi}\sqrt{\left(\frac{\rho_{nm}}{a}\right)^2 + \left(\frac{q\pi}{d}\right)^2} \quad \text{for TM}_{nmq} \text{ modes} \tag{11.13}$$

A mode chart for the circular cavities is depicted in Figure 11.7. The mode chart helps designers to select the proper resonator dimensions (a and d) to obtain the desired resonant frequency and to design for a wide spurious-free window.

Detailed analyses of coaxial, rectangular, and circular waveguides are well covered in several textbooks on microwaves [1, 2], providing information on the field distribution of the various modes. Filter designers need to be familiar with the field distribution of these modes to identify the optimum location for the input/output coupling as well to determine the optimum interresonator coupling configuration.

11.2.2 Resonance Frequency Calculation Using the Transverse Resonance Technique

Loading the transmission-line resonator with a capacitive/inductive element or with any type of discontinuity such as a screw alters the resonant frequency of the resonator. Figure 11.8 shows

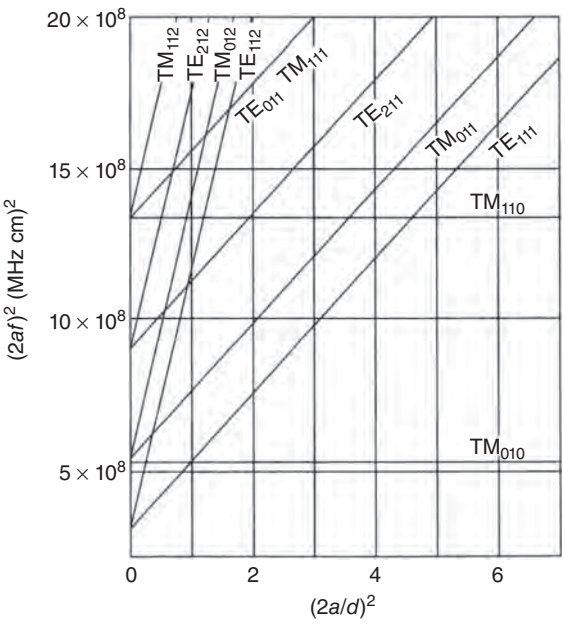


Figure 11.7 Mode chart for the resonant modes in circular waveguide resonators [1].

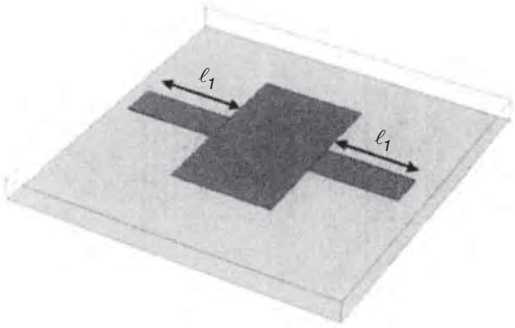


Figure 11.8 A microstrip transmission line loaded with a rectangular strip.

a microstrip transmission-line resonator loaded with a rectangular strip along the resonator's length. The rectangular strip represents a discontinuity in the transmission line. The field distribution in the resonator must change from that of the transmission line in order to satisfy the new boundary conditions caused by the discontinuity. This, in turn, changes the stored electric and magnetic energy, causing a shift in the resonant frequency. The new resonant frequency can be calculated with the use of the transverse resonance technique [12]. The discontinuity is represented by a two-port network, whose *S*-parameters can be calculated using a commercial EM simulator such as HFSS [13] or sonnet em [14], or any other software tool. The resonator can be represented as a two-port network connected to two open-circuited transmission lines of lengths l_1 and l_2 as shown in Figure 11.9. Γ_L is given by

$$\Gamma_L = e^{-2j\beta l_2} \tag{11.14}$$

With knowledge of the *S*-parameters of the discontinuity and in view of Section 5.1.5, we can express the input reflection coefficient Γ_{in} as follows:

$$\Gamma_{\text{in}} = \frac{V_1^-}{V_1^+} = S_{11} + \frac{S_{12}\Gamma_L S_{21}}{1 - \Gamma_L S_{22}} \tag{11.15a}$$

However, the incident and reflected voltages V_1^+ and V_1^- at the input of the two-port network are related as follows:

$$V_1^+ = e^{-2j\beta l_1} V_1^- \tag{11.15b}$$

Combining equations (11.15a) and (11.15b) gives

$$\left[1 - e^{-2j\beta l_1}\Gamma_{\text{in}}\right] V_1^- = 0 \tag{11.16}$$

Equation (11.16) is known as the "characteristic equation." At resonance, the coefficient of V_1^+ must be equal to zero. Then, the resonant frequency of the resonator is obtained by solving the following equation:

$$\left[1 - e^{-2j\beta l_1}\left(S_{11} + \frac{S_{12}e^{-2j\beta l_2}S_{21}}{1 - e^{-2j\beta l_2}S_{22}}\right)\right] = 0 \tag{11.17}$$

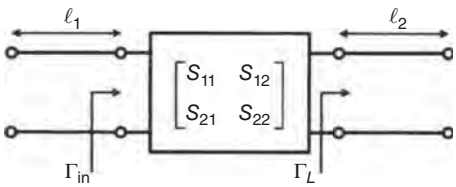


Figure 11.9 A two-port representation of the discontinuity shown in Figure 11.8.

Note that both the propagation constant β and the scattering parameters of the discontinuity are functions of frequency. For a first approximation, the S-parameters are calculated at the resonant frequency of the unperturbed resonator and are assumed to be frequency independent. Solving equation (11.17) provides β, which in turn gives an approximate value for the new resonant frequency. For a more accurate evaluation, equation (11.17) needs to be solved numerically. Alternatively, the adaptive frequency sampling technique, described in Chapter 15, can be used to find frequency-dependent polynomial expressions as an approximation for the scattering parameters.

11.2.3 Resonance Frequency of Arbitrarily Shaped Resonators

For resonators with arbitrarily shaped structures, an EM simulation tool is required for accurate calculation of the resonant frequency. This is achieved by using either an eigenmode analysis [15] or an S-parameter analysis. Both techniques are capable of calculating the resonant frequency of the fundamental resonant mode, as well as the resonant frequencies of the higher-order modes.

11.2.3.1 Eigenmode Analysis

Several EM packages such as HFSS [13] and CST [16] have eigenmode solution tools that are capable of analyzing any microwave resonator providing information on the resonant frequency, unloaded Q factor, and field distribution. The user specifies the frequency range, and the eigenmode tool identifies all the resonant modes that exist within this frequency range. Figure 11.10 shows the eigenmode solution of a top-loaded coaxial cavity obtained by HFSS [13]. The figure gives the first two resonant frequencies and the unloaded Q and the field distribution of the first resonant TEM mode.

The results show that the resonant frequency of the fundamental mode is 1.406 GHz, whereas that of the second mode is 6.537 GHz; thus, the resonator has a spurious-free window of 5.131 GHz.

	Resonant frequency (GHz)	Q factor
First eigenmode	1.406	3640
Second eigenmode	6.537	6300

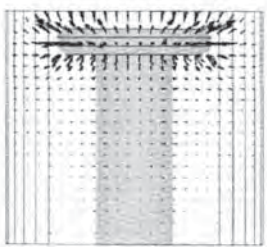


Figure 11.10 The field distribution of the first eigenmode solution of a top-loaded coaxial resonator.

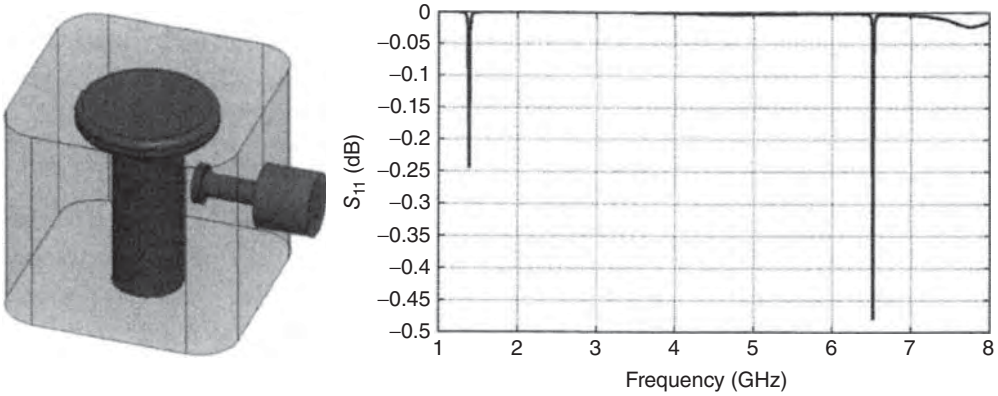


Figure 11.11 One-port *S*-parameter analysis of a top-loaded coaxial resonator.

11.2.3.2 *S*-Parameter Analysis

An alternative to the eigenmode analysis tool is to make use of an EM simulation tool to calculate the scattering parameters. Figure 11.11 shows the same resonator given in Figure 11.10 fed by a probe. The input reflection coefficient S_{11} seen at the probe versus frequency is also shown in Figure 11.11. For most of the frequency range, the RF energy is reflected back. The location of the dips indicates the frequencies where the RF energy is stored in the resonator, specifically the resonant frequencies of the various modes. The S_{11} response shows only the modes that the probe can excite inside the resonator. Thus, we need to have a rough idea about the field distribution of the resonant modes in order to ensure the proper probe excitation.

The width of the dips reflects the losses in the cavity. For lossless cavities this width is infinitesimally small, and the dips may not show up in simulation. Therefore, in using the one-port *S*-parameter analysis, we need to account for the losses. Having losses widens the dips, allowing them to be shown in the S_{11} response. It should also be mentioned that for an accurate calculation of the resonant frequency, we need to use very loose coupling between the probe and the resonator to minimize the loading on the resonator. This explains the slight shift in the resonant frequency usually obtained between the results of the eigenmode analysis and those of the one-port *S*-parameter analysis.

Alternatively, one can use a two-port analysis where the resonator is fed by an input/output port. However, the accuracy of the two-port analysis deteriorates when the input/output coupling is strong. Figure 11.12 compares the results of a microstrip resonator fed by one port and by two ports. It can be seen that the two-port analysis provides results that are slightly different from those of the one-port analysis, which is attributed to loading caused by the output probe.

The one-port *S*-parameter analysis is very useful in measuring the resonant frequency. A simple S_{11} measurement directly can provide information on the resonant frequencies of the resonator and its spurious performance.

Dual mode and triple mode of resonant frequencies can also be calculated by using either eigenmode analysis or the *S*-parameter analysis. The eigenmode solution would display almost identical resonant frequencies with two different field distributions, while in the one-port *S*-parameter analysis, the dual mode shows up in the S_{11} simulation as two dips very close to each other. Figure 11.13 illustrates S_{11} as simulated by HFSS for a dielectric resonator fed by an input probe. The first dip represents the resonant frequency of the single TE_{01} mode, while the two dips represent the resonant frequency of the hybrid HE dual mode [5]. As shown in

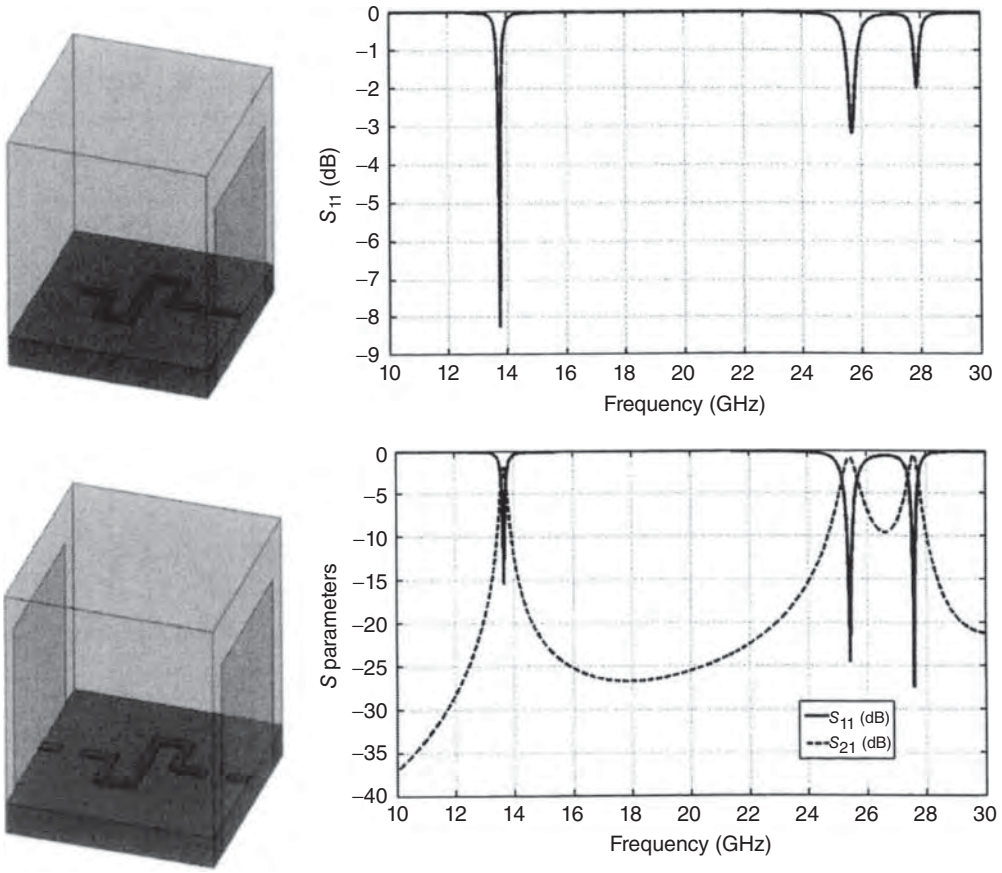


Figure 11.12 Comparison between the one-port *S*-parameter analysis and the two-port *S*-parameter analysis for a microstrip resonator as calculated by HFSS.

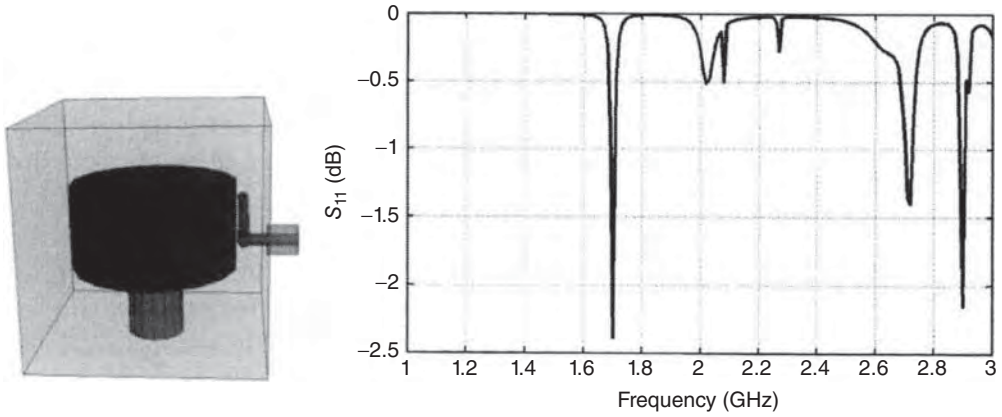


Figure 11.13 One-port analysis of a dielectric resonator illustrating the resonant frequencies of single and dual modes.

Chapter 14, the spacing between the two dips represents the coupling between the two dual modes. The coupling, in this case, is caused by the input probe.

11.3 Resonator Unloaded *Q* Factor

The unloaded Q_u factor of a resonator describes its frequency-selective properties, namely, the sharpness of the resonance curve. It is defined as

$$Q_u = \omega_0 \frac{W_T}{P_{\text{loss}}} \tag{11.18}$$

where ω_0 is the radian resonance frequency, W_T is the average energy stored, and P_{loss} is the power dissipated in the resonator. The energy stored is the sum of the stored electric energy W_e and stored magnetic energy W_m:

$$W_T = W_e + W_m = 2W_e = 2W_m \quad \text{(since } W_e = W_m \text{ at resonance)} \tag{11.19}$$

$$W_e = \frac{\varepsilon}{4}\int_v E \cdot E^* dv \tag{11.20}$$

$$W_m = \frac{\mu}{4}\int_v H \cdot H^* dv \tag{11.21}$$

The total power dissipated in the resonator is the sum of the power dissipated in resonator conducting walls P_c and the power dissipated in dielectric P_d, which can be written as

$$P_c = \frac{R_s}{2}\int_{\text{walls}} |H_t|^2 ds \tag{11.22}$$

$$P_d = \frac{\omega_0 \varepsilon''}{2}\int_v E \cdot E^* dv \tag{11.23}$$

Here, H_t is the tangential magnetic field at the surface; R_s is the surface resistance of the conducting walls, $R_s = \sqrt{\omega\mu_0/2\sigma}$, where σ is the conductivity of the resonator metallic walls; and ε'' is the imaginary part of the permittivity, which is related to the loss tangent $\tan\delta$ by

$$\varepsilon = \varepsilon' - j\varepsilon'' = \varepsilon_r\varepsilon_0(1 - j\tan\delta) \tag{11.24}$$

By factoring in the conducting and dielectric losses, we can express the unloaded Q_u as follows:

$$Q_u = \omega_0 \frac{W_T}{(P_c + P_d)} \tag{11.25}$$

By defining Q_c as $Q_c = \omega_0 W_T/P_c$ and Q_d as $Q_d = \omega_0 W_T/P_d$, we can further express the unloaded Q_u by

$$\frac{1}{Q_u} = \frac{1}{Q_c} + \frac{1}{Q_d} \tag{11.26}$$

In view of equations (11.20) and (11.23), Q_d is written as follows:

$$Q_d = \omega_0 \frac{2W_e}{P_d} = \frac{\varepsilon'}{\varepsilon''} = \frac{1}{\tan\delta} \tag{11.27}$$

From equations (11.26) and (11.27), it is concluded that for resonators that have dielectric loading, the overall unloaded Q will be less than $1/(\tan\delta)$.

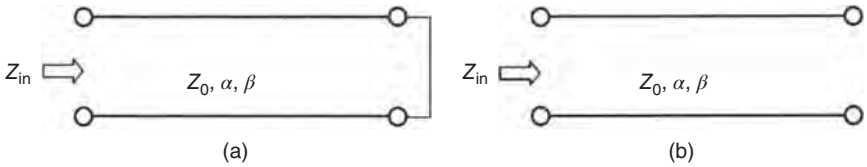


Figure 11.14 (a) A $\lambda/2$ short-circuit transmission line; (b) $\lambda/2$ open-circuit transmission line.

11.3.1 Unloaded *Q* Factor of Conventional Resonators

11.3.1.1 A Coaxial $\lambda/2$ Resonator Terminated in a Short Circuit

For coaxial resonators we adopt equations (11.18)–(11.23) and the field distribution inside the coaxial line [1, 2] to determine the exact analytical expression for Q_c. However, an approximate expression for Q_c can be obtained by considering the resonator as a transmission line terminated in a short circuit as shown in Figure 11.14a. The input impedance near resonance is approximated as a series *RLC* resonant circuit:

$$Z_{in} = Z_0 \tanh(\alpha l + j\beta l) = Z_0 \frac{\tanh \alpha l + j \tan \beta l}{1 + j \tan \beta l \tanh \alpha l} \tag{11.28}$$

Around resonance, $\omega = \omega_0 + \Delta\omega$, and $\beta l = \pi + \pi\Delta\omega/\omega_0$. For small losses we can approximate $\tanh \alpha l$ as $\tanh \alpha l \approx \alpha l$; Z_{in} can then be approximated by

$$Z_{in} \approx Z_0 \left(\alpha l + j \frac{\pi \Delta\omega}{\omega_0} \right)$$

which is in the form of the input impedance of a series *RLC* resonant circuit with *R*, *L*, and *C* given by [1, 2]

$$R = Z_0 \alpha l, \quad L = \frac{Z_0 \pi}{2\omega_0}, \quad C = \frac{1}{\omega_0^2 L} \tag{11.29}$$

where Z_0 is the characteristic impedance of the coaxial transmission line, $Z_0 = \eta/2\pi \ln b/a$, and α is the attenuation constant. The unloaded Q of the equivalent series resonant circuit is represented by

$$Q = \frac{\omega_0 L}{R} = \frac{\pi}{2\alpha l} = \frac{\beta}{2\alpha} \tag{11.30}$$

Thus Q_c in coaxial lines is thus approximated as

$$Q_c \approx \frac{\beta}{2\alpha_c} \tag{11.31}$$

Assuming a coaxial line of an inner radius a and an outer radius b, α_c is given by

$$\alpha_c = \frac{R_s}{2\eta \ln b/a} \left(\frac{1}{a} + \frac{1}{b} \right) \tag{11.32}$$

$$Q_c \approx \frac{ab\beta\eta \ln b/a}{(a+b)R_s} \tag{11.33}$$

η is the free-space impedance, $\eta = 376.7\ \Omega$. Note that the expression of Q_c given in equation (11.33) does not account for the losses in the short-circuit end plates.

11.3.1.2 A $\lambda/2$ Microstrip Resonator

Since microstrip lines are inhomogeneous structures, there are no accurate closed-form expressions for the field distribution in microstrip lines. Therefore, it is difficult to use equations (11.18)–(11.23) to obtain an exact analytical expression for Q_c in microstrip lines. The same concept applied for coaxial line resonators can also be applied for microstrip line resonators to obtain an approximate value for Q_c. By considering the resonator as $\lambda/2$ transmission line terminated in an open circuit (see, e.g., Figure 11.14b), the input impedance around the resonant frequency can be approximated as a parallel *RLC* resonant circuit. The input impedance Z_{in} is given by

$$Z_{\text{in}} = Z_0 \coth(\alpha l + j\beta l) = Z_0 \frac{1 + j \tan \beta l \tanh \alpha l}{\tanh \alpha l + j \tan \beta l} \tag{11.34}$$

For small losses and near resonance, we have $\omega = \omega_0 + \Delta\omega$, $\beta l = \pi + \pi\Delta\omega/\omega_0$, and $\tanh \alpha l \approx \alpha l$. We can then approximate Z_{in} as

$$Z_{\text{in}} \approx \frac{Z_0}{\alpha l + j\frac{\pi\Delta\omega}{\omega_0}}$$

which is in the form of the input of a parallel *RLC* resonant circuit with *R*, *L*, and *C* given by [1, 2].

$$R = \frac{Z_0}{\alpha l}, \quad C = \frac{\pi}{2\omega_0 Z_0}, \quad L = \frac{1}{\omega_0^2 C} \tag{11.35}$$

The unloaded *Q* of the equivalent parallel resonant circuit is given by

$$Q = \omega_0 RC = \frac{\pi}{2\alpha l} = \frac{\beta}{2\alpha} \tag{11.36}$$

Thus Q_c and Q_d of a $\lambda/2$ microstrip resonator can be approximated as

$$Q_c \approx \frac{\beta}{2\alpha_c}, \quad Q_d \approx \frac{\beta}{2\alpha_d} \tag{11.37}$$

There are no accurate closed-form expression for α_c and α_d in microstrip lines. They can be approximated as [2]

$$\alpha_c \approx \frac{R_s}{Z_0 W} \tag{11.38}$$

$$\alpha_d = \frac{k_0 \varepsilon_r (\varepsilon_{\text{eff}} - 1) \tan \delta}{2\sqrt{\varepsilon_{\text{eff}}}(\varepsilon_r - 1)} \tag{11.39}$$

R_s is the surface resistivity, $R_s = \sqrt{\omega\mu_0/2\sigma}$. This expression for Q_c does not factor in the effects of the fringing field and the losses of the metallic enclosure that houses the resonator.

Rectangular Waveguide Resonators In rectangular waveguide resonators (Figure 11.3b) there exist analytical expressions for the field distribution inside the cavity. Equations (11.18)–(11.23) can then be employed to derive an analytical expression for the unloaded Q_c. For a rectangular cavity of dimensions *a*, *b*, *d*, operating in TE_{10q} modes, Q_c is given by [1, 2]

$$Q_c = \omega_0 \frac{W_T}{P_c} = \frac{(k_{10q} ad)^3 b\eta}{2\pi^2 R_s} \frac{1}{[2q^2 a^3 b + 2bd^3 + q^2 a^3 d + ad^3]} \tag{11.40a}$$

where

$$k_{10q} = \left[\left(\frac{\pi}{a}\right)^2 + \left(\frac{q\pi}{b}\right)^2\right]^{1/2}, \quad \eta = \sqrt{\frac{\mu}{\varepsilon}} \tag{11.40b}$$

Circular Cavities Similarly, analytical expressions for the field distribution are employed in equations (11.18)–(11.23) to obtain an exact expression for Q_c. The unloaded Q_c of a circular cavity of radius a and length d, operating in TE_{nmq} modes, is given by [1]

$$Q_c = \omega_0 \frac{W_T}{P_c} = \frac{\lambda_0}{\delta_s} \frac{\left[1 - \left(\frac{n}{p'_{nm}}\right)^2\right]\left[(p'_{nm})^2 + \left(\frac{q\pi a}{d}\right)^2\right]^{3/2}}{2\pi\left[(p'_{nm})^2 + \frac{2a}{d}\left(\frac{q\pi a}{d}\right)^2 + \left(1 - \frac{2a}{d}\right)\left(\frac{nq\pi a}{p'_{nm}d}\right)^2\right]} \tag{11.41}$$

where δ_s is the skin depth and $\delta_s = 1/\sqrt{\pi f \mu_0 \sigma}$. Note that Q_c varies as λ_0/δ_s and thus the unloaded Q is inversely proportional to the square root of the frequency.

Figure 11.15 displays a plot of Q for various modes for circular waveguide cavities (Figure 11.3c) [1]. Note that mode TE_{011} has a Q value higher than that of the TE_{111} resonant mode. It is maximum around $d = 2a$. However, as seen from the mode chart given in Figure 11.7, the TE_{011} mode is degenerate with the TM_{111} mode; therefore, one needs to be careful in selecting the coupling scheme that does not excite the TM_{111} mode [17].

11.3.2 Unloaded *Q* of Arbitrarily Shaped Resonators

The eigenmode analysis is used to calculate both the resonant frequency and the unloaded Q of the resonator. With the eigenvalue (resonant frequency) determined, the eigenvector (field distribution) can be calculated, which in turn can be used to calculate the Q values. The process has been automated in most commercial EM software packages directly offering the frequency, unloaded Q, and field distribution. As shown in Figure 11.10, the eigenmode analysis for the top-loaded coaxial resonator also provides the unloaded Q values.

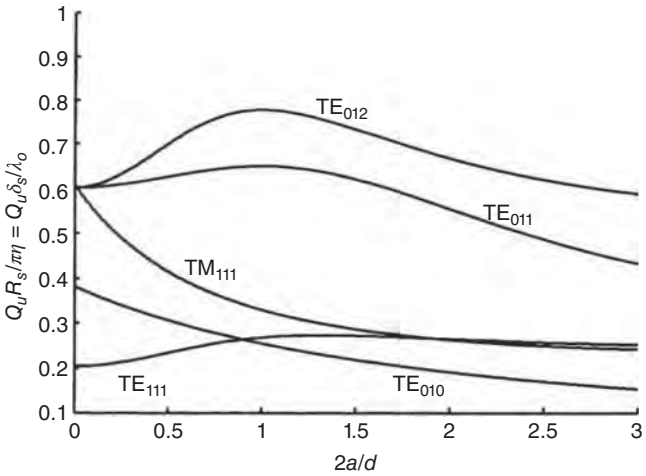


Figure 11.15 A plot of Q for various modes for circular waveguide cavities [1].

11.4 Measurement of Loaded and Unloaded *Q* Factor

An accurate calculation of the unloaded Q factor of microwave resonators requires knowledge of the conductivity of the metal walls as well as the dielectric loss tangent for resonators that are loaded with dielectric materials. Typically, high-conductivity metals such as gold, copper, and silver are used to achieve high Q values. Such metals are usually deposited or electroplated on metal surfaces or dielectric surfaces in the case of planar filters. The achievable conductivity can vary from sample to sample, depending on the deposition technique used and the surface roughness achieved.

In addition, there is a considerable uncertainty regarding the loss tangent value of dielectric materials. Often, the loss tangent of dielectric substrates and dielectric resonators is specified by the manufacturer at a particular frequency that may be different from the operating frequency. Most designers assume the product [frequency $\cdot$ $(1/\tan\delta)$] constant and scale the loss tangent according to the operating frequency. Although this scaling yields a good approximation, the scaling does not give the exact value of the loss tangent at the operating frequency. Moreover, like the dielectric constant ε_r, the loss tangent also varies from batch to batch. Furthermore, in actual filter implementations, resonators are usually loaded with tuning elements that have a significant impact on the Q value. Because of these factors, measurement is the only approach that can provide an accurate evaluation of the unloaded Q factor of resonators. The measurement is achieved through one- or two-port measurement [18–28]. In this chapter we address the one-port measurement, where reflection-type measurements are taken at an input port, coupled capacitively or inductively to the resonator.

To illustrate the concept, let us consider the circuit shown in Figure 11.16, which represents the equivalent circuit of a resonator, coupled to an input transmission line by a reactance X_s. The unloaded resonator has a resonant frequency of $f_0 = 1/2\pi\sqrt{LC}$ and an unloaded Q_0 of $Q_0 = \dfrac{2\pi f_0 C}{G_0}$. The unloaded resonator at port 2 is loaded by an external admittance Y_{ex} as depicted in Figure 11.17 and is given by

$$Y_{\text{ex}} = G_{\text{ex}} + jB_{\text{ex}} = \frac{1}{R_c + jX_s} \tag{11.42}$$

The effect of the loading is to change the resonant frequency to a new resonant frequency f_L due to the addition of B_{ex} and to reduce the Q factor to a loaded value of Q_L due to the addition of G_{ex}.

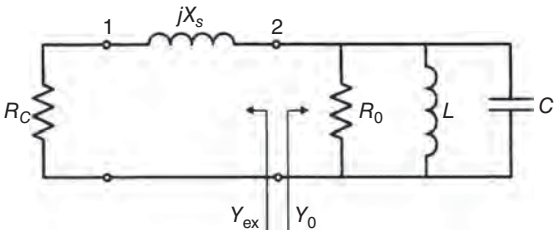


Figure 11.16 A resonator inductively coupled to an input port.

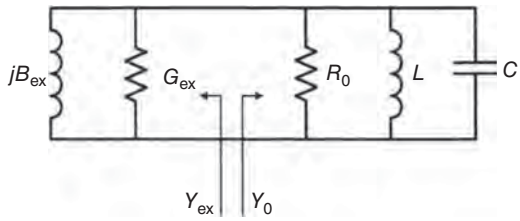


Figure 11.17 Resonator loading caused by input coupling.

Near resonance, the admittance $Y_L = (Y_0 + Y_{ex})$ of the loaded resonator can be approximated as

$$Y_L \approx G_{ex} + G_0 + j\left(2G_0Q_0\frac{f-f_0}{f_0} + B_{ex}\right) \tag{11.43}$$

The loaded resonant frequency f_L is then given by equating the imaginary part of Y_L to zero as follows:

$$f_L = f_0\left(1 - \frac{B_{ex}}{2Q_0G_0}\right) \tag{11.44}$$

The loaded $Q_L = \frac{2\pi f_0 C}{G_0+G_{ex}}$ is written in terms of Q_0 as

$$Q_L = Q_0\frac{G_0}{G_0+G_{ex}} = Q_0\frac{1}{1+\frac{G_{ex}}{G_0}} \tag{11.45}$$

$$Q_L = \frac{Q_0}{1+k} \tag{11.46}$$

where k is the coupling coefficient, representing the ratio of the power dissipated in the external load to the power dissipated in the resonator:

$$k = \frac{P_{loss-ex}}{P_{loss-res}} = \frac{V^2G_{ex}}{V^2G_0} = \frac{G_{ex}}{G_0} \tag{11.47}$$

When the amount of power dissipated in the external circuit equals to that dissipated in the resonator, the coupling is said to be critical and the coupling coefficient $k = 1$. Undercoupling ($k < 1$) corresponds to the case where the power dissipated in the external circuit is less than that dissipated in the resonator, whereas overcoupling ($k > 1$) corresponds to the case where the power dissipated in the external circuit is more than that dissipated in the resonator.

In view of equation (11.42), the coupling coefficient is expressed in terms of the coupling reactance element X_s as

$$k = \frac{\frac{R_0}{R_c}}{1+\frac{X_s^2}{R_c^2}} \tag{11.48}$$

The resonant frequency of the loaded resonator is written in terms of k, the unloaded Q_0 factor, and the equivalent circuit elements X_s and R_c as

$$f_L = f_0\left(1 + \frac{kX_s}{2Q_0R_c}\right) \tag{11.49}$$

The measurement of the reflection coefficient S_{11} versus the frequency at port 1 in Figure 11.16 provides the required information to calculate Q_L and k and Q_u. In view of Figure 11.16, the input reflection coefficient Γ at port 1 is written as

$$S_{11} = \Gamma = \frac{jX_s + \frac{1}{Y_0} - R_c}{jX_s + \frac{1}{Y_0} + R_c} \tag{11.50}$$

Kajfez [20] has shown that equation (11.50) can be reduced to

$$\Gamma = \Gamma_d + \frac{2k}{1+k}\frac{e^{-j2\delta}}{1+2jQ_L\frac{f-f_L}{f_0}} \tag{11.51}$$

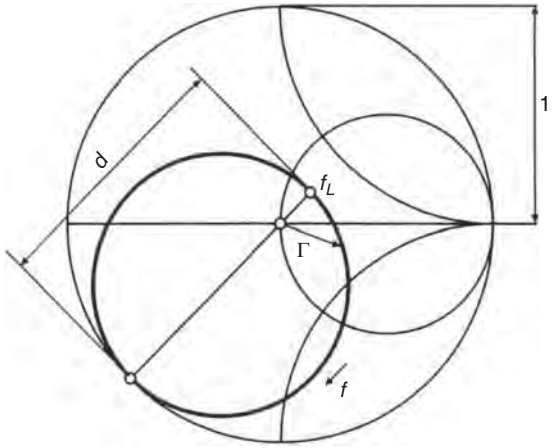


Figure 11.18 The *Q* circle on the Smith chart.

where

$$\Gamma_d = \frac{jX_s - R_c}{jX_s + R_c} = -\frac{Y_{\text{ex}}}{Y_{\text{ex}}^*} = -e^{-j2\delta}, \quad \delta = \tan^{-1}\frac{X_s}{R_c} \tag{11.52}$$

The vector $\Gamma - \Gamma_d$ describes a circle on the Smith chart, as shown in Figure 11.18. The circle is known as the *Q* circle [20]. For frequencies far from resonance (i.e., detuned $f > f_L$), the second term in equation (11.51) is small, and we can approximate the reflection coefficient as $\Gamma_{\text{untuned}} \approx \Gamma_d$. At resonance $f = f_L$, the reflection coefficient is given by

$$\Gamma_{\text{res}} = \Gamma_d + \frac{2k}{1+k}e^{-j2\delta} = \frac{k-1}{k+1}e^{-j2\delta} \tag{11.53}$$

The following observations are based on equations (11.51)–(11.53):

1) The magnitude of the reflection coefficient for $f > f_L$ is almost 1. The reflection coefficient corresponding to these frequencies lies on the circumference of the Smith chart.
2) The reflection coefficient at resonance has the angle $(-j2\delta)$. Depending on the k value, it is either in phase or 180° out of phase with Γ_{untuned} or Γ_d.
3) The reflection coefficient represents a circle on the Smith chart with Γ_{res} corresponding to the point on the circle with the shortest distance to the center of the Smith chart. As the frequency increases, δ also increases, and the input reflection coefficient moves along the *Q* circle in a clockwise direction.
4) The difference between $\Gamma_d - \Gamma_{\text{res}} = [2k/(1+k)]$, which represents the diameter d of the *Q* circle. Thus the diameter d is written as

$$d = \frac{2k}{1+k} \tag{11.54}$$

With measurement of the diameter d on the Smith chart, the coupling coefficient k is calculated as follows:

$$k = \frac{d}{2-d} \tag{11.55}$$

A glance at the *Q* circle on the Smith chart will reveal whether the coupling is under- or overcoupled. A small *Q* circle indicates undercoupling ($k < 1$); a large *Q* circle signifies overcoupling ($k > 1$). A *Q* circle of diameter 1 denotes critical coupling ($k = 1$).

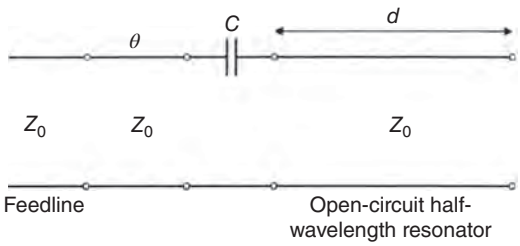


Figure 11.19 A $\lambda/2$ transmission-line resonator capacitively coupled to an input port.

For example, Figure 11.19 illustrates a $\lambda/2$ resonator capacitively coupled to an input port. The resonator is a $\lambda/2$ transmission line, terminated in an open circuit. The transmission line has a length of 53 mm and a dielectric constant of 2 resonating at 2 GHz. The transmission line is assumed to have an attenuation of 1 dB/m. A simple calculation shows that the structure is critically coupled ($k = 1$) when the capacitance C is 0.125 pF. The reflection coefficient is simulated using $\theta = 0$ with $C = 0.1$ pF (undercoupling), $C = 0.2$ pF (overcoupling), and $C = 0.125$ pF (critical coupling). The Q charts of these three cases are shown in Figure 11.20a–c, respectively.

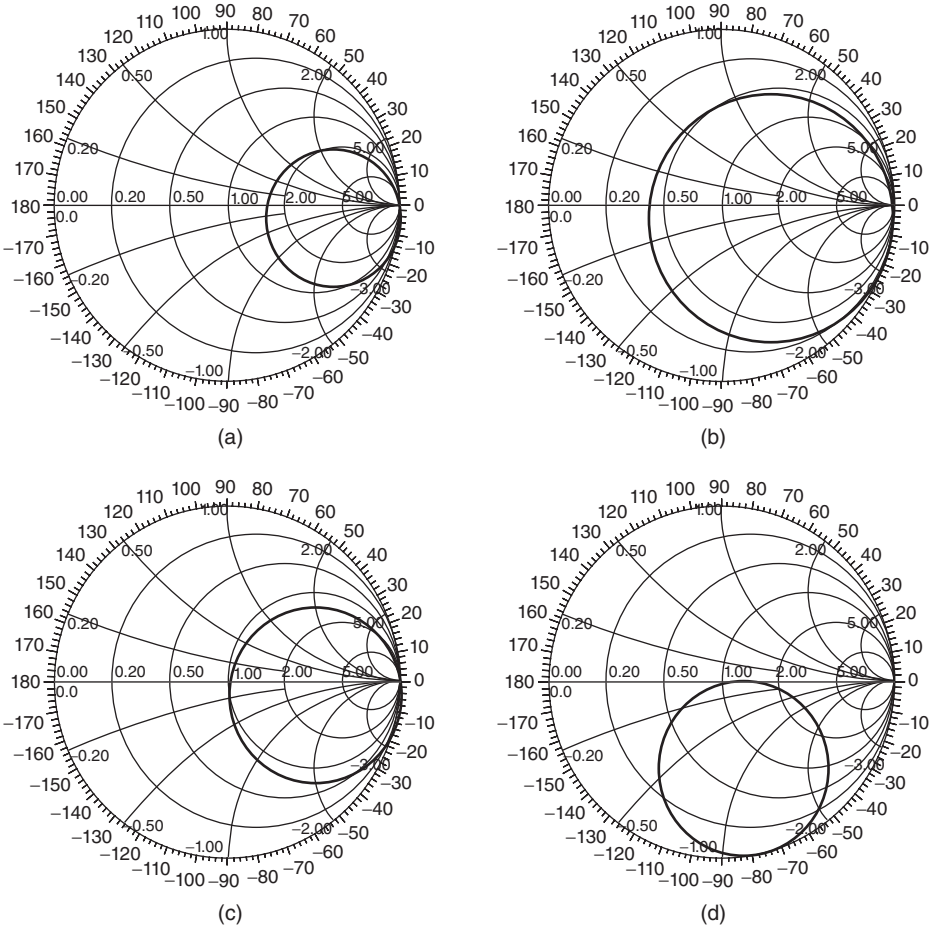


Figure 11.20 The Q circles for the circuit shown in Figure 11.19: (a) $\theta = 0$, undercoupled; (b) $\theta = 0$, overcoupled; (c) $\theta = 0$, critical coupled; (d) $\theta = 35$, critical coupled.

It is clear that the diameter of the Q circle of the reflection coefficient depends on the coupling coefficient k. It is also noted that the circles are shifted from the 0° line on the Smith chart by -2δ. The circuit is then simulated by assuming a capacitance of $C = 0.125$ pF (critical coupling) and $\theta = 35°$. The corresponding Q chart is shown in Figure 11.20d. The Q circle now is shifted from 0° by $-(2\delta + 2\theta)$. This case illustrates the fact that the Q circle can be located anywhere in the Smith chart depending on the reference plane, where measurements of the reflection coefficient are taken.

To determine the loaded Q factor, it is necessary to identify the resonant frequency of the loaded resonator f_L and the phase values of the reflection coefficient at two frequency points around the resonant frequency. The frequency on the Q circle that provides the shortest length for Γ (i.e., minimum reflection) is the resonant frequency. By using equation (11.51) and in view of Figure 11.21, we can readily show that by moving the marker to frequencies f_1 and f_2, respectively, below and above the resonant frequency f_L, the corresponding phases ϕ_1, and ϕ_2 become

$$\tan\phi_1 = -2Q_L\frac{f_1 - f_L}{f_0} \tag{11.56}$$

$$\tan\phi_2 = -2Q_L\frac{f_2 - f_L}{f_0} \tag{11.57}$$

Subtracting equations (11.56) and (11.57) gives

$$Q_L = \frac{1}{2}\frac{f_L}{f_2 - f_1}(\tan\phi_1 - \tan\phi_2) \tag{11.58}$$

If we select f_1 and f_2 such that $\phi_1 = 45°$ and $\phi_2 = -45°$, we get

$$Q_L = \frac{f_L}{f_2 - f_1} \tag{11.59}$$

Note that the angle ϕ cannot be read directly from the Smith chart display, but it can be calculated indirectly from the angle of the reflection coefficient. Therefore, for the polar display, it may be easier to use equation (11.58) rather than equation (11.59).

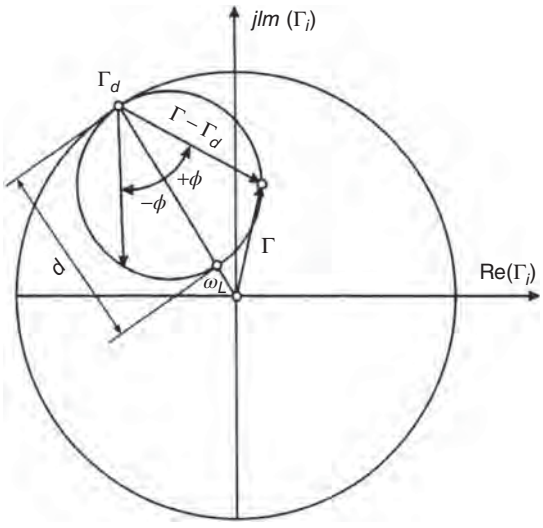


Figure 11.21 The reflection coefficient at two frequencies around f_L on the Q circle [20].

In view of the analysis aforementioned, the measurement of unloaded Q is achieved by using either the polar display (Smith chart) or the linear display (dB measurements) of the reflection coefficient. The steps are now summarized.

11.4.1 Measurement of Unloaded *Q* Using Polar Display of Reflection Coefficient

Step 1. Find f_L by identifying the frequency on the Q circle that corresponds to the minimum reflection coefficient.
Step 2. Calculate the coupling coefficient by measuring the diameter of the Q circle $[k = d/(2 - d)]$.
Step 3. Identify two frequencies f_1 and f_2 around f_L, and calculate Q_L by using either equation (11.58) or equation (11.59).
Step 4. Calculate Q_u by using $Q_u = Q_L(1 + k)$.

11.4.2 Measurement of Unloaded *Q* Using Linear Display of Reflection Coefficient

The reflection coefficient on the linear scale is depicted in Figure 11.22. In view of Figure 11.21, it can be easily shown that when $\phi = 45°$, we have

$$|\Gamma|^2 = \frac{1 + |\Gamma_{\text{res}}|^2}{2} \tag{11.60}$$

Step 1. Find f_L by locating the frequency marker at the dip of the reflection coefficient.
Step 2. With the marker located at the dip, read $S_{11}^{\min}$ in dB. Note that $S_{11}^{\min}$ is the reflection coefficient Γ_{res} given in equation (11.53):

$$S_{11}^{\min} = 20\log|\Gamma_{\text{res}}| = 20\log\left|\frac{k-1}{k+1}\right| \tag{11.61}$$

Thus, the coupling coefficient k can be calculated in terms of $S_{11}^{\min}$ as follows:

$$k = \frac{1 - 10^{S_{11}^{\min}/20}}{1 + 10^{S_{11}^{\min}/20}} \quad \text{if the resonator is undercoupled} \tag{11.62}$$

$$k = \frac{1 + 10^{S_{11}^{\min}/20}}{1 - 10^{S_{11}^{\min}/20}} \quad \text{if the resonator is overcoupled} \tag{11.63}$$

To determine whether the coupling is undercoupled or overcoupled, one needs to switch to the polar display to see the radius of the Q circle. Note that no calculation is done in the polar display. All the calculations are obtained by using the linear display.
Step 3. Determine S_{11}^{ϕ} in dB at $\phi = 45°$ using the formula

$$S_{11}^{\phi} = 10\log\frac{1 + 10^{S_{11}^{\min}/10}}{2} \tag{11.64}$$

Corresponding to $S_{11} = S_{11}^{\phi}$, determine the frequencies f_1 and f_2 as shown in Figure 11.22. Calculate Q_L as

$$Q_L = \frac{f_L}{f_2 - f_1} \tag{11.65}$$

Step 4. Calculate Q_u by using $Q_u = Q_L(1 + k)$.

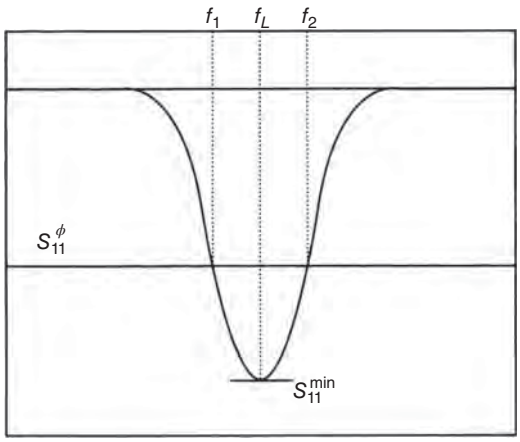


Figure 11.22 The reflection coefficient on a linear scale.

Summary

This chapter provided an overview of the theoretical and experimental techniques for evaluating the resonant frequency and unloaded Q factor of microwave resonators. Closed-form expressions are included for the resonant frequency and unloaded Q of standard rectangular and circular waveguide resonators. The use of the transverse resonance technique for the evaluation of the resonant frequency of transmission-line resonators is described by considering a microstrip resonator loaded with a capacitive discontinuity. The chapter also illustrated two approaches for calculating the resonant frequency of arbitrarily shaped resonators: the eigenmode analysis and the S-parameter analysis. Examples are given illustrating the implementation of these two techniques using EM-based commercial software tools such as HFSS.

The computation of the Q value is affected by the uncertainties in the properties of materials used in the construction of microwave filters. For example, in dielectric resonators filters, there is usually an uncertainty in the value of the loss tangent of the dielectric materials. Moreover, tuning screws can degrade the Q value by a large amount. For such resonator structures, Q measurement is perhaps the only approach that can provide reliable results. A step-by-step procedure is outlined for measuring the loaded and unloaded Q values using either the polar display of a vector network analyzer or the linear display of a scalar network analyzer.

References

1 Collin, R.E. (1966) *Foundation for Microwave Engineering*, McGraw-Hill, New York.
2 Pozar, D. (1998) *Microwave Engineering*, 2nd edn, Wiley, New York.
3 Atia, A.E. and Williams, A.E. (1972) Narrow bandpass waveguide filters. *IEEE Transactions on Microwave Theory and Techniques*, **20**, 238–265.
4 Atia, A.E. and Williams, A.E. (1971) New types of waveguide bandpass filters for satellite transponders. *COMSAT Technical Reviews*, **1** (1), 21–43.
5 Fiedziusko, S.J. (1982) Dual-mode dielectric resonator loaded cavity filter. *IEEE Transactions on Microwave Theory and Techniques*, **30**, 1311–1316.
6 Kajfez, D. and Guilon, P. (1986) *Dielectric Resonators*, Artech House, Norwood, MA.
7 Mansour, R.R. (1994) Design of superconductive multiplexers using single-mode and dual-mode filters. *IEEE Transactions on Microwave Theory and Techniques*, **42**, 1411–1418.

8 Walker, V. and Hunter, I.C. (2002) Design of triple mode TE_{01} resonator transmission filters. *IEEE Microwave and Wireless Components Letters*, **12**, 215–217.

9 Mansour, R.R. (1972) Microwave superconductivity. *IEEE Transactions on Microwave Theory and Techniques*, **50**, 750–759.

10 Hammerstad, E.O. and Jensen, O. (1980) Accurate models for microstrip computer-aided design. *IEEE-MTT-S*, 407–409.

11 Bahl, I. and Bhartia, P. (April 2003) *Microwave Solid State Circuit Design*, 2nd edn, Wiley, New York.

12 Itoh, T. (April 1989) *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*, Wiley, New York.

13 ANSYS HFSS, http://www.ansys.com/products/electronics/ansys-hfss.

14 Sonnet Software, http://www.sonnetsoftware.com.

15 Collin, R.E. (1960) *Field Theory of Guided Waves*, McGraw-Hill, New York.

16 CST Microwave Studio, www.cst.com.

17 Atia, A.E. and Williams, A.E. (1976) General TE01 mode waveguide bandpass filters. *IEEE Transactions on Microwave Theory and Techniques*, **24**, 640–648.

18 Ginzton, E.L. (1957) *Microwave Measurements*, McGraw-Hill, New York.

19 Sucher, M. and Fox, J. (eds) (1963) *Handbook of Microwave Measurements*, Polytechnic Press, New York.

20 Kajfez, D. (1994) Q *Factor*, Vector Forum, Oxford, MS.

21 Kajfez, D. and Hwan, E.J. (1984) *Q*-factor measurement with network analyzer. *IEEE Transactions on Microwave Theory and Techniques*, **32**, 666–670.

22 Wheless, W. P. and Kajfez, D. (1986) Microwave resonator circuit model from measured data fitting. IEEE MTT-S Symposium Digest, Baltimore, June 1986, pp. 681–684.

23 Asija, A. and Gundavajhala, A. (1994) Quick measurement of unloaded *Q* using a network analyzer. *RF Design*, **17** (11), 48–52.

24 Kajfez, D. (1995) *Q*-factor measurement with a scalar network analyzer. *IEE Proceedings Microwave, Antennas and Propagation*, **142**, 369–372.

25 Sanchez, M.C., Martin, E., and Zamarro, J.M. (1990) Unified and simplified treatment of techniques for characterizing transmission, reflection, or absorption resonators. *IEE Proceedings*, **137** (4), 209–212.

26 Miura, T., Takahashi, T., and Kobayashi, M. (1994) Accurate Q-factor evaluation by resonance curve area method and it's application to the cavity perturbation. *IEICE Transactions on Electronics*, **C** (6), 900–907.

27 Leong, K. and Mazierska, J. (2001) Accurate measurements of surface resistance of HTS films using a novel transmission mode *Q*-factor technique. *Journal of Superconductivity*, **14** (1), 93–103.

28 Leong, K. and Mazierska, J. (1998) Accounting for lossy coupling in measurements of Q-factors in transmission mode dielectric resonators. Proceedings of Asia Pacific Microwave Conference, APMC'98, Yokohama, December 8–11, pp. 209–212.

12

Waveguide and Coaxial Lowpass Filters

Synthesis methods presented in Chapters 7–9 are based on lossless lumped-element prototype filters. Such a model is suitable for realizing practical narrowband (typically < 2% bandwidth) filters at microwave frequencies. The bulk of filter applications in a communication system is for bandpass filters with bandwidths of less than 2%. The lumped-element model serves well to satisfy this need. However, communication systems also require lowpass filters (LPFs). The bandwidth requirement for LPFs is much higher (typically in the GHz range), and therefore prototype models based on lumped elements are not suitable for realization at microwave frequencies. It requires use of distributed elements for the prototype filters and, as a consequence, modification of the synthesis techniques.

In this chapter, we study the methods for synthesizing the transfer and reflection polynomials, suitable for two important classes of LPF: the distributed *stepped impedance* (SI) filter and the *mixed lumped/distributed* filter. The theory behind the synthesis and realization of LPFs is due mainly to Levy [1–5], who produced a series of papers in the 1960s and 1970s. He introduced a class of prototype polynomials that is suitable for representing the LPF transfer and reflection performance, and from which a lowpass prototype (LPP) circuit may then be synthesized. These include the all-pole Chebyshev function, the Chebyshev function of the second kind, and the Achieser–Zolotarev function [2, 6]. Once a circuit is synthesized from these functions, a microwave LPF can be realized in rectangular waveguide, coaxial transmission-line, or planar (TEM) technology, for example, stripline.

The following is a description of the types of prototype polynomial suitable for realization of LPFs in a waveguide, coaxial, or planar technology and some methods for their generation. Then we shall examine the commensurate-line elements that go into the LPP network.

Firstly the transforms that are used for the SI LPF will be studied, and the resultant transfer polynomials matched with those derived from the structure itself, which leads to the network synthesis method. A useful by-product of this synthesis method, the short-step impedance transformer, which has a filter-like transfer performance, will also be presented. The slightly more complex synthesis procedure for the lumped/distributed (L/D) LPF will then be detailed and the methods for dimensioning the structure outlined.

12.1 Commensurate-Line Building Elements

An essential component in the realization of microwave filters is the *commensurate-line element*. These are short-length transmission line, all of the same electrical length θ_c. With these elements, the distributed equivalents of lumped capacitors and inductors may be created.

Commensurate elements have the same sign and value as those of their lumped equivalents, but a frequency dependence that varies as $t = j \tan\theta$ instead of $s = j\omega$, where $\theta = \omega l / v_p =$

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

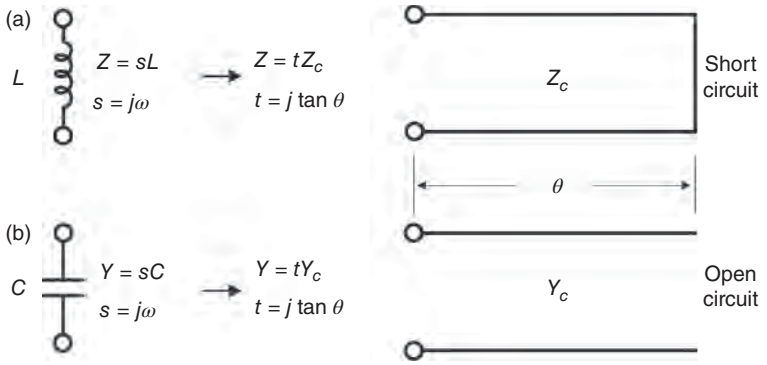


Figure 12.1 Commensurate-line equivalents for (a) lumped inductor and (b) lumped capacitor.

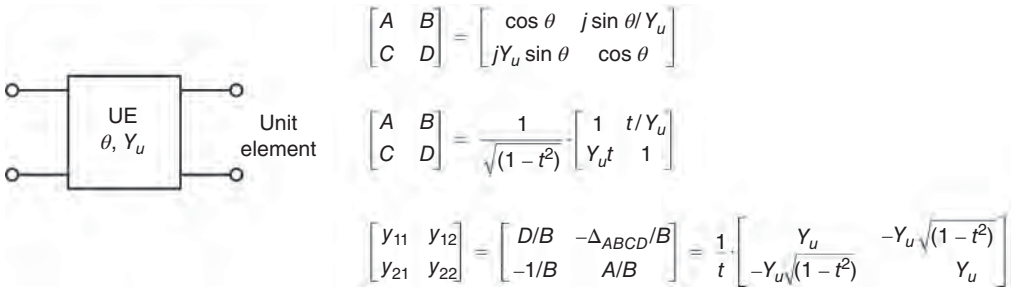


Figure 12.2 Unit element (UE) and associated [*A B C D*] transfer and admittance matrix.

$2\pi l/\lambda = \beta l$ is the electrical length of the element at frequency ω, l is the physical length of the commensurate line, and v_p is the velocity of propagation in the transmission-line medium. Assuming that v_p is constant at all frequencies, $\theta_0 = \omega_0 l/v_p$, where θ_0 is the electrical length at a reference frequency ω_0. This yields $\theta = (\omega/\omega_0)\theta_0$, the frequency variable θ, in terms of ω. $t = j\tan\theta$ is known as *Richard's transform* [7], and is used extensively as the frequency variable in commensurate-line networks.

The lumped components and their commensurate-line equivalents shown in Figure 12.1 have the same value of impedance or admittance at the frequency f_c at which the length of line is θ_c radians, but deviate from each other as the frequency moves away from f_c. Notably, the lumped element's reactance changes monotonically as the frequency moves toward infinity, whereas the distributed component's reactance is cyclic, repeating every π radians.

Another transmission-line component that is used extensively is the unit element (UE), which has no direct lumped-element equivalent. Figure 12.2 shows a UE of phase length θ and characteristic admittance Y_u and gives the associated [*A B C D*] transfer and admittance matrices in terms of $t = j\tan\theta$, obtained with the use of the identities $\cos\theta = 1/\sqrt{1+\tan^2\theta}$ and $\sin\theta = \tan\theta/\sqrt{1+\tan^2\theta}$.

These elements are also widely used for the synthesis and modeling of combline, interdigital bandpass filters, couplers, impedance transformers, and matching devices.

12.2 Lowpass Prototype Transfer Polynomials

For the synthesis of SI and lumped/distributed LPFs, the microwave structures can support only certain types of filter function.

***Stepped Impedance*:**

- All-pole functions (no transmission zeros)
- Even- or odd-degree functions
- Equiripple all-pole Chebyshev, the most common class although others can be used
- All-pole Zolotarev functions.

***Lumped/Distributed*:**

- Only odd-degree polynomials can be realized
- One pair of "half-zeros" necessary in the transfer polynomials
- Chebyshev functions of the second kind with a "half-zero" pair
- Zolotarev functions, also with a half-zero pair.

12.2.1 Chebyshev Polynomials of the Second Kind

The equiripple all-pole Chebyshev polynomial of the first kind is the regular polynomial that is used for the design of bandpass filters. The methods for the synthesis of these polynomials are detailed in Chapter 6. The transfer and reflection functions are given by

$$S_{21}(\omega) = \frac{1}{\varepsilon E(\omega)}, \quad S_{11}(\omega) = \frac{F(\omega)}{E(\omega)} \tag{12.1}$$

where $E(\omega)$ is an Nth-degree monic Chebyshev polynomial of the first kind. The lumped/distributed LPF functions require Chebyshev polynomials of the second kind, with one symmetric half-zero pair of transmission zeros. The transfer and reflection functions have the form

$$S_{21}(\omega) = \frac{\sqrt{\omega^2 - a^2}}{\varepsilon E(\omega)}, \quad S_{11}(\omega) = \frac{F(\omega)}{E(\omega)} \tag{12.2}$$

where $\pm a$ are the positions of the half-zeros and $E(\omega)$ is now an Nth-degree Chebyshev polynomial of the second kind. This is compared with the more familiar Chebyshev function of the first kind with one symmetric "full-zero" pair, for which the synthesis is established in Chapter 6, represented by

$$S_{21}(\omega) = \frac{(\omega^2 - a^2)}{\varepsilon E(\omega)}, \quad S_{11}(\omega) = \frac{F(\omega)}{E(\omega)} \tag{12.3}$$

Figure 12.3 compares the performances of the two kinds of Chebyshev function, showing that the transmission zeros of the second kind are "thinner" than the first kind, although the far-out-of-band rejection is better. In these seventh-degree cases, the positions of the full-zero position of the first kind is adjusted until the rejection lobe level is 50 dB. Then the positions of the half-zeros of the second kind are set in the same positions. It is evident that the lobe levels of the second kind are lower and close-to-band selectivity is less. The in-band return loss (RL) performance levels of both kinds are almost identical.

The Nth-degree numerator and denominator polynomials for the Chebyshev function of the second kind for up to N half-zero pairs is generated by a zero-mapping formula and recursive method, similar to that used in Chapter 6 for generation of the $F(\omega)$ polynomial of the first kind from the given $P(\omega)$ polynomial. Using the same symbols as previously, we obtain

$$C_N(\omega) = \frac{F(\omega)}{P(\omega)} = \cosh\left[\sum_{n=1}^{N} \cosh^{-1}(x_n)\right] \tag{12.4}$$

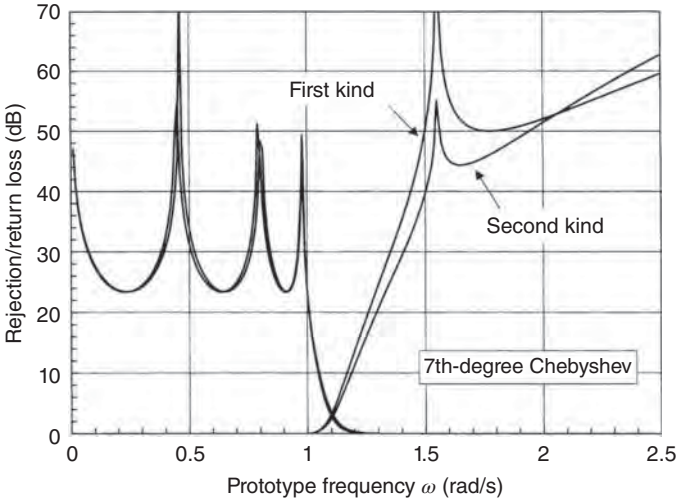


Figure 12.3 Comparison of Chebyshev functions of the first and second kinds, with a full-zero pair (first kind), and a half-zero pair (second kind).

where

$$x_n = \omega\sqrt{\frac{(1-1/\omega_n^2)}{(1-\omega^2/\omega_n^2)}}$$

The factors a_n and b_n are calculated by

$$a_n = x_n = \omega\sqrt{\frac{(1-1/\omega_n^2)}{(1-\omega^2/\omega_n^2)}}$$

$$b_n = \sqrt{x_n^2 - 1} = \frac{\omega'}{\sqrt{(1-\omega^2/\omega_n^2)}} \tag{12.5}$$

where $\omega' = \sqrt{\omega^2 - 1}$, the transformed frequency variable.

Finally, the C_N polynomial is formed

$$C_N(\omega) = \frac{\prod_{n=1}^{N}\left[\omega\sqrt{(1-1/\omega_n^2)} + \omega'\right] + \prod_{n=1}^{N}\left[\omega\sqrt{(1-1/\omega_n^2)} - \omega'\right]}{2\prod_{n=1}^{N}\sqrt{(1-\omega^2/\omega_n^2)}}$$

$$= \frac{\prod_{n=1}^{N}\left[c_n + d_n\right] + \prod_{n=1}^{N}\left[c_n - d_n\right]}{2\prod_{n=1}^{N}\sqrt{(1-\omega^2/\omega_n^2)}} \tag{12.6}$$

where $C_n = \omega\sqrt{(1-1/\omega_n^2)}$ and $d_n = \omega'$.

Using the same recursion method as that in Section 6.3, the $F(\omega)$ polynomial is formed, and, knowing the prescribed half-zero-position polynomial $P(\omega)$, one can determine the denominator polynomial $E(\omega)$.

The routine to generate the $F(\omega)$ polynomial is similar to the one in Chapter 6. Again, the complex array XP contains the prescribed half-zero positions, including those at $\omega = \infty$:

```
                X=1.0/XP(1)**2                  initialize with the first prescribed
                                                half-zero ω1
                Y=CDSQRT(1.0-X)
                U(l)=0.0
                U(2)=Y
                V(1)=1.0
                V(2)=0.0
C
                DO 10 K=3, N+1                  multiply by the second and
                                                subsequent prescribed half-zeros
                X=1.0 / XP(K-1)**2
                Y=CDSQRT(1.0-X)
                U2(K)=0.0
                V2(K)=0.0
                DO 11 J=1, K-1                  multiply by the constant terms
                U2 (J)=-V(J)
11              V2 (J)=U(J)
                DO 12 J=2, K                    multiply by the terms in ω
                U2(J)=U2(J)+Y * U(J-1)
12              V2(J)=V2(J)+Y * V(J-1)
                DO 13 J=3, K                    multiply by the terms in ω2
13              U2(J)=U2(J)+V(J-2)
                DO 14 J=1, K                    update Un and Vn
                U(J)=U2(J)
14              V(J)=V2(J)
10              CONTINUE                        reiterate for 3rd, 4th, … , Nth
                                                prescribed zero
```

As before, the array U now contains the coefficients of the $F(\omega)$ polynomial, and computing its roots yields the N singularities of the numerator of the reflection function $S_{11}(\omega)$. If it is required, $N - 1$ in-band reflection maxima can be found by determining the roots of polynomial V. The alternating-pole principle (see Section 6.3) cannot be used here directly to create the coefficients of $E(\omega)$ for odd numbers of half-zero pairs, because of the square root in the numerator of $S_{21}(\omega)$. However, it is shown in the section on lumped/distributed LPFs that after transformation to the t-plane, $S_{21}(t)$ will always have an even number of half-zeros. Thus, the square root sign in the numerator of $S_{21}(\omega)$ disappears, the $P(t)$ polynomial is formed, and the alternating principle can be used as before to find $E(t)$.

12.2.2 Achieser–Zolotarev Functions

Achieser–Zolotarev [2] functions are similar to the Chebyshev functions of the second kind in that they have an equiripple in-band amplitude characteristic. However, they possess an extra design parameter that allows the peak nearest to the origin to exceed the preset equiripple level. Figure 12.4 illustrates the in-band amplitude behavior of an even-degree (eighth) and an odd-degree (ninth) Achieser–Zolotarev function.

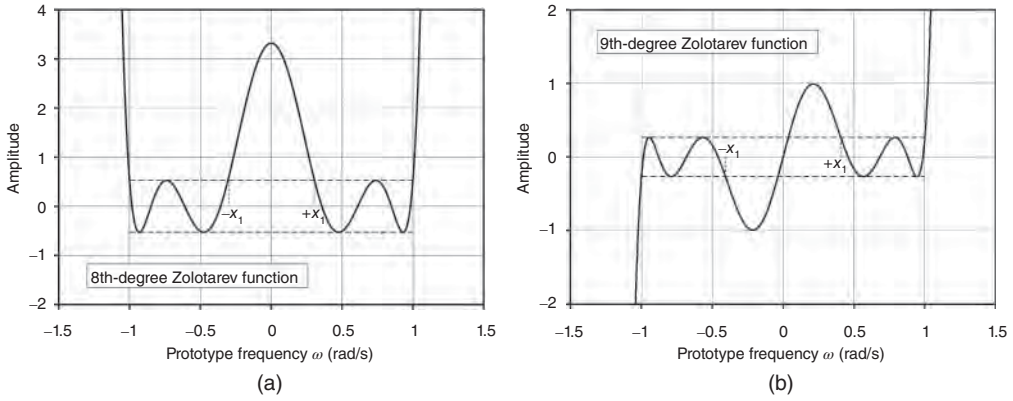


Figure 12.4 In-band (voltage) amplitude of (a) an eighth-degree function and (b) a ninth-degree Zolotarev function.

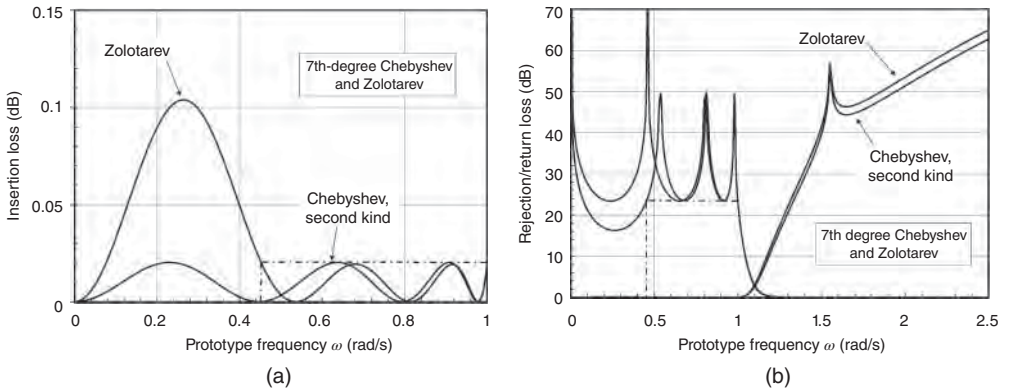


Figure 12.5 Comparison of seventh-degree Chebyshev and Zolotarev functions: (a) in-band amplitude and (b) out-of-band rejection.

Figure 12.5 compares the in-band and rejection performances of the seventh-degree Chebyshev, second kind, and Zolotarev functions. In Figure 12.5a, the higher level first ripple level of the Zolotarev function is clearly seen, whereas the out-of-band rejection of Figure 12.5b indicates that the rejection performances are quite similar, with the Zolotarev somewhat superior. However, it is not for this reason that the Zolotarev function is valuable in the design of LPFs. The function tends to yield better element values with less abrupt transitions and greater internal gap dimensions that help with high-power design.

12.2.2.1 Synthesis of Achieser–Zolotarev Functions

The Nth-degree (N even) all-pole Achieser–Zolotarev function (used for the SI LPF, but not for the lumped/distributed types) is easily generated from the zeros of the regular Chebyshev function of half degree $N_2 = N/2$ and prescribed RL level through the following mapping formula

$$s'_k = j\sqrt{\frac{1}{2}[(1+x_1^2) - js_k(1-x_1^2)]}$$

and

$$s'_{N+k} = s'^{*}_k, \quad k = 1,2,\ldots,N_2 \tag{12.7}$$

where s_k is the original position of the singularity in the complex s plane, s'_k is the transformed position, and x_1 ($|x_1| < 1$) is the frequency point in the band at which the equiripple behavior

starts (see Figure 12.4a). For an all-pole Chebyshev function of degree N_2 (N_2 even or odd) with a prescribed RL level (dB), the poles of $S_{21}(s)$ [the roots s_{pk} of $E(s)$] and the zeros of $S_{11}(s)$ [the roots s_{zk} of $F(s)$] are easily determined from [8, 9]:

$$s_{pk} = j\cosh(\eta + j\theta_k)$$
$$s_{zk} = j\cos\theta_k$$

where

$$\eta = \frac{1}{N_2}\ln\left(\varepsilon_1 + \sqrt{\varepsilon_1^2 + 1}\right), \quad \varepsilon_1 = \sqrt{10^{\mathrm{RL}/10} - 1}, \quad \text{and} \quad \theta_k = \frac{(2k-1)\pi}{2N_2},$$
$$k = 1,2,\ldots,N_2, \quad N_2 = N/2 \tag{12.8}$$

When $x_1 = 0$, the even-degree Zolotarev function degenerates to the pure all-pole Chebyshev function. These same even-degree Zolotarev functions are sometimes used for the design of dual-passband filter prototypes.

In contrast to the ease with which the polynomials of the even-degree Zolotarev function are determined, those for the odd-degree Zolotarev function, with or without half-zero pairs, are significantly more difficult to generate. The polynomial synthesis procedure is described as clearly as it can be by Levy [2], and tables of element values are also included in Ref. [6].

12.3 Synthesis and Realization of the Distributed Stepped Impedance Lowpass Filter

The prototype SI LPF is realized with a cascade of commensurate-line elements, all of the same electrical length θ_c, but with adjacent elements having different impedances in the direction of propagation. The objective of this section is to demonstrate that the polynomials that represent the transfer and reflection characteristics of such a cascade are in the same form as the polynomials that represent the design transfer and reflection characteristics, transformed by the ω-plane to θ-plane mapping function

$$\omega = \frac{\sin\theta}{\sin\theta_c} = a\sin\theta \tag{12.9}$$

where θ_c is the commensurate-line length and $a = 1/\sin\theta_c$.

The effect of applying this mapping function to the ω-plane transfer function is shown in Figure 12.6a, where, the transfer and reflection characteristics of an all-pole Zolotarev function are seen in the ω plane with its cutoff frequency at $\omega = \pm 1$ and RL in the equiripple region of 30 dB. Figure 12.6b exhibits the characteristics after mapping to the θ plane by the function equation (12.9) with a commensurate length $\theta_c = 20°$, that is, $a = 2.9238$.

As θ increases from zero, the corresponding frequency variable in the ω plane increases, reaching band edge at $\omega = +1$ when $\theta = \theta_c$ (20°). The range $\theta_c \le \theta \le 90°$ maps into the range $1 \le \omega \le a$. As θ increases beyond 90° toward 180°, ω retraces its path back to zero and then on to $-a$ as θ reaches 270°. Thus, the repeating pattern in Figure 12.6b is mapped from the portion of the ω-plane characteristic between $\omega = \pm a$.

12.3.1 Mapping the Transfer Function S_{21} from the ω Plane to the θ Plane

By using equation (12.9) (mapping function) and applying the identity $\sin\theta = \tan\theta/\sqrt{1+\tan^2\theta}$, we obtain

$$\omega = \frac{a\tan\theta}{\sqrt{1+\tan^2\theta}} \tag{12.10}$$

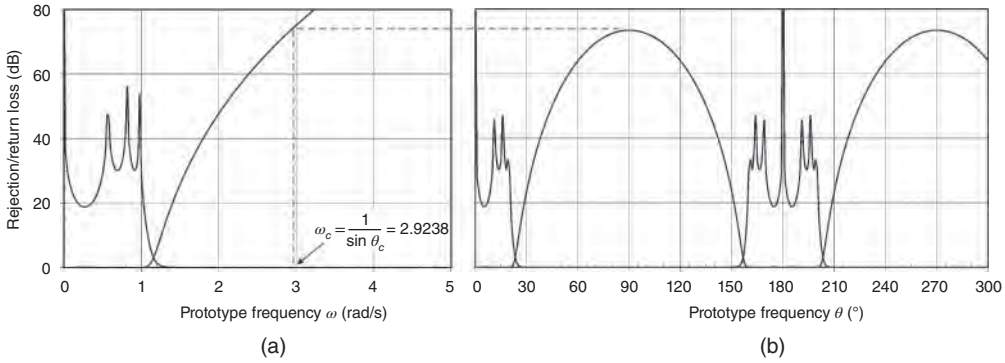


Figure 12.6 Stepped impedance LPF mapping of an all-pole transfer and reflection function from the ω plane to the θ plane through the mapping function $\omega = a \sin\theta$.

Putting $t = j \tan\theta$, it follows that

$$s = j\omega = \frac{at}{\sqrt{1-t^2}} \tag{12.11a}$$

or

$$t = \frac{\pm s}{\sqrt{a^2+s^2}} = \frac{\pm s \sin\theta_c}{\sqrt{1+(s\sin\theta_c)^2}} \tag{12.11b}$$

The s-plane all-pole transfer function has the form $S_{21}(s) = 1/(\varepsilon E(s))$, where $E(s)$ is the Nth-degree denominator polynomial of the all-pole Chebyshev [equation (12.8)] or Zolotarev transfer function. By changing the frequency variable from s to t using equation (12.11a), we obtain

$$S_{21}(t) = \frac{1}{\varepsilon E\left(\dfrac{at}{\sqrt{1-t^2}}\right)} = \frac{\left[\sqrt{1-t^2}\right]^N}{\varepsilon' E'(t)} \tag{12.12}$$

where $E'(t)$ is another Nth-degree polynomial in variable $t = j\tan\theta$. The most accurate method for determining $E'(t)$ is to transform the N s-plane singularities of $E(s)$ to the t plane by using equation (12.11b) and then factorizing them to form the new polynomial in t. Then, normalizing constant ε' is found by evaluating it at $\theta = \theta_c$ (corresponding to $\omega = 1$), where the RL level is known. Note that the N transmission half-zeros of $S_{21}(t)$ occur at $t = \pm 1$, equivalent to $s = \pm j\infty$ in the s plane.

12.3.1.1 Form of $S_{21}(t)$ for the Circuit Approach

As mentioned, the Nth-degree SI LPF consists of a cascade of N commensurate-line lengths with different characteristic impedances Z_i, $i = 1,2,\ldots,N$. One such line length is shown in Figure 12.7a, and its $[A\,B\,C\,D]$ transfer matrix is given by

$$[A\,B\,C\,D] = \begin{bmatrix} \cos\theta & jZ_i\sin\theta \\ \dfrac{j\sin\theta}{Z_i} & \cos\theta \end{bmatrix} = \frac{1}{\sqrt{1-t^2}} \begin{bmatrix} 1 & Z_i t \\ \dfrac{t}{Z_i} & 1 \end{bmatrix} \tag{12.13}$$

since $t = j\tan\theta$, $\cos\theta = 1/\sqrt{1-t^2}$, and $j\sin\theta = t/\sqrt{1-t^2}$.

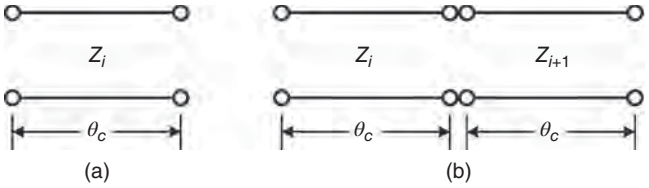


Figure 12.7 Commensurate-line lengths (UEs): (a) single UE and (b) two UEs in cascade.

Now, by cascading two UEs of impedance Z_i and Z_{i+1} (Figure 12.7b), we obtain

$$[A\,B\,C\,D] = \left[\frac{1}{\sqrt{1-t^2}}\right] \cdot \begin{bmatrix} 1 + t^2\dfrac{Z_i}{Z_{i+1}} & t(Z_i + Z_{i+1}) \\ t\left(\dfrac{1}{Z_i} + \dfrac{1}{Z_{i+1}}\right) & 1 + t^2\dfrac{Z_{i+1}}{Z_i} \end{bmatrix}$$

$$= \left[\frac{1}{\sqrt{1-t^2}}\right]^2 \cdot \begin{bmatrix} A_2(t) & B_1(t) \\ C_1(t) & D_2(t) \end{bmatrix} \tag{12.14}$$

For N lines

$$[A\,B\,C\,D] = \left[\frac{1}{\sqrt{1-t^2}}\right]^N \cdot \begin{bmatrix} A_N(t) & B_{N-1}(t) \\ C_{N-1}(t) & D_N(t) \end{bmatrix} \quad (N \text{ even})$$

$$= \left[\frac{1}{\sqrt{1-t^2}}\right]^N \cdot \begin{bmatrix} A_{N-1}(t) & B_N(t) \\ C_N(t) & D_{N-1}(t) \end{bmatrix} \quad (N \text{ odd}) \tag{12.15}$$

where the subscripts indicate the degree of the polynomials in the variable t. The transfer function associated with this $[A\,B\,C\,D]$ matrix is derived from

$$S_{21}(t) = \frac{2[1-t^2]^{N/2}}{A(t) + B(t) + C(t) + D(t)} = \frac{[1-t^2]^{N/2}}{\varepsilon_t E'(t)} \tag{12.16a}$$

$$= \frac{[(1-t)(1+t)]^{N/2}}{\varepsilon_t E'(t)} = \frac{\sqrt{P(t)}/\varepsilon_t}{E'(t)} \tag{12.16b}$$

where $P(t)$ is the Nth-degree numerator polynomial of $S_{21}(t)$ and constants are consolidated into ε_t. It is obvious at this stage that equation (12.16a) is in the same form as equation (12.12), derived by directly transforming the all-pole Chebyshev or Zolotarev function to the t plane with the mapping function (12.11). Note that the numerator polynomial $P(t)$ is of degree N, with N transmission half-zeros at $t = \pm 1$, which means that the function is fully canonical. The corresponding reflection function is $S_{11}(t) = F(t)/(\varepsilon_{Rt} E'(t))$, and because the conservation of energy condition must still be satisfied, ε_{Rt} is nonunity (although very close to unity for the higher degrees).

Although it is not possible to form the polynomial $\sqrt{P(t)}$ for odd-degree functions, it is actually not needed for the synthesis procedure. ε_t is evaluated at a convenient frequency point in the filter function where $S_{21}(t)$ and $S_{11}(t)$ are known, such as at the cutoff frequency, where $\theta = \theta_c$, the cutoff angle. Then

$$\varepsilon_t = \left.\frac{\left[1-t^2\right]^{N/2}}{\left(\sqrt{1-10^{-\mathrm{RL}/10}}\right) \cdot |E'(t)|}\right|_{t=t_c} \quad \text{and} \quad \varepsilon_{Rt} = \frac{\varepsilon_t}{\sqrt{\varepsilon_t^2 - 1}} \tag{12.17}$$

where $t_c = j \tan \theta_c$.

12.3.2 Synthesis of the Stepped Impedance Lowpass Prototype Circuit

The synthesis of the SI LPF cascade follows lines very similar to those used for the LPP networks for bandpass filters, with a knowledge of the *form* of the $A(t)$, $B(t)$, $C(t)$, and $D(t)$ polynomials [see equation (12.15)]. As a symmetric characteristic about $t = 0$, the $F(t)$ polynomial is even or odd, depending on whether N is even or odd, and both $E'(t)$ and $F(t)$ have pure real coefficients.

The impedance Z_{in}, looking in at the input of the circuit, is

$$Z_{\text{in}} = \frac{A(t)Z_L + B(t)}{C(t)Z_L + D(t)} = \frac{1 + S_{11}(t)}{1 - S_{11}(t)} = \frac{E'(t) + F(t)/\varepsilon_{Rt}}{E'(t) - F(t)/\varepsilon_{Rt}} \tag{12.18}$$

where Z_L is the load impedance terminating the output of the network, and it is assumed that the source impedance $Z_S = 1$. In the synthesis procedure, we shall be taking ratios such as $A(t)/C(t)$ and $B(t)/D(t)$ to evaluate the Z_i, the characteristic impedances of the UE cascade. Therefore, it is unimportant whether or not Z_L is included in equation (12.18). For reasons of clarity, we omit Z_L for the synthesis procedure described next.

Knowing that $A(t)$ and $D(t)$ are even polynomials, and $B(t)$ and $C(t)$ are odd for all N [see equation (12.15)], the $A(t)$, $B(t)$, $C(t)$, and $D(t)$ polynomials are constructed from the coefficients of $E'(t)$ and $F(t)/\varepsilon_{Rt}$ polynomials as follows:

$$\begin{aligned}
A(t) &= (e_0 + f_0) + (e_2 + f_2)t^2 + (e_4 + f_4)t^4 + \cdots \\
B(t) &= (e_1 + f_1)t + (e_3 + f_3)t^3 + (e_5 + f_5)t^5 + \cdots \\
C(t) &= (e_1 - f_1)t + (e_3 - f_3)t^3 + (e_5 - f_5)t^5 + \cdots \\
D(t) &= (e_0 - f_0) + (e_2 - f_2)t^2 + (e_4 - f_4)t^4 + \cdots
\end{aligned} \tag{12.19}$$

where e_i and f_i, $i = 0, 1, 2, \ldots, N$ are the coefficients of $E'(t)$ and $F(t)/\varepsilon_{Rt}$ polynomials, respectively. Writing the polynomials in this way ensures that the form is correct and that the transfer and reflection functions are satisfied [see equation (12.16a)]:

$$S_{12}(t) = S_{21}(t) = \frac{\sqrt{P(t)}/\varepsilon_t}{E'(t)} = \frac{2[1 - t^2]^{N/2}/\varepsilon_t}{A(t) + B(t) + C(t) + D(t)} \tag{12.20a}$$

$$S_{11}(t) = S_{22}(t) = \frac{F(t)/\varepsilon_{Rt}}{E'(t)} = \frac{A(t) + B(t) - C(t) - D(t)}{A(t) + B(t) + C(t) + D(t)} \tag{12.20b}$$

12.3.2.1 Network Synthesis

Following the methods described in Section 7.2.1 to extract inverters, UEs are removed in sequence from the $[A\,B\,C\,D]$ matrix, leaving a remainder matrix after each extraction. First, the overall $[A\,B\,C\,D]$ matrix is decomposed into the first UE and a remainder matrix expressed as

$$\begin{aligned}
\frac{\varepsilon_t}{[1 - t^2]^{N/2}} \cdot \begin{bmatrix} A(t) & B(t) \\ C(t) & D(t) \end{bmatrix} &= \frac{1}{[1 - t^2]^{1/2}} \cdot \begin{bmatrix} 1 & Z_1 t \\ t/Z_1 & 1 \end{bmatrix} \frac{\varepsilon_t}{[1 - t^2]^{(N-1)/2}} \begin{bmatrix} A_{\text{rem}}(t) & B_{\text{rem}}(t) \\ C_{\text{rem}}(t) & D_{\text{rem}}(t) \end{bmatrix} \\
&= \frac{\varepsilon_t}{[1 - t^2]^{N/2}} \begin{bmatrix} A_{\text{rem}}(t) + tZ_1 C_{\text{rem}}(t) & B_{\text{rem}}(t) + tZ_1 D_{\text{rem}}(t) \\ C_{\text{rem}}(t) + \dfrac{tA_{\text{rem}}(t)}{Z_1} & D_{\text{rem}}(t) + \dfrac{tB_{\text{rem}}(t)}{Z_1} \end{bmatrix}
\end{aligned} \tag{12.21}$$

The open-circuit impedance of this circuit may be calculated in two ways:

$$z_{11} = \left.\frac{A(t)}{C(t)}\right|_{t=1} = \left.\frac{A_{\text{rem}}(t) + tZ_1 C_{\text{rem}}(t)}{C_{\text{rem}}(t) + tA_{\text{rem}}(t)/Z_1}\right|_{t=1} = Z_1 \tag{12.22a}$$

$$z_{11} = \left.\frac{B(t)}{D(t)}\right|_{t=1} = \left.\frac{B_{\text{rem}}(t) + tZ_1 D_{\text{rem}}(t)}{D_{\text{rem}}(t) + tB_{\text{rem}}(t)/Z_1}\right|_{t=1} = Z_1 \tag{12.22b}$$

Having evaluated Z_1 from either of the formulas in equation (12.22), the first UE is extracted from the initial $[A\,B\,C\,D]$ matrix, leaving the remainder matrix $[A\,B\,C\,D]_{\text{rem}}$

$$\frac{\varepsilon_t}{[1-t^2]^{(N-1)/2}}\begin{bmatrix}A_{\text{rem}}(t) & B_{\text{rem}}(t)\\ C_{\text{rem}}(t) & D_{\text{rem}}(t)\end{bmatrix} = \frac{1}{[1-t^2]^{1/2}}\cdot\begin{bmatrix}1 & -Z_1 t\\ \dfrac{-t}{Z_1} & 1\end{bmatrix}\cdot\frac{\varepsilon_t}{[1-t^2]^{N/2}}\cdot\begin{bmatrix}A(t) & B(t)\\ C(t) & D(t)\end{bmatrix}$$

$$= \frac{\varepsilon_t}{[1-t^2]^{(N+1)/2}}\begin{bmatrix}A(t)-tZ_1C(t) & B(t)-tZ_1D(t)\\ C(t)-\dfrac{tA(t)}{Z_1} & D(t)-\dfrac{tB(t)}{Z_1}\end{bmatrix} \tag{12.23}$$

Finally, the right-hand matrix in equation (12.23) must be divided top and bottom by $(1-t^2)$ to reduce the denominator term to the same power as that of the remainder matrix in the left-hand side (LHS) of equation (12.21). After the extraction of Z_1, the numerator polynomials will be exactly divisible by $(1-t^2)$, leaving the remainder polynomials $A_{\text{rem}}(t)$, $B_{\text{rem}}(t)$, $C_{\text{rem}}(t)$, and $D_{\text{rem}}(t)$. Polynomials $A_{\text{rem}}(t)$ and $D_{\text{rem}}(t)$ will decrease in degree by two if the overall degree of the network has changed from even to odd as a consequence of the extraction of the UE Z_1, or $B_{\text{rem}}(t)$ and $C_{\text{rem}}(t)$ will decrease if the change is from odd to even (see equation (12.15)). The process is repeated on the remainder matrix to extract the second UE Z_2 and so on until all N of the UEs are extracted.

The load termination is calculated by recognizing that at zero frequency, the cascade of UEs is effectively transparent, and $Z_{\text{in}} = Z_L$. Thus, we evaluate equation (12.18) at $t = 0$ by

$$Z_{\text{in}}|_{t=0} = \left.\frac{E'(t)+F(t)/\varepsilon_{Rt}}{E'(t)-F(t)/\varepsilon_{Rt}}\right|_{t=0} = \frac{e_0+f_0}{e_0-f_0} = Z_L \tag{12.24}$$

For odd-degree characteristics, $f_0 = 0$ (point of perfect transmission at $t = 0$), and so Z_L is equal to unity and equal to the source impedance Z_S. For even-degree characteristics, $f_0 \neq 0$ and $Z_L > 1$.

It can also be shown [3] that the UE values of odd-degree networks are equal about the center of the network, whereas those of even-degree networks are antimetric:

$$Z_k = Z_{N-k+1}(N\,\text{odd}) \quad \text{and} \quad Z_k Z_{N-k+1} = Z_L(N\,\text{even}), \quad k = 1,2,\ldots,N \tag{12.25}$$

This useful feature, where the second half of UE values are calculated from the first half, is used to improve or check the accuracy of the synthesis.

Either of the expressions for z_{11} in equation (12.22) is used to carry out the network synthesis. However, to avoid the accumulation of errors during the synthesis process, it is best to use only one or the other throughout the synthesis, not to mix them. Best results are usually obtained by carrying out parallel independent synthesis by using both formulas, and then using symmetry (odd degree) or antimetry (even degree) to determine which synthesis route has diverged the least.

12.3.3 Realization

The cascade of UEs that form the prototype SI LPF can, in theory, be realized directly in waveguide, coaxial, or planar structures as a series of transmission lines, all of the same length. The characteristic impedances are set according to Z_i, obtained during the synthesis procedure. For a rectangular waveguide realization, this means that the waveguide height dimension (b) varies in proportion to the value of Z_i. When Z_i is high, then b is high, and vice versa, b is normalized to the incoming waveguide dimension, either full height or transformed down. A sixth-degree example is given in Figure 12.8.

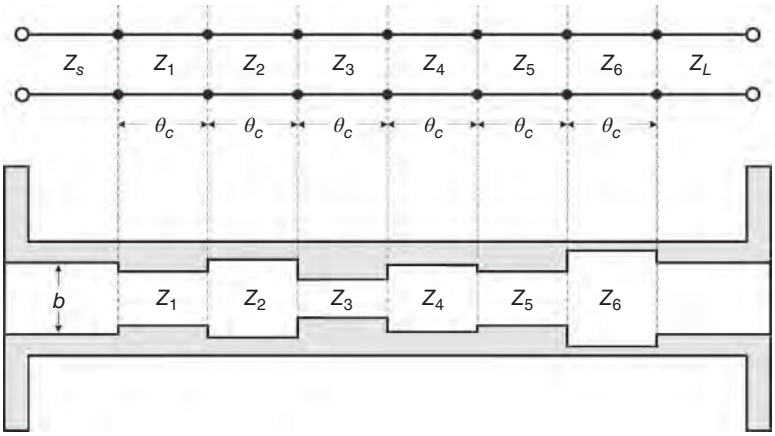


Figure 12.8 Sixth-degree stepped impedance LPF prototype circuit and direct rectangular waveguide realization.

However, this gives some quite abrupt impedance changes associated with stray capacitances at the junctions, and performance can be quite severely degraded from the ideal. Better results are obtained if redundant impedance inverters are introduced at the junctions by using the dual-network theorem on alternate UEs. This tends to equalize the line impedances and gives a degree of freedom, allowing the lines impedances to be prescribed.

Once the redundant inverters are introduced, their impedances are scaled (see Figure 12.9), keeping the coupling coefficient $k_{i,i+1}$ constant:

$$k_{i,i+1} = \frac{K'_{i,i+1}}{\sqrt{Z'_i \cdot Z'_{i+1}}} = \frac{K''_{i,i+1}}{\sqrt{Z''_i \cdot Z''_{i+1}}} \tag{12.26}$$

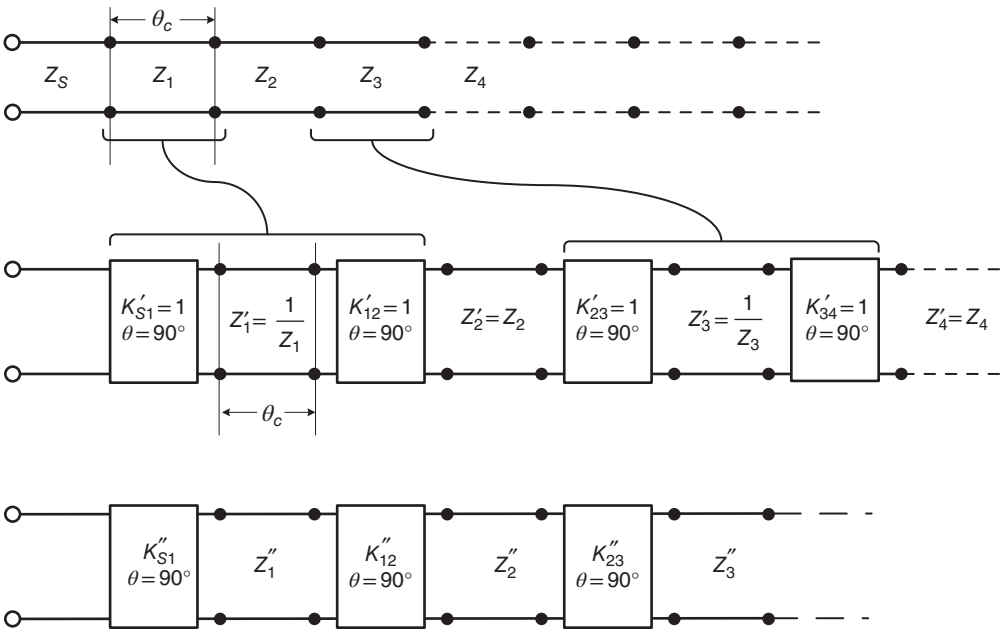


Figure 12.9 Introduction of redundant inverters and impedance scaling.

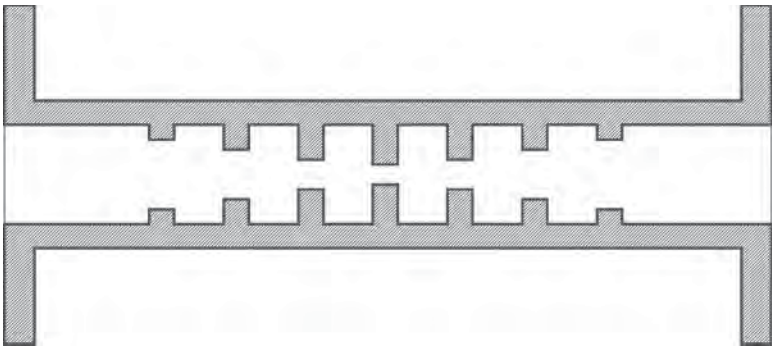


Figure 12.10 Sixth-degree stepped impedance LPF in a rectangular waveguide.

For example, if all Z_i'' are set at unity (a common case), the impedance values of the inverters are given by

$$K_{i,i+1}'' = \frac{1}{\sqrt{Z_i' \cdot Z_{i+1}'}}, \quad i = 0, 1, 2, \ldots, N \tag{12.27}$$

where $Z_0 \equiv Z_S = 1$ and $Z_{N+1} \equiv Z_L$. Then, $Z_1'' = Z_2'' = \cdots = Z_N'' = Z_L'' = 1$. The inverters now have nonunity values and are realized as capacitive irises in a rectangular waveguide structure. If the network is antimetric (even degree) to start with, it becomes symmetric, operating between the unity terminations as in the odd-degree cases. Figure 12.10 illustrates a possible realization in the waveguide for a sixth-degree filter.

To illustrate the design procedure, the design of the prototype circuit for a sixth-degree Zolotarev SI LPF, with a RL of 26 dB, the start of the equiripple region θ_1 set at 8°, and the cutoff angle $\theta_c = 22°$, is selected.

Initially, equation (12.8) is applied to generate the all-pole Chebyshev singularities [roots of $E(\omega)$ and $F(\omega)$], and then by using the mapping function (12.7), the ω-plane singularities are transformed to those for the even-degree Zolotarev function, as shown in the first two columns of Table 12.1. These singularities are then transformed to the t plane using equation (12.11) and are listed in the second two columns of Table 12.1.

Reforming the t-plane polynomials, creating the $[A\ B\ C\ D]$ polynomials, and synthesizing the cascade of six UEs by using the evaluation and extraction method gives the values for the Hi–Lo cascade of UEs in the first two columns of Table 12.2. Now, inserting unity impedance inverters

Table 12.1 Poles and zeros of sixth-degree Zolotarev function singularities in ω and t planes.

ω Plane		t Plane	
Poles [roots of $E(\omega)$]	**Zeros [roots of $F(\omega)$]**	**Mapped poles**	**Mapped zeros**
$-0.1486 \pm j1.1328$	$\pm j0.9707$	$-0.0747 \pm j0.4654$	$\pm j0.3904$
$-0.3952 \pm j0.8516$	$\pm j0.7543$	$-0.1707 \pm j0.3234$	$\pm j0.2946$
$-0.4907 \pm j0.3429$	$\pm j0.4425$	$-0.1851 \pm j0.1230$	$\pm j0.1681$
		$\varepsilon = 704.58$	$\varepsilon_R = 1.000001$

Table 12.2 Element values of stepped impedance lowpass filter.

Hi–Lo cascade		After inclusion of inverters $K'_{i,i+1}$		
	Unit element impedances	UE impedances Z'_i ($K'_{i,i+1} = 1.0$)	Unit elements Z''_i and Z_L scaled to unity	
Z_1	2.62571	2.62571	K''_{S1}	0.61713
Z_2	0.32087	3.11657	K''_{12}	0.34957
Z_3	5.22472	5.22472	K''_{23}	0.24782
Z_4	0.32208	3.10479	K''_{34}	0.24829
Z_5	5.24455	5.24455	K''_{45}	0.24782
Z_6	0.64089	1.56033	K''_{56}	0.34957
Z_L	1.68279		K''_{6L}	0.61713

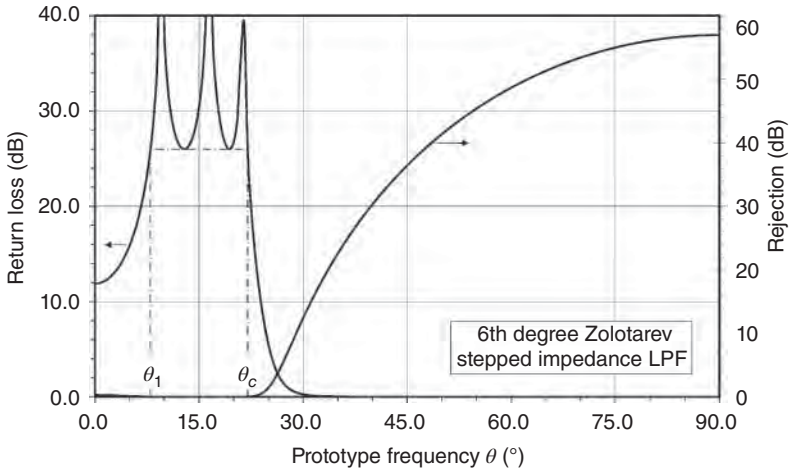


Figure 12.11 Rejection and return loss of the sixth-degree 26 dB return loss stepped impedance lowpass filter, with $\theta_1 = 8°$ and $\theta_c = 22°$.

$K_{i,i+1}$ between the elements (and one at the beginning of the cascade) has the effect of reciprocating every second UE (dual-network theorem), as shown in the third column. Finally, the impedances of the UEs and the terminating impedances are scaled to unity by equation (12.26) (see the final column in Table 12.2). It is evident that the values of the network are now symmetric about the center.

A circuit analysis of this network yields the lowpass RL and rejection characteristics shown in Figure 12.11. It clearly shows the reduced level of the first ripple of the RL characteristic up to $\theta = 8°$ but maintaining 26 dB between $\theta = 8°$ and 22°.

12.3.3.1 Realization of the Impedance Inverters

The impedance inverters of the prototype network are equated to a capacitive iris in a rectangular waveguide, situated between two phase planes (see Figure 12.12). Then, the characteristic impedance of the prototype inverter, which is proportional to the height of the iris b_0, can be

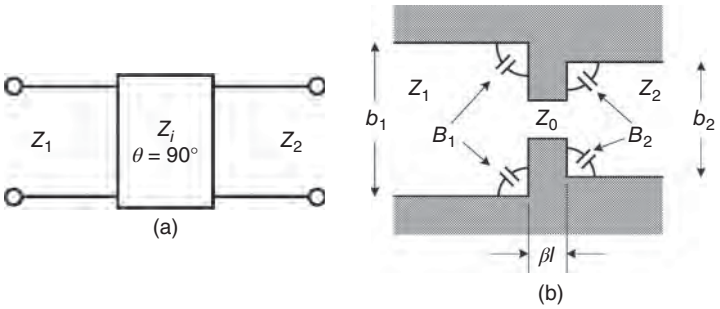


Figure 12.12 Equivalent circuits: (a) impedance inverter and (b) capacitive iris in a rectangular waveguide.

related to the impedance Z_0 of the iris according to the following procedure:

$$\begin{bmatrix} a'' & jb'' \\ jc'' & d'' \end{bmatrix} \qquad \begin{bmatrix} a' & jb' \\ jc' & d' \end{bmatrix}$$

$$a'' = d'' = 0 \qquad a' = Z_{n1}(c - sB_2Z_0)$$

$$b'' = \frac{K_{i,i+1}}{Z_{n2}} \qquad b' = s\frac{Z_0}{Z_{n2}}$$

$$c'' = \frac{Z_{n2}}{K_{i,i+1}} \qquad c' = Z_{n2}\left[c(B_1 + B_2) + s\left[\frac{1}{Z_0} - B_1B_2Z_0\right]\right]$$

$$d' = \frac{c - sB_1Z_0}{Z_{n1}} \tag{12.28}$$

where

Z_{n1} = $\sqrt{Z_{i+1}/Z_i}$ and $Z_{n2} = \sqrt{Z_iZ_{i+1}}$

B_1, B_2 = the fringing capacitances of the capacitive steps [10]

l = thickness of the iris (usually preset)

β = $2\pi/\lambda_{gi}$, λ_{gi} wavelength in iris section

c = $\cos \beta l$, $s = \sin \beta l$

Z_i, Z_{i+1} and $K_{i,i+1}$ = values of characteristic impedances of UEs and inverters from prototype.

The value of the unknown, the impedance Z_0 of the capacitive step, may be found by equating the matrix elements in the following equation and iteratively solving

$$1 + \frac{1}{4}(a' - d')^2 + \frac{1}{4}(b' - c')^2 = 1 + \frac{1}{4}(a'' - d'')^2 + \frac{1}{4}(b'' - c'')^2 \tag{12.29}$$

The reference phase planes are then calculated by [11]

$$\tan 2\phi_1 = \frac{2(b'd' - a'c')}{(a'^2 - d'^2) + (b'^2 - c'^2)}$$

$$\tan 2\phi_2 = \frac{2(a'b' - c'd')}{(d'^2 - a'^2) + (b'^2 - c'^2)} \tag{12.30}$$

It should be noted that inclusion of the inverters in the prototype network has no effect on its performance, because the inverters are frequency independent. However, the inverters are converted into capacitive irises to allow realization of the lowpass characteristic in a waveguide structure. These irises are frequency dependent and therefore behave as additional elements of

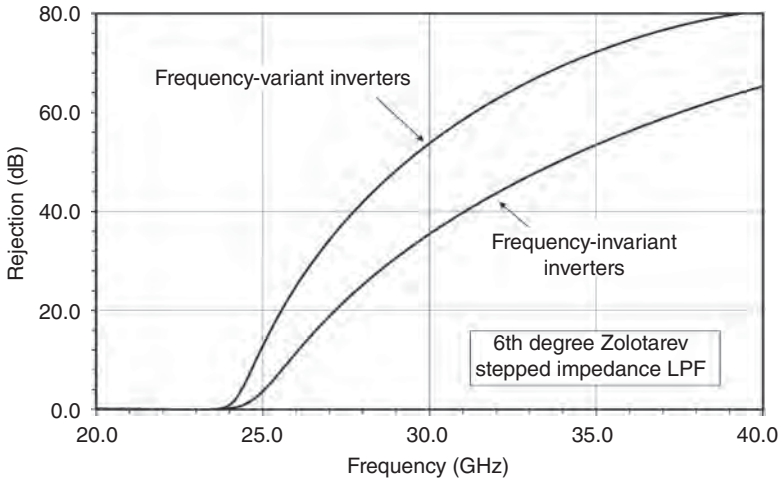


Figure 12.13 Comparison of the rejection performance of a sixth-degree stepped impedance LPF with and without frequency variation in the UEs.

filter design. Thus, the structure in Figure 12.10, although derived from a 6th-degree prototype network, exhibits the performance of a 13th-degree (6 + 7) filter. Figure 12.13 illustrates the effect for an LPF with a cutoff frequency of 23.5 GHz.

12.4 Short-Step Transformers

The short-step transformer is introduced here because of the similarity of its design procedure to that for an SI LPF, using even-degree Zolotarev functions and the same commensurate-line synthesis method. The short-step transformer is a useful matching element between transmission lines of different impedance levels, from a full-height to reduced-height waveguide [12]. A special case of the short-step transformer is noted in which it degenerates to the more familiar quarter-wavelength transformer.

It was noted earlier that when the synthesis procedure for SI LPFs is applied to an even-degree LPF filter function, the resulting UE characteristic impedances Z_i are antimetric in value about the center of the filter. Consequently, the load impedance Z_L is unequal in value to the source Z_S. This is so because, at zero frequency, the even-degree filtering function has a nonzero insertion loss, and the cascade of UEs realizing it is transparent; that is, the source of impedance $Z_S = 1\ \Omega$ is directly connected to the load Z_L. Thus, it is the nonunity Z_L and the resultant mismatch between the source and load that is responsible for the finite RL at $\omega = 0$, according to the relationship

$$S_{11}(\omega)|_{\omega=0} = \frac{f_0}{e_0} = 10^{-\mathrm{RL}_0/20}$$

$$Z_L = \left.\frac{1 + S_{11}(\omega)}{1 - S_{11}(\omega)}\right|_{\omega=0} = \frac{e_0 + f_0}{e_0 - f_0} \tag{12.31}$$

where RL_0 is the RL of the LPF function at zero frequency and e_0 and f_0 are the constant coefficients of the polynomials $E(\omega)$ and $F(\omega)$, respectively.

The large and adjustable first ripple of the even-degree Zolotarev function is ideal for providing the necessary RL RL_0 at zero frequency for a given degree. This value can be

adjusted by altering the equiripple RL level or the bandwidth of the Zolotarev function [the equiripple region between $\omega = x_1$ and $\omega = 1$ in the ω plane (see Figure 12.4) or the equivalents $\theta = \theta_1$ and $\theta = \theta_c$ in the θ plane]. Usually, the RL level is chosen as the variable, and a simple single-variable solving routine is set up to converge at the value of RL that gives the desired Z_L through equation (12.31), fixing the degree and bandwidth. If the resultant value for RL_0 is unacceptably low or inappropriately high, the degree or bandwidth may then be adjusted and the process repeated.

The design process for the short-step transformer allows for a tradeoff between wide equiripple bandwidth and out-of-band rejection levels on the upper side (the rejection level on the lower side is defined by the required transformer ratio).

Defining $\theta_b = \theta_c - \theta_1$ as the bandwidth and $\theta_0 = (\theta_c + \theta_1)/2$ as the center frequency of the transformer, setting θ_0 below 45° tends to increase the rejection on the upper side to levels above those on the lower side, and vice versa for θ_0 above 45°. Widening the equiripple bandwidth θ_b also tends to lower the rejection levels on the upper side, but, in practice, all of these considerations of rejection are usually secondary to those of a wide bandwidth of operation, good RL, and ease of realization.

We illustrate the procedure by taking an example of a sixth-degree transformer that is required to transform to an impedance level of $Z_L = 4Z_S$, and $\theta_1 = 20°$, and $\theta_c = 50°$, that is, $\theta_0 = 35°$. The iterative procedure results in an equiripple RL level of 21.57 dB between $\theta = 20°$ and 50°. At this RL level, the sixth-degree Zolotarev function gives the desired insertion loss of 1.94 dB (4.44 dB RL) at zero frequency, corresponding to the specified 1:4 source: load impedance ratio. The insertion loss and RL characteristics are shown in Figure 12.14a, and the six synthesized line impedance values are given in Table 12.3.

An interesting feature of the short-step transformer is that if the center frequency is set at 45°, the impedance values of the adjacent pairs of the UEs along the length of the transformer become equal in value. Let us now take the same example, where $\theta_1 = 30°$ and $\theta_c = 60°$, that is, $\theta_0 = 45°$. For this case, the equiripple RL level is 30.8 dB, and the insertion loss and RL characteristics are shown in Figure 12.14b. Note that they have become symmetric about 45°. The synthesized line impedance values for this symmetric case are listed in Table 12.3. It should be noted that the adjacent internal steps have become equal valued. Now, the short-step transformer has become, in effect, a regular quarter-wave transformer, with three steps, each $2 \times 45° = 90°$ at the center frequency. The impedance levels are still antimetric about the center, with the value of the middle step equal to $\sqrt{Z_L}$, if $N/2$ is odd.

12.5 Synthesis and Realization of Mixed Lumped/Distributed Lowpass Filters

Like the SI circuit, the lumped/distributed (L/D) LPF circuit is composed of a series of UEs, but is distinct from the SI LPF circuit in that the UEs are in pairs, each pair having the same characteristic impedance Z_i, and at the junction of each pair, a distributed shunt capacitor is located. Figure 12.15 shows a cascade of one such UE pair with an open-circuited UE connected in shunt at each end, forming the distributed capacitors (see Figure 12.15b). As such, this circuit represents a third-degree prototype of a tapered–corrugated LPF.

At a frequency where the lengths θ_c of the UEs become 90°, it can be seen that the open circuits at the ends of the shunt UEs transform to short circuits across the mainline, and the double-length UEs become 180°. Thus, this third-degree circuit gives a transmission zero at the frequency at which θ_c of the UEs become 90° (or 90° plus an integer multiple of 180°).

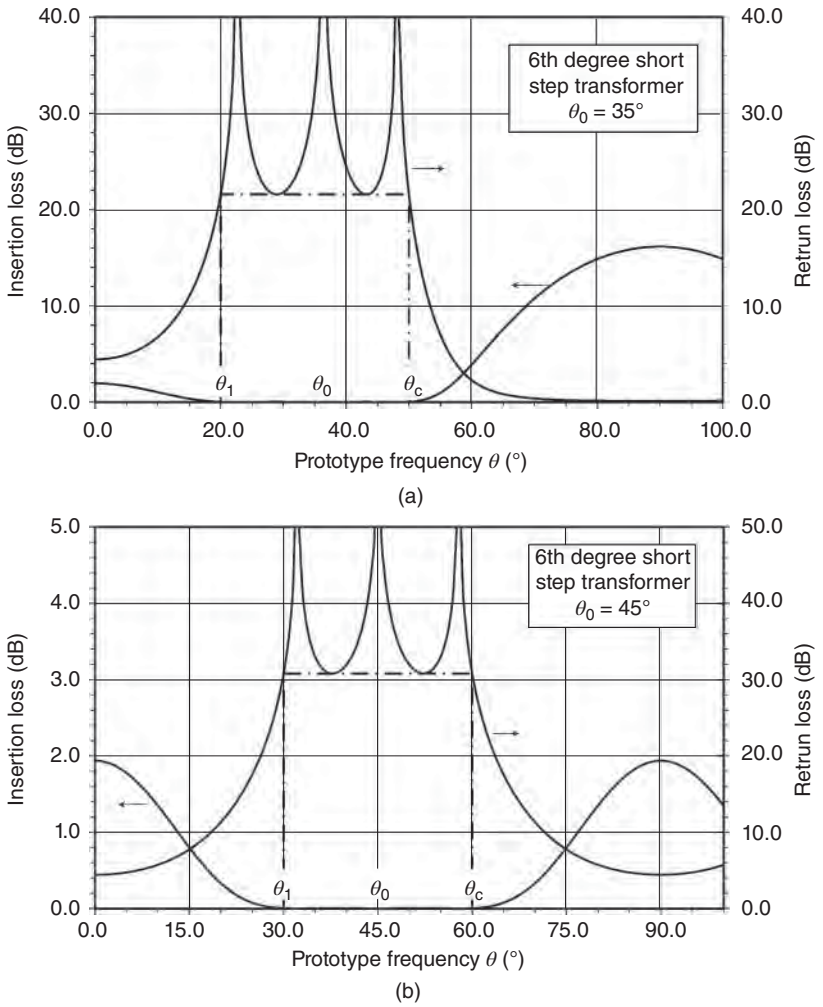


Figure 12.14 Insertion loss and return loss characteristics of the 1:4 short-step transformer: (a) $\theta_0 = 35°$ and (b) $\theta_0 = 45°$.

Table 12.3 Short-step transformer line impedances for center frequencies $\theta_0 = 35°$ and $\theta_0 = 45°$.

	Line impedances, $\theta_0 = 35°$	**Line impedances, $\theta_0 = 45°$**
Z_S	1.0	1.0
Z_1	1.6427	1.2417
Z_2	0.9661	1.2417
Z_3	2.9717	2.0000
Z_4	1.3460	2.0000
Z_5	4.1403	3.2215
Z_6	2.4350	3.2215
Z_L	4.0	4.0

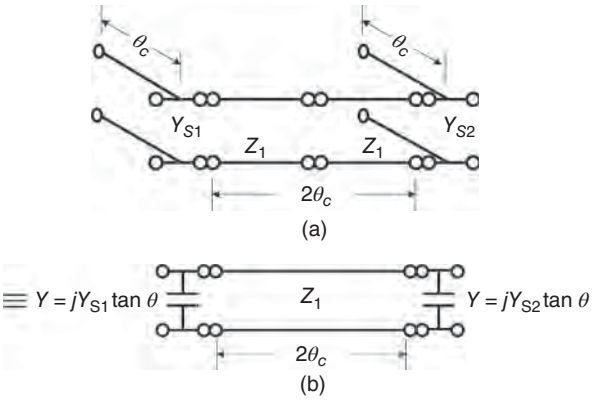


Figure 12.15 Basic prototype elements of the lumped/distributed filter: (a) double-length unit element with a shunt-connected open-circuited UE at each end (distributed capacitors) and (b) with the distributed capacitors represented by their lumped equivalents.

Like the SI LPF, we demonstrate that the class of circuit as shown in Figure 12.15 has the same form of transfer function as a Chebyshev or Zolotarev function of the second type, with a pair of transmission half-zeros at a frequency of $\omega = 1/(\sin\theta_c)$. Here, θ_c is the chosen cutoff frequency for the distributed prototype and is transformed with the same mapping formula as the one for the SI LPF.

12.5.1 Formation of the Transfer and Reflection Polynomials

The mapping formula for the SI LPF is repeated here

$$\omega = \frac{\sin\theta}{\sin\theta_c} = a\sin\theta \tag{12.32}$$

where θ_c is the commensurate-line length and $a = 1/\sin\theta_c$. In this case, a represents the position of the half-zero pair in the ω plane, which maps to 90° in the θ plane (see Figure 12.16). Again, it can be seen that the repeating pattern in Figure 12.16b is mapped from the portion of the ω-plane characteristic between $\omega = \pm a$. The transfer function for the Chebyshev function of the second kind or Zolotarev function with a single half-zero pair has the following form in the ω plane:

$$S_{21}(\omega) = \frac{\sqrt{\omega^2 - a^2}}{\varepsilon E(\omega)} \tag{12.33}$$

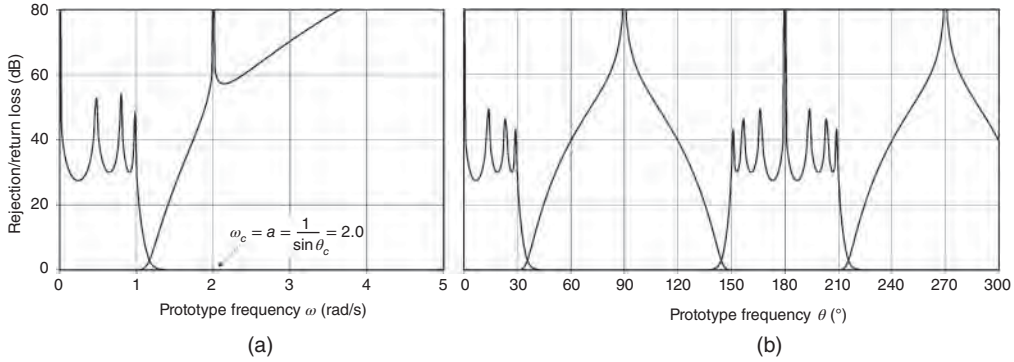


Figure 12.16 Mapping of the seventh-degree Zolotarev transfer and reflection functions from the ω plane (a) to the θ plane (b) through the mapping function $\omega = a\sin\theta$, where $a = 1/\sin(\theta_c)$. In this case $\theta_1 = 10°$, $\theta_c = 30°$, and the return loss = 30 dB.

Applying the transformation to $S_{21}(\omega)$, we obtain

$$S_{21}(\theta) = \frac{j\sqrt{a^2 - a^2\sin^2\theta}}{\varepsilon E(a\sin\theta)} = \frac{j\cos\theta}{\varepsilon_t E(a\sin\theta)} \tag{12.34}$$

where the constants are consolidated into ε_t.

Applying the identities $\sin\theta = \tan\theta/\sqrt{1+\tan^2\theta}$ and $\cos\theta = 1/\sqrt{1+\tan^2\theta}$ and letting $t = j\tan\theta$, we obtain

$$S_{21}(t) = \frac{1}{\sqrt{1-t^2}} \cdot \frac{1}{\varepsilon_t E\left(-jta/\sqrt{1-t^2}\right)} = (-1)^{(N+1)/2} \frac{\left[\sqrt{1-t^2}\right]^{N-1}}{\varepsilon_t' E'(t)} \tag{12.35}$$

where ε_t' is another constant and $E'(t)$ is an odd-degree polynomial in the variable t. The numerator is a polynomial $[\sqrt{1-t^2}]^{N-1}$, which, because $N-1$ is an even number, can be formed into the $(N-1)$th-degree even polynomial $P(t)$, the numerator of $S_{21}(t)$. From this, we observe that $S_{21}(t)$ has $N-1$ zeros at $t = \pm 1$.

Knowing the form of the $S_{21}(t)$ function from the mapped prototype polynomials that represent the lowpass transfer function, we can check whether the transfer function of the commensurate-line circuit, as shown in Figure 12.15, has the same form. Once this is established, the $[A\,B\,C\,D]$ polynomials are formed from the prototype polynomials and the synthesis of the circuit can begin.

If we take a cascade of two UEs of commensurate length θ_c and both of the same characteristic impedance Z_1 (Figure 12.17a), then the transfer $[A\,B\,C\,D]$ matrix for the UE pair is

$$[A\ B\ C\ D] = \begin{bmatrix} \cos\theta & jZ_1\sin\theta \\ \dfrac{j\sin\theta}{Z_1} & \cos\theta \end{bmatrix}^2 = \cos^2\theta \begin{bmatrix} 1 & jZ_1\tan\theta \\ \dfrac{j\tan\theta}{Z_1} & 1 \end{bmatrix}^2$$

$$= \frac{1}{1-t^2} \cdot \begin{bmatrix} 1+t^2 & 2Z_1 t \\ \dfrac{2t}{Z_1} & 1+t^2 \end{bmatrix} \tag{12.36}$$

As before, $t = j\tan\theta$. Now we pre- and postmultiply by the distributed capacitive stubs (Figure 12.17b) and obtain

$$\frac{1}{1-t^2} \cdot \begin{bmatrix} 1 & 0 \\ Y_{S1}t & 1 \end{bmatrix} \cdot \begin{bmatrix} 1+t^2 & 2Z_1 t \\ \dfrac{2t}{Z_1} & 1+t^2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ Y_{S2}t & 1 \end{bmatrix} = \frac{1}{1-t^2} \cdot \begin{bmatrix} A_2(t) & B_1(t) \\ C_3(t) & D_2(t) \end{bmatrix} \tag{12.37}$$

where again the subscripts represent the degree of the polynomial in the variable t. This is now the $[A\,B\,C\,D]$ matrix for a third-degree lumped/distributed lowpass circuit as shown in Figure 12.15.

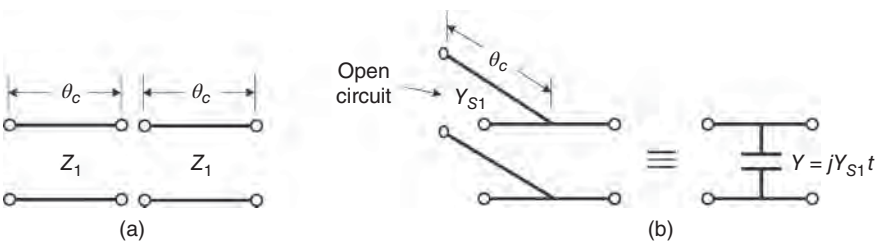


Figure 12.17 Components of the tapered–corrugated LPF: (a) double unit element and (b) distributed shunt capacitor.

With the addition of more UE pairs and shunt capacitors, the general form for an Nth-degree $[A\,B\,C\,D]$ matrix becomes

$$[A\ B\ C\ D] = \frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A_{N-1}(t) & B_{N-2}(t) \\ C_N(t) & D_{N-1}(t) \end{bmatrix} \tag{12.38}$$

from which the transfer function $S_{21}(t)$ is derived as

$$S_{21}(t) = \frac{2[1-t^2]^{(N-1)/2}}{A_{N-1}(t) + B_{N-2}(t) + C_N(t) + D_{N-1}(t)} = \frac{[1-t^2]^{(N-1)/2}}{\varepsilon_t E_N(t)} \tag{12.39}$$

This is in the same form as equation (12.35), which results from the transform of the transfer function polynomials.

Figure 12.15 confirms that the class of circuit has the same transfer characteristic (and because it is lossless, it also has the same reflection characteristic) as the Chebyshev or Zolotarev polynomials. At this point, we can generate the $E(t)$ and $F(t)$ polynomials to satisfy a particular specification, relate them to the $[A\,B\,C\,D]$ polynomials of the L/D prototype circuit, and then synthesize the values of the circuit elements.

12.5.2 Synthesis of the Tapered–Corrugated Lowpass Prototype Circuit

Using the methods presented in Section 12.2, the reflection function numerator polynomial $F(\omega)$, corresponding to the transfer function $S_{21}(\omega)$ with a single half-zero pair $P(\omega) = \sqrt{\omega^2 - a^2}$ as the numerator, can be found. The next task is to find $E(\omega)$, the denominator polynomial common to both $S_{11}(\omega)$ and $S_{21}(\omega)$, so that the network synthesis for the lumped/distributed prototype network can begin. However, because the numerator polynomial of $S_{21}(\omega)$ is in the form $P(\omega) = \sqrt{\omega^2 - a^2}$, the alternating-pole principle cannot be used directly to find $E(\omega)$.

However, we now know that with transformation of the frequency variable from ω to $j\tan\theta$, a rational function for $S_{21}(t)$ in the form of equation (12.35) is formed. This has $P(t) = [1-t^2]^{(N-1)/2}$ as its numerator, which, because N is an odd number, represents a polynomial in the variable t.

So the method used to find the polynomials of the reflection function $S_{11}(t) = F(t)/E(t)$ would proceed as follows:

1) With the desired lowpass cutoff angle θ_c, which determines the width of the reject band before the second harmonic passband cuts in (see Figure 12.16), calculate the Tx half-zero position $a = 1/\sin\theta_c$. Then, with the RL equiripple level that is desired over the passband, form the polynomial $F(\omega)$ by using the method of Section 12.2 for the Chebyshev (second kind) case or [2] for the odd-degree Zolotarev function, both with the single half-zero pair at $\omega = \pm a$ in the numerator.
2) Root the polynomial $F(\omega)$ and transform its singularities to the t plane by the transform (12.11b). Form the polynomial $F(t)$ from the t-plane singularities and the polynomial $P(t)$ from $P(t) = [1-t^2]^{(N-1)/2}$.
3) Derive ε_t at a convenient frequency where the RL level is known, such as at the cutoff frequency where $t = t_c = j\tan\theta_c$:

$$\varepsilon_t = \frac{1}{|F(t)_c|} \cdot \frac{[1-t_c^2]^{(N-1)/2}}{\sqrt{10^{\mathrm{RL}/10} - 1}} \tag{12.40}$$

4) Form the polynomial $P(t) \pm \varepsilon_t F(t)$, root it, and reflect any singularities that are in the right half-plane back to the left half-plane to preserve the Hurwitz condition. Now the Hurwitz polynomial $E(t)$ can be formed.

With $E(t)$ and $F(t)$ polynomials, we begin the synthesis of the network, employing the same methods as for the SI LPF. Noting the form of the polynomials in the cascade matrix [equation (12.38)], we can form the $A(t)$, $B(t)$, $C(t)$, and $D(t)$ polynomials of the t plane $[A\,B\,C\,D]$ matrix as follows:

$$\begin{aligned}
A(t) &= (e_0 + f_0) + (e_2 + f_2)t^2 + (e_4 + f_4)t^4 + \cdots \\
B(t) &= (e_1 + f_1)t + (e_3 + f_3)t^3 + (e_5 + f_5)t^5 + \cdots \\
C(t) &= (e_1 - f_1)t + (e_3 - f_3)t^3 + (e_5 - f_5)t^5 + \cdots \\
D(t) &= (e_0 - f_0) + (e_2 - f_2)t^2 + (e_4 - f_4)t^4 + \cdots
\end{aligned} \tag{12.41}$$

where e_i and f_i, $i = 0, 1, 2, \ldots, N$, are the coefficients of the $E(t)$ and $F(t)/\varepsilon_{Rt}$ polynomials, respectively. ε_{Rt} is unity here because the numerator polynomial is one less in degree than the denominator (i.e., the filtering function is noncanonical).

12.5.2.1 Network Synthesis

For the first synthesis cycle, the overall ladder network is assumed to consist of a shunt capacitor, followed by a double-length UE and the remainder matrix $[A\,B\,C\,D]^{(2)}$.

The first component to be extracted is the shunt capacitor Y_{S1}, to prepare the network for the extraction of the double-length UE of characteristic impedance Z_1. The overall network is assumed at first to be composed of the shunt capacitor in cascade with the remainder matrix $[A\,B\,C\,D]^{(1)}$ (see Figure 12.18):

$$\begin{aligned}
\frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A(t) & B(t) \\ C(t) & D(t) \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ tY_{S1} & 1 \end{bmatrix} \cdot \frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A^{(1)} & B^{(1)} \\ C^{(1)} & D^{(1)} \end{bmatrix} \\
&= \frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A^{(1)} & B^{(1)} \\ C^{(1)} + tY_{S1}A^{(1)} & D^{(1)} + tY_{S1}B^{(1)} \end{bmatrix}
\end{aligned} \tag{12.42}$$

The short-circuit admittance y_{11}, looking into this network, is

$$y_{11} = \frac{D(t)}{B(t)} = y_{11R} + tY_{S1} = \frac{D^{(1)}}{B^{(1)}} + tY_{S1} \tag{12.43}$$

The problem now is to evaluate Y_{S1}. Unlike most other extractions, this one cannot be done by simply evaluating at a convenient value of t, for example, at $t = 0, 1, t_c$ or ∞. However, we can exploit a feature of a network that consists of a double UE as the leading component (e.g., $[A\,B\,C\,D]^{(1)}$ in Figure 12.18) and the *differential* of its input admittance with respect to t is zero at $t = \pm 1$. This is easily proved by differentiating the input admittance y_{11R} of the cascaded

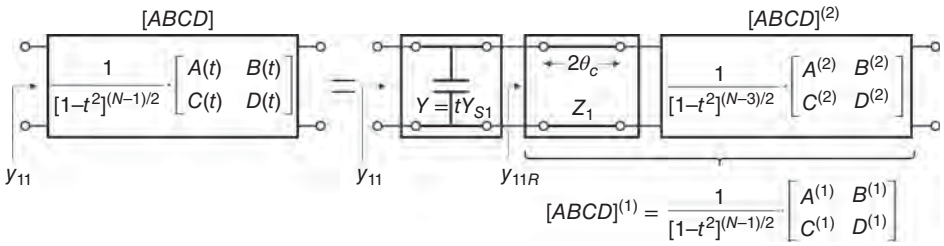


Figure 12.18 Initial synthesis cycle for the lumped/distributed LPF ladder network; the original matrix comprises a shunt capacitor + double UE + remainder matrix $[A\,B\,C\,D]^{(2)}$.

double-length line plus remainder network ($[A\,B\,C\,D]^{(1)}$ in Figure 12.18) with respect to t, when it is evaluated at $t = 1$.

The differential y'_{11} of the input admittance y_{11} of the overall network is derived from equation (12.43) as

$$y'_{11} = \frac{dy_{11}}{dt} = \left[\frac{D(t)}{B(t)}\right]' = \left[\frac{D^{(1)}}{B^{(1)}}\right]' + Y_{S1} \tag{12.44}$$

where the prime here denotes differentiation with respect to t. Since the first term on the right-hand side (RHS) of equation (12.44) is equal to zero when it is evaluated at $t = 1$, we have

$$y'_{11}|_{t=\pm 1} = \left[\frac{D(t)}{B(t)}\right]'\Bigg|_{t=\pm 1} = Y_{S1} \tag{12.45}$$

Now, with the value of Y_{S1} known, the synthesis continues rather more conventionally:

1) Extract the shunt admittance Y_{S1} from the overall matrix:

$$\begin{aligned}
&\begin{bmatrix} 1 & 0 \\ -tY_{S1} & 1 \end{bmatrix} \cdot \frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A & B \\ C & D \end{bmatrix} \\
&\quad = \frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A & B \\ C - tY_{S1}A & D - tY_{S1}B \end{bmatrix} \\
&\quad = \frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A^{(1)} & B^{(1)} \\ C^{(1)} & D^{(1)} \end{bmatrix}
\end{aligned} \tag{12.46}$$

2) The remainder matrix $[A\,B\,C\,D]^{(1)}$ contains a double UE plus a remainder matrix $[A\,B\,C\,D]^{(2)}$ (see Figure 12.18). Evaluate the characteristic impedance Z_1 of the first double UE by

$$\begin{aligned}
\frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A^{(1)} & B^{(1)} \\ C^{(1)} & D^{(1)} \end{bmatrix} &= \frac{1}{(1-t^2)} \cdot \begin{bmatrix} 1+t^2 & 2tZ_1 \\ 2t/Z_1 & 1+t^2 \end{bmatrix} \cdot \frac{1}{[1-t^2]^{(N-3)/2}} \cdot \begin{bmatrix} A^{(2)} & B^{(2)} \\ C^{(2)} & D^{(2)} \end{bmatrix} \\
&= \frac{1}{[1-t^2]^{(N-1)/2}} \\
&\quad \cdot \begin{bmatrix} (1+t^2)A^{(2)} + 2tC^{(2)}Z_1 & (1+t^2)B^{(2)} + 2tD^{(2)}Z_1 \\ (1+t^2)C^{(2)} + 2tA^{(2)}/Z_1 & (1+t^2)D^{(2)} + 2tB^{(2)}/Z_1 \end{bmatrix}
\end{aligned} \tag{12.47}$$

The admittance y_{11R}, looking in at the input of $[A\,B\,C\,D]^{(1)}$, is

$$y_{11R} = \frac{D^{(1)}}{B^{(1)}} = \frac{(1+t^2)D^{(2)} + 2tB^{(2)}/Z_1}{(1+t^2)B^{(2)} + 2tD^{(2)}Z_1} \tag{12.48}$$

By evaluating at $t = 1$, we obtain

$$y_{11R}\big|_{t=1} = \frac{D^{(1)}}{B^{(1)}}\bigg|_{t=1} = \frac{2B^{(2)}/Z_1 + 2D^{(2)}}{2B^{(2)} + 2D^{(2)}Z_1} = \frac{1}{Z_1} \tag{12.49}$$

3) Knowing the value of Z_1, we can now extract the double-length line from $[A\,B\,C\,D]^{(1)}$ such that

$$\begin{aligned}
&\frac{1}{(1-t^2)} \cdot \begin{bmatrix} 1+t^2 & -2Z_1t \\ -2t/Z_1 & 1+t^2 \end{bmatrix} \cdot \frac{1}{[1-t^2]^{(N-1)/2}} \cdot \begin{bmatrix} A^{(1)} & B^{(1)} \\ C^{(1)} & D^{(1)} \end{bmatrix} \\
&\quad = \frac{1}{[1-t^2]^{(N+1)/2}} \cdot \begin{bmatrix} A^{(2)} & B^{(2)} \\ C^{(2)} & D^{(2)} \end{bmatrix}
\end{aligned} \tag{12.50}$$

4) Finally, divide the top and bottom of the remainder matrix $[A\,B\,C\,D]^{(2)}$ by $(1-t^2)^2$ to obtain the correct degree $(N-3)/2$ for the denominator polynomial (see $[A\,B\,C\,D]^{(2)}$ in Figure 12.18). Having extracted the first shunt capacitor Y_{S1}, the numerator polynomials $A^{(2)}$, $B^{(2)}$, $C^{(2)}$, and $D^{(2)}$ are exactly divisible by $(1-t^2)^2$.

After the first shunt capacitor and first double-length line are extracted, the cycle is repeated by using the remainder matrix $[A\,B\,C\,D]^{(2)}$, and so on until the last shunt capacitor, which is extracted at $t = 1$, as usual.

12.5.3 Realization

Figure 12.19 shows the network for a seventh-degree prototype network for a lumped/distributed LPF, with shunt-distributed capacitors Y_{Si} and the double-line lengths with characteristic impedances Z_i. The shunt capacitors (Figure 12.20a) in the prototype network translate into capacitive irises (Figure 12.20b) in the waveguide or coaxial structure, and the matrix matching procedure to determine the value of the iris impedance, which is very similar to the method used for the SI LPF, is outlined in Figure 12.11 and equation (12.28) [11]:

$$
\begin{array}{ll}
\begin{bmatrix} a'' & jb'' \\ jc'' & d'' \end{bmatrix} & \begin{bmatrix} a' & jb' \\ jc' & d' \end{bmatrix} \\
a'' = Z_{n1} & a' = Z_{n1}(c - sB_2Z_0) \\
b'' = 0 & b' = s\dfrac{Z_0}{Z_{n2}} \\
c'' = t_c Y_S Z_{n2} & c' = Z_{n2}\left[c(B_1+B_2) + s\left[\dfrac{1}{Z_0} - B_1B_2Z_0\right]\right] \\
d'' = \dfrac{1}{Z_{n1}} & d' = (c - sB_1Z_0)/Z_{n1}
\end{array}
\tag{12.51}
$$

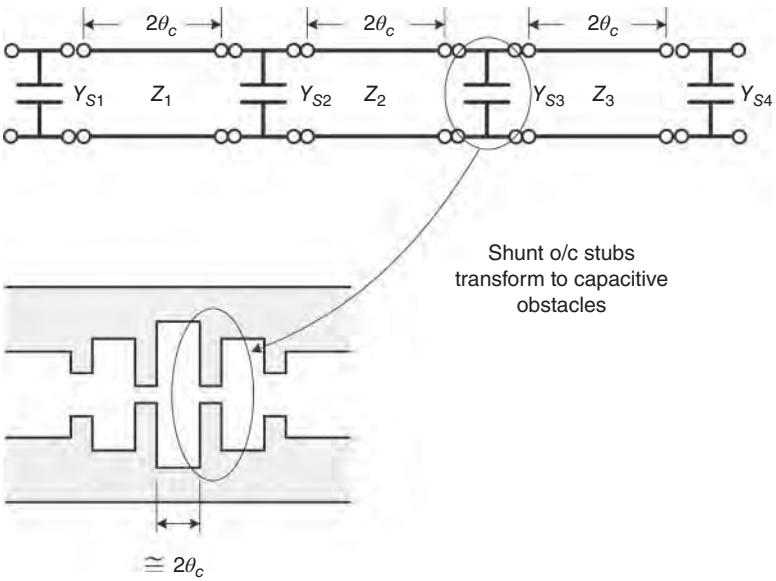


Figure 12.19 Relationships of the components of the prototype lumped/distributed LPF network to a waveguide structure (o/c = open circuit).

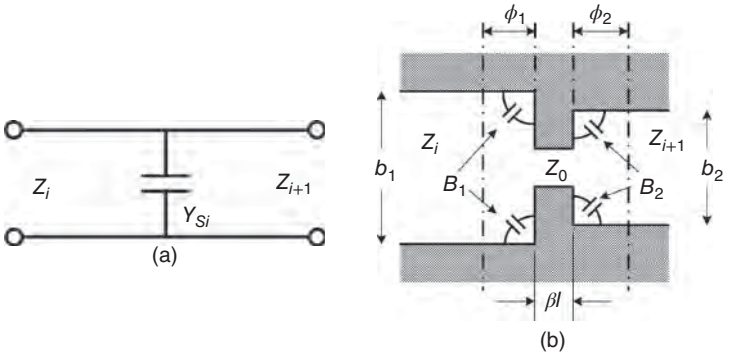


Figure 12.20 Equivalent circuits: (a) shunt-distributed capacitor and (b) capacitive iris in rectangular waveguide.

where

$$\begin{aligned} Z_{n1} &= \sqrt{Z_{i+1}/Z_i} \text{ and } Z_{n2} = \sqrt{Z_i Z_{i+1}} \\ B_1, B_2 &= \text{fringing capacitances of capacitive steps [10]} \\ l &= \text{thickness of iris (usually preset)} \\ \beta &= 2\pi/\lambda_{gi},\ \lambda_{gi} = \text{wavelength in iris section} \\ c &= \cos\beta l,\ s = \sin\beta l,\ t_c = \tan\theta_c \\ Z_i, Z_{i+1}, Y_{Si} &= \text{values of double unit element impedances and shunt capacitors from prototype.} \end{aligned}$$

In a method similar to that used for the SI LPF, the value of the unknown, the impedance Z_0 of the capacitive step, is found by equating the matrix elements in the following equation and iteratively solving

$$1 + \frac{1}{4}(a' - d')^2 + \frac{1}{4}(b' - c')^2 = 1 + \frac{1}{4}(a'' - d'')^2 + \frac{1}{4}(b'' - c'')^2 \tag{12.52}$$

A first estimate for the Z_0 of the $(i+1)$th iris, $i = 0, 1, \ldots, N$, may be obtained for this iterative process if the capacitive step fringing capacitors are ignored. Solving equation set (12.51) by using equation (12.52) with B_1 and $B_2 = 0$ yields

$$Z_0 = \frac{Z_{n2}}{\sqrt{2}}\left[\sqrt{x+1} - \sqrt{x-1}\right]$$

where

$$x = \frac{1}{2}\left[\frac{Z_{i+1}}{Z_i} + \frac{Z_i}{Z_{i+1}} + Z_i Z_{i+1}\left(\frac{t_c Y_{s,i+1}}{\sin\beta l}\right)^2\right] \tag{12.53}$$

The reference phase planes, which tend to shorten the inter-iris phase lengths, are calculated from equation (12.30). Small adjustments can be applied to Z_0 to compensate for proximity effects, which are the capacitive couplings that occur between adjacent irises [13, 14]. A very similar method is used to design the capacitive sections for coaxial LPFs.

12.5.3.1 Synthesis Accuracy Considerations

As described in the previous section, the synthesis of the prototype network for a lumped/distributed or SI LPF is a progressive process: extracting the electrical components one by one in sequence from the defining polynomials, building up the ladder network, and finishing when

the final component has been extracted and the polynomials are of zero degree. Accompanying such processes, computation errors accumulate, which sooner or later cause the synthesis to become unstable, usually in a spectacular and sudden fashion. More accurate input data and double-precision computing help the situation, but other simple measures can also improve the accuracy:

1) *Parallel synthesis.* The synthesis procedure made use of the short-circuit parameter y_{11}, calculated from the B and D polynomials of the t-plane $[A\,B\,C\,D]$ matrix. y_{11} can also be calculated from the A and C polynomials ($y_{11} = C(t)/A(t)$) [equation (12.42)] and the element values of the network may be determined using the same procedure. This gives two sets of values for Y_{Si} and Z_i. Since we know that the values should be symmetric about the center of the filter, the set that gives the better symmetry is selected. After the set that is the most accurate is determined, the values of the components in the second half of the network can be made to be symmetric with those in the first half. It is not advisable to average the values from the two sets, particularly after each step of the synthesis process.
2) *Higher-degree LPFs.* By using the symmetry property, an accurate synthesis may be carried out for LPFs up to degrees as high as 25 before synthesis failures start to occur. By using an adapted z-plane synthesis technique [15–17], filters with degrees up to 31 can be synthesized with reasonable accuracy.

To design LPFs of degrees greater than 31, we take advantage of the observation that the dimensions of the central regions of higher-degree LPFs tend to be uniform, that is, the heights of the high and low impedance sections are not significantly different from each other in the central regions of the LPF. Thus, for the high degree LPF (degree > 31), it is feasible to design a degree 31 LPF (or whatever degree the limit of accuracy allows), and then "copy and paste" the dimensions of its central section repeatedly until the desired degree of LPF is obtained. Analysis indicates that this usually achieves good results with only slight degradation in RL level near the cutoff regions. Care should be taken to ensure that after "copying and pasting," the network is still physically symmetric about the center.

Figure 12.21 shows the results of extending the central dimensions of a degree 31 Zolotarev LPF, with a design RL of 30 dB and cutoff frequency of 12 GHz, to a degree 49 LPF. We can see that the 30 dB level is largely preserved, except for the regions near to the cutoff frequency.

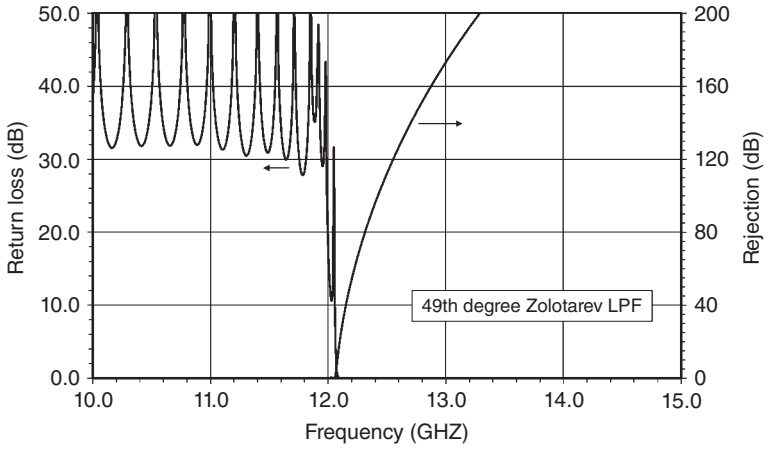


Figure 12.21 Rejection and return loss performance of a degree 49 Zolotarev LPF, designed by extrapolation from a degree 31 prototype.

12.5.3.2 Spurious Responses

Since the main function of an LPF is to reject noise and interfering signals up to frequencies that can be several times the cutoff frequency, preservation of the integrity of the reject bands is critical. Spurious responses occur for three main reasons:

1) The repeat passbands of the commensurate-line network
2) Propagation of the higher-order TE_{n0} modes, $n = 2,3, \ldots$
3) Mode conversion within the filter structure.

12.5.3.3 Repeat Passbands

The repeat passbands of a commensurate-line ladder network are an inherent feature of this class of microwave circuits and cannot be altogether avoided. Because there exist nonideal commensurate-line elements within the network such as the capacitive irises in rectangular waveguide that are realizing the distributed capacitors in lumped/distributed LPFs, the repeat passbands tend to occur at lower frequencies than the prototype predicts. This effect may be counteracted to a certain extent by choosing a lower cutoff frequency for the prototype, reducing the passband width, but this tends to make the variations in the values of the synthesized elements more extreme.

12.5.3.4 Higher-Order TE_{n0} Modes in Waveguide

As the frequency of the signal increases, the width of the waveguide will become progressively cut on for the higher ordered modes, first the TE_{20} mode, then the TE_{30}, and so on. If present at the input to the LPF, these modes will propagate through the filter at the higher frequencies with theoretically the same transfer characteristics as the fundamental TE_{10} mode, although with reduced bandwidths. The spurious bands are approximately predicted as follows [5].

In the lower passband edge f_{0n}, $f_{0n} = nf_{01}$, where $f_{01} = c/2a \approx 15/a$ (GHz) is the fundamental waveguide cutoff frequency, c is the velocity of light, a (cm) is the waveguide broad dimension, and f_{0n} (GHz) is the cutoff frequency of the waveguide of the TE_{n0} mode passband. The upper band edge f_{cn} is given by

$$f_{cn} = \sqrt{f_{cl}^2 + (n^2 - 1)f_{01}^2} \tag{12.54}$$

where f_{c1} and f_{cn} are the upper edges of the fundamental and the TE_{n0}-mode passbands, respectively. For waveguide WR42, which has a broad dimension of $a = 1.0668$ cm, the higher-order passbands for an LPF characteristic with a cutoff frequency of 23.5 GHz are calculated as shown in Table 12.4.

These higher-order passbands are shown in Figure 12.22 for the TE_{20} and TE_{30} modes. If these modes are incident at the input of the LPF, in theory, they should propagate through without attenuation within their respective passbands. Ideally, there should be no such incident modes,

Table 12.4 Bands of propagation for higher-order waveguide modes.

n	Mode TE_{n0}	Lower band edge f_{0n} (GHz)	Upper band edge f_{cn} (GHz)
1	TE_{10}	14.06	23.50
2	TE_{20}	28.12	33.84
3	TE_{30}	42.18	46.19
4	TE_{40}	56.24	59.31

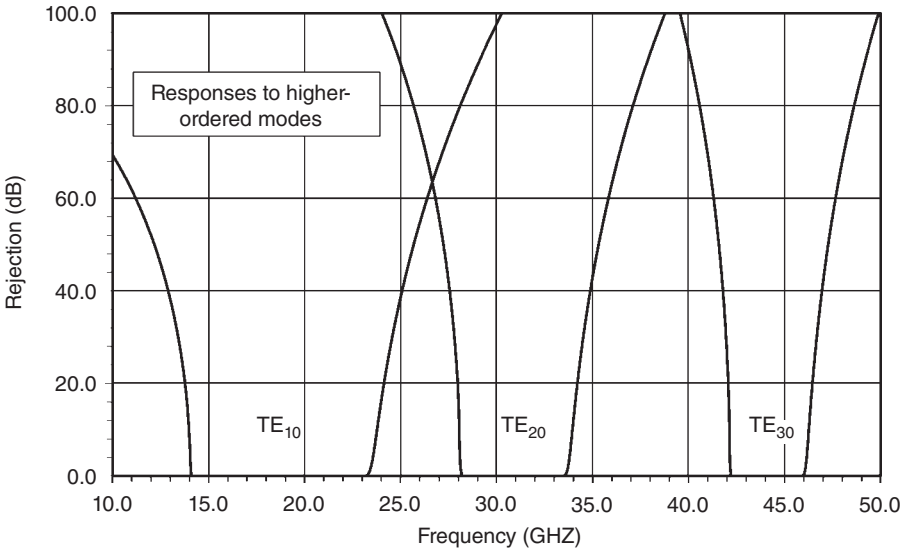


Figure 12.22 Theoretical passbands of the lowpass filter to higher-order TE modes. The TE_{10} mode forms the primary passband.

but they can be generated by certain coaxial-to-waveguide launch methods, waveguide bends, and misalignments between sections of waveguides. The effects can be reduced to a certain extent by tapering the inner sections of the filter to scatter the higher passband responses, but the amount of tapering is limited by the need to avoid creating sections within the filter that are cut off to the fundamental mode.

12.5.3.5 Mode Conversion

The internal generation and propagation of higher-order modes can severely limit the performance of LPFs in coaxial or waveguide, if the dimensions of the inner regions of the LPF allow them to exist. The higher ordered modes are generated by conversion of the fundamental mode at discontinuities within the filter and carry energy through the filter before reconverting to the fundamental mode nearer to the output. In effect the rejection properties of the LPF are being bypassed.

These higher-order modes can be substantially reduced by ensuring that the widths and heights of all sections within the structure are no greater than those of the incoming waveguide. Usually, this necessitates the introduction of step transformers at either end to reduce the height of the waveguide at the input. Because all of the internal heights are referred to the height of this waveguide feeding the first LPF iris, this dimension is chosen to ensure that all the internal heights are less than those of the main waveguide from the RF source.

Scaling the internal heights means that the minimum heights within the filter, usually those of capacitive irises near the center, are also reduced, compromising power handling capability. This is where the advantages of using a Zolotarev LPF function becomes apparent; for the same degree of filter function, the gap sizes tend to be greater than for the Chebyshev, as well as rather less variation in the heights of the corrugations.

12.5.3.6 Example of Synthesis

To illustrate the procedure, the design of the prototype circuit for a degree 17 Chebyshev lumped/distributed LPF with a RL level of 26 dB and the cutoff angle $\theta_c = 22°$ is traced. The

Table 12.5 Poles and zeros of 17th degree Chebyshev reflection function $S_{11}(\omega)$ in ω and t planes.

ω Plane		t Plane	
	Zeros [roots of $F(\omega)$]	**Zeros [roots of $F(t)$]**	**Poles [roots of $E(t)$]**
1	± 0.9958	$\pm j0.4020$	$-0.0095 \pm j0.4130$
2	± 0.9622	$\pm j0.3864$	$-0.0276 \pm j0.3965$
3	± 0.8960	$\pm j0.3563$	$-0.0437 \pm j0.3649$
4	± 0.7993	$\pm j0.3138$	$-0.0568 \pm j0.3206$
5	± 0.6753	$\pm j0.2615$	$-0.0669 \pm j0.2665$
6	± 0.5281	$\pm j0.2018$	$-0.0739 \pm j0.2052$
7	± 0.3626	$\pm j0.1371$	$-0.0786 \pm j0.1392$
8	± 0.1845	$\pm j0.0693$	$-0.0811 \pm j0.0703$
9	0.0000	0.0000	$-0.0819 \pm j0.0000$

LPF is realized in waveguide WR42 ($1.0668 \times 0.4318\,\text{cm}^2$) with iris lengths of 0.1 cm and a design cutoff frequency of 23.5 GHz. The input/output to the LPF is in reduced-height (0.2 cm) waveguide.

In this case, the ω-plane position of the half-zero is at $\omega = a = 1/\sin\theta_c = 2.6695$. By using the procedure given in Section 12.2.1, the zeros of the reflection function $S_{11}(\omega)$ are generated and are listed in the second column of Table 12.5. The third column gives the zeros after mapping to the t plane by using equation (12.11b). The final column provides the reflection poles [denominator of $S_{11}(t)$], derived by using the alternating singularity principle.

With the t-plane poles and zeros of the reflection function, polynomials $E(t)$ and $F(t)$ are computed by using equation (12.41), and the synthesis procedure for the lumped/distributed ladder network is applied. The resultant prototype values for the admittances of the shunt capacitors

Table 12.6 Prototype values for degree 17 Chebyshev lumped/distributed lowpass filter network.

	Synthesized admittances			**Element lengths (°)**	
Element i	**Shunt capacitor Y_{si}**	**Double UE $1/Z_i$**	**Capacitive iris admittance $1/Z_{0i}$**	**Capacitive iris**	**Double UE**
1, 17	1.8532	—	2.2329	22.6190	—
2, 16	—	0.4948	—	—	40.2709
3, 15	4.0907	—	4.3459	22.6190	—
4, 14	—	0.4219	—	—	41.8787
5, 13	4.5219	—	4.7863	22.6190	—
6, 12	—	0.4086	—	—	42.0657
7, 11	4.6289	—	4.8965	22.6190	—
8, 10	—	0.4051	—	—	42.1092
9	4.6533	—	4.9217	22.6190	—

Table 12.7 Lumped/distributed lowpass filter element values and electrical lengths corrected for the fringing capacitance and proximity effects, and dimensions.

Element i	Capacitive irises		Double UE		Heights (cm)	Lengths (cm)	Widths (cm)
	Y_{Si}	θ_i (°)	Y_i	θ_i (°)			
1, 17	1.7021	22.6190	—	—	0.1175	0.1000	1.0668
2, 16	—	—	0.4948	39.2847	0.4042	0.1737	1.0668
3, 15	2.7701	22.6190	—	—	0.0722	0.1000	1.0668
4, 14	—	—	0.4219	40.5960	0.4741	0.1795	1.0668
5, 13	3.0036	22.6190	—	—	0.0666	0.1000	1.0668
6, 12	—	—	0.4086	40.8383	0.4894	0.1805	1.0668
7, 11	3.0658	22.6190	—	—	0.0652	0.1000	1.0668
8, 10	—	—	0.4051	40.9016	0.4937	0.1808	1.0668
9	3.0804	22.6190	—	—	0.0649	0.1000	1.0668

and double UEs are given in columns 2 and 3 of Table 12.6. Column 4 lists the characteristic admittances of the capacitive irises, calculated from the shunt capacitance values Y_{Si}, but ignoring for the moment, the fringing capacitances B_1 and B_2. The corresponding electrical lengths of the elements are given in columns 5 and 6. Note that the lengths of the double UEs are less than $2\theta_c = 44°$, due to the difference in the phase reference planes of the shunt capacitors and the capacitive irises.

Table 12.7 summarizes the characteristic admittances of the irises after correction for the fringing capacitances and proximity effects, which also means adjustments to the lengths of the double UEs. The final three columns give the actual dimensions of the elements, where the heights are referred to the incoming waveguide b dimension (0.2 cm), and the lengths are calculated from the wavelength in WR42 at 23.5 GHz. The widths here are constant, but can be tapered to scatter the higher-frequency spurious responses (Figure 12.23).

Analyzing these dimensions with EM-mode matching software, we obtain the RL and rejection characteristics as shown in Figure 12.24. It may be seen that the RL is at or above the design

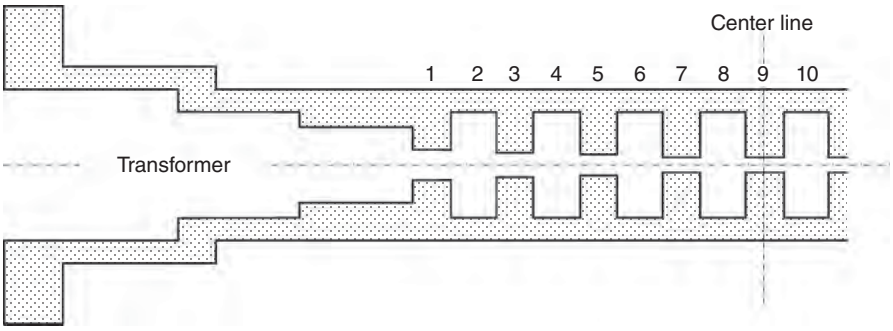


Figure 12.23 Seventeenth-degree lumped/distributed lowpass filter with transformers.

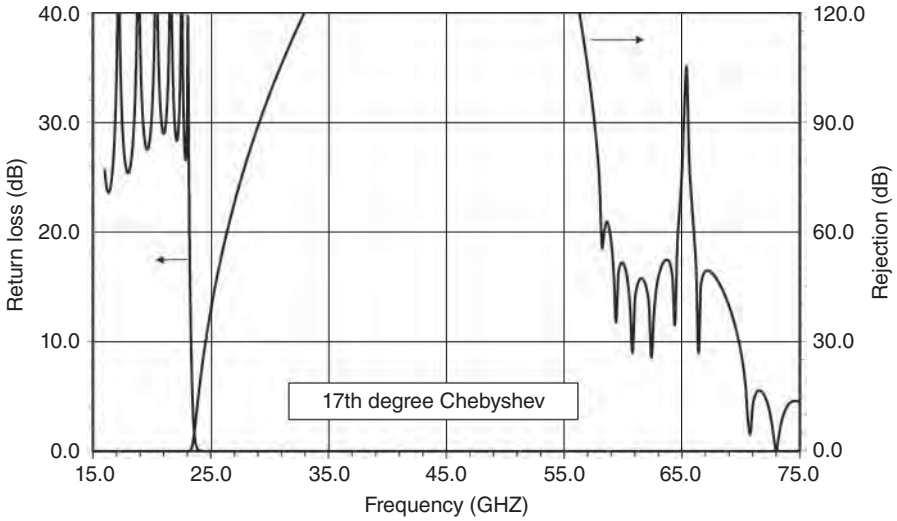


Figure 12.24 Mode-matching analysis of the 17th-degree Chebyshev lumped/distributed LPF.

26 dB for about 6 GHz within the passband. An outline sketch of the 17th-degree Chebyshev lumped/distributed LPF with transformers is shown in Figure 12.23.

Summary

Synthesis methods in Chapters 7–9 are based on lumped-element parameters. The prototype models derived from these techniques are suitable for realizing narrowband (typically < 2% bandwidth), bandpass, and bandstop filters at microwave frequencies. In the microwave region, say, between 1 and 12 GHz, a 2% bandwidth translates to 20 and 240 MHz, respectively. The LPFs, which are required to suppress the harmonics generated by the high-power amplifiers, require bandwidths in the GHz range. As a consequence, the lumped-element prototype model is not applicable for the design of such LPFs. It is necessary to introduce distributed elements in the prototype filter, leading to the so-called distributed LPP model.

This chapter is devoted to the synthesis techniques for the realization of LPFs at microwave frequencies. The chapter begins with a review of the properties of distributed circuit elements, basically lengths of transmission line called a *unit element*. It should be noted that there is no lumped-element equivalent to a UE. At the prototype design stage, these UEs can differ in length, but more commonly, they are chosen to be the same length, the "commensurate-line" length. The class of transfer and reflection polynomials appropriate for realization in the commensurate-line circuit are then introduced.

Chebyshev functions of the second kind and the Achieser–Zolotarev function represent the prime candidates for derivation of the characteristic polynomials for a LPF design. A method for evaluation of such functions is described. This is followed by the detailed procedure for the synthesis of the distributed SI LPF and the mixed lumped/distributed LPF. Examples of SI and lumped/distributed (tapered–corrugated design) filters in waveguide structures are included to illustrate the design procedure. Brief discussions of the spurious response of such filters and methods to minimize it are included.

References

1 Levy, R. and Cohn, S.B. (1984) A history of microwave filter research, design and development. *IEEE Transactions on Microwave Theory and Techniques*, **32**, 1055–1066.
2 Levy, R. (1970) Generalized rational function approximation in finite intervals using Zolotarev functions. *IEEE Transactions on Microwave Theory and Techniques*, **18**, 1052–1064.
3 Levy, R. (1965) Tables of element values for the distributed low-pass prototype filter. *IEEE Transactions on Microwave Theory and Techniques*, **13**, 514–536.
4 Levy, R. (1970) A new class of distributed prototype filters with application to mixed lumped/distributed component design. *IEEE Transactions on Microwave Theory and Techniques*, **18**, 1064–1071.
5 Levy, R. (1973) Tapered corrugated waveguide low-pass filters. *IEEE Transactions on Microwave Theory and Techniques*, **21**, 526–532.
6 Levy, R. (1971) Characteristics and element values of equally terminated Achieser–Zolotarev quasi-lowpass filters. *IEEE Transactions on Circuit Theory*, **18**, 538–544.
7 Richards, P.I. (1948) Resistor-transmission-line circuits. *Proceedings of the IRE*, **36**, 217–220.
8 Rhodes, J.D. and Alseyab, S.A. (1980) The generalized Chebyshev low-pass prototype filter. *IEEE Transactions on Circuit Theory*, **8**, 113–125.
9 Hunter, I.C. (2001) *Theory and Design of Microwave Filters*, Electromagnetic Waves Series 48, IEE, London.
10 Marcuvitz, N. (1986) *Waveguide Handbook*, Electromagnetic Waves Series 21, IEE, London.
11 Levy, R. (1973) A generalized design technique for practical distributed reciprocal ladder networks. *IEEE Transactions on Microwave Theory and Techniques*, **21**, 519–525.
12 Matthaei, G.L. (1966) Short-step Chebyshev impedance transformers. *IEEE Transactions on Microwave Theory and Techniques*, **14**, 372–383.
13 Somlo, P.I. (1967) The computation of coaxial line step capacitances. *IEEE Transactions on Microwave Theory and Techniques*, **15**, 48–53.
14 Green, H.E. (1965) The numerical solution of some important transmission line problems. *IEEE Transactions on Microwave Theory and Techniques*, **13**, 676–692.
15 Saal, R. and Ulbrich, E. (1958) On the design of filters by synthesis. *IRE Transactions on Circuit Theory*, **5**, 284–327.
16 Bingham, J.A.C. (1964) A new method of solving the accuracy problem in filter design. *IEEE Transactions on Circuit Theory*, **11**, 327–341.
17 Orchard, H.J. and Temes, G.C. (1968) Filter design using transformed variables. *IEEE Transactions on Circuit Theory*, **15**, 385–407.

13

Waveguide Realization of Single- and Dual-Mode Resonator Filters

In the preceding chapters, we have studied methods for generation of the polynomials for filtering functions that have the widest practical application in today's microwave systems for terrestrial and airborne telecommunications, radar, and earth observation and scientific systems. In the great majority of cases, the Chebyshev equiripple filtering function is employed, as it has the optimal balance between in-band linearity, close-to-band selectivity, and far-out-of-band rejection. Specialist variants are available where high performance is required—asymmetric, group delay equalized, prescribed transmission zeros (TZs), and combinations of the foregoing. Other classes of filtering function were briefly touched on—including elliptic Butterworth, where very linear group delay performance is required and predistorted for linear amplitude and compact size.

From the generation of transfer and reflection polynomials, attention was then focused on exact synthesis techniques for the electrical networks corresponding to the original filtering function polynomials, concentrating on those configurations that may be realized with a microwave structure of some kind, such as the coaxial resonator or waveguide cavity resonator. From the electrical network, or directly from the transfer and electrical polynomials themselves, a coupling matrix (CM) may be generated for most configurations. Having a direct correspondence with individual elements in the filter structure, the CM is a very useful representation of the filter's electrical components, bridging the gap between the pure mathematics of the transfer/reflection polynomials and the real world of mainline couplings, resonator tuning offsets, and so on.

The prototype CM may be generated using the methods of Chapter 8 or by optimization methods. Once the CM has been generated, purely mathematical operations may be applied to it [e.g., similarity transforms (plane rotations) or "scaling at nodes"] to create, annihilate, or change coupling element values or tuning states. These operations preserve the CM's eigenvalues, and so an arbitrarily large series of them may be applied to the CM without destroying the unique one-to-one relationship between the CM elements in the eventually realized hardware.

In this chapter we discuss some of the more practical aspects of the filter design process, making use of the methods that have been developed to create useful filtering functions, then their realization in a microwave structure, and where possible to give design information relating the electrical values in the CM to actual dimensions in the filter structure.

In particular, dual-mode waveguide and dielectric realizations will be addressed. Dual-mode filters have found very wide application in both (1) the low-power channel filtering stages of a telecommunication system, where high selectivity and in-band linearity are required, and (2) the high-power output stages, where low loss is a priority. For satellite systems in particular, the advanced electrical performance has to be combined with lowest possible mass and volume, high stability of the electrical performance in the presence of environmental temperature fluctuations, and (for output filters) high power handling capability (conduction of dissipated

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

RF energy to a cooling plate, ability to withstand multipactor breakdown in vacuum, corona in partial pressure conditions). Moreover, the filters and multiplexer assemblies need to be robust enough to survive the rigors of the launch (vibration, shock).

Since their introduction to practical use in the early 1970s by Atia and Williams [1–4], waveguide dual-mode filters have been developed by the space industry [5–8] and refined to deliver advanced performance in the harshest of environmental conditions. More recently the methods for realizing asymmetric filtering characteristics without an accompanying increase in complexity have been introduced [9]. Asymmetric responses are useful for correcting the inherent group delay slopes that appear in wideband filters, or some of the distortions encountered in the channels at the edge of a contiguous group in an output multiplexer, or providing an efficient "tailor-made" rejection characteristic where rejection specifications are themselves asymmetric.

13.1 Synthesis Process

In Chapter 7, a typical process for the design and realization of a microwave Chebyshev bandpass filter in coaxial resonator technology was described (see Figure 7.5). The procedure can be broadly broken up into three stages:

1) Design and optimization of the filter function (polynomials, singularities) by using a suitable simulator to ensure compliance with the required specifications.
2) Synthesis of the prototype electrical network, usually in the form of a CM. The elements of the CM correspond to individual physical elements or resonator tuning states in the filter structure—diagonal elements represent tuning offsets (if any); off-diagonal elements correspond to coupling apertures, probes, loops or screws, or input/output couplings.
3) Lowpass prototype to RF bandpass mapping, dimensioning of resonant cavities, coupling apertures, and inclusion of dispersions (frequency dependencies) specific to each individual coupling element and the unloaded Q factor (Q_u) specific to each resonator.

13.2 Design of the Filter Function

Design of the filter function and tradeoffs are described in Sections 3.7, 4.4, and 6.2. In Chapter 6, the design method for the general Chebyshev filter function is illustrated by including an example. The general Chebyshev function, in conjunction with TZs, is most often used in the design of filters for a wide range of applications. The prescribed TZs can be on the imaginary axis of the complex plane (creating TZs and rejection lobes in the filter's transfer characteristic) or on the real axis (group delay equalization of symmetric characteristics) or complex (group delay equalization of symmetric or asymmetric characteristics), or combinations of the above. The recursion method described in Chapter 6 for the generation of Chebyshev polynomials will deal with TZ numbers up to the maximum possible (N zeros in an Nth-degree function, i.e., fully canonical), and by bypassing the recursion routine for the generation of the reflection function numerator polynomial $F(s)$ and putting all the reflection zeros at $s = 0$ [i.e., $F(s) = s^N$], Butterworth filtering functions may also be generated.[1]

1 Pure Butterworth functions are sometimes used when high far-out-of-band rejection slopes are required. Butterworth filters have the highest asymptotic rejection slopes of any filtering function—6 N dB per octave. Also, because of the relatively slow cutoff rates near to the band edges, Butterworth filters have relatively linear in-band group delay characteristics and are sometimes used where exceptionally flat group delay performance over a greater part of the passband is required.

When TZs are required, it is necessary to optimize the location of these zeros with respect to the specification requirement of the amplitude or group delay or both.

13.2.1 Amplitude Optimization

In Chapter 3, we described the concept of unified design charts (UDCs) to design filters on the basis of their amplitude response. For the Butterworth, Chebyshev, and elliptic function filters, UDCs can be readily generated and are included as Appendix 3A. Use of these charts provides a good first approximation in optimizing the amplitude response of a filter. In Chapter 4, we introduced a computer-aided optimization procedure to generate filter functions with arbitrary amplitude response. A subset of this procedure is the ability to optimize the amplitude response of a filter function, given its degree and the number of TZs. This implies optimization of the location of the transmission and reflection zeros to achieve the desired filter response.

The flow of the process to generate characteristics with prescribed rejection lobe levels and/or in-band group delay equalization is not always routine, depending sometimes on the particular filtering function under consideration. The process is often a matter of personal preference for the individual designer, but one method for optimizing rejection lobe levels and for group delay equalization is suggested here.

13.2.2 Rejection Lobe Optimization

For rejection lobes a cost function may be built up from the difference between the target lobe levels that have been set and those actually analyzed on the prototype. Because $E(s)$ is a complex polynomial with its zeros in the complex plane, its turning points (which indicate the frequency positions of the rejection lobes) cannot be directly derived by differentiating the function $S_{21}(s) = P(s)/\varepsilon E(s)$. However all the turning points may be found by differentiating the scalar power function

$$S_{21}(s)S_{21}(s)^* = \frac{P(s)P(s)^*}{P(s)P(s)^* + \varepsilon^2 F(s)F(s)^*/\varepsilon_R^2} \tag{13.1}$$

Because the zeros of $P(s)$ lie in mirror-image pairs about the imaginary axis or on the imaginary axis itself, $P(s)^* = -(-1)^N P(s)$ (see Section 6.1.2), and so differentiating equation (13.1) results in a polynomial quotient, the numerator of which may be determined as $(-1)^N [P^2(FF'^* + F'F^*) - 2PP'FF^*]/(\varepsilon\varepsilon_r)^2$. The prime indicates differentiation with respect to s. The zeros of this general expression will correspond to the in-band and out-of-band turning points of the power characteristic, including fully canonical and asymmetric types, and those with off-axis reflection zeros. If it is now assumed that the zeros of the reflection polynomial $F(s)$ are all positioned on the imaginary axis (and are therefore coincident with the zeros of $F_{22}(s)$), then $F(s)^* = (-1)^N F(s)$ and the numerator may be simplified as follows:

$$2P(s)F(s)[P(s)F'(s) - P'(s)F(s)]/(\varepsilon\varepsilon_r)^2 \tag{13.2}$$

By equating (13.2) to zero and rooting it, all the $2(N + n_{fz}) - 1$ turning points of the power transfer function may be found.

However, the computational effort may be reduced by recognizing that the polynomial $P(s)$ contains the n_{fz} turning points associated with the positions of the n_{fz} TZs; likewise, $F(s)$ contains the N turning points associated with the positions of the N reflection zeros (points of perfect transmission), and so the remaining $(N + n_{fz}) - 1$ turning points may be found by solving the polynomial equation

$$F'(s)P(s) - P'(s)F(s) = 0 \tag{13.3}$$

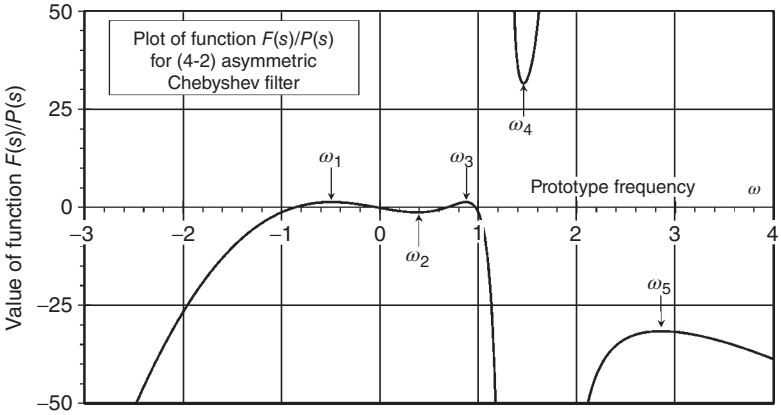


Figure 13.1 Plot of function $F(s)/P(s)$ for a (4–2) asymmetric filtering function.

which is the same as evaluating the following differential equation:

$$\frac{d}{ds}\left(\frac{F(s)}{P(s)}\right) = 0 \tag{13.4}$$

Thus it is necessary only to construct the polynomial function (13.3) and root it to find the $(N + n_{fz}) - 1$ turning points and then to select the out-of-band real frequencies that will correspond to the rejection lobe positions. Evaluating $-20\log(S_{21}(s))$ at these out-of-band turning frequencies will give the levels of the associated rejection lobes in dB. For a Chebyshev function, there will be $N - 1$ in-band real frequency turning points, corresponding to the points of reflection maxima within the passband. Figure 13.1 gives a plot of the function $F(s)/P(s)$ for the (4–2) asymmetric filtering function that was designed in Chapter 6, indicating the three in-band turning points and the two rejection lobe positions.

This method of finding out-of-band rejection lobe levels involves rooting polynomials of degree $(N + n_{fz}) - 1$ instead of $2(N + n_{fz}) - 1$ and so will be faster and more accurate. For fully canonical functions ($n_{fz} = N$), the highest-degree coefficients will cancel in (13.3), and there will be one less root in total $(2N - 2)$. This means that if all N TZs of the fully canonical filtering function are on the imaginary axis, the rejection characteristic will have $N - 1$ lobes produced by the N TZs, in contrast to the n_{fz} lobes that are produced by n_{fz} TZs in a noncanonical filtering function.

The method depends on the polynomials $F(s)$ and $P(s)$ having symmetry about the imaginary axis. For $P(s)$ this will always be the case, but for the $F(s)$ polynomial, the restriction must be applied that all its zeros must lie on the imaginary axis itself. It may be recalled that the same constraint applied when using the alternating pole principle (Chapter 6). An alternative method for the calculation of the positions of rejection lobes for Chebyshev characteristics is presented in Section 6.3.4.

By associating the calculated lobe levels with their associated target levels, the cost function may be evaluated. The optimizer may then move the positions of the TZs, the new $F(s)$ polynomial may be determined, and then the new lobe levels and cost function may be evaluated. If there is an improvement from the previous iteration, the zero positions may be moved again in the same direction, and so on until the cost function is arbitrarily close to zero.

Constraints must be placed on the optimizer to prevent it from stepping the TZs too close to the band edge, or within the passband itself ($|s| > j$). Otherwise the process seems to be stable and converges rapidly even for fully canonical functions. The (4–2) asymmetric filtering

function with 30 dB lobe levels that has been used as an example in previous chapters was designed using this method.

If the computer optimization routines being used to optimize the amplitude function require gradients with respect to the TZ locations, these may be computed analytically as described in Section 4.6. The relationships are summarized as follows:

$$\frac{\partial T}{\partial b_i} = -40 \log e \cdot |\rho|^2 \cdot \frac{b_i}{(s^2 \pm b_i^2)}$$

$$\frac{\partial R}{\partial b_i} = 40 \log e \cdot |t|^2 \cdot \frac{b_i}{(s^2 \pm b_i^2)}$$

where the zero locations are at $s^2 = \pm b_i^2$. The positive sign applies to the zeros on the imaginary axis, and the negative sign applies to zeros located along the real axis. T and R represent the transmission and reflection loss in dB, respectively. In terms of the scattering parameters, $T = -10 \log |S_{21}(s)|^2$ dB and $R = -10 \log |S_{11}(s)|^2$ dB. The complex zeros, located symmetrically about the imaginary axis, are defined as

$$\begin{aligned} Q(s) &= (s + \sigma_1 \pm j\omega_1)(s - \sigma_1 \pm j\omega_1) \\ &= s^2 \pm j2s\omega_1 - (\sigma_1^2 + \omega_1^2) \end{aligned}$$

The gradients with respect to the complex zero locations are given by

$$\frac{\partial T}{\partial \sigma_1} = -40 \log e \cdot |\rho|^2 \sigma_1 \frac{(s^2 - \sigma_1^2 - \omega_1^2)}{|Q(s)|^2}$$

and

$$\frac{\partial R}{\partial \omega_1} = -40 \log e |t|^2 \omega_1 \frac{(3s^2 - \sigma_1^2 - \omega_1^2)}{|Q(s)|^2}$$

The inclusion of these analytic gradients improves the efficiency of the optimization process.

13.2.3 Group Delay Optimization

Phase and group delay responses of minimum phase filters are uniquely related to their amplitude response [10]. Minimum phase filters are characterized by the distribution of the zeros of the transfer function along the imaginary axis (elliptic and quasielliptic filters) or at infinity (maximally flat and Chebyshev filters). For such filters, once the amplitude response is optimized, we must accept the group delay response that is unique to that amplitude response. It is possible to improve the group delay without affecting the amplitude by the use of all-pass equalizers that are discussed in Chapter 17.

An alternative way to improve the group delay response is by locating some of the zeros of the transfer function [zeros of $P(s)$] along the real axis or as pairs distributed symmetrically about the imaginary axis. Zeros of $P(s)$ must have symmetry about the imaginary axis to satisfy the realizability conditions, as described in Sections 3.11 and 6.1. Such filters are referred to as *nonminimum phase*, *linear phase*, or *self-equalized filters*. The amplitude and phase of this class of filters are no longer uniquely related. Linear phase filters can be optimized to provide an improved group delay performance at the expense of a poorer amplitude response. This implies that we need to resort to a higher-degree filter to obtain the desired amplitude response and improved group delay. As always, there is a tradeoff between the filter complexity and performance. In order to identify different types of zeros of $P(s)$, often referred to as TZs, a notation that is widely used indicates the degree of the filter, TZs along the imaginary axis followed by TZs elsewhere. As an example, (10–2–4) represents a 10th-degree filter with two TZs along the

$j\omega$ axis (not necessarily symmetrically located) and four TZs distributed symmetrically about the $j\omega$ axis. The last index in this notation is always an even integer.

Next, the problem is to determine the location of the zeros to minimize the group delay variations. Because they exist only in pairs symmetrically located about the imaginary axis, the zeros themselves do not contribute to the overall phase characteristic of the filter as s varies along the imaginary axis (except for integer multiples of π radians) and therefore have no effect on group delay. However, the introduction of zeros changes the characteristic polynomials and thus the transfer and reflection performance of the filter. Polynomials $F(s)$ and $E(s)$ need to be regenerated in order to restore the original features of the filter function (e.g., equiripple passband in the case of the Chebyshev family). As described in Section 3.4, the phase and group delay characteristics are determined entirely by the $E(s)$ polynomial.

13.2.3.1 Group Delay and Attenuation Slope Calculation

The group delay and attenuation slope characteristics of a filter function can be conveniently calculated from the transfer polynomials as follows.

If $S_{21}(s) = P(s)/\varepsilon E(s)$ is a transfer function and $S'_{21}(s)$ is its first derivative with respect to the frequency variable s, then the real part of the quantity $-S'_{21}(s)/S_{21}(s)$ is the group delay function $\tau(s)$ and the imaginary part is the gain slope of the transfer characteristic. The proof follows directly from the fact that

$$-\frac{S'_{21}(s)}{S_{21}(s)} = -\frac{d}{ds}[\ln S_{21}(s)] = -\frac{d}{ds}[\ln(|S_{21}(s)|e^{j\phi})]$$
$$= -\frac{d}{ds}[\ln|S_{21}(s)| + j\phi]$$

where ϕ is the phase of $S_{21}(s)$. Because $s = j\omega$ and therefore $ds = jd\omega$, it follows that

$$-\frac{S'_{21}(s)}{S_{21}(s)} = j\frac{d}{d\omega}[\ln|S_{21}(s)| + j\phi] = -\frac{d\phi}{d\omega} + j\frac{dA}{d\omega} \tag{13.5}$$

where $A = \ln|S_{21}(s)|$ is the gain in nepers. Thus $\mathrm{Re}[-S'_{21}(s)/S_{21}(s)]$ is the group delay in seconds and $-8.6859\ \mathrm{Im}[-S'_{21}(s)/S_{21}(s)]$ is the attenuation slope in dB per rad/s (radians per second). It can be easily shown that

$$-\frac{S'_{21}(s)}{S_{21}(s)} = \frac{E'(s)}{E(s)} - \frac{P'(s)}{P(s)} \tag{13.6}$$

where $E'(s)$ and $P'(s)$ are the differentials of the polynomials $E(s)$ and $P(s)$ with respect to s. Because any TZs are either on the imaginary axis or are in mirror-image pairs symmetrically positioned about the imaginary axis, the second term in (13.6) will always be purely imaginary and need not be calculated to obtain group delay values.

13.2.3.2 Group Delay Optimization Procedure

Although methods exist to analytically design certain types of delay-equalized filters (also known as *linear phase filters*) [10, 11], the most generally useful procedure is with optimization. With the speed and power of today's PCs, the optimization process converges within a few seconds, even for the most complex cases. Moreover, with the optimization method, it is possible to simultaneously optimize TZs for prescribed rejection lobe levels and to deal with asymmetric cases. A flow diagram showing one possible organization of the group delay optimization process is given in Figure 13.2.

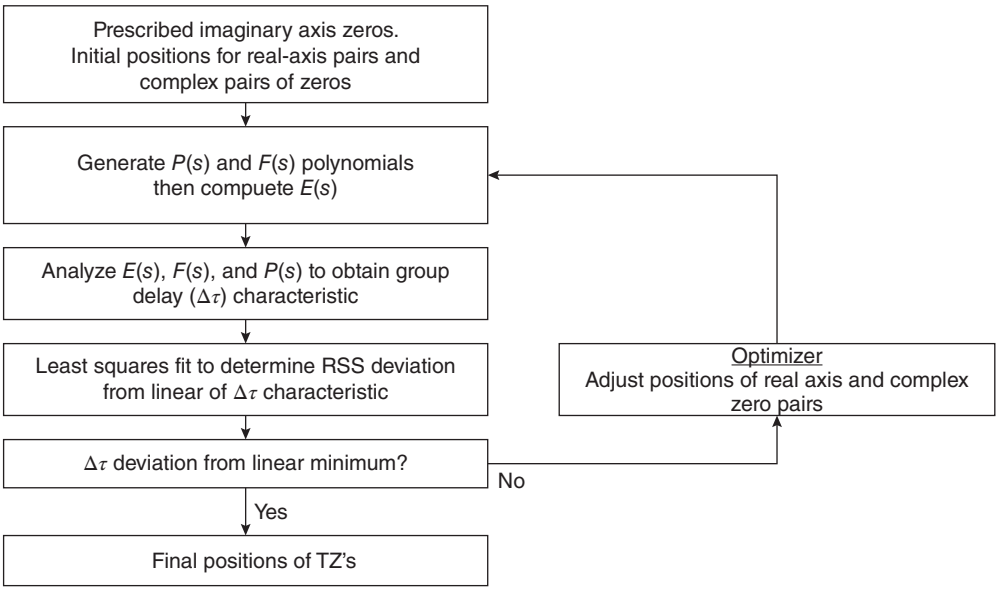


Figure 13.2 Flow diagram for the group delay equalization process.

The procedure may be summarized as follows:

1) Check that the total number of TZs (for group delay equalization and rejection lobes) complies with the minimum path rule for the intended filter topology. The next step is to select a fraction of the bandwidth over which group delay equalization is to be carried out, 50% for a real-axis pair, and 65–70% for a quad is usually realistic. A certain amount of trading off may be done between the equalization bandwidth and the amplitude of the group delay ripple within that bandwidth.
2) Postulate initial positions for the complex TZs: $\pm 0.8 + j0.0$ for real-axis pairs and $\pm 0.8 \pm j0.5$ for quads are generally successful for symmetric characteristics.
3) With the prescribed TZ positions, form the $P(s)$ polynomial, and generate the $F(s)$ and $E(s)$ polynomials. This is accomplished by optimizing the amplitude response using the analytic gradients and optimization procedure as explained in Chapter 4. For Chebyshev-type filters, it is more expedient to generate $F(s)$ and $E(s)$ polynomials by using the method described in Chapter 6.
4) Calculate the group delay τ_i at a number N_s sample points s_i across the equalization bandwidth (or half the bandwidth from the center to the edge for symmetric characteristics). With polynomial $E(s)$, the group delay at a sample frequency s_i is calculated from equation (13.6):

$$\tau_i = \mathrm{Re}\left[\frac{E'(s_i)}{E(s_i)}\right] \tag{13.7}$$

where the prime denotes differentiation.
5) Calculate the least-squares best-fit straight line (BFSL) to the data:

$$\tau(\omega) = m\omega + c$$

$$\text{where } m = \frac{\sum_{i=1}^{N_S} \tau_i(\omega_i - \omega_{av})}{\sum_{i=1}^{N_S} (\omega_i - \omega_{av})^2}$$

$$c = \tau_{av} - m\omega_{av}$$

$$\omega_{av} = \text{average of the frequency sampling point } s = \left[\sum_{i=1}^{N_S} \omega_i\right] / N_S$$

$$\tau_{av} = \text{average delay} = \left[\sum_{i=1}^{N_S} \tau_i\right] / N_S$$

$$\omega_i = -js_i = \text{frequency of the } i\text{th sample point}$$

$$m = s\text{lope of the BFSL, } c \text{ is its constant, and } N_S \text{ is the number of sample points} \tag{13.8}$$

Once the slope and the constant of the BFSL have been computed, the root-mean-square (rms) deviation of the group delay data from the BFSL is calculated by

$$\Delta\tau_{\text{rms}} = \left[\sum_{i=1}^{N_S} [\tau_i - (m\omega_i + c)]^2\right]^{1/2} \tag{13.9}$$

Both $\Delta\tau_{\text{rms}}$ and the slope of the BFSL m can be built into the error function to be minimized by the optimization routine to be as close to zero as possible. In addition, the slope bias can be built into the error function to counter any dispersion effects introduced by the technology that the filter is realized in, for example, a positive slope to counter the natural negative group delay slope in an iris-coupled waveguide bandpass filter. With the appropriate weightings, the imaginary-axis TZs can also be built into the error function along with the complex and real-axis equalization pairs to form a symmetric or asymmetric quasielliptic self-equalized filter function. As an example, the response of an eighth-degree filter with 23 dB return loss is given in Figure 13.3. Here, there is a group delay slope of 1 s over the prototype bandwidth and two symmetric amplitude lobes of 40 dB each on either side of the passband. This is achieved with two TZs at $+j1.3527$ and $-j1.3692$ and a group delay self-equalization pair at $\pm 0.8592 + j0.1091$.

In some cases, it is possible to optimize the characteristic with the CM model of the filter rather than the poles and zeros of the filter function. Because of the direct correspondence between the CM entries and the actual coupling devices in the filter, individual dispersion behavior can be assigned to each coupling, and the frequency variation characteristics of each resonator can be modeled more precisely. Although the overhead of synthesizing and analyzing the matrix at each iteration adds to the computation time, a closer fit with the real performance is achieved.

13.3 Realization and Analysis of the Microwave Filter Network

Once the prototype filter function has been designed, a prototype network may be synthesized. If the filter function is an all-pole type, such as a pure Chebyshev or Butterworth with all TZs at $s = \pm j\infty$, then there are no cross-couplings, and the lowpass prototype circuit is a simple ladder structure. Typically, such all-pole networks are synchronously tuned, meaning that the FIRs at the nodes are zero in value. For filter functions with finite-position TZs, it is necessary to have cross-couplings between nonsequentially numbered resonator nodes.

In the majority of practical cases, a CM is formed, either directly from the filter function polynomials or indirectly via a prototype circuit synthesis. The circuit route is slightly more flexible, allowing the synthesis of multiple cascaded trisections and extracted poles. The configurations of both the trisection and the direct-coupled extracted pole are amenable to a representation by a CM, but only the main body of the phase-coupled extracted pole type [12] is represented by a CM.

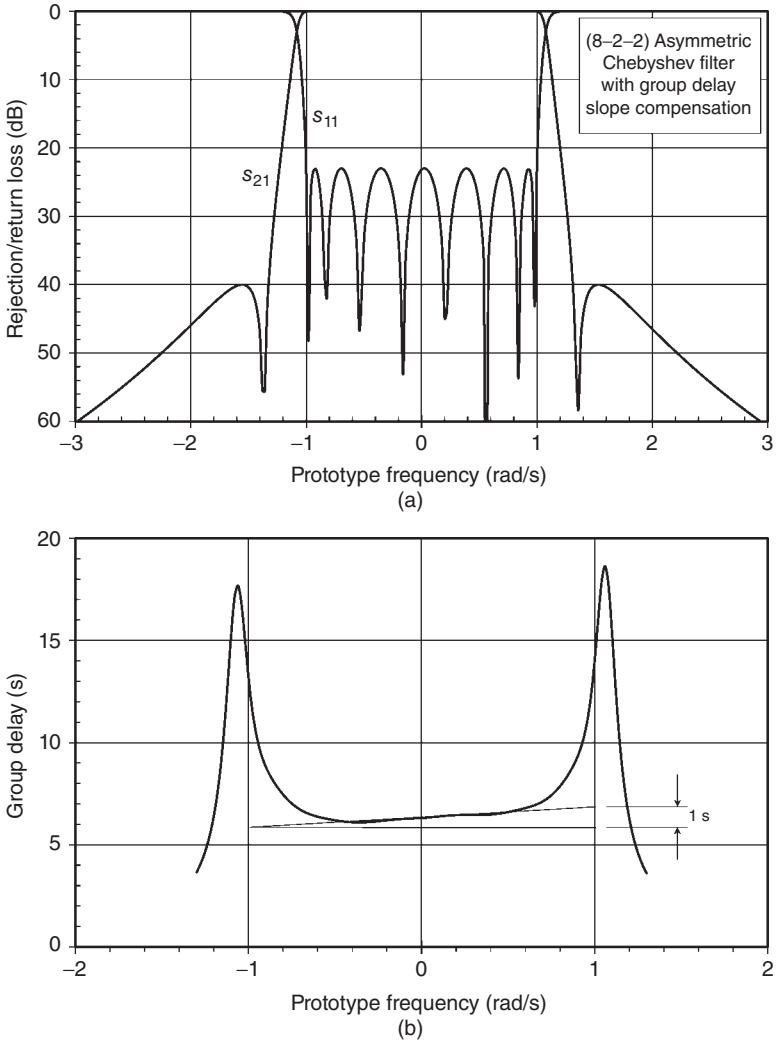


Figure 13.3 An eighth-degree filter function with two rejection lobes and a group delay compensation slope: (a) rejection and return loss; (b) group delay.

The next step in the realization process is to apply similarity transforms to the prototype coupling matrices to reconfigure them to the desired topology. Measurable electrical values for each RF coupling element of the filter can be calculated using as a basis the corresponding entry from the prototype CM. Figure 13.4 illustrates the procedure for a sixth-degree symmetric filter function with two TZs on either side of the passband creating two rejection lobes and realized in rectangular waveguide "inline" configuration.

For the rectangular waveguide example in Figure 13.4, the mainline couplings are realized as inductive apertures and the cross-couplings (some of which can be negative in the CM, depending on the positions and type of TZs) as inductive apertures or loops (mainly in coaxial resonator filters), or capacitive probes if negative. The next step in the realization process for the couplings is to assign RF electrical values to them, which are then dimensioned either by measurements on the test sample pieces or by electromagnetic techniques discussed in later chapters.

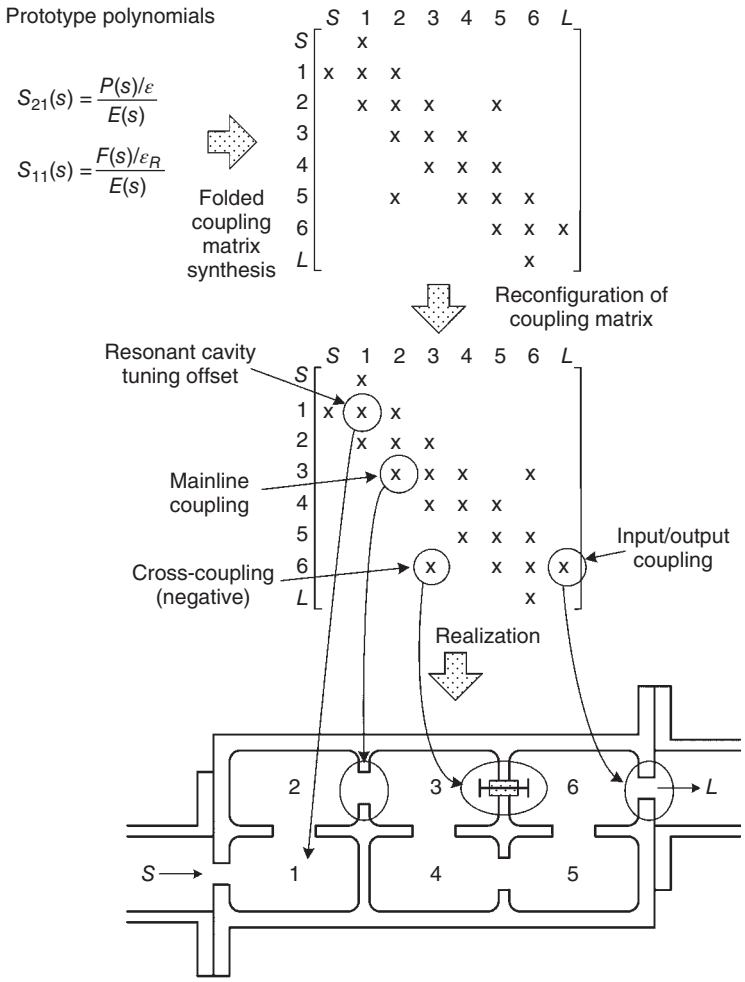


Figure 13.4 Realization process, from prototype filtering function via a coupling matrix to the realization in a rectangular waveguide.

For a waveguide coupling in the direction of propagation of the waveguide, the shunt inductance is a useful equivalent to the impedance or admittance inverter taken from the CM. Matthaei *et al*. [13] show, using matrix matching techniques, that the electrical equivalent of an impedance inverter is a shunt inductor X symmetrically placed in a length ϕ of transmission line of characteristic impedance Z_0 (see Figure 13.5). (The values of all the impedances or admittances of the inverters or lumped components are normalized to the characteristic impedance Z_0 or admittance Y_0, respectively.)

The phase length ϕ is found by equating the "A" elements of the [*ABCD*] matrices of the two circuits:

$$\phi = -\tan^{-1} 2X = -\cot^{-1} \frac{B}{2} \tag{13.10a}$$

and the susceptance B is found by equating and then subtracting the "*B*" and "*C*" elements:

$$B = \frac{1}{|X|} = K - \frac{1}{K} \tag{13.10b}$$

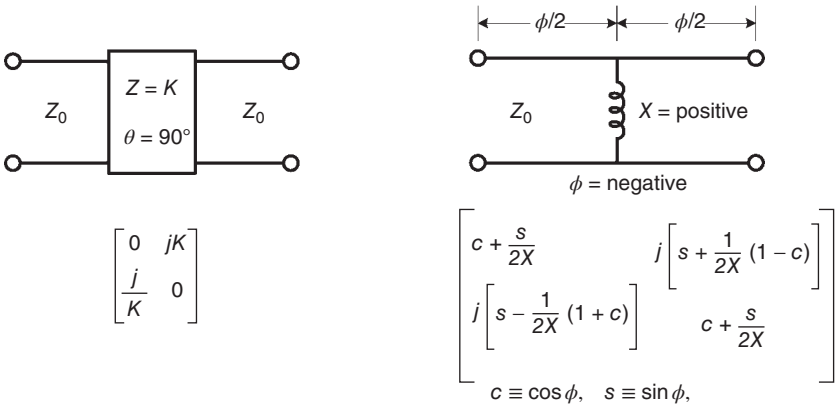


Figure 13.5 Equivalence of impedance inverter and shunt reactance symmetrically placed in a length of transmission line.

This equivalent circuit for waveguide coupling is particularly useful because the lengths of line can be absorbed into the adjacent line lengths (the lengths of waveguide that form the resonant cavities), shortening them slightly. Moreover, the RF susceptance B of a waveguide aperture or iris is a directly measurable parameter and is directly related to the prototype coupling values taken from the CM.

For input and output couplings [13],

$$K_{S1} = \frac{M_{S1}}{\sqrt{\alpha}}, \quad K_{LN} = \frac{M_{LN}}{\sqrt{\alpha}} \tag{13.11a}$$

and for the interresonator couplings,

$$K_{ij} = \frac{M_{ij}}{\alpha} \tag{13.11b}$$

where M_{ij} are the coupling values taken from the prototype CM; K_{ij} are the RF inverter values; $\alpha = (\lambda_{g1} + \lambda_{g2})/[n\pi(\lambda_{g1} - \lambda_{g2})]$, λ_{g1}, and λ_{g2} are the guide wavelengths at the lower and upper band-edge frequencies, respectively; and n is the number of half-wavelengths of the waveguide resonator cavity. For example, for a TE_{10} mode, resonating in a rectangular waveguide cavity that is three half-wavelengths long, the resonant mode is TE_{103}, and so $n = 3$. α is the lowpass to waveguide bandpass scaling factor[2] and is usually >1.

Equation (13.11) can now be applied to find the susceptances of the waveguide coupling elements, normalized to the waveguide admittances. For endwall couplings between resonator cavities R_i and R_j and between cavities R_j and R_k in the direction of propagation in the waveguide (see Figure 13.6), the mainline susceptances B_{ij} and B_{jk} are given by [13]

$$B_{ij} = K_{ij} - \frac{1}{K_{ij}}, \quad B_{jk} = K_{jk} - \frac{1}{K_{jk}} \tag{13.12}$$

Because α is usually >1, the RF coupling susceptances B_{ij} have negative values when the corresponding prototype coupling M_{ij} is positive (they can be represented electrically as shunt inductors or physically as inductive apertures or irises). The total electrical length of resonator cavity R_j is then given by

$$\theta_j = n\pi + \frac{1}{2}\left(\cot^{-1}\frac{B_{ij}}{2} + \cot^{-1}\frac{B_{jk}}{2}\right) \text{ radians} \tag{13.13}$$

2 For coaxial bandpass filters, α is the reciprocal of the bandwidth ratio: $\alpha = f_0/BW$.

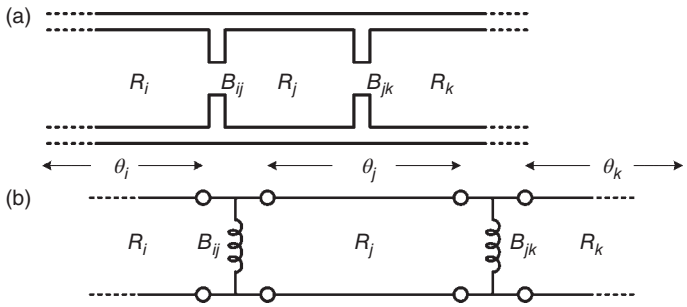


Figure 13.6 Rectangular waveguide resonators where the endwall is coupled: (a) rectangular waveguide resonator R_j; (b) open-wire electrical equivalent circuit.

Sidewall couplings between resonator cavities can be represented by a series coupling from approximately the midpoint of the two resonators (or the voltage maximum points; see Figure 13.6). The effect of the series coupling is to insert a shunt FIR approximately halfway along the length of the cavities R_j and R_l [14]. Equation (13.13) for the electrical length of the cavity can be modified to account for the sidewall coupling, as well as for the endwall couplings. Also, it is convenient now to include any tuning offset for the resonance, represented by the diagonal elements of the CM (asynchronous tuning):

$$\theta_{1j} = \frac{n\pi}{2} + \frac{1}{2}\left(\cot^{-1}\frac{B_{ij}}{2} - \sin^{-1}(B_{jl} - B_{jj})\right)$$
$$\theta_{2j} = \frac{n\pi}{2} + \frac{1}{2}\left(\cot^{-1}\frac{B_{jk}}{2} - \sin^{-1}(B_{jl} - B_{jj})\right) \tag{13.14}$$

where $B_{jl} = M_{jl}/\alpha$ is the sidewall coupling between resonant cavities R_j and R_l (see Figure 13.7), $B_{jj} = M_{jj}/\alpha$ is the tuning offset factor, and $\theta_{1j} + \theta_{2j}$ is the total electrical length of the resonator cavity R_j.

The sidewall coupling component B_{jl} is shown as an inductor in Figure 13.7b, indicating that this coupling has the same sign as that of the endwall couplings, such as B_{ij}, and so it can also be realized as an inductive iris (Figure 13.7a). If the sign of B_{jl} is opposite to the signs of the aperture couplings, it can be represented electrically as a capacitor and realized as a capacitive probe. It will be explained in the next section how resonances R_j and R_l can be accommodated by two orthogonally polarized waveguide modes within a single cylindrical or square cross-sectional

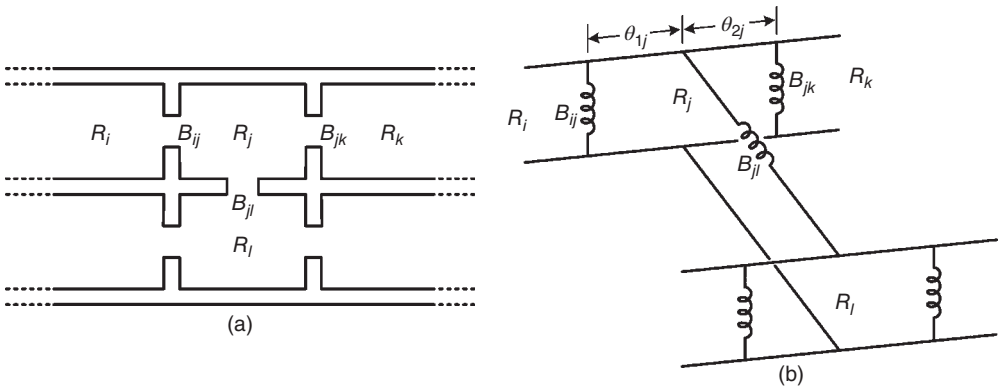


Figure 13.7 Rectangular waveguide resonators with sidewall coupling: (a) sidewall coupling between rectangular waveguide resonators R_j and R_i; (b) open-wire electrical equivalent circuit.

dual-mode cavity. In a dual-mode cavity, the sidewall coupling B_{jl} is realized as a 45° offset screw on the side of the cavity, and the intercavity couplings B_{ij}, B_{jk}, and so on are realized as slots in the cavity endwalls.

It should be remembered that when the prototype circuit is mapped to the waveguide circuit, a transform from the *parallel* resonator of the prototype circuit (shunt-connected $C_j + jB_j$) to the *series* resonator of the waveguide circuit ($\approx n\lambda/2$ lengths of transmission line) needs to be included. This is done by applying the dual-network theorem; add an inverter at both ends of the prototype network to form the input/output irises B_{S1} and B_{NL}, as shown in Figure 13.9.

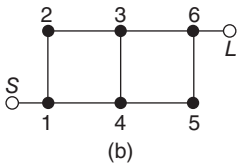
To demonstrate that this open-wire model, representing a waveguide bandpass filter, offers a reasonably accurate analysis, an example is taken of a sixth-degree asymmetric filter realized in extended box topology. Electrically, the filtering function is a (6–2) asymmetric Chebyshev with 23 dB return loss and two positive TZs, producing two rejection lobes of 50 dB each on the upper side of the passband. At RF, the passband width is 40 MHz centered at 6.65 GHz. Physically, the filter is realized in rectangular TE_{103}-mode cavities of broad dimension 3.485 cm (WR137).

Following the procedures laid out in earlier chapters, the prototype polynomials are determined, and the folded CM is synthesized. Then the folded matrix can be reconfigured to the extended box form by using the methods outlined in Chapter 10. The final $N + 2$ CM and the corresponding coupling and routing diagram are shown in Figure 13.8. Note that in the CM the signs of the sequential (mainline) couplings have been organized such that after the application of equation (13.12), the RF values B_{ij} are all negative so that they may be realized as apertures. This leaves the cross-couplings B_{14} and B_{36} positive in value, and they are realized with capacitive probes (see Figure 13.9).

Applying equations (13.11), (13.12), and (13.14), $\alpha = 20.5219$, and then the coupling values and cavity lengths are calculated as shown in Table 13.1. An analysis of the open-wire circuit with these values is given in Figure 13.10, which shows that the return loss level and rejection lobe levels are reasonably well preserved. The return loss characteristic is not perfectly equiripple, due to the dispersion effects that are included in the analysis.

	S	1	2	3	4	5	6	*L*
S	0.0	1.0638	0	0	0	0	0	0
1	1.0638	0.0399	–0.1513	0	–0.8887	0	0	0
2	0	–0.1513	–0.7214	0.5328	0	0	0	0
3	0	0	0.5328	0.3023	–0.5576	0	–0.7572	0
4	0	–0.8887	0	–0.5576	0.0732	0.4282	0	0
5	0	0	0	0	0.4282	–0.5516	–0.4892	0
6	0	0	0	–0.7572	0	–0.4892	0.0399	1.0638
L	0	0	0	0	0	0	1.0638	0.0

(a)



(b)

Figure 13.8 Extended box (6–2) filter prototype: (a) prototype coupling matrix; (b) coupling–routing diagram.

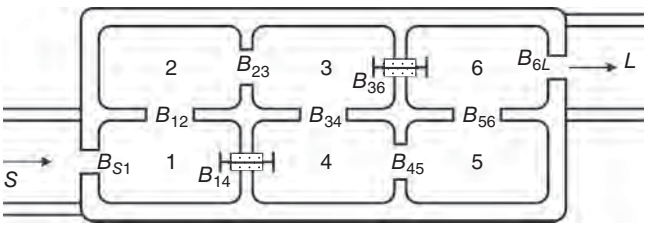


Figure 13.9 Realization of (6–2) extended box filter in rectangular waveguide TE_{103}-mode cavities.

Table 13.1 (6–2) Asymmetric extended box filter with RF values of coupling susceptances[a] and cavity lengths.

	Couplings		Cavity lengths (°)		
	Endwall	**Sidewall**	j	θ_{1j}	θ_{2j}
Mainline couplings	$B_{S1} = -4.0234$	—	1	257.05	272.75
	—	$B_{12} = -0.0074$	2	269.20	267.72
	$B_{23} = -38.4940$	—	3	269.71	273.31
	—	$B_{34} = -0.0272$	4	273.36	269.69
	$B_{45} = -47.9005$	—	5	268.72	269.91
	—	$B_{56} = -0.0238$	6	272.85	257.52
	$B_{6L} = -4.0234$	—	—	—	—
Cross-couplings	$B_{14} = +23.0492$	—	—	—	—
	$B_{36} = +27.0672$	—	—	—	—

a) Normalized to the rectangular waveguide characteristic admittance Y_0.

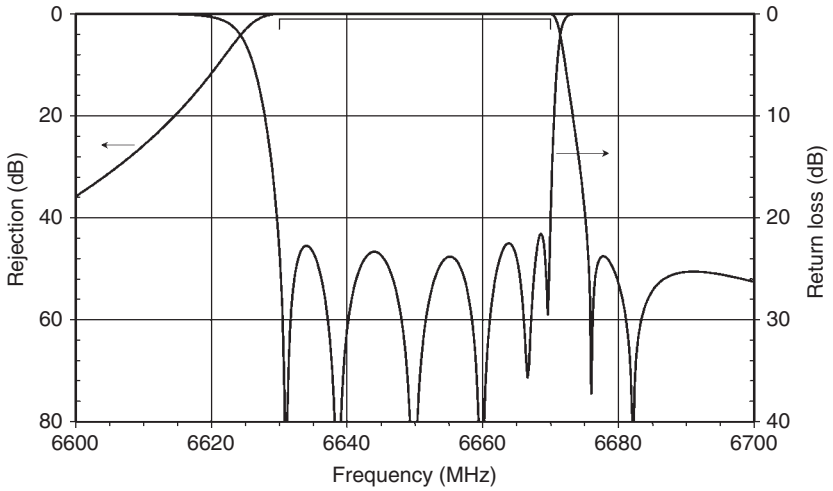


Figure 13.10 RF analysis of open-wire circuit model of (6–2) asymmetric filter in extended box configuration.

13.4 Dual-Mode Filters

Dual-mode filters are based on a polarized mode degeneracy within the same cavity rather than two *different* modes within the cavity [15]. *Mode degeneracy* means that two modes with the same mode indices can exist independently, that is, orthogonally, until the modes are coupled by some means such as a coupling screw that perturbs the fields. Both TE and TM modes can be degenerate, and dual TE_{11n} mode in cylindrical cavities (see Figure 13.11) is used most often. Dual-mode filters became popular in the early 1970s for satellite transponders because of their compact size and ability to realize advanced filtering functions [1–4].

Each cylindrical cavity supports two orthogonally polarized TE_{11n}-mode resonances. They are tuned by the two adjustable screws set 90° apart around the circumference of the cavity, which define and stabilize the polarization directions. The two resonances exist independently; that is, if vertical resonance 1 is excited by the horizontal input iris, then resonance 2 (the

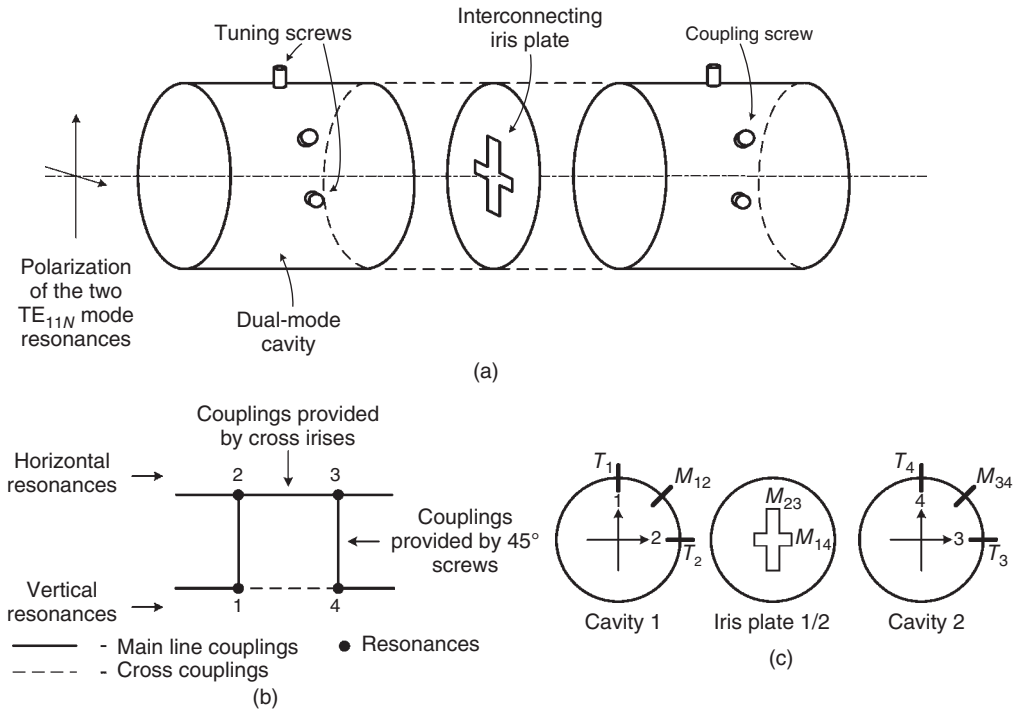


Figure 13.11 Dual-mode resonator cavities: (a) two cylindrical waveguide dual-mode cavities and interconnecting iris plate; (b) corresponding coupling–routing diagram; (c) relations between the mode polarization vectors and tuning screws, coupling screws, and coupling irises.

horizontal) does not exist until the circular symmetry of the cavity is perturbed, for example, by an adjustable screw at 45°. This asymmetry distorts the field of the first resonance, producing field components in the direction of the second resonance's field lines such that it is sustained. In effect, the coupling screw is realizing the coupling M_{12} from the CM, and the depth of its penetration into the cavity determines the strength of that coupling.

The intercavity couplings are provided by the cross-iris plates between the cavities. Again, if we assume that resonance 1 in cavity 1 is vertically polarized, vertical resonance 4 in cavity 2 is generated through the horizontal iris slot. Thus, this iris slot realizes the coupling M_{14} from the CM; the strength of the coupling depends on the length, width, and thickness of the slot. Similarly, the vertical iris slot realizes the coupling M_{23} between the horizontal resonance 2 in cavity 1 and horizontal resonance 3 in cavity 2. Because the iris slots are usually longer than they are wide, the cross-polarization couplings are negligibly small. For example, M_{14}, which couples resonances 1 and 4, will not couple strongly between resonances 2 and 3. There are other iris configuration possibilities, various input/output coupling schemes, and many useful design formulas discussed in the literature [15].

13.4.1 Virtual Negative Couplings

When we synthesize and reconfigure the CM for a filter function that includes finite-position TZs, we find that certain patterns of those TZs unavoidably cause some of the couplings to be negative. An example is a (4–2) filter characteristic with two symmetrically placed TZs located on the imaginary axis. In the CM for this characteristic, the cross-coupling M_{14} is negative

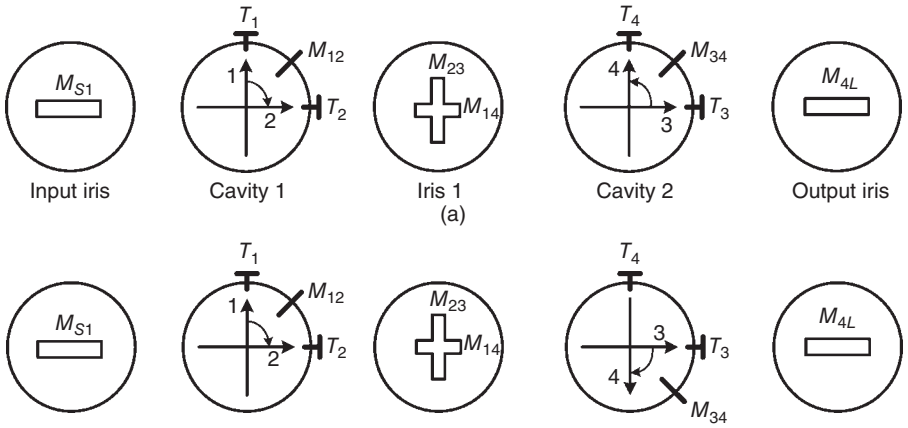


Figure 13.12 Creation of a virtual negative coupling between dual-mode cavities: (a) M_{14} positive; (b) M_{14} negative.

with respect to M_{12}, M_{23}, and M_{34}. In a filter realized with single-mode cavities, M_{14} would be realized with a capacitive probe, while the others are inductive loops or apertures.

For filters with dual-mode cavities, the negative coupling is achieved by the appropriate positioning of the 45° coupling screws relative to each other, obviating the need to provide special capacitive coupling devices that increase the complexity considerably and reduce Q_u (unloaded Q), the reliability, and the stability.

To explain how this "virtual" negative coupling is formed, a convention is adopted where the positive polarization direction (i.e., the heads of the vector arrows) of the horizontally and vertically polarized modes in a dual-mode cavity remains adjacent to the 45° screw that couples them in the same cavity. With this convention, it is evident from Figure 13.12a, where the coupling screws M_{12} and M_{34} are in the same axial position around the peripheries of cavities 1 and 2, that resonances 1 and 4 are polarized in the same direction as seen through their intercavity coupling iris M_{14}. However, if the coupling screw M_{34} in the second cavity is moved around the periphery of the cavity to be axially displaced 90° relative to coupling screw M_{12}, as shown in Figure 13.12b, then resonance 4 is polarized opposite to resonance 1, as seen through the coupling iris M_{14}. Thus, the phase of resonance 4 is 180° different from that of resonance 1. The overall effect is that iris M_{14} behaves as a virtual negative coupling, even though it is electrically inductive.

When we analyze such couplings with an open-wire model, the negative sign should be allocated to the shunt inductor B_{14} that represents the coupling M_{14} (its susceptance B_{14} is positive) but retains the frequency dependency of an inductive coupling (λ_g/λ_{g0}).

13.5 Coupling Sign Correction

The ability to realize both positive and negative couplings within a structure with only inductive couplings is one of the main attractions of the dual-mode resonant cavity. It not only offers mass and volume reductions over the single-mode equivalent but is also able to realize the couplings of advanced filter functions that, in general, require some negative couplings, without the need for capacitive probe structures.

However, before embarking on the design of a dual-mode filter, the CM that embodies the filter characteristic to be realized by the dual-mode structure needs to be conditioned to ensure

that the negative couplings are assigned to one of the two arms of the coupling iris, that is, the horizontal couplings in coupling–routing diagrams such as those shown in Figure 13.11. After synthesizing and reconfiguring a CM for a particular filter function and topology, negative couplings tend to appear rather randomly, sometimes corresponding to 45° coupling screws in the dual-mode cavities. These couplings need to be shifted to a coupling to be realized by an intercavity iris, where they can be realized "virtually" by 90° coupling screw axial offsets as explained in Section 13.4.1.

The sign of any coupling M_{ij} in the CM can be "shifted" (i.e., changed to positive) by applying a scaling of −1 to all the entries in row and column i, or row and column j of the CM. The effect of doing this is to shift the negative sign from M_{ij} to the other couplings in the row and column that the scaling is applied to. For example, if the coupling M_{12} (which is accomplished with a 45° screw in a dual-mode structure) is negative in the CM of Figure 13.8, then multiplying all the entries in row 2 and column 2 by −1 shifts the negative sign to mainline coupling M_{23} and any other cross-coupling in row/column 2. Coupling M_{23} is realized as one of the slots in the iris plate between cavities 1 and 2 and so can be "virtually" realized as a 90° axial offset in the position of the coupling screws in cavities 1 and 2. Note that the diagonal entry M_{22} representing the frequency offset of resonance 2 is unaffected by this "scaling" because it is multiplied by −1 twice.

Because the scaling has no effect on the electrical performance of the CM, the −1 scaling can be applied as many times as is necessary to shift the negative couplings to the right positions in the CM, although multiple scalings become quite cumbersome and prone to error. However, an easier graphical method, working with the coupling–routing diagram, may be used instead.

The method involves drawing arbitrarily shaped "enclosures" on the coupling–routing diagram. Each coupling line that the boundary of the enclosure crosses changes the sign of that coupling. Figure 13.13 gives an example of an eighth-degree longitudinal dual-mode filter, where the synthesis/reconfiguration process leaves a negative sign on M_{56} (a screw coupling); all others are positive. The enclosure drawn in Figure 13.13b changes the sign of M_{56} to positive, and M_{45} and M_{58}, both of which correspond to an iris coupling, to negative. Now the filter is realized as a dual-mode structure, as shown in Figure 13.13c. This example is equivalent to multiplying row/column 5 in the CM by −1.

Multiple enclosures or a single enclosure enveloping more than one resonator node and correcting several negative signs at once can be drawn. Also, the boundary of an enclosure can cut

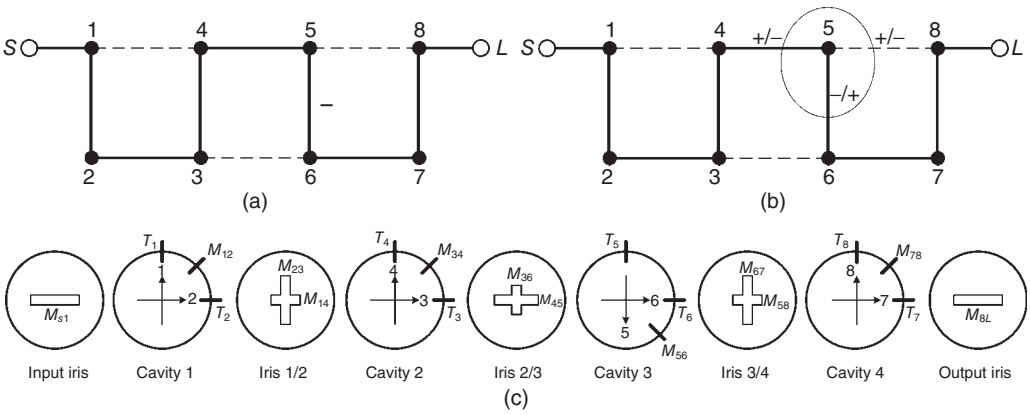


Figure 13.13 Coupling sign correction by the method of enclosures: (a) negative sign on M_{56}; (b) moving the negative sign to iris couplings—every coupling the enclosure cuts changes the sign; (c) dual-mode realization.

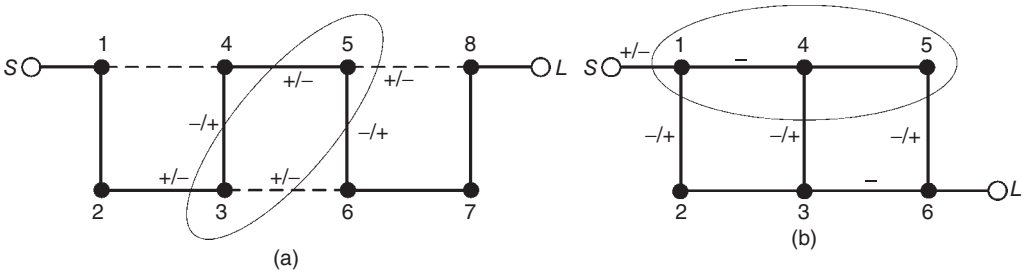


Figure 13.14 Two examples of multiple sign corrections on coupling–routing diagrams by the method of enclosures: (a) M_{34} and M_{56} negative; (b) M_{12}, M_{34}, and M_{56} negative.

single- or multiple-input/output couplings, changing their signs, if required. Figure 13.14 gives two further examples, the second of which is the correction of the sixth-degree extended box CM shown in Figure 13.8 into a form suitable for realization with dual-mode cavities.

13.6 Dual-Mode Realizations for Some Typical Coupling Matrix Configurations

In previous chapters, the synthesis and topology reorganization procedures for several types of filter configuration were described. Typically, the canonical form that emerges from the direct CM synthesis procedure is the folded array (sometimes known as the *reflex array*). Similarity transforms (rotations) can be applied to the matrix to form other topologies (minimum path rule permitting), many of which can be realized as a dual-mode structure. These configurations include the following:

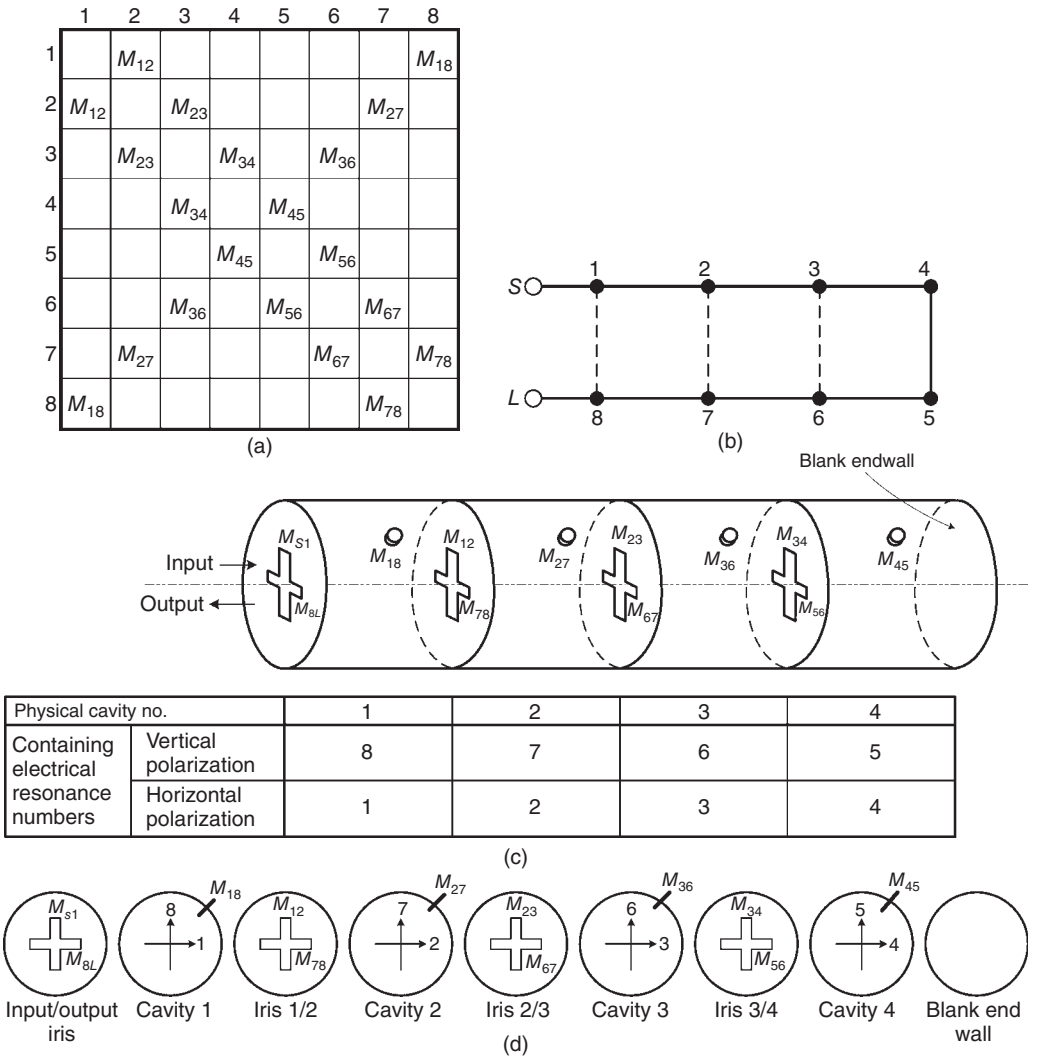
- Propagating or inline configuration
  Symmetric structure
  Asymmetric structure
- Pfitzenmaier configuration
- Cascade quartet
- Extended box

Each of these configurations has its own advantages and disadvantages, which are outlined along with the corresponding diagrammatic arrangement of dual-mode cavities and intercoupling iris plates and screws.

13.6.1 Folded Array

The canonical folded CM can be directly realized with a dual-mode structure, as shown in Figure 13.15. The polarized mode enters the filter in one polarization that propagates through the resonators in that polarization toward the cavity at the other end of the structure. The 45° coupling screw in this cavity generates the orthogonal polarization of resonance in this end resonator, and then the signal propagates back through the cavities toward the front end of the filter in this orthogonal polarization and emerges from the same cavity as it entered. The 45° coupling screws in the cavities provide the cross-couplings.

The advantages of the reflex dual-mode filter are that it is capable of realizing up to $N-2$ finite-position TZs ($n_{fz}=6$ in the eight-pole filter example) and that the coupling values are



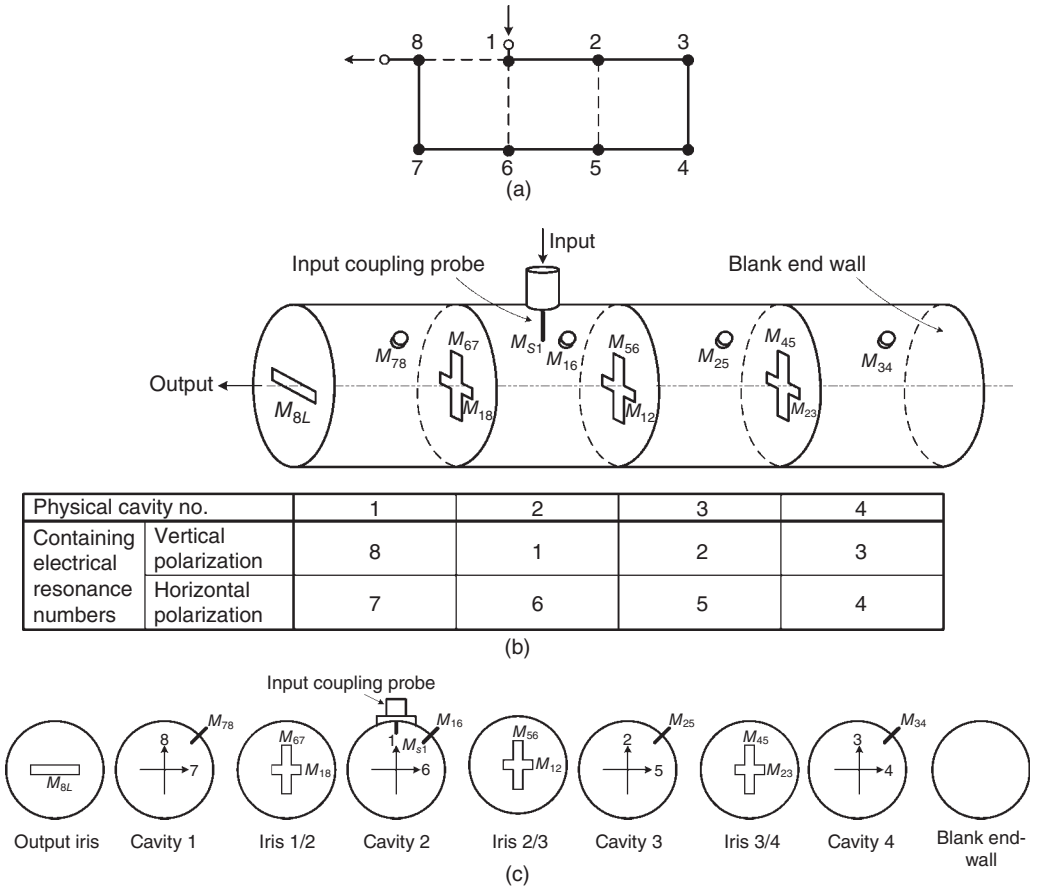
Physical cavity no.		1	2	3	4
Containing electrical resonance numbers	Vertical polarization	8	7	6	5
	Horizontal polarization	1	2	3	4

Figure 13.15 Realization of an eighth-degree folded configuration coupling matrix in dual-mode filter structure: (a) folded $N \times N$ coupling matrix; (b) corresponding coupling–routing diagram; (c) dual-mode cavity arrangement; (d) cavity–iris arrangement.

symmetric about the cross-diagonal of the CM. This means that the values of the two coupling slots in each iris plate are equal (and therefore, it can be circular hole). However, the big disadvantage is the existence of the input/output couplings in the same cavity, leading to poor input/output isolation. Furthermore, an orthomode transducer or sidewall coupling to the first cavity is needed to separate the input and output signals.

13.6.2 Pfitzenmaier Configuration

The Pfitzenmaier configuration (Figure 13.16) overcomes the poor isolation problem of the folded structure by having the input/output couplings located in adjacent cavities but can realize the same maximum numbers of finite-position TZs.



Physical cavity no.		1	2	3	4
Containing electrical resonance numbers	Vertical polarization	8	1	2	3
	Horizontal polarization	7	6	5	4

Figure 13.16 Pfitzenmaier dual-mode filter arrangement—eighth-degree example: (a) coupling–routing diagram; (b) dual-mode structure; and (c) cavity–iris relationships.

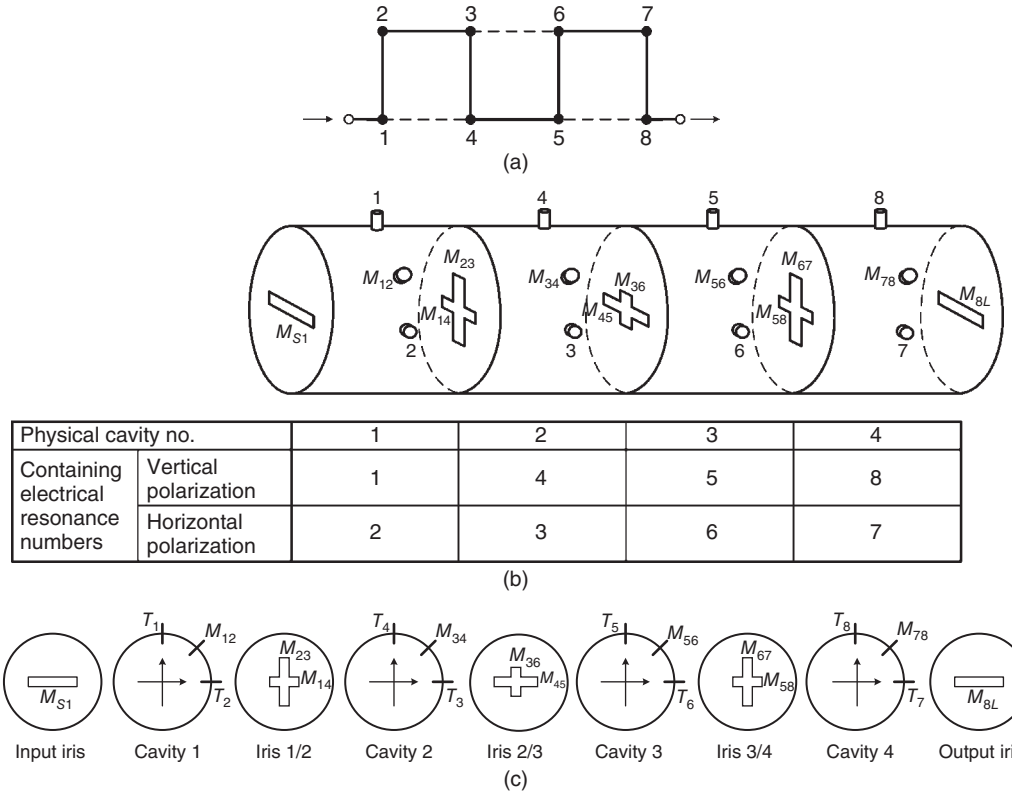
13.6.3 Propagating Forms

Both the symmetric and asymmetric propagating forms can be realized with the dual-mode arrangement, indicated in Figure 13.17, as an example of an eighth-degree filter.

The asymmetric propagating structure can realize any symmetric filter function (minimum path rule permitting), whereas the symmetric form has some restrictions on the type of filter function that it can realize. With the input and output at the opposite ends of the filter, a very good input/output isolation is attained, but the structure is able to realize fewer finite-position TZs as compared with the folded structure (e.g., $n_{fz} = 4$ for the above eighth-degree example). Note that if $N/2$ is odd, the output coupling slot is 90° axially rotated with respect to the input slot.

13.6.4 Cascade Quartet

The cascade quartet is similar to the propagating form but simplified in that some of the internal irises have single slots. Tuning is also easier, because each quartet is uniquely associated with the production of a symmetric pair of TZs, either on the imaginary axis or on the real axis (see Figure 13.18).



Physical cavity no.		1	2	3	4
Containing electrical resonance numbers	Vertical polarization	1	4	5	8
	Horizontal polarization	2	3	6	7

(b)

(c)

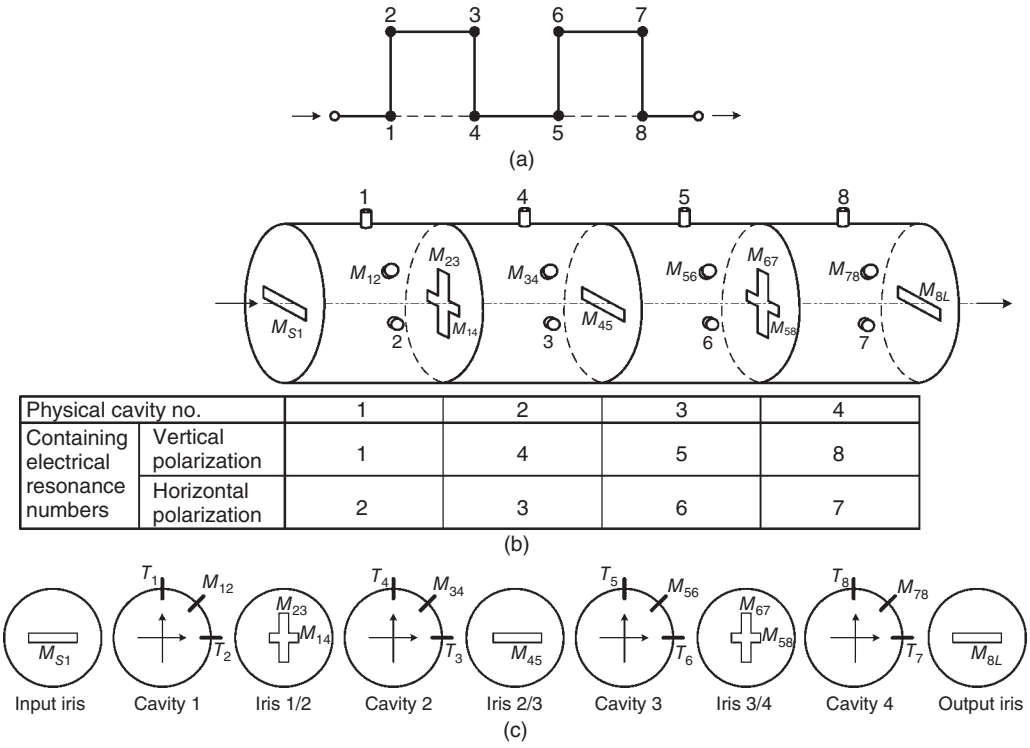
Figure 13.17 "Propagating" dual-mode filter arrangement for an eighth-degree filter; (a) coupling–routing diagram; (b) dual-mode structure; (c) cavity–iris relationships.

13.6.5 Extended Box

Extended box filters are unique in their ability to realize asymmetric filtering functions and in retaining the simple construction of the regular propagating dual-mode structure for symmetric filtering characteristics. The output coupling slot is always 90° axially rotated with respect to the input slot for any (even) degree of filter. Also, the resonances are asynchronously tuned, which can complicate the design of the dual-mode cavities in some cases. The example in Figure 13.19 is for the (6–2) asymmetric filter function for which the coupling–routing diagram is given in Figure 13.8. Here, M_{14} and M_{36} are both negative, and it may be seen in Figure 13.19 that to implement these negative couplings, the coupling screw M_{34} is 90° rotated with respect to coupling screws M_{12} and M_{56}.

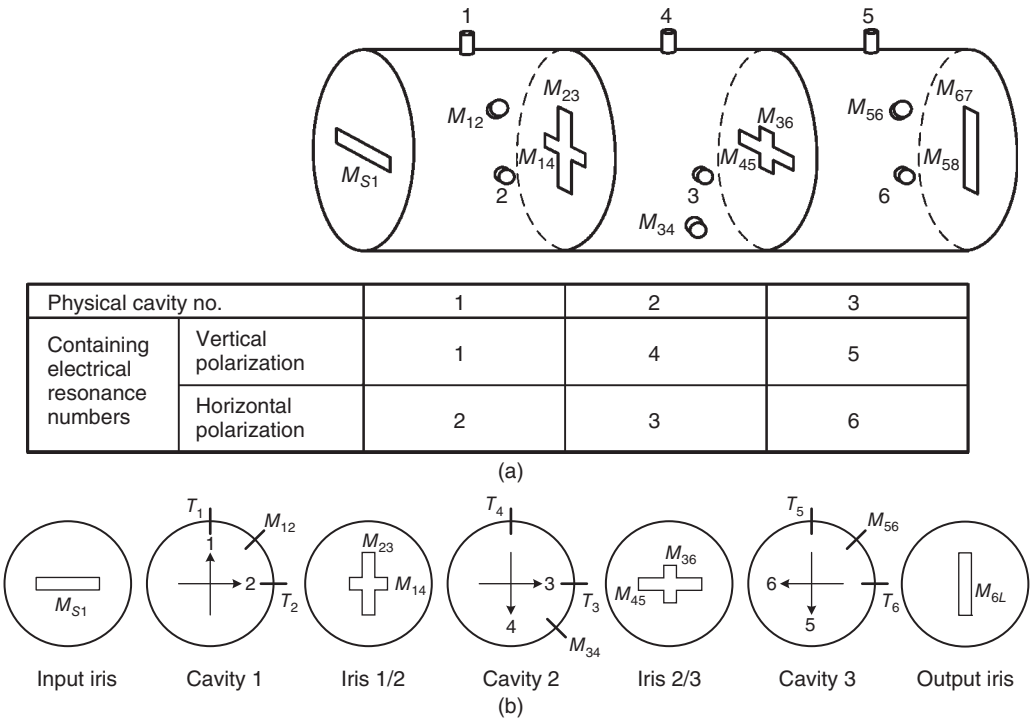
13.7 Phase- and Direct-Coupled Extracted Pole Filters

In Chapter 10, the method for extracting the TZs from a network and realizing them as band-stop cavities was detailed. Once the TZs are extracted, the main body of the filter can be synthesized and reconfigured as required, in the same way as for a nonextracted pole filter. With the TZs extracted, the cross-couplings in the main body, if any, are positive, which is useful for single-mode cavities, such as the TE_{011}-mode resonant cavity, where negative couplings must



Physical cavity no.		1	2	3	4
Containing electrical resonance numbers	Vertical polarization	1	4	5	8
	Horizontal polarization	2	3	6	7

Figure 13.18 Cascade quartet dual-mode filter arrangement for an eighth-degree filter: (a) coupling–routing diagram; (b) dual-mode structure; (c) cavity–iris relationships.



Physical cavity no.		1	2	3
Containing electrical resonance numbers	Vertical polarization	1	4	5
	Horizontal polarization	2	3	6

Figure 13.19 Extended box dual-mode filter arrangement for a sixth-degree filter: (a) dual-mode structure; (b) cavity–iris relationships.

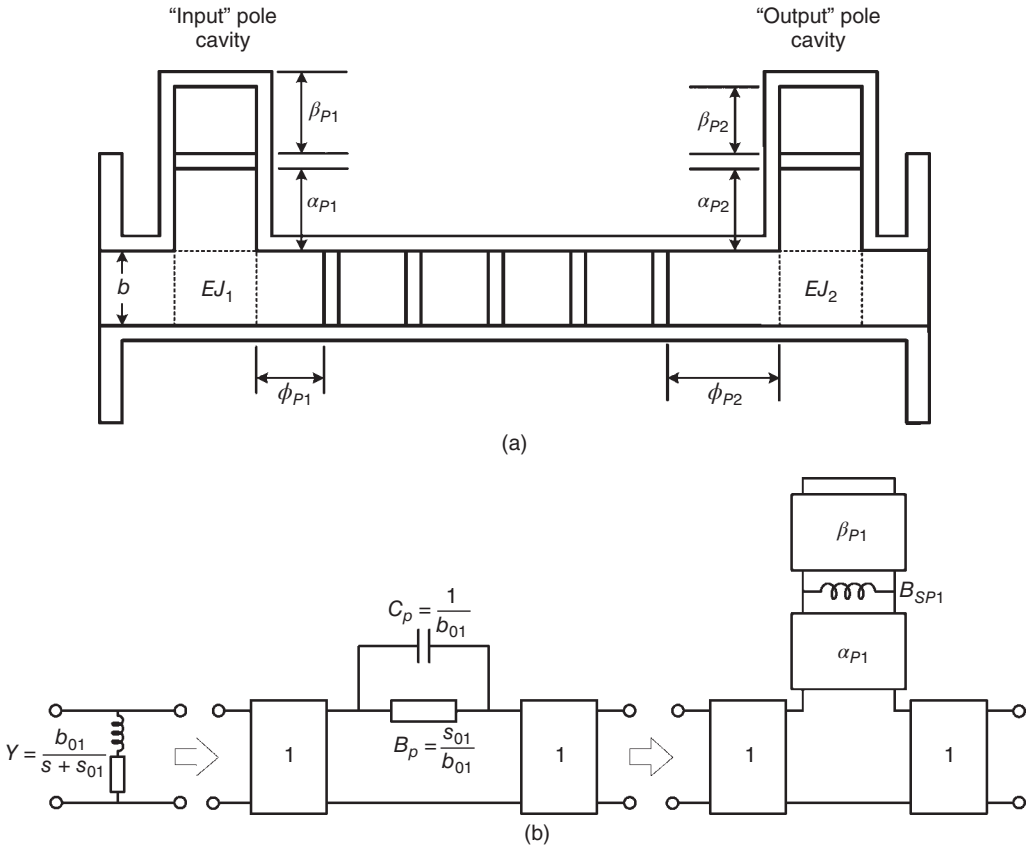


Figure 13.20 Sixth-degree rectangular waveguide filter with two *E*-plane extracted poles: (a) waveguide realization; (b) development of one of the prototype extracted poles to the waveguide equivalent circuit.

be formed as capacitive probes with all the associated consequences of poor power handling and reliability.

Figure 13.20a is an outline sketch of a rectangular waveguide phase-coupled extracted pole filter for a (6–2) filter function, where the TZs are realized with two extracted pole cavities, situated at either end of the main body and connected to the body through lengths of waveguide. Because there are only the two TZs in the transfer function (i.e., no real axis or complex pairs), the main body of the filter has no cross-couplings and is realized as a straight cascade of cavities. Figure 13.20b illustrates the development of one of the shunt-connected series resonant pairs from the extracted pole prototype circuit to the rectangular waveguide equivalent circuit for a bandstop cavity, connected to the main waveguide through an *E*-plane (series-connected) junction [12]. Equations (13.10), (13.11), and (13.12) are adapted to determine the electrical values by treating the extracted pole cavity as the first resonator of a regular filter but with no output coupling and no cross-coupling and resonant at the frequency of the TZ being considered.

13.7.1 Pole Cavity and Interconnecting Stub

Using equations (13.11) and (13.12), the susceptance of the input iris to the pole cavity is given by [12]

$$B_{SP1} = \frac{1}{\sqrt{\alpha C_{P1}}} - \sqrt{\alpha C_{P1}} \tag{13.15}$$

where $C_{P1} = 1/b_{01}$ and b_{01} is the residue of extracted pole 1. The phase length of the pole cavity is given by

$$\beta_{P1} = n\pi + \frac{1}{2}\cot^{-1}\frac{B_{SP1}}{2} - \sin^{-1}(B_{P1}) \tag{13.16}$$

where $B_{P1} = \omega_{01}/\alpha$ and $s_{01} = j\omega_{01}$ is the prototype position of the relevant TZ on the imaginary axis. This ensures that the pole cavity resonates at the RF frequency of the TZ. The shortening effect of the inductive iris means that the interconnecting length of line α_{P1}, between the pole cavity and the wall of the waveguide, must be negative for proper electrical operation. Consequently, a length equivalent to 180° phase is added to bring it to a positive value. Since this length is loaded by the same inductive iris as the main pole cavity and is normalized to the pole frequency, equation (13.16) applies, and $\alpha_{P1} = \beta_{P1}$.

13.7.2 Phase Lengths between the Filter Body and Pole Cavities

The unity admittance length of waveguide between the filter body and the T junction ϕ_{P1} is required to accommodate three factors:

1) Phase length ψ_{P1} in the prototype circuit between the first shunt $C_1 + jB_1$ of the main body of the filter and the adjacent extracted pole.
2) The short negative length associated with the input coupling iris of the main body of the filter.
3) If the pole cavity is connected to the main waveguide run via an E-plane junction, a 90° inverter associated with the equivalent circuit needs to be added, as shown in Figure 13.20.

Thus, the prototype phase length between the body of the filter and the nearest extracted pole is modified by the following equation [12]:

$$\phi_{P1} = \psi_{P1} + \frac{\pi}{2} - \frac{1}{2}\left(\cot^{-1}\frac{B_{S1}}{2}\right) \tag{13.17}$$

where ψ_{P1} is the phase length between the extracted pole and the main body of the filter obtained from the prototype circuit and B_{S1} is the susceptance of the input iris of the main body of the filter.

13.7.3 Phase Lengths between Adjacent Pole Cavities

For the phase lengths between the adjacent pole stubs, adjacent admittance inverters (if any) cancel, and the effect of the first body iris is nil. The phase lengths need only be modified by an arbitrary number of half-wavelengths, which should be added or subtracted for a convenient physical layout:

$$\phi_P \rightarrow \phi_P \pm k\pi \quad \text{radians}, \quad k = \text{integer} \tag{13.18}$$

13.8 The "Full-Inductive" Dual-Mode Filter

The dual-mode resonant cavity operates on the basis of two orthogonally polarized modes of the same kind in each cavity, resonating independently until intercoupled by a 45° screw. A novel "full-inductive" (all inductive couplings) dual-mode form, introduced by Guglielmi *et al*. [16], is based on specially dimensioned rectangular waveguide resonator cavities. Two different rectangular waveguide modes exist independently within each cavity, launched by a single critically dimensioned and positioned aperture at the input end. At the other end of the cavity, the two

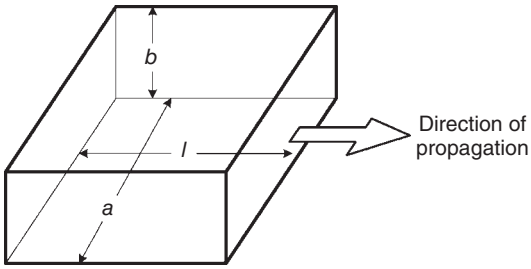


Figure 13.21 Dual-mode rectangular waveguide resonant cavity.

modes are coupled through one or possibly two apertures to the same two modes in the second cavity.[3] The couplings between the same modes through the intercavity aperture are referred to as "straight" couplings, and those between the unlike modes are "cross-mode" couplings.

Figure 13.21 is a sketch of a rectangular waveguide resonator cavity, showing the cross-sectional broad and narrow dimensions a and b, respectively, and the length l in the direction of propagation.

For a mode pair TE_{m0n} and TE_{p0q} to resonate simultaneously at frequency f_0 in a cavity of width a and length l,

$$a = \frac{c}{2f_0}\sqrt{\frac{(mq)^2 - (np)^2}{q^2 - n^2}} \tag{13.19a}$$

$$l = \frac{c}{2f_0}\sqrt{\frac{(mq)^2 - (np)^2}{m^2 - p^2}} \tag{13.19b}$$

$$\frac{a}{l} = \sqrt{\frac{m^2 - p^2}{q^2 - n^2}} \tag{13.19c}$$

where c is the velocity of propagation.

For the proper operation of the true dual-mode cavity, the first and last mode indices must be different between the two modes: $m \neq p$ and $n \neq q$. If the middle index of the modes is zero, then the b dimension of the waveguide is arbitrarily chosen for the best Q_u value and spurious mode rejection. The resonant frequencies of the two modes track each other, if there is a linear expansion of the cavity dimensions caused by a rise in the temperature of cavities.

Figure 13.22 shows a possible configuration for a (4–2) dual-mode filter with two resonator cavities, each supporting one TE_{102}-mode resonance and one TE_{201}-mode resonance. The

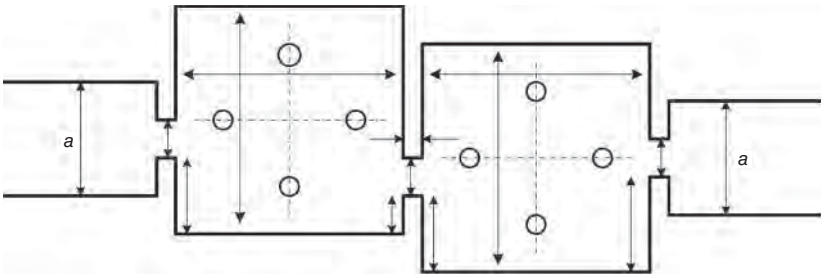


Figure 13.22 An example of a fourth-degree filter with two full-inductive dual-mode resonator cavities.

3 It is not strictly necessary to have the same modes in the second cavity as in the first; the principles still apply if the modes are different, provided that they can couple through the aperture(s) provided.

tuning screws are placed to be symmetric with respect to each mode to prevent an internal cross-mode coupling and at points which are at an electric field maximum for one mode, while minimal or zero for the other mode. The dimensional and positional variables of the input and output apertures and the intercavity apertures are as indicated. These dimensions can be used by an EM-based optimization routine to give the required straight and cross-mode coupling values simultaneously. We can also employ the same routine to fine-tune the dimension of the cavities. However, in practice, tuning screws tend to perturb the fields and upset the balance of the multiple coupling values provided by each aperture.

There is a certain small amount of coupling between the two modes in the same cavity at the iris interfaces, which can also be brought into the aperture optimization process. Also, extra degrees of freedom can be brought into play by introducing a second aperture and possibly a third aperture, for example, to couple exclusively between like modes in the two cavities. Here, the first aperture is designed to provide the correct cross-mode couplings, which will unavoidably also introduce some straight coupling as well. The second and third iris is then used to fill in and bring the straight couplings to their required values.

13.8.1 Synthesis of the Equivalent Circuit

Figure 13.23 shows the equivalent circuit of a fourth-degree dual-mode filter and also a three-cavity cascade forming a sixth-degree dual-mode filter.

The dual-input/output couplings from the interfacing rectangular waveguide are indicated, as well as the straight and cross-mode couplings that are provided by the intercavity iris aperture(s). This topology can be created from any of the inline coupling matrices, including the asymmetric dual-mode, the cascade quartet, or even the asymmetric extended box section form.

In a true dual-mode cavity, there is no direct coupling between the two modes that it is supporting. This is in contrast to the regular dual-mode configuration where intracavity mode

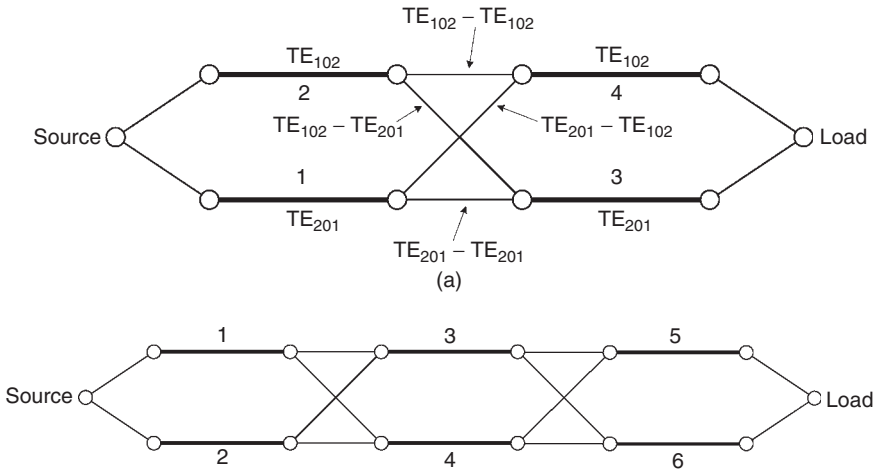
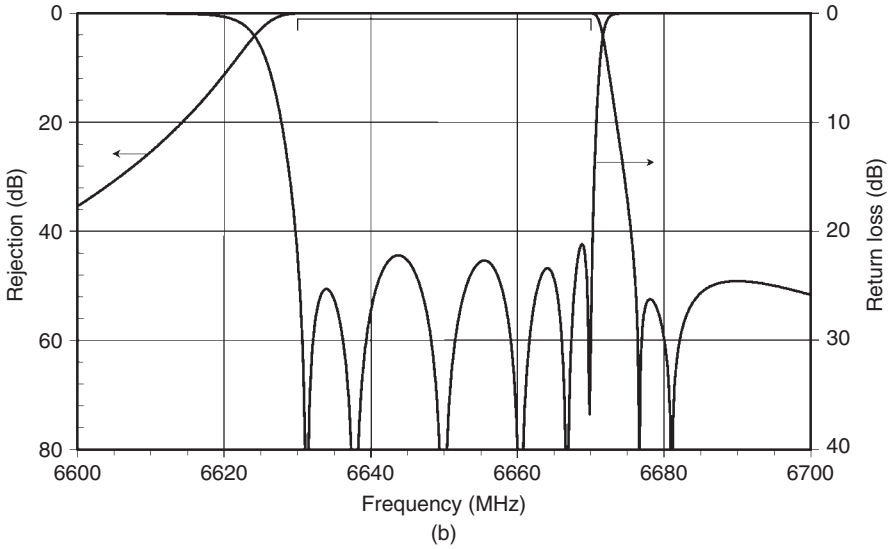


Figure 13.23 Full-inductive dual-mode filter equivalent circuits: (a) two cascaded cavities (4th degree) with TE_{102}/TE_{201}-mode dual resonances; (b) three cascaded cavities (6th degree) with TE_{m0n}/TE_{p0q}-mode dual resonances.

	S	1	2	3	4	5	6	*L*
S	0.0	0.2000	1.0449	0	0	0	0	0
1	0.2000	–0.7504	0.0	0.2012	0.5111	0	0	0
2	1.0449	0.0	0.0689	0.7397	–0.4740	0	0	0
3	0	0.2012	0.7397	–0.3815	0.0	0.0540	0.5798	0
4	0	0.5111	–0.4740	0.0	0.7571	0.5240	–0.3782	0
5	0	0	0	0.0540	0.5240	–0.8275	0.0	0.5226
6	0	0	0	0.5798	–0.3782	0.0	0.3158	0.9266
L	0	0	0	0	0	0.5226	0.9266	0.0

(a)



(b)

Figure 13.24 Full-inductive dual-mode realization for asymmetric (6–2) filter characteristic: (a) the coupling matrix; (b) analysis of an equivalent circuit at RF.

coupling is provided by the 45° coupling screw. Therefore, the synthesis procedure for the true dual-mode CM involves starting with an inline form and applying a series of $N/2$ cross-pivot rotations [see equation (8.37 g)] at pivots $[1, 2], [3, 4], \ldots, [N-1, N]$, at each stage, annihilating the mainline coupling $M_{12}, M_{34}, \ldots, M_{N-1,N}$, where N is the degree of the filtering characteristic. In the process of eliminating M_{12}, a second input coupling M_{S2} is created, as well as M_{13} and M_{24}. The second rotation annihilates M_{34} but creates M_{35} and M_{46} and so on until all the intra-cavity couplings are annihilated, and the second output coupling $M_{N-1,L}$ is created.

The process is illustrated for the (6–2) extended box configuration that was used as an example for the network analysis earlier in this chapter. Figure 13.8 gives the original extended box CM for the (6–2) prototype, and Figure 13.24a gives the matrix after the application of three cross-pivot rotations at pivots [1,2], [3,4], and [5,6], annihilating M_{12}, M_{34}, and M_{56} in sequence and some sign correction. Figure 13.24b presents the analysis of an open-wire equivalent circuit of this true dual-mode filter by assuming a TE_{102} and a TE_{201} mode in each cavity.

Summary

Dual-mode filters employ two orthogonally polarized modes in the physical realization of the filter structure. This implies that each physical cavity or resonator can support two electrical resonances, independent of each other. This leads to a size reduction of almost a factor of 2 and a concomitant reduction in mass as well. However, these advantages come at the expense of reduced spurious free frequency bands and some degradation in the unloaded Q. Owing to the large savings in mass and volume, dual-mode filters are widely used in communication systems.

In addition, dual-mode filters can be configured in waveguide, dielectric-loaded, and planar structures in a variety of topologies. This chapter describes the process of practical filter design by using dual-mode resonators that operate in the dominant mode, as well as in the higher-degree propagation modes. A variety of examples are included to illustrate the design procedure: longitudinal and canonical configurations, the extended box design, the extracted pole filter, and filters with all inductive couplings. The examples also include symmetric and asymmetric response filters. The steps involved in the simultaneous optimization of amplitude and group delay response are described for the design of a filter that may be realized in a variety of technologies. Examples in this chapter span the analysis and synthesis techniques described in Chapters 3–10.

References

1 Atia, A.E. and Williams, A.E. (1971) New types of bandpass filters for satellite transponders. *COMSAT Technical Review*, **1**, 21–43.

2 Williams, A.E. (1970) A four-cavity elliptic waveguide filter. *IEEE Transactions on Microwave Theory and Techniques*, **18**, 1109–1114.

3 Atia, A.E. and Williams, A.E. (1972) Narrow bandpass waveguide filters. *IEEE Transactions on Microwave Theory and Techniques*, **20**, 258–265.

4 Atia, A.E. and Williams, A.E. (1974) Non-minimum phase optimum-amplitude bandpass waveguide filters. *IEEE Transactions on Microwave Theory and Techniques*, **22**, 425–431.

5 Kudsia, C.M., Kallianteris, S., Ainsworth, K.R. and O'Donovan, M.V. (1978) Status of filter and multiplexer technology in the 12 GHz frequency band for space application. Communications Satellite Systems Conference, 7th, San Diego, CA, April 24–27, 1978, Technical Papers (A78-32876 13-32). American Institute of Aeronautics and Astronautics, Inc., New York, pp. 194–202.

6 Fiedziuszko, S.J. (1982) Dual mode dielectric resonator loaded cavity filter. *IEEE Transactions on Microwave Theory and Techniques*, **30**, 1311–1316.

7 Cameron, R.J., Tang, W.C. and Kudsia, C.M. (1990) Advances in dielectric loaded filters and multiplexers for communications satellites. Proceedings of the AIAA 13th International Satellite System Conference, March 1990.

8 Kudsia, C., Cameron, R., and Tang, W.C. (1992) Innovations in microwave filters and multiplexing networks for communications satellite systems. *IEEE Transactions on Microwave Theory and Techniques*, **40**, 1133–1149.

9 Cameron, R.J., Harish, A.R., and Radcliffe, C.J. (2002) Synthesis of advanced microwave filters without diagonal cross-couplings. *IEEE Transactions on Microwave Theory and Techniques*, **50**, 2862–2872.

10 Rhodes, J.D. (1976) *Theory of Electrical Filters*, Wiley, New York.

11 Rhodes, J.D. (1973) Filters approximating ideal amplitude and arbitrary phase characteristics. *IEEE Transactions on Circuit Theory*, **20**, 150–153.

12 Rhodes, J.D. and Cameron, R.J. (1980) General extracted pole synthesis technique with applications to low-loss TE_{011} mode filters. *IEEE Transactions on Microwave Theory and Techniques*, **28**, 1018–1028.

13 Matthaei, G., Young, L., and Jones, E.M.T. (1980) *Microwave Filters, Impedance Matching Networks and Coupling Structures*, Artech House, Norwood, MA.

14 Rhodes, J.D. (1970) The generalized direct-coupled cavity linear phase filter. *IEEE Transactions on Microwave Theory and Techniques*, **18**, 308–313.

15 Uher, J., Bornemann, J., and Rosenberg, U. (1993) *Waveguide Components for Antenna Feed Systems—Theory and CAD*, Artech House, Norwood, MA.

16 Guglielmi, M., *et al*. (2001) Low-cost dual-mode asymmetric filters in rectangular waveguide. IEEE MTT-S International Microwave Symposium Digest, Phoenix, pp. 1787–1790.

14

Design and Physical Realization of Coupled Resonator Filters

This chapter addresses the physical realization of microwave filters illustrating how one can combine filter circuit models with electromagnetic (EM) simulation tools to determine the filter physical dimensions. The techniques illustrate a direct approach to synthesis filter physical dimensions from the elements of the coupling matrix model, the K-impedance inverter model, or the J-admittance inverter model. The approach is fast, yet it provides reasonably accurate results. Numerical examples are given in this chapter to demonstrate, step by step, the application of the approach to the design of dielectric resonator, waveguide, and microstrip filters.

Bandpass filters typically consist of resonators coupled to each other by inductive or capacitive coupling elements. Figure 14.1 shows some of the common microwave filter structures. In waveguide, coaxial, and dielectric resonator filters, interresonator coupling is achieved with the use of an iris, whereas input/output coupling is realized with the use of a probe or an iris. In microstrip filters, interresonator and input/output coupling can take various forms depending on the circuit layout. The objective is to determine the physical dimensions of the individual resonators, interresonator couplings, and input/output couplings. The filter design process usually involves the following four main steps:

Step 1. Identification of the filter order and filter function according to specification requirements
Step 2. Synthesis of the coupling matrix [M]-, K-, or J-inverter circuit models that can realize the required filter function
Step 3. Identification of the filter type (waveguide, coaxial dielectric, microstrip, etc.) on the basis of size requirements, Q value, and power handling capability
Step 4. Identification of the filter physical dimensions

Chapters 3, 4, and 6–10 dealt with steps 1 and 2, whereas Chapter 11 addressed some aspects of step 3. This chapter and Chapter 15 deal with step 4. In Chapter 15 several advanced EM-based techniques for filter design are presented. The techniques given in Chapter 15 are very accurate, yet they are very computationally intensive. In this chapter, we employ an approach that combines filter circuit models with an EM simulator to get a fast yet reasonably accurate design. The approach employs a "divide and conquer strategy": to divide the filter structure into the design of interresonator couplings and input/output couplings. The accuracy of this approach is limited by the inherent approximation of the circuit models and by the fact that elements of the circuit model are assumed to be frequency independent. While the solution obtained using this approach is not expected to yield an exact design, it usually suffices as an initial solution. The exact design can be obtained by fine-tuning of the initial design or by using any of the advanced EM-based techniques presented in Chapter 15.

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

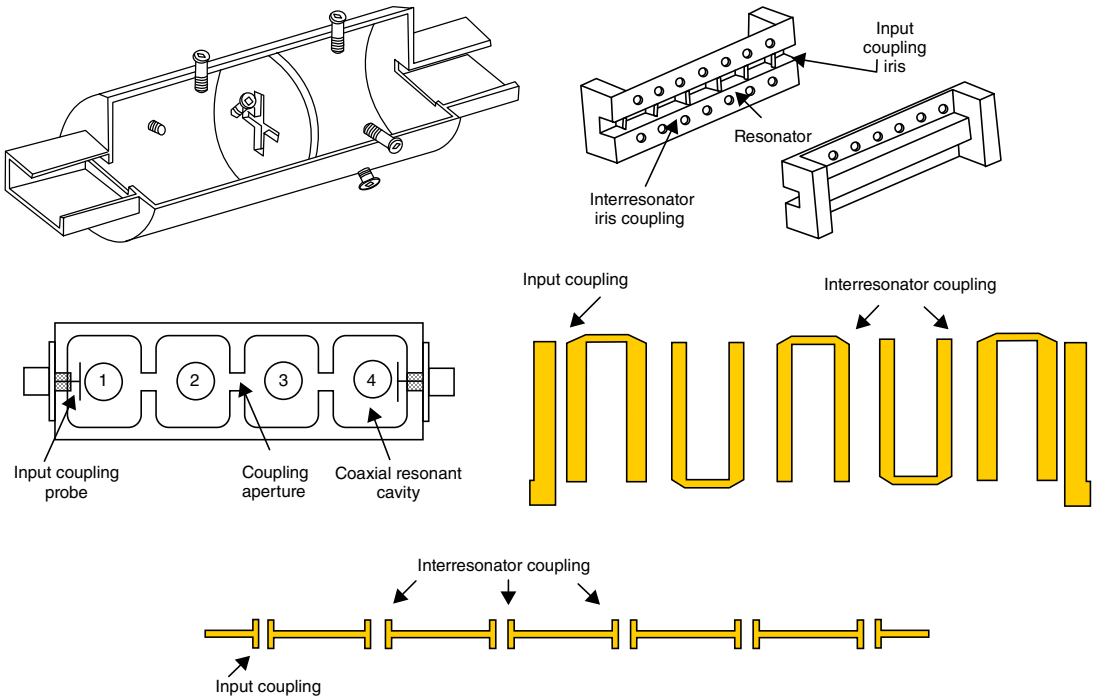


Figure 14.1 Several waveguide, dielectric resonator, and coaxial and microstrip filter components illustrating interresonator coupling and input/output coupling.

14.1 Circuit Models for Chebyshev Bandpass Filters

We use Chebyshev filters in this chapter to demonstrate the concept of using filter circuit models and an EM simulator to synthesize the physical dimensions of the filter. This concept is not limited to Chebyshev filters; it can be applied to other filter functions as well. Figure 14.2 illustrates three circuit models for Chebyshev filters: the coupling matrix model, the K-impedance inverter model, and the J-admittance inverter model. Chapters 7 and 8 provide details of these models. The three filter circuit models are equivalent to each other. The coupling elements $M_{n,n+1}$, the impedance inverters $K_{n,n+1}$, and the admittance inverters $J_{n,n+1}$ are related to the lumped-element lowpass prototype elements $g_0, g_1, g_2, \ldots, g_{N+1}$ as [1] follows:

Coupling matrix model:

$$M_{j,j+1} = \frac{1}{\sqrt{g_j g_{j+1}}}, \qquad j = 1, 2, \ldots, N-1 \tag{14.1}$$

$$R_1 = \frac{1}{g_0 g_1}, \qquad R_N = \frac{1}{g_N g_{N+1}} \tag{14.2}$$

K-Impedance Inverter Model:

$$\frac{K_{j,j+1}}{Z_0} = \frac{\pi\Delta}{2\sqrt{g_j g_{j+1}}}, \qquad j = 1, 2, \ldots, N-1 \tag{14.3}$$

$$\frac{K_{01}}{Z_0} = \sqrt{\frac{\pi\Delta}{2g_0 g_1}}, \qquad \frac{K_{N,N+1}}{Z_0} = \sqrt{\frac{\pi\Delta}{2\ g_N g_{N+1}}} \tag{14.4}$$

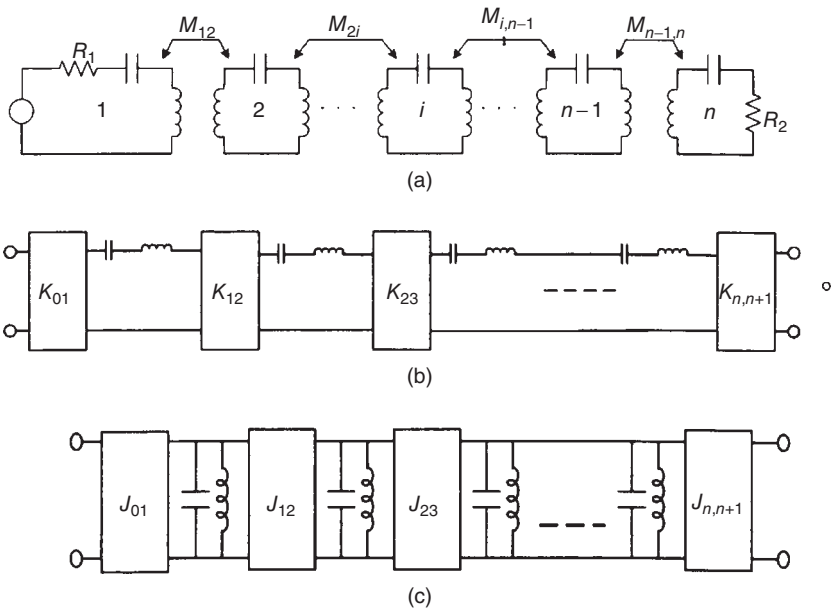


Figure 14.2 (a) Coupling matrix model for Chebyshev filters, (b) impedance inverter model for Chebyshev filters, and (c) admittance inverter model for Chebyshev filters.

J-Admittance Inverter Model:

$$\frac{J_{j,j+1}}{Y_0} = \frac{\pi\Delta}{2\sqrt{g_j g_{j+1}}}, \qquad j = 1, 2, \ldots, N-1 \tag{14.5}$$

$$\frac{J_{01}}{Y_0} = \sqrt{\frac{\pi\Delta}{2g_0 g_1}}, \qquad \frac{J_{N,N+1}}{Y_0} = \sqrt{\frac{\pi\Delta}{2g_N g_{N+1}}} \tag{14.6}$$

The parameter Δ in equations (14.3)–(14.6) is the fractional bandwidth. For a dispersive medium, Δ is given by $\Delta = (\lambda_{g1} - \lambda_{g2})/\lambda_{g0}$, where λ_{g0} is the guide wavelength at the center frequency and λ_{g1} and λ_{g2} are the guide wavelengths at the band-edge frequencies. The lowpass g_k values for maximally flat response and Chebyshev response are determined using equations (14.7) and (14.8), respectively [1, 2]:

Maximally Flat Filters:

$$\begin{aligned} g_0 &= 1 \\ g_k &= 2\sin\left[\frac{(2k-1)\pi}{2n}\right], \quad k = 1, 2, \ldots, n \\ g_{n+1} &= 1 \end{aligned} \tag{14.7}$$

Chebyshev Filters:

$$\begin{aligned} g_0 &= 1 \\ \beta &= \ln\left(\coth\left(\frac{L_{\mathrm{AR}}}{17.37}\right)\right) \\ \gamma &= \sinh\left(\frac{\beta}{2n}\right) \\ a_k &= \sin\left[\frac{(2k-1)\pi}{2n}\right], \quad k = 1, 2, \ldots, n \\ b_k &= \gamma^2 + \sin^2\left(\frac{k\pi}{n}\right), \quad k = 1, 2, \ldots, n \\ g_1 &= \frac{2a_1}{\gamma} \\ g_k &= \frac{4a_{k-1}a_k}{b_{k-1}g_{k-1}} \quad k = 2,3, \ldots, n \\ g_{n+1} &= \begin{cases} 1 & n \text{ odd} \\ \coth^2\left(\frac{\beta}{4}\right) & n \text{ even} \end{cases} \end{aligned} \tag{14.8}$$

where n is the filter order and L_{AR} is the ripple level in the passband that can be calculated from the filter return loss. The g_k values for Chebyshev filters with various ripple levels are given in Table 14.1. As illustrated in Chapter 7, the impedance and admittance inverters are basically impedance and admittance transformers, respectively. These inverters can be realized using any of the lumped-element circuits shown in Table 14.2 [1, 3]. Equations (14.9)–(14.16) provide the relations between the values of the circuit elements and their equivalent inverter values.

Table 14.1 Element values for Chebyshev filters for various ripple levels ($g_0 = 1$, $\omega_c = 1$).

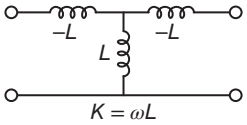
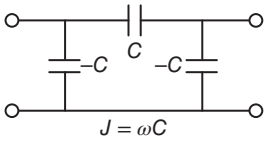
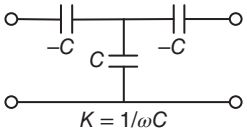
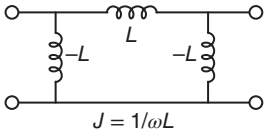
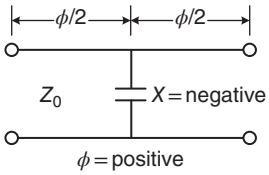
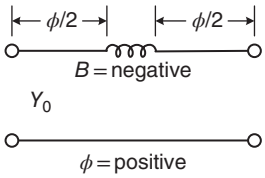
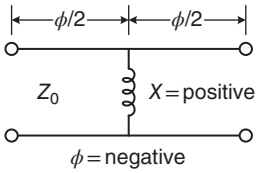
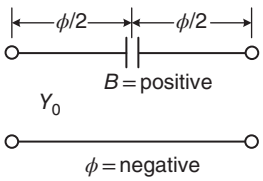
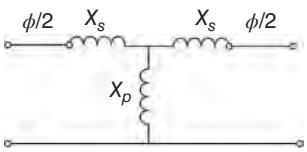
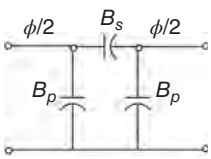
	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
					0.01 dB ripple						
1	0.0960	1.0000									
2	0.4488	0.4077	1.1007								
3	0.6291	0.9702	0.6291	1.0000							
4	0.7128	1.2003	1.3212	0.6476	1.1007						
5	0.7563	1.3049	1.5773	1.3049	0.7563	1.0000					
6	0.7813	1.3600	1.6896	1.5350	1.4970	0.7098	1.1007				
7	0.7969	1.3924	1.7481	1.6331	1.7481	1.3924	0.7969	1.0000			
8	0.8072	1.4130	1.7824	1.6833	1.8529	1.6193	1.5554	0.7333	1.1007		
9	0.8144	1.4270	1.8043	1.7125	1.9057	1.7125	1.8043	1.4270	0.8144	1.0000	
10	0.8196	1.4369	1.8192	1.7311	1.9362	1.7590	1.9055	1.6527	1.5817	0.7446	1.1007
					0.0138 dB ripple						
1	0.1128	1.0000									
2	0.4886	0.4365	1.1194								
3	0.6708	1.0030	0.6708	1.0000							
4	0.7537	1.2254	1.3717	0.6734	1.1194						
5	0.7965	1.3249	1.6211	1.3249	0.7965	1.0000					
6	0.8210	1.3770	1.7289	1.5445	1.5414	0.7334	1.1194				
7	0.8362	1.4075	1.7846	1.6368	1.7846	1.4075	0.8362	1.0000			
8	0.8463	1.4269	1.8172	1.6837	1.8847	1.6234	1.5973	0.7560	1.1194		
9	0.8533	1.4400	1.8380	1.7109	1.9348	1.7109	1.8380	1.4400	0.8533	1.0000	
10	0.8583	1.4493	1.8521	1.7281	1.9636	1.7542	1.9344	1.6546	1.6223	0.7668	1.1194
					0.1 dB ripple						
1	0.3052	1.0000									
2	0.8430	0.6220	1.3554								
3	1.0315	1.1474	1.0315	1.0000							
4	1.1088	1.3061	1.7703	0.8180	1.3554						
5	1.1468	1.3712	1.9750	1.3712	1.1468	1.0000					
6	1.1681	1.4039	2.0562	1.5170	1.9029	0.8618	1.3554				
7	1.1811	1.4228	2.0966	1.5733	2.0966	1.4228	1.1811	1.0000			
8	1.1897	1.4346	2.1199	1.6010	2.1699	1.5640	1.9444	0.8778	1.3554		
9	1.1956	1.4425	2.1345	1.6167	2.2053	1.6167	2.1345	1.4425	1.1956	1.0000	
10	1.1999	1.4481	2.1444	1.6265	2.2253	1.6418	2.2046	1.5821	1.9628	0.8853	1.3554
					0.2 dB ripple						
1	0.4342	1.0000									
2	1.0378	0.6745	1.5386								
3	1.2275	1.1525	1.2275	1.0000							
4	1.3028	1.2844	1.9761	0.8468	1.5386						

(Continued)

Table 14.1 (Continued)

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
5	1.3394	1.3370	2.1660	1.3370	1.3394	1.0000					
6	1.3598	1.3632	2.2394	1.4555	2.0974	0.8838	1.5386				
7	1.3722	1.3781	2.2756	1.5001	2.2756	1.3781	1.3722	1.0000			
8	1.3804	1.3875	2.2963	1.5217	2.3413	1.4925	2.1349	0.8972	1.5386		
9	1.3860	1.3938	2.3093	1.5340	2.3728	1.5340	2.3093	1.3938	1.3860	1.0000	
10	1.3901	1.3983	2.3181	1.5417	2.3904	1.5536	2.3720	1.5066	2.1514	0.9034	1.5386

Table 14.2 Equivalent circuits for impedance and admittance inverters.

Impedance inverter	Admittance inverter
 (a)	 (f)
 (b)	 (g)
 (c)	 (h)
 (d)	 (i)
 (e)	 (j)

For the impedance inverter circuits shown in Table 14.2c and d, we have

$$K = Z_0 \tan\left|\frac{\phi}{2}\right|, \quad \phi = -\tan^{-1}\frac{2X}{Z_0} \tag{14.9}$$

$$\left|\frac{X}{Z_0}\right| = \frac{\dfrac{K}{Z_0}}{1-\left(\dfrac{K}{Z_0}\right)^2} \tag{14.10}$$

For the admittance inverter circuits shown in Table 14.2h and i, we have

$$J = Y_0 \tan\left|\frac{\phi}{2}\right|, \quad \phi = -\tan^{-1}\frac{2B}{Y_0} \tag{14.11}$$

$$\left|\frac{B}{Y_0}\right| = \frac{\dfrac{J}{Y_0}}{1-\left(\dfrac{J}{Y_0}\right)^2} \tag{14.12}$$

For the impedance inverter circuit shown in Table 14.2e, we have

$$\frac{K}{Z_0} = \left|\tan\left(\frac{\phi}{2} + \arctan\frac{X_s}{Z_0}\right)\right| \tag{14.13}$$

$$\phi = -\arctan\left(2\frac{X_p}{Z_0} + \frac{X_s}{Z_0}\right) - \arctan\frac{X_s}{Z_0} \tag{14.14}$$

For the admittance inverter circuit shown in Table 14.2j, we have

$$\frac{J}{Y_0} = \left|\tan\left(\frac{\phi}{2} + \arctan\frac{B_p}{Y_0}\right)\right| \tag{14.15}$$

$$\phi = -\arctan\left(2\frac{B_s}{Y_0} + \frac{B_p}{Y_0}\right) - \arctan\frac{B_p}{Y_0} \tag{14.16}$$

14.2 Calculation of Interresonator Coupling

14.2.1 The Use of Electric Wall and Magnetic Wall Symmetry

We consider first the coupling between two resonators, where symmetry is used to divide the problem into single resonators terminated by a magnetic wall and an electric wall. The coupling then is determined from the knowledge of the resonant frequencies of the two individual resonators. The equivalent circuit of two resonators separated by an inductive coupling element is shown in Figure 14.3. The coupling element is represented by a T network consisting of a shunt inductance L_m and two series inductances of value $-L_m$. The network effectively represents the K-impedance inverter given in Table 14.2a. The coupling coefficient M can be computed by evaluating the resonant frequencies of the even- and odd-mode circuits shown in Figure 14.3.

In view of Figure 14.3, the resonant frequencies f_m and f_e of the two circuits shown in Figures 14.3b and c are given by [4].

$$f_m = \frac{1}{2\pi\sqrt{(L+L_m)C}} \tag{14.17}$$

$$f_e = \frac{1}{2\pi\sqrt{(L-L_m)C}} \tag{14.18}$$

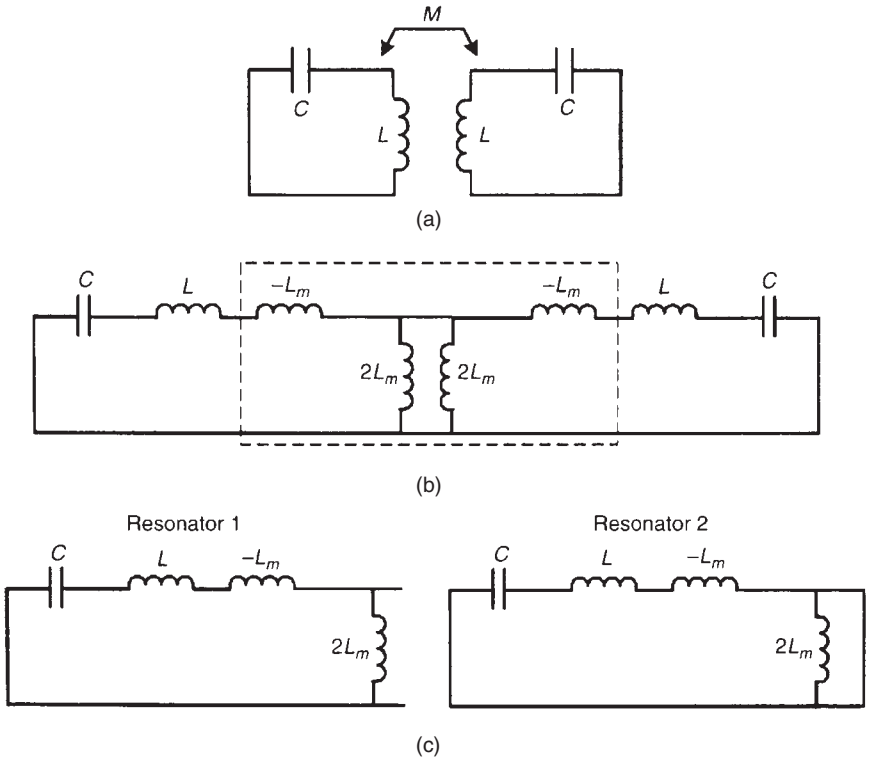


Figure 14.3 (a) Impedance inverter—equivalent circuit of two resonators separated by an inductive coupling, (b) half of the circuit terminated in a magnetic wall (even mode; open circuit), and (c) half of the circuit terminated in an electric wall (odd mode; short circuit).

Solving equations (14.17) and (14.18) gives the inductive coupling coefficient k_M as

$$k_M = \frac{L_m}{L} = \frac{f_e^2 - f_m^2}{f_e^2 + f_m^2} \tag{14.19}$$

In view of Chapter 8, the elements of the coupling matrix $[M]$ are normalized to the fractional bandwidth. The coupling element M between two adjacent resonators is then given by

$$M = \frac{f_0}{\text{BW}} \frac{f_e^2 - f_m^2}{f_e^2 + f_m^2} \tag{14.20}$$

where f_0 is the filter center frequency and BW is the filter bandwidth.

Similarly, for capacitive coupling, the equivalent circuit can be presented as shown in Figure 14.4. The π network shown in this figure effectively represents the admittance inverter J shown in Table 14.2f.

The resonant frequencies f_m and f_e of the two circuits shown in Figures 14.4b and c are given as

$$f_e = \frac{1}{2\pi\sqrt{(C + C_m)L}} \tag{14.21}$$

$$f_m = \frac{1}{2\pi\sqrt{(C - C_m)L}} \tag{14.22}$$

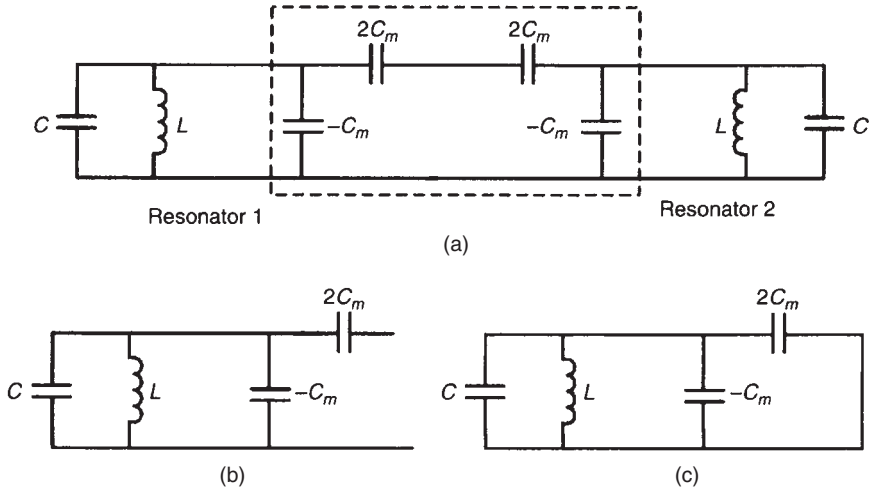


Figure 14.4 (a) Admittance inverter—equivalent circuit of two resonators capacitively coupled, (b) magnetic wall (even mode), and (c) electric wall (odd mode).

Solving equations (14.21) and (14.22) yields the electric coupling coefficient k_e as

$$k_e = \frac{C_m}{C} = \frac{f_m^2 - f_e^2}{f_m^2 + f_e^2} \tag{14.23}$$

The normalized coupling element M is given by

$$M = \frac{f_0}{\text{BW}} \frac{f_m^2 - f_e^2}{f_m^2 + f_e^2} \tag{14.24}$$

In general, $f_e > f_m$ for inductive coupling and $f_m > f_e$ for capacitive coupling. Only the magnitude of the coupling is needed to calculate the physical dimensions of the coupling iris as illustrated in the example given in Section 14.4. In the analysis above, we assume a synchronous design, where all resonators in the filter have the same resonance frequency. Expressions for the case of nonsynchronous designs are given in [4].

14.2.2 Interresonator Coupling Calculation Using *S*-Parameters

The use of electric and magnetic wall symmetry to calculate interresonator coupling is difficult to implement experimentally. It can be only theoretically calculated and requires the use of an EM simulator that has an eigenvalue calculation capability such as HFSS [5] or CST [6]. Alternatively, the two resonant frequencies f_e and f_m can be calculated by connecting the two resonators in a two- or one-port network configuration as shown in Figure 14.5. The ports need to be weakly coupled to the resonator for this approach to work. A sketch of S_{11} is given in Figure 14.6. The two peaks of S_{11} represent the two frequencies f_e and f_m. The nature of the coupling, whether magnetic or electric, can be determined from the phase information of S_{11} [4].

In waveguide and microstrip filters where the resonators are in the form of transmission lines, the interresonator coupling can be calculated directly from the *S*-parameters by treating the coupling element as a discontinuity between two transmission lines. For example, in waveguide filters where interresonator coupling is realized using an iris, the equivalent circuit of the

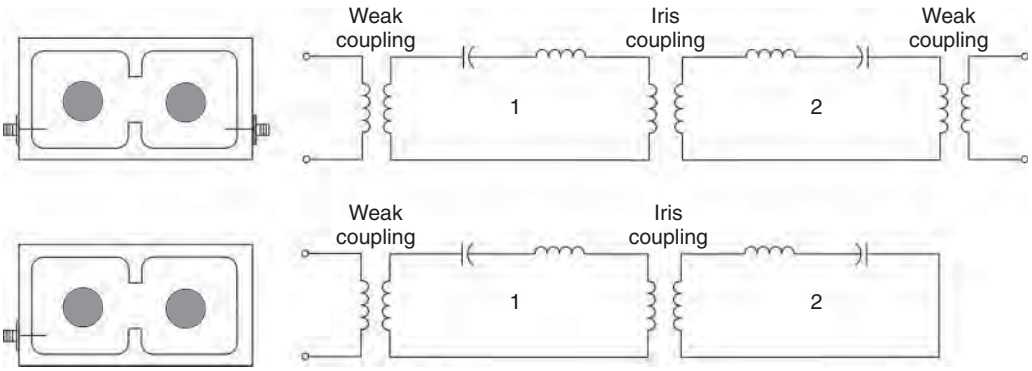


Figure 14.5 Measurement of interresonator coupling using a two-port network and a one-port network.

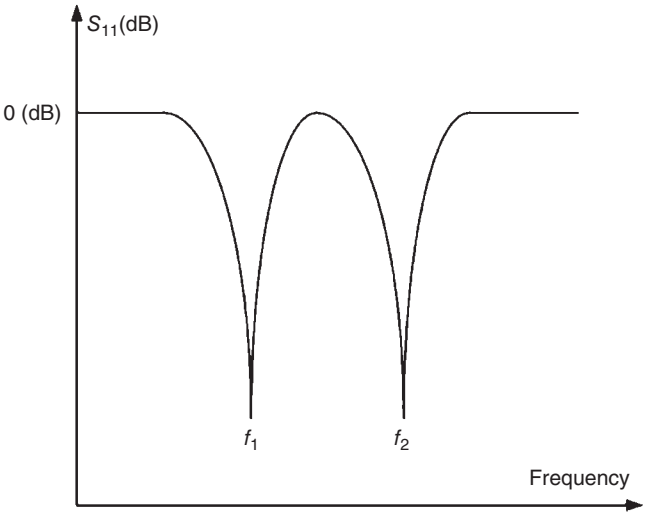


Figure 14.6 The return loss of the circuits shown in Figure 14.5.

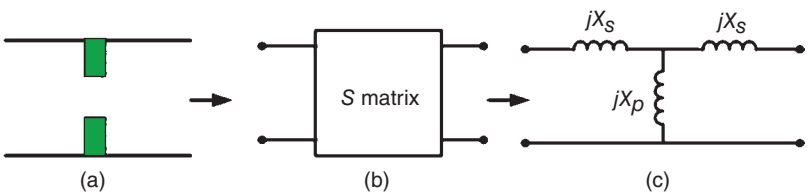


Figure 14.7 Interresonator coupling calculation using *S*-parameters: (a) coupling iris in a rectangular waveguide, (b) equivalent *S* matrix, and (c) equivalent T network that is used to calculate the coupling between two resonators.

discontinuity formed by the iris can be represented by a T network as shown in Figure 14.7. The elements of the T network are obtained by converting the two-port *S* matrix of the waveguide discontinuity to the *Z* matrix using the formula given in Table 5.2. The shunt inductance determines the value of the *K*-inverter, while the series inductance represents the loading on the adjacent resonators. Section 14.5 gives a detailed example of the use of this approach in calculating interresonator coupling.

14.3 Calculation of Input/Output Coupling

14.3.1 Frequency Domain Method

Figure 14.8 shows the equivalent circuit for the first resonator when coupled to the input feed source. The input coupling is represented by the conductance G. The reflection coefficient looking into the single resonator, with respect to a feed line of characteristic admittance G, is given by

$$S_{11} = \frac{G - Y_{\text{in}}}{G + Y_{\text{in}}} = \frac{1 - Y_{\text{in}}/G}{1 + Y_{\text{in}}/G} \tag{14.25}$$

The impedance Y_{in} seen looking into the single resonator is

$$Y_{\text{in}} = j\omega C + \frac{1}{j\omega L} = j\omega_0 C\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \tag{14.26}$$

where $\omega_0 = 1/\sqrt{LC}$. For frequencies near resonance, $\omega = \omega_0 + \Delta\omega$, where $\Delta\omega \ll \omega_0$, Y_{in} is approximated by

$$Y_{\text{in}} \approx j\omega_0 C\frac{2\Delta\omega}{\omega_0} \tag{14.27}$$

Substituting equation (14.27) in the reflection coefficient S_{11} of the resonator and noting that $Q_e = (\omega_0 C/G)$, we get

$$S_{11} = \frac{1 - jQ_e(2\Delta\omega/\omega_0)}{1 + jQ_e(2\Delta\omega/\omega_0)} \tag{14.28}$$

When the frequency offset from resonance is $\Delta\omega_{\mp} = \mp\omega_0/2Q_e$, the phase of S_{11} takes $\pm 90°$. Thus the Q_e is related to the $\pm 90°$ bandwidth in the phase of S_{11} by $\Delta\omega_{\pm 90} = \Delta\omega_+ - \Delta\omega_- = \omega_0/Q_e$:

$$Q_e = \frac{\omega_0}{\Delta\omega_{\pm 90}} \tag{14.29}$$

On the other hand, in view of Chapter 8, the relation between the external quality factor Q_e and normalized input impedance R of the coupling matrix model is given by $Q_e = \omega_0/[R \cdot (\omega_2 - \omega_1)]$, where $(\omega_2 - \omega_1)$ is the filter bandwidth in radians per second. Thus R can be written as

$$R = \frac{\Delta\omega_{\pm 90}}{(\omega_2 - \omega_1)} \tag{14.30}$$

The normalized impedances R_1 and R_2 of the coupling matrix $[M]$ model can then be extracted from simulation or measurement of the $\pm 90°$ phase of the reflection coefficient of the input and output resonators, respectively.

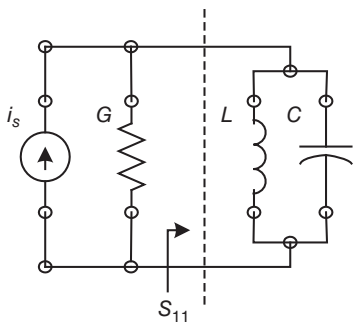


Figure 14.8 Equivalent circuit of the input coupling and first resonator.

14.3.2 Group Delay Method

The group delay method for determining the input coupling to a single resonator is based on analysis of the group delay of the reflection coefficient S_{11}. Equation (14.28) can be rewritten as

$$S_{11} = \left| \frac{1 - jQ_e(2\Delta\omega/\omega_0)}{1 + jQ_e(2\Delta\omega/\omega_0)} \right| \angle\phi \tag{14.31}$$

where

$$\phi = -2 \arctan\left(Q_e \left(2\frac{(\omega - \omega_0)}{\omega_0} \right) \right) \tag{14.32}$$

Using

$$\frac{d}{dx}(\arctan(x)) = \frac{1}{1 + x^2}$$

the group delay $\tau = -(\partial\phi/\partial\omega)$ is given by

$$\tau = \frac{4Q_e}{\omega_0} \frac{1}{1 + (2Q_e(\omega - \omega_0)/\omega_0)^2} \tag{14.33}$$

Note that the group delay has its maximum value at resonance when $\omega = \omega_0$: $\tau_{\max} = \tau(\omega_0) = (4Q_e/\omega_0)$. Since Q_e is related to the normalized impedance R of the coupling matrix model as $Q_e = \omega_0/[R \cdot (\omega_2 - \omega_1)]$, the normalized impedance R can then be written as

$$R = \frac{4}{(\omega_2 - \omega_1)} \frac{1}{\tau(\omega_0)} \tag{14.34}$$

Thus, the external Q_e values and hence the normalized impedances R_1 and R_2 of the coupling matrix model can be extracted by calculating the group delay of the reflection coefficient of the input and output resonators, respectively, at ω_0.

14.4 Design Example of Dielectric Resonator Filters Using the Coupling Matrix Model

To demonstrate the technique for using the coupling matrix filter circuit model and an EM simulator to synthesize the physical dimensions of the filter, we consider the design of a four-pole dielectric resonator filter. Initially, the design is specified in terms of the Chebyshev bandpass response with a particular center frequency, bandwidth, and return loss. The filter specifications translate into desired coupling matrix elements M_{ij}, R_1, and R_2. The objective is to translate these coupling elements into physical dimensions.

Figure 14.9 shows the four-pole dielectric resonator filter structure. It consists of four dielectric resonators operating in single $TE_{01\delta}$ modes. The coupling between the resonators is realized using irises, whereas the input/output coupling is achieved with the use of probes. The filter is assumed to have the following specifications:

Filter order	$n = 4$
Center frequency	$f_0 = 1930$ MHz
Bandwidth	BW = 15 MHz (fractional bandwidth = BW/f_0 = 0.007772)
Return loss	RL = 20 dB (equivalent to a passband ripple of 0.0436 dB)

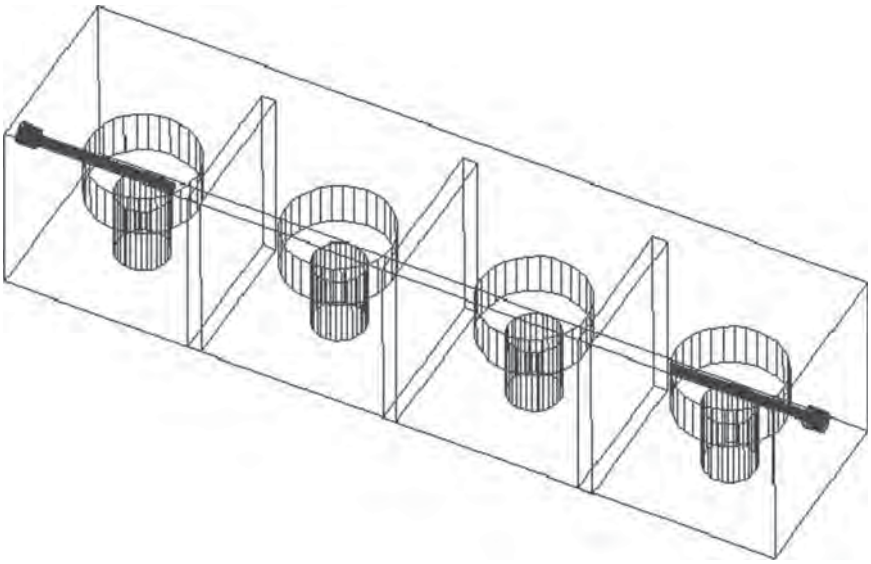


Figure 14.9 The four-pole dielectric resonator filter.

From equation (14.8), the g_k values corresponding to a return loss of 20 dB are $g_0 = 1$, $g_1 = 0.9332$, $g_2 = 1.2923$, $g_3 = 1.5795$, $g_4 = 0.7636$, and $g_5 = 1.2222$.

Using equations (14.1) and (14.2), the elements of the coupling matrix are given by

$$R_1 = \frac{1}{g_0 g_1} = 1.0715$$

$$R_n = \frac{1}{g_4 g_5} = 1.0715$$

$$M_{12} = \frac{1}{\sqrt{g_1 g_2}} = 0.9106$$

$$M_{23} = \frac{1}{\sqrt{g_2 g_3}} = 0.6999$$

$$M_{34} = \frac{1}{\sqrt{g_3 g_4}} = 0.9106$$

The diagonal elements $M_{11} = M_{22} = M_{33} = M_{44} = 0$, and $M_{ij} = M_{ji}$.

In view of Chapter 8, with knowledge of the coupling matrix $[M]$, the filter S-parameters are given by

$$S_{11} = 1 + 2jR_1[\lambda \mathbf{I} - j\mathbf{R} + \mathbf{M}]_{11}^{-1} \tag{14.35}$$

$$S_{21} = -2j\sqrt{R_1 R_n}[\lambda \mathbf{I} - j\mathbf{R} + \mathbf{M}]_{n1}^{-1} \tag{14.36}$$

where $\mathbf{R}$ is an $n \times n$ matrix with all entries zero except $[\mathbf{R}]_{11} = R_1$ and $[\mathbf{R}]_{nn} = R_n$, $\mathbf{M}$ is the $n \times n$ symmetric coupling matrix, $\mathbf{I}$ is the $n \times n$ identity matrix, and

$$\lambda = \frac{f_0}{\mathrm{BW}}\left(\frac{f}{f_0} - \frac{f_0}{f}\right)$$

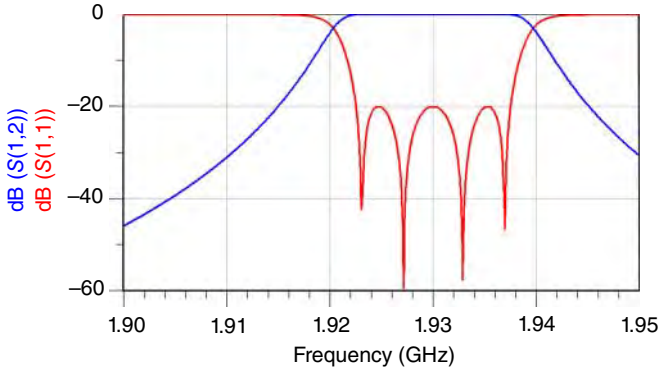


Figure 14.10 The simulated ideal response of the filter based on the coupling matrix model.

Figure 14.10 illustrates the simulated ideal response of the filter using the coupling matrix elements of the four-pole filter.

14.4.1 Calculation of Dielectric Resonator Cavity Configuration

Figure 14.11 illustrates the cavity configuration and associated dimensions. The dielectric resonator is assumed to have $\varepsilon_r = 34$ and is mounted inside a square cavity of dimensions $(50.8 \times 50.8 \times 51.5\,\text{mm}^3)$ on a dielectric cylindrical support post with $\varepsilon_r = 10$, a diameter of $D_s = 14.22$ mm and a height L_s of 20.32 mm. Using HFSS (version 8.5), the EM simulation results show that a dielectric resonator operating in $\text{TE}_{01\delta}$ mode, with a diameter of $D = 29.87$ mm and a height of $L_d = 12.21$ mm, has a resonant frequency of 1.931 GHz. The second mode would have a resonant frequency of 2.292 GHz; thus, the resonator exhibits a spurious-free window of about 300 MHz.

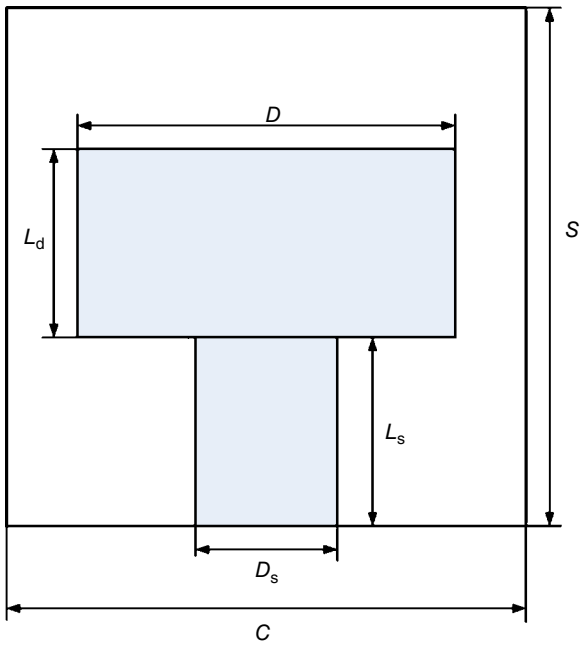


Figure 14.11 Dielectric resonator configuration, $\varepsilon_r = 34$; $D = 29.87$ mm; $L_d = 12.21$ mm; support, $\varepsilon_{rs} = 10$; $D_s = 14.22$ mm; $L_s = 20.32$ mm; cavity, $S = 51.5$ mm; $C = 50.8$ mm; mode 1, $f_0 = 1.931$ GHz (HFSS version 8.5); and mode 2, $f_0 = 2.292$ GHz (HFSS version 8.5).

14.4.2 Calculation of Iris Dimensions for Interresonator Coupling

We use the electric wall and magnetic wall symmetry analysis presented in Section 14.2.1 for the calculation of interresonator coupling. The HFSS simulator [5] is employed to calculate the resonant frequency of the even and odd modes. Figure 14.12 illustrates the two-coupled resonator structure used in the HFSS simulations. Note that in calculating the even- and odd-mode resonant frequencies, we assumed the original cavity dimensions (i.e., $C = 50.8\,\text{mm}$, $S = 51.5\,\text{mm}$). We assume the same iris thickness of $T = 3.81$ mm for all irises. Table 14.3 shows the calculated values for f_e and f_m and the corresponding coupling factor k.

The physical coupling k_{ij} can be calculated from the normalized coupling M_{ij} as follows:

$$M_{ij} = \frac{f_0}{\text{BW}} \times k_{ij} \tag{14.37}$$

In view of Figure 14.13, and by using interpolation, we can determine the iris dimensions required to achieve the coupling elements M_{12}, M_{23}, and M_{34}. Table 14.4 gives the iris dimensions and associated coupling values for the structure shown in Figure 14.12.

One particular consideration in using this approach is the loading effects of irises and input and output couplings on the adjacent resonators. The result of this loading is a shift in the resonant frequency of the resonator. Adjusting the dimensions of the dielectric resonator to account for the loading is necessary to get good results from this approach.

We can possibly derive a theoretical expression for the shift in resonance frequency due to iris loading of adjacent cavities. We need to adjust the dimensions of the basic resonator

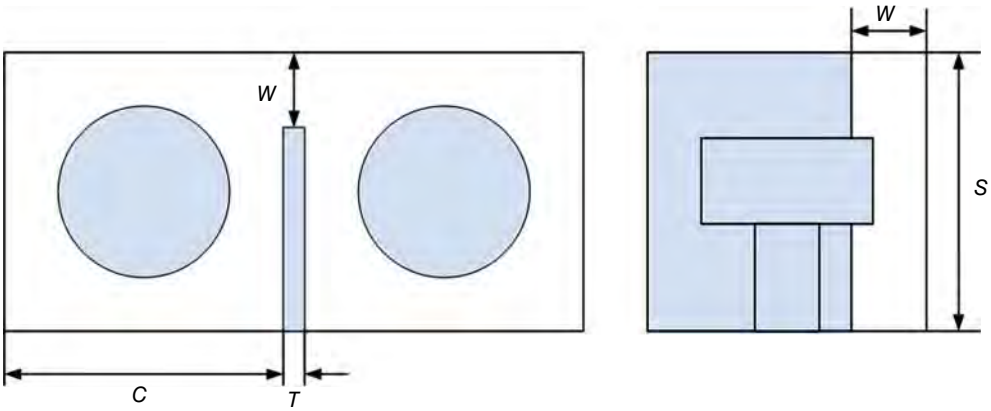


Figure 14.12 The two-coupled dielectric resonator structure.

Table 14.3 Calculated values for f_e and f_m for various values of iris width W.

W (in.)	f_m (GHz)	f_e (GHz)	$k = (f_e^2 - f_m^2)/(f_e^2 + f_m^2)$
0.05	1.9324	1.9355	0.001603
0.07	1.9285	1.9336	0.002641
0.1	1.9261	1.9319	0.003007
0.2	1.9232	1.9319	0.004513
0.25	1.9211	1.9313	0.005295
0.45	1.9170	1.9319	0.007742

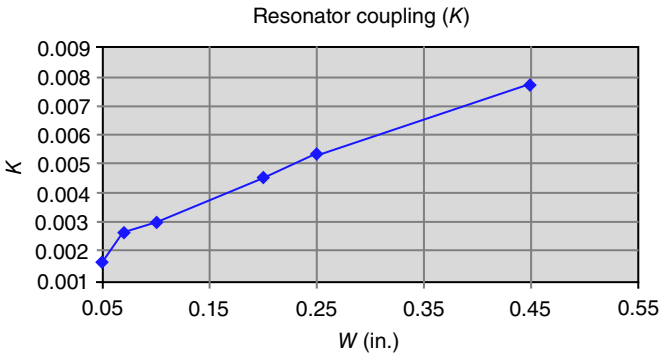


Figure 14.13 Resonator coupling (k) versus iris window (W), calculated from the data given in Table 14.2.

Table 14.4 Interresonator coupling and iris of the four-pole dielectric resonator filter.

M_{ij}	k	W (in.)
M_{12} (0.9106)	0.007077	0.396 (W_{12})
M_{23} (0.6999)	0.00544	0.262 (W_{23})
M_{34} (0.9106)	0.007077	0.396 (W_{34})

configuration shown in Figure 14.11 to account for this frequency shift. Alternatively, as a good approximation, the adjustment of the dielectric resonator dimensions can be obtained by solving the eigenvalue of the structure shown in Figure 14.14, which consists of the dielectric resonator, two irises, and two adjacent cavities with no dielectric resonators. The dielectric resonator dimensions (diameter or height or both) are then adjusted such that the cavity, with the two irises, resonates at the filter center frequency.

The structure shown in Figure 14.14 uses the iris dimensions that were calculated using the original resonator configuration given in Figure 14.11. This assumption is made since the interresonator coupling is not highly sensitive to a small change in resonator resonance frequency. With the use of HFSS [5], the diameter of the resonator is slightly adjusted such that the cavity with the two irises shown in Figure 14.14 resonates at 1.930 GHz. The diameter is found to be 29.66 mm; thus, in comparison with the starting resonator diameter given in Figure 14.11, the resonator diameters of the second and third dielectric resonators need to be reduced by 0.21 mm.

14.4.3 Calculation of Input/Output Coupling

The input coupling Q_e can be calculated from the values of the normalized input/output coupling resistance R_1 and R_n:

$$Q_e = \frac{f_0}{\text{BW} \times R_1} \tag{14.38}$$

For $R_1 = 1.07154$, the required $Q_e = 120.08$. We will calculate the Q_e using the group delay method outlined in Section 14.3.2. With the use of HFSS, the group delay is calculated from

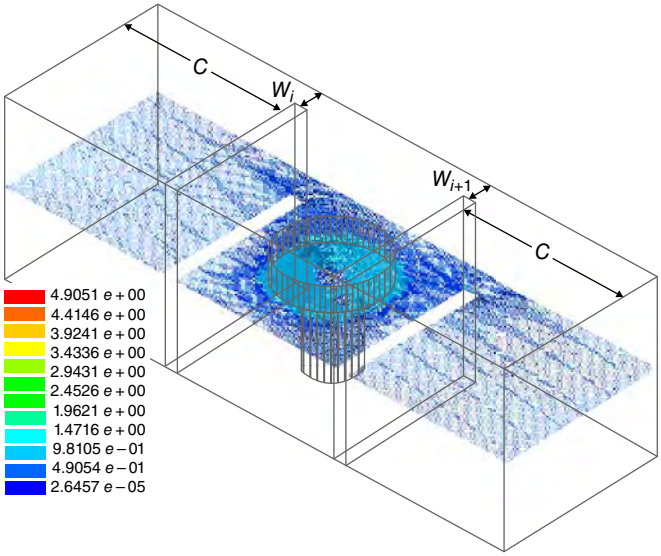
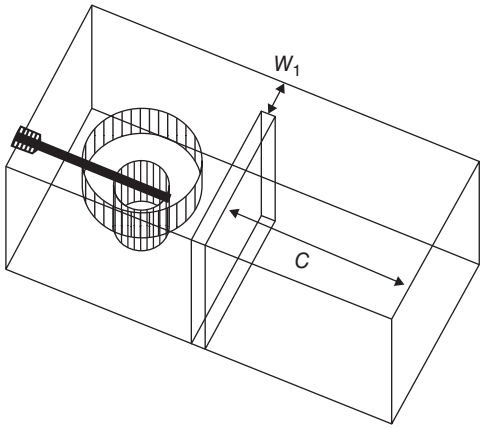


Figure 14.14 Calculation of the resonance frequency of the resonator including the effect of the iris. The adjacent cavities are detuned by removing the dielectric resonators.

Figure 14.15 Calculation of the Q_e and resonance frequency of the first resonator, factoring in the effect of iris loading.



the input reflection coefficient S_{11} of the structure shown in Figure 14.15. In order to factor in iris loading effects, we used the structure shown in Figure 14.15 rather than a single-cavity resonator. The first cavity is terminated by an iris W_{12}, and the second cavity is detuned by removing the dielectric resonator.

Figure 14.16 shows the details of the probe configuration we used in this filter, as well as the probe's dimensions. We need to vary the probe penetration depth H until we get the proper external Q_e value. However, the probe penetration depth H also affects the resonant frequency of the cavity. Thus, in the presence of the iris W_{12}, we need to adjust both the probe penetration H and dielectric resonator diameter D to achieve the required Q_e and the resonant frequency f_0. This can be achieved by following an iterative approach. First we assume the unloaded original resonator diameter of $D = 29.87$ mm and then use the group delay calculation and adjust H to

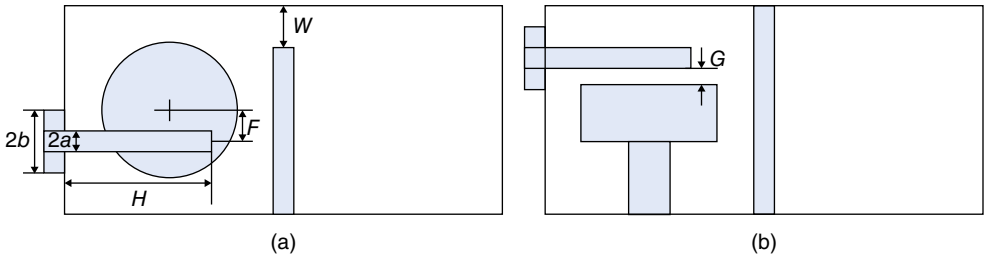


Figure 14.16 Physical dimensions of probe used for the first and last cavities: $a = 0.044$ in.; $b = 0.1$ in.; $G = 0.01$ in., $F = 7.47$ mm; $H =$ length of the probe inside cavity. (a) Top view and (b) side view.

get the required Q_e. Assuming this value of H, the dielectric resonator diameter is then adjusted using HFSS such that the cavity with the iris and probe resonates at the filter center frequency f_0. For that new value of D, the probe penetration H is readjusted to get the desired Q_e and the iterative process continues; H and D are adjusted individually to get Q_e and f_0 iteratively until the process converges. Alternatively one may change the probe dimensions and location (H, G, and F shown in Figure 14.16) until both the required Q_e and resonance frequency f_0 are achieved. The required Q_e is 120.014 and the center frequency is 1.930 GHz. On the basis of HFSS simulations, these values are achieved with the use of probe penetration depth H of 37.84 mm and a dielectric resonator diameter D of 29.87 mm.

Table 14.5 summarizes the dimensions of the designed four-pole filter. Note that we used the iris dimensions that we calculated from the model given in Figure 14.12. This is an acceptable approximation since the interresonator coupling is not highly sensitive to the resonance frequency. Figure 14.17 illustrates the HFSS simulated results of the filter based on the physical dimensions given in Table 14.5. A comparison between the simulated and the ideal response is given in the same figure. The results obtained are in reasonable agreement with the ideal response and can be considered a reasonable initial solution. A more accurate solution can be obtained by fine-tuning or applying any of the advanced EM modeling techniques described in Chapter 15.

In this example, we have changed the dielectric resonator diameters to factor in the effect of iris and input/output probe loading. This leads to a solution where the diameters of resonators 1 and 4 differ from those of resonators 2 and 3. Machining and grinding of dielectric resonators

Table 14.5 Dimensions of designed four-pole dielectric resonator filter (cavity dimensions are given in Figure 14.11).

Parameters	Dimensions (mm)
Irises	W_{12} and $W_{34} = 10.058$
	$W_{23} = 6.6548$
DR resonator diameters	$D_1 = D_4 = 29.87$
	$D_2 = D_3 = 29.66$
Input/output probe dimensions	$H = 37.84$
	$G = 0.254$
	$a = 1.117$
	$b = 2.54$
	$F = 7.47$

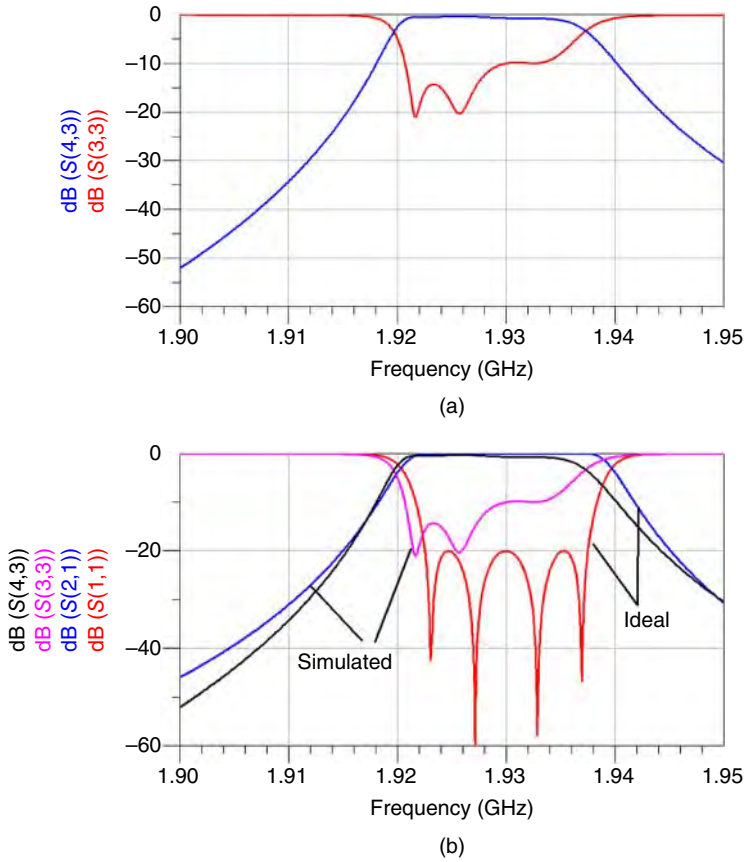


Figure 14.17 (a) HFSS simulation results based on physical dimensions given in Table 14.5 and (b) comparison with ideal response.

to get the proper length is usually an expensive process. In practical applications, the same diameter and same height are used for the four dielectric resonators and tuning screws are usually used to compensate for design imperfections as well as iris and probe loading effects.

14.5 Design Example of a Waveguide Iris Filter Using the Impedance Inverter Model

In this section we consider the design of the four-pole Chebyshev inductive iris-coupled waveguide filter shown in Figure 14.18a. We combine the K-impedance inverter circuit model shown in Figure 14.2b with an EM simulator to calculate the filter physical dimensions. The filter consists of rectangular half-wavelength resonators operating in TE_{101} modes separated by inductive irises. A three-dimensional (3D) illustration of a five-pole version of this filter is shown in Figure 14.1. With the use of an EM simulator, the scattering matrix of the iris and consequently its equivalent T network can be calculated. Each iris is represented by two series inductances denoted by X_s and a shunt inductance denoted by X_p. The waveguide iris filter can then be deduced to the equivalent circuit shown in Figure 14.18b. To arrange the circuit in the form of the K-inverter impedance model shown in Figure 14.2b, we use the impedance inverter circuit

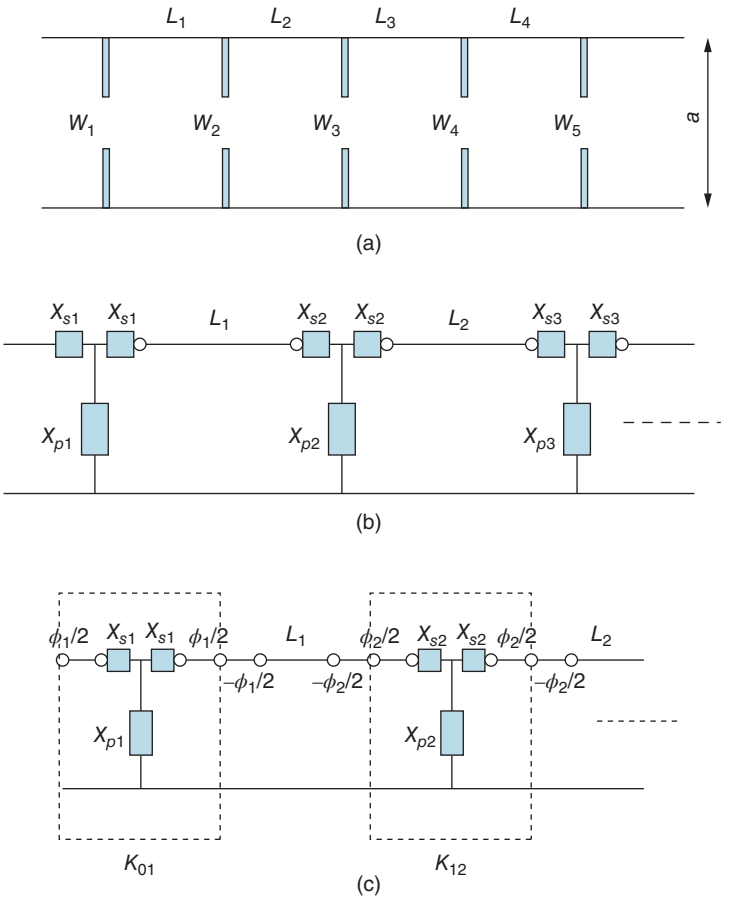


Figure 14.18 Four-pole waveguide iris filter: (a) top view of filter, (b) equivalent circuit, and (c) modified equivalent circuit.

shown in Table 14.2e. It consists of an inductive T network and two sections of $\phi/2$ on each side. The inverter is created by adding a length of $\phi/2$ and a length of $-\phi/2$ on each side of the discontinuity as shown in Figure 14.18c. Note that adding the length of $\phi/2$ and a length of $-\phi/2$ does not change the original circuit.

The circuit of Figure 14.18c is exactly in the same form as the K-impedance inverter circuit shown in Figure 14.2b. The resonators in this case are waveguide transmission lines of length L_n connected to two transmission lines of artificial lengths $-\phi_n/2$ and $-\phi_{n+1}/2$. These lengths represent the loading on the resonator from the adjacent coupling inverters. The filter is assumed to have the following specifications:

Filter center frequency	$f_0 = 11$ GHz
Bandwidth	BW = 300 MHz (fractional bandwidth = 0.02727)
Return loss	RL = 25 dB (0.01376 dB ripple)
Waveguide dimensions	WR90: $a = 0.9$ in., $b = 0.4$ in. ($a = 22.86$ mm, $b = 10.16$ mm)
Iris thickness	2 mm (the same for all irises)

Using equation (14.8) or Table 14.1, the lowpass g_k values corresponding to a return loss of 25 dB are $g_0 = 1$, $g_1 = 0.75331$, $g_2 = 1.22520$, $g_3 = 1.37121$, $g_4 = 0.67310$, and $g_5 = 1.11917$.

The filter center frequency and band-edge frequencies are related by $f_0 = \sqrt{f_1 f_2}$, $\text{BW} = f_1 - f_2$, which gives

$$f_1 = 10.85\,\text{GHz}, \quad f_2 = 11.15\,\text{GHz}$$

$$\Delta = \frac{\lambda_{g1} - \lambda_{g2}}{\lambda_{g0}} = 0.0423$$

$$\frac{K_{01}}{Z_0} = \frac{K_{45}}{Z_0} = \sqrt{\frac{\pi\Delta}{2g_0g_1}} = 0.297$$

$$\frac{K_{12}}{Z_0} = \frac{K_{34}}{Z_0} = \frac{\pi\Delta}{2\sqrt{g_1g_2}} = 0.0692$$

$$\frac{K_{23}}{Z_0} = \frac{\pi\Delta}{2\sqrt{g_2g_3}} = 0.0513$$

Using an EM simulator based on the mode matching technique [7], the scattering parameters of the iris can be calculated. From Table 5.2, the elements of the T network shown in Figure 14.7, X_s and X_p are related to the scattering parameters as

$$j\frac{X_s}{Z_0} = \frac{1 - S_{12} + S_{11}}{1 - S_{11} + S_{12}} \tag{14.39}$$

$$j\frac{X_p}{Z_0} = \frac{2S_{12}}{(1 - S_{11})^2 - S_{12}^2} \tag{14.40}$$

where S_{11} and S_{21} are the scattering parameters of the dominant TE_{10} mode. For the K-impedance inverter shown in Table 14.2e, X_s and X_p are related to K and ϕ as follows:

$$\frac{K}{Z_0} = \left|\tan\left(\frac{\phi}{2} + \arctan\frac{X_s}{Z_0}\right)\right| \tag{14.41}$$

$$\phi = -\arctan\left(2\frac{X_p}{Z_0} + \frac{X_s}{Z_0}\right) - \arctan\frac{X_s}{Z_0} \tag{14.42}$$

Both X_s and X_p and consequently K and ϕ are functions of the iris width W. Since these functions are not explicit, one needs to run the EM simulator for a range of iris widths to calculate S-parameters and the corresponding X_s, X_p, ϕ, and K. This will allow the construction of a lookup table or a curve for K versus W. With knowledge of the required K value, we can identify the iris width W and then use the associated ϕ_i values to calculate the resonator lengths:

$$l_r = \frac{\lambda}{2\pi}\left[\pi + \frac{1}{2}(\phi_r + \phi_{r+1})\right], \quad r = 1, \ldots, N \tag{14.43}$$

Table 14.6 summarizes the physical dimensions of the designed filter. Figure 14.19 shows the mode matching EM simulation results. A comparison obtained using the dimensions given in Table 14.6 with the ideal response that is generated from the coupling matrix is also given in the same figure. The good results obtained are attributed to the fact that the K-impedance inverter model makes it possible to accurately include the effect of iris loading on the resonator as seen in equation (14.43). The analysis yields good results for narrowband filters, and for wideband filters we need to follow the analysis given in [8] for accurate calculation of the midband guide wavelength λ_g.

Table 14.6 Synthesized physical dimension of four-pole waveguide filter.

Parameters	X_S	X_P	Phase	Width (mm)
$K_{01} = K_{45} = 0.297$	0.358	0.120	−0.816	10.499
$K_{12} = K_{34} = 0.0692$	0.071	0.070	−0.279	6.706
$K_{23} = 0.0513$	0.052	0.063	−0.228	6.147
Cavity length (mm)			$L_1 = L_4 = 14.022$	
			$L_2 = L_3 = 15.611$	

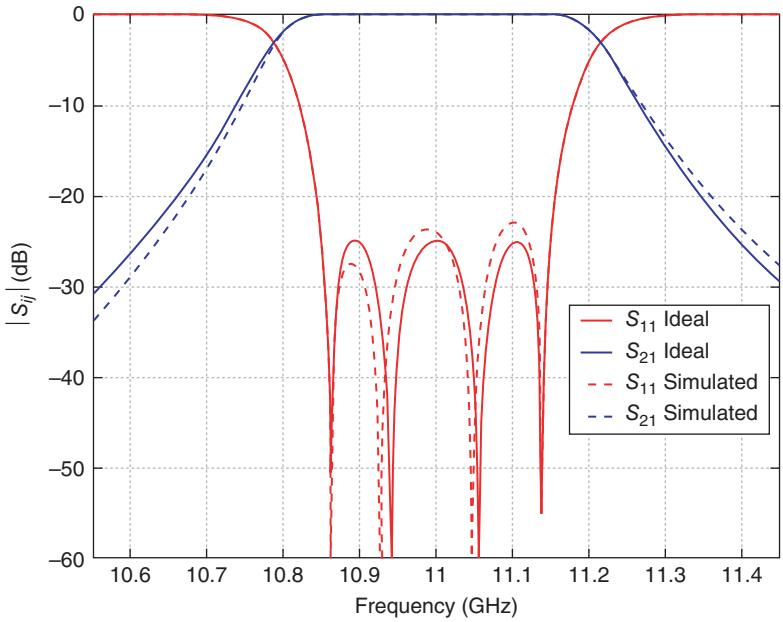


Figure 14.19 A comparison between the ideal response and the mode matching EM simulation results of the designed filter whose dimensions are given in Table 14.6.

14.6 Design Example of a Microstrip Filter Using the *J*-Admittance Inverter Model

In this section we consider the design of the six-pole microstrip filter shown in Figure 14.20a. The filter consists of six microstrip resonators capacitively coupled to each other. We use the *J*-admittance inverter filter model given in Figure 14.2c together with an EM simulator for planar circuits to design this filter. One can easily show that with the use of an EM simulator, the capacitive coupling section can be represented by a π network consisting of a series capacitance C_s and two parallel capacitances of C_p as shown in Figure 14.21. The equivalent circuit of the filter can then be represented as shown in Figure 14.20b. The circuit needs to be modified to conform to the *J*-admittance model given in Figure 14.2c. We use in this example the *J*-admittance inverter model shown in Table 14.2j, which consists of a π network and two

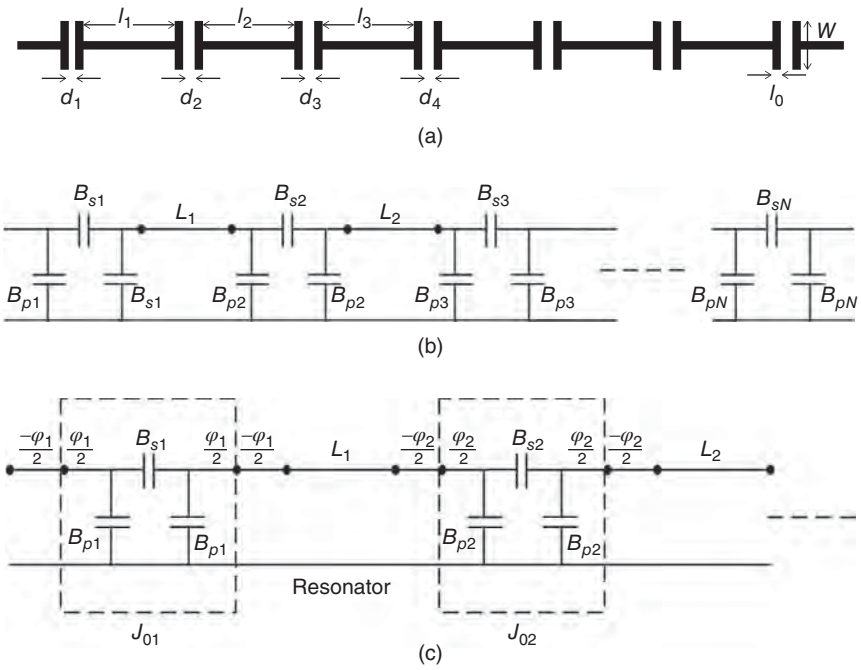


Figure 14.20 The six-pole capacitive coupled microstrip filter: (a) top view of filter, (b) equivalent circuit, and (c) modified equivalent circuit.

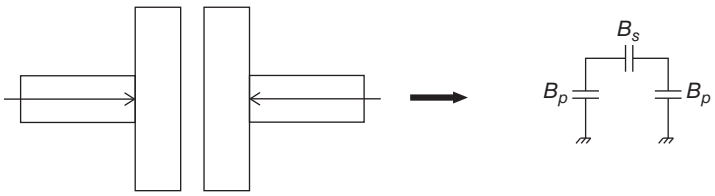


Figure 14.21 The equivalent circuit of the capacitive microstrip discontinuity.

transmission-line lengths attached to the sides. In order to modify the circuit to fit the selected model, we add a negative and a positive length of an electrical length ϕ on each side of the capacitive discontinuity. The resultant circuit is shown in Figure 14.20c. It now has the form of the J-admittance inverter filter model, where the resonators consist of microstrip transmission lines of lengths $L_1 + \phi_i + \phi_{i+1}$. The microstrip filter is assumed to have the following specifications:

Center frequency	$f_0 = 2$ GHz
Bandwidth	BW = 40 MHz (relative bandwidth = 0.02)
Return loss	20 dB (equivalent to a ripple of 0.0436 dB)
Substrate of input/output impedance	$\varepsilon_r = 10$ and thickness 1.27 mm
	$Z_0 = 50\ \Omega$ (microstrip width = 1.234 mm)

With the knowledge of the lowpass g_k values, $g_0 = 1, g_1 = 0.994, g_2 = 1.4131, g_3 = 1.8933, g_4 = 1.5506, g_5 = 1.7253, g_6 = 0.8141$, and $g_7 = 1.2210$, we can express the J-admittance inverters as

$$J_{01} = J_{67} = Y_0\sqrt{\frac{\pi\Delta}{2g_0g_1}} = 3.55 \times 10^{-3}\,\text{s}$$

$$J_{12} = J_{56} = \frac{Y_0\pi\Delta}{2\sqrt{g_1g_2}} = 5.3 \times 10^{-4}\,\text{s}$$

$$J_{23} = J_{45} = \frac{Y_0\pi\Delta}{2\sqrt{g_2g_3}} = 3.84 \times 10^{-4}\,\text{s}$$

$$J_{34} = \frac{Y_0\pi\Delta}{2\sqrt{g_3g_4}} = 3.67 \times 10^{-4}\,\text{s}$$

where Δ is the fractional bandwidth, which is given by

$$\Delta = \frac{f_2 - f_1}{f_0} = 0.02$$

Note that we used the frequency rather than λ_g in defining Δ since the effective dielectric constant of the microstrip line ε_{eff} can be assumed to be constant over such narrow frequency band.

Using em sonnet [9] or Agilent Momentum [10], the scattering parameters of the capacitive discontinuity can be calculated. The elements of the Y-network parameters can be written in terms of the scattering matrix as

$$\frac{jB_p}{Y_0} = \frac{1 - S_{12} - S_{11}}{1 + S_{11} + S_{12}} \tag{14.44}$$

$$\frac{jB_s}{Y_0} = \frac{2S_{12}}{(1 + S_{11})^2 - S_{12}^2} \tag{14.45}$$

where S_{11} and S_{21} are the scattering parameters of the capacitive microstrip discontinuity. In view of equations (14.15) and (14.16), the J and ϕ in terms of B_s and B_p are given by

$$\frac{J}{Y_0} = \left|\tan\left(\frac{\phi}{2} + \arctan\frac{B_p}{Y_0}\right)\right| \tag{14.46}$$

$$\phi = -\arctan\left(2\frac{B_s}{Y_0} + \frac{B_p}{Y_0}\right) - \arctan\frac{B_p}{Y_0} \tag{14.47}$$

We need to find the dimensions of the capacitive discontinuity that give the required J value. There are three variables, which we can adjust in the microstrip capacitive discontinuity shown in Figure 14.21 to get the proper J: the width l_0, the width W, and the gap d. We assumed the same width W and the same width l_0 for all capacitive discontinuities. The values selected for W and l_0 are $W = 6$ mm and $l_0 = 1.234$ mm.

We only need to change the capacitive gap d in order to get the required J-inverter value. The discontinuity is simulated for a range of capacitive gaps d, and a lookup table is obtained relating d versus J and ϕ. The capacitive gaps $d_0, d_1, \ldots, d_N$ are then determined according to the J values given in Table 14.9. With knowledge of ϕ, one can calculate the resonator lengths as

$$l_r = \frac{\lambda_{g0}}{2\pi}\left[\pi + \frac{1}{2}(\phi_r + \phi_{r+1})\right], \quad r = 1, \ldots, N \tag{14.48}$$

Table 14.7 Lookup table for finding desired gap sizes using Agilent ADS circuit simulator.

d (mm)[a)]	Y_{11}	Y_{12}	J	ϕ
0.04	j0.0159294	$-j$0.0057781	0.0036	−1.296
0.06	—	—	0.0035	—
0.08	—	—	0.0034	—
0.10	—	—	0.0033	—
1.00	—	—	9.4097e−004	—
1.35	—	—	5.6546e−004	—
1.39	j0.0137513	$-j$0.0007852	5.3333e−004	−1.2037
1.40	—	—	5.2562e−004	—
1.50	—	—	4.5418e−004	—
1.60	—	—	3.9241e−004	—
1.61	j0.0137280	$-j$0.0005688	3.8671e−004	−1.2026
1.63	—	—	3.7558e−004	—
1.65	j0.0137252	$-j$0.0005365	3.6479e−004	−1.2024
1.70	—	—	3.3906e−004	—

a) The gaps that correspond to the required J values.

Table 14.8 Final design dimensions (mm) obtained using Agilent ADS circuit simulator.

	1	2	3	4
l	17.312	17.739	17.745	—
d	0.040	1.390	1.610	1.650

The lookup table that relates the gap d to J and ϕ is generated using both the Agilent ADS circuit simulator [11] and the Agilent Momentum EM simulator [10]. Table 14.7 shows the lookup table generated by the ADS circuit simulator. Table 14.8 shows the filter dimensions as calculated based on Table 14.7, whereas Figure 14.22 shows the simulated results using the ADS circuit simulator. A slight deviation from the ideal response is observed and is attributed to the fact the equivalent circuit of the gap capacitance is assumed to be frequency independent in this approach.

A more accurate analysis can be obtained if the lookup table is generated using an EM simulator. The results for d versus J and ϕ as calculated by the Agilent Momentum EM simulator are shown in Table 14.9. The corresponding filter design dimensions are given in Table 14.10. The EM simulation performance of the filter using the dimensions given in Table 14.9 is shown in Figure 14.23. The degradation in the roll-off at the lower band edge is attributed to radiation, since the circuit is simulated in Agilent Momentum with no box. It is noted also that the filter response exhibits a transmission zero on the left side. The deviation from the ideal response is due to cross-coupling between resonators, which is not accounted for in the approach presented in this chapter. However, the solution obtained is considered a good initial solution and can be further optimized using the advanced EM-based design techniques described in Chapter 15.

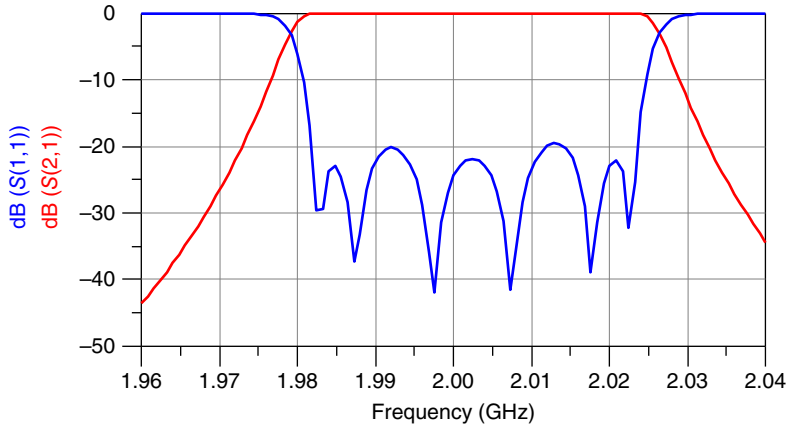


Figure 14.22 Agilent ADS circuit simulation of the design given in Table 14.7.

Table 14.9 Lookup table for finding desired gap sizes using Agilent Momentum EM simulator.

d(mm)	Y_{11}	Y_{12}	J	ϕ
0.10	—	—	0.0039	—
0.11	—	—	0.0038	—
0.133	j0.0165340	$-j$0.0059586	0.00361	−1.3304
0.138	—	—	0.00357	—
0.14	—	—	0.00356	—
0.15	—	—	0.0035	—
1.56	—	—	5.3439e−004	—
1.565	j0.0137929	$-j$0.0007837	5.3128e−004	−1.2065
1.58	—	—	5.2217e−004	—
1.60	—	—	5.1022e−004	—
1.61	—	—	5.0441e−004	—
1.70	—	—	4.5482e−004	—
1.80	—	—	4.0571e−004	—
1.82	—	—	3.9664e−004	—
1.83	—	—	3.9210e−004	—
1.85	j0.0137662	$-j$0.0005648	3.8330e−004	−1.2052
1.89	j0.0137637	$-j$0.0005397	3.6630e−004	−1.2050
1.90	—	—	3.6224e−004	—

Table 14.10 Final design dimensions (mm) obtained using Agilent Momentum EM simulator.

	1	**2**	**3**	**4**
l	17.188	17.763	17.770	—
d	0.133	1.565	1.850	1.890

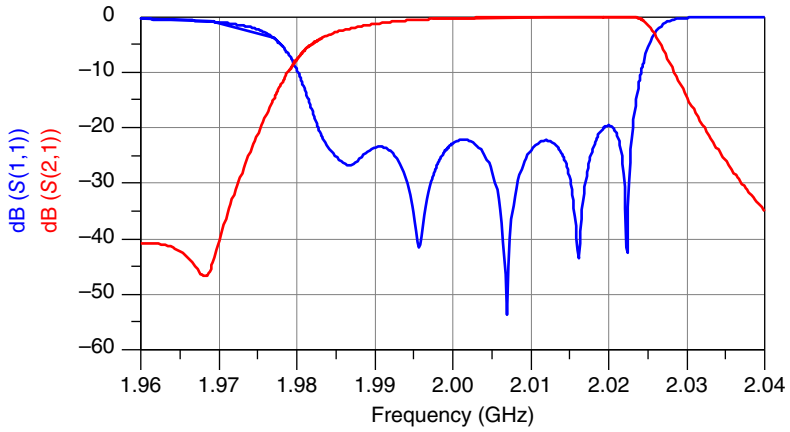


Figure 14.23 Agilent Momentum EM simulation results of the design given in Table 14.10.

Summary

This chapter illustrates how one can couple the filter circuit models with EM simulation tools to synthesize the physical dimensions of microwave filters. Three coupled resonator filter circuit models are briefly outlined: the coupling matrix model, the impedance inverter model, and the admittance inverter model. For Chebyshev filters, designers only need to know the low-pass g values to calculate the M_{ij} elements of the coupling matrix model and the impedance and admittance inverter values of the other two models. For other complicated filter functions, Chapters 7–10 can be used to construct the elements of these circuit models.

With the lack of an EM simulation tool, filter designers may use references such as Microwave Engineer's Handbook or refer to a large pool of papers published in 1970s and early 1980s where microwave discontinuity dimensions were approximated using closed-form expressions. These simple expressions were obtained either empirically or using approximations such as infinitely thin irises in waveguide realizations and quasi-static analyses in TEM structures.

This chapter demonstrates the use of a synthesis approach to translate the coupling matrix elements to physical dimensions using a dielectric resonator filter as an example. The approach is quite general and can be employed to single-mode or dual-mode filters realized using any technology (microstrip, coaxial, waveguide, or dielectric). However, it is ideal for dealing with microwave filters where the resonators are not in the form of coupled transmission-line resonators such as dielectric resonator filters, combline filters, or other filters with arbitrarily shaped resonators. Synthesis approaches that use the impedance and admittance inverter models are ideal for use in filters where the resonators have the form of coupled transmission-line resonators. Examples of such filters are iris-coupled waveguide filters, E-plane septum filters, microstrip capacitively coupled filters, and microstrip parallel-coupled filters. These two approaches are demonstrated by considering the design of a waveguide inductive iris filter and a capacitively loaded microstrip filter.

The three approaches described in this chapter provide reasonably accurate designs that can be used as initial solutions for further optimization by tuning or by using any of the advanced EM-based techniques presented in Chapter 15. The accuracy of these approaches is limited by the fact that the elements of the filter circuits are assumed to be frequency independent. Therefore, they are more applicable to narrowband designs. These approaches also do not take into consideration the cross-talk between various filter elements that is a pronounced problem in microstrip filters.

Practical filter designers may find the materials given in this chapter very useful to come up with a reasonably good initial design for Chebyshev filters. With a minimum tuning effort, this initial design can be easily turned into a perfect design.

References

1 Matthaei, G., Young, L., and Jones, E.M.T. (1980) *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*, Artech House, Norwood, MA.
2 Saad, T.S. *et al.* (eds) (1971) *Microwave Engineer's Handbook*, Artech House, Norwood, MA.
3 Collin, R.E. (1966) *Foundation for Microwave Engineering*, McGraw-Hill, New York.
4 Hong, J. and Lancaster, M.J. (2001) *Microstrip filters for RF/Microwave Applications*, Wiley, New York.
5 ANSYS HFSS, http://www.ansys.com/products/electronics/ansys-hfss.
6 CST, Microwave Studio, www.cst.com.
7 Cogollos, S. (2005) Design of Iris Filters Using Mode Matching Technique. Technical report. University of Waterloo.
8 Bui, L.Q., Ball, D., and Itoh, T. (1984) Broadband millimeter-wave *E*-plane bandpass filter. *IEEE Transactions on Microwave Theory and Techniques*, **32**, 1655–1658.
9 em sonnet, http://www.sonnetusa.com.
10 Agilent Momentum, http://eesof.tm.agilent.com.
11 Agilent ADS, http://eesof.tm.agilent.com.

15

Advanced EM-Based Design Techniques for Microwave Filters

A number of software packages are now commercially available for the electromagnetic (EM) simulation of microwave circuits. These EM simulation software tools provide very accurate results, allowing microwave filter designers to innovate and practice on various filter configurations in a very cost-effective approach without the need to build and test the hardware. However, commercial EM software packages are classified as "simulation tools" and not as "design tools." Although they may help users to check the validity of their designs, they cannot be used directly to come up with the design itself. Filter designers must employ other tools and techniques, together with the EM simulator, to carry out the design process. Several EM-based techniques are available for the design of microwave filters; these techniques are divided into three categories:

- EM-based synthesis techniques (EM simulator + circuit models)
- EM-based optimization techniques (EM simulator + optimization tools)
- EM-based advanced design techniques (EM simulator + circuit models + optimization tools)

Chapter 14 addresses the first category of EM-based synthesis techniques. In this chapter we address the other two categories with a focus on the third category of EM-based advanced design techniques.

15.1 EM-Based Synthesis Techniques

Chapter 14 is dedicated to description of the category of EM-based synthesis design techniques. Three examples are given in Chapter 14, where we demonstrate how EM simulators can be used to synthesize the physical dimensions of dielectric resonator, waveguide, and microstrip filters. In view of the results given in Chapter 14, it is clear that EM-based synthesis techniques do not provide results that fully meet the design specifications. This is attributed to the fact that the cross-coupling between nonadjacent resonators is neglected in these techniques and also that the coupling elements are calculated at the filter center frequency. Although these techniques provide reasonable results for waveguide filters, they may fail to provide accurate results for planar filter structures, where the cross-coupling between the resonators is substantial. These techniques also fail to provide accurate results for coaxial and dielectric resonator filters, where it is difficult to accurately account for the effect of iris loading over the filter bandwidth. However, EM-based synthesis techniques are useful in obtaining an initial design that can be further optimized by manual tuning or by applying EM-based optimization as described in the following section.

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

15.2 EM-Based Optimization Techniques

In these techniques, an accurate EM-based simulator is linked to an optimization software tool that tunes the filter dimensions to achieve the desired response. There are various optimization techniques available in the literature [1–3]. The two main classes are optimization techniques that require an initial solution such as a gradient-type approach [4] and techniques that do not require an initial solution such as a genetic algorithm [5]. Usually a good initial solution reduces the optimization process time and provides a superior design. In the optimization process, the physical parameters of the filter are iteratively modified by an optimization package until the requirements are satisfied. The flowchart of the design process is shown in Figure 15.1. The initial solution is easily obtained from any synthesis technique such as those described in Chapter 14.

The principal engine of the process is the EM-based simulation tool. The accuracy of the design is determined mainly by the accuracy of this tool. Various forms for the accurate simulator tool are outlined in Figure 15.1. The tool could be a commercial EM simulator or any full-wave analysis tool that is available to the designer. Several variations of this tool can be employed, including the semi-EM simulator or the EM simulator with adaptive frequency sampling, or the use of function interpolations or networks that emulate the EM simulator. Examples of these EM-based interpolations are neural networks, fuzzy logic, and multidimensional Cauchy techniques.

The parameters to be optimized are the filter physical dimensions, which are varied by the optimization routine to minimize an objective function. In determining the error function (objective function) for the optimization process, the filter performance is specified by simulating several frequency points within the passband and stopband as depicted in Figure 15.2.

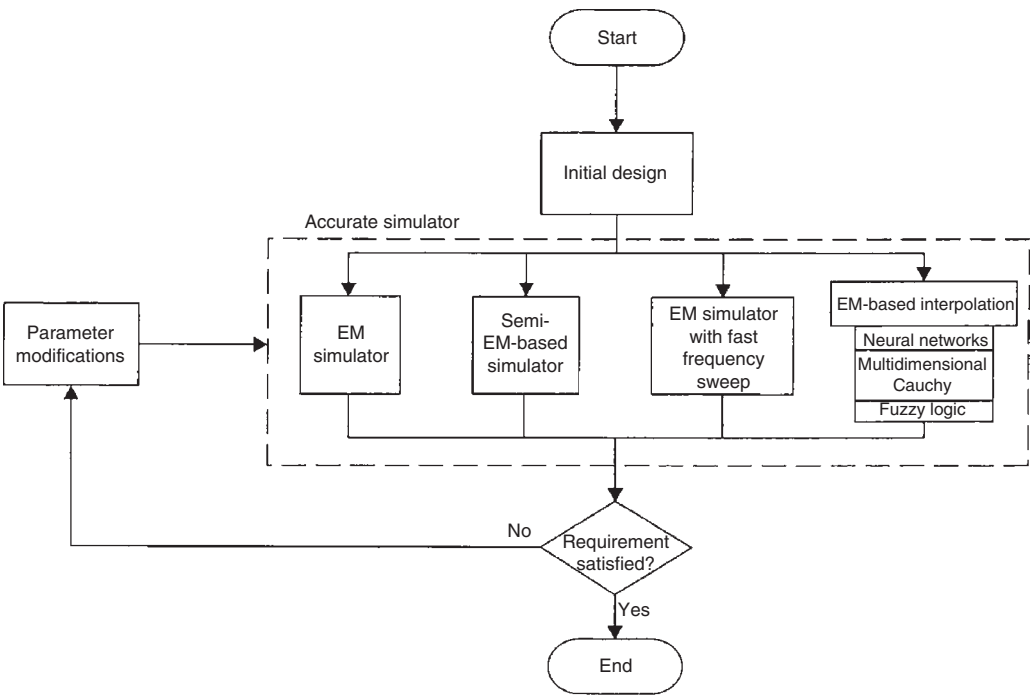


Figure 15.1 EM-based optimization process.

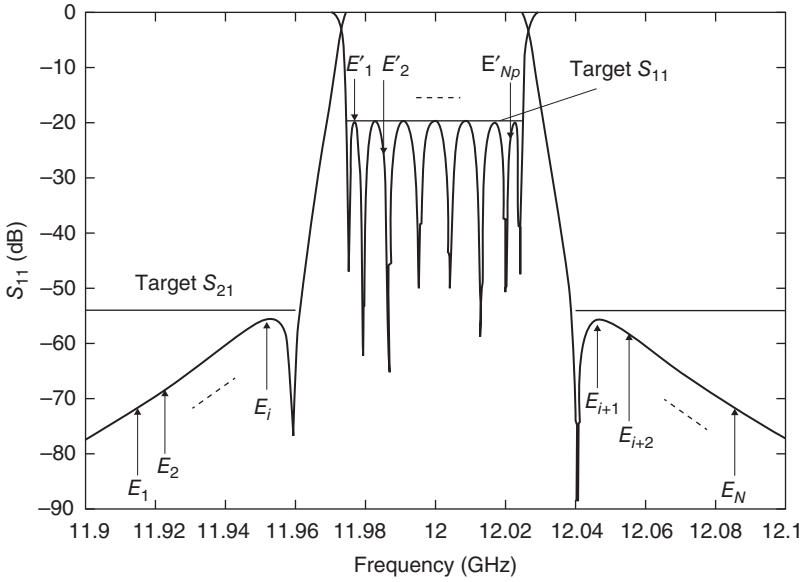


Figure 15.2 Sample frequency points taken over the passband and stopband.

The filter response at these frequency points, along with the given filter specification, is used to generate two sets of error functions within the passband and stopband. The response can be the magnitude of the *S*-parameters, phase, or group delay. In most cases, however, the magnitude of the *S*-parameters is adopted at least in the initial optimization stage. A possible definition for the error functions is

$$\text{Passband}: \quad F_j = \left| \frac{S_{11(j)\text{spec.}}}{S_{11(j)\text{resp.}}} \right|, \quad j = 1,2,\ldots,N_p \tag{15.1}$$

$$\text{Stopband}: \quad F'_j = \left| \frac{S_{21(j)\text{spec.}}}{S_{21(j)\text{resp.}}} \right|, \quad j = 1,2,\ldots,N_s \tag{15.2}$$

where $S_{11(j)\text{resp.}}$ and $S_{21(j)\text{resp.}}$ are the computed *S*-parameters in dB using the accurate simulator at the selected sampling frequency points, while $S_{11(j)\text{spec.}}$ and $S_{21(j)\text{spec.}}$ are the desired *S*-parameters in dB at the same sampling frequency points. For least-squares [6] and minimax optimization [7], the error function may be defined as

$$\text{Least-square:} \quad \text{Minimize } U = \left(\sum_{j=1}^{Np} F_j^2 + \sum_{j=1}^{Ns} F'^2_j \right) \tag{15.3}$$

$$\text{Minimax:} \quad \text{Minimize } U = \text{Max of } (F_1, F_2, \ldots, F_{Np}, \ F'_1, F'_2, \ldots, F'_{Ns}) \tag{15.4}$$

Several other approaches are available for constructing optimization error functions [1, 2].

15.2.1 Optimization Using an EM Simulator

In this case, the EM simulator can be a commercial EM simulator such as HFSS [8], em sonnet [9], Agilent Momentum [10], IE3D [11], CST Microwave Studio [12], Microwave Wizard [13], or any other full-wave simulation analysis software tool. Few commercial EM simulators have built-in optimization tools that allow the designer to specify the parameters to be optimized and offer various ways to build the objective functions. However, at the present time the majority of

these commercial software packages still do not offer complete flexibility in selecting the design parameters and typically require an experienced user to exploit the optimization tools.

The use of optimization with a rigorous EM simulation provides the ultimate design accuracy. However, the approach is very computationally intensive. For example, using today's computer, the direct EM-based optimization of an eight-pole dielectric resonator filter can take weeks to complete. For gradient-type optimization, the typical CPU time required for optimization can be approximated by

$$\text{Total time} = N_{\text{freq}} \times N_{\text{var}} \times N_{\text{ite}} \times I_{\text{drev}} \times T_{\text{sim}} \tag{15.5}$$

where

N_{freq} = number of frequency sampling points needed to construct the objective function ($N_{\text{freq}} = N_p + N_s$)
N_{var} = number of variables involved in the optimization process
N_{ite} = number of optimization iterations
I_{drev} = number of function evaluations for numerical gradient calculation
T_{sim} = time taken by the EM simulator to calculate performance of the complete filter circuit at one frequency point.

For example, let us consider a symmetric six-pole Chebyshev filter structure, where there are seven variables for optimization: $N_{\text{var}} = 7$. The typical values for the other parameters in equation (15.5) are $N_{\text{freq}} = 20$, $N_{\text{ite}} = 20$, $I_{\text{drev}} = 2$ for a second-order gradient. The total CPU time taken to optimize this filter is then given by

$$\text{Total optimization time} = 20 \times 7 \times 20 \times 2 \times T_{\text{sim}} = 5600 \;\; T_{\text{sim}} \tag{15.6}$$

The time T_{sim} is dependent on the filter configuration, the EM simulator, and the speed of the computer workstation. With the assumption of one minute per frequency point (i.e., $T_{\text{sim}} =$ 1 min), the total CPU time to optimize this filter is 5600 min (close to 4 days). It can be seen that although it only takes 1 min per frequency point to simulate the filter performance, it takes close to 4 days of CPU time to optimize its performance to the given requirements. Moreover, in some cases, two optimization runs may be required to fine-tune the filter to the desired performance; thus close to 8 days of CPU time may be required to design a six-pole filter of this type.

As computer speed continues to increases, the computation time T_{sim} is expected to be significantly reduced. Thus, this brute-force straightforward approach may emerge as the simplest preferred approach for designers in the future. However, at the present time, such approach is useful only for low-order filters. For more complicated filter configurations such as high-order dielectric resonator filters, T_{sim} is very high, prohibiting the use of this technique in the design of such filters. Moreover, with the increasing complexity of filter circuit layouts, the required CPU time increases exponentially, making this approach prohibitively expensive. In addition, there is always a need for advanced techniques to reduce the CPU time needed for the design of more complex structures.

15.2.2 Optimization Using Semi-EM-Based Simulator

The objective of using the semi-EM-based approach is to reduce the simulation time T_{sim} given in equation (15.5) without a major sacrifice in accuracy. In the semi-EM-based simulation approach, the overall filter circuit is divided into subcircuits, whose S-parameters are calculated by an EM simulator. The scattering parameters of the overall filter circuit are then computed by using any of the circuit cascading techniques presented in Chapter 5. The reduction in

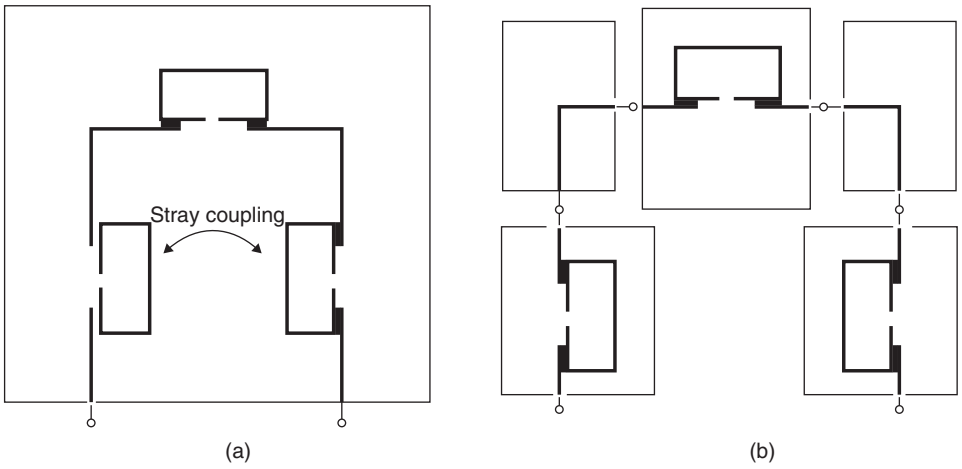


Figure 15.3 (a) The complete filter circuit; (b) the filter circuit divided into subsections cascaded by transmission lines.

T_{sim} stems from the fact that for most EM simulators the CPU time is roughly proportional to the square of the circuit volume, namely, the CPU $\propto$ (volume^2). Thus, in dividing a circuit into N equal subcircuits, the CPU time taken to simulate these N subcircuits is approximately $(N \cdot T_{\text{overall}}/N^2) \approx (T_{\text{overall}}/N)$, where T_{overall} is the time needed to simulate the overall circuit. For example, the filter circuit in Figure 15.3a is divided into small circuits connected by transmission lines, as shown in Figure 15.3b. In this case, the term T_{sim} given in equation (15.5) is the time required to calculate the scattering matrices of the individual three subsections by the EM simulator and the time needed to cascade these subcircuits at one frequency point. This CPU time is considerably less than that required to simulate the overall circuit, which, in turn, reduces the overall optimization time.

However, the improvement in speed is obtained at the price of a less accurate design. The semi-EM simulator approach does not account for the cross-coupling between the nonadjacent resonators. While the problem may not be an issue for waveguide filters, it impacts the accuracy of the scattering parameter results in planar filters, where the cross-resonator coupling between nonadjacent resonators is more pronounced. Figure 15.4 compares the results of the two approaches: EM simulation results of the overall filter circuit and the results obtained from

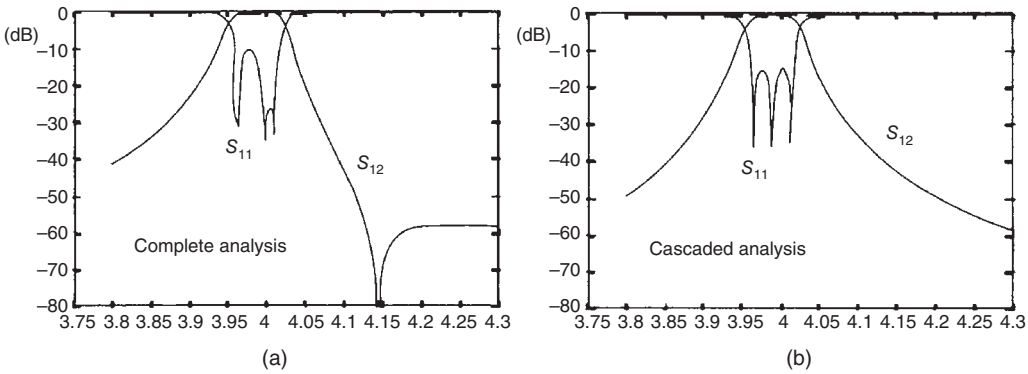


Figure 15.4 (a) The EM simulation results of circuit shown in Figure 15.3a; (b) the simulation results obtained using the semi-EM simulation approach for the circuit shown Figure 15.3b.

using the cascade of EM-based scattering parameters of the cascaded subcircuits. The transmission zero in Figure 15.4a is attributed to the coupling between nonadjacent resonators. It is clear that the cross-coupling effect is not captured by the semi-EM simulator.

15.2.3 Optimization Using an EM Simulator with Adaptive Frequency Sampling

This approach targets the reduction of the number of frequency points N_{freq} used in equation (15.5). The idea here is that instead of simulating the filter at N_{freq} points to build the objective functions, we use the EM simulator for only few frequency points (obviously fewer than N_{freq}). These few points are in turn used to create an interpolation polynomial function to describe the filter response over the entire frequency band. This polynomial function is then used to calculate the filter response at the other frequency points. The adaptive frequency sampling is a feature that is available today in the majority of commercial EM simulators (sometimes known as "fast frequency sweep"). The feature allows the designer to get the results over the whole specified band of simulation, whereas the actual simulations are done at only few frequency points.

Using the Padé approximation [14], the filter scattering parameters of passive microwave circuits are accurately described by a rational polynomial function [15–17] in the form of

$$S(f) = \frac{a_0 + a_1 f + a_2 f^2 + \cdots + a_N f^N}{1 + b_1 f + b_2 f^2 + \cdots + b_M f^M} = \frac{a_0 + \sum_{j=1}^{N} a_j f^j}{1 + \sum_{j=1}^{M} b_j f^j} \tag{15.7}$$

The polynomial given in equation (15.7) is referred to as Cauchy interpolation. The order of the polynomial depends on the sampling simulation points. With the use of K simulation points ($K = 1 + N + M$), a linear system of equations for the polynomial coefficients $a_1, a_2, \ldots, a_N$ and $b_1, b_2, \ldots, b_M$ can be solved by matrix inversion. In order to avoid matrix inversion problems, the Bulirsch–Stoer algorithm [18] was proposed in [17] to calculate the polynomial coefficients of equation (15.7).

To demonstrate the efficiency of the fast frequency sweep, the results of the spline and Cauchy interpolation schemes for a four-pole microstrip filter are compared in Figure 15.5 [17]. The response of the filter is sampled at only a few frequency points. It is noted that the Cauchy interpolation results in almost an exact response.

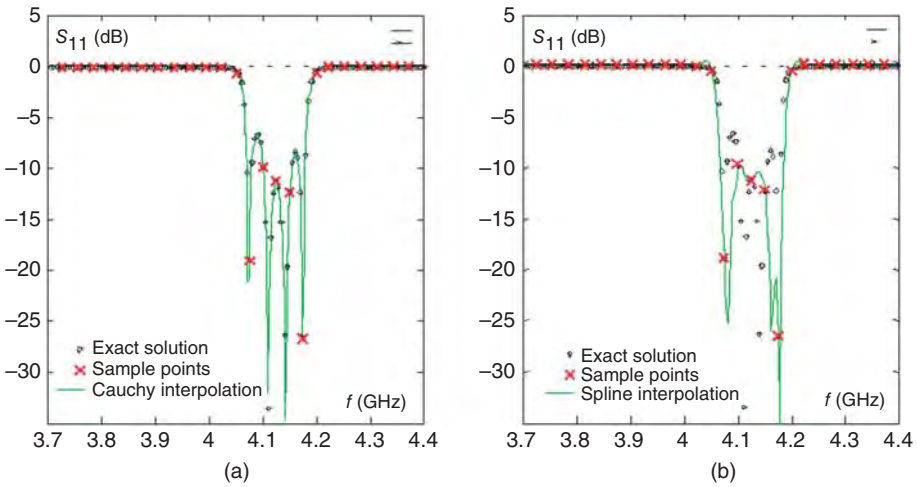


Figure 15.5 Comparison between Cauchy (a) and spline (b) interpolation for four-pole filter.

The number of sampling points can be significantly reduced by the use of adaptive frequency sampling technique [15–17]. In the adaptive approach, a few sample points are first selected uniformly across the band, and then more sample points are selected adaptively until convergence is achieved. The adaptive frequency sampling technique minimizes the number of sample points needed to construct the Cauchy polynomial. An example is given in [17], where only nine sampling frequency points are used to build a polynomial that accurately describes the filter. The first five frequency points are selected uniformly across the band of interest. These five sampling points are used to build two different rational polynomials: polynomial 1 with $N = 2$, $M = 2$ and polynomial 2 with $N = 1$, $M = 3$. These two polynomial approximations are compared over the entire frequency band. Additional samples are selected at the frequency points with the most mismatch. Then new polynomials are calculated and the procedure is repeated until both models agree. The procedure shown in Figure 15.6 illustrates an example where an additional sample frequency point is added at each iteration. It is noted that the two polynomials agree with each other after the adaptive addition of four sample points to the original five sample points.

15.2.4 Optimization Using EM-Based Neural Network Models

This approach targets the reduction of the simulation time T_{sim} in equation (15.5). The EM model of a microwave circuit can be described as a system with an input vector p and an output vector y. The input vector represents the circuit physical dimensions and frequency, whereas the output vector represents the scattering parameters of the circuit. The relationship between p and y is multidimensional and nonlinear. Neural networks [19, 20] can be trained by employing EM simulation results to represent the relationship between p and y vectors. Once properly trained, the neural network becomes the model for the circuit and can be used in place of the computationally intensive EM simulator. Typically, the EM-based neural network design approach usually involves the following steps: (1) selecting the appropriate neural network structure, (2) data generation and network training, and (3) network application.

Step 1: Selection of the Appropriate Neural Network Structure. A typical neural network has two basic sets of components: the processing elements and the interconnections among them [19, 20]. The processing elements are called *neurons* and the connections between the neurons are known as *links*. Every link has a corresponding weight parameter associated with it. Each neuron receives stimulus from other neurons connected to it, possesses the information, and produces an output. The neurons that receive stimuli from outside the network are called *input neurons*; neurons whose outputs are externally used are called *output neurons*. Neurons that receive stimuli from other neurons and whose outputs are stimuli for other neurons in the network are known as "hidden neurons." The multilayer perceptions (MLP) format, shown in Figure 15.7, is a popular neural network structure.

In the MLP neural network, the neurons are grouped into layers in the form of an input layer, an output layer, and hidden layers. The neurons in the hidden layers are activated by the sigmoid function as

$$a_h = \frac{1}{1 + \exp\left(-\sum_{l=1}^{n} a_l w_{lh} - \theta_h\right)} \tag{15.8}$$

where a_h is the activated neuron, w_{ih} are the adjustable weights between the ith and hth neurons, and θ_h is the threshold. A good starting point for the neural network is to use a three-layer network. The universal approximation theorem [20] states that there always exists

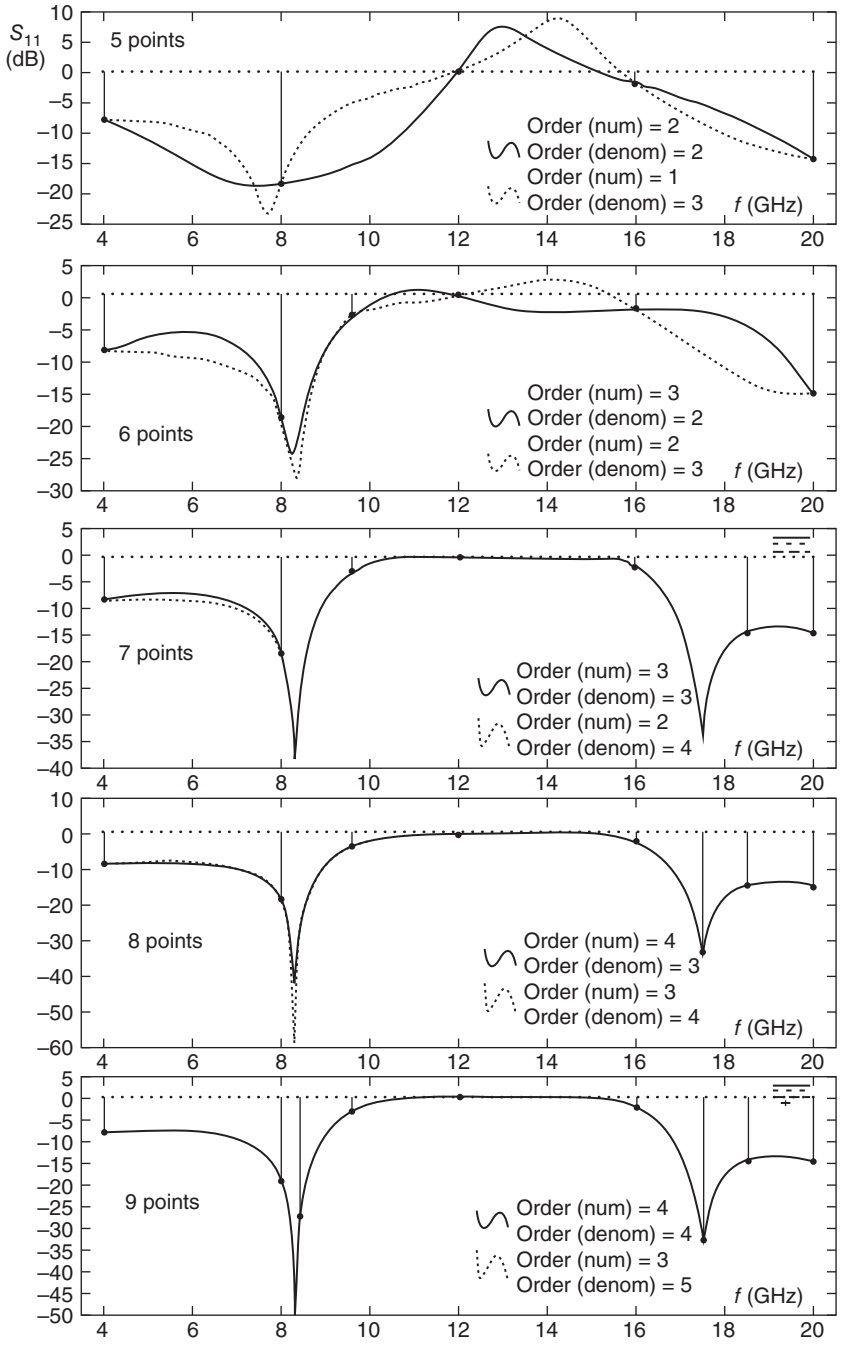


Figure 15.6 Adaptive frequency sampling (num—numerator; denom—denominator).

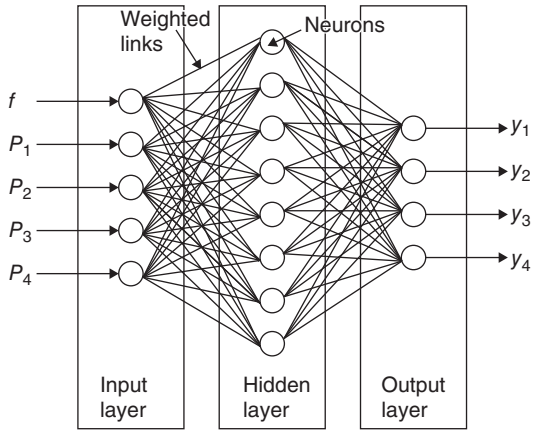


Figure 15.7 A neural network consisting of three layers.

a three-layer MLP neural network that can approximate any arbitrary, nonlinear, continuous, and multidimensional functions. The number of hidden neurons must be carefully determined. A small number of neurons cannot model the system precisely, whereas too many neurons tend to overstrain the network. More details on the selection of the neural network structure can be found in [19, 20].

Step 2: Data Generation and Network Training. The network must be trained by a number of sample input/output vector data pairs (p, y) in order to construct the network. These data are computed by the EM simulator. The generated data are divided into two sets: training data and test data. The training data are used to guide the training process and to determine the weight factors W_{ih} and threshold θ_h. The test data are used to examine the quality of the trained neural network.

During the training process, the neural network automatically adjusts its weights W_{ik} and threshold values θ_h to minimize the error between the predicted output (by the neural network) and the actual output (by the EM simulator). Typically, the neural training algorithm uses gradient-based training techniques such as backpropagation, conjugate gradient, and quasi-Newton techniques [20]. Global optimization methods such as simulated annealing and genetic algorithms can be used for improving quality of neural network training, but this increases the CPU time associated with the training process.

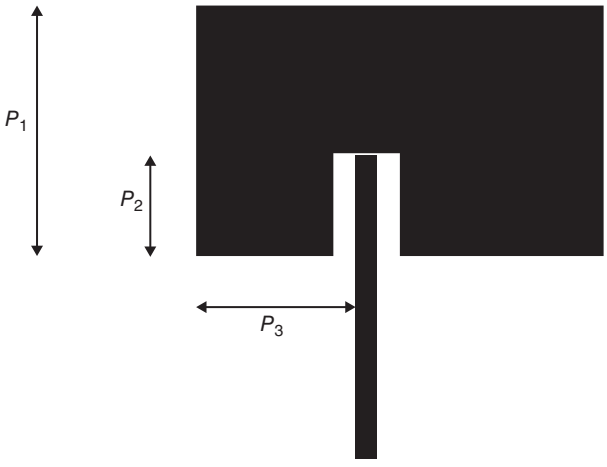


Figure 15.8 An inset-fed microstrip patch resonator.

Step 3: Network Application. After training and testing, the network is ready for use as a substitute for the EM simulator. The network is capable of predicting output y for any given input parameters p in the trained region. The calculation is very fast since only a few basic algebraic operations are required. Once constructed, the neural network can be used also to replace the EM simulator in the optimization process, reducing the design time considerably.

An example [21] illustrating the use of EM-based neural networks in the design of the inset-fed microstrip patch resonator is shown in Figure 15.8. The patch is designed on a substrate of $\varepsilon_r = 2.3$ and a height of $h = 2$ mm. The width of the patch is fixed at 28.7 mm. The objective is to find the dimensions $\vec{p} = (p_1, p_2, p_3)$ such that the microstrip patch resonates at 4 GHz. A total of 500 training data points are obtained by varying the three geometric parameters p_1, p_2, p_3 and frequency over the following ranges: p_1, 17–29 mm; p_2, 0–12 mm; p_3, 0–12 mm (f—3.4–4.4 GHz). The MLP network is constructed with input neurons: frequency and 3 geometric parameters, 10 hidden neurons, and 2 output neurons (magnitude and phase of S_{11}).

The network is tested at two points, and the results are shown in Figure 15.9 in comparison with EM simulation. A good agreement is observed. The neural network is then implemented in the optimization process. The optimized dimensions for the patch to resonate at 4 GHz are $\vec{p} = (23.38,\ 3.5,\ 1.19\,\text{mm})$. The S_{11} simulation results obtained with the EM simulator (exact solution) and the neural network are also compared in Figure 15.9. The optimization process takes only a few seconds. The same optimization process would have taken several hours had the EM simulator been used instead of the neural network model [21].

There are only a few examples reported in literature on the successful use of neural network in the design of low-order (three- or four-pole) microwave filters, where the neural network is used to model the whole filter [21, 22]. For high-order filters, the size of vectors p and y can become too large, requiring a large volume of training/test data to properly train the network. The neural network approach can be applied to model subsections of the filters, which can then be cascaded to form the filter. For example, in the semi-EM-based technique presented

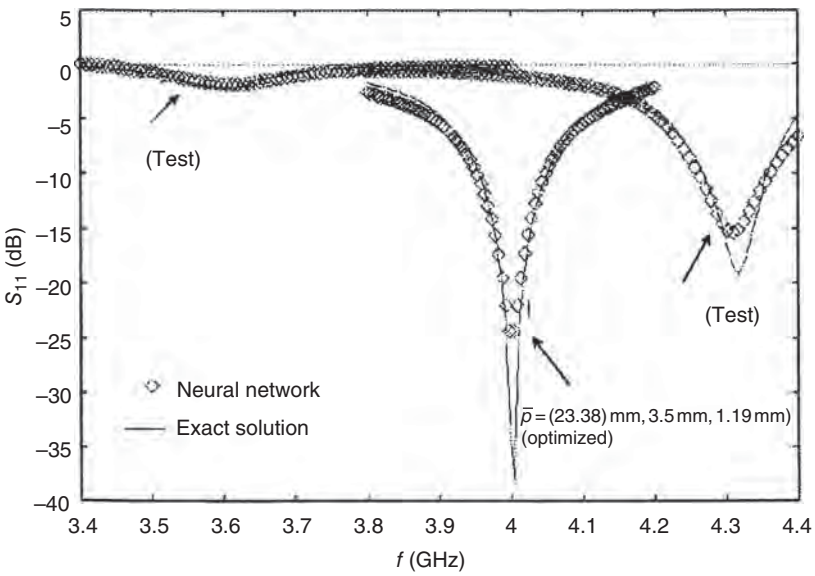


Figure 15.9 Correlation between the S_{11} results obtained by the neural network and EM simulator (exact solution) for the resonator shown Figure 15.8 [21].

in Section 15.2.2, the capacitive subsections can be easily modeled using EM-based neural networks. The overall scattering matrix of the filter can then be obtained by cascading these subsections.

A neural network software package is commercially available [23] that has a built-in interface to several commercial EM simulators. Neural network tools are also available from MATLAB [24]. The user, in this case, needs to perform several EM simulations to collect the data pairs. The performance of EM-based neural networks can be enhanced by incorporating available knowledge about the circuit into the training process. Several techniques known as knowledge-based neural networks have been reported [25, 26]. More details on EM-based neural network tools for microwave circuit design are given in [19].

15.2.5 Optimization Using EM-Based Multidimensional Cauchy Technique

The multidimensional Cauchy technique is similar to neural networks in that it requires the use of the EM simulator to generate data pairs of inputs (physical parameters and frequency) and output (filter circuit scattering parameters) to build the model. These data pairs are then used to generate a multidimensional Cauchy interpolation that can replace the EM simulator for the analysis and optimization of this particular filter circuit.

The multidimensional Cauchy technique enables the interpolation of the circuit performance with respect to frequency and dimensions. It builds on the one-dimensional (1D) Cauchy interpolation with respect to frequency, described in Section 15.2.3. The multidimensional Cauchy technique, proposed in [17], uses a recursive algorithm, which solves the multidimensional interpolation by using the 1D Cauchy interpolation. The multidimensional technique also allows the application of adaptive sampling, which reduces the required number of sampling data pairs [17]. The technique is applied to the design of the three-pole filter circuit shown in Figure 15.10. Because of symmetry, the filter has four variables: $\vec{p} = (p_1, p_2, p_3, p_4)$. EM simulation training data is obtained by varying these three variables and frequency. The results obtained are shown in Figure 15.11. Excellent results are obtained with only 625 training data points [17].

The multidimensional Cauchy technique is useful to accurately model small circuits, where the number of input variables is relatively low (five or less). As the case for neural networks, the filter circuit can be divided into small subsections. Then, each subsection is represented by a multidimensional Cauchy model to replace the EM simulator. The overall performance of the filter is obtained by cascading these subsections.

15.2.6 Optimization Using EM-Based Fuzzy Logic

The feasibility of using fuzzy logic systems (FLSs) in diagnosis and tuning of microwave filters was demonstrated for the first time in [27] and is covered in detail in Chapter 19. The FLS is basically a function approximator. It has been shown in [28] that an FLS can approximate any real continuous nonlinear function to arbitrary degree of accuracy. With the use of data generated from an EM simulator, an FLS model can be built to approximate the EM simulator—similar to what can be achieved by neural networks. The FLS model can then replace the accurate simulator in the optimization process shown in Figure 15.1.

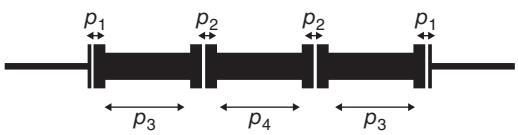


Figure 15.10 Schematic of a three-pole microstrip filter designed by the multidimensional Cauchy technique [17].

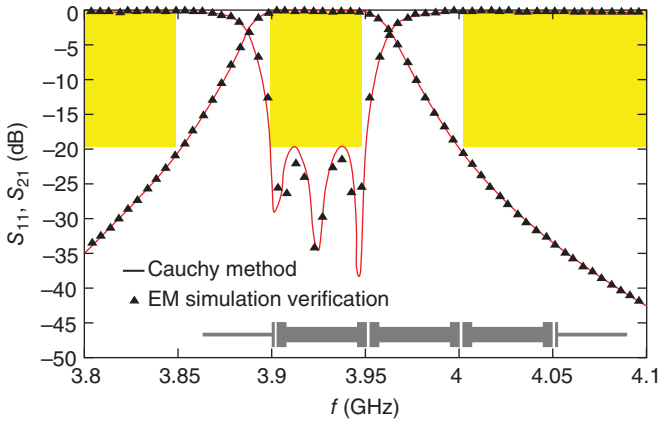


Figure 15.11 Simulation results for the three-pole microstrip filter shown in Figure 15.10 [17].

The function approximation problem $\vec{y} = f(\vec{x})$ of building a neural network model or a fuzzy logic model, where the input $\vec{x}$ represents the filter physical dimensions and frequency while the output $\vec{y}$ represents the filter scattering parameters, is known as the "forward problem." In [29] the FLS technique was used to deal with the "reverse problem" $\vec{x} = \phi(\vec{y})$, where the data points from the EM simulator were used to build an FLS system that predicts the filter physical dimensions for given specifications or given scattering parameters. This, in turn, eliminates the need to use optimization. It is a direct approach for an approximate design of low-order microwave filters. More details are given in [29, 30].

15.3 EM-Based Advanced Design Techniques

Several approaches have been reported with the goal of reducing the CPU time of the EM-based design process. The majority of these approaches employ two models: a fine model and a coarse model. The fine model is an accurate but computationally intensive, whereas the coarse model is fast but less accurate. An EM simulator can serve as the fine model, while the coarse model can be a circuit model. The idea in these approaches is to make use of the accuracy of the fine model and the speed of the coarse model to devise a solution that is fast and accurate. These approaches also require the use of an optimization tool. However, in this case the optimization is done using only the coarse model. The fine model is used only a few times during the design process.

Thus the fine model is combined with the coarse model into a hybrid model that is effectively both fast and accurate. While the bulk of computation is done using the fast coarse model, yet such design approaches successfully result in an accurate design. In this chapter, we focus on two techniques: the space mapping (SM) technique and the calibrated coarse model (CCM) technique.

15.3.1 Space Mapping Techniques

The SM technique has been successfully employed in the design of microwave filters and other engineering problems [31–35]. The SM technique establishes a mathematical link (mapping) between the spaces of the design parameters of the two models: the accurate and time-intensive

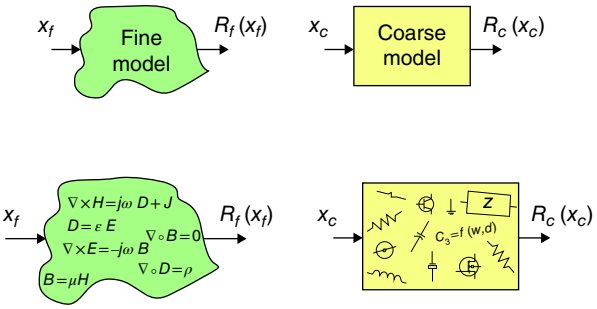


Figure 15.12 Representation of the fine and coarse models. (*Source*: Bandler *et al*. 2004 [34]. Reproduced with the permission of IEEE.)

fine model and the less accurate but fast coarse model. The aim is to avoid the computationally expensive calculations encountered in optimizing the filter structure using the time-intensive fine model. SM must use a coarse model that is capable of simulating the filter structure for a wide range of parameters. The coarse model is not updated or changed during the SM process. The SM technique involves two steps: (1) optimizing the design parameters of the coarse model to satisfy the original design specifications and (2) establishing a mapping between parameters spaces of the two models. The space-mapped design is then taken as the mapped image of the optimal coarse model.

We refer to the coarse model parameters as x_c and the fine model parameters as x_f. The optimal coarse model design is denoted as x_c^*. We also denote the response of the coarse model as $R_c(x_c)$ and that of the fine model as $R_f(x_f)$, as shown in Figure 15.12. The mapping between the fine model space and the coarse model space is written as $x_c = P(x_f)$ such that the difference $|R_f(x_f) - R_c(x_c)|$ is minimized. Usually the mapping P is local over a region of space parameters. It is calculated in an iterative way. If we denote $P^{(j)}$ as the mapping calculated after jth iteration, the corresponding fine model design is given by

$$x_f^{(j+1)} = (P^{(j)})^{-1}(X_c^*)$$

If $R_f(x_f^{(j+1)})$ satisfies the specification within certain acceptable accuracy criteria, the solution $x_f^{(j+1)}$ is accepted as the space-mapped design for the coarse solution x_c^*; otherwise, the mapping is updated and a new iteration for P is evaluated.

Bandler *et al*. introduced the first SM approach in 1994 [31]. In this original SM approach, a linear mapping is assumed between the two parameter spaces of the two models. The mapping is evaluated by a least-squares solution of the linear equations, resulting from the corresponding data points in the two spaces. This mapping is simple but may not converge to a proper solution if a significant misalignment exists between the response of the coarse model and that of the fine model. The misalignment is typically due to the use of very rough coarse model, whose response is far away from the accurate response of the fine model.

A modified version of the original SM technique was also proposed by Bandler *et al*. [32, 33] and is known as *aggressive space mapping* (ASM). ASM does not assume that the mapping is necessarily linear and usually converges to the proper design in a few iterations. Several other SM-based techniques have been also proposed to deal with the nonuniqueness problem of parameter extraction and to minimize the number of iterations. The list includes trust region aggressive space mapping (TRASM), hybrid ASM, and neural SM. A comprehensive overall review on SM techniques is given in [34].

In this chapter, we discuss in details the original SM approach and the ASM approach. We focus particularly on the ASM approach, since it is a more reliable approach. Two examples are given to demonstrate the application of the ASM concept including an example on the design of a six-pole microstrip filter.

15.3.1.1 The Original Space Mapping Approach

The steps in this approach are summarized as follows:

Step 1. Optimize the coarse model to find optimal solution x_c^*. In this case, x_c^* is the optimum solution based on the coarse model; that is, $R_c(x_c^*)$ is the desired performance.

Step 2. Set $x_f^{(1)} = x_c^*$, and select an initial set of points $x_f^{(2)}, x_f^{(3)}, \ldots, x_f^{(j)}$ that are arbitrarily selected in the vicinity of $x_f^{(1)}$. Note that $x_f^{(1)}, x_f^{(2)}, \ldots, x_f^{(j)}$ are vectors representing the filter physical parameters. The set $S_f^{(j)}$ is defined as

$$S_f^{(j)} = \left\{ x_f^{(1)}, x_f^{(2)}, \ldots, x_f^{(m_j)} \right\} \tag{15.9}$$

Step 3. Calculate the fine model response for every point in the set $S_f^{(j)}$.

Step 4. Optimize the coarse model to match the fine model responses calculated in step 3. Use a parameter extraction technique such as optimization for $\min||R_c(x_c^{(i)}) \simeq R_f(x_f^{(i)})||$ for $i = 1,2,\ldots,j$ to extract $x_c^{(1)}, x_c^{(2)}, \ldots, x_c^{(j)}$, which are vectors representing the physical parameters in the coarse model space. A set of coarse model points $S_c^{(j)}$ can be written as

$$S_c^{(j)} = \left\{ x_c^{(1)}, x_c^{(2)}, \ldots, x_c^{(m_j)} \right\} \tag{15.10}$$

Step 5. The two sets of points $S_c^{(j)}$ and $S_f^{(j)}$ are then used to calculate the mapping $x_c = P^j(x_f)$. Bandler *et al*. [31] assumed that the mapping between the two spaces is linear. Thus

$$x_c = P^{(j)}(x_f) = B^{(j)} x_f + C^{(j)} \tag{15.11}$$

If we define the $A^{(j)}$ as $\left[C^{(j)} \;\; B^{(j)}\right]$, it follows that equation (15.11) can be written as

$$\begin{bmatrix} x_c^{(1)} & x_c^{(2)} & \cdots & x_c^{(m_j)} \end{bmatrix} = A^{(j)} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_f^{(1)} & x_f^{(2)} & \cdots & x_f^{(m_j)} \end{bmatrix} \tag{15.12}$$

A least-squares solution for $A^{(j)}$ is given [35] by

$$A^{(j)T} = (D^T D)^{-1} D^T Q \tag{15.13}$$

where

$$D = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_f^{(1)} & x_f^{(2)} & \cdots & x_f^{(m_j)} \end{bmatrix}^T \tag{15.14}$$

and

$$Q = \begin{bmatrix} x_c^{(1)} & x_c^{(2)} & \cdots & x_c^{(m_j)} \end{bmatrix}^T \tag{15.15}$$

Once $A^{(j)}$ is obtained, the space-mapped design is

$$x_f^{(j+1)} = P^{(j)^{-1}}(x_c^*) = B^{(j)^{-1}}(x_c^* - C^{(j)}) \tag{15.16}$$

Step 6. Take $x_f^{(j+1)}$ as an approximation to the optimal fine point design x_f, if the condition $||R_f(x_f^{(j+1)}) - R_c(x_c^*)|| \leq \varepsilon$ is satisfied; otherwise, the set $S_f^{(j)}$ is augmented to $S_f^{(j+1)}$ by adding

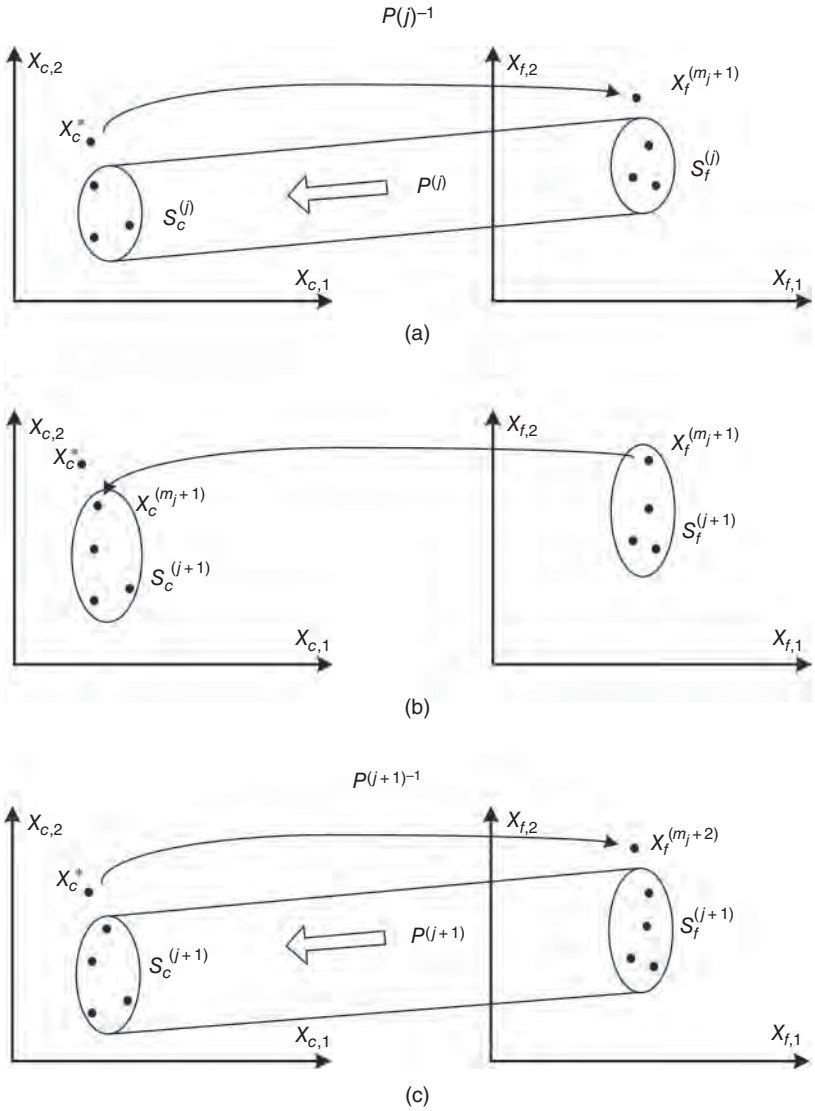


Figure 15.13 The original space mapping technique. (*Source*: Bakr *et al*. 2000 [35]. Reproduced with the permission of Springer.)

point $x_f^{(j+1)}$ and the set $S_c^{(j)}$ is augmented to $S_c^{(j+1)}$ by adding the corresponding point $x_c^{(j+1)}$. A new mapping is then found, and the fine model response of the mapped point is checked. The process is repeated until the fine model response equals the desired response $R_c(x_c^*)$. Figure 15.13 illustrates the original SM approach.

This approach assumes a linear mapping between the two models, and this may not be accurate in many problems. Therefore this original SM approach is recommended only if the coarse model is very close to the fine model such that a linear mapping can be assumed. The ASM technique is preferred, since it is not highly sensitive to the misalignment problems that might exist between the coarse model and the fine model.

15.3.1.2 Aggressive Space Mapping (ASM) Approach

In the ASM approach, the solution x_f^* is obtained by solving the following nonlinear equation:

$$P(x_f) - x_c^* = 0 \tag{15.17}$$

By using the quasi-Newton method, the solution is obtained in an iterative manner. If $x_f^{(j)}$ is the jth iterate, the next iteration $x_f^{(j+1)}$ is written as

$$x_f^{(j+1)} = x_f^{(j)} + h^{(j)} \tag{15.18}$$

where $h^{(j)}$ is the solution of $B^{(j)}h^{(j)} = -f^{(j)}$, where

$$f^{(j)} = P(x_f^{(j)}) - x_c^* = x_c^{(j)} - x_c^* \tag{15.19}$$

$$P(x_f^{(j)}) = x_c^{(j)} \quad \text{such that} \quad \|R_c(x_c^{(j)}) - R_f(x_f^{(j)})\| \leq \varepsilon \tag{15.20}$$

The vector $x_c^{(j)}$ is obtained by using parameter extraction (optimization) of the coarse model such that the coarse model response equals the fine model response for $x_f^{(j)}$. $B^{(j)}$ is an approximation to the Jacobian of f with respect to x_f at $x_f^{(j)}$. Using Broyden's rank 1 formula [35], $B^{(j+1)}$ can be written as

$$B^{(j+1)} = B^{(j)} + \frac{f^{(j+1)}h^{(j)^T}}{h^{(j)^T}h^{(j)}} \tag{15.21}$$

The initial value for $B^{(1)}$ is taken as the identity matrix: $B^{(1)} = U$. The steps of the ASM approach are summarized as follows:

Step 1. Extract x_c^*, such that $R_c(x_c^*) \approx$ desired performance. Initialize $j = 1$, and set $x_f^{(1)} = x_c^*, B^{(1)} = U$.
Step 2. Evaluate $R_f(x_f^{(1)})$.
Step 3. Extract $x_c^{(1)}$ such that $R_c(x_c^{(1)}) \approx R_f(x_f^{(1)})$.
Step 4. Evaluate $f^{(1)} = x_c^{(1)} - x_c^*$. Stop if $||f^{(1)}|| \leq \eta$.
Step 5. Solve $B^{(j)}h^{(j)} = -f^{(j)}$ for $h^{(j)}$.
Step 6. Set $x_f^{(j+1)} = x_f^{(j)} + h^{(j)}$.
Step 7. Evaluate $R_f(x_f^{(j+1)})$.
Step 8. Extract $x_c^{(j+1)}$ such that $R_c(x_c^{(j+1)}) \approx R_f(x_f^{(j+1)})$.
Step 9. Evaluate $f^{(j+1)} = x_c^{(j+1)} - x_c^*$. Stop if $\|f^{(j+1)}\| \leq \eta$.
Step 10. Update $B^{(j+1)} = B^{(j)} + \dfrac{f^{(j+1)}h^{(j)^T}}{h^{(j)^T}h^{(j)}}$.
Step 11. Set $j = j + 1$; go to step 4.

The steps are repeated until a solution is found. To illustrate the ASM technique, we will consider the application of the concept to two examples: two Rosenbrock functions and the design of a six-pole microstrip filter.

Example 15.1 ***An Illustrative Example Demonstrating the Concept of ASM (Two Rosenbrock Functions)***

In this example, we apply the ASM technique to find the solution to the Rosenbrock function [36]. The example allows the reader to follow the design process step by step and to duplicate

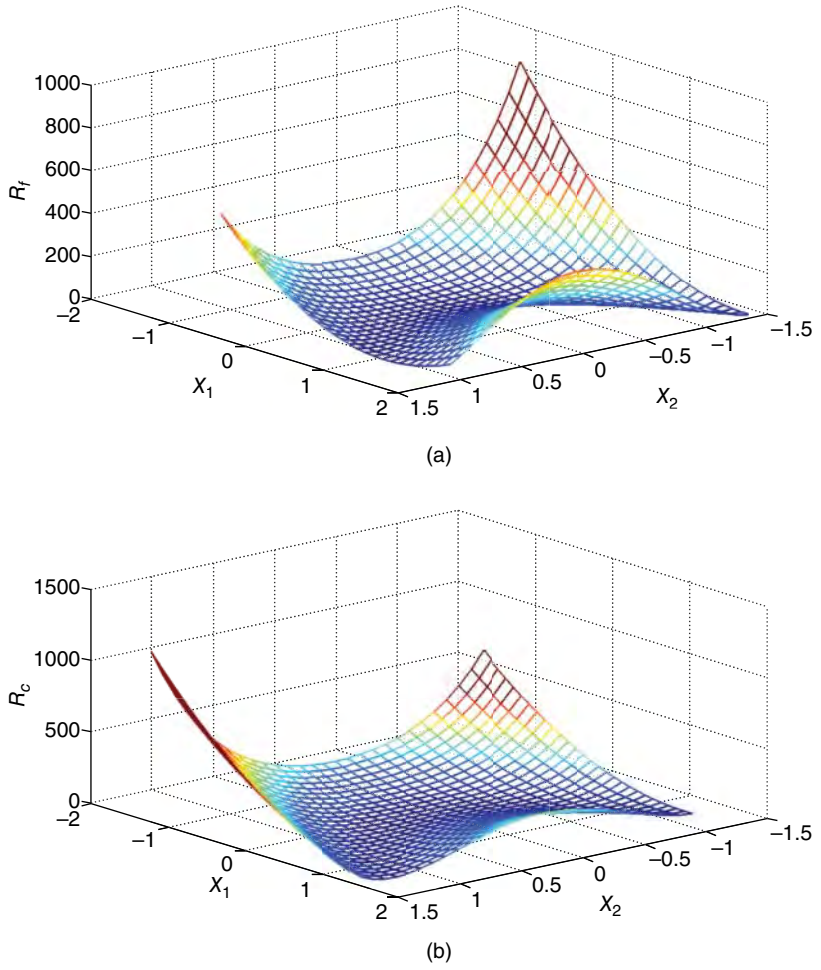


Figure 15.14 The Rosenbrock functions: (a) the fine model $R_f(x_1, x_2)$; (b) the coarse model $R_c(x_1, x_2)$.

the results by using a calculator. Let us consider that we have the two Rosenbrock functions, $R_c(x)$ and $R_f(x)$, given as follows.

The coarse model is expressed as $R_c(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$, where

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The fine model is expressed as $R_f(x) = 100(u_2 - u_1^2)^2 + (1 - u_1)^2$, where

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 1.1 & -0.2 \\ 0.2 & 0.9 \end{bmatrix} x + \begin{bmatrix} -0.3 \\ 0.3 \end{bmatrix}$$

The problem is to find the values of x_1 and x_2 that is the solution of $R_f(x) = 0$. The two functions are shown in Figure 15.14.

Solution Using the ASM Technique

1) Extract x_c^* such that $R_c^* = R_c(x_c^*) = 0$, $x_c^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and set $x_f^{(1)} = x_c^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $B^{(1)} = U, j = 1$.

2) $R_f(x_f^{(1)}) = 108.32$.

3) Extract $x_c^{(1)}$ for $R_c(x_c^{(1)}) = R_f(x_f^{(1)}) = 108.32$, $x_c^{(1)} = \begin{bmatrix} 0.6 \\ 1.4 \end{bmatrix}$.

4) $f^{(1)} = \begin{bmatrix} 0.6 \\ 1.4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.4 \\ 0.4 \end{bmatrix}$.

5) Since $B^{(1)} = 1$, it follows that $h^{(1)} = \begin{bmatrix} 0.4 \\ -0.4 \end{bmatrix}$.

6) Set $x_f^{(2)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.4 \\ -0.4 \end{bmatrix} = \begin{bmatrix} 1.4 \\ 0.6 \end{bmatrix}$.

7) $R_f(x_f^{(2)}) = 1.8207$.

8) Extract $x_c^{(2)}$ such that $R_c(x_c^{(2)}) = R_f(x_f^{(2)}) = 1.8207, \rightarrow x_c^{(2)} = \begin{bmatrix} 1.12 \\ 1.12 \end{bmatrix}$.

9) $f^{(2)} = \begin{bmatrix} 1.12 \\ 1.12 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.12 \\ 0.12 \end{bmatrix}$.

10) $B^{(2)} = B^{(1)} + \dfrac{f^{(2)} h^{(1)^T}}{h^{(1)^T} h^{(1)}} = \begin{bmatrix} 1.15 & -0.15 \\ 0.15 & 0.85 \end{bmatrix}$.

11) $h^{(2)} = -B^{(2)^{-1}} f^{(2)} = \begin{bmatrix} -0.12 \\ -0.12 \end{bmatrix}$.

12) Set $x_f^{(3)} = \begin{bmatrix} 1.4 \\ 0.6 \end{bmatrix} + \begin{bmatrix} -0.12 \\ -0.12 \end{bmatrix} = \begin{bmatrix} 1.28 \\ 0.48 \end{bmatrix}$.

13) $R_f(x_f^{(3)}) = 0.1308$.

14) Extract $x_c^{(3)}$ such that $R_c(x_c^{(3)}) = R_f(x_f^{(3)}), \rightarrow x_c^{(3)} = \begin{bmatrix} 1.012 \\ 0.988 \end{bmatrix}$.

15) $f^{(3)} = \begin{bmatrix} 1.012 \\ 0.988 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.012 \\ -0.012 \end{bmatrix}$.

16) $B^{(3)} = B^{(2)} + \dfrac{f^{(3)} h^{(2)T}}{h^{(2)^T} h^{(2)}} = \begin{bmatrix} 1.1 & -0.2 \\ 0.2 & 0.9 \end{bmatrix}$.

17) $h^{(3)} = -B^{(3)^{-1}} f^{(3)} = \begin{bmatrix} -0.0082 \\ -0.0151 \end{bmatrix}$.

18) $x_f^{(4)} = \begin{bmatrix} 1.28 \\ 0.48 \end{bmatrix} + \begin{bmatrix} -0.0082 \\ 0.0151 \end{bmatrix} = \begin{bmatrix} 1.2718 \\ 0.4951 \end{bmatrix}$.

19) $R_f(x_f^{(4)}) = 9.2 \times 10^{-8}$, then $x_f = x_f^{(4)}$, and we can end the algorithm.

The solution to $R_f(x) = 0$ then is $x_1 = 1.2718$, and $x_2 = 0.4951$.

Example 15.2 *Design of a Six-Pole Filter Using Aggressive Space Mapping*

Let us examine the design of a six-pole microstrip filter with the layout shown in Figure 15.15. The filter has a center frequency of 2 GHz, a bandwidth of 2%, and a return loss of 20 dB. The

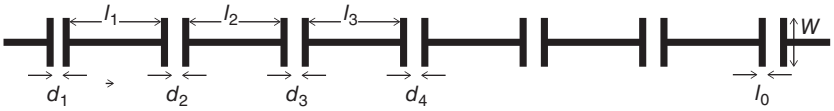


Figure 15.15 The six-pole filter layout.

filter is realized on an alumina substrate with a height $h = 0.025$ in. (25 mil) and a dielectric constant of $\varepsilon_r = 10.2$. The structure consists of six resonators and seven coupling gaps. All resonators are assumed to have a 50 Ω impedance. The problem is to find the circuit physical dimensions: w, l_0 d_1, d_2, d_3, d_4, l_1, l_2, and l_3.

Solution Using the ASM Technique

In this case, the fine model is the EM simulator Agilent Momentum [10], whereas the coarse model is the Agilent ADS circuit simulator ADS [37]. The circuit schematic is shown in Figure 15.16 where the capacitive coupling is presented by two-step discontinuities, two transmission lines and a gap capacitor. The coarse model is fast since ADS uses closed-form empirical models for these discontinuities. We follow the steps of the ASM technique and give the details for the six-pole filter design. The coarse model parameters x_c^* are obtained by optimizing the schematic circuit shown in Figure 15.16. We include all the nine parameters shown in Figure 15.15 in the optimization process. The optimal x_c^* is given in Table 15.1, and Figure 15.17 shows the corresponding coarse model optimal response.

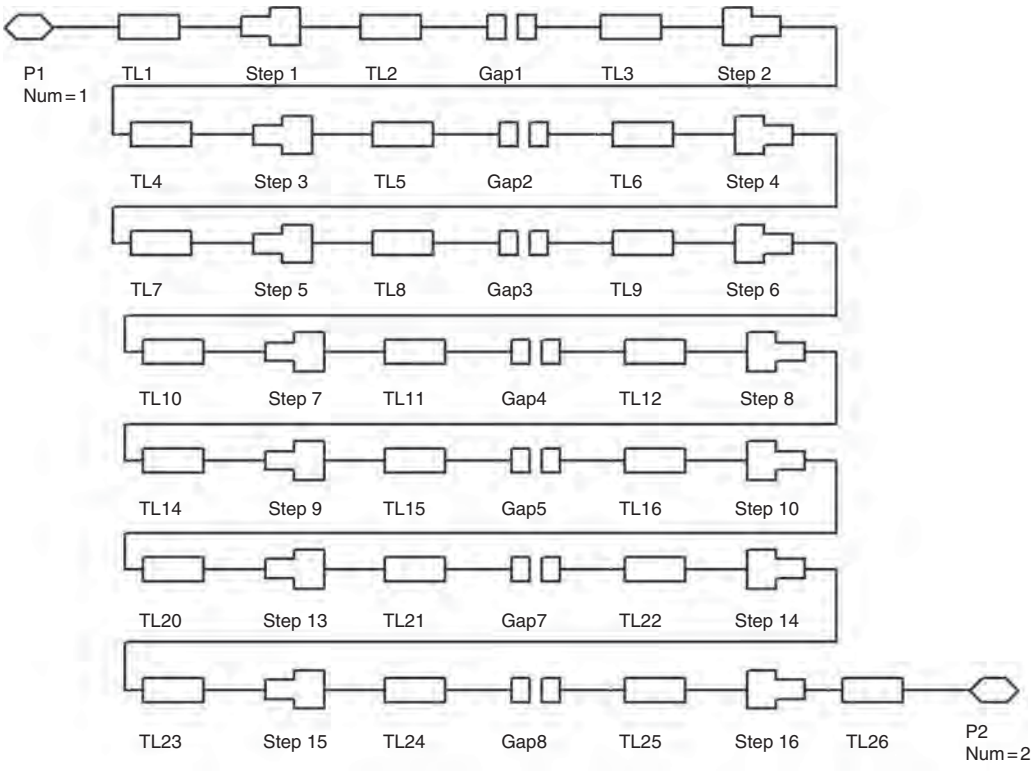


Figure 15.16 The coarse model of the microstrip filter shown in Figure 15.15.

Table 15.1 The design dimensions of the six-pole filter circuit shown in Figure 15.16 after ADS coarse model optimization.

d_1 (mil)	d_2 (mil)	d_3 (mil)	d_4 (mil)	l_0 (mil)	l_1 (mm)	l_2 (mm)	l_3 (mm)	w (mil)
1.01337	27.9174	32.1179	32.7377	24.9124	17.6166	18.0528	18.061	225.02

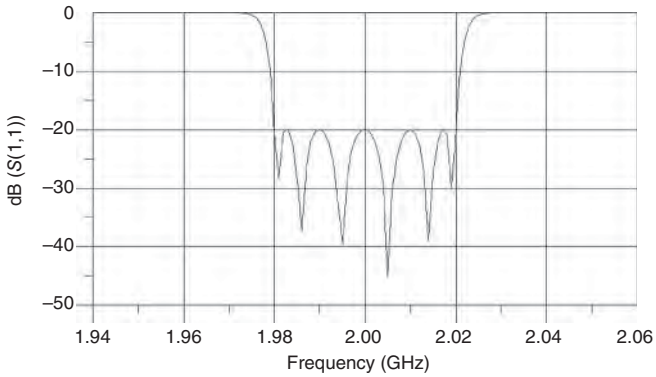


Figure 15.17 The optimized filter response using the coarse model.

The application of the ASM technique iteratively leads to a set of $x_c^{(j)}$, $x_f^{(j)}$, $f^{(j)}$, $h^{(j)}$. Table 15.2 gives the numerical values for these parameters. The ASM converges in eight iterations. Figure 15.18 shows the filter response for the eight iterations.

15.3.2 Calibrated Coarse Model (CCM) Techniques

The CCM technique for filter design was introduced in [38]. A more general version of this technique that makes use of the filter coupling matrix model is presented [39–41]. The CCM technique also employs two models: a coarse model and a fine model. The CCM technique, however, allows the designer to add physical insights into the coarse model. This helps to significantly reduce the required number of EM simulations and gives the designer the flexibility to dynamically change the coarse model to maintain a reasonable alignment with the fine model. For example, in the case of filters, where coupling between nonadjacent resonators usually contributes to transmission zeros, the conventional SM technique may not converge, if the coarse model does not include such effects. On the other hand, in the CCM technique, we can begin with the same coarse model. However, it is modified from one iteration to the next to obtain a proper alignment with the fine model.

The most fundamental difference between the conventional SM technique and the CCM technique is that the coarse model in the SM technique is static; the same coarse model is used for all the iterations. The coarse model in the CCM technique is dynamic in the sense that a different coarse model is used in each iteration.

In fact, the CCM technique is similar to the calibration concept employed in microwave testing. In order to test a device, the measurement setup needs to be calibrated to compensate for losses, phase variations, and mismatch. Typically, a set of adjustments are stored and used to evaluate the performance of the device under test. In the CCM technique, the fine model is used to extract the "standards" needed for calibration.

The circuit, simulated by the coarse model, can be "calibrated" locally so that the coarse model simulations are the same as the fine model simulations. In other words, the CCM technique allows dynamic adjustments to be added to the coarse model to improve its accuracy, so that its solution matches that of the fine model. Since the calibration is done in a local sense, fine readjustments may be needed to the corrected coarse model after each optimization process. However, the end result is that the CCM technique requires only a few EM simulations by the fine model.

Table 15.2 Fine and coarse model variables obtained at eight iterations.

Aggressive SM iterations		d_1 (mil)	d_2 (mil)	d_3 (mil)	d_4 (mil)	l_0 (mil)	l_1 (mm)	l_2 (mm)	l_3 (mm)	w (mil)
Iteration 1	$x_f^{(1)}$	1.01337	27.9174	32.1179	32.7377	24.9124	17.6166	18.0528	18.061	225.02
	$x_c^{(1)}$	0.954706	26.8626	29.9675	30.5151	16.8989	18.3516	18.4402	18.4455	267.01
	$f^{(1)}$	−0.0587	−1.0548	−2.1504	−2.2226	−8.0135	0.7350	0.3874	0.3845	41.9900
	$h^{(1)}$	−0.0207	0.0282	0.0014	−0.0258	−0.3206	0.0055	0.0009	0.0006	1.2830
Iteration 2	$x_f^{(2)}$	1.0720	28.9722	34.2683	34.9603	32.9259	16.8816	17.6654	17.6767	183.0280
	$x_c^{(2)}$	1.23176	28.6173	32.9179	33.5163	20.9102	17.0009	17.5378	17.5515	246.608
	$f^{(2)}$	0.2184	0.6999	0.8000	0.7786	−4.0022	−0.6157	−0.5150	−0.5095	21.5880
	$h^{(2)}$	−0.0124	−0.0399	−0.0456	−0.0444	0.2281	0.0351	0.0293	0.0290	−1.2302
Iteration 3	$x_f^{(3)}$	0.6285	27.5506	32.6434	33.3789	41.0547	18.1321	18.7114	18.7117	139.1787
	$x_c^{(3)}$	1.04541	27.7568	30.6978	30.4684	21.5204	18.3551	18.7148	18.6829	212.054
	$f^{(3)}$	0.0320	−0.1606	−1.4201	−2.2693	−3.3920	0.7385	0.6620	0.6219	−12.9660
	$h^{(3)}$	−0.0732	−0.0990	0.5234	0.9662	2.7851	−0.2204	−0.2071	−0.1879	1.0712
Iteration 4	$x_f^{(4)}$	0.6868	27.9074	33.8485	35.1238	41.7942	17.4332	18.0983	18.1265	155.3479
	$x_c^{(4)}$	0.8533	28.1703	32.5873	33.5311	28.4139	17.1644	17.5678	17.5959	217.48
	$f^{(4)}$	−0.1601	0.2529	0.4694	0.7934	3.5015	−0.4522	−0.4850	−0.4651	−7.5400
	$h^{(4)}$	0.0497	−0.2537	−0.4371	−0.6397	−1.3282	0.3691	0.3666	0.3536	−0.1173
Iteration 5	$x_f^{(5)}$	1.2047	27.6410	32.4441	32.4860	29.4390	18.5305	19.2887	19.2519	178.3964
	$x_c^{(5)}$	1.44391	27.6133	31.3112	31.3195	20.1401	18.2778	18.7656	18.7294	239.325
	$f^{(5)}$	0.4305	−0.3041	−0.8067	−1.4182	−4.7723	0.6612	0.7128	0.6684	14.3050
	$h^{(5)}$	−0.1009	0.1580	0.3879	0.5922	0.6341	−0.2699	−0.2636	−0.2482	−0.01757
Iteration 6	$x_f^{(6)}$	0.8174	27.7587	33.3458	34.2027	36.9616	17.8706	18.5708	18.5795	163.1318
	$x_c^{(6)}$	0.9733	28.0252	32.2081	32.8322	25.1093	17.5888	18.0159	18.0251	225.869
	$f^{(6)}$	−0.0401	0.1078	0.0902	0.0945	0.1969	−0.0278	−0.0369	−0.0359	0.8490
	$h^{(6)}$	0.0358	−0.0689	0.0084	0.0492	−0.0678	−0.0380	−0.0244	−0.0227	−0.0593
Iteration 7	$x_f^{(7)}$	0.8584	27.6370	33.2676	34.1409	37.0008	17.8909	18.5998	18.6080	162.0220
	$x_c^{(7)}$	1.03403	27.8892	32.1165	32.7635	25.233	17.6111	18.0519	18.0604	223.737
	$f^{(7)}$	0.0207	−0.0282	−0.0014	0.0258	0.3206	−0.0055	−0.0009	−0.0006	−1.2830
	$h^{(7)}$	−0.0428	0.0286	−0.0038	−0.0201	−0.0686	0.0016	−0.0091	−0.0085	0.0105
Iteration 8	$x_f^{(8)}$	0.8469	27.6602	33.2659	34.1182	36.7479	17.8937	18.6005	18.6087	162.6640
	$x_c^{(8)}$	—	—	—	—	—	—	—	—	—

The steps of the CCM technique are summarized as follows:

Step 0. Extract x_c^*, such that $R_c(x_c^*) \approx$ desired performance. Set $x_f^{(1)} = x_c^{(1)} = x_c^*$.

Step 1. Evaluate $R_f(x_f^{(1)})$.

Step 2. Extract $Y_c^{(1)}$ such that $R_c(x_c^{(1)}, Y_c^{(1)}) = R_f(x_f^{(1)})$.

Step 3. Extract $x_c^{(2)}$ such that $R_c(x_c^{(2)}, Y_c^{(1)}) =$ desired performance.

Step 4. Set $x_f^{(2)} = x_c^{(2)}$.

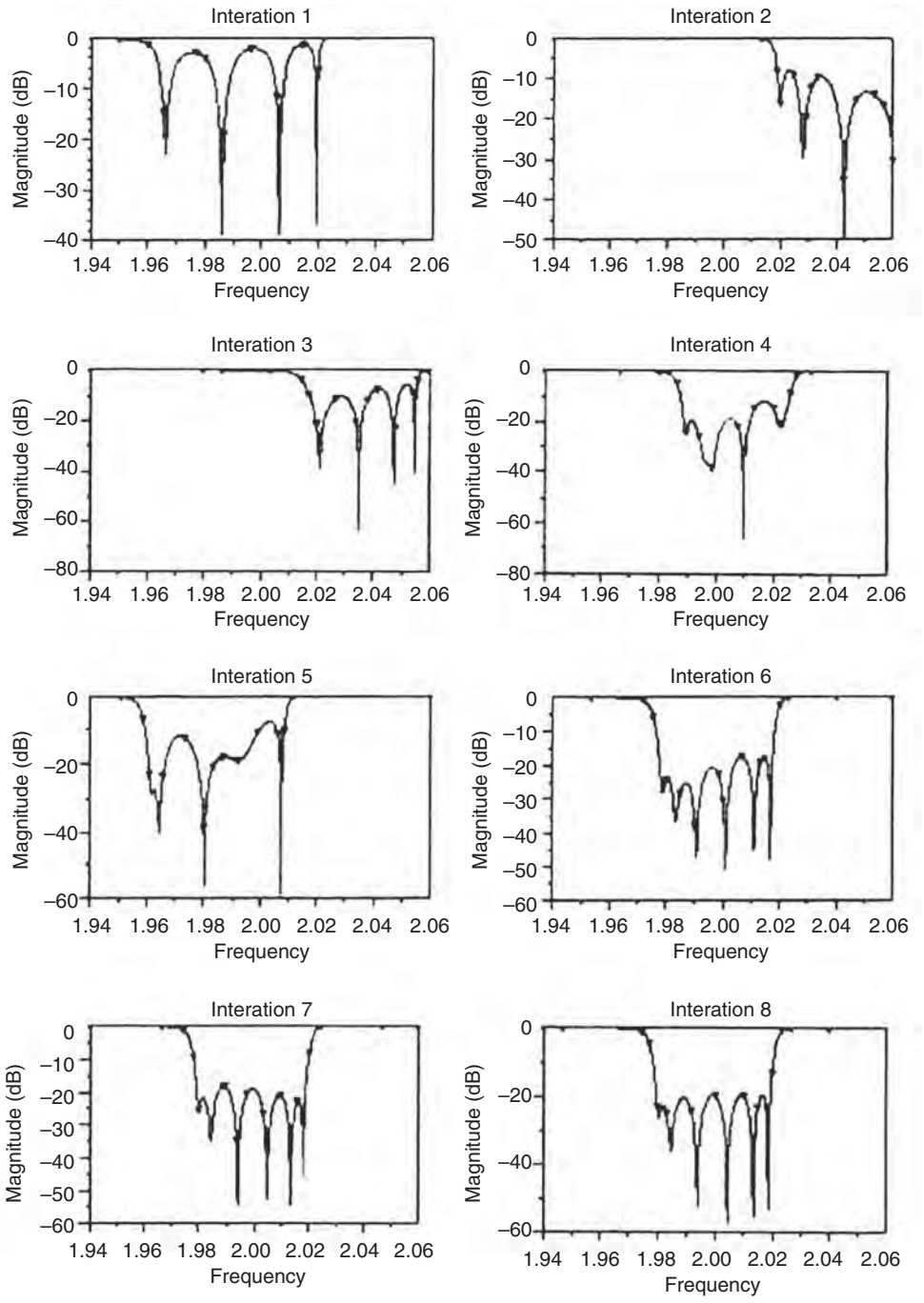


Figure 15.18 The filter responses after each iteration of the ASM technique.

Step 5. Evaluate $R_f(x_f^{(2)})$.

Step 6. Extract $Y_c^{(2)}$ such that $R_c(x_c^{(2)}, Y_c^{(2)}) = R_f(x_f^{(2)})$.

Step 7. Extract $x_c^{(3)}$ such that $R_c(x_c^{(3)}, Y_c^{(2)})$ = desired performance.

Step 8. Set $x_f^{(3)} = x_c^{(3)}$.

Step 9. Evaluate $R_f(x_f^{(3)})$.

Steps 6–9 are repeated j times, until $R_f(x_f^{(j)})$ meets the desired performance. The solution is then $x_f^{(j)}$.

Note that the flexibility of the CCM technique lies in the ability of adding the term $Y_c^{(j)}$, which allows the CCM to provide a solution that matches that of the fine model in a local sense [i.e., around the point $x_c^{(j)}$]. The following two examples demonstrate this concept.

Example 15.3 ***An Illustrative Example Demonstrating the Concept of the CCM Technique***

In this example, we consider the simple functions as initial coarse and fine models to illustrate the CCM concept:

$$R_f(x) = \frac{1}{x^2 + 0.1}$$

The problem is to find x for $R_f(x) = 5.0$:

$$R_c(x) = \frac{1}{x^2} + 1$$

In the CCM technique, we need to add another parameter Y_c to the coarse model that is varied from one iteration to another for a better alignment with the fine model. Figure 15.19 illustrates a plot for the two functions $R_f(x)$ and $R_c(x)$. It is clear that the two functions can be close to each other locally (i.e., over a given Δx) by a linear scaling; therefore, the CCM is selected as

$$R_c(x, Y_c) = \frac{c}{x^2} + 1$$

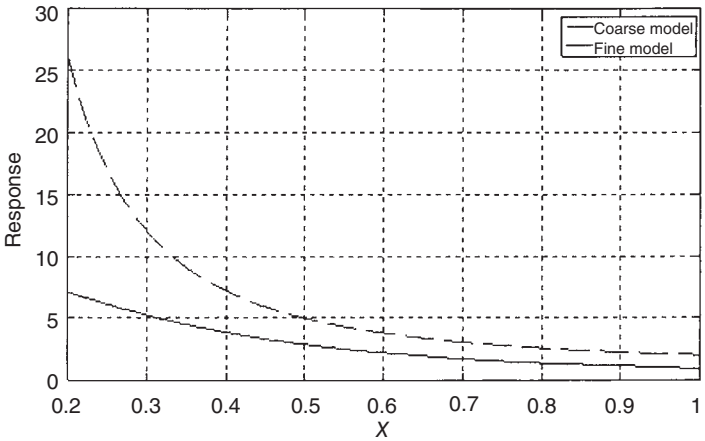


Figure 15.19 Fine and coarse model functions.

Application of the CCM technique then gives the following.

Step 0. Extract x_c^* such that $R_c(x_c^*) = 5, \rightarrow x_c^* = 0.5$. Then set $x_f^{(1)} = x_c^{(1)} = x_c^* = 0.5$.

Step 1. Evaluate $R_f(0.5) = \dfrac{1}{0.25 + 0.1} = 2.8$.

Step 2. Extract $Y_c^{(1)}$ such that $R_c(x_c^{(1)}, Y_c^{(1)}) = R_f(x_f^{(1)}), \dfrac{c}{x^2} + 1 = 2.8 \rightarrow c = 0.45$.

Step 3. Extract $x_c^{(2)}$ such that $R_c(x_c^{(2)}, Y_c^{(1)}) = 5, \dfrac{0.45}{x^2} + 1 = 5, \rightarrow x_c^{(2)} = 0.3354$.

Step 4. Set $x_f^{(2)} = x_c^{(2)} = 0.3354$.

Step 5. Evaluate $R_f(x_f^{(2)}) = \dfrac{1}{0.3354^2 + 0.1} = 4.7060$.

Step 6. Extract $Y_c^{(2)}$ such that $R_c(x_c^{(2)}, Y_c^{(2)}) = R_f(x_f^{(2)}), \dfrac{c}{0.3354^2} + 1 = 4.7060 \rightarrow c = 0.4169$.

Step 7. Extract $x_c^{(3)}$ such that $R_c(x_c^{(3)}, Y_c^{(2)}) = 5, \dfrac{0.4169}{x^2} + 1 = 5, \rightarrow x_c^{(3)} = 0.3228$.

Step 8. Set $x_f^{(3)} = x_c^{(3)} = 0.3228$.

Step 9. Evaluate $R_f(x_f^{(3)}) = \dfrac{1}{0.3228^2 + 0.1} = 4.8972$.

Step 10. Extract $Y_c^{(3)}$ such that $R_c(x_c^{(3)}, Y_c^{(3)}) = R_f(x_f^{(3)}), \dfrac{c}{0.3228^2} + 1 = 4.8972 \rightarrow c = 0.4061$.

Step 11. Extract $x_c^{(4)}$ such that $R_c(x_c^{(4)}, Y_c^{(3)}) = 5, \dfrac{0.4061}{x^2} + 1 = 5 \rightarrow x_c^{(4)} = 0.3186$.

Step 12. Set $x_f^{(4)} = x_c^{(4)} = 0.3186$.

Step 13. Evaluate $R_f(x_f^{(4)}) = \dfrac{1}{0.3186^2 + 0.1} = 4.9626$. (We can end the algorithm here.)

After four iterations the CCM technique's solution is $x = 0.3186$. The exact solution for $R_f(x) = 0.5$ is $x = 0.3162$. It is noted that the CCM technique provided a solution that is less than 1% in only four iterations; thus the fine model is used for simulation only 4 times.

Example 15.4 ***Design of the Three-Pole Filter Using the CCM Technique [38]***
Consider the three-pole microstrip filter shown in Figure 15.20. The fine model is selected to be the full EM simulation of the overall filter circuit as shown in Figure 15.20, while the coarse model is the semi-EM simulation of the circuit as shown in Figure 15.21. The capacitive parts are simulated by using an EM simulator, and the scattering matrix of the overall filter is obtained by cascading the filter subsections. A comparison between the fine model EM simulation and the coarse model semi-EM simulations is given in Figure 15.22. Although the capacitive parts are calculated by using the same EM simulator, it is evident that there is a significant difference between the results of the two models.

The coarse model does not account for the effect of cross-coupling between resonators 1 and 3, which is strong in this case. The intuitive addition of a two-port network, as shown in

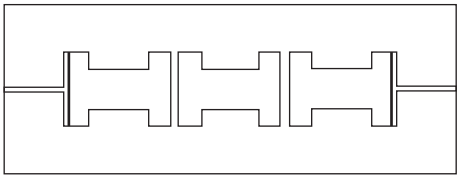


Figure 15.20 The fine EM model of the three-pole microstrip filter.

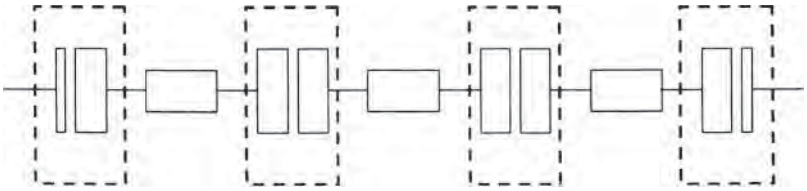


Figure 15.21 The coarse model of the three-pole filter.

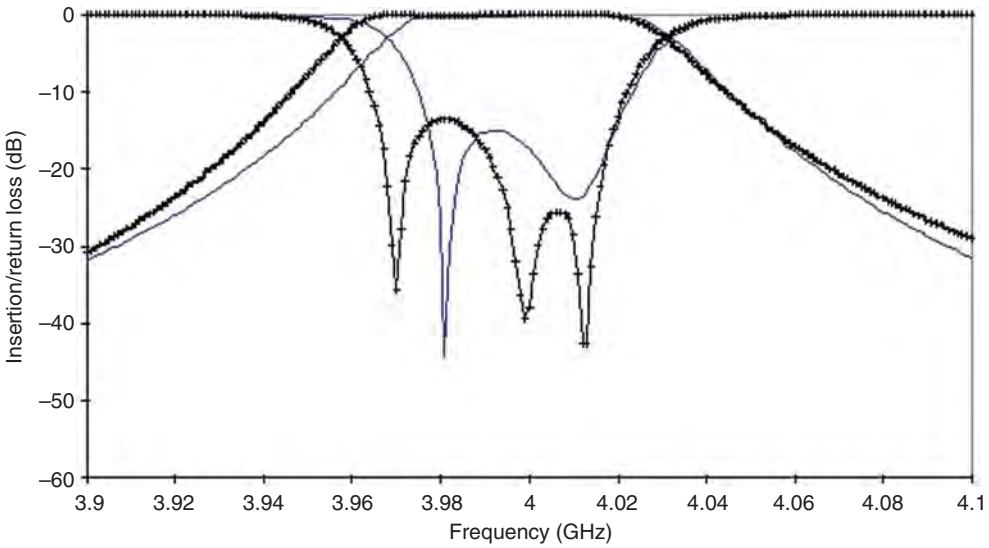


Figure 15.22 Comparison between results of the coarse model and the fine model.

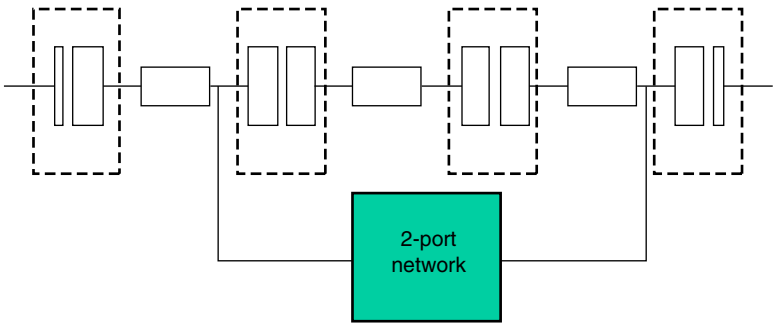


Figure 15.23 The calibrated coarse model (CCM) technique. (*Source*: Ye and Mansour 1997 [38]. Reproduced with the permission of IEEE.)

Figure 15.23, can emulate the cross-coupling. This two-port network represents the term Y_c that is added to the coarse model in the CCM technique. This two-port network is varied from iteration to another to get better alignment with the fine model. To illustrate the importance of this element in restoring alignment, Figure 15.24 shows a comparison between the fine model and the CCM in one of the iterations. Very good alignment is observed. Details of this example are presented in [38], where it is shown that only two EM simulations were needed to complete the design process.

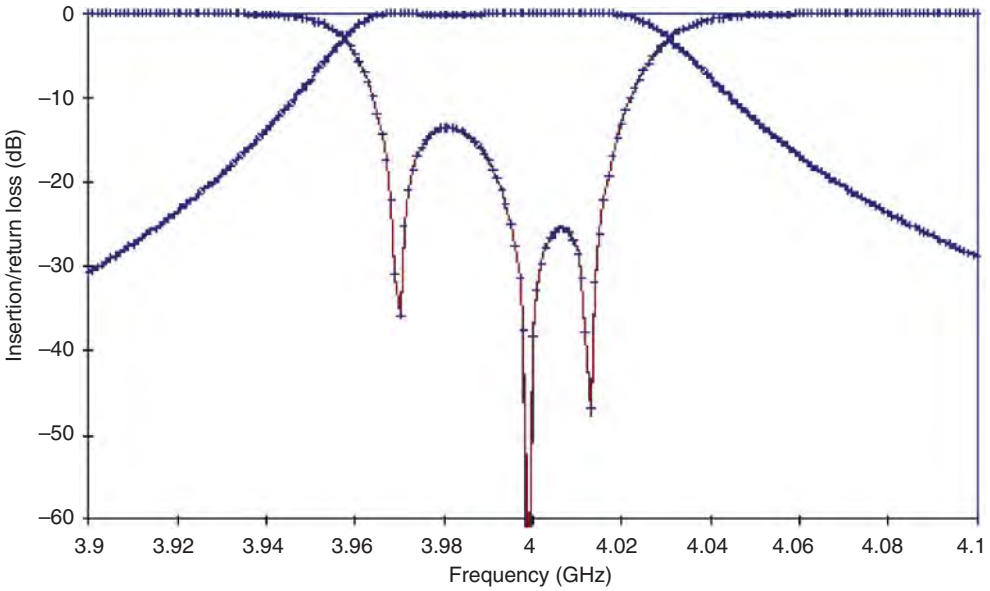


Figure 15.24 A comparison between the fine model (solid line) and the calibrated coarse model (+ + +).

15.3.3 Generalized Calibrated Coarse Model Technique for Filter Design

The CCM technique, presented in Section 15.3.2, requires some degree of engineering intuition in selecting the configuration of the term Y_c. A proper selection of Y_c may speed up the design process. As we have seen in the example of the three-pole filter, only two iterations were needed to complete the design process by properly selecting the location of the two-port network. However, the approach may not be applicable to complicated filter circuits, where it is very difficult to select the proper Y_c that can accurately calibrate the coarse model. A more systematic approach based on the CCM technique was introduced in [39, 40] to circumvent this problem.

The approach uses the EM simulator as the fine model and a circuit simulator as the coarse model. The results of these two models are then used to extract the filter coupling matrix. The calibration is done at the coupling matrix level. The coupling matrix model is selected since it provides direct interpretation of parasitic interactions among the filter elements. The technique therefore circumvents the problems in the SM approach that could arise when there is a significant misalignment between the coarse model and fine model, particularly as a result of the stray coupling between nonadjacent resonators.

For the generalized filter configuration, the filter scattering matrix can be written in terms of the M matrix as

$$S_{11} = 1 + 2jR_1[\lambda I - jR + M]^{-1}_{11} \tag{15.22}$$

$$S_{21} = -2j\sqrt{R_1 R_n}[\lambda I - jR + M]^{-1}_{n1} \tag{15.23}$$

where I is the identity matrix and λ, R, and M are as defined in Chapter 8.

We define $[S]_c$ as the filter response (filter scattering matrix) obtained from the coarse model and $[S]_f$ as the filter response obtained from the fine model. We define $[S^*]_c$ as the scattering matrix of the CCM. $[S^*]_c$ is calculated from the calibrated coupling matrix after modifying the coupling matrix of the coarse model as shown in Figure 15.25.

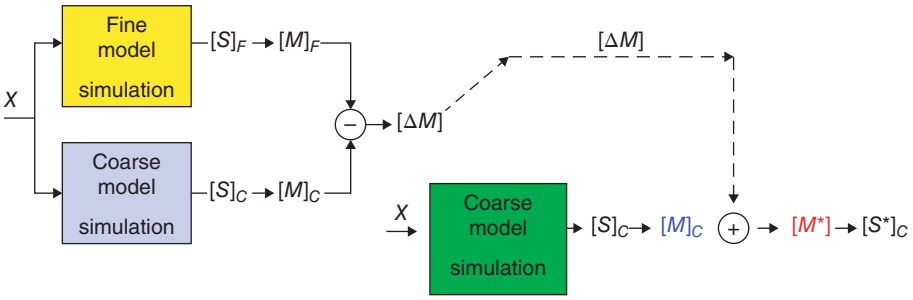


Figure 15.25 The calibrated coupling matrix technique.

With knowledge of the filter scattering parameters, we can extract the coupling matrix by optimization or by using any of the parameter extraction techniques (see Chapter 19). Thus, associated with the scattering matrices $[S]_c$, $[S]_f$, and $[S^*]_c$, we define the coupling matrices $[M]_c$, $[M]_f$, and $[M^*]_c$, respectively. The steps of the generalized CCM techniques are as follows:

Step 0. Extract x_c^* such that $[S]_c(x_c^*)$ = desired performance. Set $x_f^{(1)} = x_c^*$.
Step 1. Evaluate $[S]_f(x_f^{(1)})$.
Step 2. Extract the coupling matrix $[M]_c(x_c^*)$ from $[S]_c(x_c^*)$.
Step 3. Extract the coupling matrix $[M]_f(x_f^{(1)})$ from $[S]_f(x_f^{(1)})$.
Step 4. Define $[\Delta M]^{(0)} = [M]_f(x_f^{(1)}) - [M]_c(x_c^*)$.
Step 5. Use $[\Delta M]^{(0)}$ to modify the coarse model as illustrated in Figure 15.25.
Extract $x_c^{(1)}$ such that $[S^*]_c(x_c^{(1)})$ = Desired performance.
Step 6. Set $x_f^{(2)} = x_c^{(1)}$. Evaluate $[S]_f(x_f^{(2)})$.
Step 7. Extract the coupling matrix $[M]_c(x_c^{(1)})$ from $[S]_c(x_c^{(1)})$.
Step 8. Extract the coupling matrix $[M]_f(x_f^{(2)})$ from $[S]_f(x_f^{(2)})$.
Step 9. Define $[\Delta M]^{(1)} = [M]_f(x_f^{(2)}) - [M]_c(x_c^{(1)})$.
Step 10. Use $[\Delta M]^{(1)}$ to modify the coarse model as illustrated in Figure 15.27.
Extract $x_c^{(2)}$ such that $[S^*]_c(x_c^{(2)})$ = desired performance.
Step 11. Set $x_f^{(3)} = x_c^{(2)}$. Evaluate $[S]_f(x_f^{(3)})$ and so on until $[S]_f(x_f^{(j)})$ meets the desired performance. The solution is then $x_f^{(j)}$.

Example 15.5 ***Design of a Six-Pole Filter Using the Generalized Coupling Matrix Technique [40]***
Figure 15.26 shows an example of a six-pole microstrip filter [40] designed by using the generalized calibrated coarse model technique. The fine model is the EM simulation of the

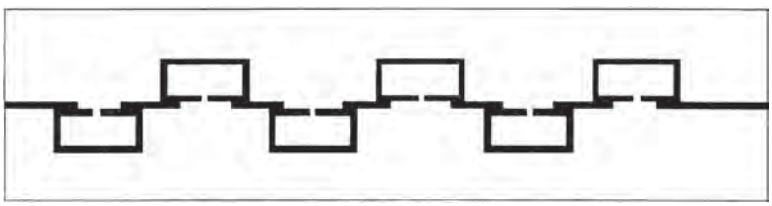


Figure 15.26 A six-pole filter designed using the generalized CCM technique. The EM simulation of the whole filter represents the fine model.

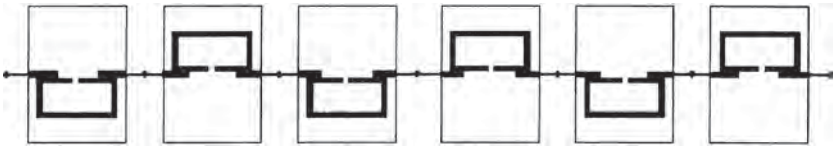


Figure 15.27 The coarse model.

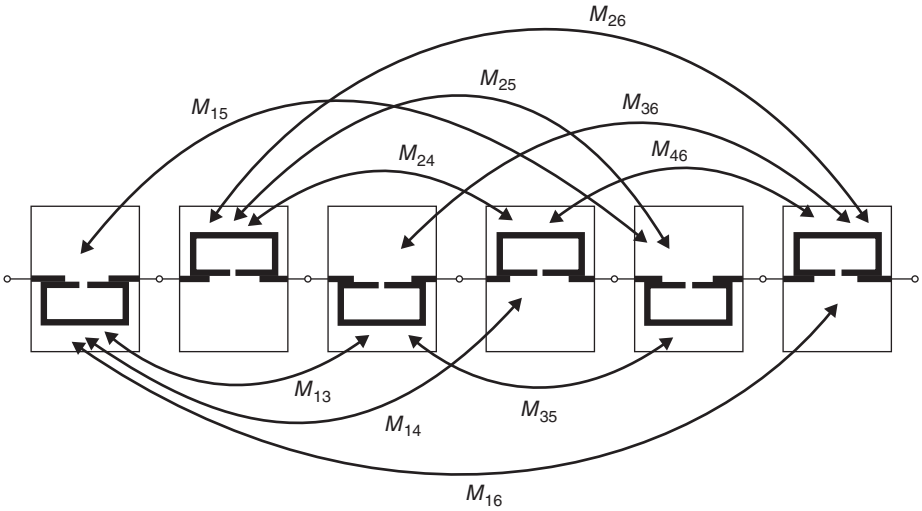


Figure 15.28 Coupling between various resonators of the microstrip filter shown in Figure 15.20.

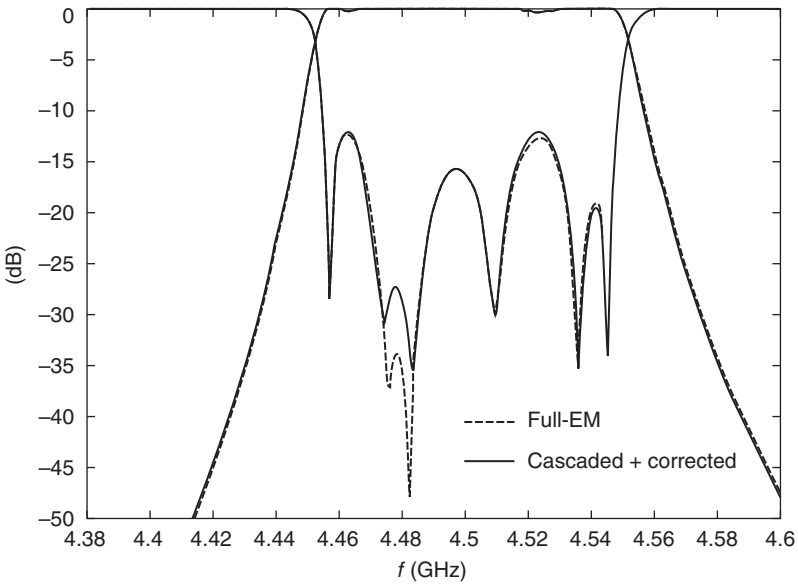


Figure 15.29 Comparison between the solution of the calibrated model (corrected model) and the fine model [40].

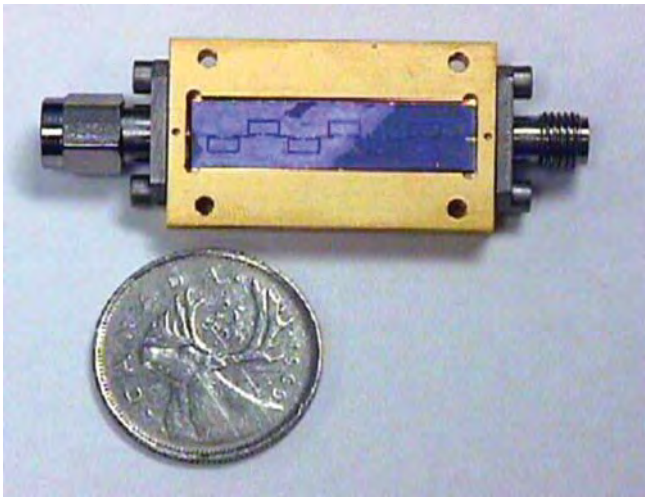


Figure 15.30 Photograph of a superconductive filter designed using the CCM technique [40].

overall filter shown in Figure 15.26. The coarse model is a semi-EM simulation and is obtained by dividing the filter into subsections as shown in Figure 15.27. The scattering matrices of these subsections are calculated by the EM simulator and then cascaded to construct the overall scattering matrix of the whole filter. The two models yield different simulation results due to the stray coupling among the nonadjacent resonators as shown Figure 15.28. This stray coupling is captured by the difference matrix $[\Delta M]$. It should be noted that the difference matrix $[\Delta M]$ does not only capture the stray coupling but also any other deviations between the two models. Figure 15.29 shows a comparison between the results of the EM simulator and the calibrated coarse model after a few iterations. A good agreement is obtained. A picture of the actual filter is given in Figure 15.30. Details of this design example are given in [39, 40].

Summary

This chapter presents several techniques for EM-based design of microwave filters. The most direct approach is to combine an accurate EM simulation tool with an optimization software package and then optimize the physical dimensions of the filter to achieve the desired performance. This is effectively a tuning process where the tuning is done by the optimization package rather than a technologist. A good initial design is needed for the optimization process to converge. Such initial designs can be obtained using the techniques described in Chapter 14. Combining the knowledge of the filter with the optimization procedure is also highly recommended. In other words, one should evaluate the relative sensitivities of the couplings and resonance frequencies of the microwave structure and make use of this information in constraining the optimization procedure. For complex filters and multiplexing networks with a large number of variables, such an approach is deemed essential in reaching a viable solution.

Direct optimization approach without any simplifying assumptions can be very computation intensive. The approach becomes impractical when dealing with large-order filters or three-dimensional filter structures. For example, in view of today's computer workstation speeds, the design of an eight-pole dielectric resonator filter using this approach can take

weeks. Several methods are offered to reduce the computation time of this brute-force technique. It includes the use of a semi-EM simulator, EM simulators with fast frequency sweep or EM-based interpolation and polynomials to replace the actual EM simulator. Although the use of these approaches may help in reducing the optimization CPU time, it certainly impacts the accuracy of the design.

In this chapter, we also include two advanced EM-based techniques to optimize the physical dimensions of microwave filters: the SM technique and the CCM technique. These techniques are described in detail, outlining, step by step, the implementation procedure. To demonstrate the concepts, simple examples are given that should allow the reader to duplicate numerical results using a calculator. Details of a design example of a six-pole microstrip filter using the ASM technique are included to demonstrate the benefits of using such technique in microwave filter design. Two examples for designing filters using the CCM technique are outlined. While the direct optimization process may take 100s of EM simulations to complete the optimization process, only a few EM simulations are needed when using the ASM or the CCM techniques. Several variations of the SM technique have been reported in the literature. This technique is emerging as the most useful approach for accurate EM-based design of microwave filters.

References

1 Bandler, J.W. and Chen, S.H. (1988) Circuit optimization: the state of the art. *IEEE Transactions on Microwave Theory and Techniques*, **36**, 424–443.
2 Fletcher, R. (1987) *Practical Methods of Optimization*, 2nd edn, Wiley, New York.
3 Steer, M.B., Bandler, J.W., and Snowden, C.M. (2002) Computer-aided design of RF and microwave circuits and systems. *IEEE Transactions on Microwave Theory and Techniques*, **50**, 996–1005.
4 Bandler, J.W., Chen, S.H., Daijavad, S., and Madsen, K. (1988) Efficient optimization with integrated gradient approximations. *IEEE Transactions on Microwave Theory and Techniques*, **36**, 444–455.
5 Haupt, R.L. (1995) An introduction to genetic algorithms for electromagnetics. *IEEE Antennas and Propagation Magazine*, **37**, 7–15.
6 Hald, J. and Madsen, K. (1985) Combined LP and quasi-Newton methods for nonlinear LI optimization. *SIAM Journal on Numerical Analysis*, **22**, 68–80.
7 Bandler, J. *et al.* (1994) Microstrip filter design using direct EM field simulation. *IEEE Transactions on Microwave Theory and Techniques*, **42**, 1353–1359.
8 ANSYS HFSS, http://www.ansys.com/products/electronics/ansys-hfss.
9 em sonnet, http://www.sonnetusa.com.
10 Agilent Momentum, http://eesof.tm.agilent.com.
11 Zeland Software, http://www.microwavejournal.com/authors/206-zeland-software-inc.
12 CST, Microwave Studio, www.cst.com.
13 MiCIAN, Microwave Wizard, http://www.mician.com.
14 Brezinski, C. (1980) *Padé-Type Approximation and General Orthogonal Polynomials*, Birkhauser Verlag, Basel, Switzerland.
15 Ureel, J., *et al*. (1994) Adaptive frequency sampling algorithm of scattering parameters obtained by electromagnetic simulation. IEEE AP-S Symposium Digest, pp. 1162–1167.
16 Dhaene, T., Ureel, J., Fache, N. and De Zutter, D. (1995) Adaptive frequency sampling algorithm for fast and accurate S-parameter modeling of general planar structures. IEEE MTT-S Symposium Digest, pp. 1427–1431.

17 Peik, S.F., Mansour, R.R., and Chow, Y.L. (1998) Multidimensional Cauchy method and adaptive sampling for an accurate microwave circuit modeling. *IEEE Transactions on Microwave Theory and Techniques*, **46** (12), 2364–2371.

18 Stoer, J. and Bulirsch, R. (1980) *Introduction to Numerical Analysis*, Springer-Verlag, Berlin, Sec. 2.2.

19 Zhang, Q.J. and Gupta, K.C. (2000) *Neural Networks for RF and Microwave Design*, Artech House, Norwood, MA.

20 Zhang, Q.J., Deo, M., and Xu, J. (2005) Neural network for microwave circuits, in *Encyclopedia of RF and Microwave Engineering*, Wiley.

21 Peik, S.F., Coutts, G. and Mansour, R.R. (1998) Application of neural networks in microwave circuit modeling. IEEE Canadian Conference of Electrical and Computer Engineering, 24–28 May 1998, pp. 928–931.

22 Burrascano, P., Dionigi, M., Fancelli, C. and Mongiardo, M. (1998) A neural network model for CAD and optimization of microwave filters. IEEE-MTT-S Microwave Symposium Digest, June 1998, pp. 13–16.

23 Zhang, Q.J. (2004) *Neuromodeler Software*', Department of Electronics, Carleton University, Ottawa, Canada.

24 MATLAB software, www.mathworks.com/.

25 Watson, P.M., Gupta, K.C. and Mahajan, R.L. (1998) Development of knowledge based artificial neural networks models for microwave components. IEEE MTT-S International Microwave Symposium Digest, Baltimore, pp. 9–12.

26 Zhang, Q.J. (1997) Knowledge based neuromodels for microwave design. *IEEE Transactions on Microwave Theory and Techniques*, **45**, 2333–2343.

27 Miraftab, V. and Mansour, R.R. (2002) Computer-aided tuning of microwave filters using fuzzy logic. *IEEE Transactions on Microwave Theory and Techniques*, **50** (12), 2781–2788.

28 Wang, L.X. and Mendel, J. (1992) Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Transactions on Neural Networks*, **3** (5), 807–814.

29 Miraftab, V. and Mansour, R.R. (2003) EM-based design tools for microwave circuits using fuzzy logic techniques. IEEE MTT-S International Microwave Symposium Digest, June 2003, pp. 169–172.

30 Miraftab, V. (2006) Computer-aided design and tuning of microwave circuits using fuzzy logic. Ph.D. thesis. University of Waterloo, Ontario, Canada.

31 Bandler, J.W., Biernacki, R.M., Chen, S.H. *et al.* (1994) Space mapping technique for electromagnetic optimization. *IEEE Transactions on Microwave Theory and Techniques*, **42**, 2536–2544.

32 Bandler, J.W., Biernacki, R.M., Chen, S.H. *et al.* (1995) Electromagnetic optimization exploiting aggressive space mapping. *IEEE Transactions on Microwave Theory and Techniques*, **43**, 2874–2882.

33 Bandler, J.W., Biernacki, R.M., Chen, S.H., and Huang, Y.F. (1997) Design optimization of interdigital filters using aggressive space mapping and decomposition. *IEEE Transactions on Microwave Theory and Techniques*, **45**, 761–769.

34 Bandler, J.W. *et al.* (2004) Space mapping: the state-of-the-art. *IEEE Transactions on Microwave Theory and Techniques*, **52**, 337–361.

35 Bakr, M.H., Bandler, J.W., Madsen, K., and Søndergaard, J. (2000) Review of the space mapping approach to engineering optimization and modeling. *Optimization and Engineering*, **1**, 241–276.

36 Rayas-Sanchez, J.E. (1999) Space mapping optimization for engineering design: a tutorial presentation. Proceedings of the Workshop Next Generation Methodologies for Wireless and Microwave Circuit Design, McMaster University, June 1999.

37 Agilent ADS, http://eesof.tm.agilent.com.
38 Ye, S. and Mansour, R.R. (May 1997) An innovative CAD technique for microstrip filter design. *IEEE Transactions on Microwave Theory and Techniques*, **45**, 780–786.
39 Peik, S.F. (1999) Efficient design and optimization techniques for planar microwave filters. Ph.D. thesis. University of Waterloo, Ontario, Canada.
40 Peik, S.F. and Mansour, R.R. (2002) A novel design approach for microwave planar filters. IEEE MTT-S International Microwave Symposium Digest, June 2002, pp. 1109–1112.
41 Bila, S. *et al.* (2001) Direct electromagnetic optimization of microwave filters. *IEEE Microwave Magazine*, **2**, 46–51.

16

Dielectric Resonator Filters

The term *dielectric resonator* was first reported in 1939 by Richtmyer [1] of Stanford University, who showed that dielectric objects can function as microwave resonators. It was not, however, before the 1960s that several pioneering works were published on the behavior of dielectrics at microwave frequencies, including the first reports on microwave dielectric resonator filters by Cohn in 1965 [2, 3] and by Harrison [4]. However, such filters were not implemented in practical applications because of the poor thermal stability of the dielectric resonator materials available at that time. The dielectric constant of these materials exhibited a significant change in their values with temperature variation. This in turn caused the temperature drift of the filter center frequency to be as large as 500 ppm/°C.

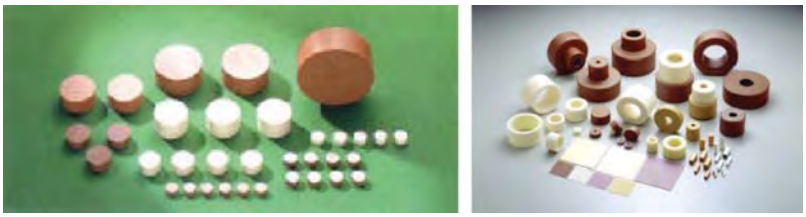
Advances in the 1970s and 1980s in dielectric materials made it possible to combine high unloaded Q, high dielectric constants, and a small temperature drift in materials suitable for use at microwave frequencies. High-Q dielectric materials with dielectric constants ranging from 20 to 90, and a temperature drift from −6 to +6 ppm/°C, are now commercially available from various manufacturers. Dielectric resonators with $\varepsilon_r = 29$ are commercially available with quality factor-frequency product $Q \times f$ values of 90,000; thus, an unloaded Q value of more than 50,000 can be achieved at 1.8 GHz. As the dielectric constant increases, the achievable unloaded Q typically decreases. For materials with a dielectric constant of 45, the $Q \times f$ value reduces to 44,000 [5]. Figure 16.1 illustrates various dielectric materials available commercially for dielectric resonator filter applications.

Typically, a dielectric resonator consists of a high-dielectric-constant material in a cylindrical form (known as a *dielectric puck*), mounted inside a metal enclosure using a low-dielectric-constant support as shown in Figure 16.2. The dimensions of the metal enclosure are chosen such that metal enclosure (without dielectric) operates in an evanescent mode at the operating frequency. The electromagnetic (EM) field is concentrated mainly in the dielectric puck, and the Q factor of the resonator is determined largely by the loss tangent of the dielectric materials.

16.1 Resonant Frequency Calculation in Dielectric Resonators

In contrast to conventional rectangular and circular waveguide resonators, an exact solution of Maxwell's equations in dielectric resonators can be computed only by a numerical EM method such as the mode matching technique [6, 7], finite element analysis [8], or integral equation technique [9]. Commercial software packages such as HFSS [8] and CST Microwave Studio [10] can be readily used to calculate the resonant frequency, field distribution, and resonator Q of dielectric resonators having any arbitrary shape. A software package, based on

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.



(a) (b)

Figure 16.1 Various dielectric materials commercially available for dielectric resonator filter applications: (a) resonators and (b) resonators, supports, and substrates.

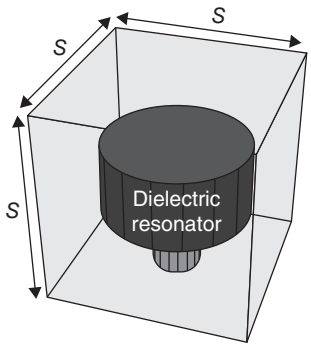


Figure 16.2 A dielectric resonator consisting of a dielectric puck mounted inside a metal enclosure.

the mode matching technique, for the analysis of dielectric resonators of cylindrical shapes is also commercially available [11]. The package is capable of accurate calculation of both the resonant frequency and Q at a relatively fast speed.

For $TE_{01\delta}$ modes, one can assume that a magnetic wall surrounds the resonator and deals with the problem as a homogeneously filled cylindrical resonator [12]. The magnetic wall assumption is based on the fact that an incident wave from a high-dielectric-constant medium on an air-filled medium sees a reflection coefficient of approximately 1. This assumption simplifies the analysis and provides reasonable results for the resonant frequency and field distribution of $TE_{01\delta}$ modes [12].

Consider the dielectric resonator shown in Figure 16.3. Because of the cylindrical symmetry of the structure, the equations for the radial and azimuthal components of EM fields with $e^{-j\beta z}$ propagating in the z direction are given by [12, 13]

$$E_r = \frac{-j}{k_c^2}\left(\beta\frac{\partial E_z}{\partial r} + \frac{\omega\mu}{r}\frac{\partial H_z}{\partial \phi}\right) \tag{16.1}$$

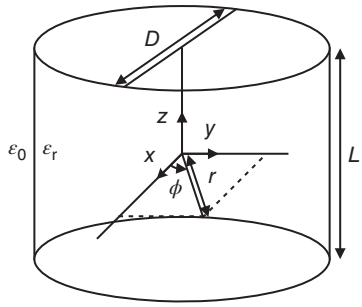
$$H_r = \frac{j}{k_c^2}\left(\frac{\omega\varepsilon}{r}\frac{\partial E_z}{\partial \phi} - \beta\frac{\partial H_z}{\partial r}\right) \tag{16.2}$$

$$E_\phi = \frac{-j}{k_c^2}\left(\frac{\beta}{r}\frac{\partial E_z}{\partial \phi} - \omega\mu\frac{\partial H_z}{\partial r}\right) \tag{16.3}$$

$$H_\phi = \frac{-j}{k_c^2}\left(\omega\varepsilon\frac{\partial E_z}{\partial r} + \frac{\beta}{r}\frac{\partial H_z}{\partial \phi}\right) \tag{16.4}$$

where $k_c^2 = k^2 - \beta^2$, $k = \sqrt{\varepsilon_r}k_0$. The magnetic field of the TE mode with $E_z = 0$ satisfies the Helmholtz equation, $(\nabla^2 + k^2)H_z = 0$. Expressing the Helmholtz equation in the cylindrical

Figure 16.3 Cylindrical dielectric resonator.



coordinate system, we have

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}H_z\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\phi^2}(H_z)+\frac{\partial^2}{\partial z^2}(H_z)+k^2H_z=0 \tag{16.5}$$

Continuing with the assumption of a propagation in the z direction implies that H_z must vary as $H_z = R(r)\Phi(\phi)e^{-j\beta z}$. Substituting the expression for H_z into equation (16.5) and dividing both sides by r^2H_z results in

$$\underbrace{\frac{r}{R(r)}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}R(r)\right)+r^2(k^2-\beta^2)}_{\text{function of } r \text{ only}}+\underbrace{\frac{1}{\Phi(\phi)}\frac{\partial^2}{\partial\phi^2}(\Phi(\phi))}_{\text{function of } \phi \text{ only}}=0 \tag{16.6}$$

The first and second terms are functions of r and ϕ, respectively. The sum of these two terms must be a constant number, and hence the terms themselves must have constant values, since the equation must hold for any combination of the two independent degrees of freedom r and ϕ. Letting

$$\frac{1}{\Phi(\phi)}\frac{\partial^2}{\partial\phi^2}(\Phi(\phi))=-k_\phi^2 \tag{16.7}$$

with $\Phi(\phi)=\Phi(\phi+2\pi)$ for self-consistency, we have $\Phi(\phi)=A\cos(k_\phi\phi)+B\sin(k_\phi\phi)$, and with $k_\phi = n = 1,2,3,\ldots$ with $k_c^2 = k^2-\beta^2$, the radial equation becomes

$$r^2\frac{\partial^2R(r)}{\partial r^2}+r\frac{\partial R(r)}{\partial r}+(k_c^2r^2-n^2)R(r)=0 \tag{16.8}$$

The physically acceptable (finite) solutions to equation (16.8) are the Bessel functions of the first kind [9]:

$$R(r)=CJ_n(k_cr) \tag{16.9}$$

A plane wave traveling in a dielectric medium with a relative permittivity of $\varepsilon_r \gg 1$, incident normally at a planar dielectric boundary, has very low characteristic impedance, $\eta_D=\eta_0/\sqrt{\varepsilon_r}$, in comparison with the air characteristic impedance. As a result, the field will be reflected according to

$$\vec{E}_{\text{ref}}=\vec{E}_{\text{inc}}\frac{(\eta_0-\eta_D)}{(\eta_0+\eta_D)}\sim E_{\text{inc}} \tag{16.10}$$

In analogy with electrical open circuits, the approximation is called the *magnetic wall boundary condition* [12] and simplifies the boundary condition analysis significantly. Imposing the magnetic wall boundary condition further implies that $H_z=0$ at $r=a$ and quantizes k_c so that

$$J_n(k_ca)=0,\quad k_c^{(nm)}=\frac{p_{nm}}{a} \tag{16.11}$$

where p_{nm} is the value of the mth root of the nth Bessel function $J_n(k_c r)$. With this constraint, the total solution for (nm)th mode is

$$H_z = H_0 J_n(k_c^{(nm)} r)(A\cos(n\phi) + B\sin(n\phi))e^{-j\beta z} \tag{16.12}$$

The derivation of the resonance condition for the $TE_{01\delta}$ mode requires an additional boundary condition on the dielectric–air interfaces at $Z = L/2$ and $Z = -L/2$. Setting $n = 0, m = 1$, we have $H_z = AJ_0(k_c^{(01)} r)e^{-j\beta z}$, which implies $(\partial H_z)/(\partial\phi) = 0$. Using the fact that $E_z = 0$ for TE modes and $(\partial E_Z)/(\partial r) = (\partial E_z)/(\partial\phi) = 0$ for the $TE_{01\delta}$ mode, equations (16.1)–(16.4) reduce to

$$E_\phi = AJ_0'(k_c^{(01)} r)\cos(\beta z) \tag{16.13}$$

$$H_r = -\frac{jA\beta}{\omega\mu_0} J_0'(k_c^{(01)} r)\sin(\beta z) \tag{16.14}$$

with $E_r = H_\phi = 0$ for $|Z| \le L/2$ (i.e., in the dielectric medium). The evanescent decaying solutions in air for $|Z| > L/2$ are therefore [12]

$$E_\phi = BJ_0'(k_c^{(01)} r)\exp(-\alpha|z|) \tag{16.15}$$

$$H_r = -\frac{jB\alpha}{\omega\mu_0} J_0'(k_c^{(01)} r)\exp(-\alpha|z|) \tag{16.16}$$

where $\alpha = \sqrt{k_c^2 - k_0^2}$. The dielectric interface boundary condition dictates that the transverse components of the electric field be equal on both sides of the boundary: $E_\phi(L/2^+) = E_\phi(L/2^-)$. Similarly, the radial component of the magnetic flux density must be equal, $H_r(L/2^+) = H_r(L/2^-)$, since both materials are nonmagnetic. Substituting (16.13)–(16.16), we have

$$A\cos\left(\frac{\beta L}{2}\right) = B\exp\left(-\frac{\alpha L}{2}\right) \tag{16.17}$$

$$A\beta\sin\left(\frac{\beta L}{2}\right) = B\alpha\exp\left(-\frac{\alpha L}{2}\right) \tag{16.18}$$

with

$$\beta = \sqrt{\varepsilon_r k_0^2 - \left(\frac{2p_{01}}{D}\right)^2} \tag{16.19}$$

$$\alpha = \sqrt{\left(\frac{2p_{01}}{D}\right)^2 - k_0^2} \tag{16.20}$$

The solution for diameter D of the dielectric resonator, assuming $D/L = 3$, $\varepsilon_r = 45$, at a particular resonant frequency, is found by numerically solving equations (16.17)–(16.20). A plot of the solutions for several different resonant frequencies is given in Figure 16.4.

The field distribution for the $TE_{01\delta}$ resonant mode is important when designing the input/output coupling. The field equations associated with the $TE_{01\delta}$ mode are

$$E_\phi = AJ_0'(k_c^{(01)} r)\cos(\beta z) \tag{16.21}$$

$$H_r = -\frac{jA\beta}{\omega\mu_0} J_0'(k_c^{(01)} r)\sin(\beta z) \tag{16.22}$$

$$H_z = AJ_0(k_c^{(01)} r)e^{-j\beta z} \tag{16.23}$$

$$E_r = E_z = H_\phi = 0 \tag{16.24}$$

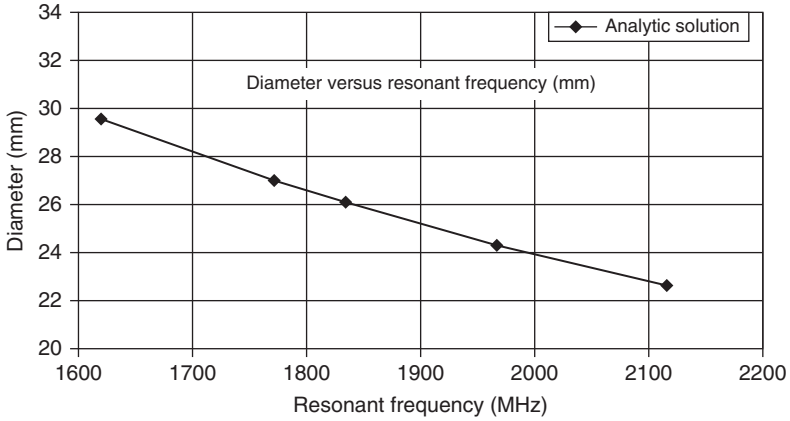


Figure 16.4 Analytic solution of diameter versus resonant frequency, $D/L = 3$ and $\varepsilon_r = 45$.

In view of these field equations, it is noted that $E_r = 0$ and $E_z = 0$; that is, the E-field has only a ϕ component. The magnetic field is a linear combination of vectors pointing in the $\hat{\mathbf{e}}_r$ and $\hat{\mathbf{e}}_z$ directions. Because of the sinusoidal and cosinusoidal variation of $\mathbf{H}_r$ and $\mathbf{H}_z$, respectively, their linear combination produces circular H-field lines perpendicular to the electric field lines.

16.2 Rigorous Analyses of Dielectric Resonators

Because of the simplification of the boundary condition on the radial and azimuthal fields at the dielectric boundary, the analysis given in Section 16.1 provides only an approximate solution. In addition, dielectric resonators are usually placed inside a metal enclosure to shield the resonator from outside radiation. A dielectric support of low relative permittivity is used in order to physically suspend the dielectric resonator away from the surfaces of the metal box as shown in Figure 16.2.

These additional parameters are not considered in the approximate analytical solution given in Section 16.1, but they can be included in a full EM simulation such as the mode matching technique [6, 7] or finite element analyses [8]. The finite element analysis in particular can deal with any arbitrary shaped dielectric objects inside the enclosure. Figure 16.5 shows a comparison [13] between the results obtained for the resonator structure given in Figure 16.3 using the analytical solution given in Section 16.1 and the finite element technique HFSS [8]. The resonator is centered inside the cavity. No support structure is assumed in computing the results given in Figure 16.5. The comparison shows that there is about 15% deviation in resonant frequency between the results of the analytical solution and those of the finite element HFSS solution [8].

16.2.1 Mode Charts for Dielectric Resonators

Consider the cylindrical dielectric resonator shown in Figure 16.6. The parameters of this resonator are given in Table 16.1.

The modes in dielectric resonators can be categorized as TEE, TEH, TME, TMH, HEE, and HEH. The third index was used in [7, 11] to differentiate between modes with even and odd symmetry along the axis of the resonator at $Z = L/2$. E is used to denote electric wall symmetry (i.e., electric wall at $Z = L/2$), while "H" is used to denote magnetic wall symmetry (i.e., magnetic wall at $Z = L/2$). The dominant TEH mode is basically the $TE_{01\delta}$ mode. Figures 16.7–16.10

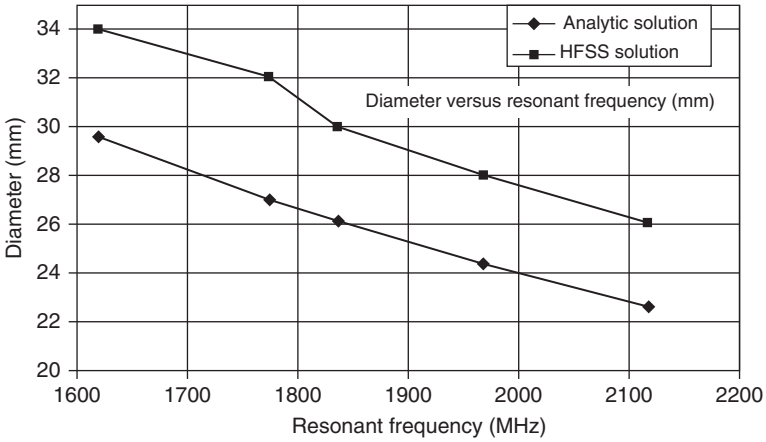


Figure 16.5 Comparison plot of resonant frequency versus diameter, $D/L = 3$ and $\varepsilon_r = 45$.

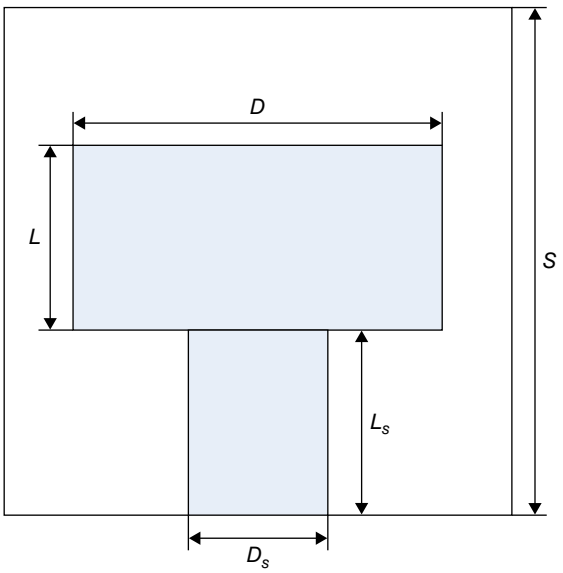


Figure 16.6 Dielectric resonator.

Table 16.1 Parameters and dimensions of the resonator given in Figure 16.6.

	Parameters and dimensions
Dielectric resonator	$\varepsilon_r = 34$, $D = 1.176$ in.; $L = 0.481$ in.
Support	$\varepsilon_r = 10$, $D_s = 0.56$ in.; $L_s = 0.8$ in.
Enclosure	$2 \times 2 \times 2.03$ in.^{3} ($C = 2$ in., $S = 2.03$ in.)

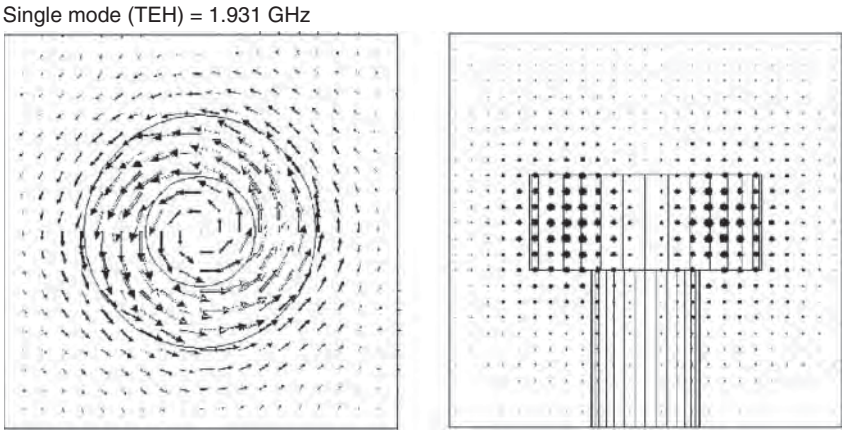


Figure 16.7 Electric field distribution of the TEH mode.

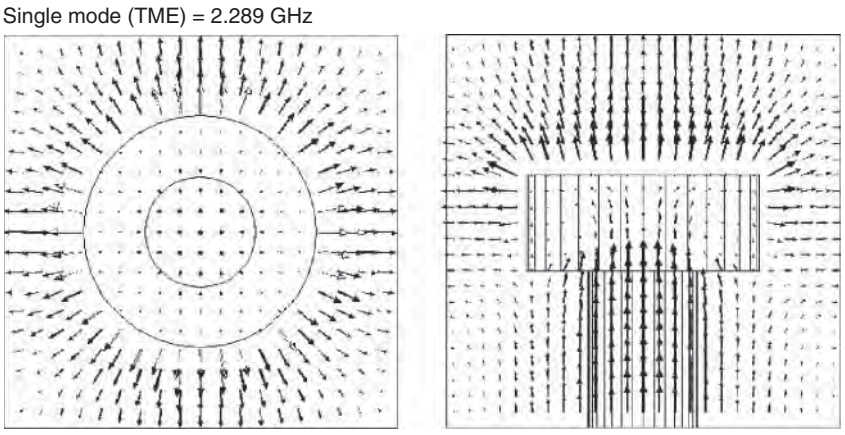


Figure 16.8 Electric field distribution of the TME mode.

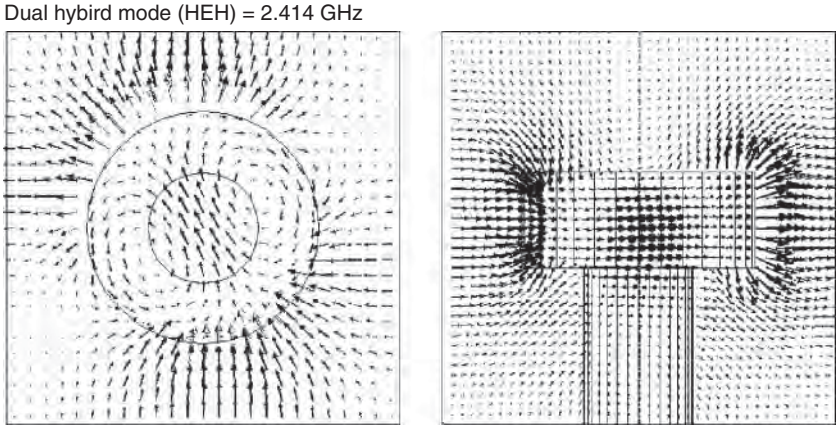


Figure 16.9 Electric field distribution of the HEH mode.

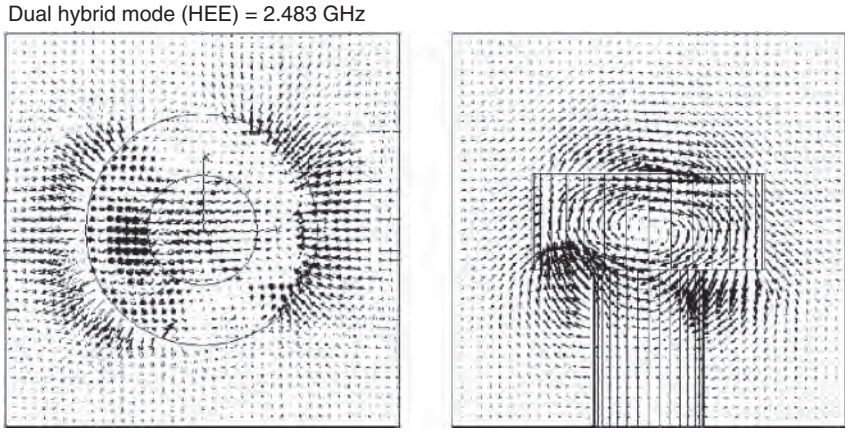


Figure 16.10 Electric field distribution of the HEE mode.

Table 16.2 Resonance frequency in GHz of first four modes in dielectric resonators.[a)]

Ratio (*D*/*L*)	*D* (in.)	*L* (in.)	TEH (Single mode)	HEH (Dual mode)	HEE (Dual mode)	TME (Single mode)
1	1.176	1.176	1.672	1.909	1.605	1.521
1.5	1.176	0.784	1.739	2.218	1.883	2.027
2	1.176	0.588	1.841	2.331	2.188	2.193
2.445	1.176	0.481	1.931	2.414	2.484	2.289
3	1.176	0.392	2.039	2.518	2.838	2.381
3.5	1.176	0.336	2.131	2.597	3.111	2.437
4	1.176	0.294	2.220	2.669	3.209	2.481

a) *D* is fixed while *L* is varied. Resonator parameters and dimensions are given in Table 16.1.

show the field distribution of the first four modes (TEH, TME, HEH, and HEE) and the corresponding resonant frequencies for $D = 1.176$ in. and $L = 0.481$ in. as obtained by HFSS [8]. Both the TEH and TME are single modes, whereas both HEH and HEE are dual hybrid modes. The resonant frequencies of these modes, calculated by assuming a constant diameter D of 1.176 in. for various values of resonator thickness L, are given in Table 16.2. The results given in this table translate to the mode chart given in Figure 16.11. Details of theoretical and experimental evaluation of the resonant frequency and unloaded Q factor are given in Chapter 11.

Note that for TME modes, the E-field is in the z direction and exists in the support. The use of a support of a lower dielectric constant will shift the resonant frequency of the TME mode to a higher frequency range away from that of the TEH mode.

16.3 Dielectric Resonator Filter Configurations

Several dielectric resonator configurations have been reported in the literature over the past two decades [14–19]. A typical dielectric resonator filter consists of a number of dielectric resonators mounted inside cavities operating below cutoff. The cavities are separated by irises to

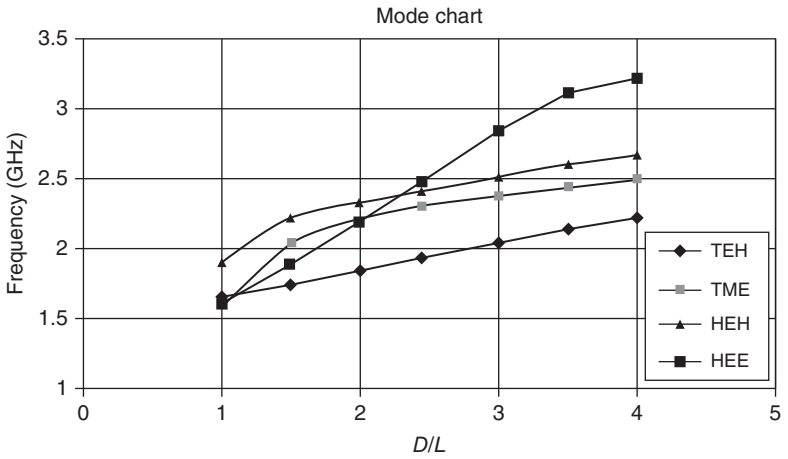
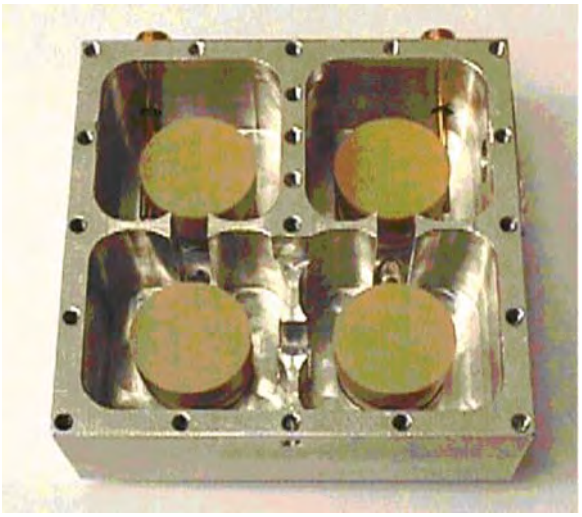


Figure 16.11 The mode chart for the first four modes of the dielectric resonator. (Generated from Table 16.2.)

Figure 16.12 A dielectric resonator filter in a planar configuration.



provide the necessary coupling between resonators. The dielectric resonators are supported inside the cavity in either planar or axial configuration as shown in Figures 16.12 and 16.13, respectively. While the axial configuration is analogous to cylindrical cavity filters, which makes it easy to select the proper iris shape for coupling, the planar configuration offers simplicity in designing the resonator support structure. The planar configuration also provides flexibility in folding the filter structure and providing coupling between nonadjacent resonators, which in turn makes it possible to create filters with advanced functions such as quasi-elliptic and linear phase filters.

As shown in Section 16.2, there are several possible operating modes for dielectric resonator filters. These modes have an impact on the filter size, unloaded Q, and spurious performance. Single-mode DR filters that operate in $TE_{01\delta}$ (TEH) modes are the most commonly used designs, particularly in wireless applications. The design offers the highest achievable filter Q in comparison with other modes of operation. For a dielectric constant of 45, the size of a dielectric resonator cavity (DR resonator + enclosure) at 1.8 GHz is about $1.6 \times 1.6 \times 1.7$ in.^{3}, or

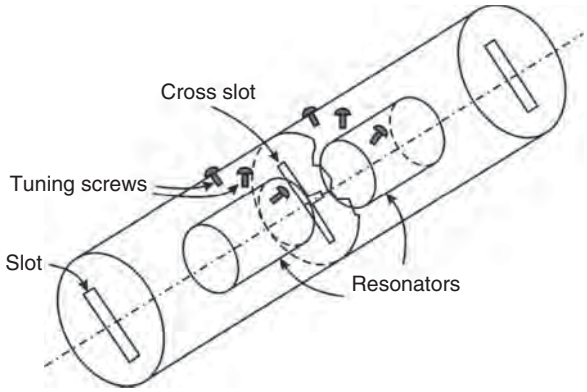


Figure 16.13 A dielectric resonator filter in an axial configuration.

a volume of 4.4 in.^{3} The volume would be increased by about 5–10% if a cylindrical dielectric resonator with a hole in the center were used rather than a solid disk resonator. The hole in the center would slightly improve the resonator spurious performance, but it is preferred mainly because it facilitates alignment and mounting of the resonator inside the enclosure. Chapter 14 gives a step-by-step numerical example on the design of single-mode dielectric resonator filters.

Dual-mode dielectric resonator filters offer roughly 30% volume saving in comparison with single-mode dielectric resonator filters. Assuming a dielectric constant of 45, the size of a dual-mode DR resonator would be $1.8 \times 1.8 \times 1.7$ in.^{3}, a volume of 5.5 in.^{3} for two (2) electrical resonators. Figure 16.14 illustrates an eight-pole dual-mode filter designed for wireless code-division multiple-access (CDMA) applications [20]. It consists of four resonators operating in dual HEH modes. The resonators are supported on a plate using low-loss dielectric supports, which in turn are mounted on a housing having four cavities separated by irises.

Triple-mode DR resonators [21] offer roughly 50% volume saving in comparison with single-mode DR filters. Figure 16.15 shows a triple-mode dielectric resonator [21] realized by using a cubic dielectric resonator inserted inside a cubic cavity. The resonator consists of three orthogonal $TE_{01\delta}$ modes that are coupled to each other by tuning screws. Quadruple-mode dielectric resonator filters have been also investigated. Figure 16.16 illustrates a quadruple-mode resonator proposed in [22]. The resonator operates in four orthogonal

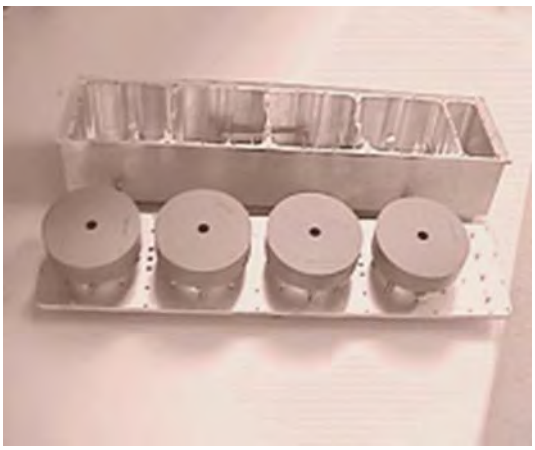


Figure 16.14 An eight-pole dual-mode filter for CDMA applications [20].

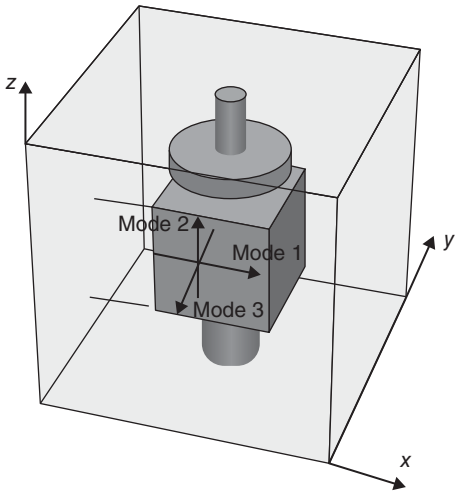


Figure 16.15 A triple-mode dielectric resonator. *Source*: Walker and Hunter [21] 2002. Reproduced with the permission of IEEE.

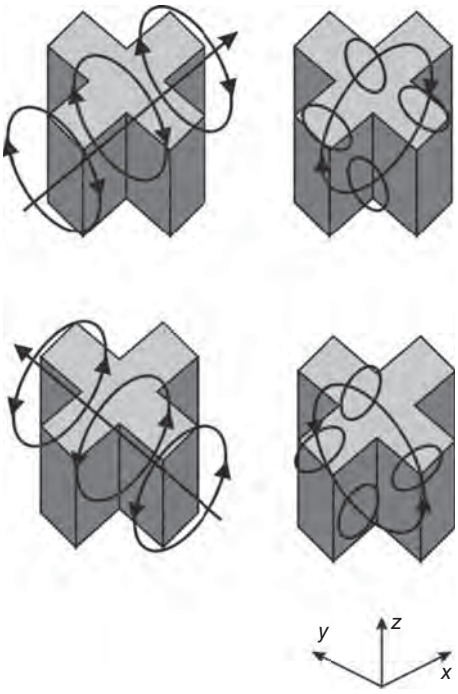


Figure 16.16 The quadruple-mode dielectric resonator illustrating the four orthogonal modes. *Source*: Hattori *et al*. [22] 2003. Reproduced with the permission of IEEE.

modes (two TM modes and two TE modes). Details of the coupling between these modes are given in [22]. Both triple-mode and quadruple-mode designs require the use of dielectric resonators of irregular shapes, which are not readily available from dielectric resonator suppliers. Therefore, the cost associated with resonator machining and tuning of such types of filters may overshadow the benefits of the volume reduction.

Table 16.3 summarizes the size and achievable loaded Q values of the above three designs at 1.8 GHz assuming dielectric resonators of $\varepsilon_r = 29, 45$. It is noted that the achievable filter Q value is typically 70–80% that of the unloaded Q value. The reduction in Q is attributed to losses in the metallic enclosure, support structure, and tuning screws.

Table 16.3 Size and achievable filter *Q* for dielectric resonator filters operating in single modes, dual modes, and triple modes at 1.8 GHz.

	Unloaded *Q* (1/tan δ)	Loaded filter *Q*	Volume (in.^{3})	Volume per electric resonator (in.^{3})
Single mode ($TE_{01\delta}$) $\varepsilon_r = 29$	50,000	35,000	7.8	7.8
Dual mode (HEH) $\varepsilon_r = 29$	50,000	30,000	10.0	5.0
Dual mode (HEH) $\varepsilon_r = 29$	50,000	28,000	11.1	3.7
Single mode ($TE_{01\delta}$) $\varepsilon_r = 45$	24,000	20,000	4.4	4.4
Dual mode (HEH) $\varepsilon_r = 45$	24,000	18,000	5.5	2.8
Triple mode ($3TE_{01\delta}$) $\varepsilon_r = 45$	24,000	17,000	6.3	2.1

16.4 Design Considerations for Dielectric Resonator Filters

16.4.1 Achievable Filter *Q* Value

The *Q* value quoted by the dielectric resonator suppliers is the unloaded *Q* and not the actual *Q* that determines the dielectric resonator filter insertion loss performance. The unloaded *Q* quoted by the suppliers is typically the inverse of the loss tangent ($1/\tan\delta$) of the dielectric material. In addition to the loss tangent, the achievable filter loaded *Q* value is determined by other factors such as size of enclosure, support structure, and tuning screws.

16.4.1.1 Size of Enclosure

Increasing the enclosure size will naturally increase the *Q* value. The size, however, is usually limited by the bandwidth of the filter. The walls of the enclosure, where the irises are located, must be close enough to the dielectric resonators for the irises to provide the necessary coupling. Typically the enclosure has a square shape with a side width that is 1.5–1.7 times the diameter of the dielectric resonator. The conductivity of the materials used for the enclosure would also have an impact on the achievable filter loaded *Q* value.

16.4.1.2 Support Structure

The dielectric resonators are typically mounted inside the cavity using a support structure that is made of dielectric material with a low dielectric constant; for example, Trans-Tech support has ε_r of 4.5 with a $\tan\delta$ of 0.0002. A layer of adhesive material is typically used between the resonator and the dielectric support as shown in Figure 16.17. The combined dielectric resonator–dielectric support is typically mounted inside the cavity using a metallic screw or a plastic screw. The support structure can significantly degrade the *Q* if it is not properly designed.

16.4.1.3 Effect of Tuning Elements

The tolerances on the dielectric constant are typically ±1 from batch to batch and ±0.5 for the same batch. For dielectric resonators with $\varepsilon_r = 29$, a variation in dielectric constant of ±1 translates to a roughly ±30 MHz shift in the resonant frequency. Filter suppliers reduced the impact of the tolerance problem by carrying out in-house measurement on each batch. Then a grinding machine is used to adjust the resonator length to compensate for the tolerances in the dielectric constant. However, a full readjustment cannot be made and screws still must be used to tune the resonant frequencies. The design example given in Section 14.4 also highlights the need to use tuning screws for the dielectric resonators. In addition, the tolerances on the filter housing are usually large enough to affect the coupling values that are provided by the irises.

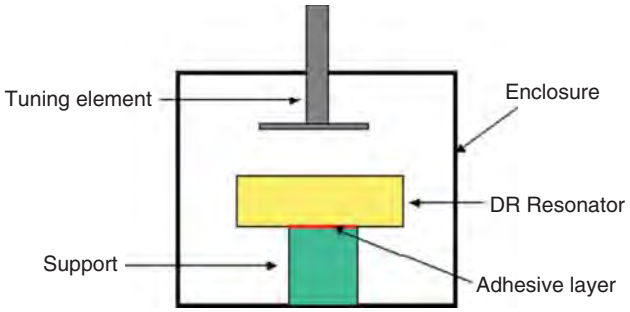


Figure 16.17 The typical support structure of a DR resonator.

The coupling values are usually readjusted by using tuning screws mounted inside the irises. The tuning screws for the resonators and irises would certainly degrade the filter Q.

16.4.2 Spurious Performance of Dielectric Resonator Filters

Compared with other microwave filters (waveguide, coaxial, planar, etc.), dielectric resonator filters have the worst spurious performance. At 1.8 GHz, the spurious-free window is around 400–500 MHz. Dielectric resonator filter suppliers have solved this problem either by employing the mixed-resonator approach [17], where coaxial cavity resonators are mixed with dielectric resonators as shown in Figure 16.18, or by using a coaxial cavity filter connected in cascade with the dielectric resonator filter as shown in Figure 16.19. The coaxial cavity filter has a very wide spurious-free window at least up to the third harmonic, namely, 3.6 GHz at a center frequency of 1.8 GHz. The addition of such coaxial cavities would certainly solve the spurious problem but would increase the overall filter volume by 15–20% and the filter insertion loss by about 10–15%.

Another approach was proposed [23] to improve the spurious performance of dielectric resonator filters operating in dual modes. The experimental results given in [23] demonstrate that reshaping the resonator structure from the traditional cylindrical shape to the modified resonator structure shown in Figure 16.20 improves the spurious performance by more than 40%. As can be seen from the mode charts given in Figure 16.11, the proximity of the HEE mode limits the spurious performance of resonators that operate at HEH dual modes. The electric field of the HEE mode is maximum near the top and bottom surfaces of the resonator, whereas that of HEH mode is almost zeros at the two surfaces. Thus the reduction of the diameter at top and bottom surfaces, as shown in Figure 16.20, should shift the resonant frequency of the

Figure 16.18 Mixing of dielectric resonators with coaxial resonators to improve the filter spurious performance.

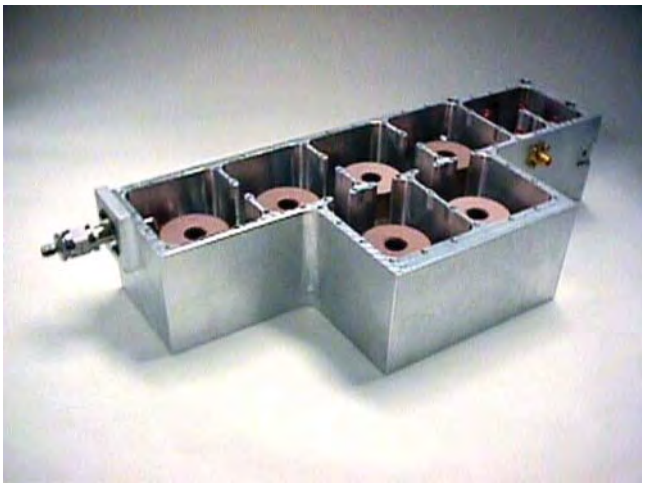


Figure 16.19 A coaxial cavity filter is added to improve the filter spurious performance [20].

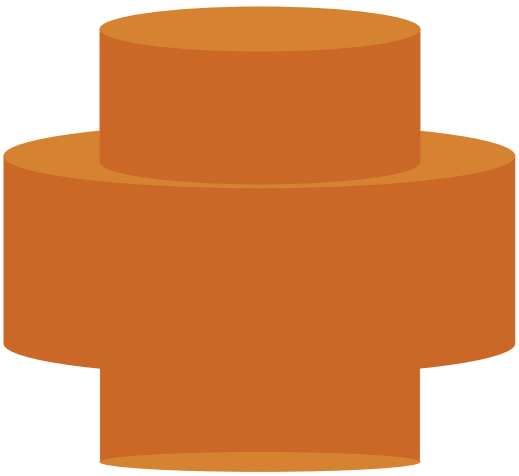


Figure 16.20 Modified resonator configuration to improve the spurious performance of resonators operating in dual HEH modes.

HEE mode up with very little impact on that of the HEH mode. More details on the concept are given in [23].

16.4.3 Temperature Drift

The overall temperature drift of the filter is determined by the temperature coefficient of the resonator, the support structure, and the thermal expansion coefficient of the tuning screws and enclosure. Dielectric resonators are offered with a wide range of temperature coefficients (−6 ppm/°C to +6 ppm/°C) to allow designers to compensate for the combined temperature drift that results from the aforementioned factors. Such combined temperature drift could be positive or negative depending on the type of material used. With the proper choice of temperature coefficient of dielectric resonator materials, the overall temperature drift of the filter can be reduced to 1 ppm/°C. (e.g., over a ΔT of 50°C, the frequency drift at a center frequency of 1.8 GHz would be only 90 kHz).

16.4.4 Power Handling Capability

Dielectric resonator filters operating in the $TE_{01\delta}$ mode are the ideal choice for high-peak-power applications because of their field distribution and low-loss performance. The thermal conductivity of the support and the adhesive material are important factors in designing high-power dielectric resonator filters. The support structure is designed to allow the heat generated inside the dielectric resonator to be easily transferred to the filter housing.

16.5 Other Dielectric Resonator Configurations

Several other novel dielectric resonator configurations have been reported in the literature. A half-cut resonator was proposed in [18]. The resonator is almost half the size of the conventional single-mode resonator, with a small Q degradation. The field distribution of the HEH mode in conventional dual-mode dielectric resonators is given in Figure 16.21a. With the use of a metal plate (electric wall) as an image plate [24], half of the resonator can be made, as shown in Figure 16.21b, to resonate at the same resonance frequency of the original full-size resonator. The use of a perfect magnetic wall, as shown in Figure 16.21c [18], should also yield the same result, leading to a resonator with the same resonance frequency f_0. The perfect magnetic wall could be approximated by an dielectric–air interface (nonperfect magnetic wall), leading to a half-cut resonator with a slightly higher resonant frequency $f_0 + \Delta f$ as shown in Figure 16.21d.

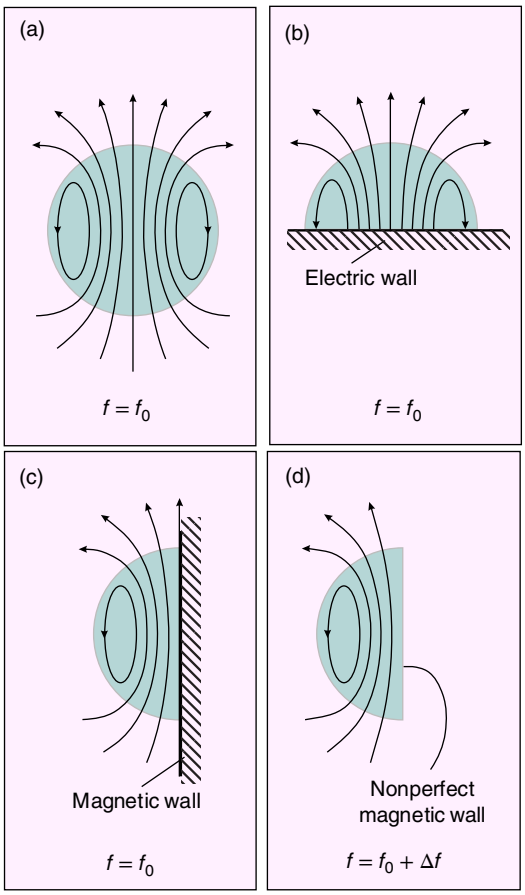


Figure 16.21 Field distribution in dielectric resonators: (a) HE dual modes; (b) half-cut resonator terminated in a perfect electric wall; (c) half-cut resonator terminated in a perfect magnetic wall; and (d) half-cut resonator, with dielectric–air interface equivalent to termination with a nonperfect magnetic wall. *Source*: Mansour *et al*. [18] 2000. Reproduced with the permission of IEEE.

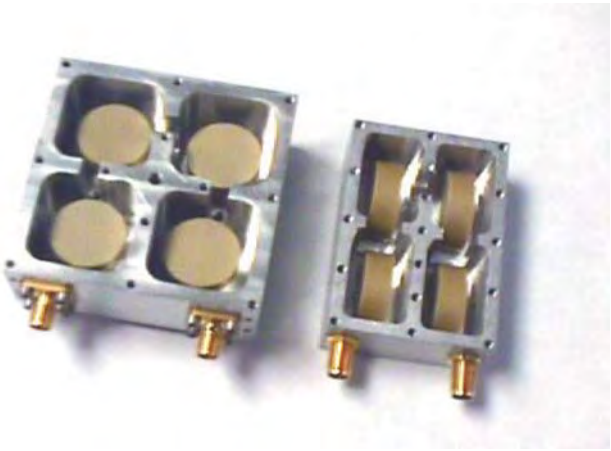


Figure 16.22 Size comparison between traditional single-mode DR resonator filters and single-mode half-cut DR resonator filters. *Source*: Mansour *et al*. [18] 2000. Reproduced with the permission of IEEE.

The half-cut resonator was further reshaped to improve its spurious performance [18]. A comparison between a four-pole filter built using a half-cut resonator and a four-pole filter built using a conventional cylindrical single-mode resonator is given in Figure 16.22. It can be seen that the half-cut resonator filter yields 40% reduction in size. The achievable loaded filter Q is 7000, while the achievable loaded Q of the traditional single-mode $TE_{01\delta}$ resonator is 9000 at 4 GHz [18].

The concept of using two single-mode resonators to construct a quasi-dual-mode resonator that has all the features of traditional dual-mode resonators has been proposed [18]. Two half-cut resonators can be combined in a single cavity to form a quasi-dual-mode resonator. A four-pole filter employing four half-cut resonators placed in two cavities is shown in Figure 16.23. The measured results obtained are also shown in the figure.

Other different filter configurations provide a compromise between the small size advantages of metallic combline resonator filters and the high Q values of dielectric resonator filters. Among these filters are the combline dielectric resonator filters [19] and TM dielectric resonator filters. Figure 16.24 illustrates an eight-pole coaxial DR resonator filter [20]. It consists of six DR resonators mounted directly on the housing. The input and output resonators are just

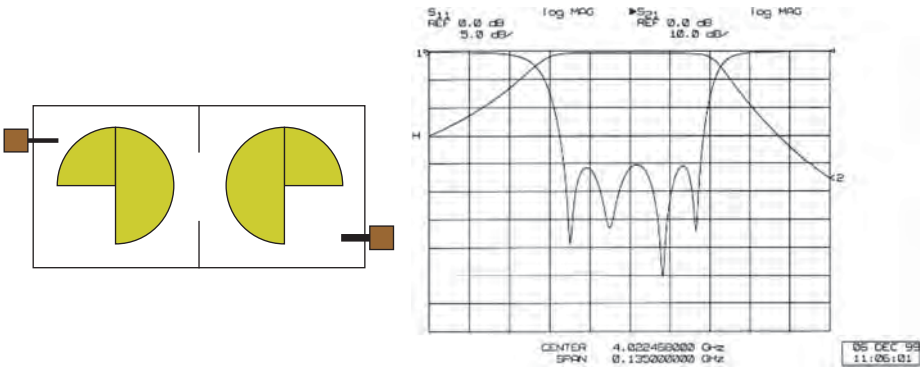


Figure 16.23 A four-pole quasi-dual-mode DR resonator filter constructed by using two half-cut DR resonators in each cavity. *Source*: Mansour *et al*. [18] 2000. Reproduced with the permission of IEEE.

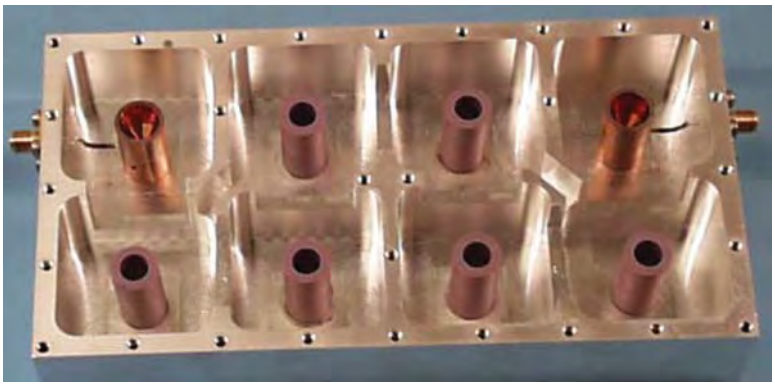


Figure 16.24 A coaxial TM dielectric resonator filter [20].

normal conductor combline resonators used to improve the filter spurious performance. In comparison with standard metallic combline resonator filters, the DR combline resonator filter offers 50% increase in Q while maintaining the same size. The TM dielectric resonator filter [20] is shown in Figure 16.25. It consists of dielectric resonators that are in contact with metal housing on both ends. It offers reasonably good Q for the given volume. It is useful in applications where the emphasis is on size reduction.

More recently, a novel configuration of DR filters has been reported in [25]. The filter structure is amenable to low-cost mass production while providing reasonably high Q value. The dielectric resonators are constructed as one piece from a substrate of high-K materials. All of the resonators are connected to one another by the same high-K dielectric material, resulting in a whole piece that can be efficiently and accurately cut to the desired shape. The substrate is commercially available from Trans-Tech [5] and can be easily cut using low-cost waterjet machining technology. The one-piece dielectric resonator filter is assembled inside the filter housing using a support made of low-K substrate such as Teflon. Thus the assembly of an N-pole filter reduces from the assembly of N individual resonators to the assembly of one dielectric substrate. This simplifies the assembly by avoiding the need for alignment of the resonators, resulting in considerable time and cost savings. Figure 16.26 is a photograph of the high-K ceramic substrate used to construct the filter, whereas Figure 16.27 illustrates the assembled four-pole filter. This type of filter promises to be useful in wireless base station filter applications. More details are given in [25].

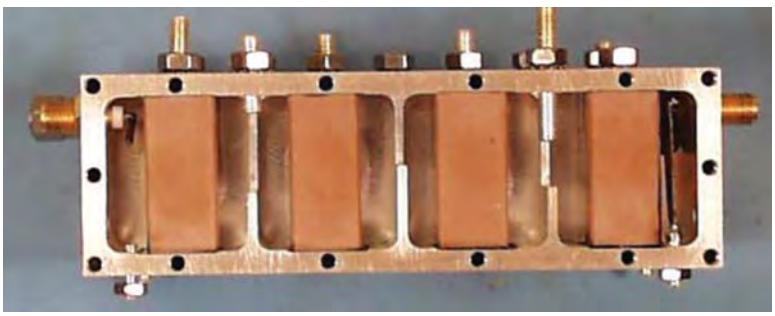


Figure 16.25 A TM dielectric resonator filter.

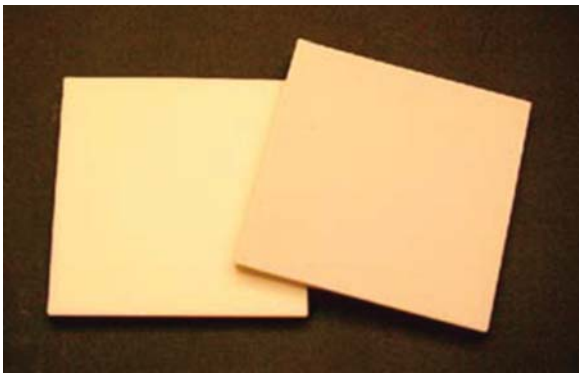


Figure 16.26 Trans-Tech high-*K* ceramic substrates used to construct low-cost DR filters [25].

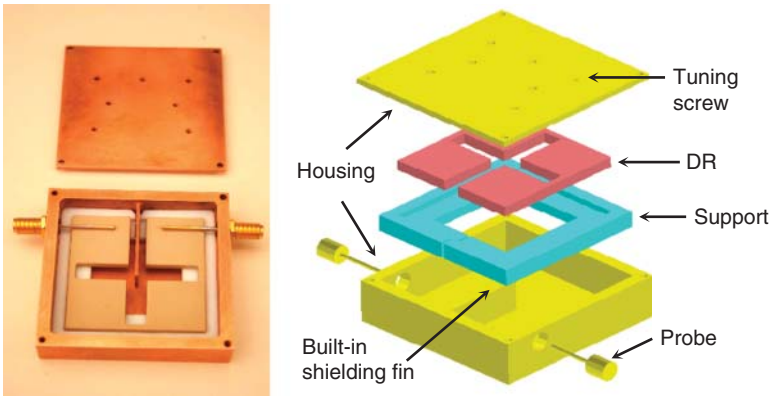


Figure 16.27 An assembled four-pole filter made of high-*K* ceramic substrates [25].

16.6 Cryogenic Dielectric Resonator Filters

Various dielectric materials exhibit a considerable improvement in loss tangent when cooled to cryogenic temperatures. For example, high-purity sapphire ($\varepsilon_r = 9.4$) at X-band frequencies exhibits a $\tan\delta$ of 10^{-5} at 300 K and a $\tan\delta$ of 10^{-7} at 77 K [21]. Other dielectric materials such as magnesium oxides (MgO; $\varepsilon_r = 9.7$), rutile ($\varepsilon_r = 105$), and lanthanum aluminate ($LaAlO_3$; $\varepsilon_r = 23.4$) exhibit a noticeable loss tangent improvement at cryogenic temperatures [26] as compared with room-temperature operation.

Some of the dielectric resonator materials, which were developed for conventional room-temperature filter applications, exhibit low loss at cryogenic temperature while maintaining very small temperature drift. An example is the BaMgTaO (BMT) compound manufactured by Murata [27] with a permittivity of approximately 22. Figure 16.28 illustrates the measured loss tangent versus temperature [28] of the BMT materials. Other dielectric materials manufactured by Murata and other suppliers exhibit a similar loss tangent improvement at cryogenic temperatures. These results clearly show the potential of improving the loss tangent of dielectric materials filters when operating at cryogenic temperatures.

A fully automated measuring system to measure Q and dielectric constant, over the temperature range 10–310 K, has been proposed [26]. Figure 16.29 illustrates the DR cavity used in the measurement. It consists of a dielectric resonator supported on a quartz spacer in a copper cavity. Tuning is achieved by using a top plate that is adjustable by means of a screw. The $TE_{01\delta}$

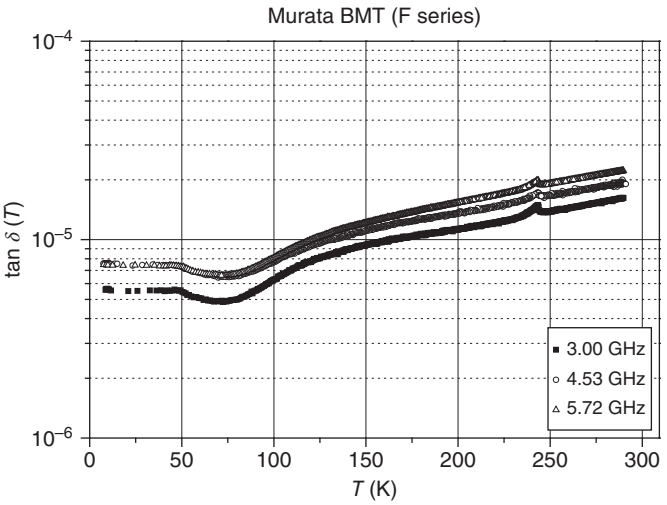
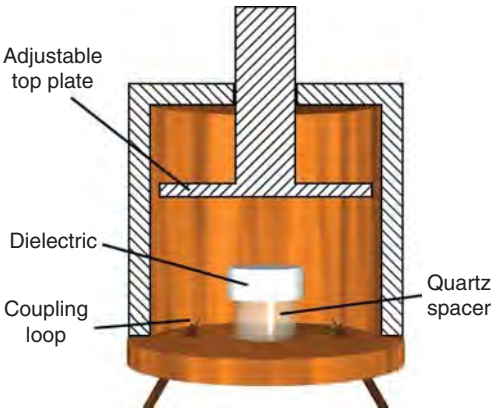


Figure 16.28 The measured loss tangent versus temperature of BMT material.

Figure 16.29 A dielectric resonator cavity for measuring unloaded *Q*. *Source*: Penn and Alford [26]. Reproduced with the permission of South Bank University.



fundamental mode is used for measurement. This mode is used because the Q of the mode is easily related to the dielectric loss tangent. Measurements are made using an RF probe that is weakly coupled to the cavity. The RF probe is weakly coupled so that errors in the insertion loss due to uncertainty of the cable loss at different temperatures do not contribute to an error in the unloaded Q. The assembly is placed on the cold head of a cryocooler, which allows for temperature variation between 20 and 300 K. Table 16.4 illustrates the measured Q of four different ceramic compounds at 20, 77, and 300 K [26].

Table 16.4 Measured Q values for different ceramic compounds [26].

Material	Q_u at 20 K	Q_u at 77 K	Q_u at 300 K
Alumina	704,000	269,000	40,000
BaMgTaO (BMT)	87,000	37,000	21,000
ZrSnTiO	56,000	17,000	5,300
TiO_2	77,000	30,000	4,400

Any conventional dielectric resonator filter can be designed to operate at cryogenic temperatures. The key design considerations are the proper selection of dielectric materials and resonator mechanical structure.

16.7 Hybrid Dielectric/Superconductor Filters

For the conventional dielectric resonator shown in Figure 16.30a, the tangential electric field of the HEE_{11} mode vanishes at the plane $z = 0$. As a consequence, introducing a conducting wall at this plane will slightly perturb the field distribution of this mode. Thus the size of the dielectric resonator filter operating in the HEE mode may be reduced by supporting the resonator directly on a conducting surface (image plate) as shown in Figure 16.30c. With the use of dielectric support that fills the whole cavity as shown in Figure 16.30d, further size reduction can be achieved.

The use of a normal-conducting image plate, however, degrades the Q of the resonator, which in turn makes the approach unfeasible for high-Q applications. However, if the image plate is replaced by a surface made of high-temperature superconductor (HTS) materials, the original unloaded resonator Q will be only slightly affected, allowing the use of this resonator structure to realize compact size filters with a superior loss performance for cryogenic applications. The reduction in size is attributed to the use of the image plate, while the improvement in loss performance is attributed to the use of the HTS image plate as well as to the increase in the unloaded Q of the dielectric resonator as the temperature is lowered from 300 to 77 K.

Several hybrid four-pole/eight-pole DR/HTS filters, which utilize the concept illustrated in Figure 16.30, have been built and tested [28–30]. Figure 16.31 shows a detailed configuration of an eight-pole hybrid DR/HTS filter [30]. The filter consists of four dielectric resonators operating in image-type dual modes. The resonators are supported inside the filter housing using blocks made of low-loss low-dielectric-constant ceramic materials. The HTS material is in the form of small wafers (0.5×0.5 in.^{2} at C band), which are diced from a standard 2 in. wafer. The HTS films are kept in contact with the dielectric resonators using springs and metallic plates bolted to the filter housing. Both the housing and the ceramic blocks have holes to allow the filter to be tuned using standard metallic screws. The insertion loss of this type of filter is determined mainly by the loss tangent of the dielectric resonators and the ceramic blocks as well as the surface resistance of the cavity walls. Both the conductivity of the cavity walls and the loss tangent of the dielectric material improve at lower temperatures, which

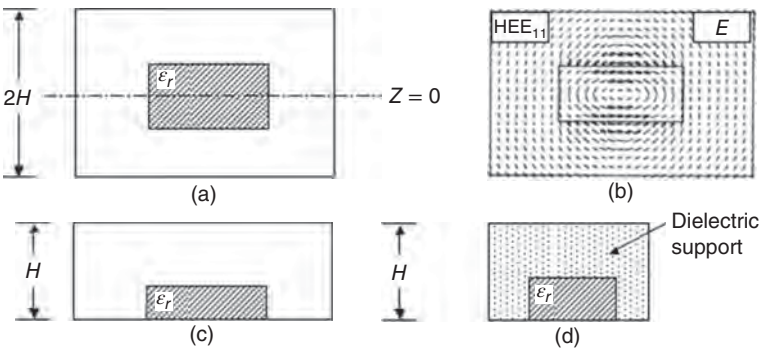


Figure 16.30 Transition from conventional DR cavity to an image DR cavity: (a) conventional DR resonator; (b) field distribution; (c) image DR resonator; and (d) image DR resonator with a dielectric support [28].

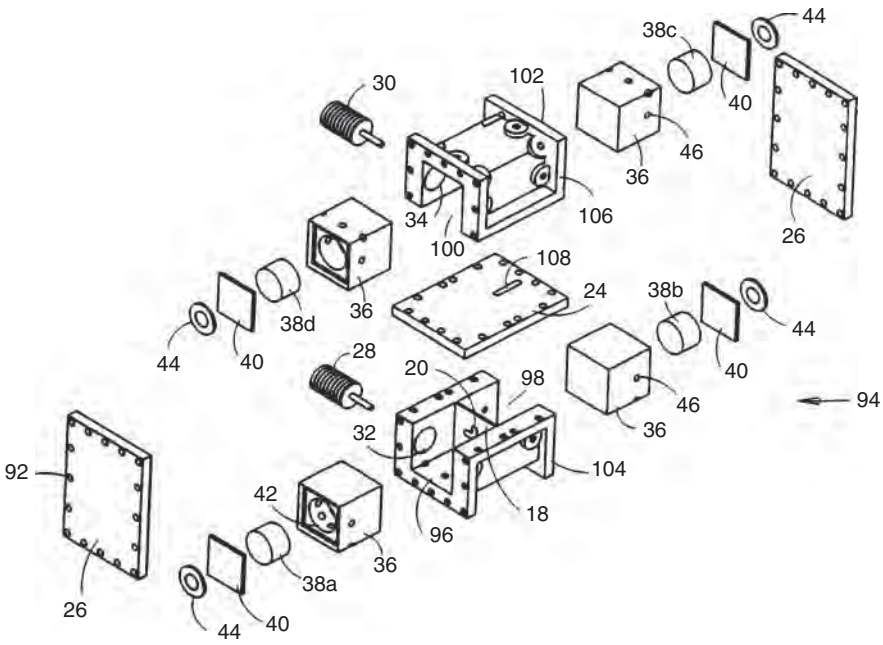


Figure 16.31 A hybrid DR/HTS filter. Components: 38s, dielectric resonators; 40s, HTS wafers diced from a standard 2 in. wafer; 36s, dielectric supports; 44s, spring washers; 26s-24, 102, 104, elements of filter housing; 28s, input/output probes. *Source*: Mansour and Dokas [30] 1996. Reproduced with the permission of US Patent US 05498771.

helps to lower the insertion loss of the filter. A detailed description of this type of filter is given in [30].

The major drawback of this class of filter is its mechanical design complexity. However, hybrid DR/HTS filters have several advantages over conventional HTS thin-film planar filters:

1) No etching is required for the HTS material. Etching may reduce the power handling capability of the HTS material and degrade its surface resistance. The design also eliminates the need to deposit gold contacts, which are required for thin-film filters. Also, there is no need to generate circuit and contact masks.
2) In patterned microstrip thin-film filters, a good portion of the HTS materials is wasted in the etching process, while in hybrid DR/HTS filters the wafer is diced into small pieces, used as "shorting plates." The wafer is used efficiently.
3) The filters can be easily tuned using conventional tuning screws.
4) The filter has a good spurious performance. The image plate helps not only reduce the filter size but also eliminate spurious modes, making it possible to design filters of this type with a spurious-free window of more than 2 GHz for a filter centered at a frequency of 4 GHz.

With metal plates replacing the HTS wafers, the filter can be used in conventional room-temperature applications. The advantages in this case are significant size reduction in comparison with traditional dual-mode filters. The unloaded filter Q is certainly much lower than that of traditional DR filters resonators, but such filters are useful in applications where the emphasis is on size reduction. Figure 16.32 illustrates an eight-pole filter realized using hybrid DR/HTS resonators [28, 30]. The filter volume is one eighth the volume of a conventional dual-mode DR filter.

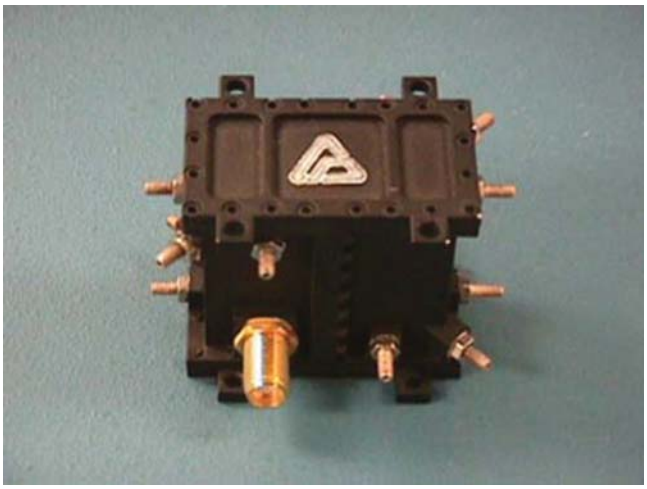


Figure 16.32 A layout of an eight-pole dual-mode hybrid DR/HTS filter at 4 GHz. The filter volume is one eighth the size of a conventional dual-mode DR filter [28, 30]. *Source*: Mansour 1996 [28]. Reproduced with the permission of IEEE, Mansour and Dokas 1996 [30]. U.S. Patent 0549,8771.

16.8 Miniature Dielectric Resonators

16.8.1 A Quadruple-Mode Dielectric Resonator Using a Single Traditional Cylindrical Dielectric Resonator

The mode chart of a dielectric resonator of diameter $D = 17.78$ mm and $\varepsilon_r = 38$ mounted on a low dielectric constant support inside a cubic cavity of size $1'' \times 1'' \times 1''$ is shown in Figure 16.33. The chart gives the change in resonance frequencies of the first three modes (THE, HEH_{11}, and HEE_{11}) for different ratios of D/L. We observe that at D/L ratio of approximately 2, the frequency of the dual-mode HEH_{11} coincides with that of dual-mode HEE_{11}. It should be mentioned that the resonant frequencies of the two modes may coincide at a different ratio other than 2 if other materials are used for the dielectric resonator and the support material. The idea of realizing a quadruple-mode filter is to size the cylindrical dielectric resonator at this particular ratio of D/L, to have four modes at the same frequency [31, 32]. In view of the field distribution of the two dual-modes HEH_{11} and HEE_{11} shown in Figure 16.34, it can be seen that the maximum E-field associated with the four modes does not lie in the same planes. While the four modes coexist at the same frequency, conceivably they can be controlled independently. Thus, a single cylindrical dielectric resonator can be used as a building block for a "quad mode" DR filter, yielding a considerable size reduction.

Figure 16.35 illustrates details of the quad-mode dielectric resonator filter. A cylindrical resonator of diameter D and height L is placed inside a cylindrical cavity of diameter D_c and height L_c. Probes are placed away from the center of the cavity and at 90° from each other, extending from one of the top or bottom walls of the cavity. The length of the probes Hp and the offset distance from the resonator determine the amount of input/output coupling. Several tuning and coupling screws are placed as shown in Figure 16.35. The screws opposite to the two probes are for tuning and for adjusting the coupling between the two components of HEH_{11} and HEE_{11} that align with the probes. The screws that are positioned at 45° are for coupling between the two orthogonal components of the mode pair. The vertical screws are located away from the center of the cavity, while the horizontal screws are exactly in the middle of the resonator. Figure 16.36a

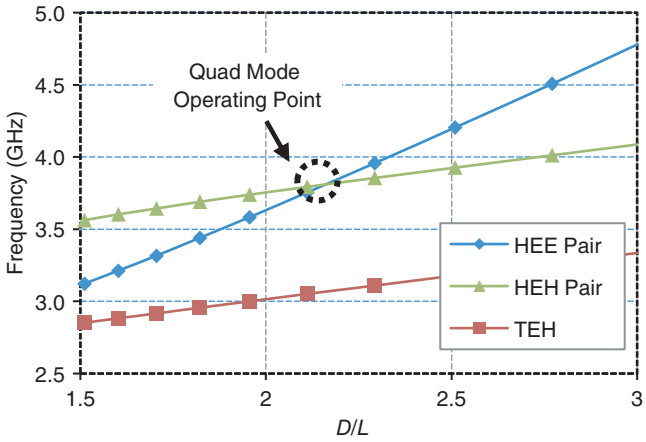
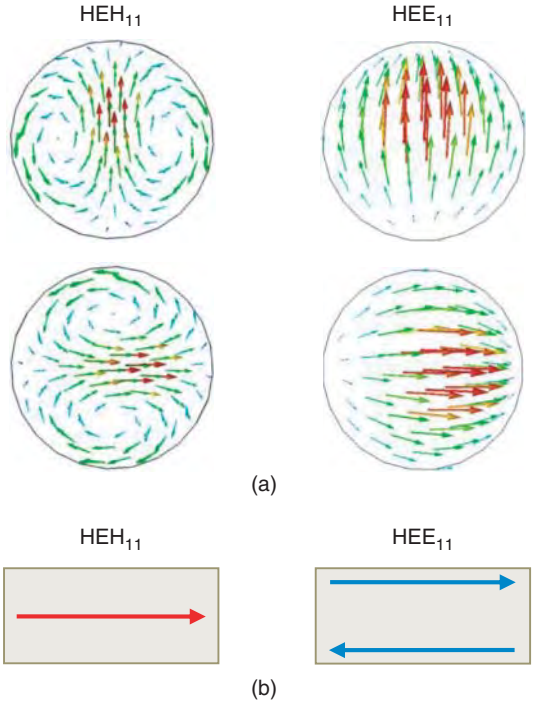


Figure 16.33 Mode chart of a cylindrical dielectric resonator. $D = 17.78$ mm, $\varepsilon_r = 38$, cavity size; 25.4 mm × 25.4 mm × 25.4 mm. The EM simulation results are obtained by varying resonator height L.

Figure 16.34 (a) Top view of E-field distribution of the dual-mode HEH_{11} and HEE_{11}. Each mode has two orthogonal components, which are rotated 90°; (b) side view of E-field of HEH_{11} mode that is concentrated mainly near the middle and that of the HEE_{11} mode that is concentrated mainly at the top and bottom of the resonator.



shows the fabricated quad-mode filter, whereas Figure 16.36b illustrates the measured results obtained [32].

16.8.2 A Dual-Mode Dielectric Resonator Using a Half-Cut Dielectric Resonator

As explained in Section 16.5, in case of the dual-mode HEH_{11} and HEE_{11}, one can cut the cylinder dielectric resonator in half at the plane where the normal electric field to the plane is zero and can still yield almost the same field distribution in the remaining half. We refer to these two modes as $\frac{1}{2}HEH_{11}$ and $\frac{1}{2}HEE_{11}$, which will have the same resonant frequency if we select

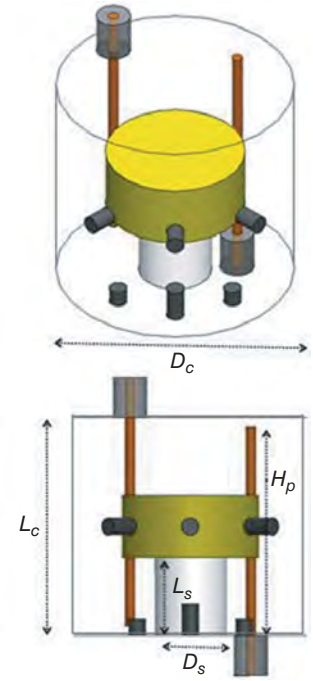
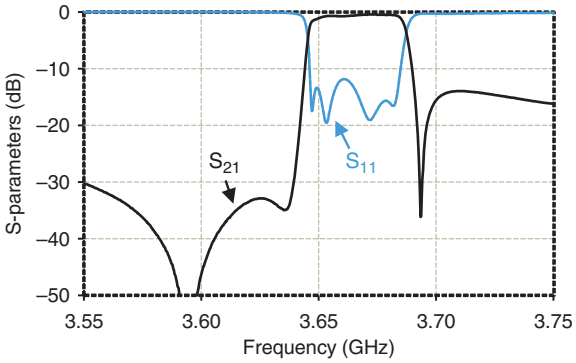


Figure 16.35 Structure of the quad-mode four-pole filter.
Source: Memarian [32] 2009. Reproduced with the permission of IEEE.

(a)



(b)

Figure 16.36 (a) A four-pole filter realized using a quad-mode DR resonator, (b) the measured results. *Source*: Memarian [32] 2009. Reproduced with the permission of IEEE.

D/L close to 2 as explained above in the case of the quad-mode design. Figure 16.37 illustrates the field distribution of these two modes. It can be seen that the two modes have field components that can be controlled independently. The idea of realizing dual-mode filters here is to use $\frac{1}{2}HEH_{11}$ and $\frac{1}{2}HEE_{11}$ in a half-cut dielectric resonator as a dual-mode resonator [31, 32]. Figure 16.38 shows the configuration for the tuning and coupling screws within a single cavity to couple the two modes: $\frac{1}{2}HEH_{11}$ and $\frac{1}{2}HEE_{11}$. It also shows how coupling between these modes in adjacent cavities is achieved through the use of irises. Figure 16.39 shows a four-pole filter realized using half-cut dual-mode resonators, whereas Figure 16.40 illustrates its measured results. The spurious of the half-cut dual-mode resonator can be improved by including slots in the resonator as shown in [32]. More details on the quad-mode and dual-mode designs of the type presented in this section can be found in [31–33].

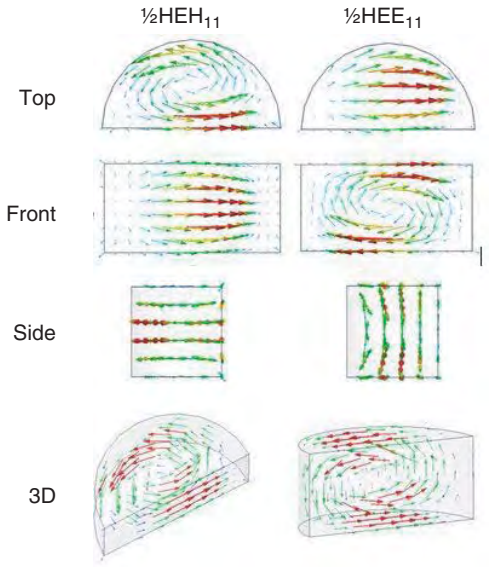


Figure 16.37 Views of the E-field distribution of $\frac{1}{2}HEH_{11}$ and $\frac{1}{2}HEE_{11}$ modes of the half-cylinder dielectric resonator. The modes are orthogonal.

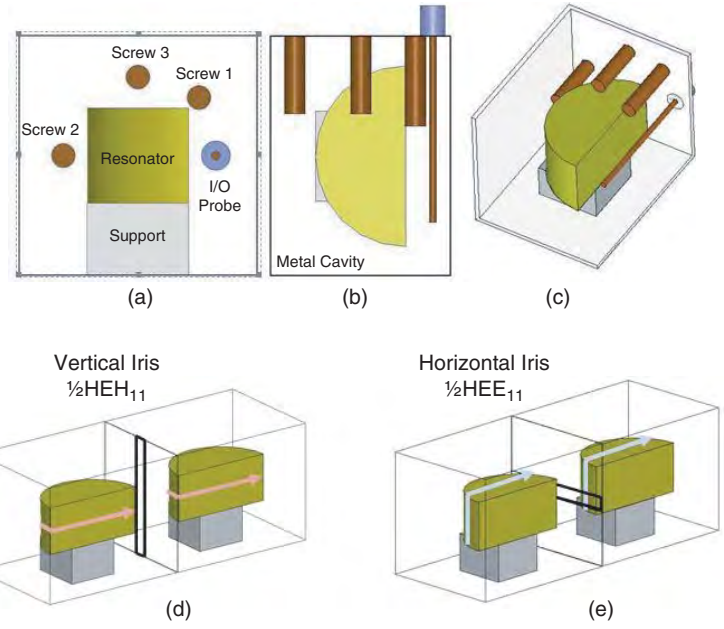


Figure 16.38 (a) Front view, (b) top view, and (c) 3D view of the dual-mode cavity with coupling and tuning mechanisms. Screw 1 for coupling the two modes, Screw 2 for tuning the $\frac{1}{2}HEH_{11}$ mode, and Screw 3 for tuning the $\frac{1}{2}HEE_{11}$ mode. The *I/O* probe placement as shown only couples mostly to the $\frac{1}{2}HEH_{11}$ mode, (d) vertical iris to couple the $\frac{1}{2}HEH_{11}$ modes in adjacent cavities, and (e) horizontal iris to couple $\frac{1}{2}HEE_{11}$ modes in adjacent cavities. *Source*: Memarian [32] 2009. Reproduced with the permission of IEEE.

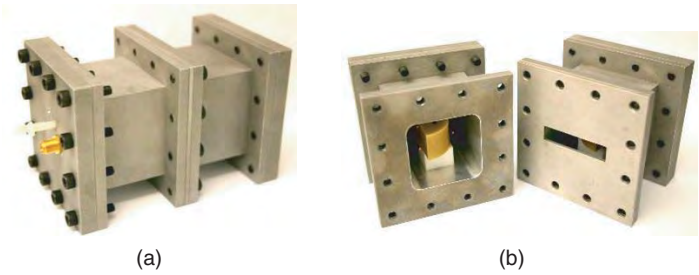


Figure 16.39 (a) Fabricated four-pole filter using half-cut dual-mode resonators with the two cavities attached, (b) with the two cavities separated to illustrate half-cut resonator mounting and iris coupling. *Source*: Memarian [32] 2009. Reproduced with the permission of IEEE.

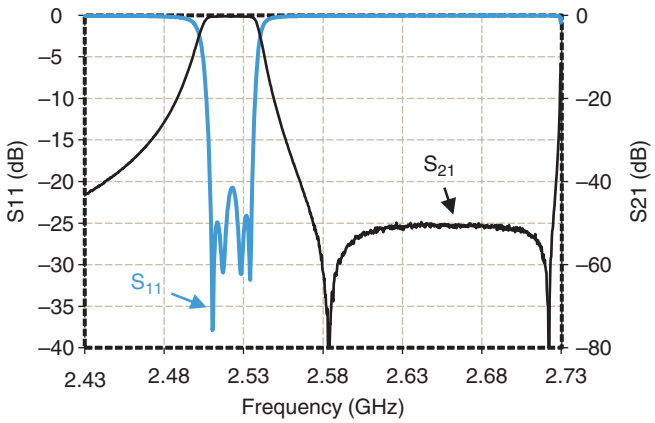


Figure 16.40 The measured results of the four-pole half-cut dual-mode resonator shown in Figure 16.39. *Source*: Memarian [32] 2009. Reproduced with the permission of IEEE.

Summary

Dielectric resonator filters are widely employed in wireless and satellite applications. Continuing advances in the quality of dielectric materials is a good indication of the growing application of this technology. This chapter has been devoted to the design of dielectric resonator filters in a variety of configurations. Mode charts along with plots illustrating the field distribution of the first four modes in dielectric resonators are included. Commercial software packages such as HFSS and CST Microwave Studio can be readily utilized to calculate the resonant frequency, field distribution, and resonator Q of dielectric resonators having any arbitrary shape.

Major design considerations of dielectric resonator filters illustrating two different approaches for improving their spurious performance have been described. Several unconventional designs for dielectric resonator filters are included. These designs provide a tradeoff between size reduction and achievable Q value. This tradeoff bridges the gap in terms of size and Q value between conventional dielectric resonator filters and conventional combine coaxial filters. Design configuration of a dielectric resonator filter made of high-K ceramic substrates is also included. The structure is attractive since the fabrication, assembly, and integration of the dielectric resonators are much simpler than that of conventional DR filters. The structure combines the low-cost feature of the combine coaxial filters and the high-Q

feature of dielectric resonator filters. It promises to be useful in mass production of filters for wireless base station applications.

The chapter also addressed the improvement in the Q value of dielectric resonator filters when cooled to cryogenic temperatures. A hybrid dielectric/superconductor filter that offers a significant size reduction compared with conventional design is presented. The structure uses the image concept and can be employed for room-temperature applications by replacing the image superconductor plates with normal metallic plates.

References

1 Richtmyer, R.D. (1939) Dielectric resonators. *Journal of Applied Physics*, **10**, 391–398.
2 Cohn, S.B. and Targow, E.N. (1965) Investigation of Microwave Dielectric Resonator Filters. Internal report. Rantec Div., Emerson Electric Co.
3 Cohn, S.B. (1967) Microwave bandpass filters containing high-Q dielectric resonators. *IEEE Transactions on Microwave Theory and Techniques*, **16**, 218–227.
4 Harrison, W.H. (1968) Microwave bandpass filters containing high-Q dielectric resonators. *IEEE Transactions on Microwave Theory and Techniques*, **16**, 218–227.
5 Trans-Tech, http://www.trans-techinc.com.
6 Zaki, K.A. and Chunming, C. (1986) New results in dielectric-loaded resonators. *IEEE Transactions on Microwave Theory and Techniques*, **34** (7), 815–824.
7 Xiao-Peng, L. and Zaki, K.A. (1993) Modeling of cylindrical dielectric resonators in rectangular waveguides and cavities. *IEEE Transactions on Microwave Theory and Techniques*, **41** (12), 2174–2181.
8 Ansoft HFSS, http://www.ansoft.com.
9 Kajfez, D. and Guillon, P. (1986) *Dielectric Resonators*, Artech House, Norwood, MA.
10 CST, Microwave Studio, www.cst.com.
11 Zaki, K.W. University of Maryland, College Park, MD, USA.
12 Pozar, D. (1998) *Microwave Engineering*, 2nd edn, Wiley, New York.
13 Salefi, J. (2004) Design Engineering Fourth-Year Design Project Report, University of Waterloo.
14 Fiedziuzko, J. *et al.* (2002) Dielectric materials, devices and circuits. *IEEE Transactions on Microwave Theory and Techniques*, **50**, 706–720.
15 Kudsia, C., Cameron, R., and Tang, W.C. (1992) Innovations in microwave filters and multiplexing networks for communications satellite systems. *IEEE Transactions on Microwave Theory and Techniques*, **40** (6), 1133–1149.
16 Noshikawa, T. (1998) Comparative filter technologies for communications systems. Proceedings of 1998 MTT-S Workshop Comparative Filter Technologies for Communications Systems, Baltimore.
17 Ji-Fuh, L. and William, B. (1998) High-Q TE_{01} mode DR filters for PCS wireless base stations. *IEEE Transactions on Microwave Theory and Techniques*, **46**, 2493–2500.
18 Mansour, R.R. *et al.* (2000) Quasi dual-mode resonators. *IEEE Transactions on Microwave Theory and Techniques*, **48** (12), 2476–2482.
19 Wang, C., Zaki, C.K.A., Atia, A.E., and Dolan, T. (1998) Dielectric combline resonators and filters. *IEEE Transactions on Microwave Theory and Techniques*, **3** (7–12), 1315–1318.
20 Mansour, R.R. (2004) Filter technologies for wireless base station filters. *IEEE Microwave Magazine*, **5** (1), 68–74.

21 Walker, V. and Hunter, I.C. (2002) Design of triple mode TE_{01} resonator transmission filters. *IEEE Microwave and Wireless Components Letters*, **12**, 215–217.

22 Hattori, J. *et al*. (2003) 2 GHz band quadruple mode dielectric resonator filter for cellular base stations. IEEE-MTT IMS-2003 Digest, Philadelphia.

23 Mansour, R.R. (1993) Dual-mode dielectric resonator filters with improved spurious performance. IEEE MTT International of Microwave Symposium Digest, pp. 439–442.

24 Nishikawa, T. *et al.* (1987) Dielectric high-power bandpass filter using quarter-cut TE_{01} image resonator for cellular base stations. *IEEE Transactions on Microwave Theory and Techniques*, **35**, 1150–1154.

25 Zhang, R. and Mansour, R.R. (2006) Dielectric resonator filters fabricated from high-K ceramic substrates. IEEE-IMS

26 Penn, S. and Alford, N. High Dielectric Constant, Low Loss Dielectric Resonator Materials. EPSRC final report GR/K70649, EEIE, South Bank University, London, 2000.

27 Murata. http://www.murata.com.

28 Mansour, R.R. *et al*. (1996) Design considerations of superconductive input multiplexers for satellite applications. *IEEE Transactions on Microwave Theory and Techniques*, **44**, 1213–1228.

29 Mansour, R.R. (2002) Microwave superconductivity. *IEEE Transactions on Microwave Theory and Techniques* (Special issue—50th Anniversary of *IEEE MTT Trans*, **52**, 750–759.

30 Mansour, R.R. and Dokas, V. (1996) Miniaturized dielectric resonator filters and method of operation thereof at cryogenic temperatures. US Patent 0549,8771.

31 Memarian, M. and Mansour, R.R. (2012) Methods of operation and construction of dual-mode filters, dual-band filters and diplexers/multiplexers using half-cut and full dielectric resonators. US Patent 8,111,115, issued Feb. 2012.

32 Memarian, M. and Mansour, R.R. (2009) Quad-mode and dual-mode dielectric resonator filters. *IEEE Transactions on Microwave Theory and Techniques*, **57** (12), 3418–3426.

33 Memarian, M. (2009) Compact dielectric resonator filters and multiplexers. Master thesis. University of Waterloo, Waterloo, Ontario, Canada.

17

Allpass Phase and Group Delay Equalizer Networks

The ideal characteristics for a filter network call for a brick-wall amplitude response and a linear phase across the band of interest. Such a response is not feasible, even with idealized components. This leads into tradeoffs between amplitude and phase response of filters as a function of its order, response function and design complexity.

Most filter designs are based on the minimum phase networks since they provide the highest rate of attenuation. The filter response is optimized on the basis of amplitude response, providing a specified level of attenuation outside its passband. The phase response (and group delay) of such a filter in the passband is uniquely related to the filter's amplitude response. This implies that once the amplitude response has been optimized, the resulting phase and the group delay response must be accepted as is. For many applications, this tradeoff is acceptable.

However, it is possible to improve the phase and group delay response of a filter without sacrificing the amplitude performance. This can be accomplished by either incorporating internal equalization (linear phase filters) or using allpass external phase/group delay equalizer networks. Either method adds complexity and increases the passband loss for the overall filter network. In this chapter, we deal with the external allpass phase and group delay equalizers. It also includes a section in which we discuss the practical tradeoffs between the linear phase and externally equalized filter networks.

17.1 Characteristics of Allpass Networks

A *network* is defined as an allpass network if it transmits all the incident power and reflects none. Such a network implies a transmission coefficient of unity and the reflection coefficient of zero. The pole–zero configuration of the network is determined by analyzing the scattering matrix of a two-port network, given by

$$[S] = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} \tag{17.1}$$

The network is defined as an allpass network if

$$S_{11} = S_{22} = 0 \quad \text{and} \quad |S_{12}| = 1 \ \text{ evaluated at } \ s = j\omega \tag{17.2}$$

The transfer function S_{12} of a lossless two-port network is represented by

$$S_{12}(s) = \frac{N(s)}{E(s)} \tag{17.3}$$

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

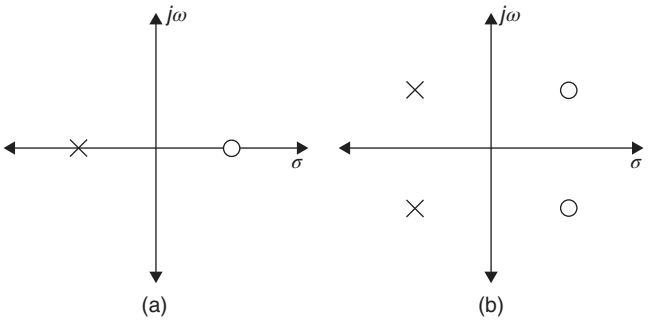


Figure 17.1 Pole–zero location of lumped-element allpass networks: (a) first-order network; (b) second-order network.

where $E(s)$ is a strict Hurwitz polynomial and $N(s)$ is a polynomial with real coefficients whose order does not differ from that of $E(s)$ by more than 1. For an equalizer network, polynomials $N(s)$ and $E(s)$ must be equal in magnitude by virtue of equation (17.2). Furthermore, for an allpass network to be able to manipulate the phase response, the two polynomials cannot have an identical phase response, because it leads to the trivial solution representing just a matched transmission network. The only other possibility within the constraint of equal magnitude is for the two polynomials to have an equal and opposite phase. Thus, a two-port lossless allpass network is represented by

$$S_{12}(s) = \frac{E(-s)}{E(s)} \tag{17.4}$$

The Hurwitz polynomial $E(s)$ is formed by three factors:

$$\begin{aligned} &(s+a), && a \text{ real and positive} \\ &(s^2+b^2), && b \text{ real} \\ &(s^2+2cs+c^2+d^2), && c \text{ real and positive; } d \text{ real} \end{aligned} \tag{17.5}$$

It is readily seen that for the allpass functions, the allowable locations for the zeros of Hurwitz polynomial are given by

$$(s+a) \quad \text{and} \quad (s^2+2cs+c^2+d^2) \tag{17.6}$$

The zeros pertaining to (s^2+b^2) cancel out and are thus excluded from the allpass functions. This leads to the conclusion that allpass networks consist of a cascade of networks that can be represented by a symmetric pair of pole–zero locations on the real axis or a complex quad, as depicted in Figure 17.1. Therefore, the general form of an allpass function consisting of k first-order sections and m second-order sections is given by

$$E(s) = \prod_{a=1}^{k}(s+\sigma_a)\prod_{b=1}^{m}(s^2+2s\sigma_b+|s_b|^2), \quad \text{where } s_b = \sigma_b + j\omega_b \tag{17.7}$$

and σ_a, σ_b, ω_b are positive and real numbers.

The magnitude and phase response for such a network is given by

$$\begin{aligned} &\text{Magnitude:} \quad \alpha = |S_{12}(s)| = 1 \\ &\text{Phase:} \quad \beta = -\sum_{a=1}^{k} 2\tan^{-1}\frac{\omega}{\sigma_a} - \sum_{b=1}^{m} 2\tan^{-1}\frac{2\sigma_b\omega}{|s_b|^2-\omega^2} \end{aligned} \tag{17.8}$$

The group delay response is given by

$$\tau = -\frac{d\beta}{d\omega} = 2\sum_{a=1}^{k}\frac{\sigma_a}{\sigma_a^2+\omega^2} + 2\sum_{b=1}^{m}\frac{2\sigma_b(|s_b|^2+\omega^2)}{\omega^4+2\omega^2(\sigma_b^2-\omega_b^2)+|s_b|^4} \tag{17.9}$$

17.2 Lumped-Element Allpass Networks

The lumped-element allpass networks are thoroughly covered by many authors [1, 2]. Here, we present a summary of such networks as a background to the design of microwave allpass networks.

Typically, lumped-element allpass networks are realized by symmetric lattice structures. The general form of a lattice structure is shown in Figure 17.2a, and its equivalent bridge network is shown in Figure 17.2b.

The two alternative representations are identical. However, the bridge representation lends itself easily for the analysis of a single lattice section. On the other hand, if a number of lattice networks are connected in tandem, then the general series-type representation is a more practical description. The open-circuit impedance of the lattice network, as shown in Figure 17.2b, is given by

$$z_{11} = z_{22} = \frac{1}{2}(Z_a + Z_b) \tag{17.10}$$

This simply represents the impedance of two identical branches in parallel. The open-circuit transfer impedance z_{12} is determined as follows.

Let V_1 and V_2 be the voltages across terminals 1 and 2, and let I_1 and I_2 be the corresponding currents. The voltage of terminal 2 with respect to 1 is

$$V_2 = \frac{Z_b V_1}{Z_a + Z_b} - \frac{Z_a V_1}{Z_a + Z_b} \tag{17.11}$$

The voltage V_1 is related to current I_1 with terminals $2-2'$ open according to equation (17.10):

$$V_1 = I_1\left[\frac{1}{2}(Z_a + Z_b)\right] \tag{17.12}$$

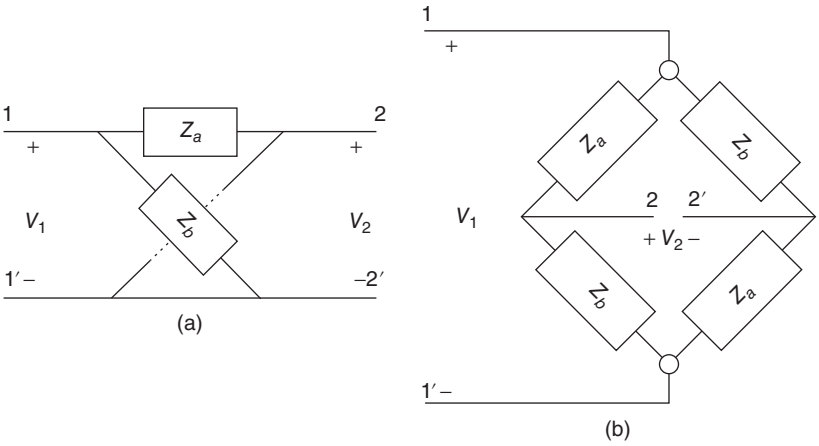


Figure 17.2 Lumped-element symmetric lattice networks: (a) standard representation of symmetric lattice; (b) bridge form of symmetric lattice.

From equations (17.11) and (17.12), we obtain

$$\frac{V_2}{I_1} = z_{12} = \frac{1}{2}(Z_b - Z_a) \tag{17.13}$$

As a result, the impedances Z_a and Z_b are determined to be

$$\begin{aligned} Z_a &= z_{11} - z_{12} \\ Z_b &= z_{11} + z_{12} \end{aligned} \tag{17.14}$$

In a similar manner

$$\begin{aligned} y_{11} &= y_{22} = \frac{1}{2}(Y_a + Y_b) \\ y_{12} &= \frac{1}{2}(Y_b - Y_a) \end{aligned} \tag{17.15}$$

where

$$Y_a = \frac{1}{Z_a}, \quad Y_b = \frac{1}{Z_b}$$

Many different methods are used to synthesize open-circuit lattice networks according to the relationships in equations (17.14) and (17.15). However, our interest lies in networks that can be used to equalize the phase and group delay of filters that are terminated in a matched load. Consequently, we next examine lattice networks that are terminated in a resistance.

17.2.1 Resistively Terminated Symmetric Lattice Networks

Let us start with a general network terminated in a resistive load as shown in Figure 17.3. The equations for the network, in terms of its open-circuit parameters, are given by

$$\begin{aligned} V_1 &= z_{11}I_1 + z_{12}I_2 \\ V_2 &= z_{12}I_1 + z_{22}I_2 \end{aligned} \tag{17.16}$$

For an allpass network, the driving point impedance must equal the terminating impedance of the filter network; therefore

$$Z(s) = \frac{V_1}{I_1} = R \tag{17.17}$$

By substituting $V_2 = -RI_2$ into equation (17.16), it is possible to determine $Z(s)$ in terms of the open-circuit parameters. Assuming a symmetric lattice (i.e., $z_{11} = z_{22}$) and a network normalized with $R = 1$, equation (17.17) is evaluated as

$$Z(s) = \frac{z_{11}^2 - z_{12}^2 + z_{11}}{z_{11} + 1} = 1 \quad \text{or} \quad z_{11}^2 = 1 + z_{12}^2 \tag{17.18}$$

Combining equation (17.18) with equations (17.10) and (17.13), we obtain

$$Z_a Z_b = 1 \tag{17.19}$$

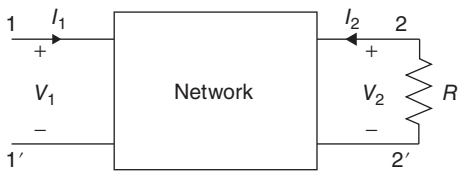


Figure 17.3 General network terminated in the resistor, *R*.

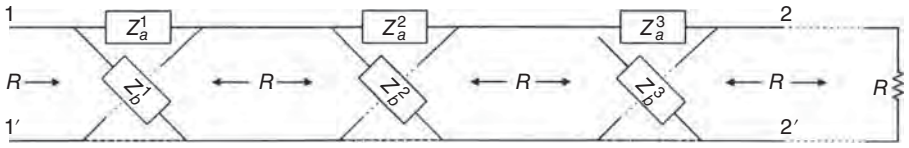


Figure 17.4 Cascade connection of constant-resistance lumped-element lattice networks. (*Source*: Van Valkenburg 1960 [2]. Reproduced with the permission of John Wiley & Sons.)

If the lattice is terminated in R (Ω), then we have the more general form $Z_a Z_b = R^2$. Thus, Z_a and Z_b need to be reciprocal impedances for the input impedance to be a constant value. For the network terminated in R, the transfer impedance $Z_{12}(= V_2/I_1)$ is related to Z_a and Z_b as follows [2]:

$$Z_{12} = R\left(\frac{R - Z_a}{R + Z_a}\right) \tag{17.20}$$

For the network normalized with $R = 1$, we obtain

$$Z_{12} = \left(\frac{1 - Z_a}{1 + Z_a}\right), \quad Z_a = \frac{1 - Z_{12}}{1 + Z_{12}}, \quad Z_b = \frac{1}{Z_a} \tag{17.21}$$

This rather important relationship makes it possible to find Z_a and Z_b from a given Z_{12} specification. As always, Z_{12} must satisfy the conditions of physical realizability, which means that Z_{12} has no poles in the right half plane or on the imaginary axis and that $|Z_{12}(j\omega)| \leq 1$.

Let us now consider lattice networks composed entirely of lossless inductors and capacitors except for the termination. For such a network, evaluated at real frequencies, we have

$$|Z_{12}(j\omega)| = \left|\frac{1 - Z_a(j\omega)}{1 + Z_a(j\omega)}\right| = \left|\frac{1 - jX_a}{1 + jX_a}\right| = 1 \tag{17.22}$$

Thus, a symmetric lattice network, composed of L and C and terminated in a matched resistor, represents an allpass network and satisfies the condition $|Z_{12}(j\omega)| = 1$ for all ω. The condition $|Z_{12}(j\omega)| < 1$ implies a loss in the lattice structure.

There is one more important property of the constant-resistance lattice network. Given a single lattice terminated in R (Ω), its input impedance is also R (Ω); thus, this terminated lattice can serve as a proper termination of R (Ω) for another lattice. This argument can be repeated any number of times for the cascade connection of lattices in Figure 17.4. For lattice networks with the same constant-R specification, the input impedance of the cascade connection of any number of lattices is still R (Ω).

17.2.2 Network Realizations

An allpass function can be represented as a cascade of the first- and second- order allpass functions. Furthermore, an allpass function can be realized as a cascade of individual lattice networks. Thus, all we need to do is to relate the transfer function of a first- and a second-order allpass function to the impedances of a single lattice network, shown in Figure 17.5.

The transfer function $H(s)$ is defined by the output to input voltage ratio V_2/V_1. For the allpass network, $V_1 = I_1 R$, and therefore, the transfer function is given by

$$H(s) = \frac{Z_{12}}{R} = \frac{R - Z_a}{R + Z_a} \tag{17.23}$$

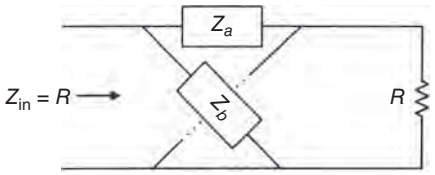


Figure 17.5 Constant-resistance lumped-element lattice network.

17.2.2.1 First-Order Lattice Section

For the first-order section, $Z_a = sL$, and substituting this value in equation (17.23) yields

$$H(s) = \frac{R/L - s}{R/L + s}$$

Using equation (17.7), we obtain

$$L = \frac{R}{\sigma_a}$$

$$Z_b = \frac{R^2}{Z_a} = \frac{R^2}{sL} = \frac{1}{sC} \tag{17.24}$$

The value of the capacitor is

$$C = \frac{L}{R^2} = \frac{1}{\sigma_a R} \tag{17.25}$$

Thus, the reactive elements of the lattice are determined in terms of the pole–zero location of the first-order allpass function. Note that L and C resonate at σ_a.

17.2.2.2 Second-Order Lattice Section

For the second-order section, Z_a is the impedance of the antiresonant circuit:

$$Z_a = \frac{Ls}{LCs^2 + 1} \tag{17.26}$$

Substitution into equation (17.23) yields

$$H(s) = \frac{s^2 - (1/RC)s + 1/LC}{s^2 + (1/RC)s + 1/LC} \tag{17.27}$$

Comparing this with equation (17.7), we have

$$\frac{1}{RC} = 2\sigma_b$$

$$\frac{1}{LC} = |s_b|^2 = \sigma_b^2 + \omega_b^2 \tag{17.28}$$

The values L and C are therefore given by

$$L = \frac{2\sigma_b R}{|s_b|}$$

$$C = \frac{1}{s\sigma_b R|s_b|} \tag{17.29}$$

Since $Z_a Z_b = R^2$ and $Z_a = sL_1/(s^2 L_1 C_1 + 1)$, the series resonant circuit Z_b is given by

$$\begin{aligned} Z_b &= \frac{R^2(s^2 L_1 C_1 + 1)}{sL_1} \\ &= sR^2 C_1 + \frac{1}{s(L_1/R^2)} \\ &= sL_2 + \frac{1}{sC_2} \end{aligned} \tag{17.30}$$

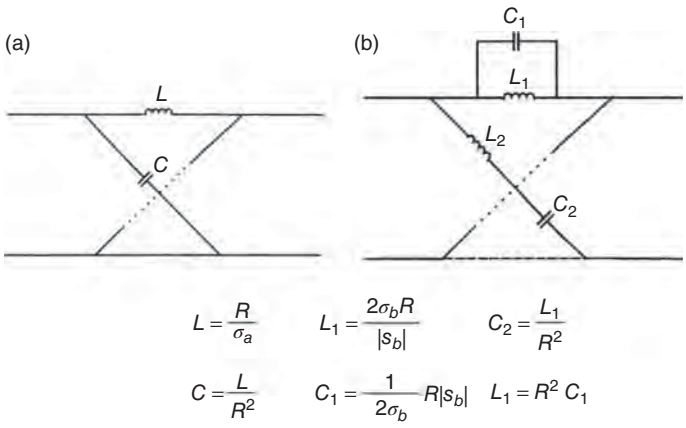


Figure 17.6 Lumped-element lattice forms of (a) first-order and (b) second-order allpass network.

This is the impendence of a series circuit, composed of L_2 andC_2, where

$$L_2 = R^2 C_1$$
$$C_2 = \frac{L_1}{R^2} \tag{17.31}$$

The design equations for the first- and second-order lattice networks are summarized in Figure 17.6.

17.3 Microwave Allpass Networks

An ideal allpass network is characterized by a constant amplitude response and a variable phase or group delay response over an infinite bandwidth. This implies that the amplitude and phase responses are completely independent of each other. Furthermore, owing to the network's unique amplitude characteristics and infinite bandwidth, there is no need to develop lowpass lumped-element prototype models as we did for the filter networks. Instead, the design can proceed directly at the required frequency and bandwidth over which equalization is required. Consequently, microwave allpass networks are dealt with directly in terms of the distributed frequency variable for their analysis and the synthesis.

Scanlon and Rhodes [3] show that any TEM allpass commensurate network must have a transfer function that has the form

$$S_{12} = \frac{E(-t)}{E(t)} \tag{17.32}$$

The distributed frequency variable t is given by

$$t = \tanh \gamma \ell = \Sigma + j\Omega \tag{17.33}$$

where γ represents the complex propagation constant and ℓ is the commensurate length of the network. This result is very similar to that of the transfer function of lumped-element allpass networks, where the real frequency variable is replaced by the distributed frequency variable t. For the lossless case, $\Sigma = 0$ and

$$t = j\Omega = j \tan \theta \tag{17.34}$$

where $\theta = (\omega \ell / v) = (2\pi/\lambda)\ell$, θ is electrical line length, ℓ is shortest commensurate-length line, v is velocity of propagation, ω is real frequency, and λ is wavelength of the signal. If n is the effective number of unit elements in the network, the expression for the transfer function is [3]

$$S_{12}(t) = \left(\frac{1-t}{1+t}\right)^{n/2} \frac{E(-t)}{E(t)} \tag{17.35}$$

The appearance of factor $[(1-t)/(1+t)]^{n/2}$ in the transfer function is equivalent to adding a delay of a transmission-line $n/4$ radians long.

The phase β of this transfer function is

$$\beta = -j\ell n S_{12}(t) \tag{17.36}$$

The group delay τ is given by

$$\tau = -\left.\frac{d\beta}{d\omega}\right|_{t=j\Omega} = -\frac{1}{2\pi}\frac{1}{f_0}\left.\frac{d\beta}{dt}\frac{dt}{dF}\right|_{t=j\Omega} \tag{17.37}$$

Here, $F = f/f_0$ is the normalized frequency variable and f_0 is the center frequency. In the physical realization of allpass networks, the commensurate-line length is usually chosen to be a quarter-length wavelength with respect to the center frequency. As a consequence, the variable t is expressed as

$$\begin{aligned} t &= j \tan \frac{2\pi}{\lambda}\frac{\lambda_{\text{cf}}}{4} \\ &= j \tan \frac{\pi}{2}\frac{\lambda_{\text{cf}}}{\lambda} \end{aligned} \tag{17.38}$$

where λ_{cf} is the wavelength corresponding to the center frequency. For TEM structures, we obtain

$$\lambda_{\text{cf}} = \lambda_0 = \frac{c}{f_0} \tag{17.39}$$

where λ_0 is the free-space wavelength at the band center frequency of f_0. Therefore

$$\begin{aligned} t &= j \tan \frac{\pi}{2}\frac{\lambda_0}{\lambda} \\ &= j \tan \frac{\pi}{2}\frac{f}{f_0} \end{aligned} \tag{17.40}$$

For waveguide networks, we must deal with the guide wavelength λ_g derived from

$$t = j \tan \frac{2\pi}{\lambda_g}\ell, \quad \lambda_g = \frac{\lambda}{\sqrt{1-\left(\frac{\lambda}{\lambda_{\text{co}}}\right)^2}} \tag{17.41}$$

where λ is the free-space wavelength and λ_{co} is the cutoff wavelength of the waveguide. By assuming ℓ to be a quarter of a wavelength at the center frequency, we obtain

$$t = j \tan \frac{2\pi}{\lambda}\sqrt{1-\left(\frac{\lambda}{\lambda_{\text{co}}}\right)^2}\frac{\lambda_{g0}}{4} \tag{17.42}$$

where λ_{g0} is the guide wavelength at the center frequency f_0, given by

$$\lambda_{g0} = \frac{\lambda_0}{\sqrt{1-\left(\frac{\lambda_0}{\lambda_{\text{co}}}\right)^2}} \tag{17.43}$$

Therefore

$$t = j \tan \frac{\pi}{2} \frac{\lambda_0}{\lambda} \frac{\sqrt{1 - \left(\frac{\lambda}{\lambda_{co}}\right)^2}}{\sqrt{1 - \left(\frac{\lambda_0}{\lambda_{co}}\right)^2}} \tag{17.44}$$

By substituting

$$F = \frac{f}{f_0} = \frac{\lambda_0}{\lambda}$$
$$F_0 = \frac{f_{co}}{f_0} = \frac{\lambda_0}{\lambda_{co}}$$

we obtain

$$t = j \tan \frac{\pi}{2} \sqrt{\frac{F^2 - F_0^2}{1 - F_0^2}}$$
$$\frac{dt}{dF} = j\frac{\pi}{2}(1 - t^2) \frac{F}{[(F^2 - F_0^2)(1 - F_0^2)]^{1/2}} \tag{17.45}$$

For the TEM-mode transmission, we obtain

$$f_{co} = 0$$
$$t = j \tan \frac{\pi}{2} F = j \tan \frac{\pi}{2} \frac{f}{f_0}$$
$$\frac{dt}{dF} = j\frac{\pi}{2}(1 - t^2) \tag{17.46}$$

Next, let us consider the general form of the Hurwitz polynomial $H(t)$ for the allpass networks given by

$$E(t) = \prod_{i=1}^{k} (t + \sigma_i) \prod_{j=1}^{m} (t^2 + 2\sigma_j t + |t_j|^2) \tag{17.47}$$

where σ_i, $\sigma_j > 0$ and $t_j = \sigma_j + j\omega_j$. Note that this equation is similar to (17.7) except that s is replaced by the distributed frequency variable t. The integer k represents the number of real zeros, and m represents the number of complex zeros; $k + m$ gives the order of the allpass network, and subscripts i and j refer to the ith and jth zeros, respectively. In microwave terminology, the first-order allpass network with a pole–zero location on the real axis is referred to as a *C section*, and the second-order network with complex conjugate pairs of pole–zero locations is referred to as a *D section*. Figure 17.7 describes the pole–zero distribution of a C/D-section allpass network.

On the basis of the relationships described in equations (17.36)–(17.47), the phase and normalized group delay for a C-section (first-order) allpass equalizer are given by

$$\beta = -2\tan^{-1} \frac{\Omega}{\sigma_c}$$
$$\tau_N = \tau f_0 = \frac{1}{2} \frac{\sigma_c (1 + \Omega^2)}{\sigma_c^2 + \Omega^2} S(F, F_0) \tag{17.48}$$

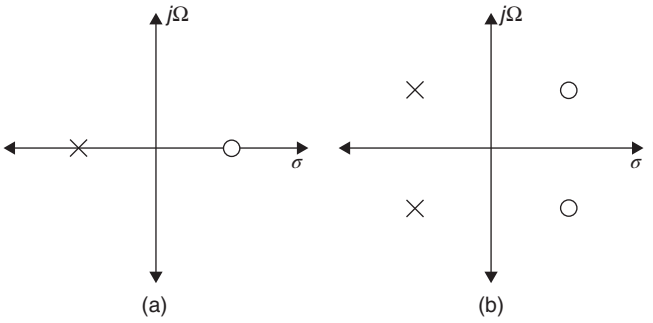


Figure 17.7 Pole–zero location of distributed allpass networks: (a) first-order allpass C section; (b) second-order allpass D section.

where

$$S(F, F_0) = \frac{F}{[F^2 - F_0^2]^{1/2}[1 - F_0^2]^{1/2}}$$

The phase and normalized time delay for a D-section (second-order) allpass equalizer are

$$\beta = -2\tan^{-1}\frac{2\sigma_d\Omega}{|t_d|^2 - \Omega^2}$$

$$\tau_N = \sigma_d(1+\Omega^2)\frac{\Omega^2 + |t_d|^2}{\Omega^4 + 2\Omega^2(\sigma_d^2 - \omega_d^2) + |t_d|^4}S(F, F_0)$$

$$t_d = \sigma_d + j\omega_d \tag{17.49}$$

Thus, in its most general form, the phase and normalized group delay of microwave allpass network, consisting of a cascade of k first-order and m second-order networks, is calculated by

$$\beta = -n\tan^{-1}\Omega - 2\left[\sum_{i=1}^{k}\tan^{-1}\frac{\Omega}{\sigma_i} + \sum_{j=1}^{m}\tan^{-1}\frac{2\sigma_j\Omega}{|t_j|^2 - \Omega^2}\right]$$

$$\tau_N = \frac{n}{4} + S(F, F_0)\left[\frac{1}{2}\sum_{i=1}^{k}\frac{\sigma_i(1+\Omega^2)}{\sigma_i^2 + \Omega^2} + \sum_{j=1}^{m}\sigma_i(1+\Omega^2)\frac{\Omega^2 + |t_j|^2}{\Omega^4 + 2(\sigma_j^2 - t_j^2) + |t_j|^4}\right] \tag{17.50}$$

From these relationships, it is possible to compute the phase and group delay responses of any allpass network. The normalized delays of C and D sections for a range of values of pole–zero locations are shown in Figures 17.8 and 17.9, respectively.

The shape of the group delays of allpass networks is nearly opposite that of the group delay of filter networks. Therefore, it is possible to equalize the delay of filters by cascading them with allpass networks. This implies that the location of the zeros of the allpass networks must be optimized to achieve the desired level of the overall phase and the group delay responses.

17.4 Physical Realization of Allpass Networks

A filter or an allpass equalizer is a two-port network. In lumped circuits, an equalizer is realized as a lattice structure, described in Section 17.2. At microwave frequencies, the equivalent of a lattice structure is complicated and often impractical. On the other hand, it is possible to

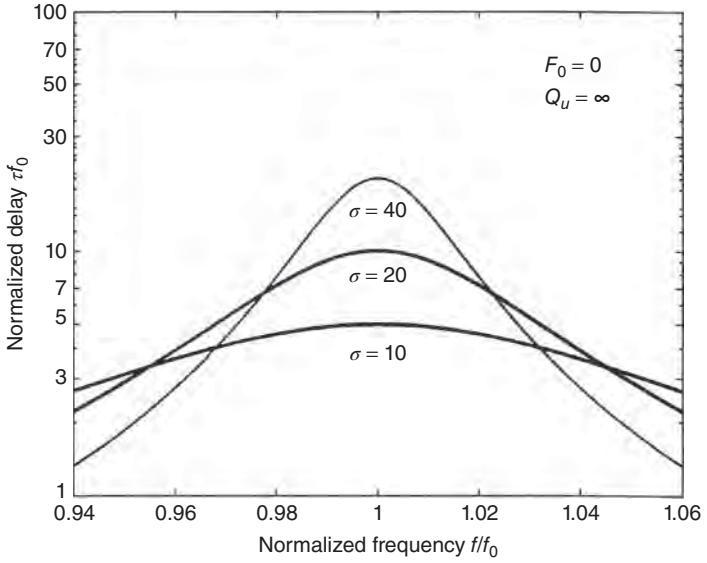


Figure 17.8 Normalized time delay of C-section TEM equalizer.

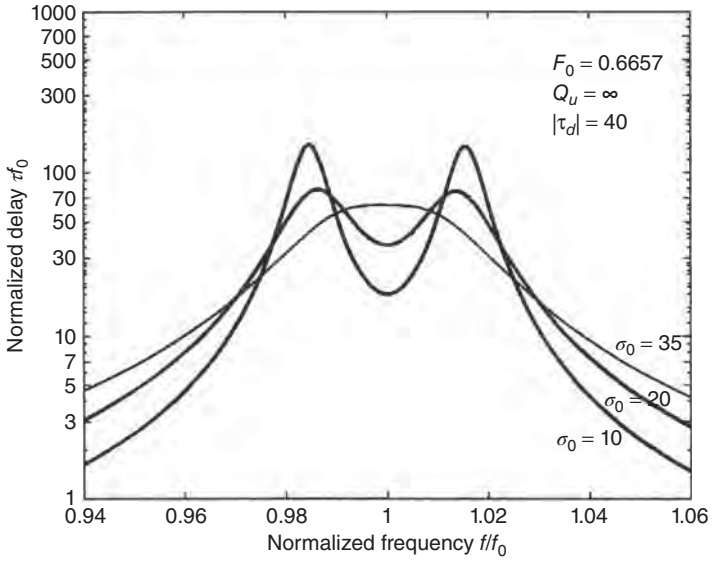


Figure 17.9 Normalized time delay of D-section waveguide equalizer.

realize allpass networks by making use of nonreciprocal ferrite circulators and isolators widely available at microwave frequencies. Such devices permit cascading of microwave circuits very effectively. The two alternative realizations are briefly described.

17.4.1 Transmission-Type Equalizers

Many authors have dealt with two-port transmission-type equalizer designs in TEM [4–6] and waveguide structures [7, 8]. A key requirement of such a network is that it maintain a

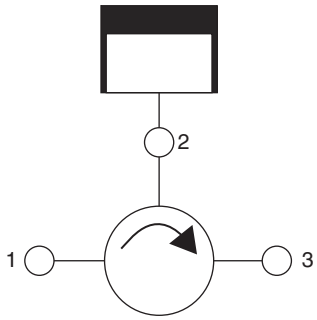


Figure 17.10 Circulator-coupled reflection-type allpass equalizer network.

near-perfect match over the band of interest, while its phase and group delay response is optimized to neutralize the filter delay. This tends to make the structure sensitive and very costly. As a result, such equalizers have found limited practical use. On the other hand, reflection-type equalizers are much simpler to implement and are described in the following section.

17.4.2 Reflection-Type Allpass Networks

The simplest configuration for an allpass microwave network is described in Figure 17.10.

This network consists of an ideal nonreciprocal three-port ferrite circulator, where port 1 is the input, port 2 is terminated with a lossless reactance network, and port 3 is the output. All the transmission lines and the circulator have the same characteristic impedance. A brief description of circulators is provided in Section 1.8.2. Due to the property of nonreciprocity and a virtual short circuit at port 2, all the energy incident at port 1 exits at port 3, providing an intrinsic allpass configuration. By designing the reactance network at port 2 as a cascade of commensurate transmission lines, we can realize an allpass network of any order. Ferrite circulators are widely available in the microwave frequency bands, in planar, coaxial, and waveguide structures. This changes the task of designing allpass networks to that of synthesizing one-port reactance networks. Thus, the design and implementation of allpass networks is much simpler and more cost effective. For some applications, the use of ferrite circulators may not be suitable owing to the requirements of a very low loss (<0.1 dB), wide operating temperature range (>40°C variation), high power (>tens or hundreds of watts), or any combination thereof. An alternative configuration that obviates the need of a nonreciprocal ferrite device is displayed in Figure 17.11.

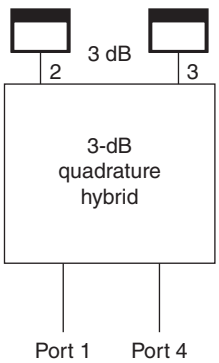


Figure 17.11 Hybrid-coupled reflection-type equalizer network.

The hybrid-coupled reflection-type equalizer network consists of a 3 dB quadrature hybrid, where port 1 is the input, port 4 is the output, and the coupled ports 2 and 3 are terminated in identical reactance networks. In such a scheme (assuming an ideal hybrid), all the power incident at port 1 exits at port 4, or vice versa, yielding an inherent allpass and reciprocal network. The design of hybrids is well established in a variety of microwave structures. In planar technology, use of hybrids offers the advantage of a compact structure, which is easy to implement. Furthermore, this configuration does not require ferrite elements and thus is not limited by the properties of ferrite materials. However, the configuration requires the use of two identical reactance networks, incurring extra hardware and a higher sensitivity in the implementation of such designs.

17.5 Synthesis of Reflection-Type Allpass Networks

Synthesis of the reflection-type microwave equalizer has been studied by several authors [4–11]. For low-order ($n \leq 3$) allpass networks and narrow bandwidths (<5%), Cristal [8] provides an excellent description and formulas for both waveguide and TEM structures. Such a range covers the majority of applications with adequate accuracy. This technique can be readily extended to higher-order networks by using modern synthesis techniques. The material presented in this section is extracted primarily from [8].

The synthesis of reflection-type equalizers is simplified by assuming an ideal circulator. An isolation of 30–40 dB or greater over the band of interest is quite adequate to justify such an assumption. Such circulators are readily available at microwave frequencies. The same holds true for the quadrature hybrid-coupled realization of an allpass network. It then boils down to the synthesis of a short-circuit reactance (one-port) network. More specifically, we need to synthesize the input reflection coefficient so that its delay characteristic has the desired response. The reflection coefficient of the reactance network is given by

$$\Gamma(j\Omega) = \frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_0} = \frac{z_{\text{in}} - 1}{z_{\text{in}} + 1} \tag{17.51}$$

where Z_0 is the characteristic impedance of the circulator, Z_{in} is the input impedance of the reactance network, and z_{in} is the normalized input impedance Z_{in}/Z_0. The next step is to relate the poles and zeros of the allpass function to the input impedance of the reactance network. It is well known that any reactance function is expressed as the ratio of the odd to even or even to odd parts of an associated Hurwitz polynomial. We may therefore express the normalized impedance as

$$z(j\Omega) = \frac{\text{even } E(j\Omega)}{\text{odd } E(j\Omega)} = \frac{E(j\Omega) - E(-j\Omega)}{E(j\Omega) + E(-j\Omega)} \tag{17.52}$$

Substituting (17.52) into (17.51), we obtain

$$\Gamma(j\Omega) = -\frac{E(-j\Omega)}{E(j\Omega)} \tag{17.53}$$

The negative sign suggests a phase change of π radians and presents no difficulty in realizing the equalizer network. The same results are derived by the dual formulation in terms of the admittance of the reactance networks. This result is identical to that of the formulation of allpass commensurate transmission-line allpass networks as described by equation (17.32). It is the mathematical proof that any arbitrary transmission-line allpass transfer function can be realized by the circulator-coupled reactance network. Once function $E(t)$ is optimized to achieve the desired delay characteristic, the $E(t)$ parameters are related to the reactance of the networks by equation (17.52).

The impedance Z_{in} of a two-port network in the distributed frequency variable t in terms of *ABCD* parameters is given by [see equation (12.18)]

$$Z_{\text{in}} = \frac{A(t)Z_L + B(t)}{C(t)Z_L + D(t)} \tag{17.54}$$

For a short-circuit reactance network, $Z_L = 0$, and

$$Z_{\text{in}} = \frac{B(t)}{D(t)} \tag{17.55}$$

Making use of equation (17.52), we have

$$Z_{\text{in}}(j\Omega) = \frac{B(t)}{D(t)} = Z_0 \frac{E(j\Omega) - E(-j\Omega)}{E(j\Omega) + E(-j\Omega)} \tag{17.56}$$

Since the optimized value of $E(t)$ is known, the input impedance, in terms of the [*ABCD*] parameters, is given by equation (17.56). Following the procedure described in Chapter 7, one can synthesize a reactance network of any order. The input impedance of C-, D-, and CD-section reactance networks is evaluated as follows:

C-section network:

$$\begin{aligned} E(t) &= t + \sigma_c \\ Z_{\text{in}}(j\Omega) &= \frac{B(t)}{D(t)} = Z_0 \frac{E(j\Omega) - E(-j\Omega)}{E(j\Omega) + E(-j\Omega)} \\ &= Z_0 \frac{j\Omega}{\sigma_c} \end{aligned} \tag{17.57}$$

D-section network:

$$\begin{aligned} E(t) &= t^2 + 2\sigma_D t + |t_D|^2 \quad \text{where } t_D = \sigma_D + j\omega_D \text{ and} \\ Z_{\text{in}}(j\Omega) &= \frac{B(t)}{D(t)} = Z_0 \frac{j2\sigma_D\Omega}{|t_D|^2 - \Omega^2} \end{aligned} \tag{17.58}$$

CD-section network:

$$\begin{aligned} E(t) &= (t + \sigma_c)(t^2 + 2\sigma_D t + |t_D|^2) \\ Z_{\text{in}}(j\Omega) &= \frac{B(t)}{D(t)} = Z_0 j\Omega \frac{(2\sigma_C\sigma_D + |t_D|^2) - \Omega^2}{\sigma_C|t_D|^2 - (\sigma_C + 2\sigma_D)\Omega^2} \end{aligned} \tag{17.59}$$

17.6 Practical Narrowband Reflection-Type Allpass Networks

As described in Chapter 3, the group delay response of filters varies inversely with absolute passband bandwidth. This implies that the group delay variation is large for filters with narrower absolute bandwidths and small for filters with larger bandwidths, irrespective of their center frequency. Typically, commercial communication systems employ narrowband channels to optimize the traffic flow and communication capacity. As a consequence, the bulk of the applications for microwave allpass networks require that the phase and group delays be equalized over narrow bandwidths, usually in the range of 0.5–5% of center frequency. The typical bandwidths of the RF channels in a communication system vary from a few megahertz up to 100 MHz. Such narrowband filters exhibit large group delays, especially at the band edges. It is, therefore, not possible to equalize the delay over the entire bandwidth of the filter without incurring a large group delay ripple. Consequently, typical specifications call for equalization of the group delay to be within a ripple of 1–3 ns in the middle 80% of the bandwidth, allowing higher group delay toward the band edges.

In a practical system, the operating thermal environment has a significant impact on the equalized delay. This is due to the fact that any misalignment between the filter and equalizer can result in a large deviation in the equalized delay. The higher the order of the equalizer, the greater is its sensitivity and thus the greater the misalignment of the group delays of the filter and

equalizer over the operating temperature range. As a result, most practical systems require no more than a third-order allpass network, with the bulk of the applications requiring a first- or second-order equalizer. For such applications, closed-form formulas are derived for narrowband one-port reactance networks [8]. A summary of these derivations for a narrowband C- and a D-section reactance networks is summarized as follows.

17.6.1 C-Section Allpass Equalizer in Waveguide Structure

The Hurwitz polynomial and the required input reactance for a single C-section equalizer are derived as follows:

$$E(t) = t + \sigma_c$$
$$z_{\text{in}} = j\left(\frac{\Omega}{\sigma_c}\right) = \frac{t}{\sigma_c} \tag{17.60}$$

The input reactance of the network, as shown in Figure 17.12, is given by

$$Z_1 = jZ_0 \tan 2\theta,$$
$$Z_2 = \frac{K_1^2}{jZ_0 \tan 2\theta}$$
$$z_{\text{in}} = \frac{Z_2}{Z_0} = \left(\frac{K_1}{Z_0}\right)^2 \frac{1}{j \tan 2\theta} \tag{17.61}$$

The waveguide cavities are π radians at the band center, and this is reflected in the choice of 2θ, as opposed to $\pi/2$ radians for cascaded transmission-line networks. For the frequency variable t, $\tan 2\theta$ corresponds to $2t/1 + t^2$, and hence

$$z_{\text{in}} = \frac{K_1^2}{Z_0^2}\left(\frac{1+t^2}{2t}\right) = \left(\frac{K_1}{Z_0}\right)^2 \left[\frac{1}{2t} + \frac{t}{2}\right] \tag{17.62}$$

In the vicinity of $t \to \infty$, near the resonant frequency, we obtain

$$z_{\text{in}} = \frac{1}{2}\left(\frac{K_1}{Z_0}\right)^2 t \tag{17.63}$$

Equating this impedance to that obtained in terms of σ_c for a C-section network in equation (17.60), we obtain

$$\frac{K_1}{Z_0} = \sqrt{2/\sigma_c} \tag{17.64}$$

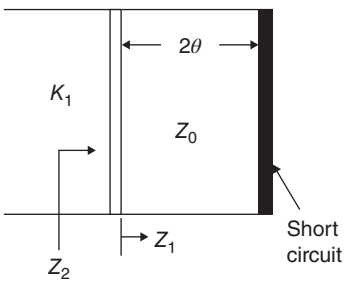


Figure 17.12 One-cavity reactance network.

The normalized reactance X_i/Z_0 and cavity length θ_i are related to the normalized values of the impedance inverters by [12]

$$\frac{X_i}{Z_0} = \frac{K_i/Z_0}{1-(K_i/Z_0)^2}, \quad i = 1,2,\ldots$$
$$\theta_i = \pi - \frac{1}{2}\left[\tan^{-1}\left(2\frac{X_i}{Z_0}\right) + \tan^{-1}\left(2\frac{X_{i-1}}{Z_0}\right)\right] \tag{17.65}$$

The reactance and cavity length are therefore evaluated as follows:

$$\frac{X_1}{Z_0} = \frac{\sqrt{2\sigma_c}}{\sigma_c - 2}$$
$$\theta_1 = \pi - \frac{1}{2}\tan^{-1}\left(2\frac{X_1}{Z_0}\right) \tag{17.66}$$

17.6.2 D-Section Allpass Equalizer in Waveguide Structure

Figure 17.13 shows a two-section reactance network in a waveguide structure. For the second-order allpass network, we have

$$E(t) = t^2 + 2\sigma_d t + |t_d|^2 \tag{17.67}$$

where $t_d = \sigma_d + j\omega_d$.

Following the same procedure as that of the C-section equalizer, the following approximate relationships for the D-section equalizer network are obtained [8]:

$$\frac{X_1}{Z_0} = \frac{2|t_d|}{|t_d|^2 - 4}, \quad \frac{K_1}{Z_0} = \frac{2}{|t_d|}$$
$$\frac{X_2}{Z_0} = \frac{2\sqrt{\sigma_d}\cdot|t_d|}{|t_d|^2 - 4\sigma_d}, \quad \frac{K_2}{Z_0} = 2\frac{\sqrt{\sigma_d}}{|t_d|}$$
$$\theta_1 = \pi - \frac{1}{2}\tan^{-1}2\frac{X_1}{Z_0}$$
$$\theta_2 = \pi - \frac{1}{2}\left[\tan^{-1}\left(2\frac{X_2}{Z_0}\right) + \tan^{-1}\left(2\frac{X_1}{Z_0}\right)\right] \tag{17.68}$$

17.6.3 Narrowband TEM Reactance Networks

For narrowband TEM network applications, the series capacitive coupled-resonator network in Figure 17.14 is useful, particularly with printed circuits and microstrip, since grounding the strips is not required. Also, it is particularly attractive to replace the circulator by a quadrature hybrid, thereby making the entire circuit relatively compact and easy to manufacture. This circuit is virtually the dual of the waveguide shunt inductive coupled cavity circuit described earlier. Therefore, equations (17.66) and (17.68) apply, provided that X_i/Z_0 is replaced by

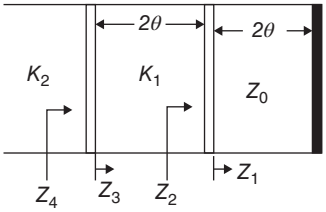


Figure 17.13 Two-cavity reactance network.

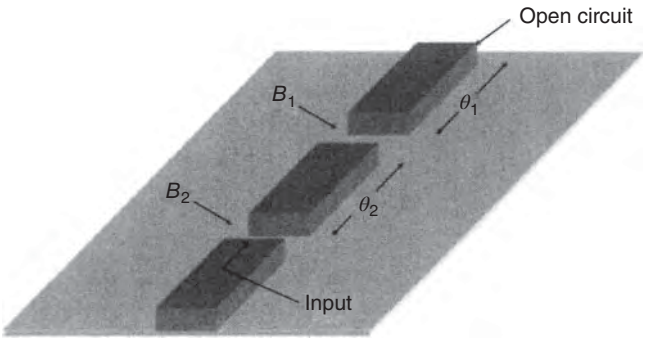


Figure 17.14 Capacitive coupled two-resonator reactance network for a second-order TEM transmission-line equalizer.

B_i/Y_0 and K_i/Z_0 is replaced by J_i/Y_0. B_i/Y_0 is the normalized series capacitive coupling susceptance, and J_i/Y_0 is the ideal admittance inverter, corresponding to the impedance inverter discussed earlier.

Numerous other types of TEM reactance networks are possible by using various distributed coupled line configurations. The design procedure for a cascade of commensurate transmission lines is well established. Cristal [8] describes an iterative relationship for determining the admittances of the cascaded reactance network, as well as a variety of TEM structures for arbitrary pole–zero locations of allpass networks.

17.6.4 Multiple Coupled Cavity Reactance Networks

A more generalized approach for the design of coupled cavity reflection-type equalizers of any order is described by Hsu *et al*. [11]. It deals directly with the reactance network composed of coupled cavity resonators. The group delay of the network is expressed in terms of the poles and zeros of the input impedance of the reactance network. The location of these poles and zeros can be readily optimized with respect to a given group delay specification mask. Once that is done, a synthesis process is used that explicitly relates the coupling coefficients of resonators to the poles and zeros of the reactance network. This procedure is well suited for synthesizing higher-order equalizer networks.

Another approach using coupling matrix synthesis for the realization of allpass networks has been proposed by Zhang and Caloz [13]. This approach leads to cross-coupled allpass filters characterized by a coupling matrix of twice the order of the transfer function of the equalizer. It is equivalent of realizing hybrid-coupled equalizers depicted in Figure 17.11. This generalized approach extends the coupling matrix synthesis to include allpass networks.

17.7 Optimization Criteria for Allpass Networks

Allpass networks are required to equalize the group delay of filters in a communication system. An ideal group delay response calls for a flat relative delay across the entire passband of the RF channel. However, such a response is not feasible. The best that can be achieved by an allpass network is an equiripple group delay over the entire passband. However, in such an approximation, the ripple can be quite large with a slow time period, introducing a large amount of distortion in the channel. An alternative is to achieve a lower value of the group delay ripple over the middle portion of the channel and let the value rise near the band edges. This allows

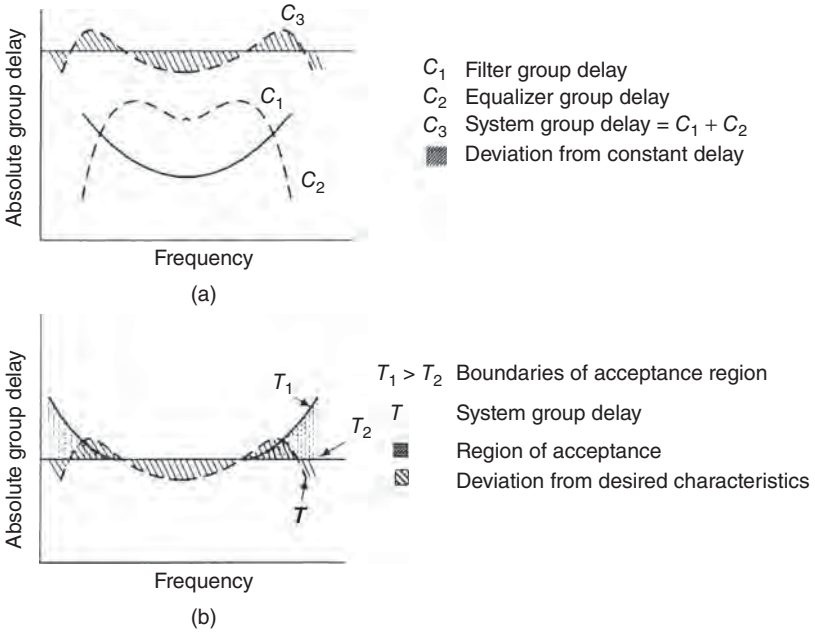


Figure 17.15 Optimization criteria for the equalizer: (a) when the system group delay is constant across the desired band and the constant itself either has a preassigned value or is allowed to float; (b) when the system group delay lies between the curves T_1 and T_2 designated as the region of acceptance.

a tradeoff between the flat portion of the delay, usually symmetric about the band center, and the delay variation at the band edges. This channel characteristic is suitable as it tends to follow the energy spectrum of most modulation schemes. Thus, the optimization constraints typically call for the total system relative group delay to lie within a prescribed region. This allows the tradeoff for flatness of group delay (i.e., control of group delay ripple) in the central region, at the expense of the increased delay toward passband edges. A typical situation in which a communication channel requires equalization is depicted in Figure 17.15.

Let the curve $C_1(f_i, \tau_i, i = 1,2, \ldots, N)$ be the set of points that describe the absolute group delay versus frequency characteristics of the filter, where f_i is the ith frequency point and τ_i is the corresponding absolute group delay. Let τ_c and τ_D be the group delays, generated at a particular frequency f_i by a cascade of a C- and D-section equalizer. Then, curve $C_2(f_i) = (\tau_C)_i + (\tau_D)_i$, where

$$(\tau_C)_i + (\tau_D)_i = \sum_{k=1}^{k} \tau_k + \sum_{m=1}^{m} \tau_m \tag{17.69}$$

represents the group delay of the equalizer consisting of k C sections and m D sections.

The overall group delay C is then given by

$$C = C_1 + C_2 \tag{17.70}$$

Defining C_0 as the overall delay at the band center, the relative delay C_R of the equalized network is

$$C_R = C - C_0, \quad C_0 = (C_1 + C_2) \quad \text{at} \quad f = f_0 \tag{17.71}$$

The minimization of C_R yields an equiripple overall delay over the desired bandwidth. Going a step further, we let the curves T_1 and T_2, as shown in Figure 17.15b, describe the prescribed

region for the overall relative delay. To meet such a criterion requires the minimization of function F, derived from

$$F = \sum_{i=1}^{N} (T_{1i} - C_{Ri})^2 + \sum_{i=1}^{N} (T_{2i} - C_{Ri})^2$$
$$T_{1i} < C_{Ri} \qquad T_{2i} > C_{Ri} \tag{17.72}$$

The minimization of F leads to determination of the optimum locations of the zeros of the C and D sections. The total number of variables involved is $(k + 2m)$. Function F is minimized by a variety of optimization procedures well documented in the literature.

Example 17.1 In this example, we determine the tradeoffs and design parameters to equalize the relative group delay of a channelizing filter in a 6/4 GHz communication satellite.

Channelizing Filter Parameters. Typical filter parameters to channelize 36 MHz transponders are as follows:

Order of the filter	9
Filter type	Chebyshev
Center frequency	4000 MHz
Passband bandwidth	38 MHz
Return loss	22 dB

Computed amplitude and group delay response of this filter is described in Figure 17.16.

Equalizer Tradeoffs. In order to generate group delay tradeoffs, we will consider a C-, D-, and CD-section allpass networks to equalize the delay of the channelizing filter. Typical system constraint requires passband group delay ripple not to exceed 2 ns peak to peak. It is, of course, not possible to equalize the delay to within 2 ns over the entire bandwidth of the filter. The intent is to maximize the equalized bandwidth within the constraint of 2 ns ripple while letting the group delay rise towards the bandedges as a tradeoff parameter.

The constraints imposed for the optimization of the equalizer network can therefore be summarized as follows:

i) $|\tau(f) - \tau(f_0)| < 1$ ns over the middle portion of the filter bandwidth
ii) Maximization of equalized bandwidth within a peak to peak group delay ripple of 2 ns

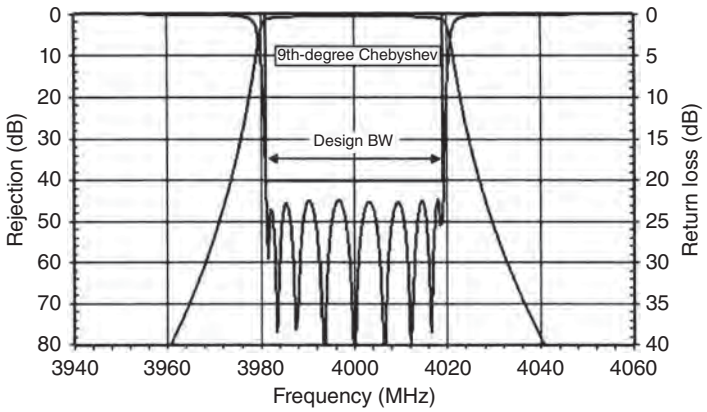


Figure 17.16 Amplitude response of a nine-pole Chebyshev filter.

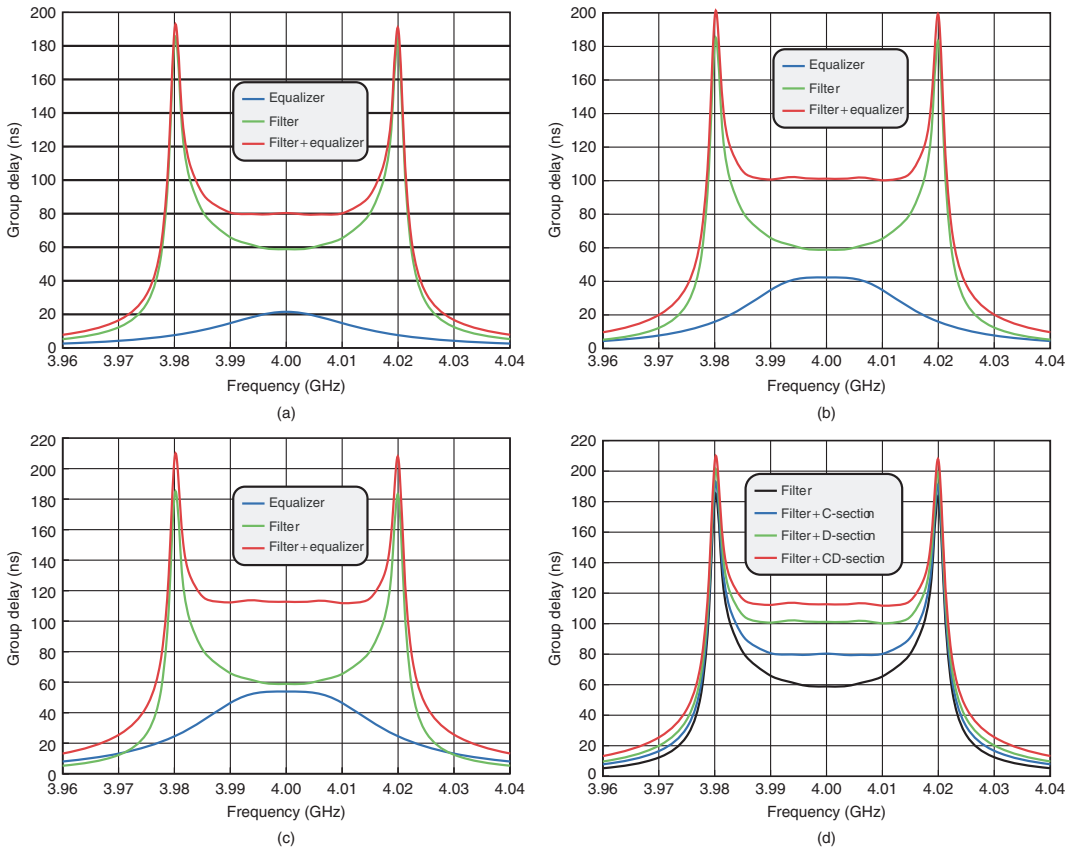


Figure 17.17 Optimized group delay of the equalized filter (a) using C-section equalizer; (b) using D-section equalizer; (c) using CD-section equalizer; and (d) comparison of the equalized group delays.

Computed Values of the Optimized Parameters

Filter & Equalizer Structure: WR 229 waveguide

$a = 2.29$ in. $= 58.166$ mm, $b = 1.145$ in. $= 29.083$ mm, cutoff frequency $= 2.577$ GHz

Optimized Equalizer Parameters

C-section equalizer: $\sigma_C = 101$, maximum equalized bandwidth $= 59\%$

D-section equalizer: $\sigma_d = 99$, $\omega_d = 57.5$, maximum equalized bandwidth $= 74\%$

C/D-section equalizer: $\sigma_C = 67$, $\sigma_d = 92.5$, $\omega_d = 57.5$, maximum equalized bandwidth $= 77.5\%$

The optimized group delay for the three equalizers, along with the filter delay, is displayed in Figure 17.17. This example clearly shows that a D-section equalizer represents the most cost-effective approach for the equalization of group delay in a communication channel. Using third- or higher-order equalizers improves the equalized bandwidth marginally at the expense of increased mass, complexity, and sensitivity of the structure. For these reasons, practical systems invariably use a second-order or at most third-order equalizer network.

Physical Realization. For computing the physical dimensions, we have chosen the D-section equalizer since it provides the most efficient practical equalizer network. Computed results following equation (17.68) are as follows (Figure 17.18):

Electrical Parameters

$$\sigma_d = 99.03 \quad \omega_d = 57.54$$
$$|t_d| = |\sigma_d + j\omega_d| = 114.53$$
$$\frac{K_1}{Z_0} = 0.0175 \quad \frac{K_2}{Z_0} = 0.1738$$
$$\frac{X_1}{Z_0} = 0.0175 \quad \frac{X_2}{Z_0} = 0.1792$$
$$\theta_1 = 179.0 \quad \theta_2 = 169.1$$

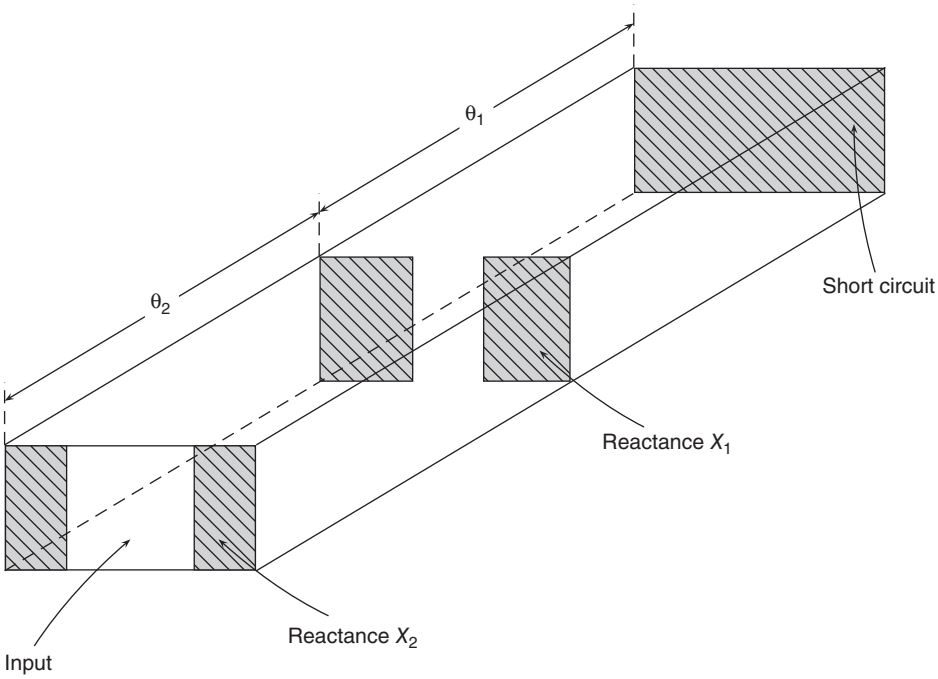


Figure 17.18 Schematic of the D-section equalizer.

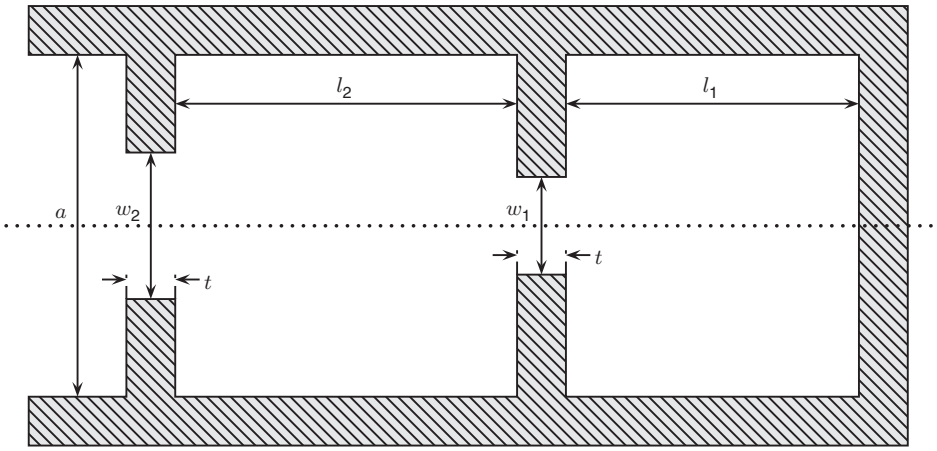


Figure 17.19 Physical realization of the D-section equalizer (top view).

Physical dimensions of the D-section equalizer in Figure 17.19.

WR-229 waveguide, $a = 58.166$ mm, $b = 29.083$ mm

- Centrally located Inductive Iris.
- Iris thickness is $t = 2$ mm.
- Widths of the irises: $w_1 = 9.067$ mm, $w_2 = 20.776$ mm.
- Cavity lengths: $l_1 = 48.53$ mm, $l_2 = 45.2814$ mm.

Physical dimensions are computed using EM techniques.

17.8 Dissipation Loss

The dissipation loss in allpass networks follows the same relationship as the filter networks. It is directly proportional to the normalized absolute group delay of the network and inversely to the unloaded Q of the equalizer structure. A good approximation for the loss is given by

$$\alpha_E = 20 \log e \frac{\pi}{Q} \tau_N \text{ dB} \tag{17.73}$$

where τ_N is the normalized group delay and Q is the unloaded of the equalizer network.

From equations (17.48) and (17.49), the normalized midband delay of a C section is calculated as $\sigma/2(1 - Fo^2)$ and for a D section by $\sigma/(1 - Fo^2)$, respectively. The dissipation loss is therefore evaluated as follows:

For C sections,

$$\alpha = 20 \log e \frac{\pi}{Q} \frac{\sigma/2}{1 - F_0^2} \text{ dB} \tag{17.74}$$

For D sections,

$$\alpha = 20 \log e \frac{\pi}{Q} \frac{\sigma}{1 - F_0^2} \text{ dB} \tag{17.75}$$

The total midband loss of an equalizer is obtained by summing the contributions from the individual C and D sections. This approximation holds for the entire passband, that is, the loss

variation tracks the group delay variation, where the constant of proportionality is $20 \log e\pi/Q$ or $27.3/Q$.

17.9 Equalization Tradeoffs

The primary application of allpass networks is to enhance the passband performance of a filter network without altering its amplitude response. However, the use of allpass networks increases the complexity, loss, mass, volume, and cost. The use of allpass networks is, therefore, dictated by overall system-level tradeoffs.

Equalization of the channel is accomplished by using external equalizers or self-equalized (linear phase) filters. Linear phase filters represent an attractive alternative to employing external equalizer networks, but the penalty involves an increase in the order of the filter and its design complexity. Such designs are sensitive to physical tolerances and difficult to tune to achieve the desired response. This is especially true for narrowband filters, where the bulk of the applications are. On the plus side, linear phase filters represent the minimum mass and volume solution to achieve a given level of performance. Their sensitivity to the operating environment is similar to that of minimum phase filters, since self-equalization implies additional resonators.

External equalizers offer simplicity in the design and implementation to achieve the desired overall (equalized) channel response. This is especially true for reflective-type equalizers. Such external equalizer networks are completely independent of the filter network. The equalizers can be employed during the final phases of integration to equalize the filter delays as well as any residual delays in the system. Equalizer networks do not need to be integrated with filters and can be separated by transmission lines or waveguide media. The design requires reactance networks that are simple to design and tune to meet a desired response. For a given overall order of filter–equalizer network, external equalization typically offers better channel characteristics than the equivalent-order linear phase filters.

Another advantage of external equalizers is that it offers the degree of freedom to change the unloaded Q of the reactance network, thereby providing a measure of control to equalize the passband amplitude with little or no impact on the group delay equalization. However, external equalization incurs extra mass and volume of the nonreciprocal (ferrite circulator) or reciprocal (3 dB quadrature hybrid) devices used to couple the reactance networks. The external equalizer incurs a higher loss owing to the losses associated with the coupling devices. For low-power applications, this extra loss is quite small and has a minimum impact on the system design. Another drawback of reflection equalizers is their sensitivity to the operating temperature range, and that can be problematic beyond ±20°C. Generally, this is due to the properties of the ferrite materials used in circulators.

From the foregoing discussion, it is clear that the detailed system-level tradeoffs are necessary to determine the best filter network design to achieve the desired channel characteristics in a communication system.

Summary

The starting point of this chapter is a brief coverage of lumped allpass networks. This serves as a background to address microwave allpass networks.

An allpass network is characterized by a cascade of the first-order (C-section) and second-order (D-section) allpass networks. The poles and zeros are located symmetrically on the real axis for C section and distributed as a complex quad for the D section. These locations can be

optimized to equalize the delay of a filter over a given bandwidth. This allows the computation of the overall group delay performance and tradeoffs independent of the physical realization of the equalizer network. For microwave applications, it is most convenient to realize the allpass network by using a three-port nonreciprocal circulator with one port terminated in a short-circuit reactance network; such a configuration provides an inherent allpass network. An alternative is the use of a 3 dB quadrature hybrid with two of its ports terminated in identical reactance networks. At microwave frequencies, circulators with 30–40 dB directivity are readily available. For practical applications, such an isolation between the successive ports of a circulator is adequate to realize a near-ideal allpass network. This reduces the synthesis procedure to the realization of an appropriate reactance network. A practical example of a satellite channel is used to demonstrate group delay equalization tradeoffs using C-, D-, and CD-section equalizers. The optimized D-section equalizer is chosen for the computation of the physical dimensions using EM techniques.

The chapter concludes with a brief discussion of the relative merits of external equalizers versus linear phase filters.

References

1 Blinchkoff, H.J. and Zverev, A.I. (1976) *Filtering in the Time and Frequency Domains*, Wiley, New York.
2 Van Valkenburg, M.E. (1960) *Modern Network Synthesis*, Wiley, New York.
3 Scanlon, J.O. and Rhodes, J.D. (1968) Microwave all-pass networks Part 1 and Part 2. *IEEE Transactions on Microwave Theory and Techniques*, **16**, 62–79.
4 Merlo, D. (1965) Development of group-delay equalizers for 4GC. *Proceedings 1EE (London)*, **112**, 289–295.
5 Cristal, E.G. (1966) Analysis and exact synthesis of cascaded commensurate transmission-line C-section all-pass networks. *IEEE Transactions on Microwave Theory and Techniques*, **14**, 285–291. [see also addendum to this article, ibid., **14**, 498–499 (Oct. 1966)].
6 Abele, T.A. and Wang, H.C. (1967) An adjustable narrow band microwave delay equalizer. *IEEE Transactions on Microwave Theory and Techniques*, **15**, 566–574.
7 Woo, K. (1965) An adjustable microwave delay equalizer. *IEEE Transactions on Microwave Theory and Techniques*, **13**, 224–232.
8 Cristal, E.G. (1969) Theory and design of transmission line all-pass equalizers. *IEEE Transactions on Microwave Theory and Techniques*, **17** (1), 28–38.
9 Kudsia, C.M. (1970) Synthesis of optimum reflection type microwave equalizers. *RCA Review*, **31** (3), 571–595.
10 Chen, M.H. (1982) The design of a multiple cavity equalizer. *IEEE Transactions on Microwave Theory and Techniques*, **30**, 1380–1383.
11 Hsu, H.T. *et al.* (2002) Synthesis of coupled-resonators group-delay equalizers. *IEEE Transactions on Microwave Theory and Techniques*, **50** (8), 1960–1967.
12 Cohn, S.B. (1957) Direct-coupled resonator filters. *Proceedings of the IRE*, **45**, 187–196.
13 Zhang, Q. and Caloz, C. (2013) Alternative construction of the coupling matrix of filters with non-paraconjugate transmission zeros. *IEEE Microwave and Wireless Components Letters*, **23** (10), 509–511.

18

Multiplexer Theory and Design

Previous chapters have dealt with the design of microwave filters as an individual component of a communication network. In a multiuser environment, communication systems employ a range of microwave filters as individual components, as well as part of a multiport network required to separate or combine a number of radio frequency (RF) channels. This is commonly referred to as a *multiplexing network* or simply a *multiplexer* (MUX). This chapter is devoted to the design and tradeoffs of MUXs for a variety of practical applications.

18.1 Background

MUXs are used in communication system applications where there is a need to separate a wideband signal into a number of narrowband signals (RF channels). Channelization of the allocated frequency band allows flexibility for the flow of communication traffic in a multiuser environment. Amplification of individual channels also eases the requirements on the high-power amplifiers (HPAs), enabling them to operate at relatively high efficiency with an acceptable degree of nonlinearity. MUXs are also employed to provide the opposite function, that is, to combine several narrowband channels into a single wideband composite signal for transmission via a common antenna. They are, therefore, referred to as *channelizers* or *combiners*. Due to the reciprocity of filter networks, a MUX can also be configured to separate the transmit and receive frequency bands in a common device, referred to as a *duplexer* or *diplexer*. Multiplexers have many applications such as in satellite payloads, wireless systems, and electronic warfare (EW) systems. A more detailed rationale for the channelization of an allocated frequency band and an overview of multiplexing schemes and their impact on channel characteristics are presented in Chapter 1.

Although the principles of combining or separating frequency-diverse signals (channels) for interfacing with a single port of an antenna system have been known for many years, it was the advent of satellite communication systems in 1970s that motivated major advances in this field. Figure 18.1 illustrates a simplified block diagram of a conventional satellite payload system. It consists of a receive and transmit antenna, a low-noise receiver, input multiplexer (IMUX) and output multiplexer (OMUX), and HPAs. The payload behaves as an orbiting repeater that receives, amplifies, and transmits signals in the allocated frequency bands. The practical constraints on HPAs require the channelization of the received signal into a number of RF channels by an IMUX. After the relatively narrowband channels are amplified separately by the HPAs, the channels are recombined by an OMUX for transmission back to the ground via a common

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

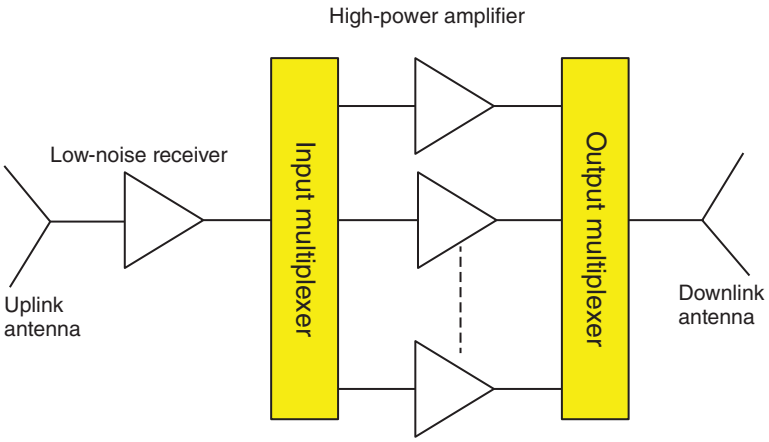


Figure 18.1 A simplified block diagram of a satellite payload.

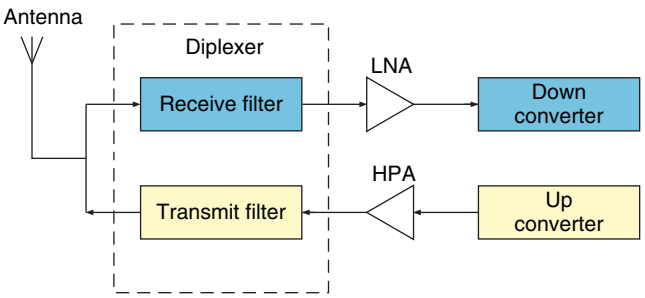


Figure 18.2 A simplified block diagram of the front end in base stations.

antenna. In satellite payloads, the IMUX and OMUX govern the characteristics of RF channels and hence have a substantial impact on the performance of the overall payload. The typical number of filters required for input and output multiplexing networks ranges from 48 to well over 100.

Other MUX structures that have found widespread applications are the transmit/receive diplexers that are used in the base stations of mobile telephony systems. These devices have the most demanding electrical and build standard requirements of all, accommodating both high- and low-power signals in close proximity within one housing. In some cases, these diplexers are located at the top of transmitter masts, where diplexers are subjected to the worst of climatic extremes. Figure 18.2 shows a simplified block diagram of the front end of a cellular base station. The purpose of the receive filter is to reject the out-of-band interference prior to the low-noise amplification and downconversion. The transmit filter is used primarily to limit the out-of-band signals generated by the transmit portion of the base station. The transmit filter must also have a very high level of rejection in the receive band in order to eliminate the possibility of intermodulation (IM) products being fed into the receiver through the common antenna.

MUXs are also used in wireless applications where the base station may need to transmit various frequency channels in different directions by using directive antennas. In this case, a multiplexer is needed to separate the overall band into separate channels, radiated in various directions. Another possible application of MUXs in wireless base stations is in cases where the

base station provides service to a number of independent operators that are licensed to operate only in specific channels within the frequency band covered by the base station.

In EW systems, MUXs are used to construct switched filter banks, which are essential building blocks in wideband receivers that operate in a hostile signal-dense environment. In this case, switches are integrated into the MUX to allow the selection of a particular channel, effectively allowing the realization of a tunable filter with a variable bandwidth and a variable center frequency.

18.2 Multiplexer Configurations

Since the mid-1970s, there have been many advances in the design and implementation of multiplexing networks [1–20]. The most commonly used configurations are hybrid-coupled MUXs, circulator-coupled MUXs, directional filter MUXs, and manifold-coupled MUXs. A summary of the advantages and disadvantages of these MUX configurations is given in Table 18.1 [5].

18.2.1 Hybrid-Coupled Approach

Figure 18.3 shows a layout of a hybrid-coupled MUX. Each channel consists of two identical filters and two identical 90° hybrids. The main advantage of the hybrid-coupled approach is its directional property, which minimizes the interaction among the channel filters. As a consequence, the hybrid-coupled MUX is amenable to a modular concept. It also allows the integration of additional channels at a later date without disrupting the existing MUX design, which is a requirement in some systems. Another key advantage of this approach is that only half of the input power goes through each filter. Thus, the filter design can be relaxed when using this type of MUX in high-power applications. On the other hand, its large size is a disadvantage, since two filters and two hybrids per channel are required. Another design consideration of such MUXs is the phase deviation between the two filter paths that the two signals undergo before they add constructively at the channel output. As a result, the structure must be fabricated with tight tolerances to minimize the phase deviation particularly in planar circuit applications where it is difficult to use tuning elements to balance the two paths.

18.2.2 Circulator-Coupled Approach

Each channel, in this approach, consists of a channel-dropping circulator and one filter, as shown in Figure 18.4. The unidirectional property of the circulator provides the same advantages as does the hybrid-coupled approach in terms of amenability to modular integration and ease of design and assembly. The insertion loss of the first channel is the sum of the insertion loss of the channel filter and the insertion loss of the circulator. The subsequent channels exhibit a relatively higher loss due to the insertion loss incurred during each trip through the channel dropping circulators. Figure 18.5 is a photo of a four-channel MUX [17]. It consists of a stack of four circulators connected to a stack of four filters by cables.

18.2.3 Directional Filter Approach

Figure 18.6 illustrates a layout of an MUX, realized by connecting directional filters in series. A direction filter is a four-port device where one port is terminated in a load. The other three ports of the directional filter essentially act as a circulator connected to a bandpass filter. The power incident at one port emerges at the second port with a bandpass frequency response, while the reflected power from the filter emerges at the third port. However, directional filters do not

Table 18.1 Comparison among various multiplexer configurations.

Hybrid-coupled MUX	Circulator-coupled MUX	Directional filter MUX	Manifold multiplexer
Advantages			
Amenable to modular concept	Requires one filter per channel	Requires one filter per channel	Requires one filter per channel
Simple to tune, no interaction between channel filters	Employs standard design of filters	Simple to tune, no interaction between channel filters	Most compact design
Total power in transmission modes as well as reflection mode is divided by the hybrid so that only 50% of the power is incident on each filter; power handling is thus increased and susceptibility to voltage breakdown is reduced	Simple to tune; no interaction between channel filters Amenable to a modular concept	Amenable to modular concept	Capable of realizing optimum performance for absolute insertion loss, amplitude and group delay response
Disadvantages			
Two identical filters and two hybrids are required for each channel	Signals must pass in succession through circulators, incurring extra loss per trip	Restricted to realize all-pole functions such as Butterworth and Chebyshev	Complex design
Line lengths between hybrids and filters require precise balancing to preserve circuit directivity	Low-loss, high-power ferrite circulators are expensive	Difficult to realize bandwidths greater than 1%	Tuning of multiplexer can be time-consuming and expensive
Physical size and weight of the multiplexer is greater than in other approaches	Higher level of passive intermodulation (PIM) products than in other configurations		Not amenable to a flexible frequency plan; that is, change of a channel frequency would require a new multiplexer design

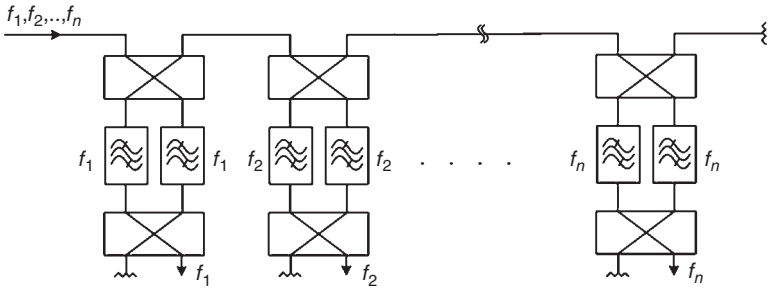


Figure 18.3 Layout of a hybrid-coupled multiplexer.

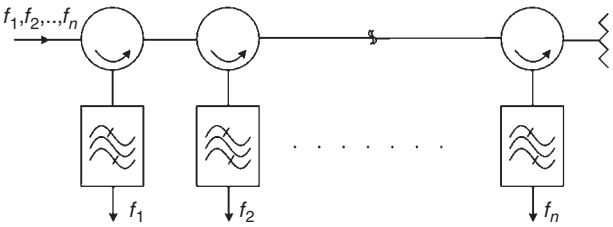


Figure 18.4 A circulator-coupled multiplexer.

Figure 18.5 A four-channel circulator-coupled multiplexer. *Source*: Mansour *et al*. [17] 1996. Reproduced with the permission of IEEE.

require the use of ferrite circulators. Figure 18.7 illustrates two possible approaches for realizing directional filters in waveguide and microstrip, respectively [1]. The waveguide directional filter is realized by coupling rectangular waveguides, operating in TE_{10} mode, to a circular waveguide filter operating in TE_{11} modes. In the microstrip version, each 360° wavelength ring resonator is coupled to another wavelength ring resonator and to two transmission lines. This multiplexing approach has the same advantages as the hybrid-coupled and circulator-coupled approaches. It is, however, limited to narrowband applications.

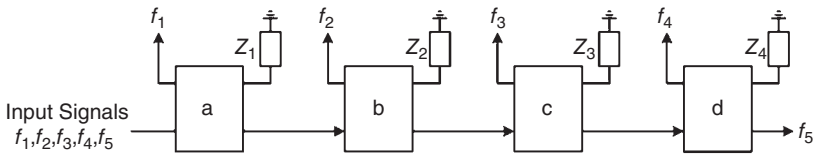


Figure 18.6 A directional filter multiplexer.

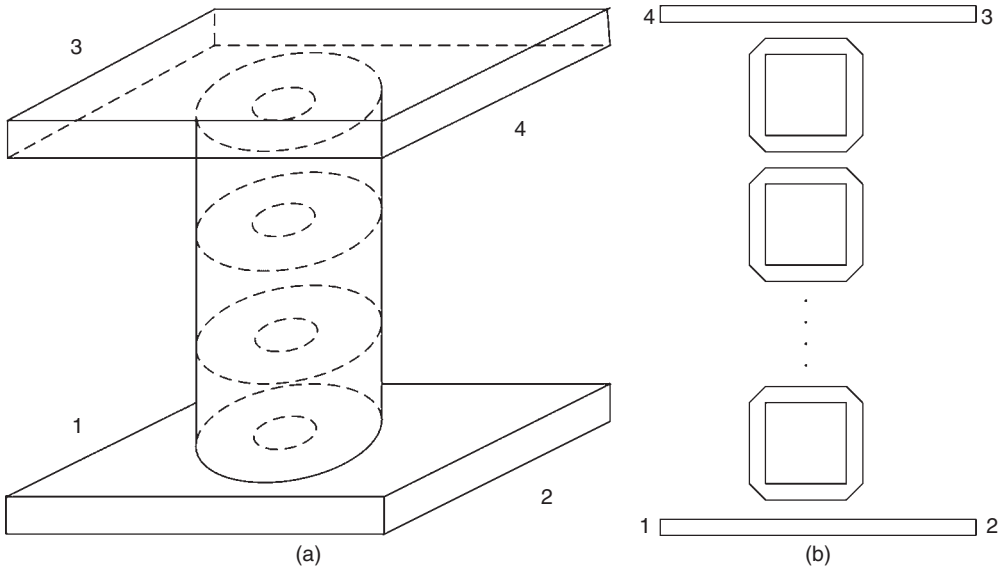


Figure 18.7 (a) A waveguide directional filter and (b) microstrip directional filter.

18.2.4 Manifold-Coupled Approach

The manifold-coupled approach, shown in Figure 18.8, is viewed as the optimum choice as far as miniaturization and absolute insertion loss are concerned; this type of MUX requires the presence of all the channel filters at the same time so that the effect of channel interactions can be compensated for in the design process. This implies that a manifold-coupled MUX is not amenable to a flexible frequency plan. Any change in the allocation of the channels requires a new MUX design. Moreover, as the number of channels increases, this approach becomes more difficult to implement. The manifold-coupled MUX shown in Figure 18.8 acts as a channelizer. The same configuration can be used as a combiner. Figure 18.9 is a photo of a four-channel MUX employing a waveguide manifold and four dual-mode circular waveguide filters. The same figure also shows a miniature MUX with a coaxial manifold and superconductive filters

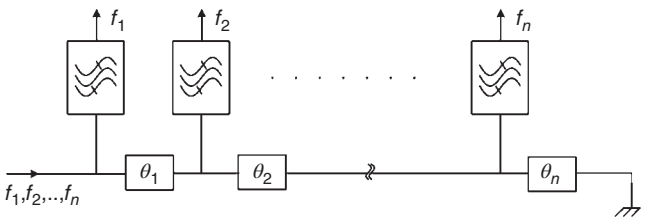


Figure 18.8 A manifold-coupled multiplexer.

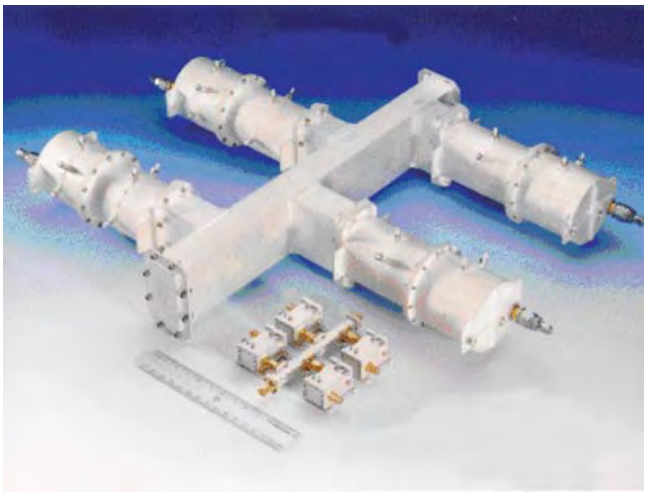


Figure 18.9 A C-band four-channel waveguide multiplexer compared with a four-channel multiplexer implemented using superconductor technology. *Source*: Mansour *et al*. [20] 2000. Reproduced with the permission of IEEE.

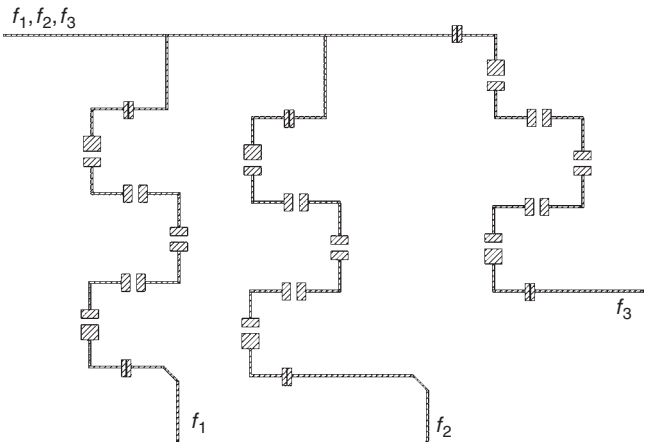


Figure 18.10 A three-channel planar multiplexer. *Source*: Mansour *et al*. [17] 1996. Reproduced with the permission of IEEE.

[20]. The manifold-coupled concept can be implemented as well in planar circuits as shown in Figure 18.10. Here, three microstrip filters are integrated with a microstrip manifold [17]. In this particular case, one of the channels is connected directly to the manifold.

There are three distinct categories of multiplexing networks required by communication systems, namely, RF channelizers, RF combiners, and transmit–receive diplexers. The design of each type of MUX is dictated by its application and system constraints.

18.3 RF Channelizers (Demultiplexers)

For broadband communication systems, the available frequency spectrum is split into a number of smaller frequency bands or RF channels. This is done to optimize the system performance,

as explained in Chapter 1. This function is carried out by a multiplexer, referred to as an *input mux* (*IMUX*) or *demultiplexer* or *RF channelizer*. Such a MUX is required at each receiving station in the system. Design constraints on such a MUX are as follows:

1) It must be able to split a frequency band into a number of discrete contiguous RF channels.
2) Each RF channel must exhibit a near-ideal response, that is, a minimum of amplitude and group delay variation across its passband and a large attenuation outside of it. This is required to minimize the interference from other channels and to provide HPA with a clean signal for amplification. In other words, a near-ideal filter at this stage prevents the amplification of the interference that may be present outside the passband of the RF channel.

In a communication system, RF channelization is carried out after amplification of the broadband signal by a low-noise amplifier (LNA). As a consequence, absolute insertion loss is not a constraint. This feature is exploited in the design of such MUXs. It should be noted that excessive loss can become a liability if it incurs the requirement of an extra amplifier. The most common designs for IMUX are the hybrid branching network and the circulator-coupled MUX.

18.3.1 Hybrid Branching Network

Figure 18.11 shows the typical arrangement for a hybrid branching network. The figure depicts the case for the demultiplexing of a group of five contiguous channels. Each 3 dB hybrid in the branch splits the incoming wideband signal into two equal paths until the number of paths equals the number of channels. At the end of each path, channel filters are located, each with an isolator at its input.

The entire spectrum of the channel group is presented to the input port of each filter, and the channel corresponding to the passband of the filter passes through; the rest are reflected and absorbed in the isolator loads. There is no restriction in the placement of filters corresponding to any channel at the end of any of the paths.

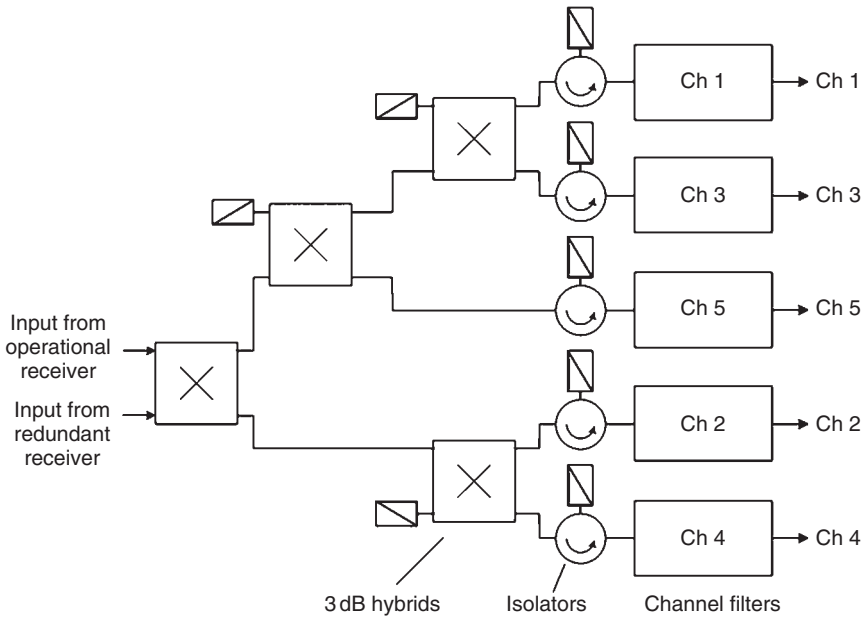


Figure 18.11 Hybrid branching input multiplexer (IMUX).

The advantages of the hybrid branching network are summarized here:

- Design is simple, requiring standard hybrids and channel filters.
- The branching network can be very lightweight, compact, and reliable, using integrated microstrip technology.
- The second port of the input hybrid can be used for the input from a redundant LNA, without the need for an RF switch.
- There is no channel interaction, since each filter receives the broadband signal directly at its input port. Each channel filter operates independently and is tuned separately without regard to the other filters in the overall MUX.
- The input return loss of the overall IMUX is good over a wide bandwidth.

A disadvantage is that insertion loss is high, typically 3.5 dB per hybrid. The worst-case insertion loss (channels 1 and 3) in the IMUX shown in Figure 18.11 is estimated to be approximately 12 dB at Ku-band, excluding the loss of the channel filter. Another disadvantage is that there is wide variation in the insertion loss of various channels depending upon the number of hybrids encountered in the path prior to the channel filters. This inequality in the power levels among channels requires equalization prior to HPAs, thus adding a degree of complexity in system design.

18.3.2 Circulator-Coupled MUX

The circulator-coupled IMUX is sometimes referred to as a *channel dropping* MUX, which gives a better indication of how it operates [15]. An example of a circulator-coupled IMUX for demultiplexing a group of five contiguous channels is shown in Figure 18.12. The hybrid at the input divides the power of the five-channel group into two branches. In each branch, the first circulator directs the group to the input of the first channel filter in the branch. The channel corresponding to the filter's passband passes (or drops) through the filter to its output, and the other channels, in out-of-band position relative to this filter, are reflected at its input and are directed by the circulator toward the second circulator/channel filter in the chain. Here, the

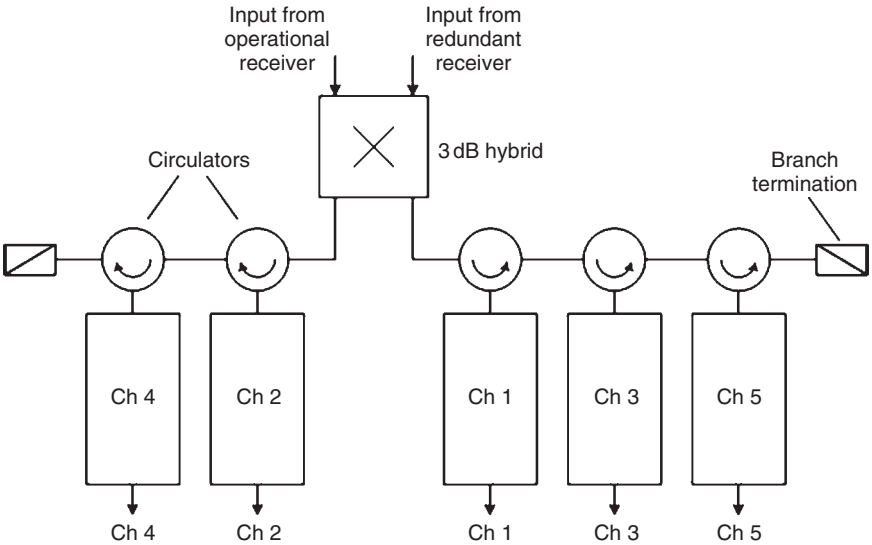


Figure 18.12 Circulator-coupled or channel dropping input demultiplexer.

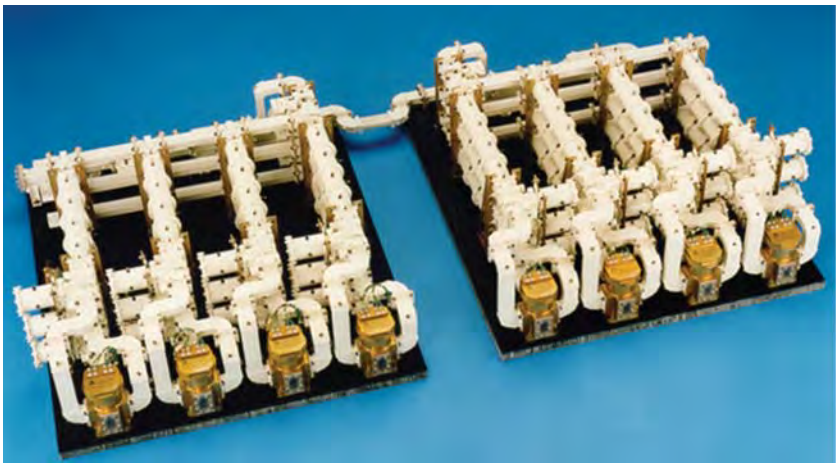


Figure 18.13 Circulator-coupled input demultiplexer subsystem. *Source*: Photo courtesy COM DEV Ltd.

second channel drops through, the rest are reflected, and so on until all the remaining channels are absorbed in the terminating load at the end of the chain.

This design approach provides an excellent way to retain nearly all the advantages of the hybrid branching scheme with a much lower insertion loss. The use of a single hybrid provides two alternative paths to extract the RF channels. Availability of the two paths allows the use of alternate (noncontiguous) channel filters in each path. Once the channels are extracted from each path, energy of all other channels, including those assigned to the other path, are absorbed by the load that terminates the port of the last circulator. Owing to the large frequency gap among the nonadjacent channels, the degradation due to reflections of channels is minimal. It can be readily seen that without the use of the input hybrid, we are forced to extract contiguous channels in the same path, simply because there is no alternative path anymore. Under such a constraint, there is significant degradation to the channel performance. The degradation, caused by reflections, is sometimes referred to as *en passant* distortion. It is dealt with in the next section.

The advantages of circulator-coupled MUX are similar to those of the hybrid branching MUX; namely, the circulator chains can be rendered very lightweight, compact, and reliable, and the channel filters can be designed and tuned individually. Their integration into the MUX subsystem has a negligible effect on their individual performance characteristics.

The losses are considerably less than those of the hybrid branching IMUX. Figure 18.13 presents a typical Ku-band waveguide circulator-coupled IMUX subsystem for a communication satellite. In this case, the filters are dual-mode bandpass with circulator-coupled external group delay equalizers.

18.3.3 En Passant Distortion

In the circulator-coupled multiplexing scheme, the transfer characteristics of the individual channel filters are not affected by their integration into the MUX subsystem. The channels themselves are affected by the reflections from the inputs of the noncorresponding filters in their path until they reach their own channel filter. The deviations introduced by such reflections in the channel characteristics are sometimes referred to as *en passant distortion*. The word *en passant* is a very apt description of such a distortion and is being gradually adopted. It is computed as follows.

The channel corresponding to the first filter in the chain (e.g., channel 1 in the right-hand branch of Figure 18.12) drops directly through to that filter's output without incurring any reflections from other filters in the chain. However, the second and subsequent channels are reflected off the input to the channel 1 filter and acquire the return loss characteristic of filter 1 before dropping through their own filter. This is expressed, approximately, by

$$S_{1c}(s) = S_{21}^{(1)}(s) \ \text{(dB)} \quad (\text{for } n = 1)$$
$$S_{nc}(s) = S_{21}^{(n)}(s) + \sum_{i=1}^{n-1} S_{11}^{(i)}(s) \ \text{(dB)} \quad (\text{for } n = 2,3,\ldots) \tag{18.1}$$

where

s = $j\omega$

n = physical filter position in chain, $n = 1$ is nearest to the common input, $n = 2$ is second along the chain, and so on

$S_{nc}(s)$ (dB) = overall transfer function of path from common input to the output of nth filter in chain

$S_{21}^{(n)}(s)$ (dB) = transfer function of nth filter in chain

$S_{11}^{(i)}(s)$ (dB) = return loss function of ith filter in chain

This formulation assumes that the circulators in the MUX are perfect, that is, they possess infinite directivity. This implies a transmission of signals in one direction and an infinite isolation in the reverse direction. The directivity of the circulators in the range of 30–40 dB is quite adequate for most practical MUXs to justify this assumption. The reflection response depends on the channel filter characteristics, as well as the frequency separation between the transmitted and reflected channels.

18.3.3.1 En Passant Distortion for a Noncontiguous Multiplexing Scheme

The IMUX configuration, described in Figure 18.12, requires noncontiguous channels in the two paths of the MUX. Taking the right-hand branch (odd-numbered channels) as an example, the channel 1 (Ch1) transfer function is that of the channel filter itself. Channel 3, en route from the common input, encounters a reflection from the channel 1 filter before it is extracted by the channel 3 filter. Thus, the transfer characteristic for the channel 3 signal through the IMUX is the return loss characteristic of the channel 1 filter plus the transfer characteristic of the channel 3 filter, expressed by

$$S_{3c}(s) = S_{21}^{(3)}(s) + S_{11}^{(1)}(s) \tag{18.2}$$

Channel 1 is unaffected by the presence of the other channel filters in the chain, but the downstream channels—Ch3 and Ch5—pick up the return loss characteristics of channel 1 and channel 1 + channel 3, respectively, in addition to their own transfer characteristics. This is illustrated in Figure 18.14a. For the noncontiguous case, the addition of the return loss characteristics of the other downstream channels in the chain has little effect on the in-band amplitude and group delay characteristics.

18.3.3.2 En Passant Distortion with Contiguous Channels

The reflection response for a contiguous channel is calculated in the same way as for that of the noncontiguous channels. We compute the reflection response over the frequency band, corresponding to the contiguous channel. This is described in Figure 18.14b, where channels 1, 3, and 5 are contiguous. Channel 1 is not affected by reflection, whereas channel 3 is reflected

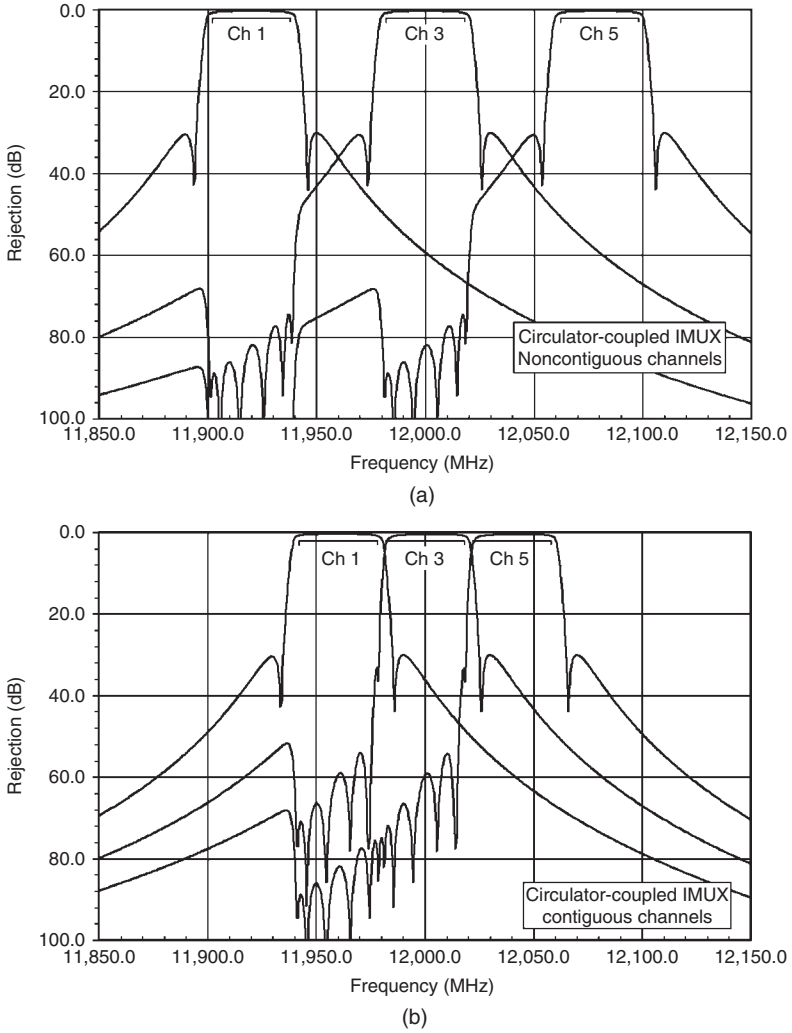


Figure 18.14 En passant distortion with (a) noncontiguous channels and (b) contiguous channels.

off channel 1. The lower edge of channel 3 is in close proximity to channel 1, and hence it suffers a large deviation in amplitude and group delay near the lower band edge as shown in Figure 18.15. The upper edge of channel 3 is far away from channel 1 and hence is relatively unaffected. Similarly, channel 5 suffers an asymmetric reflection from channel 3.

It should be noted that if the sequence of channels in the example is reversed, it is the upper edge of the channels that is adversely affected. The large deviations introduced by such reflections affect only one band edge, whereas the other edge remains relatively unaffected. As a consequence, the channel response is rendered asymmetric. Another way of looking at this is in terms of the rejection provided by the channel filters to the reflected channels. If the rejection at a given frequency is greater than 60–70 dB, there is virtually no impact on the reflected response. That is representative of noncontiguous channels. If the rejection is in the range of 5–10 dB, it can cause large deviations in the amplitude and group delay, as in the case of contiguous channels. As will be described in more detail in the following section on hybrid-coupled

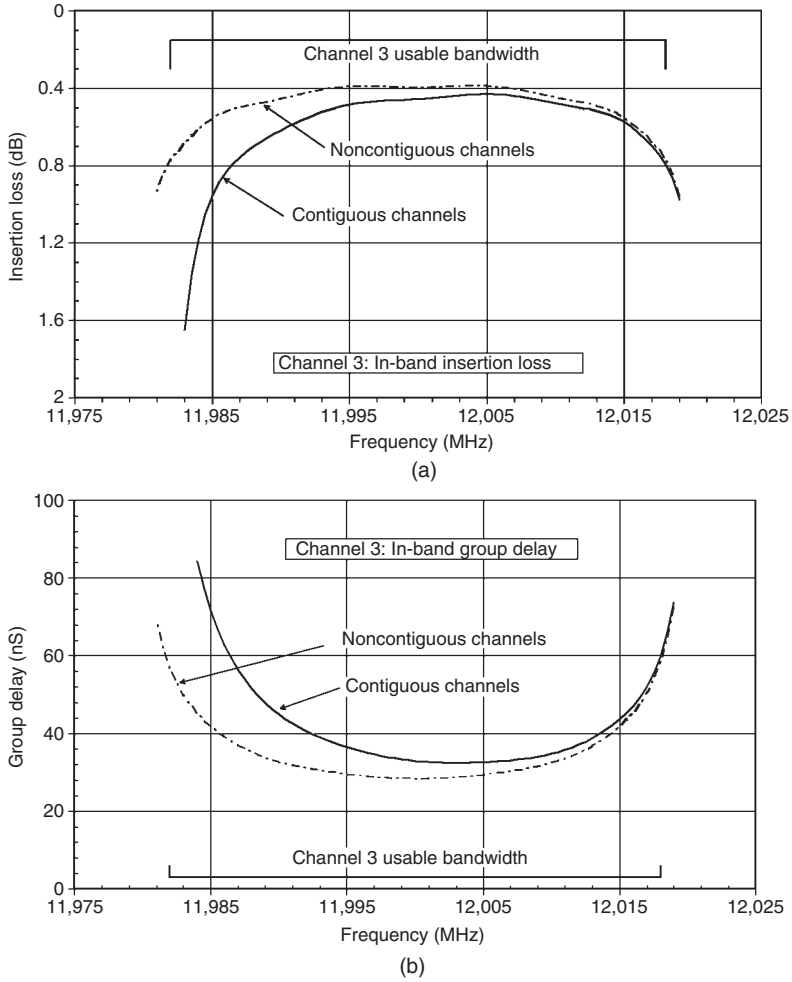


Figure 18.15 In-band distortions for channel 3: (a) insertion loss and (b) group delay.

directional filter module MUXs, these distortions can be countered to a certain extent with asymmetric filter designs.

18.4 RF Combiners

RF channel combiners, often referred to as *OMUX* or simply *MUX*, provide a function that is opposite that of RF channelizers. Such MUX is required at each transmit station in the system. The design constraints on such MUX are as follows:

1) The MUX must be able to combine the powers of a number of RF channels, which are contiguous or noncontiguous.
2) Insertion loss is a critical parameter, since it has a direct impact on the power radiated and thus the traffic capacity.
3) The equipment must be capable of handling high power in terrestrial or space environments.

4) The MUX must provide a high isolation in the receive band and suppress the harmonics generated in the HPAs.

The primary aim of an OMUX is to combine the power of a number of RF channels into a single port that can then be fed to an antenna. An antenna with a single-input port is simple to implement and its performance is simple to optimize. MUX design depends on the number of channels it handles and whether the channels are contiguous or noncontiguous. Above all, it is the power handling capability of an MUX and its operating environment that dictate the MUX design and implementation. Although all of the MUX configurations summarized in Table 18.1 are applicable for combining RF channels, it is the manifold-coupled or hybrid-coupled MUX that is used in most systems.

18.4.1 Circulator-Coupled MUX

Owing to the low-loss requirement, this configuration is viable for only a two-channel system with low to moderate power levels (tens to low hundreds of watts). It represents the simplest and cheapest configuration, but it does depend on the availability of a low-loss circulator capable of handling high power and low loss. The limitations of this configuration can be overcome by the hybrid-coupled network.

18.4.2 Hybrid-Coupled Filter Module (HCFM) Multiplexer

Instead of circulators, hybrid-coupled filter module (HCFM) derives its directional properties from an assembly of two identical bandpass filters and two 90° (quadrature) hybrids, shown schematically in Figure 18.16. In view of the scattering matrix of the 90° hybrid [21], we can readily show the following properties for such a module:

- Any out-of-band channels entering at port 3 are directed by the main-path hybrid toward the two-channel filters, are reflected off by the inputs to the two-channel filters, and are then recombined in the hybrid to emerge from port 4.
- Residual power that is not reflected passes diagonally through the module to port 2 and is absorbed by the termination at this port.
- In-band channel power entering at port 1 traverses the module diagonally and emerges at port 4, combining with any out-of-band channels that enter at port 3.

HCFMs are interconnected by arbitrary lengths of transmission line to form a MUX as shown in Figure 18.17. Within each module, the power entering at port 1 is shared by the twin filters,

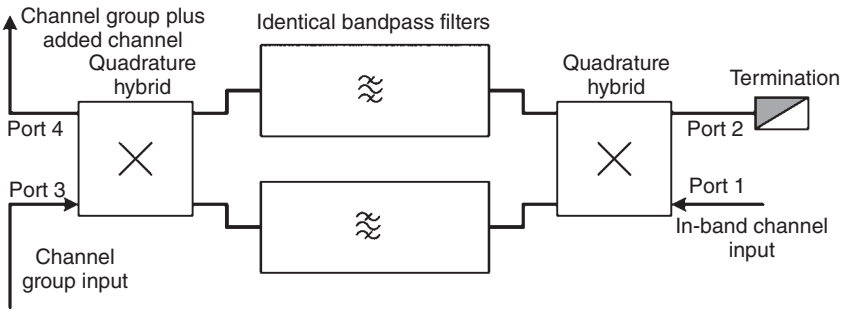


Figure 18.16 Hybrid-coupled filter module (HCFM).

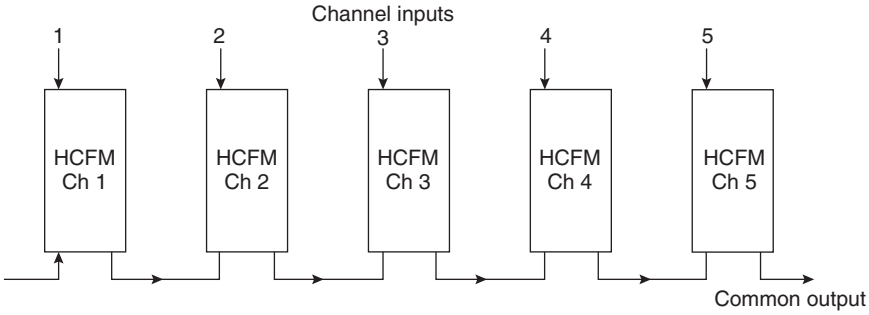


Figure 18.17 Output multiplexer with hybrid-coupled filter modules.

halving the thermal dissipation that each filter has to handle and reducing the peak voltages in the cavities by $1/\sqrt{2}$.

In waveguide networks, hybrids are typically short-slot *E*-plane couplers that are simple, compact, and internally open structures. They are ideally suited for the handling of the high-peak voltages of the combined channels in the common regions of the modules. In addition, broad wall branch line couplers can be used when the orientation of the input and output ports of the filters are at right angles to each other, for example, sixth-degree dual-mode filters. Figure 18.18 gives a possible internal structure of HCFMs with extended box filters (dual-mode realization) and with phase-coupled extracted pole filters (manufactured in a common housing).

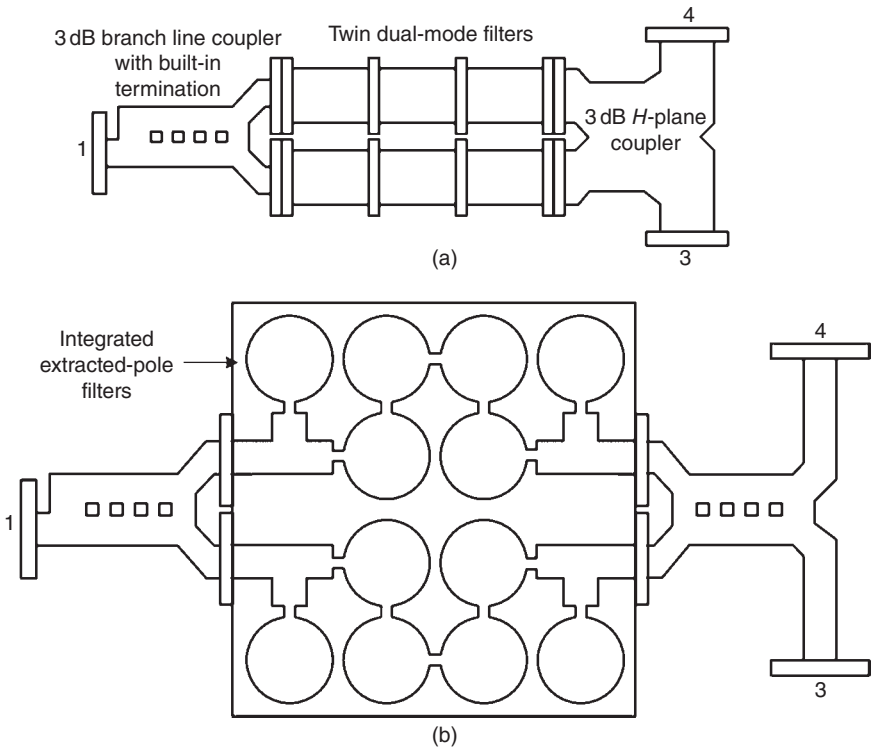


Figure 18.18 Possible HCFM configurations: (a) using dual-mode filters and (b) using extracted pole filters.

18.4.2.1 En Passant Distortion

A MUX consisting of HCFMs is similar in concept to the circulator-coupled MUX module. Here, the hybrids plus the twin filters perform the same function as the circulator and the filter. This being so, equation (18.1) may be used to characterize the transfer characteristics of each channel path, within the constraint that the hybrids used are perfect, that is, that their directivity is infinite. This corresponds to a similar assumption we made for the circulator-coupled MUX, where we assumed that the directivity of the circulator is infinite. A directivity of 30–40 dB for the hybrids is sufficient to justify such an assumption for practical systems. The transfer characteristic of each path is dominated by that of the channel filters of the module itself. However, the return loss characteristics of any other channel that may be physically located downstream from the module are also superimposed on the transfer characteristic. If it is required to combine the noncontiguous channels, this scheme provides an excellent way to accomplish it with a high-power handling capability. However, if there are any contiguous channels to be combined, then there are in-band insertion loss and group delay distortions near the band edges of the channels. The computed response for such distortions for contiguous channels is illustrated in Figure 18.15.

The large deviations in the insertion loss and group delay near the band edges renders this approach unacceptable for a practical system. For these reasons, it has been very rare to find contiguous or even semicontiguous channels in the same branch of an HCFM MUX. In satellite earth stations, even-numbered channels are combined in one branch and the odd-numbered channels in a second branch. Then the two branches are combined with a formidable hybrid whose termination must be water or forced air cooled, capable of dissipating many kilowatts of RF energy. Such an approach incurs an extra loss of 3 dB due to the hybrid, and that represents a big waste of energy since we are dealing with kilowatts of wasted power [22].

There are two ways to overcome the limitations of this design approach for contiguous channels. One way is to use manifold MUXs, described in the following section. Another way is to modify the design of the channel filters to provide a degree of compensation for the degradation of channels near the band edges and, at the same time, restore symmetry to the channel response. This requires quantitative tradeoffs between achievable performance for the channel and design complexity and sensitivity of the filter structure. This approach is described as follows.

Knowledge of filter types and design bandwidths (DBWs) of the downstream modules makes it possible to tailor the design of the filters in the modules, building in asymmetry where it is needed to compensate for a downstream neighbor. Table 18.2 provides a summary of these requirements.

The idea here is to try to keep the overall transfer characteristic as symmetric as possible by not providing built-in Tx zeros in the channel filters when the return loss characteristics of the downstream neighbors are effectively providing "virtual" Tx zeros for it.

Table 18.2 HCFM filter planning.

Downstream channels	HCFM filter solution
Noncontiguous channels or none at all	Regular symmetric with transmission zeros
One contiguous on one side of the passband	Asymmetric design with one or two Tx zeros, on side opposite that of downstream neighbor
Two contiguous on both sides	All-pole Chebyshev, downstream neighbors provide selectivity

Figure 18.19 shows the overall transfer function of an HCFM (Ch1) with a contiguous downstream neighbor (Ch2) and two others (Ch3 and Ch4) on the upper side of its passband. The filters of the HCFMs for Ch1–Ch3 are designed with one transmission zero (TZ) each on the lower side, while the HCFM for Ch4, which has no downstream neighbors, has two symmetric TZs. With contiguous downstream neighbors, it can be seen that a certain amount of symmetry is achieved for the close-to-band rejection characteristics of Ch1–Ch3. A high selectivity has been provided on the lower side by these filters' own built-in TZ and on the upper side by the return loss characteristic of the downstream neighbor. The in-band insertion loss and group delay characteristics for Ch1 are shown in Figure 18.19b and c, including a comparison with the performance of an equivalent symmetric filter (with two built-in symmetric TZs) and no downstream neighbor.

Figure 18.19d shows the effect on the rejection of Ch1 after we remove the Ch2 HCFM. In the absence of the downstream contiguous neighbor, the rejection characteristic of Ch1 tends

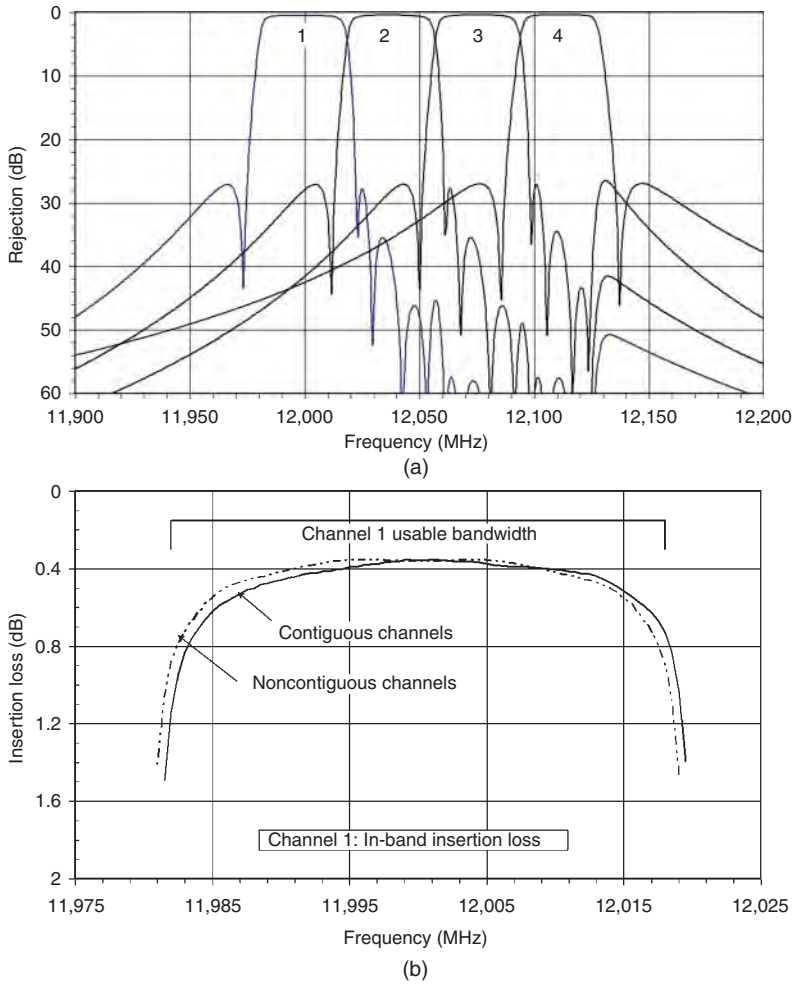


Figure 18.19 HCFM directional multiplexer with four contiguous channels: (a) overall transfer characteristics, (b) in-band insertion loss characteristic, (c) in-band group delay characteristic, and (d) effect on channel 1 rejection characteristic of removing channel 2 module.

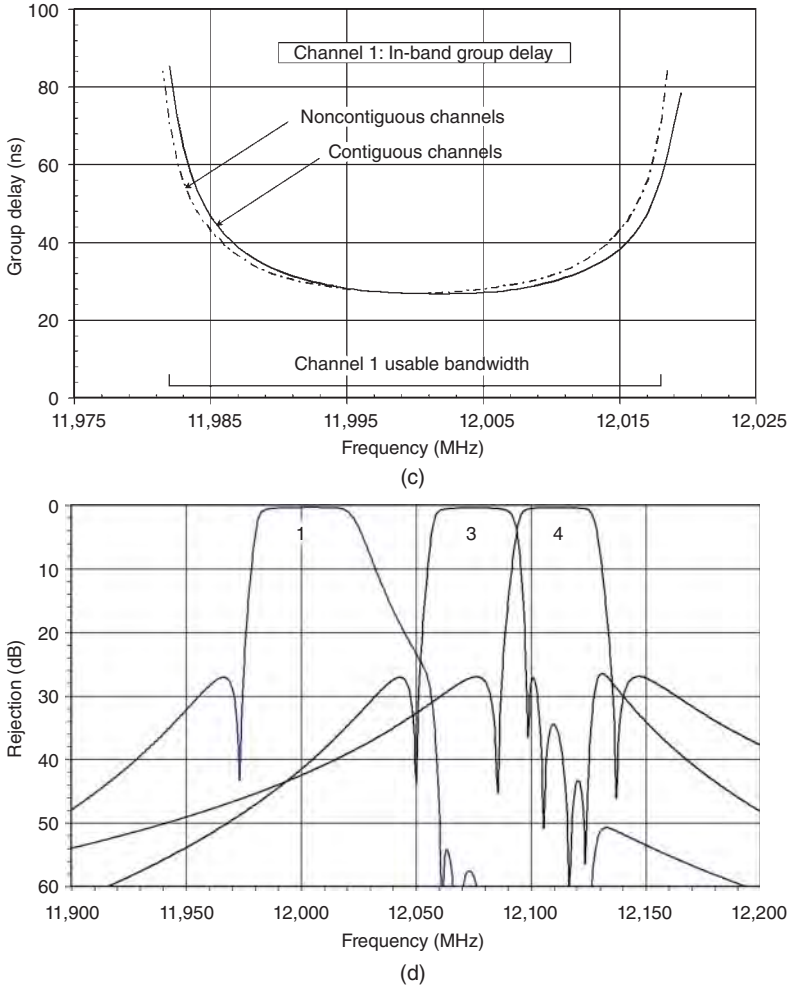


Figure 18.19 (*Continued*)

to "collapse" into the gap left by the missing module. This demonstrates Ch1's reliance on the presence of the neighboring channel for its own proper performance.

18.4.3 Directional Filter Combiner

The configuration of a MUX with directional filters is described in Section 18.2. Its advantages and disadvantages are summarized in Table 18.1. Because of the sensitivity of this structure and bandwidth limitations, this type of MUX is rarely used for high-power combining networks. An interesting aspect of this design is that the directional property of the filters can be switched by changing the polarization of the input to the filters. This feature has been exploited in satellite systems for narrowband multiplexing networks requiring switching between two antenna beams [6].

18.4.4 Manifold Multiplexer

A manifold MUX is characterized by channel filters connected to a common output port, by either a common junction or a common manifold. These two types are very similar in concept

and design. The principal difference is the presence of interchannel manifold phase lengths that separate the channel launch points in the manifold MUX. This provides the manifold MUX with extra degrees of freedom to optimize the design. Moreover, it is easier to separate the inputs to the common output transmission line, thereby rendering the structure less sensitive. In the star connection, the power of all the channels congregates at the common junction, which can cause very high voltages and a hot spot, whereas the power in the manifold of a manifold MUX tends to be more evenly distributed along the manifold. Often, the star connection is used where there are a relatively small number of channels, up to four, and is particularly favored for coaxial filter diplexing at lower frequencies and lower power where the junction can be built very compactly into the volume of the filter output cavities themselves.

18.4.4.1 Star Junction Multiplexer

Some star combining schemes are shown in Figure 18.20: a four-channel MUX with a regular star junction, a scheme where the channels are combined in stages, and the compact internal combination scheme of a Rx/Tx coaxial filter diplexer.

18.4.4.2 Manifold-Coupled Multiplexer

Three common manifold MUX configurations are shown in Figure 18.21, where the channel filters are connected to one side of the manifold (comb), to both sides (herringbone), and to end fed (applicable to either of the first two).

The design techniques for manifold MUXs underwent rapid development in the 1970s and 1980s when it was apparent that they were ideal for communication satellite payloads [3–6]. From the electrical perspective, design techniques have advanced to the point where it is possible to combine an arbitrary number of channels, regardless of their bandwidths and channel separations. In addition, there are no restrictions on the design and implementation of the channel filters in the manifold. The manifold itself is a transmission line, either a coaxial line,

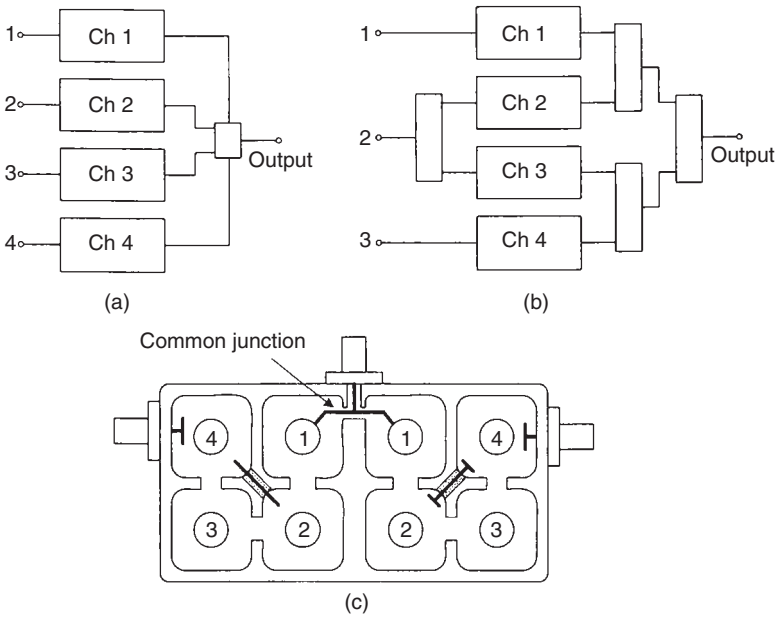


Figure 18.20 Types of star combining junctions: (a) basic star combining junction, (b) staged star junction, and (c) internal star junction in coaxial diplexer housing.

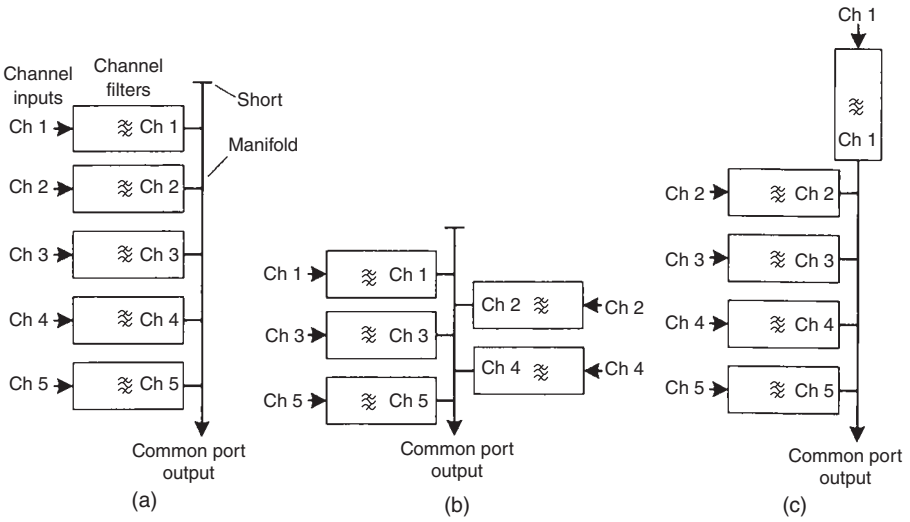


Figure 18.21 Common configurations for manifold multiplexers: (a) comb, (b) herringbone, and (c) one filter feeding directly into the manifold.

a rectangular waveguide, or some other low-loss structure. It is possible to achieve a channel performance in the MUX configuration that is close to that can be obtained from a channel filter by itself. No other MUX configuration can match this performance.

From the mechanical perspective, the MUX structure can be very lightweight and compact but, at the same time, rugged enough to withstand the vibrations and other rigors of a launch into space. By using special materials, the structure may be made to be electrically stable in the presence of large environmental temperature fluctuations and can conduct the dissipated RF thermal energy efficiently to a cooling baseplate.

Figure 18.22 shows a model of a typical C-band output MUX intended for space application. The short-circuit manifold supports twelve dual-mode channel filters arranged in a herringbone fashion (Figure 18.21b).

18.4.4.3 Design Methodology

Since there are no directional or isolating elements in the manifold MUX (circulators, hybrids), all the channel filters are electrically connected to each other through the near-lossless manifold. Moreover, the design of the manifold MUX needs to be considered as a whole, not as individual channels. In the early days, a number of ingenious techniques were invented [1–4, 7, 8] to design the individual filters such that they would properly interact with other filters on the same manifold and function virtually as if operating into a matched termination. Such quasi-analytic techniques provided a good starting point in the design process. However, for practical applications, it required extensive manual tuning to achieve acceptable performance. With the dramatic increase in computer power in recent years, MUX design procedures have utilized optimization methods to achieve the final design, in preference for the more limited analytic design techniques. Optimization techniques for MUX design have been developed to a high degree of sophistication, and worthy of note here is the space mapping technique developed by Bandler *et al*. [23], which involves optimizing a rough model (circuit or hybrid circuit model with embedded EM-modeled components) and connecting that coarse model to a fine model (e.g., full-wave electromagnetic modeling). Because of the large number of optimization variables, the fine model EM simulator takes a formidable amount of computer CPU time and

Figure 18.22 12 Channel C - Band Output Multiplexer with dual-mode filters. *Source*: Photo courtesy COM DEV Ltd.

has to be used sparingly during the optimization process. On the other hand, the rough model can be optimized by a circuit simulator very rapidly, especially if the S parameters of fixed and noninteracting elements (e.g., the manifold waveguide junctions) are modeled in advance by EM techniques and are available to the main program either as lookup tables or with the call of a rapid specialist routine. It therefore provides a means to simulate the manifold MUX with a good degree of accuracy and efficiency.

It is not the intention in this chapter to discuss the details of the EM simulation and optimization, which is thoroughly covered by Bandler *et al*. [23]. In the following sections some of the basic techniques needed for building the circuit model ready for optimization are outlined, followed by a description of the optimization strategy itself.

18.4.4.4 Analysis of Common Port Return Loss and Channel Transfer Characteristics

At the heart of any circuit optimizer is an efficient analysis routine. The optimization calls the analysis routine many times at different frequencies as it progresses with the task of optimizing the various parameters [10]. In general, the two parameters from which the overall cost function is built are the MUX common port return loss (CPRL) and the individual channel transfer characteristics.

Sampling the CPRL at frequency points corresponding to the in-band return loss zeros (points of perfect transmission) and return loss maxima (points of maximum in-band insertion loss ripple) of the original filtering function of each filter, with the aim of restoring them to their original RF frequency positions while the filter is on the manifold and is interacting with the other filters on the manifold, has been found to be effective in achieving a good overall CPRL. In addition, the two band-edge RL points are usually added, giving a total of $2N_k + 1$ frequency sampling points for each filter, where N_k is the degree of the kth filter on the manifold. Sometimes when close-to-band rejection is critical, two or four rejection corner points are added to the cost function. This necessitates that the transfer and reflection characteristics for each channel be known. Figure 18.23 shows the location of these sample points for a typical (5–2) quasielliptic filter.

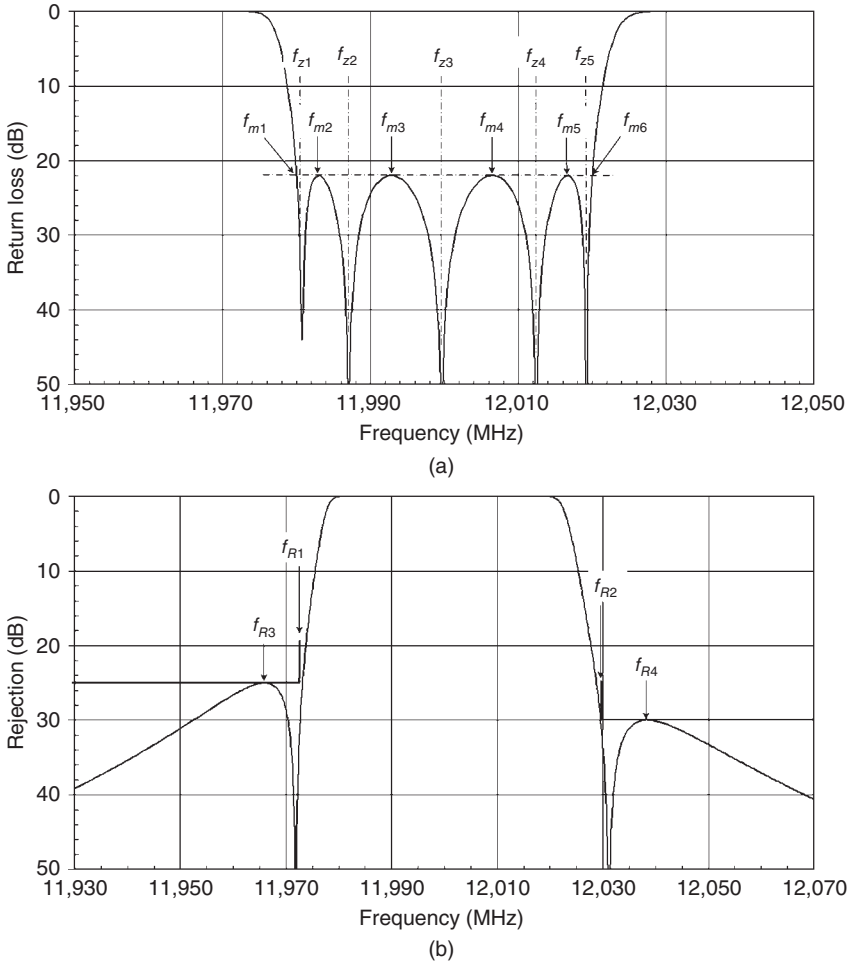


Figure 18.23 Typical sampling points for the evaluation of the optimization cost function: (a) return loss zeros and maxima and (b) close-to-band rejection points.

During the optimization process, the subroutines to calculate CPRL and channel transfer characteristics at each frequency sampling point is called many times. Although it is possible to construct an admittance matrix for the entire MUX circuit and analyze it at each frequency point to obtain the desired transfer and reflection data, this matrix is quite large for the MUX, and the large number of channel filters requires a significant amount of CPU time to invert the matrix even though it is relatively sparse. Moreover, much data is acquired that is not needed for the optimization.

Later on in this chapter, a piecewise optimization strategy, which is quite efficient in terms of computer CPU time, is described [12, 13]. Part of the strategy involves optimizing the channel filters one after the other in a repeated cycle. As the optimization parameters of each filter in turn are optimized, it is necessary to calculate only the transfer characteristic of this filter. Others are relatively unaffected by the changes (assumed small) being made to the optimized filter.

To speed up the overall optimization, it is more efficient to analyze each filter's input-to-common port transfer characteristic individually and the CPRL. The manifold of the MUX can

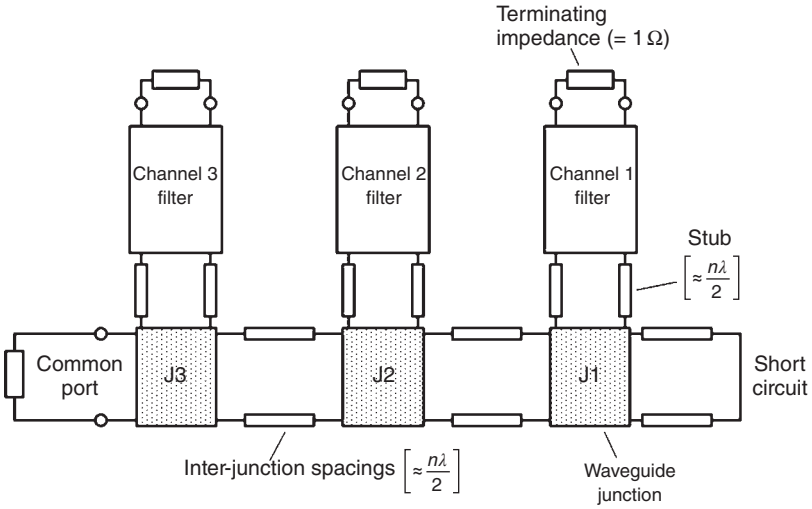


Figure 18.24 Open-wire model of waveguide manifold multiplexer (three channels).

be represented as an open-wire circuit, with a short circuit at one end and three-port *E*-plane or *H*-plane waveguide junctions spaced along the manifold's length. The channel filters are located at the third port of each junction and are separated from the junction by short lengths of transmission lines (stubs), as shown in Figure 18.24.

If the manifold is waveguide, the junctions are *E*-plane or *H*-plane. Since their intrinsic parameters do not change during the optimization, the junctions are best characterized with three-port *S*-parameters precalculated by a mode matching routine and stored over a range of frequencies that cover the bandwidths of all the channel filters or, specifically, at the frequencies of the sampling points. Typically, the junctions are symmetric about ports 1 and 2, leaving the possibility for port 3 to be of a different size as described in Figure 18.25.

Because of reciprocity and the symmetry of the junctions about the vertical axis in Figure 18.25, $S_{22} = S_{11}$ and $S_{32} = S_{31}$ (*H*-plane), and $S_{32} = -S_{31}$ and (*E*-plane). This means that only four parameters—S_{11}, S_{21}, S_{31}, and S_{33}—are needed to characterize the electrical performance of the junctions. The *S*-parameter matrices are calculated by assuming a matched

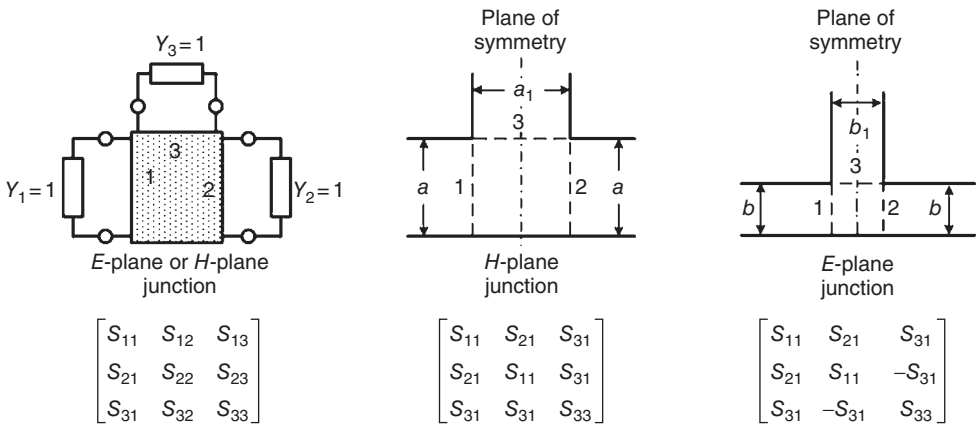


Figure 18.25 *E*-plane and *H*-plane waveguide junctions and *S*-parameter matrix representation.

termination at each port. However, if the termination at one port is arbitrary, the two-port S-parameters between the other two ports can be defined:

1) *Admittance* $Y_{L2}(\neq 1)$ *at Port 2*:

$$\begin{bmatrix} S'_{11} & S'_{13} \\ S'_{31} & S'_{33} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{bmatrix} + \frac{\Gamma_2}{1-\Gamma_2 S_{11}} \begin{bmatrix} S_{21}^2 & kS_{21}S_{31} \\ kS_{21}S_{31} & S_{31}^2 \end{bmatrix} \tag{18.3a}$$

where $\Gamma_2 = (1 - Y_{L2})/(1 + Y_{L2})$ and Y_{L2} is the admittance at port 2 of the junction and $k = 1$ for H-plane junctions, while $k = -1$ for E-plane junctions. This modified S matrix also applies if an admittance Y_{L1} is terminating port 1. It is necessary only to change the subscripts from 2 to 1 and vice versa.

2) *Admittance* $Y_{L3}(\neq 1)$ *at Port 3*:

$$\begin{bmatrix} S'_{11} & S'_{12} \\ S'_{21} & S'_{22} \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} + \frac{\Gamma_3 S_{31}^2}{1-\Gamma_3 S_{33}} \begin{bmatrix} 1 & k \\ k & 1 \end{bmatrix} \tag{18.3b}$$

where $\Gamma_3 = (1 - Y_{L3})/(1 + Y_{L3})$, Y_{L3} is the admittance at port 3 of the junction, and k has the same meaning as before.

18.4.4.5 Common Port Return Loss

CPRL computation at a given frequency point (e.g., at a sample point) proceeds as follows:

1) The channel filter input admittances Y_{F1}, Y_{F2}, and Y_{F3} are determined at the frequency sample point and stored.
2) With the known admittances at port 3 of each junction, the transfer and reflection S-parameters [S'_{21}, S'_{11}; see equation (18.3b)] for each junction are calculated and converted to [$ABCD$] matrices. From the short-circuit end, the manifold phases and the junctions are cascaded at each stage for finding and storing the along-manifold admittances Y_{M1}, Y_{M2}, … , as shown in Figure 18.26:

$$Y_{Mi} = \frac{1 + S_{11i}}{1 - S_{11i}} \tag{18.4}$$

where $i = 1,2, \ldots, n+1$ and n is the number of channels on the manifold.

The final admittance (Y_{M4} in Figure 18.26) may be used to calculate the CPRL as

$$\Gamma_{\text{CP}} = \frac{1 + Y_{Mn+1}}{1 - Y_{Mn+1}}$$
$$\text{RL}_{\text{CP}} = -20\log_{10}\Gamma_{\text{CP}} \ \text{(dB)} \tag{18.5}$$

If only the manifold spacings θ_{M1}, θ_{M2}, … are being optimized, then the filter input admittances Y_{F1}, Y_{F2}, … need to be calculated only once at each sample frequency and stored for optimization of the CPRL.

18.4.4.6 Channel Transfer Characteristics

If optimization is now focused on an individual channel filter and the transfer function between its input and the common port is needed (e.g., for a rejection sample point), the following procedure may be used. Let us consider channel 2 as an example, as shown in Figure 18.27:

1) Calculate the new Y_{F2} for channel 2 with its newly updated parameters.

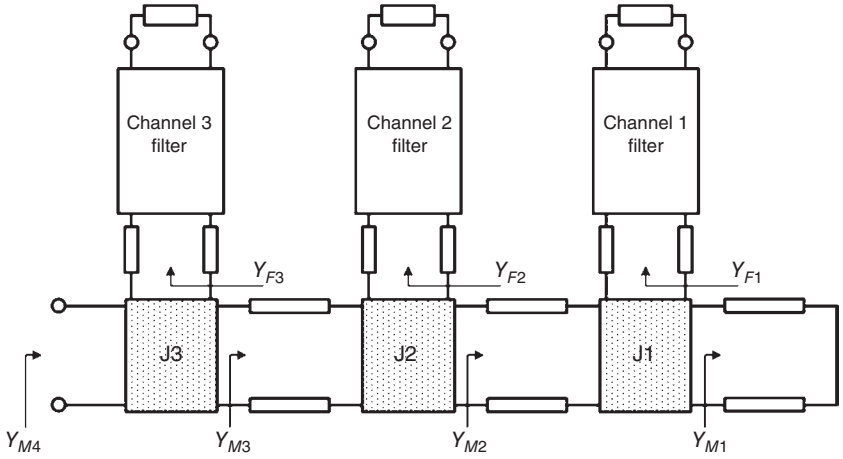


Figure 18.26 Common port return loss computation.

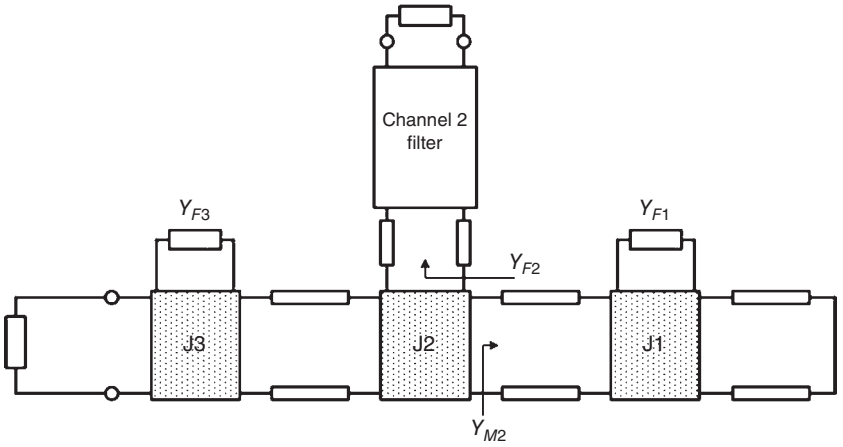


Figure 18.27 Channel 2 transfer function calculation.

2) The previously stored value of Y_{M2} is used in conjunction with equation (18.3a) to calculate S_{31} and S_{11} for junction 2 with Y_{M2} as its port 2.
3) The [*ABCD*] matrices for the channel 2 filter and the S_{31} path of J2 and then the rest of the manifold toward the common port are cascaded to calculate the transfer characteristic for channel 2. The process can be repeated for the other channels.

For both the common port reflection and the transfer characteristics, the computations are conducted by assuming lossless (purely reactive) components for the filter plus the stub networks. This allows more efficient routines for matrix multiplications, inversions, and other calculations.

18.4.4.7 Channel Filter Design

Being purely reactive (i.e., with no resistive elements between the channel inputs and the common port output), the channel filters of the manifold MUX react with each other through

the manifold itself. If the channels are widely spaced, the channel interactions are quite low, because, over one filter's passband, the other filters are well into their reject regions and present short circuits at the ports closest to the manifold. The channel filters are designed as doubly terminated networks, separately from the other filters and the manifold. The filters are integrated on the manifold; all that is needed are adjustments to the spacings and stub lengths along the manifold, and minor adjustments to the first three or four elements of the filter, to recover a good CPRL.

However, as the guard bands between the channel filters decrease toward contiguity, the filters begin to interreact strongly along the manifold. Now significant adjustments to the filter parameters are needed in addition to the manifold and stub phase lengths to reach an acceptable CPRL. Although it is possible to optimize doubly terminated filters to operate in a contiguous channel environment, a starting point much closer to the final optimal result is obtained if singly terminated (ST) filter prototypes are used in these conditions. The design methods for ST filter prototypes are outlined in Chapter 7 [1, 24]. The ST filter network is useful for the design of contiguous channel manifold MUXs, because the contiguous ST channel filters, along the manifold, tend to interact in a mutually beneficial manner, providing a conjugate match for each other at their launch points. This "natural multiplexing" effect can be explained by studying the special characteristics of the input admittance of the ST circuit, looking in at the port opposite to the terminated port.

The real part of the input admittance has the same characteristic as the filter's own power transfer characteristic. It is close to unity over the filter's passband, dropping to near zero in the out-of-band regions. As the frequency increases from minus infinity, the imaginary part of the admittance increases from zero toward a positive peak near the lower passband edge and then traverses the passband with a negative slope toward a negative peak near the upper passband edge. It then slowly returns to zero with a positive slope as the frequency goes to infinity. Figure 18.28 illustrates the variations of the real and imaginary parts of the admittance for ST and doubly terminated prototype filters.

If the contiguous band ST filters are connected to the manifold with the "zero impedance" termination closest to the manifold junction, then the negative in-band slope of the imaginary part of the admittance tends to cancel with the positive slopes of the two contiguous neighbors that extend into this filter's passband. The results are illustrated in Figure 18.29, which shows that the imaginary parts of the source (the ST filter) and the load (the other filters on the manifold) have been mostly canceled, whereas the unity-valued real part of the source mostly sees the unity load at the common port of the manifold. This is due to the impedances at the input ports of the other channel filters being almost isolated by their own rejection characteristics.

Thus, a conjugate match between the source and the load is partially achieved. Although it is not a perfect conjugate match, it is better than that if doubly terminated filters were used, which makes it a good point at which to start the optimization process. The channels at the edges of the contiguous group on the manifold, with a neighbor on one side only, have some small in-band and rejection asymmetries after the optimization process.

18.4.4.8 Optimization Strategy

The networks that model even a moderate-sized waveguide manifold MUX tend to be quite complex. An open-wire equivalent circuit of a six-channel manifold MUX with sixth-degree quasielliptic dual-mode filters has in the order of 90 frequency sampling points and 100 electrical elements of varying sensitivities and different constraints. They all need to be correctly valued before the overall MUX can operate to specification.

If all these parameters are optimized simultaneously, not only the amount of CPU time is enormous but also there is little likelihood that the global optimum can be attained; there will

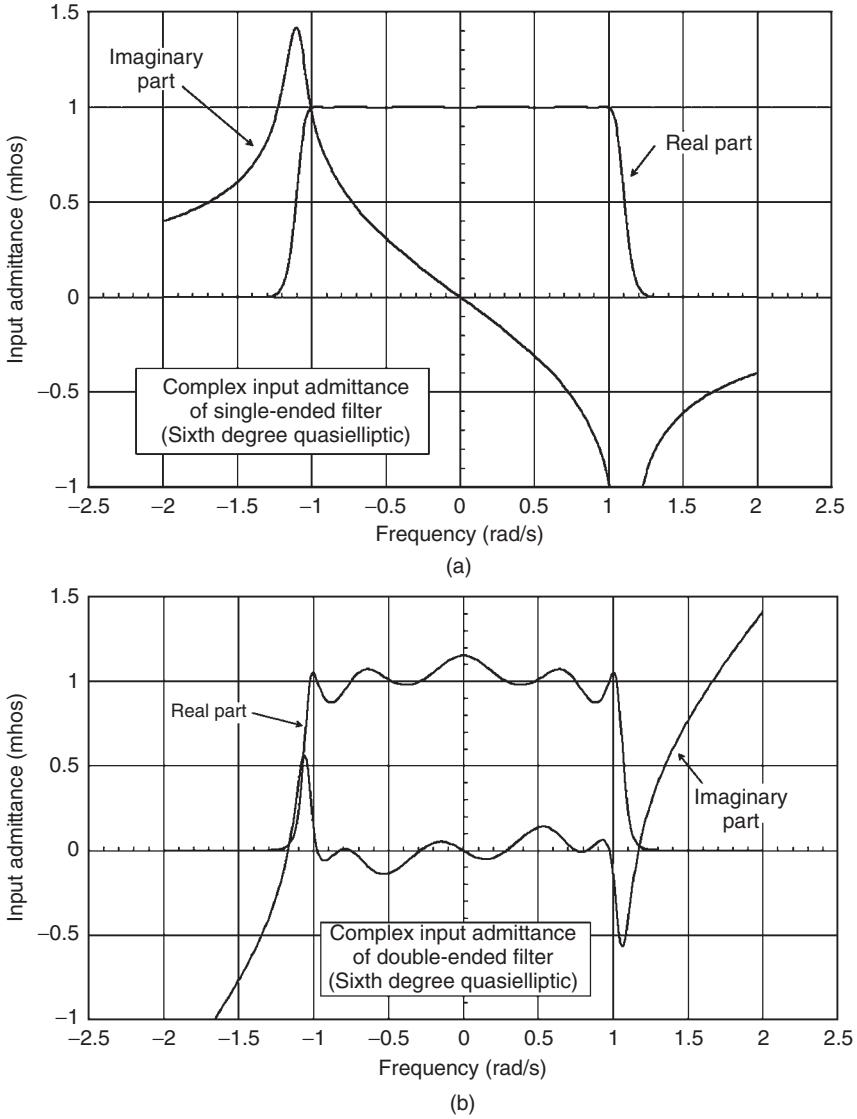


Figure 18.28 Real and Imaginary parts of input admittance of a sixth order prototype quasielliptic filter, (a) singly terminated filter, (b) doubly terminated filter.

be a myriad of shallow local optimum solutions. With manifold MUXs routinely incorporating 20 channels, and perhaps up to 30 in the future, global optimization with so many variables is clearly unsuitable.

For these reasons, most of the major satellite OMUX designers and manufacturers have developed more efficient proprietary methods for manifold MUX optimization. Among these is the "piecewise" approach, optimizing parts of the MUX separately in repeated cycles while converging upon an optimal solution. The "parts" or "parameter groups" referred to here might include the first five elements of each channel filter (narrowband domain) or all the manifold stub lengths (wideband domain). It is usual practice to commence the optimization process with the wideband sections first (parameters relating to the manifold and stubs), followed by

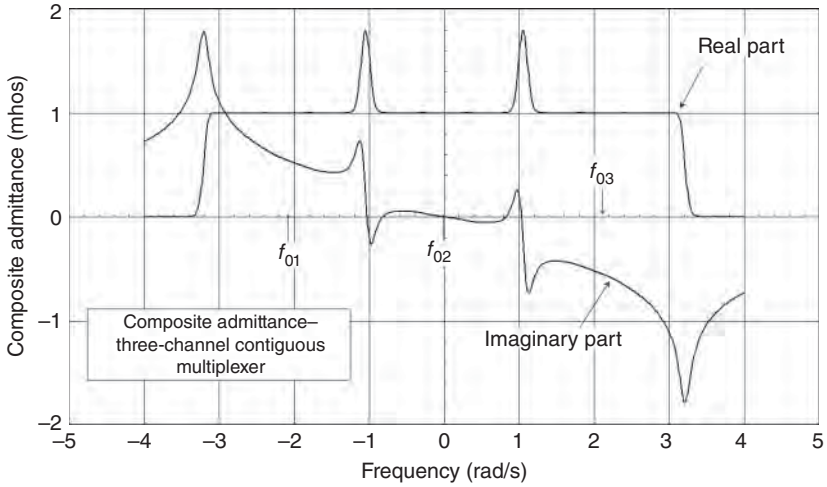


Figure 18.29 Singly terminated filter with two contiguous neighbors—composite admittances as seen from the common port of the manifold.

a shift in emphasis to the narrowband sections (filter parameters) as the CPRL begins to take shape. A typical design optimization project can proceed as follows:

Design:

1) Design the channel filter transfer/reflection functions to meet the individual in-band and rejection specifications.
2) Synthesize the corresponding coupling matrices, doubly terminated if the channel filter design bandwidths (DBW) are separated by guard bands greater than about 25% of the DBWs, ST if otherwise. It may be shown that if an ST filter is presented with a conjugate match at its unterminated port (i.e., the port closest to the manifold), its design return loss (in dB) will double. So if a design RL of 9–10 dB is chosen for the ST filter prototype, this will double to about 20 dB as the approximate imaginary admittance cancellation provided by the adjacent contiguous filters, together with the load at the output of the manifold, tend to provide a conjugate match for the channel filter. Setting such a low design RL for the ST prototype has the additional advantage of reducing its input coupling value (the coupling furthest from the manifold) closer to unity, giving a value nearer to that of the doubly terminated equivalent.
3) Set the initial manifold spacings between the *E*- or *H*-plane junctions at $m\lambda_g/2$, where m, an integer, is as low as possible for a convenient mechanical layout. Set the initial manifold short-circuit/first junction spacing at $\lambda_g/4$ (*H*-plane) or $\lambda_g/2$ (*E*-plane). λ_g is the wavelength in the manifold waveguide at the center frequency of the filter nearest to the length of the waveguide, in the direction of the common port.
4) Set the initial manifold junction with filter stub lengths at $n\lambda_g/2$. Again, n, an integer, should be as small as possible.

Optimization:

1) Wideband components. Optimize the spacings between the junctions and between the short circuit and the first junction, and the stub lengths. This often has the most dramatic effect in terms of an improvement in the CPRL.

2) Optimize the stubs and first three or four parameters of the channel 1 filter [M_{S1} (filter manifold coupling), M_{11} (first resonance tuning), and M_{12} (resonance 1 to resonance 2 coupling)].
3) Repeat the cycle for all the channels, possibly omitting the stub lengths since these do not change much, until the improvements in the cost function are beginning to become negligible.

Refinement:

1) Repeat the optimization of manifold and stub lengths.
2) Reoptimize with a fine step on all of each channel filter's parameters, most lightly on the elements farthest away from the manifold, and not at all on the final coupling M_{NL} (i.e., the input coupling to the MUX filter).

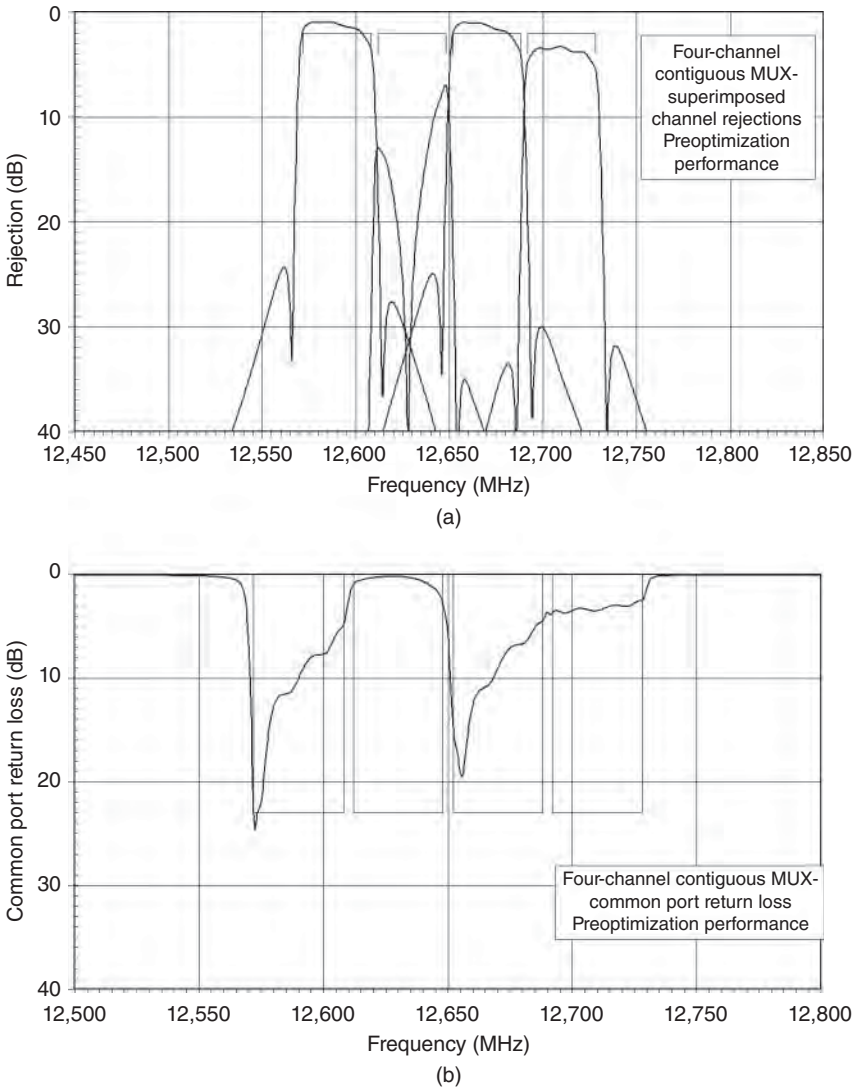


Figure 18.30 Preoptimization performance of a four-channel contiguous manifold multiplexer, (a) superimposed channel amplitude response, (b) common port return loss.

18.4.4.9 Demonstration of the Piecewise Optimization Process and Stages in the Optimization of a Four-Channel Contiguous OMUX

The piecewise optimization process is demonstrated through the optimization of a four-contiguous-channel Ku-band waveguide manifold MUX. The channel electrical specifications are satisfied with (5–2) quasielliptic filters with 30 dB rejection lobe levels and DBWs of 38 MHz and 40 MHz center frequency spacings that fall within the definition of contiguity.

Designing the filters as ST prototypes and attaching them to the manifold with the initial manifold spacings and stub lengths results in a rather disappointing performance as shown in Figure 18.30. The CPRL is poor, and one of the channels is unrecognizable.

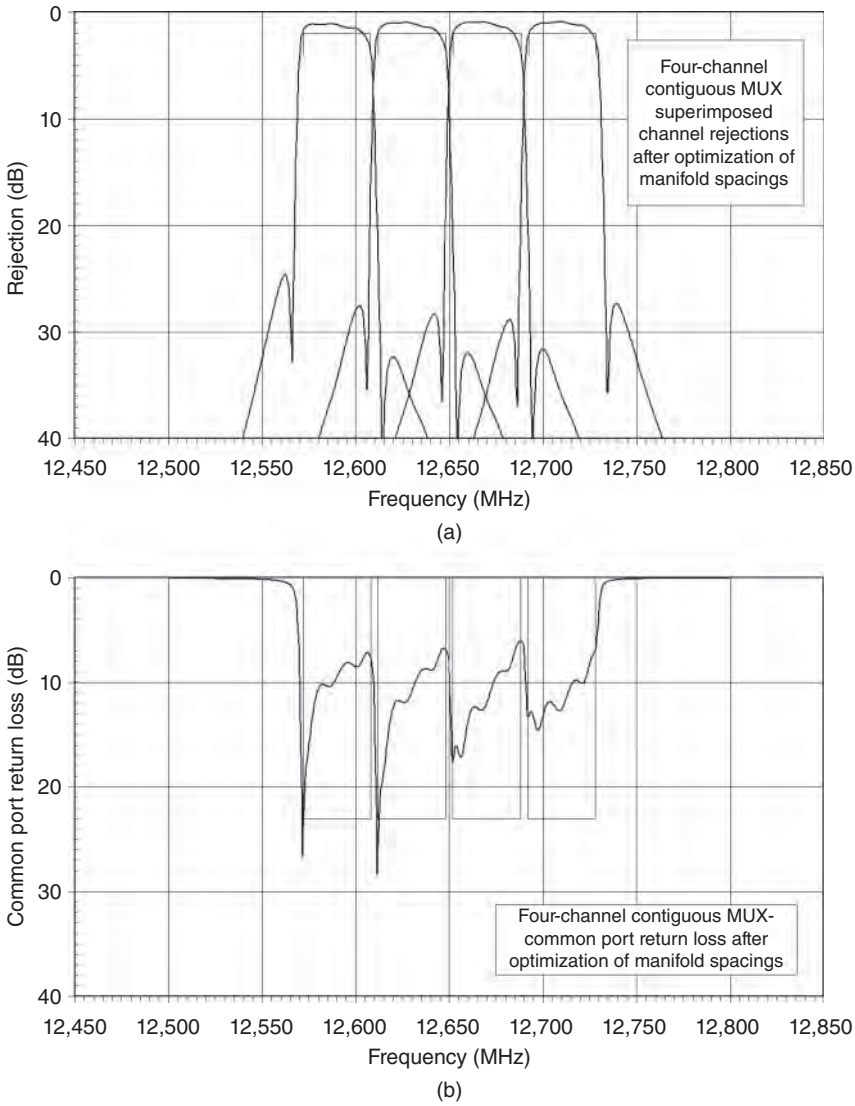


Figure 18.31 Four-channel manifold multiplexer after optimization of manifold lengths: (a) superimposed channel transfer characteristics and (b) common port return loss.

Figure 18.31 shows the dramatic improvement that results from the optimization of the manifold spacings. Now the channel rejection characteristics are close to design, and the average CPRL is in the order of 10 dB.

Further improvement in the CPRL is obtained by optimizing next the first four parameters (first coupling iris M_{S1}, first cavity tuning state M_{11}, second coupling iris M_{12}, second cavity tuning state M_{22}) of each filter. The computed response following this procedure is shown in Figure 18.32. The inner rejection lobe levels are close to the design of 30 dB, but the outer lobe levels, because there are no neighbors on the outer side of the group, have risen to about 24 dB.

Then, the refinement optimization cycles are carried out, and the final result is shown in Figure 18.33. Now the CPRL is above 23 dB over all the channel bandwidths. The effect of not

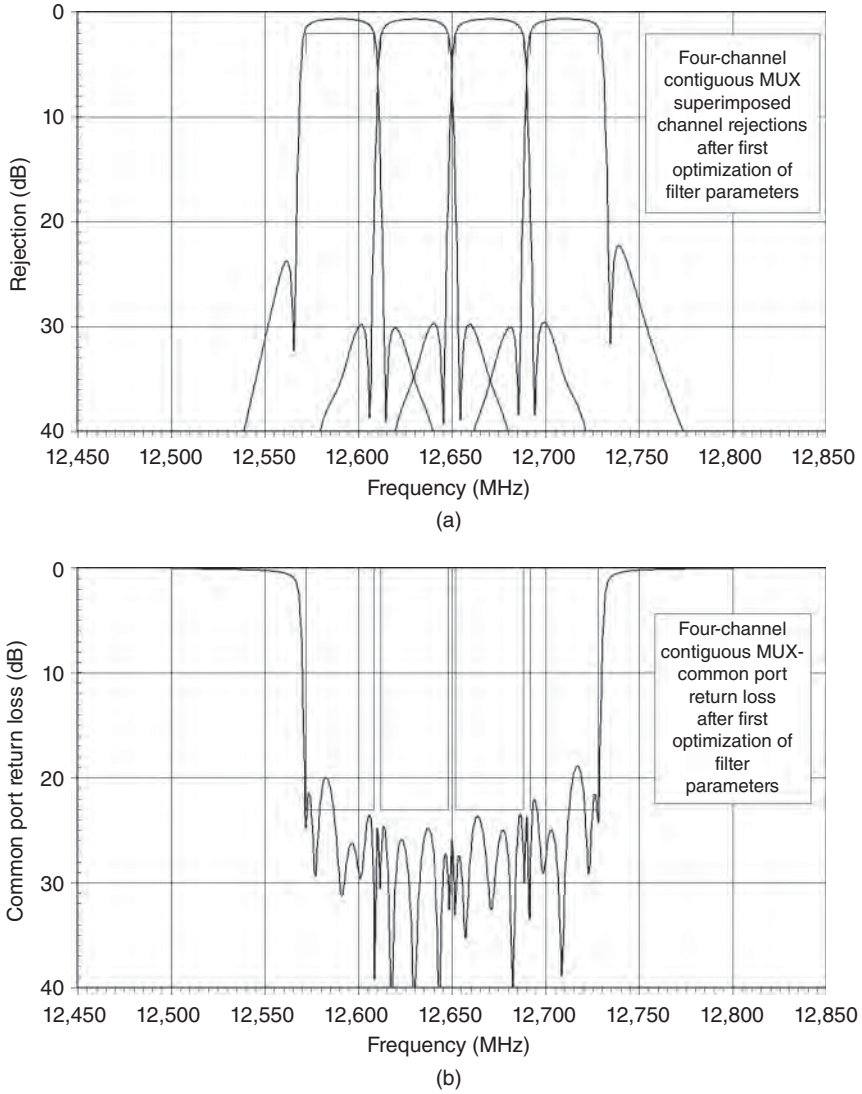


Figure 18.32 Four-channel manifold multiplexer after the first optimization of filter parameters: (a) superimposed channel transfer characteristics and (b) common port return loss.

having a contiguous neighbor on one side is evident in the rejection responses of channels 1 and 4 and the lobe levels. The rejection slopes on the outer sides are not as steep as the sides with a contiguous neighbor. This causes a small amount of asymmetric in-band distortion to the group delay and insertion loss.

The simulations have all been conducted by assuming a Q_u of 12,000, but the optimizations were carried out with a lossless network to speed up the process. With a moderate-speed PC (700 MHz), the whole optimization process took approximately 2 min.

This optimization strategy appears to work well even for a larger number of channels. Figure 18.34 shows the CPRL and rejection characteristics of a manifold MUX with 20 contiguous channels at the C-band (fourth-degree filters), which was designed using the piecewise optimization process.

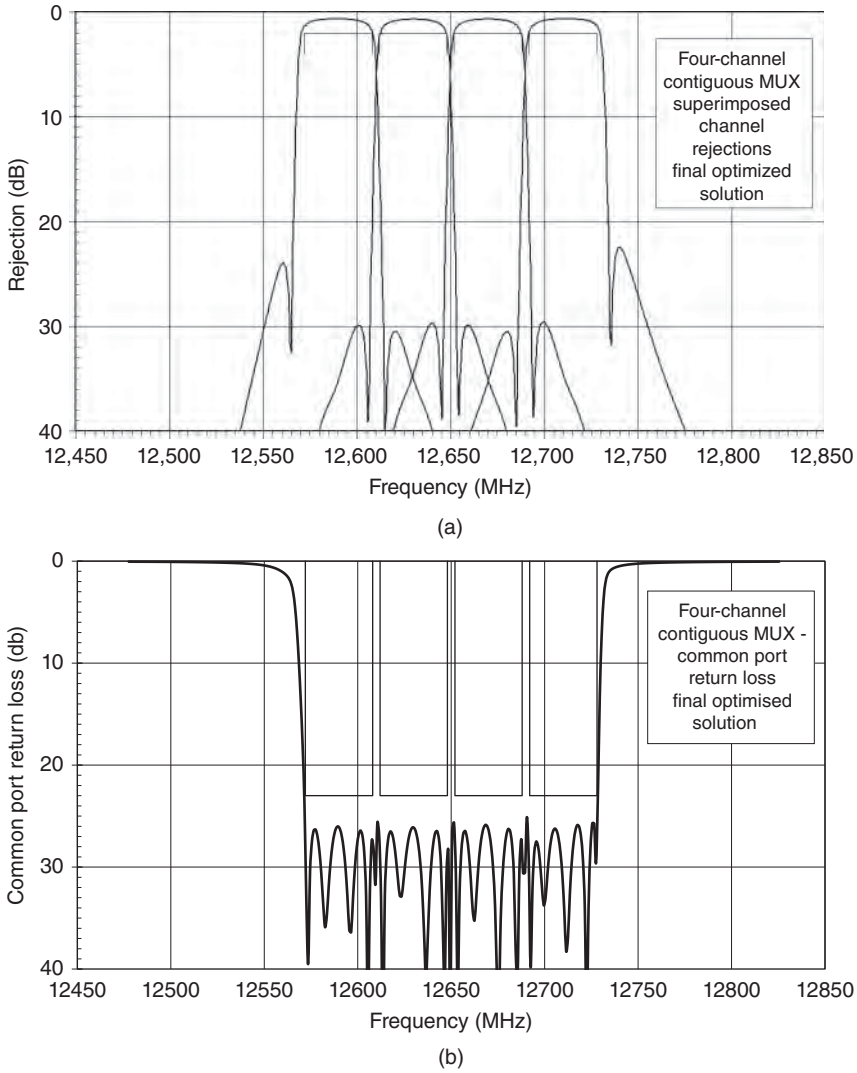


Figure 18.33 Four-channel manifold multiplexer and final performance: (a) superimposed channel transfer characteristics and (b) common port return loss.

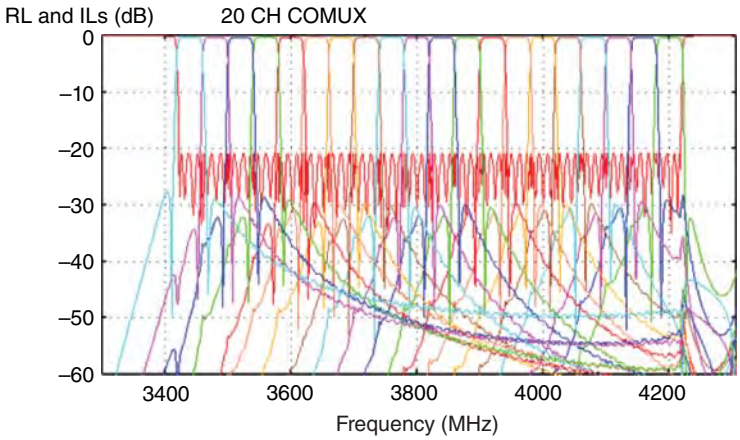


Figure 18.34 Twenty-channel manifold multiplexer superimposed channel transfer characteristics and common port return loss (RL is return loss and IL is insertion loss). *Source*: Courtesy Dr Yu, COM DEV Ltd.

18.5 Transmit–Receive Diplexers

The terms *diplexer* and *duplexer* have been used interchangeably by RF engineers for many years. The prefix *di* is defined as twice or double, while the prefix *du* is defined as two or dual. The term *plex* is generated from the Latin word *plexus*, which means an interwoven combination of parts or elements. Thus, there is no basic difference in the meaning of the two terms. Some RF engineers prefer to use the term *duplexer* for the case where the antenna is shared between the receive and transmit signals through the use of either a switch or a circulator. Others use the term *diplexer* for the same device. In view of IEEE publications on the subject over the past 20 years, one can clearly see that the two terms are often used interchangeably. In this chapter, we will use the term *diplexer*. A clear distinction exists between a diplexer, which combines the receive and transmit functions, and a simple two-channel MUX, which may be required for separating or combining two channels.

A simple form of a diplexer is shown in Figure 18.35. It consists of two 90° hybrids and two identical filters. They are designed to operate within the receive band with enough isolation to reject the transmit band. The diplexer allows the use of one antenna for the receive (Rx) and transmit (Tx) signals. In view of the scattering matrix of the 90° hybrid [21], we can readily show that the signal received from the antenna port is sent to the receive port, while the signal from the transmit port is reflected from the filters and directed to the antenna port. This diplexer configuration is bulky, requiring the use of two hybrids. However, it offers design simplicity.

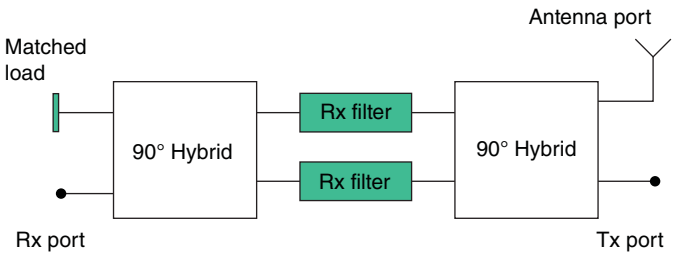


Figure 18.35 A diplexer implemented using the 90° hybrid approach.

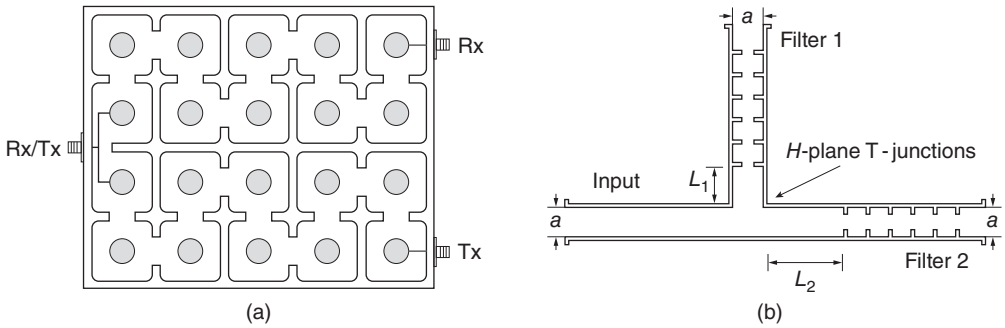


Figure 18.36 Transmit–receive diplexer configurations: (a) implemented using two coaxial filters and a wire T junction and (b) implemented using two waveguide iris filters and T waveguide junction.

Another approach to realizing the diplexer is to utilize one filter for the receive channel and another one for the transmit channel. The two filters are then combined by a T junction. Figure 18.36a illustrates two coaxial filters, combined with a wire T junction, while Figure 18.36b shows two iris waveguide bandpass filters combined together by a waveguide T junction. The interaction between the two filters, in this case, is very strong, requiring that the phase at the junction arms be adjusted to ensure that the receive and transmit signals are integrated properly. In some applications, such as in radar systems, the receive and transmit signals do not exist simultaneously. A fast-acting switch is used to replace the T junction, allowing the antenna to be connected to either the receive or the transmit circuits.

In diplexer applications, the insertion loss of the receive filter is an important factor, since the loss contributes directly to the overall noise figure of the receiver that follows the diplexer. The insertion loss of the transmit filter is equally important, since the loss impacts the power transmitted and the efficiency of the transmit system. A tradeoff exists between the loss, rejection, and size of the filter. High rejection with minimal effect on insertion loss and size can be achieved by using filters with asymmetric response. This allows the allocation of several attenuation poles in the appropriate band to increase the filter rejection. Typically, the insertion loss of the 90° hybrid design is slightly higher than that of the T-junction design. However, the 90° hybrid approach offers a higher-power handling capability because the power handling capability of the T-junction approach is determined by the power handling capability of the transmit filter alone, whereas, in the hybrid approach, the power is handled by the reflection of the two identical filters.

Tx/Rx diplexers are found in those systems where the antenna is used for transmitting and receiving signals simultaneously (as opposed to *half-duplexing* or a *simplex* system, where the receiving and transmitting are separated in time). In radar systems the Tx and Rx signals are seldom present simultaneously as they are in communication systems, and a fast-acting electronic switch (switching ferrite circulator) is often used to isolate the sensitive low-noise receiver from the high-power transmitter. In satellite communication transponders, a wideband Tx/Rx diplexer is sometimes found to cover the bandwidths of the Tx channel group in one filter and the Rx channel group in the other filter for up/downlinking by one antenna system. However, the most demanding specifications are placed on the Tx/Rx diplexers in the base stations of cellular telephony systems, where the Tx and Rx channel groups are very closely spaced.

When diplexing for a Tx/Rx system, it is essential that a very high isolation be maintained between the Tx and Rx paths, perhaps up to 120 dB, to prevent the Tx power, its harmonics, or IM products from reaching the sensitive LNA units. This means not only that there

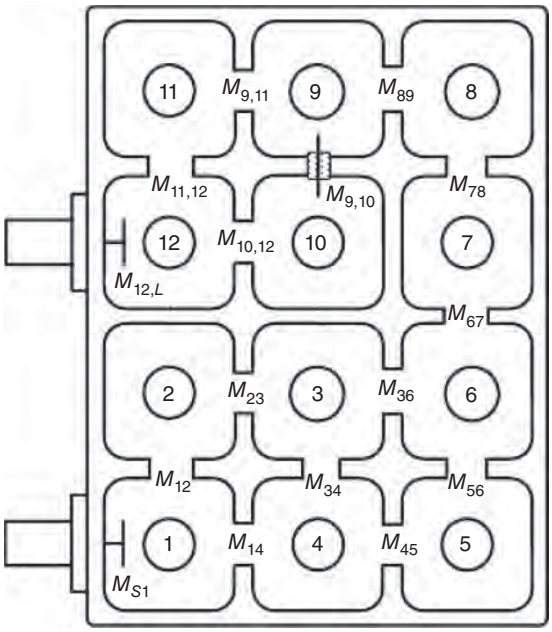


Figure 18.37 Outline sketch of a (12–3) filter realized with one sixth-degree extended box section (two TZs) and one fourth-degree box section (one TZ).

must be a very high intrinsic rejection by the Tx filter over the Rx bandwidth (and vice versa) but also that a very high build standard must be imposed in order to prevent passive intermodulation (PIM) products from being generated in the high-power regions of the diplexer. Since the rejection requirements are usually less severe on the outer sides of the filters, filters with asymmetric characteristics are often used when the Tx and Rx bands are very close, minimizing the insertion losses and overall mass, volume, and complexity. To attain a 100 dB isolation between the closely spaced receive and transmit bands of a cellular telephony system, 12th-degree coaxial filters with three TZs have been used for the filters of the base station diplexers; Figure 18.37 is a sketch of such a diplexer filter, and Figure 18.38 gives its measured response. Resonators 1–6 form a sixth-degree extended box section that realizes two of the TZs, and resonators 9–12 form a single box section realizing one more TZ. There are still some negative couplings, but these are "straight" rather than diagonal (e.g., $M_{9,10}$). Elimination of the diagonal cross-couplings reduces manufacturing complexity, resulting in improved reliability and better yield in the volume production of such diplexers.

At L- and S-band frequencies (most common for mobile communications), the reentrant coaxial resonator is usually employed for construction of the filter resonators with the star-connected common junction located within the first cavities of the two filters. This leads to a very compact structure with line lengths quite short compared with the wavelengths. Figure 18.20 gives an outline sketch of a coaxial diplexer, consisting of two asymmetric (4–2) filters, and shows the star junction for combining the two filters.

18.5.1 Internal Voltage Levels in Tx/Rx Diplexer Filters

A major system constraint in the design of Tx/Rx diplexers is the required power handling capability of the transmit section. The design must ensure that no electrical discharges are initiated under high power. Such discharge phenomena include arcing in the presence of a gas (e.g., in high-power terrestrial TV broadcast systems), corona under partial-pressure conditions (e.g.,

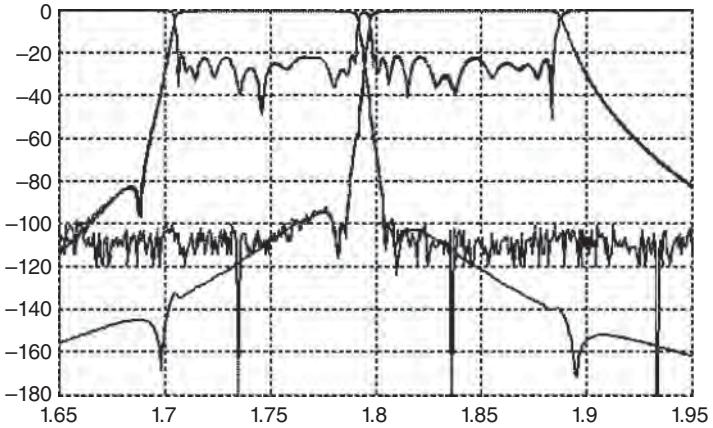


Figure 18.38 Simulated and measured rejection characteristics of a transmit–receive diplexer (duplexer) of a cellular telephony base station. *Source*: Courtesy of COM DEV Ltd.

for equipment in missiles or operating in a rocket launcher), or multipactor effect in vacuum (in spacecraft). These high-power effects are discussed in Chapter 20.

The peak voltages in the resonators of a microwave filter can be estimated at the prototype design stage, ensuring a proper mechanical design later with the dimensions of the gaps in the resonator cavities sufficient to withstand the peak voltages during operation. One procedure for determining the resonator voltages based on the $N \times N$ or $N+2$ coupling matrix for the filter may be summarized as follows:

- Build up the $N \times N$ open-circuit impedance matrix $[z]$ for the filter network as detailed in Section 8.1.2, that is, $[z'] = R + [jM + sI]$, where $[M]$ is the filter coupling matrix [including frequency offsets (self couplings) on the diagonal, if any], sI is the diagonal matrix containing the frequency variable s [plus the factor $\delta_k = f_0/(\mathrm{BW}\cdot Q_{uk})$, $k = 1,2,\ldots,N$, accounting for resonators with finite unloaded Q_u factors], and $[R]$ is the termination matrix. Adding an inverter at either end of the network to normalize the terminations to unity converts the $N \times N$ impedance matrix $[z]$ to the $N+2$ admittance matrix $[y]$ (dual-network theorem).
- Invert the $N+2$ admittance matrix $[y]$ to give $[y']^{-1} = [z'']$. Designating the input and output nodes of the circuit as 0 and $N+1$, respectively, use equation (18.6) to evaluate the voltage v_k in the kth resonator and the voltage v_0 at the input to the network [$k = 0$ in equation (18.6)]:

$$v_k = \left(z_{k,0} - \frac{z_{k,N+1} z_{N+1,0}}{1 + z_{N+1,N+1}} \right) i_0 \tag{18.6}$$

  where the elements $z_{i,j}$ are taken from the impedance matrix $[z]$. The ratio $v_k(1 + S_{11}(s))/v_0$ then gives the voltage in the kth resonator referred to the voltage incident at the input of the network [25].
- Multiply the node voltages by $\sqrt{R_s}$. If $R_S = R_L = 1\,\Omega$ and the generator voltage $e_g = 2$ V, then the voltages v_k will be referred to 1 V, which corresponds to the maximum available power from the generator P_i, of 1 W (see Section 20.5.5).

The RF voltage values for a bandpass filter are found by multiplying the voltages v_k at the internal resonator nodes by a factor F_v, the value of which depends on the applied waveform and the technology used in the realization of the resonant structure. For example, for a waveguide

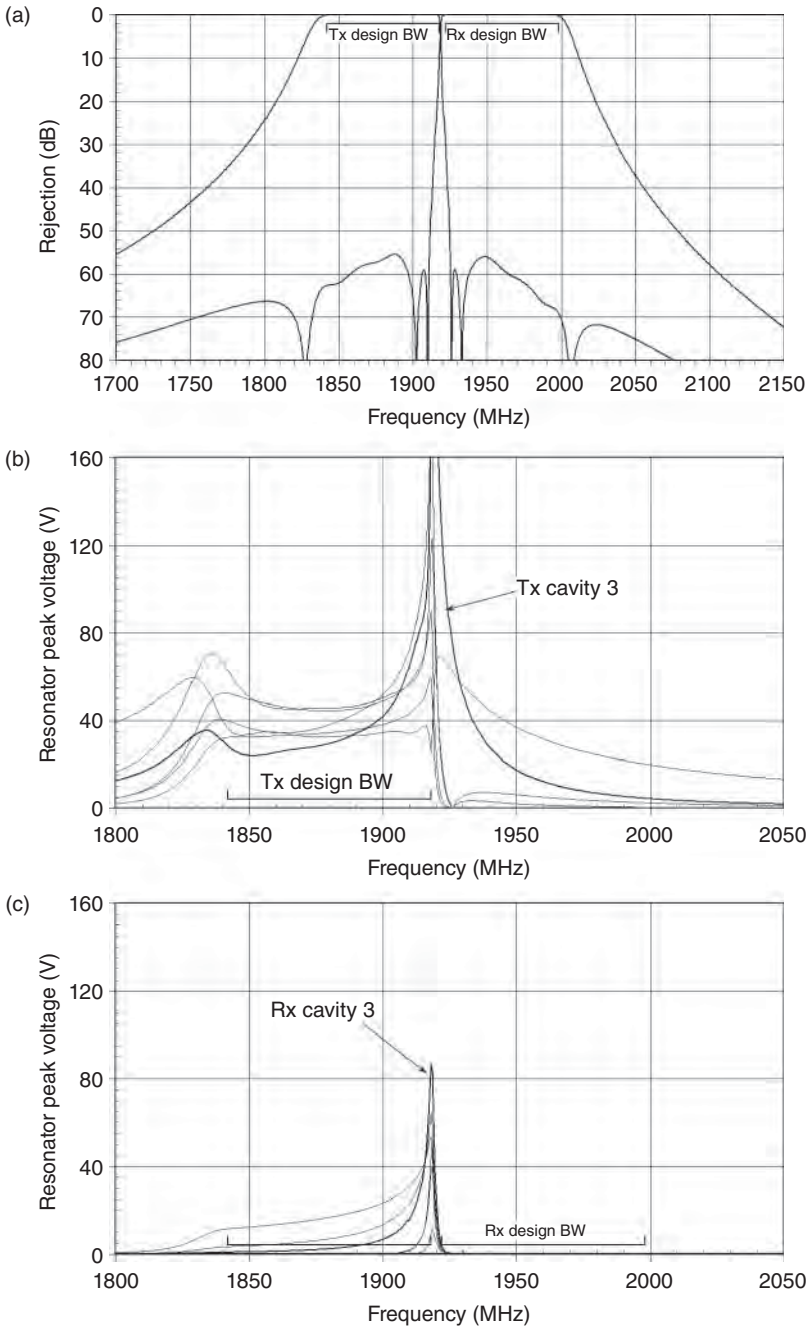


Figure 18.39 Transmit–receive diplexer with two complementary (6–2) asymmetric filters: (a) rejection characteristics, (b) voltage characteristics in transmit filter cavities, and (c) voltage characteristics in receive filter cavities.

bandpass filter,

$$F_v = \frac{\lambda_0}{\lambda_{g0}} \sqrt{\frac{2f_0}{n\pi \, \mathrm{BW}}} \tag{18.7}$$

where n is the number of half wavelengths of the waveguide resonators, f_0 is the center frequency, BW is the DBW, and λ_{g0} and λ_0 are the wavelengths at f_0 in the waveguide and in free space, respectively [25]. For a given incident power P_i watts, the peak voltage gradient across the narrow b dimension along the centerline of the cavities of a rectangular waveguide TE_{01}-mode filter are found by multiplying equation (18.6) by a further factor:

$$\hat{E} = \sqrt{\frac{480\pi \lambda_g P_i}{ab\lambda}} \quad \mathrm{V/cm} \tag{18.8}$$

where a and b are the broad and narrow dimensions, respectively, of the waveguide in centimeters and λ_g and λ are the wavelengths in the waveguide and free space, respectively.

For cavities with a more complicated geometry, such as coaxial cavities, it is more realistic to use a stored energy approach, analyzing the voltage patterns and gradients in a single resonator with a unit energy by using finite-element EM methods. Then the absolute voltage values in each resonator of the filter are estimated by scaling by the voltage found as above for each resonator of the prototype network and by the square root of the incident power [26].

Figure 18.39 shows the rejection characteristics and the estimated voltage levels within the cavities of the two (6–2) filters of an Rx/Tx diplexer, where the high-power filter is lower in frequency. The TZs of these filters were realized with two trisections each. The voltage levels are referred to 1 W of input power, and the plot, shown in bold, is the cavity with the highest-peak voltage, the third cavity from the input port in the case of the Tx filter. The plots in Figure 18.39 show that quite high voltages are also experienced by the cavities in the receive filter. This is caused by the power at the upper edge of the Tx filter bandwidth. The highest is also in the third cavity from the common port, meaning that the resonators of the receive filter also needs to be designed to withstand high voltages. For a coaxial cavity resonator, the highest voltage gradients tend to concentrate around the zone where the tuning screw is close to the top of the resonator rod [27, 28].

These voltage profiles and their levels are dependent on the filter's coupling topology. A different topology (e.g., folded instead of trisections) that gives the same rejection characteristics may give different voltage levels. Some work has been done to investigate topologies that give a more evenly distributed but lower average voltage level. However, the configurations that are best for lower voltages tend to be quite complex [29]. On the other hand, the cavity arrangements that have the minimum number of couplings, such as the cul-de-sac filter, have the worst voltage levels. It seems that simple coupling topologies are inconsistent with low internal peak voltage levels [27, 28].

Summary

The previous chapters deal with the design of a microwave filter as an individual component of a communication network. In a multiuser environment, communication systems employ a range of microwave filters as individual components, as well as parts of a multiport network required to separate or combine a number of RF channels. This is commonly referred to as a *multiplexing network* or, simply, a MUX. This chapter is devoted to the design and tradeoffs of MUXs for a variety of practical applications.

MUXs are used in communication systems where there is a need to separate a wideband signal into a number of narrowband signals (RF channels). The channelization of the allocated frequency band allows flexibility for the flow of communication traffic in a multiuser environment. This also eases the requirements on the HPAs so that they can operate at relatively high efficiency with an acceptable degree of nonlinearity. MUXs are also employed to provide the opposite function, that of combining several narrowband channels into a single wideband composite signal for transmission via a common antenna. Owing to the reciprocity of filter networks, a MUX can also be configured to separate transmit and receive frequency bands in a common device, referred to as a *duplexer* or *diplexer*. Since MUXs deal with the individual RF channels, they essentially control their characteristics. Wideband devices, such as amplifiers, switching networks, and antennas, have little impact over the narrow bandwidths of RF channels. Consequently, performance of MUXs is critical in the efficient utilization of the frequency spectrum and hence the channel capacity. MUXs find widespread use in satellite and cellular communication systems.

The starting point of this chapter is a discussion of the tradeoffs among the various types of multiplexing networks. They include circulator-coupled, hybrid-coupled, and manifold-coupled MUXs employing single-mode or dual-mode filters. It also includes MUXs based on using directional filters. This is followed by tradeoff and design considerations for each type of MUX. The design methodology and optimization strategy are dealt with in depth for the manifold-coupled MUX, by far the most complex microwave network. Numerous examples and photographs are included to illustrate the designs. The chapter concludes with a discussion of the high-power requirements and their impact on multiplexing networks.

References

1 Matthaei, G., Young, L., and Jones, M.T. (1985) *Microwave Filters, Impedance Matching Networks and Coupling Structures*, Artech House, Norwood, MA.

2 Cristal, E.G. and Matthaei, G.L. (1964) A technique for the design of multiplexers having contiguous channels. *IEEE Transactions on Microwave Theory and Techniques*, **12**, 88–93.

3 Atia, A.E. (1974) Computer aided design of waveguide multiplexers. *IEEE Transactions on Microwave Theory and Techniques*, **22**, 322–336.

4 Chen, M.H., Assal, F., and Mahle, C. (1976) A contiguous band multiplexer. *COMSAT Technical Review*, **6**, 285–307.

5 Kudsia, C.M. *et al*. (1979) A new type of low loss 14 GHz high power combining network. Proceedings of 9th European Microwave Conference, London, September 1979.

6 Kudsia, C.M., Ainsworth, K.R., and O'Donovan, M.V. (1980) Microwave filters and multiplexing networks for communication satellites in the 1980s. Proceedings of AIAA 8th Communications Satellite Systems Conference, April 1980.

7 Rhodes, J.D. and Levy, R. (1979) Design of general manifold multiplexers. *IEEE Transactions on Microwave Theory and Techniques*, **27**, 111–123.

8 Rhodes, J.D. and Levy, R. (1979) A generalized multiplexer theory. *IEEE Transactions on Microwave Theory and Techniques*, **27**, 99–110.

9 Doust, D. *et al.* (1989) Satellite multiplexing using dielectric resonator filters. *Microwave Journal*, **32**, 93–166.

10 Bandler, J., Daijavad, S., and Zhang, Q.-J. (1986) Exact simulation and sensitivity analysis of multiplexing networks. *IEEE Transactions on Microwave Theory and Techniques*, **34**, 111–102.

11 Levinson, D.S. and Bennett, R.L. (1989) Multiplexing with high performance directional filters. *Microwave Journal*, 92–112.

12 Rosowsky, D. (1990) Design of manifold multiplexers. Proceedings of ESA Workshop Microwave Filters, June 1990, pp. 145–156.

13 Rosenberg, U., Wolk, D., and Zeh, H. (1990) High performance output multiplexers for Ku-band satellites. Proceedings of 13th AIAA International Communication Satellite Conference, Los Angeles, March 1990, pp. 747–752.

14 Kudsia, C., Cameron, R., and Tang, W.C. (1992) Innovation in microwave filters and multiplexing networks for communication satellite systems. *IEEE Transactions on Microwave Theory and Techniques*, **40**, 1133–1149.

15 Uher, J., Bornemann, J., and Rosenberg, U. (1993) *Waveguide Components for Antenna Feed Systems—Theory and CAD*, Artech House, Norwood, MA.

16 Ye, S. and Mansour, R.R. (1994) Design of manifold-coupled multiplexers using superconductive lumped element filters. IEEE MTT-IMS, pp. 191–194.

17 Mansour, R.R. *et al.* (1996) Design considerations of superconductive input multiplexers for satellite applications. *IEEE Transactions on Microwave Theory and Techniques*, **44**, 1213–1228.

18 Matthaei, G., Rohlfing, S., and Forse, R. (1996) Design of HTS lumped element manifold-type microwave multiplexers. *IEEE Transactions on Microwave Theory and Techniques*, **44**, 1313–1320.

19 Mansour, R.R. *et al.* (2000) A 60 channel superconductive input multiplexer integrated with pulse-tube cryocoolers. *IEEE Transactions on Microwave Theory and Techniques*, **48** (7), 1171–1180.

20 Mansour, R.R., Ye, S., Dokas, V. *et al.* (2000) System integration issues of high power HTS output multiplexers. *IEEE Transactions on Microwave Theory and Techniques*, **48**, 1199–1208.

21 Pozar, D. (1998) *Microwave Engineering*, Wiley, New York.

22 Kudsia, C.M. (1987) High power contiguous combiners for satellite earth terminals. Proceedings of Canadian Satellite User Conference, Ottawa, Ontario, Canada, May 25–28, 1987.

23 Bandler, J., Biernacki, R., Chen, S. *et al.* (1994) Space mapping technique for electromagnetic optimization. *IEEE Transactions on Microwave Theory and Techniques*, **42**, 2536–2544.

24 Chen, M.H. (1977) Singly-terminated pseudo-elliptic function filter. *COMSAT Technical Review*, **7**, 527–541.

25 Sivadas, A., Yu, M., and Cameron, R.J. (2000) A simplified analysis for high power microwave bandpass filter structures. IEEE MTT-S International Microwave Symposium Digest, Boston, 2000, pp. 1771–1774.

26 Ernst, C. and Postoyalko, V. (1999) Comparison of the stored energy distributions in a QC-type and a TC-type prototype with the same power transfer function. IEEE MTT-S International Microwave Symposium Digest, Anaheim, CA, 1999, pp. 1339–1342.

27 Harish, A.R., Petit, J.S., and Cameron, R.J. (2001) Generation of high equivalent peak powers in coaxial filter cavities. Proceedings of 31st European Microwave Conference, London, September 2001, pp. 289–292.

28 Ernst, C. and Postoyalko, V. (2003) Prediction of peak internal fields in direct-coupled filters. *IEEE Transactions on Microwave Theory and Techniques*, **51**, 64–73.

29 Senior, B.S. (2004) Optimized network topologies for high power filter applications. Ph.D. thesis. University of Leeds.

19

Computer-Aided Diagnosis and Tuning of Microwave Filters

As a result of manufacturing and material tolerances, filter tuning is an essential postproduction process. Traditionally, filters have been tuned manually by skilled technologists. The tuning process is not only time consuming but also expensive, particularly for high-order narrowband filters with stringent requirements. It is a fact that almost all filters used for wireless base stations and satellite applications must be subjected to a postproduction tuning process. For example, in typical satellite applications, channel filters that operate at 4 GHz have very stringent in-band and out-of-band requirements with a design margin of 300 kHz, that is, <0.01%. Such a design margin is needed to accommodate the drift in filter frequency due to temperature, leaving almost a zero margin for design imperfections due to manufacturing tolerances. The specifications for wireless base station filters are as stringent as those for satellite applications.

The complexity of the tuning process depends on the technology that is employed and the filter configuration. For example, tuning is a must for narrowband dielectric resonator filters and narrowband planar filters, since the dielectric constant of the dielectric resonator or the substrate of planar structures can vary from batch to batch. For dielectric resonators with $\varepsilon_r = 38$, a deviation of ± 0.5 in the dielectric constant at 4 GHz translates into a frequency shift close to 25 MHz. In some applications, such a shift in the center frequency exceeds the filter bandwidth itself. The filter topologies and filter functions can also contribute to the complexity of the tuning process. Dual-mode elliptic filters and self-equalized filters are usually much more difficult to tune than single-mode filters with a Chebyshev response.

Typically, manual tuning is performed as a real-time iterative optimization process. To allow for tuning, filters are constructed with tuning screws or other forms of tuning elements to allow the technologist to change the resonance frequency of the filter resonators and the interresonator couplings. The technologist monitors the filter performance on the vector network analyzer (VNA) and tweaks the tuning screws iteratively, until the filter meets the specification requirements. For many technologists, the manual tuning process has been more of an art than a science. Therefore, the manual tuning of complex filter and multiplexer structures is usually performed by well-experienced technologists.

Tuning is a major factor in the overall filter production cost. Also, tuning has a significant impact on project schedules. The availability of computer-aided tuning techniques that can guide the technologists during the tuning process can be a major factor in reducing the tuning time. The use of robots has the potential to eliminate the need to use human operators, which can further reduce production costs and scheduling.

Although the concept of computer-aided tuning has been known for years [1–15], the low-cost and short-term delivery requirements of wireless base station filters in the mid-1990s has contributed to the advances and innovations in computer-aided filter tuning technology.

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

Since then, several papers have been published on the computer-aided tuning of microwave filters, employing different techniques. These techniques can be divided into five categories:

1) Sequential tuning of coupled resonator filters
2) Computer-aided tuning based on circuit model parameter extraction
3) Computer-aided tuning using poles and zeros of the input reflection coefficient
4) Time-domain tuning
5) Fuzzy logic (FL) tuning.

The following sections present details of these five approaches.

19.1 Sequential Tuning of Coupled Resonator Filters

Ness [16] has shown that the group delay of the input reflection coefficient of sequentially tuned resonators contains all the necessary information to tune filters. The paper demonstrates that the group delay value at the center frequency of the filter can be written quite simply in terms of the prototype lowpass g_k values. Consider the lowpass prototype filter shown in Figure 19.1. The Q_e and M_{ij} of the coupled matrix model are given in terms of the g_k lowpass prototype values, filter center frequency f_0, and filter bandwidth (BW) as

$$M_{j,j+1} = \frac{1}{\sqrt{g_j g_{j+1}}}, \quad j = 1,2,\ldots,N-1 \tag{19.1}$$

$$R_1 = \frac{1}{g_0 g_1}, \quad R_2 = \frac{1}{g_N g_{N+1}} \tag{19.2}$$

$$Q_e = \frac{f_0}{R_1 \cdot \mathrm{BW}} = \frac{g_0 g_1 f_0}{\mathrm{BW}} \tag{19.3}$$

$$k_{j,j+1} = \frac{\mathrm{BW}}{f_0} M_{j,j+1} = \frac{\mathrm{BW}}{f_0} \frac{1}{\sqrt{g_j g_{j+1}}} \tag{19.4}$$

The calculation from the lowpass prototype, with the proper lowpass-to-bandpass transformation, enables the group delay of the reflected signal to be expressed simply and directly in terms of the g_k values and the normalized bandwidth of the bandpass filter. Using equations (19.1)–(19.4), the external Q_e and the coupling k_{ij} can be determined from the filter specification. These coupling parameters can also be related to group delay Γ_d of the reflected signal S_{11} as successive resonators are tuned to resonance. The group delay of the $S_{11}(\omega)$ is defined as

$$\Gamma_d(\omega) = -\frac{\partial \phi}{\partial \omega} \tag{19.5}$$

For a standard lowpass-to-bandpass transformation, we obtain

$$\omega' \rightarrow \frac{\omega_0}{(\omega_2 - \omega_1)} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) \tag{19.6}$$

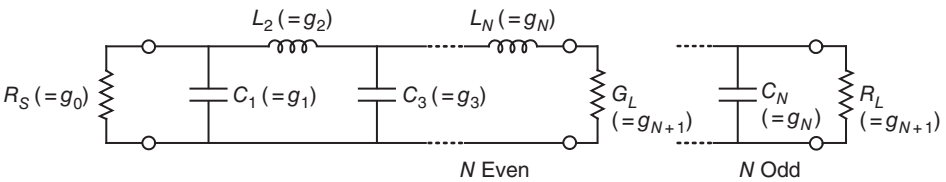


Figure 19.1 Lowpass prototype circuit.

where ω' is the angular frequency of the lowpass prototype, ω_0 is the center frequency of the bandpass filter, and ω_1 and ω_2 are the lower edge and the upper edge frequencies, respectively, of the bandpass filter. Whereas $\omega_0 = \left(\omega_1\ \omega_2\right)^{1/2}$, $\Gamma_d(\omega)$ is then written as

$$\Gamma_d(\omega) = -\frac{\partial \phi}{\partial \omega'}\frac{\partial \omega'}{\partial \omega} \tag{19.7}$$

$$\Gamma_d(\omega) = -\frac{\omega^2 + \omega_0^2}{\omega^2(\omega_2 - \omega_1)}\frac{\partial \phi}{\partial \omega'} \tag{19.8}$$

Consider the case where all the resonators are short-circuited ("shorted") (detuned) with the exception of the first resonator. This corresponds to the case where the elements $g_2, g_3, \ldots$ are disconnected from the lowpass prototype circuit. Input impedance Z_{in} and reflection coefficient S_{11} are then given by

$$Z_{\text{in}} = -\frac{j}{\omega' g_1}, \quad Z_0 = g_0 \tag{19.9}$$

$$S_{11} = \frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_0} \tag{19.10}$$

$$S_{11} = \frac{\omega' g_1 g_0 + j}{-\omega' g_1 g_0 + j} \tag{19.11}$$

$$\phi = 2\tan^{-1}\frac{1}{\omega' g_1 g_0} \tag{19.12}$$

We define $\Gamma_{d1}(\omega)$ as group delay, when all the resonators, except resonator 1, are shorted (detuned). Thus, $\Gamma_{d1}(\omega)$ can be written as

$$\Gamma_{d1}(\omega) = -\frac{2(\omega^2 + \omega_0^2)}{\omega^2(\omega_2 - \omega_1)} \quad \frac{g_0 g_1}{1 + (g_0 g_1 \omega')^2} \tag{19.13}$$

At the center frequency $\omega_0, \Gamma_{d1}(\omega_0)$ is given by

$$\Gamma_{d1}(\omega_0) = \frac{4 g_0 g_1}{(\omega_2 - \omega_1)} \tag{19.14}$$

In view of equation (19.3), we can relate the group delay $\Gamma_{d1}(\omega_0)$ to Q_e as

$$\Gamma_{d1}(\omega_0) = \frac{4Q_e}{\omega_0} \tag{19.15}$$

We then consider the case where all resonators of the bandpass filters are shorted (detuned) with the exception of the first and second resonators. In the prototype lowpass circuit, this corresponds to the case where the second element g_2 is shorted to the ground. We can then find the input impedance and use equations (19.8)–(19.10) to derive $\Gamma_{d2}(\omega_0)$, which is given as

$$\Gamma_{d2}(\omega_0) = \frac{4 g_2}{g_0(\omega_2 - \omega_1)} \tag{19.16}$$

In view of equations (19.1)–(19.4), $\Gamma_{d2}(\omega_0)$ is related to K_{12} by

$$\Gamma_{d2}(\omega_0) = \frac{16}{\omega_0^2 k_{12}^2 \Gamma_{d1}(\omega_0)} = \frac{4}{\omega_0 Q_e k_{12}^2} \tag{19.17}$$

We can follow a similar procedure to derive the group delay as the successive resonators are shorted (detuned). Table 19.1 provides the group delay and phase of the input reflection coefficient up to $N = 6$ [16]. It is interesting to note that at ω_0, the group delay of S_{11}, is determined only by the shunt elements for an odd number of resonators and by the series elements for

Table 19.1 Group delay at center frequency ω_0 as successive resonators are tuned.

$n=1$	$\Gamma_{d1}(\omega_0)=\dfrac{4g_0g_1}{(\omega_2-\omega_1)}$	$\Gamma_{d1}(\omega_0)=\dfrac{4Q_e}{\omega_0}$	$\phi\to\pm180^\circ$
$n=2$	$\Gamma_{d2}(\omega_0)=\dfrac{4g_2}{g_0(\omega_2-\omega_1)}$	$\Gamma_{d2}(\omega_0)=\dfrac{4}{\omega_0Q_ek_{12}^2}$	$\phi\to0^\circ$
$n=3$	$\Gamma_{d3}(\omega_0)=\dfrac{4g_0(g_1+g_3)}{(\omega_2-\omega_1)}$	$\Gamma_{d3}(\omega_0)=\Gamma_{d1}+\dfrac{4Q_ek_{12}^2}{\omega_0k_{23}^2}$	$\phi\to\pm180^\circ$
$n=4$	$\Gamma_{d4}(\omega_0)=\dfrac{4(g_2+g_4)}{g_0(\omega_2-\omega_1)}$	$\Gamma_{d4}(\omega_0)=\Gamma_{d2}+\dfrac{4k_{23}^2}{\omega_0Q_ek_{12}^2k_{34}^2}$	$\phi\to0^\circ$
$n=5$	$\Gamma_{d5}(\omega_0)=\dfrac{4g_0(g_1+g_3+g_5)}{(\omega_2-\omega_1)}$	$\Gamma_{d5}(\omega_0)=\Gamma_{d3}+\dfrac{4Q_ek_{12}^2k_{34}^2}{\omega_0k_{23}^2k_{45}^2}$	$\phi\to\pm180^\circ$
$n=6$	$\Gamma_{d6}(\omega_0)=\dfrac{4(g_2+g_4+g_6)}{g_0(\omega_2-\omega_1)}$	$\Gamma_{d6}(\omega_0)=\Gamma_{d4}+\dfrac{4k_{45}^2k_{23}^2}{\omega_0Q_ek_{12}^2k_{34}^2k_{56}^2}$	$\phi\to0^\circ$

an even number of resonators. The reverse applies if the dual circuit for the lowpass filter is selected.

We can easily show that for narrowband filters the group delay $\Gamma_{d1}(\omega)$, $\Gamma_{d2}(\omega)$, …, and $\Gamma_{dn}(\omega)$ can be approximated as even functions of ω around ω_0. It can be also demonstrated that the phase of S_{11} passes through $\pm180^\circ$ and 0° crossing at ω_0, as the successive resonators are tuned. Figures 19.2 and 19.3 show sketches for the phase and group delay around the center frequency, respectively, as the resonators are successively tuned for a filter, designed at a center frequency of 12 GHz. The location of the $\pm180^\circ/0^\circ$ phase crossing deviates slightly from the center frequency. Note that the group delay is not perfectly symmetric. The slight unsymmetry is also attributed to the fact that the results shown in Figures 19.2 and 19.3 are obtained from coupling matrix model M, where the values of diagonal elements M_{ii} determine how well the resonators are tuned ($M_{ii}=0$ represents the case of perfectly tuned resonators). The results given in these two figures represent the case where resonator detuning is obtained by using a large value for the diagonal elements M_{ii}, rather than perfectly shorting out (short-circuiting) the resonators. This, in fact, represents a real-case scenario since shorting the resonator is not readily achieved in most practical applications.

19.1.1 Tuning Steps

Step 1. For the given filter specifications, calculate the group delay values $\Gamma_{d1}, \Gamma_{d2}, \ldots, \Gamma_{dN}$ in terms of the lowpass g_k values or the coupling values k_{ij}.

Step 2. Short (detune) all resonators except resonator 1. Adjust the input coupling and resonance frequency of resonator 1 to set the group delay to the specified value Γ_{d1} at the filter center frequency.

Step 3. Short (detune) all the resonators except resonators 1 and 2. Adjust the resonance frequency of resonator 2 and the coupling between resonators 1 and 2 to get a symmetric group delay response about the filter center frequency and with the specified value Γ_{d2}. To maintain symmetry, it may be necessary to readjust resonator 1, if the coupling between resonators 1 and 2 is so strong to the extent that it detunes resonator 1.

Step 4. Progress through the filter by tuning each subsequent resonator as in step 3 to maintain group delay symmetry and to set the group delay at the filter center frequency to the

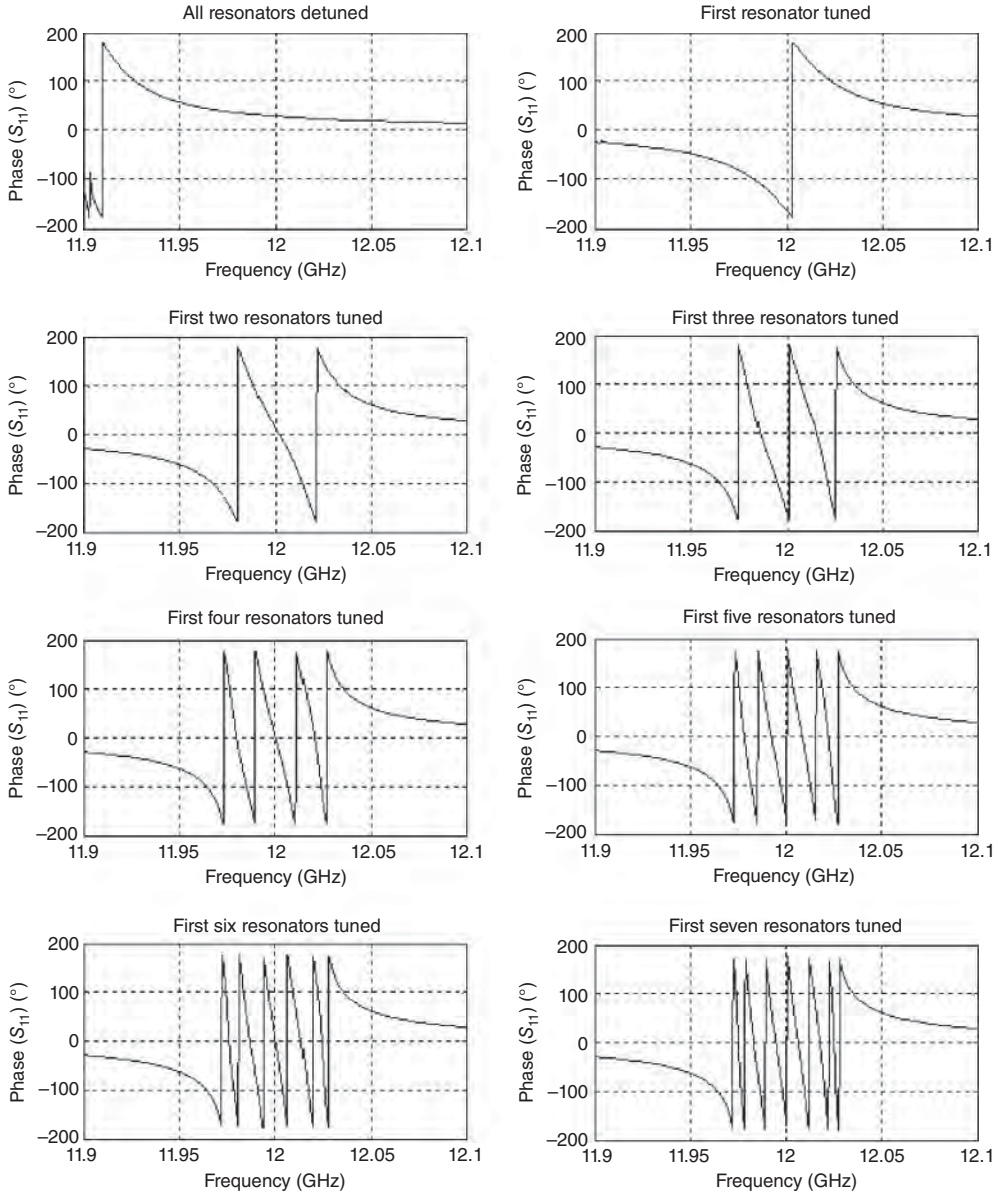


Figure 19.2 Phase of input reflection coefficient of an eight-pole filter as successive resonators are tuned.

specified values $\Gamma_{d3}, \Gamma_{d4}, \ldots, \Gamma_{dN-1}$. In each case, a slight readjustment of the resonance frequency of the preceding resonator may be required to maintain the group delay symmetry.

Step 5. When the last resonator is reached and the filter output is properly terminated, observe the amplitude response of S_{11} and tune the last resonator and the output coupling to obtain the specified return loss.

Precise results are obtained only if the resonators are properly shorted rather than simply detuned. Nevertheless, the tuning procedure presents a simple approach to tune the filter so

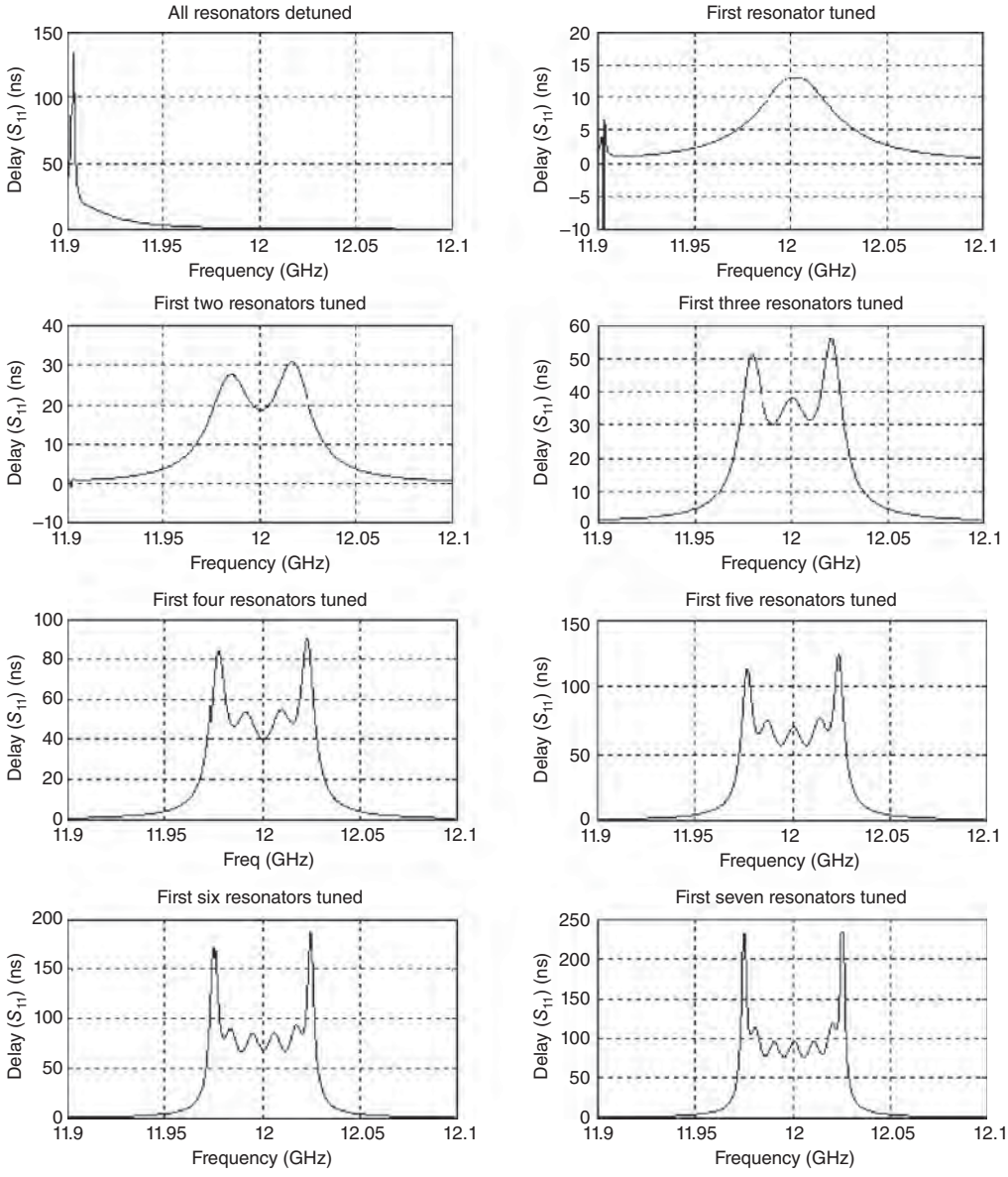


Figure 19.3 Group delay of input reflection coefficient of an eight-pole filter as successive resonators are tuned.

that its response is close to the required response. A minimum fine-tuning may be required at the end of step 5 in order to meet the exact filter specifications.

In this tuning procedure [16], each step involves the simultaneous adjustment of two parameters, the resonance frequency and interresonator coupling, requiring the group delay symmetry. Also, since the phase passes through the 0°/180° crossing at the center frequency, as the resonators are detuned, these two parameters can be successively tuned rather than simultaneously tuned. After we tune the resonator center frequency to achieve either 0° or 180° crossing at the center frequency, we adjust the interresonator coupling to set the group delay as close as

possible to the specified value as feasible. At this point, both the resonator center frequency and the interresonator coupling can be fine-tuned to set the group delay exactly to the specified value. However, in practical applications, filters are usually attached to input/output transmission lines, which affect the location where the 180°/0° phase crossing takes place, and the use of phase information requires knowledge of the input reference plane. Several approaches are available to determine the input reference planes. A simple approach is to apply step 2 to the first resonator to get group delay symmetry and to set the group delay to the value Γ_{d1}. The frequency at which this occurs directly gives the frequency shift needed to accommodate the shift in the reference plane. Thus, the use of 180°/0° phase crossing can be used by itself to rough-tune the filter to the proper frequency band.

19.2 Computer-Aided Tuning Based on Circuit Model Parameter Extraction

This approach incorporates a circuit model for the filter with an element optimization routine. The idea in this technique is to optimize the parameters of the circuit model such that the response from the circuit model best fits the measured response of the filter. After that, a comparison between the extracted parameters and the parameters of the ideal filter is used to identify the filter elements that need tuning. The technique was first introduced by Thal [7]. Several modifications of this technique have been reported [6, 17–22]. Computer-aided tuning software tools based on parameter extraction were developed by COM DEV [17] and other filter suppliers in early 1990s. These software tools were used to guide technologists through the tuning process of microwave filters.

Although several circuit models are available that can be employed in this technique, the advantage of using the coupling matrix circuit model is that the coupling elements are directly related to the position of the physically tunable elements. Once one of these coupling elements is identified and found to deviate from the desired ideal value, the coupling element can be adjusted by turning the corresponding tuning element.

Consider the generalized filter network shown in Figure 19.4. The circuit model consists of n coupled lossless resonators, where M_{ij} denotes the frequency-independent coupling between resonators i and j. The generalized matrix is given in equation (19.18). For given filter specifications, the coupling matrix is synthesized, as explained in Chapters 8–10:

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & & & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix} \tag{19.18}$$

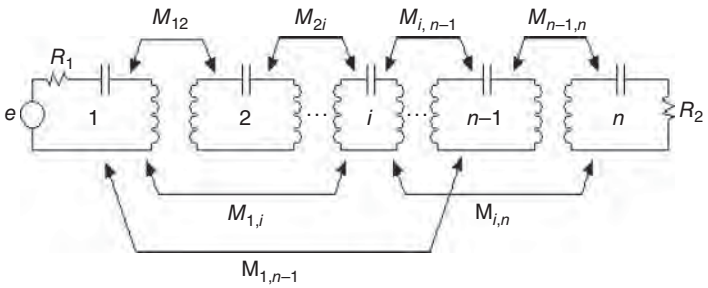


Figure 19.4 A generalized model for coupled resonator filters.

The filter scattering parameters in terms of the coupling elements matrix and input and output coupling R_1 and R_2 are given as

$$S_{21} = -2j\sqrt{R_1 R_2}[A^{-1}]_{n1} \tag{19.19}$$

$$S_{11} = 1 + 2jR_1[A^{-1}]_{11} \tag{19.20}$$

$$A = \lambda I - jR + M \tag{19.21}$$

$$\lambda = \frac{f_0}{\mathrm{BW}}\left(\frac{f}{f_0} - \frac{f_0}{f}\right) \tag{19.22}$$

where I is the unity matrix and R is a diagonal matrix with all the elements zero, except $R_{11} = R_1$ and $R_{nn} = R_2$.

The S-parameters of the measured filter are described by the same circuit model. Here, the coupling elements are related to a particular tuning element. For example, the tuning of the filter resonance frequencies correspond to varying the diagonal elements M_{ii}, while tuning of the coupling between the resonators corresponds to varying coupling elements M_{ij}. Thus, we can extract the coupling matrix, associated with the measured results, by optimizing the elements of the coupling matrix such that the S-parameters of the circuit model match the measured S-parameters. A comparison of the coupling matrix extracted from the measured results with the coupling matrix of the ideal filter reveals the elements that need tuning. The tuning procedure can then be summarized as follows:

Step 1. Measure the performance of the filter being tuned.

Step 2. Use the coupling matrix circuit model and extract the coupling elements M_{ij}, R_1, and R_2 by optimization for best fit with measured data. To construct the objective function, the measured performance of the filter is sampled at several frequency points within the band, and the objective function to be minimized can be defined as

$$\phi = \sum_{\text{freq}} \sum_{i=1}^{2} \sum_{j=1}^{2} (\text{abs}(S_{ij}^{\text{model}}) - \text{abs}(S_{ij}^{\text{measured}})) \tag{19.23}$$

where S_{ij}^{model}, $i = 1,2$ and $j = 1, 2$ are the S parameters, calculated from the coupling matrix circuit model at specific frequency points, whereas S_{ij}^{measured} $i = 1, 2, j = 1, 2$ are the measured parameters at the same frequency points. The variables to be optimized are the coupling elements M_{ij} and input/output coupling R_1 and R_2.

Step 3. Compare the coupling matrix elements, extracted from the experimental results, with the ideal coupling matrix elements. Adjust the tuning elements accordingly.

Step 4. Repeat steps 1–3 iteratively until the filter is tuned. At this point, the extracted coupling elements from the measured results should reasonably match the ideal coupling elements.

One problem associated with this technique is that very good initial values for the variables to be optimized are usually needed to reduce the possibility of running into a local minimum or failing to converge to the proper solution. The problem is particularly pronounced in the case of highly detuned filters. To overcome this problem, two approaches were proposed: one is based on sequential parameter extraction and the other is based on the FL technique. The latter technique is explained in Section 19.6.

The sequential parameter extraction procedure [22] starts with detuning all the resonators (i.e., shorting) by setting the resonance frequencies of the corresponding resonators to values outside the frequency range. Then one resonator after the other is tuned (unshorted). Thus for a filter of degree n, we obtain a sequence of n subfilters, with the subfilter i characterized by $n - i$

shorted resonators and i unshorted resonators. Then the parameter extraction is carried out at each step to extract the coupling elements of the subfilter. At step i, after unshorting resonator i, the filter return loss S_{11} is measured, and the parameters of the corresponding subfilter are extracted by minimizing the following error function:

$$\varepsilon(i) = \sum_j |S_{11}^{\text{meas}}(\omega_j) - S_{11}^{\text{mod}}(\omega_j)_{(i)}|^2 \tag{19.24}$$

Equation (19.24) represents the difference between the measured and the modeled responses for each subfilter. Once the subfilter parameters are extracted, the actual filter elements are tuned to minimize the deviation between the ideal parameters and the actual subfilter parameters before proceeding to the next step.

Let the coupling matrix of the ideal complete filter be denoted as $[M]$, and the reflection coefficient of the subfilter I, $S_{11(i)}^{\text{mod}}$ can be calculated from the physical circuit shown in Figure 19.4 by shorting $n-i$ last resonators. Alternatively, $S_{11(i)}^{\text{mod}}$ can be calculated from equations (19.19)–(19.22) by using a different coupling matrix for the subfilter at each step. Let $[M]^{(1)}$, $[M]^{(2)}, \ldots, [M]^{(n-1)}$ be the coupling matrices associated with subfilters $1,2, \ldots, n-1$. The M-matrices of the subfilters can be obtained by simply setting the corresponding diagonal elements to a very large value, which is equivalent to resonator detuning. These matrices are obtained as follows:

Subfilter 1. All the resonators with the exception of resonator 1 are shorted. The parameters to be optimized are R_1 and M_{11}. Then $[M]^{(1)} = [M]$ with $M_{22}, M_{33}, \ldots, M_{nn}$ are set to a very large value.

Subfilter 2. All the resonators with the exception of resonators 1 and 2 are shorted. The parameters to be optimized are M_{12} and M_{22}, whereas R_1 and M_{11} are fixed at the values extracted from step 1, but can also be included in the optimization for fine-tuning to achieve the best fit with the measured results of subfilter 2. Then $[M]^{(2)} = [M]$ are with $M_{33}, M_{44}, \ldots, M_{nn}$ set to a very large value.

Subfilter $n-1$. Resonator n is shorted. The parameters to be optimized are $M_{n-2,n-1}$ and $M_{n-1,n-1}$. Then $[M]^{(n-1)} = [M]$ with M_{nn} is set to a very large value.

The Complete Filter. The parameters to be optimized are M_{nn} and R_2. The other coupling elements of the coupling matrix can be set to the values obtained from the previous steps or included for the optimization as fine-tuning for the best fit with the measured results of the whole filter.

This approach provides the flexibility of controlling the number of variables to be optimized. The parameters that are extracted and tuned up to step i may be assumed fixed in the model during the next parameter extraction step, $i+1$. Alternatively, since bringing the resonator $i+1$ in-band by removing the detuning may have an impact on adjacent couplings, some of the parameters extracted in step i may be assumed to be variables in the parameter extraction step $i+1$. However, the initial values for these variables will be the values extracted in step i rather than ideal values, which in turn will speed up the optimization process.

This kind of sequential parameter extraction, although it requires several optimization steps, helps eliminate the convergence problems associated with optimizing all filter parameters at once. The number of parameter extraction processes (optimization) may be lowered by reducing the number of subfilters. As an example, if we take a 10-pole filter, we may use three subfilters: subfilter 1 with the last 7 resonators detuned; subfilter 2, with the last 3 resonators detuned; and subfilter 3, as the whole filter, with no resonator detuned. The use of sequential parameter extraction technique to tune complex filter structures has been demonstrated [22], including the tuning of a 12-pole self-equalized dual-mode waveguide filter.

At each step in the tuning process for the subfilters, a certain number of filter parameters that deviate from the ideal parameters need to be adjusted. Thus, for a particular stage, where several of the tuning elements need tuning, one of these elements must be selected for the very next tuning action. In some cases, the convergence of the tuning process depends strongly on the decision as to which tuning elements should be adjusted first. A common selection criterion is to base the action on the difference between the actual coupling matrix and ideal coupling matrix, where the elements that have a large deviation need to be tuned first. However, this approach tends to give preference to the elements that have large values such as sequential coupling over the elements with small values such as cross-coupling and resonance frequencies (the diagonal elements of the coupling matrix). Furthermore, even when using normalized values, it is difficult to compare the mistuning level of the coupling with the mistuning level of the resonators. An approach based on sensitivity analysis of the equivalent circuit has been proposed in [22] to identify the elements that need to be tuned first.

Up to this point, the parameter extraction technique has been presented as a diagnostic tool, where it is used to extract the coupling matrix so it can be compared with the ideal matrix in order to identify the elements (coupling and resonance frequencies) that need adjustment. The tool can guide the technologists to select the proper elements during the tuning process, and with repeated iterations, they can align the filter to the desired response. The question is how the deviation in coupling value (M_{ij}, R_1, R_2) or resonance frequency (M_{ii}) is translated to the penetration depth of the tuning element (screw). The information is needed to automate the parameter extraction process to construct a robot that can replace the technologist.

A sensitivity analysis is needed to determine the relationship between the deviation in the coupling value and the tuning screw, or the position of the tuning elements. The steps needed for this sensitivity analysis [19] are listed next:

Step 1. Rough-tune the filter within the tuning range.
Step 2. Measure the *S*-parameters for this basis position.
Step 3. Measure *S*-parameters with one screw turned at a time.
Step 4. Extract coupling matrix elements for the basis position and for each measurement when one screw is turned at a time.

Assume that the number of screws is m and that the screw turns are the parameters that must be controlled, when the entire process is automated. For m tuning screws, let $n_1, n_2, \ldots, n_m$ be the number of turns associated with screws 1,2, …, m. A comparison between the extracted coupling matrix of the basis position and the extracted coupling matrices provides $\Delta R_1, \Delta R_2, \Delta m_{11}$, and Δm_{ij}, which are associated with the m steps $\Delta n_1, \Delta n_2, \ldots, \Delta n_m$. We can then build the following sensitivity matrix:

$$\text{Sensitivity matrix} = \begin{bmatrix} \dfrac{\partial R_1}{\partial n_1} & \dfrac{\partial R_1}{\partial n_2} & \cdots & \dfrac{\partial R_1}{\partial n_m} \\ \dfrac{\partial m_{11}}{\partial n_1} & \dfrac{\partial m_{11}}{\partial n_2} & \cdots & \dfrac{\partial m_{11}}{\partial n_m} \\ \dfrac{\partial m_{12}}{\partial n_1} & \dfrac{\partial m_{12}}{\partial n_2} & \cdots & \dfrac{\partial m_{12}}{\partial n_n} \\ \vdots & \vdots & \cdots & \vdots \\ \dfrac{\partial m_{nn}}{\partial n_1} & \dfrac{\partial m_{nn}}{\partial n_2} & \cdots & \dfrac{\partial m_{nn}}{\partial n_m} \\ \dfrac{\partial R_2}{\partial n_1} & \dfrac{\partial R_2}{\partial n_2} & \cdots & \dfrac{\partial R_2}{\partial n_m} \end{bmatrix} \tag{19.25}$$

The relation between the deviation in the coupling value and the screw position can then be written as follows:

$$\Delta m_{ij} = \sum_{k=1}^{k=m} \frac{\partial m_{ij}}{\partial n_k} \Delta n_k$$

The sensitivity matrix can be used to construct a robot to automate the tuning process (see Section 19.6). In most filter structures, each tuning element affects not only the corresponding element in the coupling matrix but also nearby elements. The extent of the influence of each tuning screw on the other coupling elements is also found from the sensitivity matrix. Thus, this sensitivity matrix is beneficial to both the automation process and the diagnosis process.

The technique presented in this section has been successfully used in the diagnosis and automated computer-aided tuning of microwave filters. Several modifications of this technique have been reported, offering the use of other filter circuit models or an alternative approach to extract the coupling elements. For example, the following objective function has been used [18] to replace the objective function given in equation (19.23):

$$\phi = \sum_{i=1}^{N} |S_{11}^{\text{model}}(P_i)|^2 + \sum_{i=1}^{M} |S_{21}^{\text{model}}(Q_i)|^2 \tag{19.26}$$

where P_i and Q_i are the locations of the reflection and transmission zeros, respectively, which are generated from the measured response. The idea here is to construct polynomials generated from sample points from the measured filter response using the model-based parameter estimation technique [23]. The location of the reflection and transmission zeros are then extracted from these polynomials. Details on this approach are given in [18].

19.3 Computer-Aided Tuning Based on Poles and Zeros of the Input Reflection Coefficient

A general model for the determination of individual resonance frequencies and coupling coefficients of a system with cascaded coupled resonators based on knowledge of the zeros and poles of the filter is described in this section. The key idea in this approach [24] is that measurement or computation of the phase of the reflection coefficient provides all the information necessary to evaluate the couplings and resonance frequencies. To understand the concept, we consider the two-pole filter shown in Figure 19.5 with the output shorted. The loop equations are given by

$$e_1 = i_1 \left(j\omega L_1 + \frac{1}{j\omega C_1} \right) - jM_{12} i_2 \tag{19.27}$$

$$0 = i_2 \left(j\omega L_2 + \frac{1}{j\omega C_2} \right) - jM_{12} i_1 \tag{19.28}$$

If we define m_{12} as

$$m_{12}^2 = \omega_{01}\omega_{02} \frac{M_{12}^2}{Z_{01}Z_{02}} \tag{19.29}$$

where $\omega_{0i} = 1/(\sqrt{L_i C_i})$ and $Z_{0i} = \sqrt{L_i/C_i}$, then Z_{in} is written as

$$Z_{\text{in}} = \frac{e_1}{i_1} = j\frac{Z_{01}}{\omega\omega_{01}} \frac{(\omega^2 - \omega_{01}^2)(\omega^2 - \omega_{02}^2) - \omega^2 m_{12}^2}{(\omega^2 - \omega_{02}^2)} \tag{19.30}$$

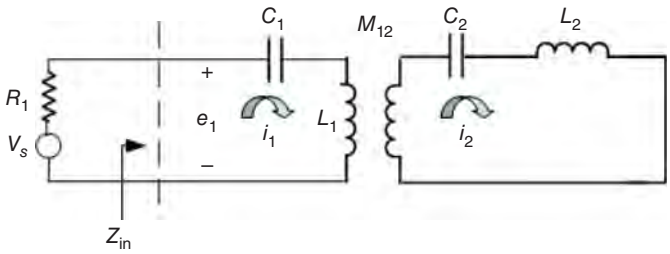


Figure 19.5 Equivalent circuit of two cascaded resonators terminated in a short circuit.

and Z_{in} can be written in a polynomial form as

$$Z_{\text{in}} = j\frac{Z_{01}}{\omega\omega_{01}}\frac{(\omega^2-\omega_{z1}^2)(\omega^2-\omega_{z2}^2)}{(\omega^2-\omega_{p1}^2)} \tag{19.31}$$

where

$$\omega_{01}^2 = \frac{\omega_{z1}^2\omega_{z2}^2}{\omega_{p1}^2}, \quad \omega_{02}^2 = \omega_{p1}^2 \tag{19.32}$$

$$m_{12}^2 = \omega_{z1}^2 + \omega_{z2}^2 - \omega_{p1}^2 - \omega_{01}^2 \tag{19.33}$$

This analysis demonstrates that the zeros and poles of the input reflection coefficient with output port terminated in a short circuit are related to individual resonance frequencies and coupling coefficients.

Consider the generalized coupled resonator model shown in Figure 19.6 with resonator n terminated in a short circuit at reference plane TT. The input impedance at loop i can be written as

$$Z_{\text{in}}^{(i)} = j\frac{Z_{0i}}{\omega\omega_{0i}}\frac{P_i(\omega^2)}{Q_i(\omega^2)}, \quad i = 1,2,\ldots,n \tag{19.34}$$

where $P_i(\omega^2)$ and $Q_i(\omega^2)$ are polynomials of orders $(n-i+1)$ and $(n-1)$, respectively, and Z_{0i} and ω_{0i} are the characteristic impedance and resonance frequency of resonator i. These two polynomials are expressed as

$$P_i(\omega^2) = \prod_{t=1}^{n-i+1}(\omega^2 - \omega_{zt}^{(i)^2}) \quad i = 1,2,\ldots,n \tag{19.35}$$

$$Q_i(\omega^2) = \prod_{q=1}^{n-i}(\omega^2 - \omega_{pq}^{(i)^2}), \quad i = 1,2,\ldots,n \tag{19.36}$$

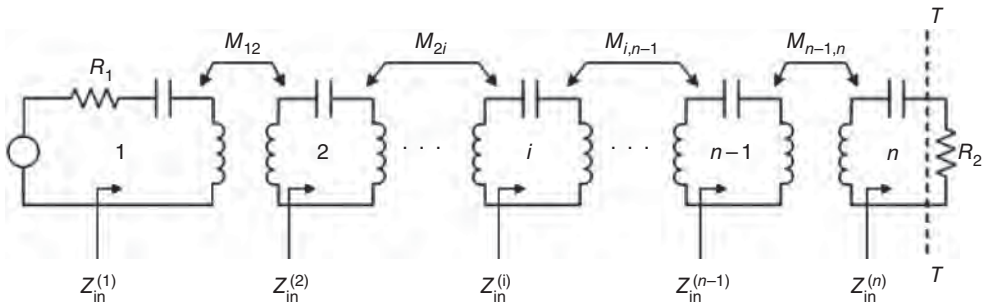


Figure 19.6 The equivalent circuit of *N*-cascaded resonators terminated in a short circuit.

where $\omega_{zt}^{(i)^2}$ $(t = 1,2,\ldots,n-i+1)$ and $\omega_{pq}^{(i)^2}(q = 1,2,\ldots,n-i)$ are the zeros of $P_i(\omega^2)$ and $Q_i(\omega^2)$ corresponding, respectively, to the zeros and poles of the input impedance of the one-port network at loop i.

It can be proved [24] that the resonance frequencies and couplings are related to the zeros and poles by the following equations:

$$\omega_{0i}^2 = \frac{\prod_{t=1}^{n-i+1} \omega_{zt}^{(i)^2}}{\prod_{q=1}^{n-i} \omega_{pq}^{(i)^2}}, \quad i = 1,2,\ldots,n \tag{19.37}$$

$$m_{i,i+1}^2 = \sum_{t=1}^{n-i+1} \omega_{zt}^{(i)^2} - \sum_{q=1}^{n-i} \omega_{pq}^{(i)^2} - \omega_{0i}^2, \quad i = 1,2,\ldots,n-1 \tag{19.38}$$

$$r_{1,n} = \left| \frac{\coprod_{i=1}^{n} (\omega_R^2 - \omega_{zi}^{(1)^2})}{\omega_R \coprod_{i=1}^{n} (\omega_R^2 - \omega_{pq}^{(1)^2})} \right| \tag{19.39}$$

where ω_R is the frequency corresponding to a $\pm 90°$ phase of the input reflection coefficient. Equations (19.37) and (19.38) provide a simple and direct correlation between the filter parameters (the resonance frequencies of the individual resonators and the interresonator couplings) and the zeros and poles of the short-circuited network. These zeros and poles can be obtained by either direct or indirect measurement. In the case of direct measurement, where the reference plan is clearly defined, the frequencies, corresponding to a phase value of $\pm 180°$ and $0°$, are the corresponding zeros and poles of the input reflection coefficient (input impedance), while in the case of indirect measurement, the locations of zeros and poles can be determined through the calculation of phase derivatives with respect to frequency. The frequencies at which the phase derivative is maximum (minimum) are the corresponding zeros (poles) of the input impedance.

This technique is based on having a short-circuited termination directly at the end of the last resonator. However, in practice, the actual short-circuit reference plane for the last resonator is seldom accessible. In most cases, the last resonator is loaded by a transmission line and/or connector, which represents a loading on the last resonator. Hsu *et al.* [24] proposed an efficient systematic method to remove the loading effect on the last resonator. The loading affects only the resonance frequency of the last resonator; therefore, the effect of the loading may not be an important step for the initial tuning stage. The reader may refer to the original paper [24] for details on removal of loading effects on the last resonator.

The tuning procedure in this approach relies on the fact that there is a direct relation between the poles and zeros of the input impedance as one set and the filter resonance frequency and interresonator coupling as another set; in other words, we can extract (extrapolate) one set from the knowledge of the other set. Let us define the desired theoretical set of interresonator couplings and resonator center frequencies as $M_{12}, M_{23}, \ldots, M_{n-1,n}, F_{01}, F_{02}, \ldots, F_{0n}$. The tuning steps are summarized here:

1) Calibrate to determine the proper location for the reference plane for accurate phase measurement. In practice, the procedure starts by inserting all the filter tuning screws deep into the cavities so that their resonance frequencies are far from the desired resonance frequencies of the individual cavities (the detuned condition). This condition is observed on the polar display of a network analyzer when the input reflection coefficient appears as a single spot, while the frequency is being swept several times the bandwidth around the desired

center frequency. The phase reference plane is then adjusted to achieve the "best" spot on the sweep, which will be used to correspond to the 0.

2) The measurement and adjustment of R_1 (input coupling) is performed by bringing the first resonator into resonance, while all the remaining resonators in the filter are still in the detuned condition. The input coupling R_1 can be determined using the procedure reported in [3] or [16]. The same approach is applied for the measurement of the output coupling R_2.
3) Terminate the output port with a short circuit and look at the phase of the input reflection coefficient. Bring all the resonators into resonance by adjusting the tuning screws until all the zeros and poles are shown on the display of VNA. The ones at the left correspond to the frequencies of zeros (±180°), and the ones at the right correspond to the poles (0°).
4) On the VNA, record the frequencies of all the zeros and poles (n zeros and $n-1$ poles). Next, extract the interresonator couplings and resonance frequencies (n resonance frequencies and $n-1$ couplings) [24]. These extracted couplings and resonance frequencies differ from the desired ones at this point. Let us identify this set as $\{M'_{12}, M'_{23}, \ldots, M'_{n-1,n}, F'_{01}, F'_{02}, \ldots, F'_{0n}\}$.
5) In the extracted set of interresonator coupling and resonator center frequencies in step 4, replace M'_{12} by the desired coupling value M_{12}. The new set $\{M_{12}, M'_{23}, \ldots, M'_{n-1,n}, F'_{01}, F'_{02}, \ldots, F'_{0n}\}$ is then used to synthesize the zeros and poles. The obtained set of zeros and poles are used as criterion for the adjustment of the coupling between resonators 1 and 2. The idea here is that by adjusting the tuning screw that corresponds to M_{12}, such that the zeros and poles are placed to the recorded position, the coupling between resonators 1 and 2 will be set to the desired value.
6) Repeat step 5 for all interresonator couplings. For example, for the coupling between resonators 2 and 3, the set $\{M_{12}, M_{23}, \ldots, M'_{n-1,n}, F'_{01}, F'_{02}, \ldots, F'_{0n}\}$ is used to synthesize the zeros and poles. Step 5 is therefore repeated $n-1$ times, until all the interresonator couplings are set to the desired value.
7) Apply step 5 to the resonance frequencies of the individual resonators; that is, for the first resonator, use the set $\{M_{12}, M_{23}, \ldots, M_{n-1,n}, F_{01}, F'_{02}, \ldots, F'_{0n}\}$ to synthesize the new set of zeros and poles. The tuning screw for the first resonator is adjusted until the measured poles and zeros match the ones that are generated by synthesis using the above set. For the resonance frequencies, step 5 is therefore repeated n times.

Overall, the number of tuning steps is $(2n-1)$ after the input and output couplings have been set correctly. Examples on the application of this approach are given in [24].

19.4 Time-Domain Tuning

This technique was first reported by Agilent [25, 26]. The idea in this tuning technique is that certain features of the time-domain response of the input reflection coefficient S_{11} correspond exactly to the resonator center frequencies and the interresonator coupling elements. Figure 19.7 shows the frequency response and the S_{11} time-domain response of a perfectly tuned five-pole Chebyshev filter. The time-domain response is the inverse Fourier transform of the frequency response of S_{11}. There are five distinct dips in the time-domain response for $t > 0$. We demonstrate that each of these five dips corresponds to a resonator in the filter. The first dip, with the shorter time delay, corresponds to the first resonator, the second dip corresponds to the second resonator, and so on. Also, we demonstrate that the peaks between the dips relate to the interresonator couplings. The time-domain response is basically an indication of the reflected energy; thus, the dips and peaks for $t < 0$ are not meaningful and have nothing to do with the filter elements. A detailed Agilent application note on the subject is given in [27].

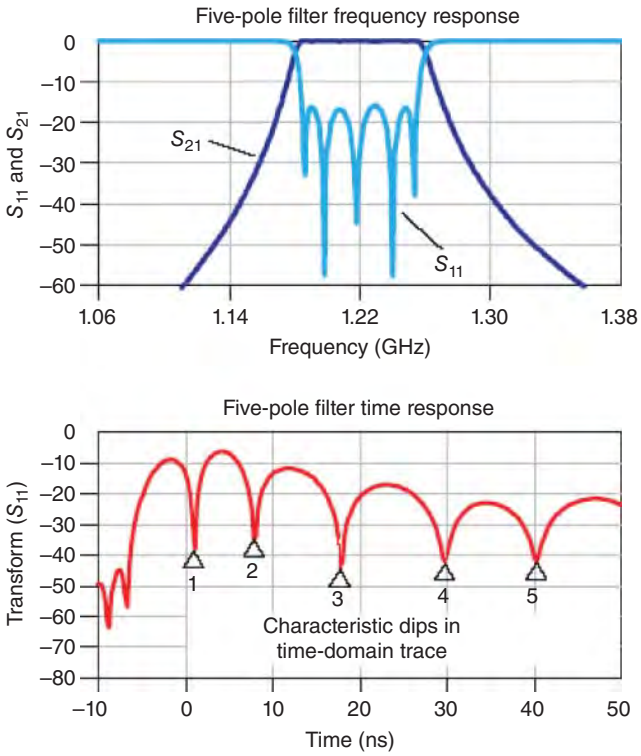


Figure 19.7 Frequency response and S_{11} time-domain response of a five-pole bandpass filter [27].

19.4.1 Time-Domain Tuning of Resonator Frequencies

In order to demonstrate the relationship between the dips in the time-domain response and the individual resonators, we detune each resonator and monitor both the simulated frequency response and the time-domain response. Figure 19.8 shows the frequency response and S_{11} time-domain response for the case of detuning of resonator 2 only, whereas Figure 19.9 illustrates the case of detuning resonator 3 only. The perfectly tuned ideal response of the filter is also included in these two figures. It is noted that, by detuning resonator 2, the first dip does not change very much; however, the second dip is no longer minimized, and neither are the following dips. In the case where resonator 3 is detuned, the first and second dips do not change, and the third, fourth, and fifth dips are no longer minimized. The same pattern is observed by detuning the other resonators in the filter. This demonstrates the relationship between the dips and the individual resonators. Detuning one resonator does not affect the dips associated with the preceding resonators. This example also illustrates the fact that the dips in the S_{11} time-domain response are minimized only when the resonators are perfectly tuned. Detuning in either direction causes the dips to deviate from the minimum value.

Alignment of the resonator frequencies is achieved by tuning the resonators successively, starting from resonator 1, until each dip is minimized (placed as low as possible). The adjustments of the individual resonators are mainly independent; however, for highly detuned filters, tuning a succeeding resonator may cause the dip of the previous resonator to deviate slightly from its minimum. Thus, after completing the tuning process of all the resonators, we may need to go through another round, starting from resonator 1, to fine-tune the dips to their minimum position. In high-order filters, the dips associated with the last few resonators may be

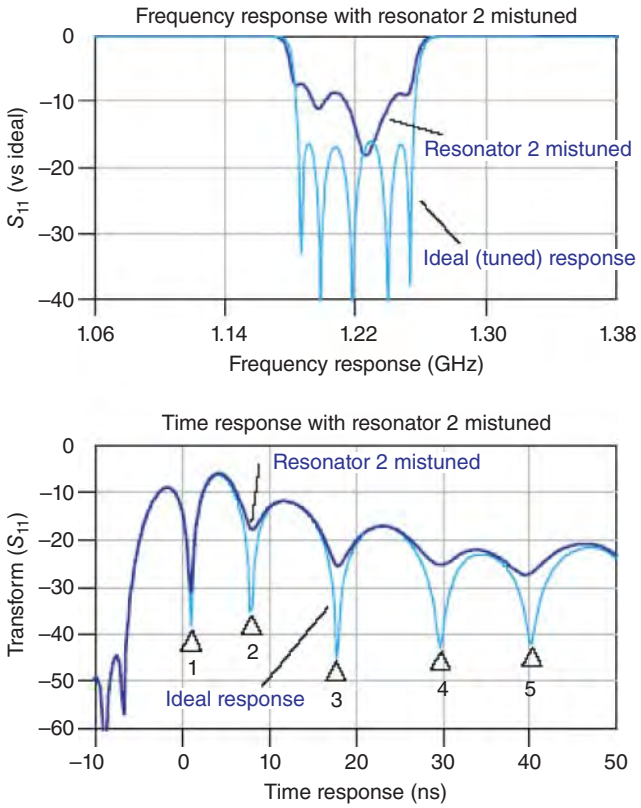


Figure 19.8 Frequency response and S_{11} time-domain response when resonator 2 is detuned (the lighter trace is the ideal response; the darker trace is after detuning) [27].

not appear as distinct as those associated with the first few resonators. In this case, we use the time-domain response of S_{22} to follow the same tuning process starting from the last resonator and moving toward the middle. The first dip in the S_{22} time domain is associated in this case with the resonance frequency of the last resonator.

19.4.2 Time-Domain Tuning of Interresonator Coupling

To demonstrate the effects of interresonator couplings on the time-domain response of S_{11}, we examine what happens to the time-domain response of the five-pole filter when the first and second interresonator coupling elements M_{12} and M_{23} are varied. Figure 19.10 shows the frequency response and S_{11} time-domain response, where the first interresonator coupling M_{12} is increased by 10%. The ideal response is also shown in this figure in the lighter trace. We see that the filter bandwidth is slightly wider and the return loss has changed when the input coupling M_{12} is increased. This is expected since increasing the coupling means more energy passing through the filter, resulting in a wider bandwidth. The time-domain response exhibits different peaks between the dips. The ones of interest to us are the peaks that are located at $t > 0$, namely, the peaks that are located between the dips are associated with the individual resonators.

It is noteworthy that increasing the coupling M_{12} causes the magnitude of the first peak (the one that is located after the dip associated with the first resonator) to decrease. The reduction in the level of this peak, when the interresonator coupling M_{12} is increased, makes sense and can

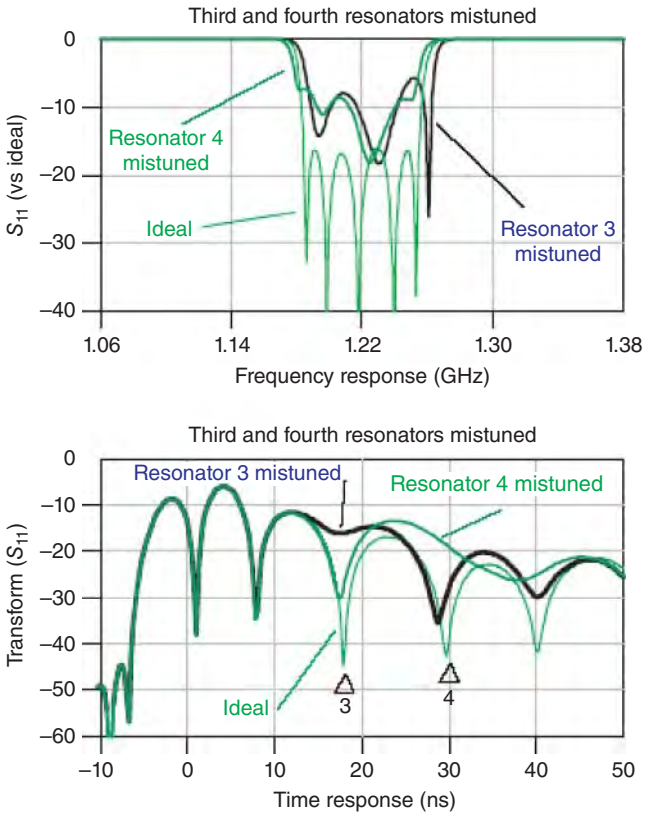


Figure 19.9 The frequency response and S_{11} time-domain response when resonator 3 is detuned (lighter trace indicates ideal response; darker trace, after detuning) [27].

be explained as follows: increasing the coupling means that more energy is coupled to the next resonator; thus less energy is reflected, and the peak corresponding to reflected energy from that coupling element should decrease. It is also noted that the level of the peaks that follow are higher in comparison with the ideal case. When energy is being coupled through the first coupling aperture, there is more energy to reflect off the remaining coupling apertures [27].

Figure 19.11 illustrates the case where the second interresonator coupling element M_{23} is reduced by 10%. We can see that although reducing M_{23} has no effect on the first peak, it causes the level of the second peak to increase (the peak that is located after the dip associated with the second resonator). This is expected since more energy is being reflected as a result of the decreased coupling M_{23}. The peaks that follow have a lower level, since the amount of energy coupled to the following apertures is reduced.

This example demonstrates how well the time-domain response separates the effects of changing each interresonator coupling, allowing the couplings to be individually adjusted. It is important, however, to recognize that changing the first interresonator coupling M_{12} does affect the responses of all the following peaks. Consequently, we begin with the tuning of the first interresonator coupling and move toward the middle of the filter for the successive tuning of the other interresonator couplings. An improperly tuned coupling near the input can mask the real response of the inner coupling elements. As in the case of tuning the resonance frequency, we can use the time-domain response of S_{11} to tune the coupling elements that are

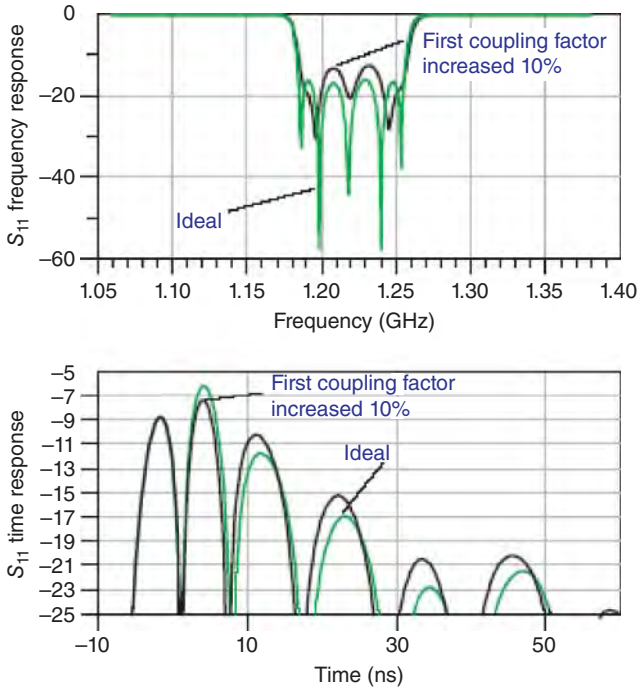


Figure 19.10 The frequency response and S_{11} time-domain response when M_{12} is increased by 10% (the lighter trace indicates the ideal response before increasing M_{12}) [27].

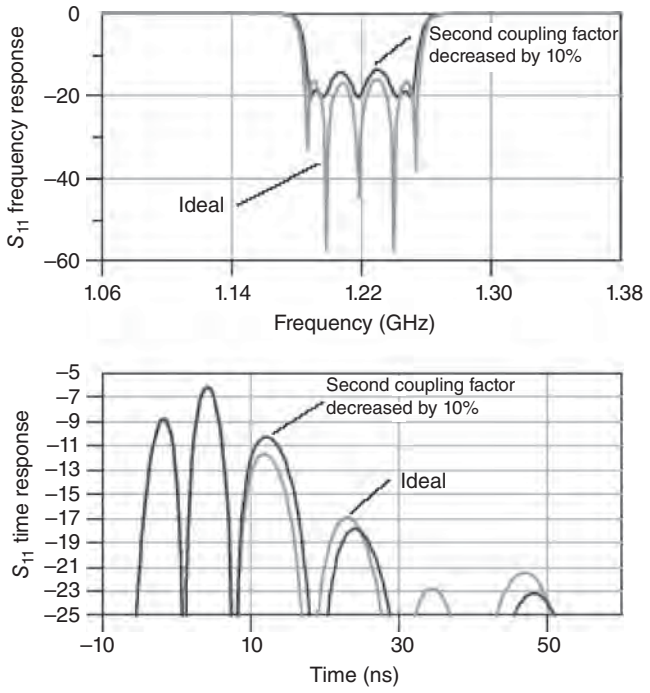


Figure 19.11 The frequency response and S_{11} time-domain response when M_{23} is reduced by 10% (the lighter trace indicates the ideal response before decreasing M_{23}) [27].

close to the input and use the time-domain response of S_{22} to tune those elements that are located near the output.

19.4.3 Time-Domain Response of a Golden Filter

It is obvious that tuning the interresonator couplings requires the availability of a target time-domain response to use as a reference to adjust the peak level. This target time-domain response can be obtained by taking the Fourier transform of the S_{11} frequency response of an ideal filter. The simulated time-domain response is then downloaded to the VNA and used as a template. Also, the target response can be obtained by measuring the time-domain response of an experimental filter (a "golden" filter) that is properly tuned. Each subsequent filter can be tuned to obtain the same response. However, it should be mentioned that the peak levels in the time-domain response depend on the ratio of the filter bandwidth to the frequency sweep used to compute the time-domain transform. The larger the frequency sweep/filter bandwidth ratio, the more total energy is reflected, which in turn will result in increased peak levels. Thus, during the tuning process, we need to maintain the same frequency sweep used to generate the target response. It is also essential to set the center frequency of the analyzer's frequency sweep during tuning equal to the desired center frequency of the filter. Typically a frequency span, that is, two to five times the bandwidth of the filter is needed. A span that is too narrow does not provide sufficient resolution to differentiate between the individual sections of the filter, whereas too wide a span causes too much energy to be reflected, reducing the tuning sensitivity.

For the network analyzer time-domain setup [27], the bandpass mode must be used. The start and stop times are set so that the individual resonators can be seen. For most filters, the start time should be set slightly before zero time, and the stop time should be set somewhat longer than twice the group delay of the filter. The correct settings can be approximated by setting the start time at $t = -(2/\pi\,\mathrm{BW})$ and the stop time at $t = (2N + 1)/(\pi\,\mathrm{BW})$, where BW is the filter's bandwidth and N is the number of filter sections. This should give a little extra time-domain response before the start of the filter and after the end of the filter time response. If both S_{11} and S_{22} responses are used in time-domain tuning of the filter, one can set the stop time to a smaller value, since the S_{22} response will be used to tune the resonators that are farther out in time (and closer to the output port). The format to use for viewing the time-domain response is log magnitude (dB). It is usually helpful to set the top of the screen at 0 dB.

It should be mentioned that tuning of the resonator frequencies can be accomplished without the need for the golden filter target time-domain response, since only one needs to minimize the dips. The golden filter time-domain response is needed only when tuning interresonator couplings. The time-domain technique does not deal with adjusting the input/output coupling elements R_1 and R_2. We need to use other tuning techniques to tune these elements. The technique is also not applicable for handling elliptic filters that have coupling between nonadjacent resonators. Detailed examples on time-domain tuning are given in [27].

19.5 Filter Tuning Based on Fuzzy Logic Techniques

The idea of using FL in filter tuning was introduced for the first time to the microwave community by Miraftab and Mansour [28–30]. The idea stems from the fact that during the manual tuning process, technologists use the concept of sets to describe the various coupling elements. Nevertheless, they converge to the proper solution and manage to align the filter perfectly. Examples of these sets for the filter coupling elements are very small, small, large, very large,

and so on. Some of these technologists have the expertise to look at the measured filter response and conclude that a specific coupling element, for example, is very small, and accordingly they take the action of adjusting the tuning screw to increase the coupling of this particular element. This process is repeated several times during the tuning process. While they do not have any means to find the exact value of the coupling elements, associated with a particular measured filter response, they manage to tune the filter and converge to the desired response. This type of human thought process can be duplicated by the FL technique, which also uses the concept of sets.

The computer-aided tuning techniques described in Sections 19.2–19.4 are based on implementing a mathematical model to guide the interpretation of the measured data. The FL technique is capable of combining mathematical models, measured data, and human experience into one single model. This is possible since FL systems interpret and process numerical information and linguistic (expert) information the same way.

The FL technique itself was conceived by Zadeh [31]. It is now well adapted in many engineering fields [32–37]. It has become a standard technology in industrial automation, pattern recognition, and medical diagnosis systems. Nevertheless, microwave researchers have been hesitant to use this technique because of the perceived fuzziness aspect of the concept. In fact, there is nothing "fuzzy" about this technique. It is simply a "function approximator"—actually a very good function approximator. It has been shown in [32] that an FLS can approximate any real continuous nonlinear function to an arbitrary degree of accuracy.

19.5.1 Description of Fuzzy Logic Systems

In classical Boolean logic, sets are defined in a crisp manner; an element either belongs to a set or does not belong to it. In FL, a membership value between binary 0 and 1 is assigned to each element of the set. A value of 0 means that the element does not belong to the set at all, whereas a value of 1 means the element totally belongs to that set. FL interprets the numerical data as linguistic rules. Then the extracted rules will be used as a kind of system specification to calculate the output values of the system.

The FL system can be described as a function approximator as shown in Figure 19.12. Overall, the FL systems map crisp inputs into crisp outputs. The system has four components: fuzzifier, rules, inference engine, and defuzzifier. Once the rules are established, the FL system can be viewed to perform the function $y = f(x)$. Rules can be expert rules, or rules extracted from numerical data. In either case, the rules are expressed as a collection of IF-THEN statements.

The fuzzifier maps crisp input numbers into fuzzy sets. The purpose of the inference engine is to generate the output fuzzy sets using the rules and the input fuzzy sets. There are many different types of FL inferential procedures. However, a small number of them are usually being used in engineering applications. This is similar to the human method of making decisions,

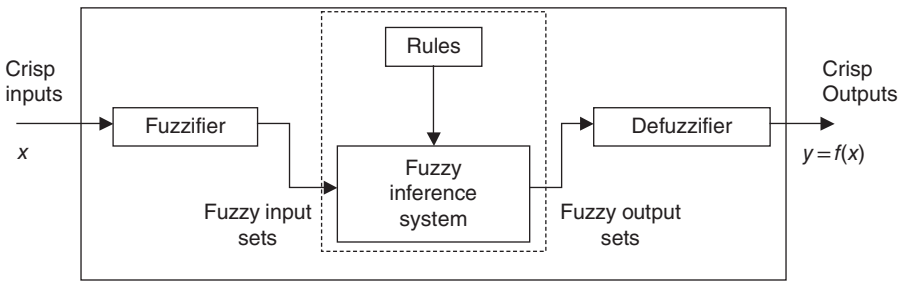


Figure 19.12 A block diagram of the fuzzy logic system.

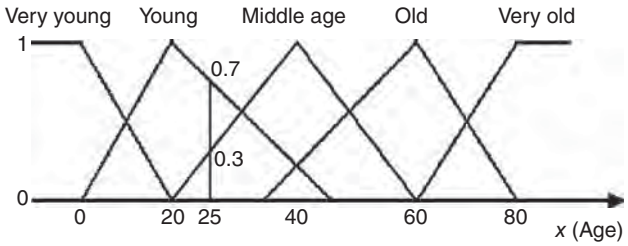


Figure 19.13 Functions for X (age) = {very young, young, middle age, old, very old}.

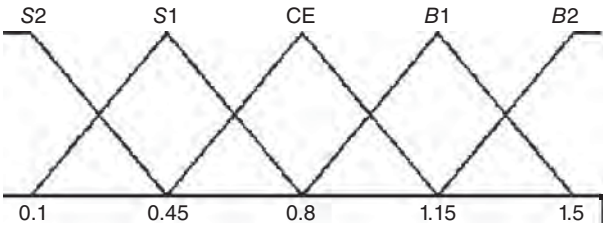


Figure 19.14 Potential membership functions for the elements of the coupling matrix.

where different types of inferential procedures are used to understand things. The defuzzifier maps fuzzy output sets into crisp output numbers. This step is necessary since we need to obtain a crisp number in most of the engineering applications.

As an example of fuzzy sets, consider Figure 19.13, which shows different membership functions for the variable "age." We can interpret "age" as a variable X. This variable can be decomposed into the following sets of terms: X (age) = {very young, young, middle age, old, very old}, where each term in X (age) is characterized by a fuzzy set in the universe of discourse $U = \{0, 80\}$. For example, as can be seen from Figure 19.13, we can interpret a person with the age 25 to belong to both fuzzy sets of "young" and "middle age," but with different membership functions of 0.7 and 0.3, respectively. This example demonstrates that in FL an element can reside in more than one set with different degrees of similarity. We can treat coupling values in a similar fashion; for example, we can select membership functions for the coupling element M_{ij}, as shown in Figure 19.14. The sets can be identified as $S2$, $S1$, CE, $B1$, and $B2$, covering the values that the coupling values M_{ij} might take.

The most commonly used shapes for the membership functions are triangular, trapezoidal piecewise linear, and Gaussian. The membership functions are designed by using an optimization procedure [30], but they are usually based on the user's experience for a particular problem. The number of membership functions is up to the designer. By choosing more membership functions, we can achieve a better resolution at the price of more computational complexity.

19.5.2 Steps in Building the FL System

To illustrate the main steps involved in the FL technique, we consider the following example. Assume that we have the set of data given in Table 19.2. We show in this example how FL can be used as a function approximator $y = f(x)$. Thus, the problem is: Given these sets of data pair points, use the FL technique to calculate the value of the output y for $x_1 = 0.5, x_2 = 0.3$. The steps for building the FL are as follows:

Step 1. Assigning the membership functions for all variables. Details on selecting and optimizing the shape of the membership functions are found in the literature [30–33].

Table 19.2 Example for a set of data.

x_1	x_2	y
0.3	0.7	0.5
0.6	0.5	0.1
0.1	0.2	0.9

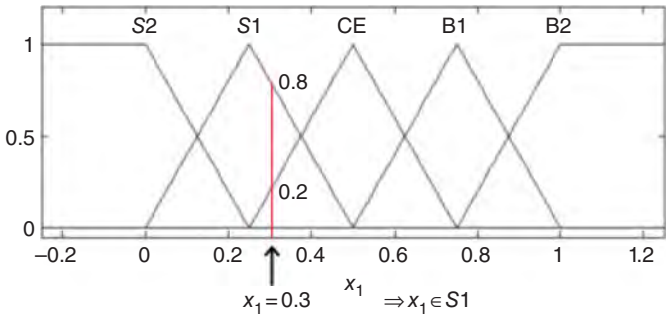


Figure 19.15 The membership function for variable x_1.

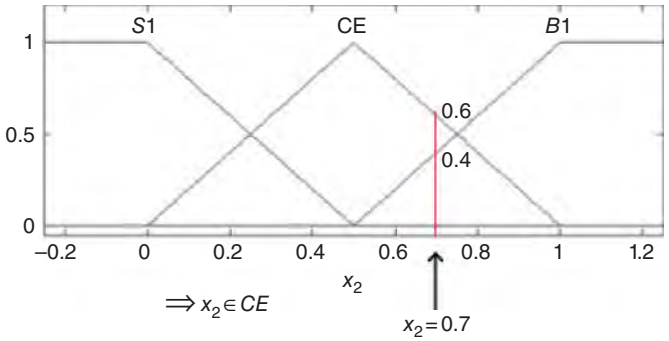


Figure 19.16 The membership function for variable x_2.

To illustrate the concept, let us select the simplest form of membership function, which is triangular, as shown in Figures 19.15–19.17. Five membership functions ($S2, S1, \text{CE}, B2, B1$) are selected for variable x_1, three membership functions are selected for variable x_2, and three membership functions ($S1$, CE, $B1$) are selected for the variable y.

Step 2. Creating the IF-THEN fuzzy rules. Many approaches are available for generating fuzzy rules from numerical data [32–34]. In this section we use the method proposed in [33], since it allows the combination of both numerical and linguistic information into a common simple framework. The rules are created by using the data pairs given in Table 19.2 as well as the membership functions. The construction of the IF-THEN rule associated with the first data pair (for $x_1 = 0.3$ and $x_2 = 0.7$, $y = 0.5$) is also illustrated in Figures 19.15–19.17. As shown in Figure 19.15, the point $x_1 = 0.7$ belongs to the set $S1$ with a membership value 0.8 and to the set CE with a membership value of 0.2. Since its membership value to set $S1$ is higher than its membership value to set CE, it is assumed

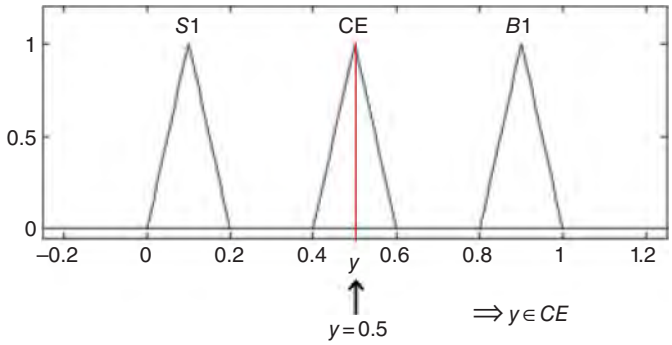


Figure 19.17 The membership function for variable *y*.

that it belongs to set *S*1 (i.e., x_1 is *S*1). Similarly, one can see from Figures 19.16 and 19.17 that the point $x_2 = 0.5$ belongs to set CE and point $y = 0.5$ belongs to the set CE. Thus the rule associated with the first data pair ($x_1 = 0.3, x_2 = 0.7$ and $y = 0.5$) is

Rule 1. IF (x_1 *is S*1) *and* (x_2 *is CE*), *THEN* (*y is CE*).

Similarly, the two other data pairs in Table 19.2 can be used to generate rules 2 and 3. Using the same concept, we can readily show that these two rules are as follows:

Rule 2. IF (x_1 *is CE*) *and* (x_2 *is CE*), *THEN* (*y is S*1).

Rule 3. IF (x_1 *is S*2) *and* (x_2 *is S*1), *THEN* (*y is B*1).

Notice that in this step, we effectively converted the three crisp data pairs given in Table 19.2 into three IF-THEN rules.

Step 3. Fuzzy inference and defuzzification process. In this step we find the output *y* for the given inputs $x_1 = 0.5$ and $x_2 = 0.3$ using the centroid defuzzification theorem [29]:

$$y_i = \frac{\sum_{j=1}^{K} m_j y_i^j}{\sum_{j=1}^{K} m_j}$$

where $m_j = m_j(x_1)m_j(x_2), \ldots$, where y_i^j denotes the center value of the fuzzy set corresponding to rule *j* and output y_i. The term $m_j(x_k)$ is the membership value of x_k to the fuzzy set corresponding to a rule *j* and input x_k. *K* is the number of rules. Note that as long as the output membership functions are symmetric, y_i^j is simply the center of each membership function. If we chose a nonsymmetric membership, we would need to calculate the center of gravity of each membership function as y_i^j. By choosing symmetric membership functions, less computation is required. In this example, $m_1(x_1)$ and $m_1(x_2)$ are calculated by looking at rule 1 and finding the membership values of $x_1 = 0.4$ and $x_2 = 0.3$. In view of rule 1 and Figures 19.15 and 19.16, we can readily show that the point $x_1 = 0.4$ belongs to the set *S*1 by a membership value of 0.4 [i.e., $m_{S1}(0.4) = 0.4$] while point $x_2 = 0.3$ belongs to the set CE with a membership value of 0.6 [i.e., $m_{\text{CE}}(0.3) = 0.6$]. Thus we have the following set of rules:

Rule 1. IF (x_1 *is S*1) *and* (x_2 *is CE*), *THEN* (*y is CE*):

$$m_1 = m_{S1}(0.4)m_{\text{CE}}(0.3) = 0.4 \times 0.6 = 0.24, \quad \bar{y}_1 = 0.5$$

Similarly, one can show that for rules 2 and 3 we have.

Rule 2. IF (x_1 *is CE*) *and* (x_2 *is CE*), *THEN* (y *is S*1):

$$m_2 = m_{CE}(0.4 rm_{CE}(0.3) = 0.6 \times 0.6 = 0.36, \quad \bar{y}_2 = 0.1$$

Rule 3. IF (x_1 *is S*2) *and* (x_2 *is S*1), *THEN* (y *is B*1):

$$m_3 = m_{S2}(0.4) m_{S1}(0.3) = 0 \times 0.4 = 0, \quad \bar{y}_3 = 0.9$$

The value of y corresponding to $x_1 = 0.4$ and $x_2 = 0.3$ is as follows:

$$y_i = \frac{\sum_{j=1}^{3} m_j y_i^j}{\sum_{j=1}^{3} m_j} = \frac{m_1 \bar{y}_1 + m_2 \bar{y}_2 + m_3 \bar{y}_3}{m_1 + m_2 + m_3}$$

$$= \frac{0.24 \times 0.5 + 0.36 \times 0.1 + 0 \times 0.9}{0.24 + 0.36 + 0} = 0.26$$

Note that rules 1 and 2 are fired, while rule 3 is not fired. In this example, the FL system acts as a function approximator. The solution for $x_1 = 0.4$ and $x_2 = 0.3$ is $y = 0.26$.

19.5.3 Comparison between Boolean Logic and Fuzzy Logic

In this section, we apply both FL and Boolean logic to approximate the simple function $y = x^2 + 1$ as an example to illustrate that the logic system is in fact a good function approximator [38]. A function approximator based on Boolean logic divides all the input values into groups, and each group is associated with an output value. Since each input value must belong to a single output value group, the output curve will often appear in "staircase" shape. The accuracy of the approximator depends on how the input values are grouped together and how many groups are used. However, the "staircase" nature of the approximator cannot be avoided by any means (since it's Boolean), as shown in Figure 19.18.

In this example, the Boolean logic function approximator divides the input values into 10 groups. The ranges of x values that each group covers are the same for design simplicity. The output value of each group is derived by taking the middle value of each group and calculates the corresponding output value using the function itself ($y = x^2 + 1$). In the example, 10 output

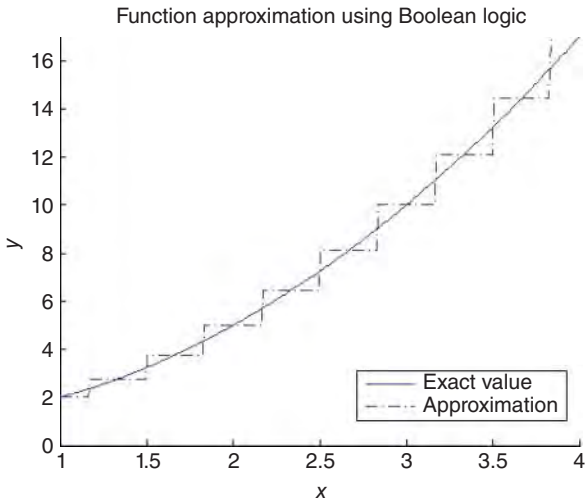


Figure 19.18 Function approximation using Boolean logic.

values are used. The input/output values used are (1, 2), (1.33, 2.78), (1.67, 3.78), (2, 5), (2.33, 6.44), (2.67, 8.11), (3, 10), (3.33, 12.11), (3.67, 14.44), and (4, 17).

With the availability of a set of data pairs, we can follow an approach similar to that outlined in Section 19.5.2 to build an FL system to approximate this function. On the other hand, FL tools are available in MATLAB [36], allowing users to easily build FL systems. Users can simply define the number of inputs and outputs of the system and then define the number of membership functions for each input and output. MATLAB FL tools offer a number of membership function types, such as the triangular function and the Gaussian function, for users to choose from. The evaluation of FL system is extremely easy once all parameters (inputs, outputs, and rules) of the fuzzy inference system (FIS) are set. MATLAB's FL tools take the input values and return the output values. All fuzzification and defuzzification processes are internal to MATLAB FL tools.

By adjusting the shape of the input membership function, the number of rules, and the shape of the output membership function, we can develop a highly accurate function approximator using just a few data points. Figure 19.19 shows the output of an FL function approximator developed by the four data points.

The FL system used to construct the approximator is a Sugeno-type fuzzy system [35] that has one input variable (x). The input x contains four triangular membership functions centered at 1, 2, 3, and 4. The four data pairs used are (1, 2), (2, 5), (3, 10), and (4, 17). Four IF-THEN rules are generated from these data points. Triangular membership functions were chosen to simplify calculation.

The FL approximator can be improved by using Gaussian membership functions. MATLAB FL tools are used to optimize these Gaussian membership functions. In this case more data pair points are needed. Some of these data pairs are used to optimize the Gaussian membership functions and the output values such that the output of the system will closely match the given data pairs. The process is called "training the fuzzy logic systems" and is described in [30]. In the optimized approximator, 10 data points (the same data points used to construct the Boolean logic approximator) are used to train and build the fuzzy system.

Figure 19.20 shows the output of the optimized function approximator, whereas Figure 19.21 shows the percentage error of the Boolean logic solution and the non-optimized and optimized FL solution. As shown in Figure 19.21, the error exists in using the Boolean logic function

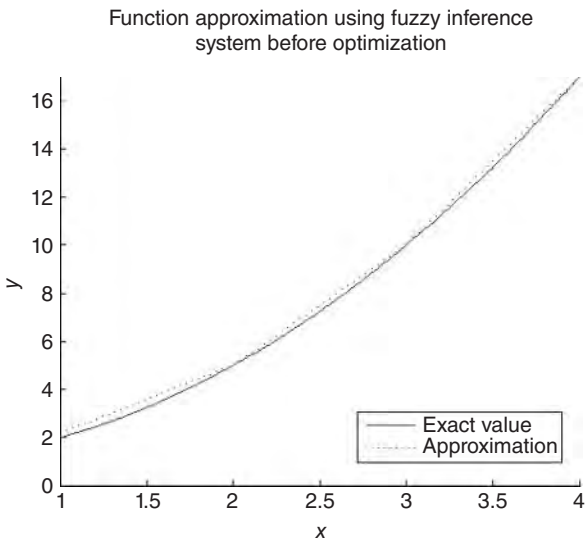


Figure 19.19 Function approximation using fuzzy logic.

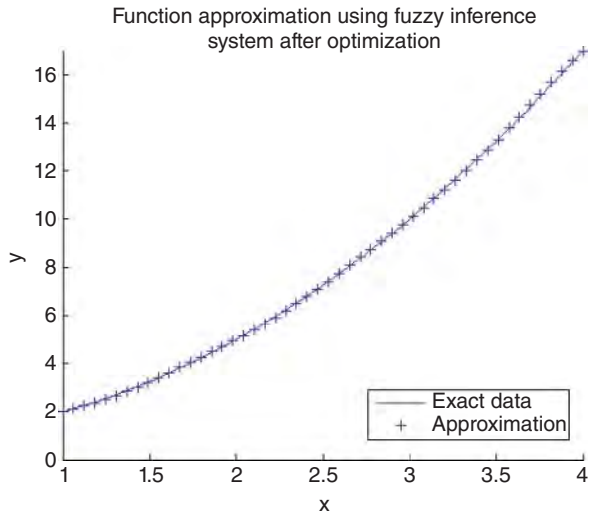


Figure 19.20 Function approximation using fuzzy logic with optimization.

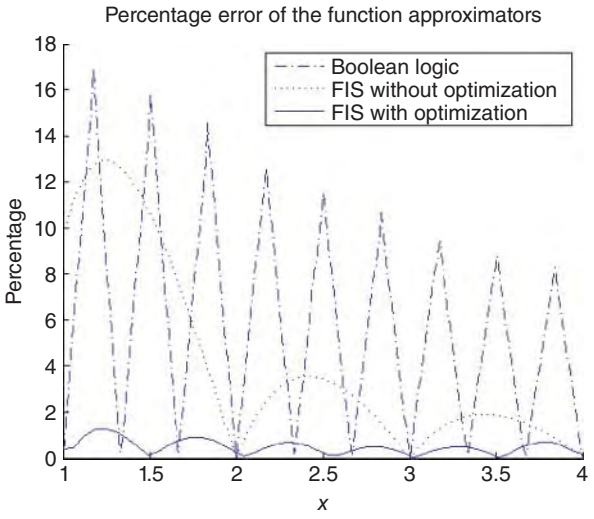


Figure 19.21 Percentage error of three different function approximators.

approximator, which fluctuates between a high value (9–17%) and 0. The FL function approximator has a maximum error of 14% and can be reduced to less than 2% with an optimized FL system. Although increasing the number of output groups in the Boolean function approximator can improve its accuracy, the FL function approximator is shown to be more accurate given the same number of data pairs.

19.5.4 Applying Fuzzy Logic to Filter Tuning

Three FL systems for tuning of microwave filters have been developed by Miraftab and Mansour [28–30, 39, 40] for filter tuning:

1) An objective FL system where the IF-THEN rules are constructed from data pairs generated from a mathematical model such as the filter circuit [29, 30].

2) An objective FL system where the IF-THEN rules are created by tracking a human expert actions during manual tuning [39].
3) A subjective FL system where the IF-THEN rules are generated directly from a human expert is proposed in [40]. This paper also demonstrates the feasibility of using the concept to construct a robot to carry out the tuning process [40].

Details of these three techniques are given in [41]. In this chapter we briefly describe the first approach where the IF-THEN rules are generated from a mathematical model [29, 30].

The mathematical model used in [29, 30] is the coupling matrix circuit model. The FL system in this case identifies the coupling matrix that is associated with a particular experimental filter response. The FL system acts as a function approximator $y = f(x)$, where x is the experimental performance and y is the coupling matrix. In Section 19.3 the same problem is solved by applying an optimization routine to the coupling matrix circuit model. The optimization involves many iterations and is considered an indirect approach method. The FL technique is a direct approach. Efforts are taken initially in building the FL system by using data pairs generated from the coupling matrix model. Once the system is constructed, it can be used directly to extract the coupling matrix associated with a particular experimental response.

We consider the M-matrix coupling coefficients as outputs and the S-parameters of the filter at different frequencies as inputs. Suppose that we have p frequency sampling points, that is, p inputs and q unknown coupling coefficients as outputs. We can extract the input information from either S_{21} or S_{11}. The inputs will then be in the form $S(f_1) \cdots S(f_p)$, which can be written in the form $x_1, x_2, \ldots, x_p$ for simplicity. The outputs, which are the coupling coefficients, could also be written in the form $y_1, y_2, \ldots, y_q$ for simplicity. The coupling matrix model [equations (19.19)–(19.22)] is used here to generate the initial data pairs. We select values for each coupling coefficient around the ideal design value and generate a number of input/output data pairs:

$$
\begin{aligned}
&(x_1^{(1)}, x_2^{(1)}, \ldots, x_p^{(1)}; \quad y_1^{(1)}, y_2^{(1)}, \ldots, y_q^{(1)}) \\
&(x_1^{(2)}, x_2^{(2)}, \ldots, x_p^{(2)}; \quad y_1^{(2)}, y_2^{(2)}, \ldots, y_q^{(2)}) \\
&\vdots \\
&(x_1^{(n)}, x_2^{(n)}, \ldots, x_p^{(n)}; \quad y_1^{(n)}, y_2^{(n)}, \ldots, y_q^{(n)})
\end{aligned}
$$

Once the data pairs are generated, we can employ the simple three steps described in Section 19.5.2 to build the FL system or use commercial FL software tools such as those of MATLAB. The IF-THEN rules are in the following form:

$$
\begin{aligned}
&IF(x_1 \; is \; fs_{x1}) \quad and \quad (x_2 \, is fs_{x2}) \ldots \\
&and \;\; (x_p \, is fs_{xp}), \quad THEN \\
&(y_1 \, is fs_{y1}) \ldots \quad and \;\; (y_q \, is fs_{yq})
\end{aligned}
$$

where f_s is a fuzzy set among the fuzzy sets of each input/output variable. To illustrate the concept, we consider the tuning of a four-pole bandpass Chebyshev filter. The coupling matrix (M-matrix) is a symmetric 4×4 matrix ($m_{ij} = m_{ji}$) with all elements zero except m_{12}, m_{23}, m_{34}. These are the variables that need tuning; R_1 and R_2 are assumed fixed in this example. The ideal coupling values for this filter are $m_{12} = 1.2$, $m_{23} = 0.95$, $m_{34} = 1.2$. Figure 19.22 depicts S_{21} versus the frequency of two detuned filters, one with a slight deviation and the other with a high deviation from the ideal filter performance.

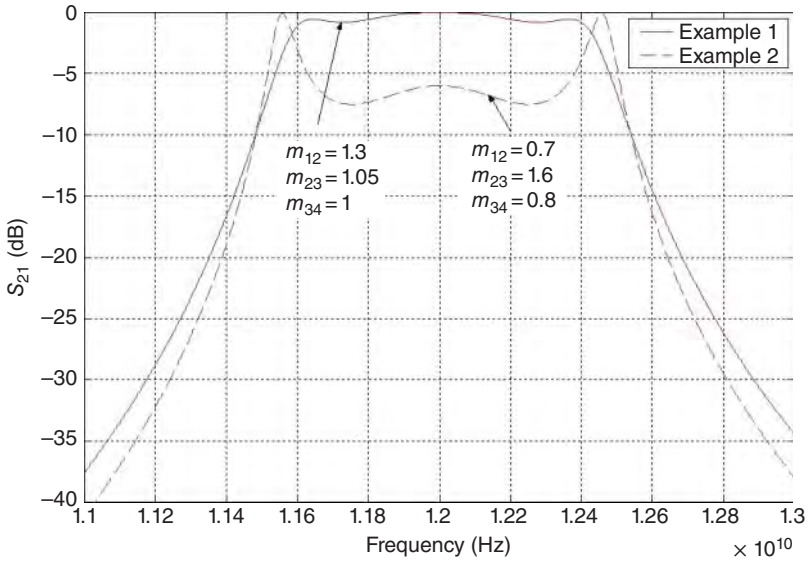


Figure 19.22 Examples of slightly detuned and highly detuned four-pole Chebyshev filter characteristics [29].

These two examples represent the experimental data of two detuned filters, each with two different deviations. In order to use the tuning procedure, we need to extract the M-matrix elements associated with the experimental results. Then, with the knowledge of the ideal coupling matrix, we can identify the elements that caused the detuning. In assigning the membership functions or fuzzy sets, we choose five fuzzy sets for each input variable. The membership functions for the output variables are selected as depicted in Figure 19.23 [28].

In this example the inputs are selected to be the magnitude of S_{21} at nine frequencies with seven frequencies inside the passband and the other two outside the passband. The FL system in this example has nine inputs and three outputs. Seventy IF-THEN rules are generated from data pairs obtained using the coupling matrix model [equations (19.19)–(19.22)]. The extracted response of the two filters (slightly detuned and highly detuned) is shown in Figures 19.24 and 19.25. The coupling elements are also shown in the same figures. The extracted responses are fairly close to the assumed experimental filter response.

The FL technique is able to integrate theoretical models, measured data, and human experience into one model. In particular, its main advantages lie in the ability to translate human experience into a model. Details on using human experience are given in [39–41].

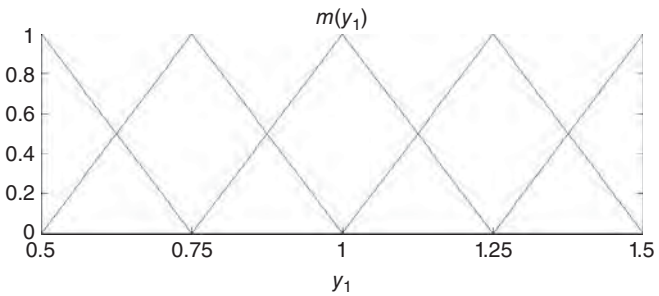


Figure 19.23 Output membership functions for the four-pole filter.

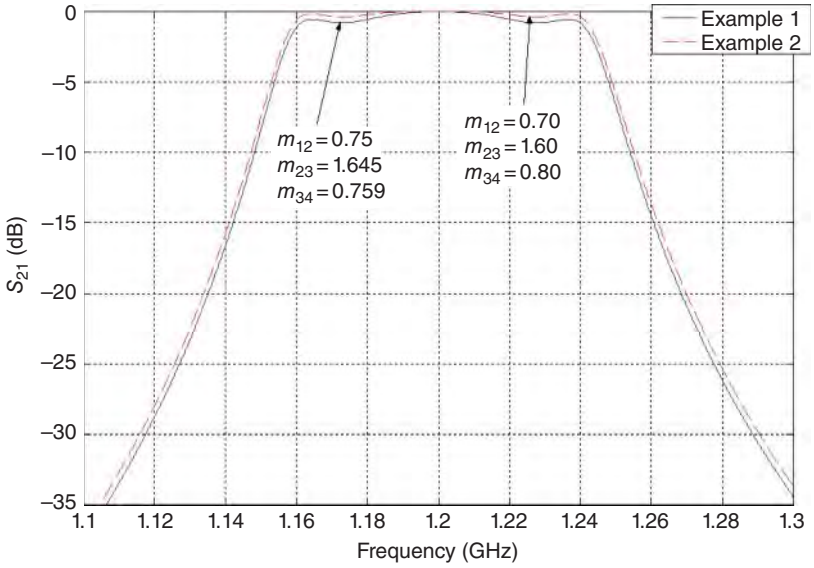


Figure 19.24 A comparison between experimental and extracted performance using fuzzy logic for the slightly detuned filter.

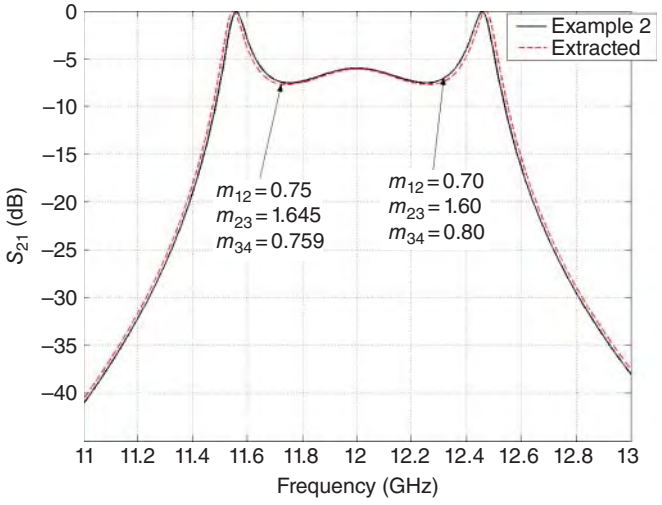


Figure 19.25 A comparison between experimental and extracted performance using fuzzy logic for the highly detuned filter.

19.6 Automated Setups for Filter Tuning

Figure 19.26 shows a schematic diagram for an automated filter tuning station (robot). It consists of a VNA, a computer, a motor with arms, and the filter under test with tuning screws. The filter measurements data are taken by the VNA and are transferred to the computer. The data are then processed by the tuning algorithm running on the computer, which in turn interfaces with the motor to provide the necessary tuning. Any of the five computer-aided tuning techniques described in this chapter can be automated to construct a robot for tuning microwave filters.

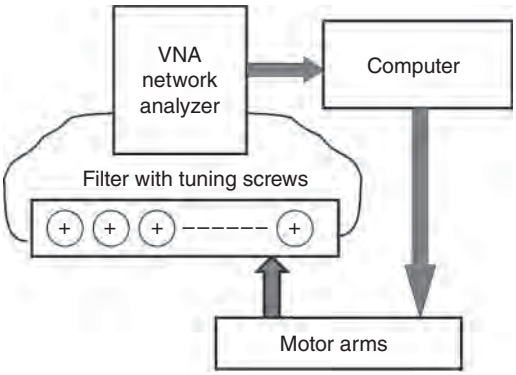
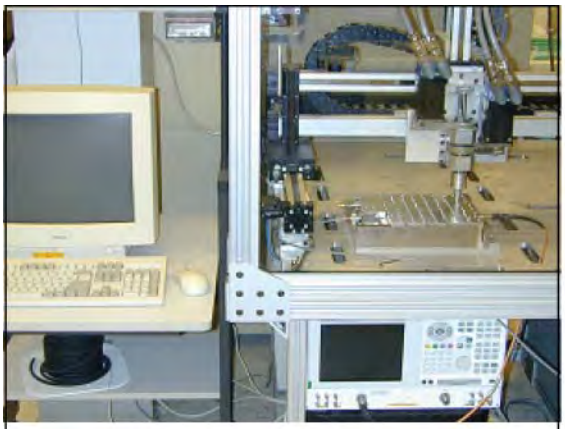
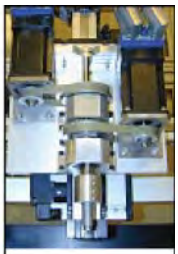


Figure 19.26 The automated tuning station.



RoboCAT system at COM DEV



Coaxial screw/nut driver handles standard tuning screw and lock nut

Figure 19.27 COM DEV's RoboCat. *Source*: Yu [43] 2006. Reproduced with the permission of Microwave Journal.

COM DEV, one of the world's leading suppliers of filters and multiplexers, has reported the first robotic system for filter tuning [42, 43]. Figure 19.27 shows a photograph of COM DEV's RoboCat system [43]. The system has a replaceable coaxial screw/nut driver head that allows simultaneous independent servo mechanically driven control of a tuning screw and its locknut. This capability is crucial in allowing the robot to perform the final tuning screw lockdown on sensitive products where the lockdown procedure can affect the device performance. The system is able to predict the effect of tightening the nut and make microscale adjustments to the tuning screw position to counteract the associated detuning. The screw/nut driver can be easily swapped to handle different screw/nut combinations and has been used on screws of various sizes. The system is fully compatible with self-locking screws and other nutless screw designs. It is capable of tuning coaxial diplexers for wireless base stations in a few minutes. More details on the system are given in [43].

Figure 19.28 shows a robotic system for filter tuning employing several motors [40]. The system is capable of tuning filters where the tuning screws are scattered over the filter body. The screws are connected to servo/step motors by custom-designed flexible leads to connect the screws to motor shafts. Universal mounting brackets are designed to arbitrarily hold and position the motors on a main plate. This enables adjustability in three dimensions to target any position on a particular filter or multiplexer. More details on this system are given in [40].

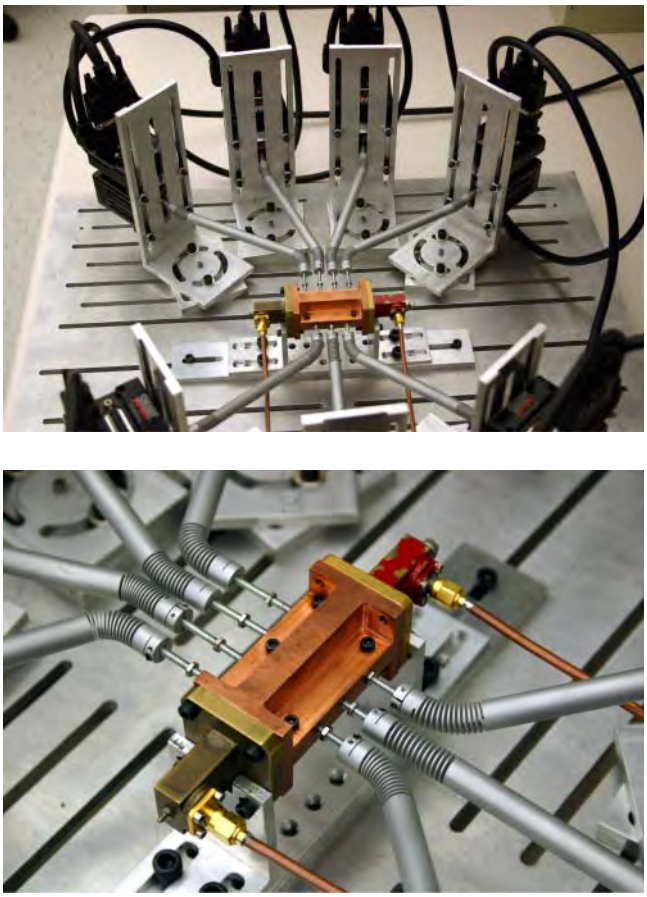


Figure 19.28 An automated filter tuning setup. *Source*: Miraftab and Mansour 2006 [40]. Reproduced with the permission of IEEE.

Summary

From a theoretical standpoint, the physical dimensions of a microwave filter can be perfected using EM-based techniques with near arbitrary accuracy. In practice, the use of EM-based tools can be very time consuming and prohibitively so for higher-order filters and multiplexing networks. Moreover, a perfect EM-based design cannot provide measured results that exactly match theoretical results. This is due to manufacturing tolerances and variations in the characteristics of the materials used to construct microwave filters. These problems are further exacerbated by the very stringent performance requirements for applications in the wireless and satellite communication systems. For these reasons, the vast majority of practical microwave filters require tuning to achieve the desired response. This chapter is devoted to the computer-aided techniques for tuning microwave filters.

By starting with a good EM-based design (Chapters 14 and 15), tuning can be invoked to achieve a measured response that meets all design requirements. Tuning is usually carried out by experienced technologists and is considered a key bottleneck in filter production. For many decades, tuning has been considered more of an art than a science. Lately, several

computer-aided tuning techniques have emerged for the purpose of augmenting the science part of the tuning process.

This chapter presents five different techniques for computer-aided tuning of microwave filters. These include filter tuning procedures based on sequentially tuned resonators, circuit model extraction parameters, pole–zero configuration of the filter function, time-domain response of filters, and FL systems. The sequential and the time-domain tuning techniques are useful in guiding the technologists to tune the resonators and coupling elements, one by one. Such techniques have been successfully implemented in tuning Chebyshev filters. The circuit model parameter extraction tuning technique has several variations. This technique has been successfully employed to tune large-order elliptic function filters as well as diplexers and multiplexers. The FL technique is investigated to deal with both the science and the art parts of the tuning process. The technique has been successfully employed to rough-tune Chebyshev filters. An outline of other concepts on computer-aided tuning is given in [44].

This chapter also includes an outline for two automated filter tuning setups. The first setup has a built-in mechanism, making it possible to lock the screws once the filter is tuned. The second setup is built with flexible arms, which make it possible to tune microwave filters and multiplexers with tuning screws scattered over the three-dimensional filter body. It also allows the tuning of several screws simultaneously, which can help considerably in speeding up the tuning process.

References

1 Dishal, M. (1951) Alignment and adjustment procedure of synchronously tuned multiple resonant circuit filters. *Proceedings of the IRE*, **39**, 1448–1455.

2 Wheeler, H.A. (1954) Tuning of waveguide filters by perturbing of individual sections. Procedings of Symposium on Modern Advances in Microwave Technology, pp. 343–353.

3 Atia, A.E. and Williams, A.E. (1974) Nonminimum phase optimum amplitude band pass waveguide filters. *IEEE Transactions on Microwave Theory and Techniques*, **22**, 425–431.

4 Atia, A.E. and Williams, A.E. (1975) Measurement of intercavity couplings. *IEEE Transactions on Microwave Theory and Techniques*, **23**, 519–522.

5 Williams, A.E., Egri, R.G., and Johnson, R.R. (1983) Automatic measurement of filter coupling parameters. IEEE MTT-S International Microwave Symposium Digest, pp. 418–420.

6 Accatino, L. (1986) Computer-aided tuning of microwave filters. IEEE MTT-S International Microwave Symposium Digest, pp. 249–252.

7 Thal, H.L. (1978) Computer aided filter alignment and diagnosis. *IEEE Transactions on Microwave Theory and Techniques*, **26**, 958–963.

8 Ishizaki, T., Ikeda, H., Uwano, T., Hatanaka, M., and Miyake, H. (1990) A computer aided accurate adjustment of cellular radio RF filters. IEEE MTT-S International Microwave Symposium Digest, pp. 139–142.

9 Mirzai, A.R., Cowan, C.F.N., and Crawford, T.M. (1989) Intelligent alignment of waveguide filters using a machine learning approach. *IEEE Transactions on Microwave Theory and Techniques*, **37**, 166–173.

10 Antreich, K., Gleissner, E., and Muller, G. (1975) Computer aided tuning of electrical circuits. *Nachrichtentech. Z.*, **28** (6), 200–206.

11 Adams, R.L. and Manaktala, V.K. (1975) An optimization algorithm suitable for computer-assisted network tuning. Proceedings of IEEE International Symposium on Circuits and Systems, Newton, MA, pp. 210–212.

12 Bandler, J.W. and Salama, A.E. (1985) Functional approach to microwave postproduction tuning. *IEEE Transactions on Microwave Theory and Techniques*, **33**, 302–310.

13 Lopresti, P.V. (1977) Optimum design of linear tuning algorithms. *IEEE Trans. Circuits Syst.*, **24**, 144–151.

14 Alajajian, C.J., Trick, T.N., and El-Masry, E.I. (1980) On the design of an efficient tuning algorithm. Proceedings of IEEE International Symposium on Circuits and Systems, Houston, TX, pp. 807–811.

15 Accatino, L. and Carle, P. (1984) *Computer Aids for Designing and Simulating Microwave Multiplexer for Satellite Applications*, GLOBECOM, Atlanta, Georgia.

16 Ness, J.B. (1998) A unified approach to the design, measurement, and tuning of coupled-resonator filters. *IEEE Transactions on Microwave Theory and Techniques*, **46**, 343–351.

17 Yu, M. (1994) Computer Aided Tuning. COM DEV internal report.

18 Kahrizi, M., Safavi-Naeini, S., and Chaudhuri, S.K. (2000) Computer diagnosis and tuning of microwave filters using model-based parameter estimation and multi-level optimization. IEEE MTT-S International Microwave Symposium Digest, pp. 1641–1644.

19 Harscher, P. and Vahldieck, R. (Nov. 2001) Automated computer-controlled tuning of waveguide filters using adaptive network models. *IEEE Transactions on Microwave Theory and Techniques*, **49** (11), 2125–2130.

20 Harscher, P., Amari, S., and Vahldieck, R. (2001) Automated filter tuning using generalized low pass prototype networks and gradient-based parameter extraction. *IEEE Transactions on Microwave Theory and Techniques*, **49**, 2532–2538.

21 Yu, M. (2001) Simulation/design techniques for microwave filters—an engineering perspective. Proceedings of Workshop WSA: State-of-the-Art Filter Design Using EM and Circuit Simulation Techniques, International Symposium IEEE Microwave Theory and Techniques, Phoenix.

22 Pepe, G., Gortz, F.J., and Chaloupka, H. (2004) Computer-aided tuning and diagnosis of microwave filters using sequential parameter extraction. IEEE MTT-S International Microwave Symposium Digest.

23 Miller, E.K. and Burke, G.J. (1991) Using model-based parameter estimation to increase the physical interpretability and numerical efficiency of computational electromagnetics. *Computer Physics Communications*, **69**, 43–75.

24 Hsu, H., Yao, H., Zaki, K.A., and Atia, A.E. (2002) Computer-aided diagnosis and tuning of cascaded coupled resonators filters. *IEEE Transactions on Microwave Theory and Techniques*, **50**, 1137–1145.

25 Dunsmore, J. (1999) Simplify filter tuning using time domain transformers. Microwaves RF (March).

26 Dunsmore, J. (1999) Tuning band pass filters in the time domain. IEEE MTT-S International Microwave Symposium Digest, pp. 1351–1354.

27 Agilent application note, Agilent AN 1287–8.

28 Miraftab, V. and Mansour, R.R. (2002) Computer-aided tuning of microwave filters using fuzzy logic. IEEE MTT-S International Microwave Symposium Digest, vol. 2, pp. 1117–1120.

29 Miraftab, V. and Mansour, R.R. (2002) Computer-aided tuning of microwave filters using fuzzy logic. *IEEE Transactions on Microwave Theory and Techniques*, **50** (12), 2781–2788.

30 Miraftab, V. and Mansour, R.R. (2004) A robust fuzzy technique for computer-aided diagnosis of microwave filters. *IEEE Transactions on Microwave Theory and Techniques*, **52** (1), 450–456.

31 Zadeh, L.A. (1965) Fuzzy sets. *Information and Control*, **8**, 338–353.

32 Kim, Y.M. and Mendel, J.M. (1995) Fuzzy basis functions: comparisons with other basis functions. *IEEE Transactions on Fuzzy Systems*, **43**, 1663–1676.

33 Wang, L.X. and Mendel, J. (1992) Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Transactions on Neural Networks*, **3** (5), 807–814.

34 Mouzouris, G.C. and Mendel, J.M. (1994) Non-singleton fuzzy logic systems. Proceedings of 1994 IEEE Conference Fuzzy Systems, Orlando, FL.

35 Sugeno, M. and Tanaka, K. (1991) Successive identification of a fuzzy model and its applications to prediction of a complex system. *Fuzzy Sets and Systems*, **42** (3), 315–334.

36 Mendel, J.M. (1995) Fuzzy logic systems for engineering. *Proceedings of the IEEE* (Special Issue on Engineering Applications of Fuzzy Logic), **83** (3).

37 Larsen, P.M. (1980) Industrial applications of fuzzy logic control. *International Journal of Man-Machine Studies*, **12** (1), 3–10.

38 Lau, K.T. (2005) Function Approximation Using Fuzzy Logic. Technical report, University of Waterloo.

39 Miraftab, V. and Mansour, R.R. (2005) Tuning of microwave filters by extracting human experience using fuzzy logic. IEEE MTT-S International Microwave Symposium Digest, June 12–17, 2005, pp. 1605–1608.

40 Miraftab, V. and Mansour, R.R. (2006) Automated microwave filter tuning by extracting human experience in terms of linguistic rules using fuzzy controllers. IEEE MTT-S International Microwave Symposium Digest, San Francisco, CA, 2006, pp. 1439–1442.

41 Miraftab, V. (2006) Computer aided tuning and design of microwave circuits using fuzzy logic. Ph.D. thesis.

42 Yu, M. and Tang, W.C. (2003) A fully automated filter tuning robots for wireless basestation diplexers. Workshop Computer Aided Filter Tuning, IEEE International Microwave Symposium, Philadelphia, June 8–13.

43 Yu, M. (2006) Robotic computer-aided tuning. *Microwave Journal*, 136–138.

44 Mongiardo, M. and Swanson, D. (2003) Workshop on computer aided filter tuning. IEEE International Microwave Symposium, Philadelphia, June 8–13.

20

High-Power Considerations in Microwave Filter Networks

(Revised and augmented by Chandra M. Kudsia, Vicente E. Boria, and Santiago Cogollos)

This chapter is devoted to an overview of high-power considerations for microwave filters and multiplexing networks in terrestrial and space applications. Emphasis is placed on the practical aspects in the design of high-power filter networks.

20.1 Background

Microwave breakdown in gases under various physical conditions is a well-known and well-studied problem. A classical summary of relevant results is given by MacDonald [1]. Another phenomenon, which is dormant under normal pressures, shows up when the gas pressure is reduced to values approaching near vacuum. This is called *multipaction*. A summary of this is presented by Kudsia *et al*. [2]. Just as significant is the phenomenon of passive intermodulation (PIM) in high-power equipment. It depends on the materials used in the equipment, the mechanical design features, and the standards of work quality. Novel microwave structures and topologies, and continuing demands for higher power, give rise to new situations and parameter regimes where previous results are no longer applicable. The problem is further compounded by the use of new materials in the implementation of microwave components. Above all, when it comes to the practical aspects of high-power breakdown, PIM, and operating environment, it seems that each generation of microwave engineers tends to repeat the mistakes of the past before coming to grips with such issues. In this chapter, we briefly review the phenomenon of breakdown and then describe in more detail the practical considerations in the design of high-power filter networks.

There are two main application areas that require high power for wireless systems at microwave frequencies: radar systems and communication systems.

Radars make use of pulsed power with low duty cycles and require low to moderate bandwidths. Another feature of radars is the requirement of frequency agility and the ability to cope with jamming. Typically, peak pulsed power requirements are in the kilowatt range, while those of continuous-wave (CW) power is in the hundreds of watts range. Various pulse compression techniques are employed to achieve a good range and Doppler resolutions. High-power equipment is optimized for pulse operation, and the bulk of high-power radar systems are for military applications. Since the main focus of this book is on nonmilitary systems, we focus on the high-power equipment required for communication systems.

20.2 High-Power Requirements in Wireless Systems

As described in Section 1.3, the capacity of a communication channel is governed by the fundamental Shannon–Hartley theorem, which relates the capacity to the bandwidth and the

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

signal-to-noise ratio (SNR) within the channel bandwidth. For a given allocation of bandwidth, the communication capacity is determined by the achievable SNR. The minimization of noise is limited by the ever-present thermal noise; therefore, the ultimate capacity of a system depends on the available power for radio transmission. Does this mean that we can indefinitely increase the capacity by simply increasing the level of RF power? There are two problems associated with that: (1) the generation of RF power is expensive and (2) the physical aperture of an antenna imposes a limit on the amount of energy that can be focused toward a distant receiver. This implies that only a limited amount of radiated energy reaches the intended destination; the rest is lost in space. What this means is that there is an inherent inefficiency in the radio transmission of energy. In addition, the spillage of energy causes interference in other competing systems. The level of such interference limits the power that can be used by a system. As discussed in Section 1.2.2, frequency allocation and the allowable levels of interference are strictly controlled by government agencies to ensure a viable multiple communication system environment. Thus, generation and transmission of RF power for communication systems is both costly and a source of interference. In addition, there are always limitations of technology in the efficient generation of RF power. For handheld mobile phones and other devices, there are also health-related concerns for the potential harmful effects of microwave radiation on human beings. The typical power levels of high-power transmitters in use for terrestrial and for space systems are described in Tables 20.1 and 20.2, respectively.

Table 20.1 High-power requirements for terrestrial systems.[a)]

System type	Typical power requirements (W)
PCS	0.1–1
Cellular handheld	0.6
Mobile unit	4
Cell sites	60
Line-of-sight (LOS)	1–10
CATV	100

a) Frequency allocations are in the 400, 800, 900, 1700, 1800, and 1900 MHz frequency bands (see Tables 1.3 and 1.4 for more details).

Table 20.2 High-power requirements for typical satellite systems.

Satellite system (GHz)	Transponder bandwidths (MHz)	Typical power requirements (per transponder)	
		Uplink (earth station) (W)	Downlink (satellite) (W)
1.5	10s of kHz uplink 0.5–5 MHz downlink	1–10	150–220
6/4	36, 54, 72	500–3000	10–100
14/11	27, 54, 72	500–3000	20–200
30/20	36, 72, 112	100–600	5–80

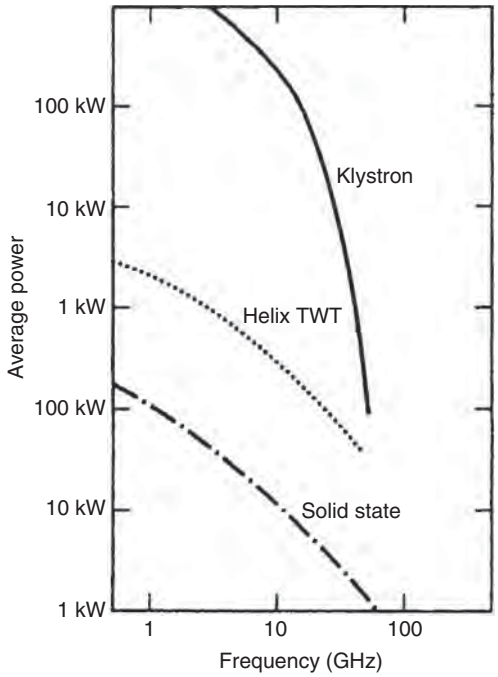


Figure 20.1 Approximate limits of microwave power sources: solid state, TWTA, and klystron. (*Source*: Gordon and Morgan 1993 [3]. Reproduced with the permission of John Wiley & Sons.)

20.3 High-Power Amplifiers (HPAs)

In a microwave system, the output amplifier can be a solid-state power amplifier (SSPA), a traveling-wave-tube amplifier (TWTA), or a klystron. They differ in the maximum RF power delivered, as plotted in Figure 20.1 [3, 4].

SSPAs are widely used for cellular and line-of-sight (LOS) wireless systems, whereas TWTAs dominate satellite systems. Klystrons are used when very high power is required. Earth stations use klystrons for high-power uplinks, and ground-based radar systems use them in a pulsed mode operation.

The efficiencies of TWTAs and klystron amplifiers are in the range of 30–50%. For space applications, TWTAs have been developed for efficiencies as high as 70%. Typically, SSPAs are limited to efficiencies of 20–35%. Even with efficiencies of 50% or more, high-power amplifiers (HPAs) dissipate a great deal of heat. This represents a major challenge for the design of high-power equipment for airborne and satellite applications.

20.4 Gas Discharge

The phenomena associated with electrical discharges in gases have attracted a great deal of attention, which has led to many discoveries. Initial studies dealt with dc voltages, or relatively low ac voltages with applications in power generation and distribution. The advent of radar and communication systems has required the use of high power at much higher frequencies. The usage of relatively high frequencies (MHz–GHz range) at high power levels leads to some new and unique effects. This section reviews the gas discharge phenomenon at microwave frequencies, and the impact on the design of high-power passive microwave components.

20.4.1 Basics

Any gas consists of atoms and molecules that are neutral particles. Also, there are a small number of electrons and charged particles present in any gaseous volume because of ionization by cosmic rays or some other phenomena, such as the photoelectric effect. When an electric field is applied to a gas, electrons and charged particles are accelerated in the direction of the field, colliding with neutral particles in their path and the sides of the container. Electron collision constitutes the dominant effect. Ions are much heavier and slow to accelerate, resulting in fewer collisions and imparting far less energy than electrons in such collisions.

Microwave breakdown in gases is explored in great detail by MacDonald [1]. The ionization of air and other gases at microwave frequencies is described in terms of collision frequency (mean free path), diffusion, attachment, and nature of the gas.

The mean free path concept is based on the classical kinetic theory of gases. It assumes that electrons, atoms, molecules, and ions are rigid elastic spheres of very small dimensions moving freely and in a random manner throughout a gas. These particles collide with each other like billiard balls. The distances between these collisions are the free paths, and the average of the lengths of these paths is the mean free path. Since we are dealing with millions and billions of such particles in an enclosed volume of gas, it is possible to compute the collision probability and frequency with a high degree of accuracy by using statistical mechanics. It is assumed that all the particles are moving at the same speed and that all the directions are equally probable. The mean free path inherently depends on the concentration of the atoms and molecules in the gas, determined by its pressure.

There are two types of collisions: elastic and inelastic. In an elastic collision, the electrons or ions bounce off an atom with an exchange of momentum, but no change in the state of the atom. Free electrons move in the direction of the field, constituting a small current. The gas envelope retains its mix of largely neutral particles. In an inelastic collision, the energy of the electron is sufficiently high to cause a change in the internal state of the atom at the expense of its kinetic energy. An excited atom normally returns to its ground state very soon after collision, and the energy that it has acquired is radiated. If the electric field is high enough, some of the electrons, on collision, knock off an electron from the atom, producing a secondary electron and a positive ion. When the electric field is strong enough for a sustained period of time, such collisions will occur often and the production rate of electrons will become greater than the loss rate due to diffusion or recombination, and breakdown will occur. The excess production rate needs to be only infinitesimally greater than the loss rate for the concentration to increase very rapidly, leading to breakdown. Figure 20.2 illustrates the variation of the electron concentration as a function of the applied electric field.

It is evident that when the electric field is gradually increased, the gas initially appears to obey Ohm's law until it starts to produce secondary electrons. When that happens, the

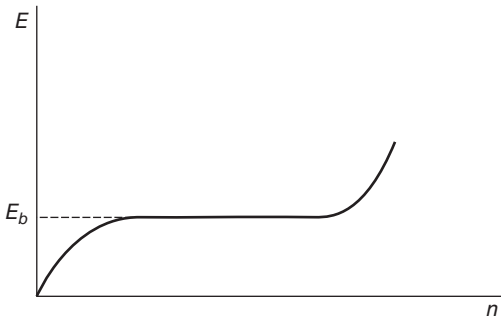


Figure 20.2 Variation of electron concentration versus electric field. (*Source*: MacDonald 1966 [1]. Reproduced with the permission of John Wiley & Sons.)

electron concentration can vary widely as shown by the flat portion of the curve, and the gas becomes conductive. Thereafter, even a minute change in the electric field can produce orders-of-magnitude change in the electron concentration and flow of current, leading to glow and breakdown.

In a gaseous medium, any gradient in particle concentration or velocity causes the flow of particles in directions that reduce such gradients. This flow of particles is called *diffusion*. In ionized gas, electrons are scattered by the atoms and molecules as they move in the direction of the applied field. The surfaces of the structure containing the gas eventually absorb the scattered electrons. This loss of electrons is attributed to the process of diffusion. The rate of diffusion depends on the electron concentration, the field gradients, the rate at which the electrons are produced, and the geometry, size, and surface conditions of the container. In addition, the rate depends on the interaction between the electrons and ions, although the dominant factor is the free diffusion of electrons.

An electron can become attached to neutral particles of gas. Once that happens, the attached electrons cease to play any significant role in the ionization process because the neutral particle is at least 2000 times heavier than an electron, and therefore its velocity is much lower than that of free electrons. It is tantamount to a loss of an electron. It should be noted that this loss of an electron is different from the diffusion process, which removes an electron from the region in which the field exists. The attachment rate varies for the various gases and is relatively high for oxygen. The nature of atoms and molecules of the gas governs the attachment process.

20.4.2 Breakdown in Air

In a given volume of gas, the electron concentration is given by [1]

$$n = n_0 \, \mathrm{e}^{((\nu_i - \nu_a) - D/\Lambda^2)t} \tag{20.1}$$

where

n = accumulated electron concentration
n_0 = initial concentration
ν_i = ionization rate
ν_a = attachment rate
D = diffusion coefficient
Λ = characteristic diffusion length

The condition where the net rate $(\nu_i - \nu_a)$ is just balanced by the rate of diffusion, D/Λ^2, represents a threshold below which breakdown cannot occur. If, on the other hand, the net ionization rate, that is, the exponent of equation (20.1), becomes greater than zero, the electrons accumulate rapidly within the volume and breakdown follows.

An analysis of the breakdown data for air [1] indicates that in a high-pressure regime, which means above one-tenth of the atmospheric pressure (i.e., 76 Torr, where 1 Torr = 1 mm Hg) or greater, the electric field for breakdown is virtually independent of frequency and the characteristic diffusion length and is represented by the value

$$\frac{E}{p} = 30\,\mathrm{V/(cm\text{-}Torr)}$$

Figure 20.3 demonstrates that at 9.4 GHz, the influence of the characteristic diffusion length Λ diminishes as the pressure is increased, whereas the breakdown strength of air converges at 30 V/(cm-Torr). The same conclusions are drawn from the data taken at 992 MHz. From this,

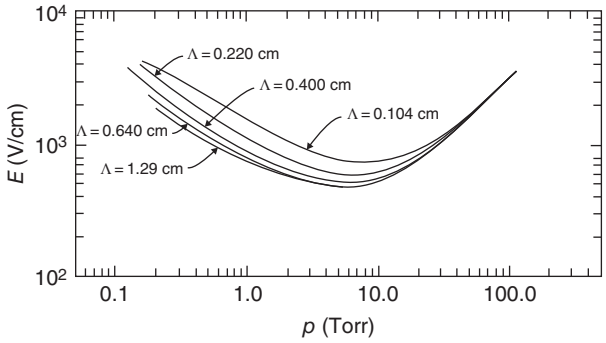


Figure 20.3 Breakdown strength of air at 9.4 GHz. (*Source*: MacDonald 1966 [1]. Reproduced with the permission of John Wiley & Sons.)

it is inferred that breakdown is an attachment-controlled process. At the normal atmospheric pressure of 760 Torr, the breakdown strength of air E_b becomes

$$E_b = 30 \times 760 = 22.8 \text{ kV/cm}$$

Figures 20.4 and 20.5 describe the experimental data on the breakdown voltages at 992 MHz and 9.4 GHz for air, oxygen, and nitrogen.

At low pressures, there are fewer collisions between electrons and gas molecules. The electrons oscillate freely, and, mostly, out of phase with the applied field. As the pressure is increased, the number of collisions also increases. This tends to reduce the phase lag between the electrons and the applied field owing to the increased number of collisions encountered by the out-of-phase electrons. The resulting motion provides a more efficient transfer of energy. The transition from a few collisions per oscillation of the field to many collisions per oscillation is pivotal and occurs when the electron collision frequency equals the frequency of the applied RF. The minimum value of the breakdown field strength occurs at a pressure where the radio frequency is equal to the gas collision frequency. For air at microwave frequencies, this pressure varies linearly from a pressure of nearly 1 mm Hg at 1 GHz to about 10 mm Hg at 10 GHz,

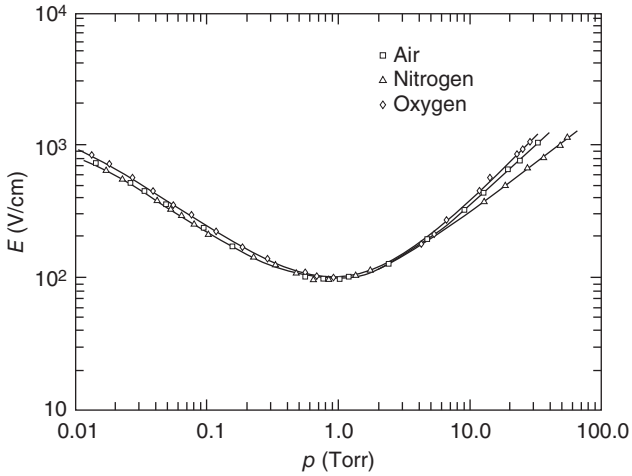


Figure 20.4 CW breakdown in air, oxygen, and nitrogen at 992 MHz. (*Source*: MacDonald 1966 [1]. Reproduced with the permission of John Wiley & Sons.)

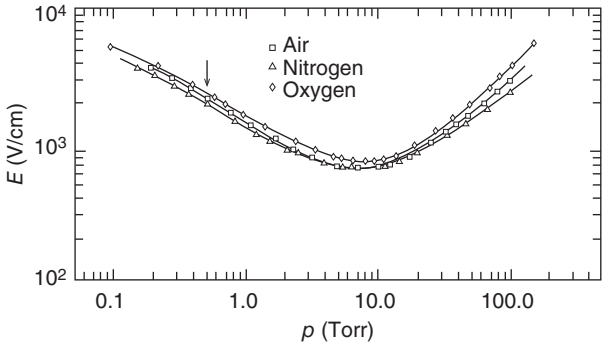


Figure 20.5 CW breakdown in air, nitrogen, and oxygen at 9.4 GHz. (*Source*: MacDonald 1966 [1]. Reproduced with the permission of John Wiley and Sons.)

as illustrated in Figures 20.4 and 20.5, respectively. At lower pressures, the breakdown field strength increases rapidly. High-power microwave equipment is often designed to withstand critical pressure even though a communication system is never intended to operate under such conditions.

In recent years, there has been a significant effort from the scientific community to develop methods capable of determining the gas breakdown power of complex microwave components [5, 6]. The main approach to accomplish this has been to solve (20.1) in the presence of arbitrary field distributions in complex component geometries. For instance, in [5], a variational method is used to solve this equation to determine the breakdown threshold. On the other hand, in [6], a full numerical approach based on finite element method (FEM) is employed. With the ongoing advances in computer capabilities, virtually any component can be rigorously analyzed in terms of breakdown.

The numerical techniques are quite accurate at pressure levels around the critical pressure. However, at intermediate or high pressures (atmospheric or higher), accuracy problems can occur because the discharge becomes extremely localized (attachment-dominated process), and the singularity of the electric field around metallic corners can play a relevant role. Owing to the uncertainty incurred in the computation of fields around such corners, the numerical analysis is rendered inaccurate. To overcome this problem, an analytical approach is proposed in [7]. However, such an approach is not general and limits the ability to simulate any type of component. The best way to solve this problem is to round the corners in high-power components, thus avoiding the field singularity. It has the added advantage of reducing the electric field amplitude, resulting in increased breakdown threshold under high power.

20.4.3 Power Rating of Waveguides and Coaxial Transmission Lines

The high-power rating of microwave equipment depends on two factors, namely, voltage breakdown and the permissible rise in temperature. Under CW operation, either one of these two conditions dominate, depending on the peak and average power, and the operating environment of the equipment. For air-filled waveguides, the voltage gradient at which breakdown occurs is approximately 22.8 kV/cm, as described in the previous section, and is based on the following assumptions [1, 8]:

1) The distance between the walls of the waveguide is much greater than the mean free path of the electrons.
2) The air inside the waveguide is still and dry.

3) The air pressure is one atmosphere.
4) Only the dominant mode is present under matched conditions [i.e., the voltage standing-wave ratio (VSWR) is unity].

Under these conditions, the theoretical power handling capability of rectangular and circular waveguides, operating in their respective dominant modes, is given by [8]

$$P_{\max}(\mathrm{TE}_{10}) = 0.34\ a\,b\,\frac{\lambda}{\lambda_g} \quad \mathrm{MW}$$
$$P_{\max}(\mathrm{TE}_{11}) = 0.26\ D^2\,\frac{\lambda}{\lambda_g} \quad \mathrm{MW} \tag{20.2}$$

where a and b are dimensions of the rectangular waveguide and D is the diameter of the circular waveguide in centimeters; λ and λ_g represent the free space and guide wavelengths. The power handling capability of a matched coaxial line is derived from [8]

$$P = \frac{E_m^2}{480}\left(\frac{b}{2.54}\right)^2 \frac{\ln b/d}{(b/d)^2} \quad \mathrm{W} = 0.17\ d^2\ \ln(b/d) \quad \mathrm{MW} \tag{20.3}$$

E_m represents the breakdown voltage of 22.8 kV/m; d and b are the inner and outer diameters of the coaxial line in centimeters.

20.4.4 Derating Factors

The power carrying the capacity of microwave equipment depends on a number of factors, none of which can be precisely assessed [9]. The basic phenomenon of gaseous breakdown is not spontaneous, but has to undergo the process of creation of sufficient electrons, which are accelerated with enough energy to produce ionizing collisions. An important stage of this process is the creation of effective electrons. The number of electrons created in a breakdown gap by cosmic rays and natural radioactivity is of the order of 10 per minute in a microwave component. Some of these electrons are captured by molecules or by the walls of the components before they can be accelerated to produce new electrons.

Whether an electron is effective depends on its initial velocity, position, and phase at the time of the electron creation. All of these factors are random. Even under CW power, it is not possible to place a precise value on either the breakdown power or the time necessary to achieve breakdown with a given power. Besides these uncertainties, we need to take into account variables such as VSWR, moisture, and other impurities inside the equipment, air pressure (altitude), presence of any spurious modes, and any discontinuities in the structure of the components. All of these factors tend to degrade the power handling capability. The cumulative effect can reduce the capacity to 5–10% of the theoretical maximum value. A practical guide of derating factors is shown in Figure 20.6.

Breakdown margins can be improved by employing pressurized systems with the concomitant penalty of increased cost and decreased reliability. The use of electronegative gases such as freon, with high attachment coefficients, can provide a much higher breakdown voltage. However, such gases are toxic and are not used in commercial systems.

20.4.5 Impact of Thermal Dissipation on Power Rating

Under CW power, the transmission lines and microwave components dissipate significant amounts of heat. As a consequence, the permissible rise of temperature is an important criterion in specifying the power handling capability. The thermal design must ensure an acceptable increase in the temperature of high-power equipment. This implies consideration

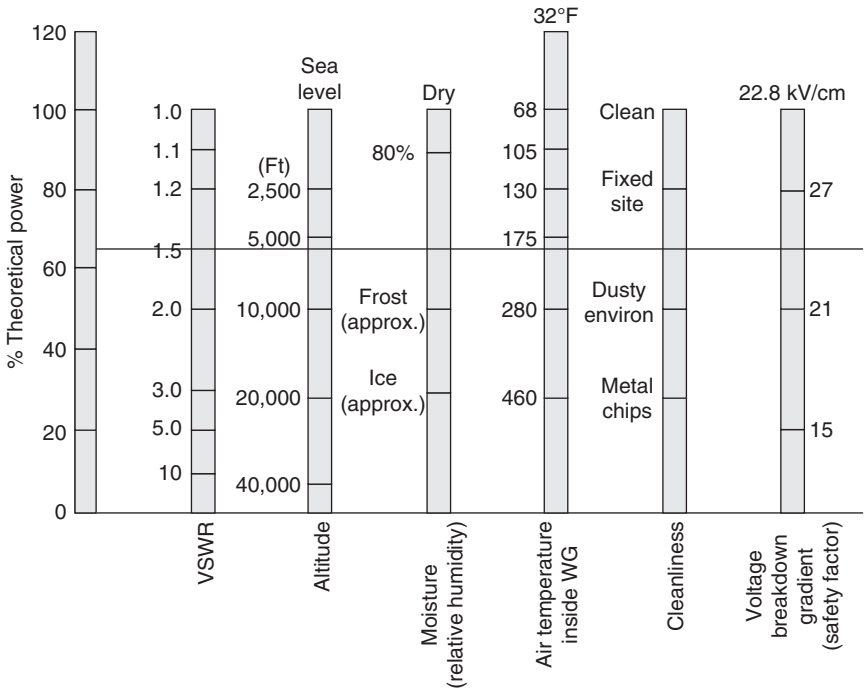


Figure 20.6 Derating factors affecting the high-power capability in waveguide structures (voltage standing-wave ratio) from Ref. [9]. (*Source*: Courtesy: J. Ciavolella.)

of techniques to cool the equipment, including radiation cooling by using enlarged cooling surfaces or by employing air or water-cooled systems in a terrestrial environment. The theoretical values of power ratings for waveguides for terrestrial application with temperature rise as a criterion are given by King [10] and are illustrated in Figure 20.7. VSWR and imperfections in the materials create a concentration of energy dissipation or hot spots, degrading the power handling capability. More in-depth coverage of the derating factors for waveguides and other microwave components are found in the literature [8, 10]. One conclusion that can be drawn from these data is that the CW power handling capability for ground-based equipment is governed by the specified temperature rise of the equipment rather than the limitation imposed by voltage breakdown.

20.5 Multipaction Breakdown

Multipaction is an RF breakdown phenomenon. It takes place under vacuum conditions when the mean free path of the electrons is larger than the gap between the walls guiding the flow of the RF power. The applied RF voltage accelerates the free electrons across the gap between the walls. Let us suppose that the transmit time is such that when the electrons hit the walls, the field is reversed and the electrons are accelerated toward the opposite wall, creating an electron resonance. If the intensity of the applied field is such that the electrons bombarding the walls cause continuous release of secondary electrons, it leads to an RF breakdown. It depends on the electron multiplication via the secondary emission of the electrons from the walls, and is often referred to as the *secondary electron resonance phenomenon*.

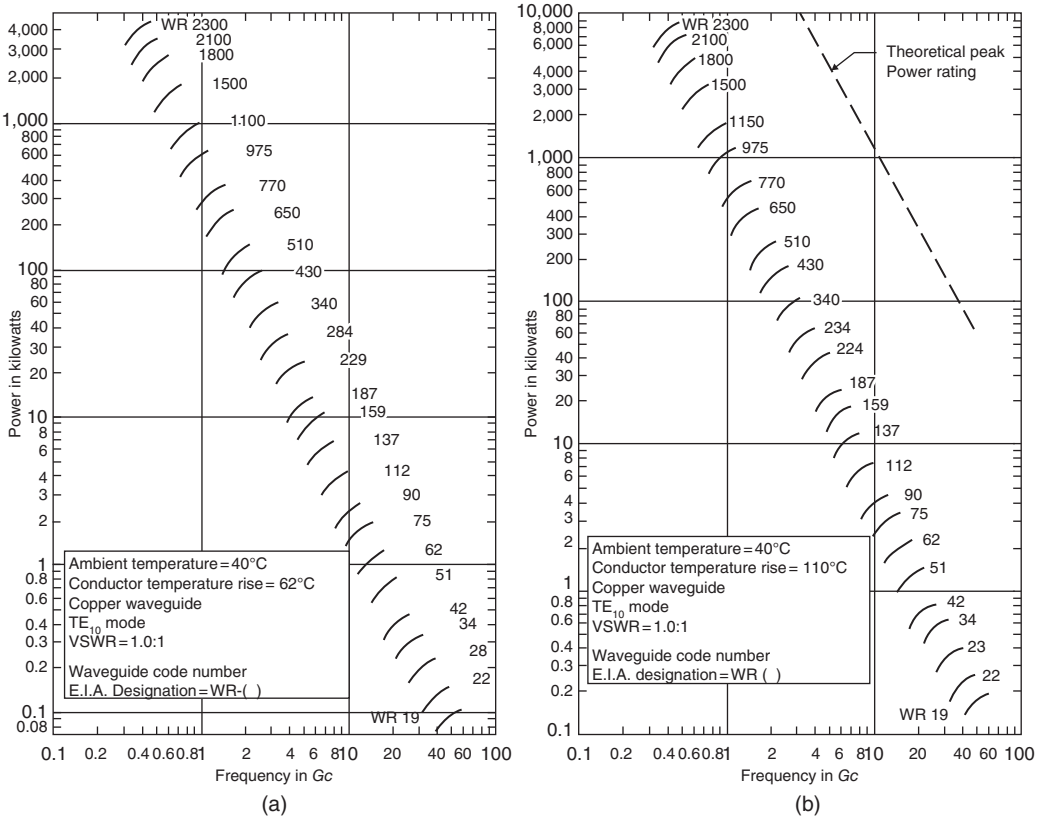


Figure 20.7 Theoretical curves of the average power rating for a copper rectangular waveguide operating in the TE_{10} mode with unity VSWR at an ambient temperature of 40°C: (a) for a temperature rise of 62°C and (b) temperature rise of 110°C. (*Source*: King 1961 [10]. Reproduced with the permission of IEEE.)

There are different types of multipaction, depending on the nature of the resonance. It is possible to have single-surface multipaction in which the electrons always impact the same surface. Single-surface multipaction appears typically in the presence of dielectric surfaces with high tangential fields [11]. On the other hand, two-surface multipaction occurs typically in between metal surfaces inside microwave components [12, 13], and it is the most common form of multipaction in communication applications. Although there do exist more complex resonances, such as multipoint multipaction, single-surface and two-surface are the most common ones.

The generation of two-surface multipaction depends on the following constraints:

1) Vacuum condition
2) Surface conditions
3) Applied RF voltage and frequency of operation in conjunction with geometry of RF components (i.e., $f \times d$ product)
4) Number of carriers and modulation

A brief description of each of these factors follows next.

20.5.1 Dependence on Vacuum Environment

For multipaction to occur, the mean free path of electrons must be long enough to permit the electrons to be accelerated between the emitting surfaces with a low probability of collision with

the ambient atoms or molecules. It has been shown [14] that for pressure of 10^{-3} Torr or less, the mean free path of electrons is in the tens of centimeters range when they are surrounded by typical gas molecules such as nitrogen, oxygen, and helium. For RF equipment, typical gap sizes are in the millimeter range. Consequently, it is safe to assume that multipaction can occur under pressures of 10^{-3} Torr or less.

20.5.2 Dependence on Surface Conditions of Materials

In order to sustain multipaction, the incident electrons on a wall must produce secondary electrons such that the secondary emission yield (SEY), defined as the ratio of the secondary to incident electrons, is greater than one. Otherwise, the generation of electrons subsides quickly and multipaction is not sustained.

For a given material, the value of the SEY is a function of the energy and angle of incidence of the primary electrons [14], as illustrated in Figure 20.8. At low incident energies, the primary electrons are unable to liberate many secondary electrons, and multipaction is not initiated. At a very high incident energy, the primary electrons penetrate so deeply into the surface that the secondary electrons produced are trapped in the substance and do not reach the surface. The energy of the primary electrons must lie between these minimum and maximum values to sustain multipaction breakdown as depicted in Figure 20.8a. These upper and lower bounds on the energy, called first and second crossover energies, respectively, permit multipaction to occur over a broad range of frequencies and applied voltages. The first crossover energy is the dominant parameter in a multipactor discharge. The value of the SEY depends also on the angle of electron incidence as shown in Figure 20.8b.

The SEY is very sensitive to the relative purity of materials. At low energy levels, there are large variations in the SEY caused by surface impurities. Most contaminants increase the value of the SEY and enhance the possibility of a multipaction occurrence. The SEY is also known to increase with time due to the aging process. This is highly dependent on the material and

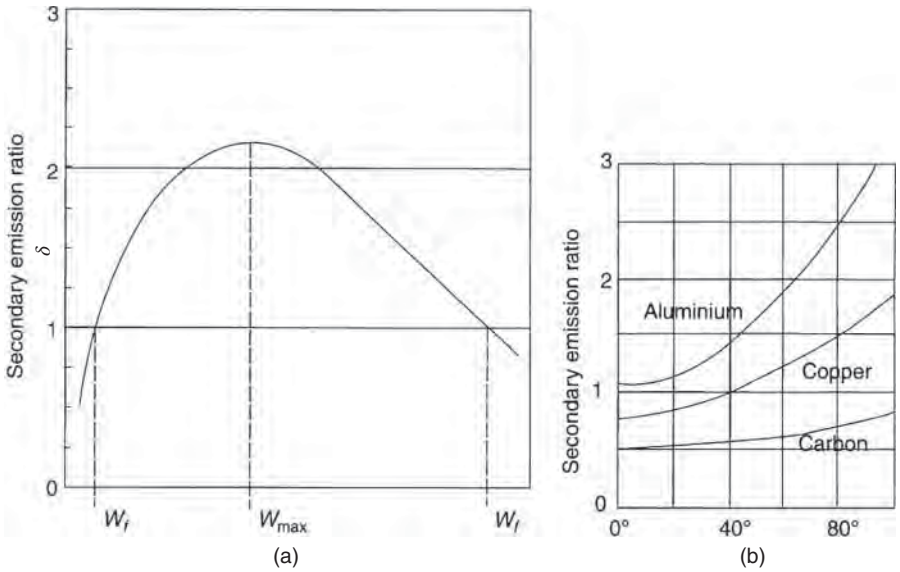


Figure 20.8 Secondary emission parameters: (a) δ versus W and (b) δ versus θ. (*Source*: Courtesy: NASA Report CR-488.)

storage conditions. Contamination-free implementation and upkeep of high-power equipment are critical for space applications.

20.5.3 Dependence on Signal Amplitude, Signal Frequency, and Geometry of the Device

The electrons are accelerated by the electric field created by the applied RF signal. Under a certain combination of signal parameters (amplitude, frequency, and phase), the electrons hit consecutively the gap surfaces in phase with the RF field and with enough energy to release secondary electrons. In order to achieve such a resonance, the gap size has to be related to the amplitude and frequency of operation. In the simplest geometry, composed of two infinite parallel plates, the multipactor breakdown voltage depends on the frequency times the gap ($f \times d$) product. A more detailed explanation of this dependence for two-surface multipaction is provided in the subsequent Section 20.5.5.

20.5.4 Dependence on Number of Carriers and Modulation

Multipaction with single-carrier signals has been widely covered in the literature, with a considerable number of theoretical and experimental results. It is the easiest case to study, since the electrondynamics can be easily modeled. This is applicable for individual components such as filters, couplers, power dividers, circulators, and transmission lines, which handle the power of a single RF channel.

The worst-case scenario occurs when the equipment must handle the combined power of a large number of RF carriers. When all the carriers are in phase, the total peak power is $N^2 P_{\text{in}}$, where N is the number of carriers and P_{in} is the power of an individual carrier. On the contrary, if the carriers are completely random in phase, the combined power is usually NP_{in}. The challenge lies in determining the duration of the peak power and whether that duration is long enough to initiate and sustain multipactor breakdown. In addition, modulation of the RF carriers introduces extra complexity in the process. The time evolution of the amplitude of the RF modulated signal depends on many factors, such as the modulation type, digital codification parameters, and nature of the information. There are very few studies on the multipaction phenomenon with modulated signals. Regardless, modulation will disturb the electron resonance, resulting in increased threshold for the breakdown power. For this reason, prediction and testing for multipactor breakdown are carried out with unmodulated signals.

20.5.5 Two-Surface Theoretical Analysis, Simulation, and Prediction

The purpose of multipactor theory and simulation is to be able to provide accurate predictions of the RF threshold power in order to design multipactor-free components and avoid the need of testing following a trial and error basis. Indeed, the inherent uncertainty in the predictions makes it necessary to impose some design margins in order to ensure that the device is multipactor-free. Two theories for the analysis of two-surface multipaction, valid for operation under a single-carrier excitation and for the case of multiple carriers, are presented next.

20.5.5.1 Single-Carrier Theory

The classical theory of multipactor [12] covers the particular case of two-surface multipactor under single-carrier operation on infinite parallel plates. Basically, the classical theory solves the electron equations of motion analytically, looking for resonance and electron growth conditions. The conditions for a multipactor discharge to occur are the following:

- The phase of electrons must have a value of an odd number of RF half periods, that is, 1/2, 3/2, 5/2, and so on, at the time of impact. Such a number is the multipactor modal order or simply the multipactor mode.
- The time or phase at which the electrons are released must be within a certain range for them to be in sync with the electric field.
- The impacting energy (velocity) must be comprised between W_1 and W_2, which are the first and second crossover energies of the material SEY, respectively.

Many researchers [14–17] have made extensive theoretical studies and measurements to obtain the combinations of RF amplitude, frequency, and gap that satisfy the above conditions. These combinations can be plotted in curves relating the $f \times d$ product and voltage in the so-called multipactor charts. A typical depiction of a multipactor chart is shown in Figure 20.9.

There is a multipaction curve for each multipactor mode. The primary mode corresponds to the multiple unity. Other modes are referred to as higher-order multipaction modes. Each multipaction curve has a minimum value and a maximum value of energy or peak voltage to initiate and sustain multipaction, as explained in Section 20.5.3. Such curves are available for a number of pure materials [17].

The classical theory assumes that the electrons are emitted with a fixed and deterministic velocity. In reality, secondary electrons are emitted with random energy and angle. This renders the classical theories very inaccurate for multipactor orders higher than one. Statistical theories incorporate such randomness to the formulation, thus yielding much more precise results [18].

The parallel-plate (PP) theory results may be extrapolated to real structures provided the gap is small compared with the rest of the device dimensions [19]. However, most practical structures are composed of discontinuities and complex geometries to meet the electrical performance requirements. In the vicinity of discontinuities and sharp edges, the electric

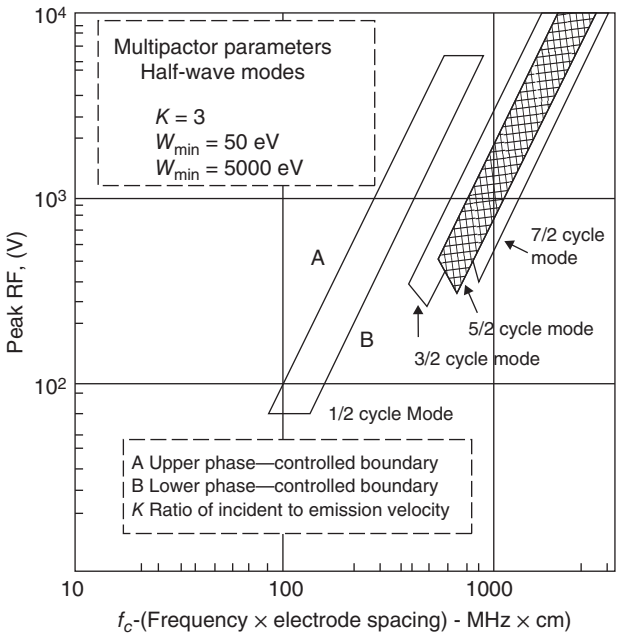


Figure 20.9 Possible regions of multipaction between parallel plates.

fields can be much different from those based on simple parallel geometry. Any deviation from homogeneous fields and parallel surfaces introduces a sort of resonance disruption. Therefore, predictions based on the PP geometry may be considered to represent the "worst-case" scenario.

The availability of fast computers has made it possible to model multipactor phenomenon numerically and analyze it using software simulation. Full-wave 3D electromagnetic solvers are coupled with particle trackers in what are called particle-in-cell (PIC) codes. The SEY properties of the surfaces are modeled numerically as well.

PIC codes have been widely used for the numerical simulation of plasmas [20]. They work with two different spaces for the resolution of the motion of particles and field equations. No restriction is placed on the movement of electrons moving in continuous space; their trajectories and interactions with the boundaries are easily calculated. However, the fields are computed in fixed locations by using mesh or grids or both, especially covering areas around corners and discontinuities. This allows performing only an initial mesh for the rest of the simulation. Both spaces are connected through some kind of particle-to-grid or grid-to-particle interpolation. PIC codes, such *Aurorasat SPARK3D* or *CST Particle Studio*, have been used for multipactor calculations, especially in 2D, but also in 3D structures [21, 22].

Numerical methods incur greater computational cost. However, they do allow for a more accurate analysis of both single-surface and two-surface multipaction in complex structures, including nonhomogeneous fields as well as different number of RF carriers and modulations.

20.5.5.2 Multi-Carrier Theory

The study of multipaction with multi-carrier signals is much more complex. A typical multi-carrier signal is composed of several modulated carriers with small frequency separation. As a result, the signal experiences periods of constructive and destructive addition of the different carriers, resulting in a time-varying envelope of the signal. Such a situation adds a dynamical behavior to the multipactor process. When the amplitude of the envelope is above the single-carrier breakdown level, the multipactor resonance conditions are met, and there is electron multiplication. When it is below that level, resonance is broken and electrons are absorbed.

For the single-carrier case, the prediction provides just a value for the breakdown voltage (RF power), given the geometry of the device, the frequency of operation and the material SEY properties. The situation in the multi-carrier counterpart is dramatically different. The multipactor discharge in multi-carrier signals depends not only on the amplitude of each of the different carriers but also on the relative phases among them. This means that for the same device, material, and carrier frequencies, a certain combination of phases implies a higher or lower breakdown level than for the others. The lowest breakdown level is referred as the worst case.

There are different ways of predicting the multipactor discharge in multi-carrier operation, and there is no general agreement on when to apply one or the other. Therefore, due to the lack of a reliable theory, industry made an effort to roughly characterize multi-carrier discharges using the single-carrier theory and making some simplistic assumptions. The result is a simple way for predicting multipaction based on peak power method and the 20-gap-crossing rule [17]. These rules, together with the applied margins, are usually safe but often tend to be excessively conservative [23].

According to the peak power method, the worst case occurs when all carriers are in phase. This method assumes that there is a discharge when the peak power equals the single-carrier power threshold. This is $N^2 P_{in} = P_{sc}$, N being the number of carriers, P_{in} the maximum power per carrier, and P_{sc} the single-carrier threshold. This method yields the most conservative prediction.

A more realistic way to predict multipaction under multi-carrier operation is to access not only the peak power but also its duration. It is the combination of these two parameters that determine the onset of multipaction. This has given rise to the 20-gap-crossing rule [17]. The rule states:

As long as the duration of the multi-carrier peak and the mode order gap are such that no more than 20 gap-crossings can occur during the multi-carrier peak, the design may be considered safe with regard to the multipaction breakdown even though the multipaction threshold may be exceeded from time to time.

The 20-gap-crossing rule establishes that in order to produce multipactor discharge, the multi-carrier envelope must be above the single-carrier threshold during a time equivalent to 20 electron gap crossings. This time is derived as follows:

$$N_c = \frac{2fT}{n} = 20$$

where N_c is the number of crossings, n is the order of the mode and, f is the lowest carrier frequency. Therefore

$$T_{20} = 10\frac{n}{f} \tag{20.4}$$

Once T_{20} is known, there are different ways to compute the worst case, which is not specified by the rule itself. The next step is to assess peak powers that can last over a duration of T_{20}. Approximate boundary curves for the prediction of the threshold power are given in [24, 25]. They are based on theoretical computations (peak power computation over a given duration by optimizing the phases of carriers for the worst-case scenario). The 20-gap-crossing rule shows two limiting cases: for small multipactor mode orders, the threshold power is close to N^2P, while for large multipactor mode orders, the threshold approaches NP.

Recently, the nonstationary theory [26], and the quasi-stationary method [23], has been proposed to fully characterize the multipactor phenomenon for a completely arbitrary set of amplitudes and phases. Such theories can model the electron growth and absorption in all intervals of the multi-carrier envelope. One can model inter-period electron accumulation or long-term discharges as well. Figure 20.10 shows an example of a six-carrier signal extracted from [27].

Full multi-carrier theories allow taking a different approach with respect to the 20-gap-crossing rule. These theories are able to compute the electron population during a period of the envelope for a certain phase combination. Therefore, it is possible to determine whether, for each phase combination and amplitude, there is multipactor discharge or not. According to these theories, there is no need for defining any number of gap crossings anymore. Given this capability, it is easy to find the phase combination that leads to the lowest breakdown level by using numerical optimization techniques.

20.5.6 Multipactor Breakdown Levels

Multipaction breakdown has been observed at very low power (of the order of a few watts) levels, and, more frequently, at higher power levels. High-Q circuits such as filters and multiplexers are especially susceptible to this breakdown. This is due to the compact structure of these devices, field enhancement at discontinuities, sharp mismatches outside the band of interest, and choice of materials. Table 20.3 is a summary of the high-power measured data for a range of microwave equipment. The measured power levels represent state-of-the-art high-power output circuits required for communications and remote sensing satellites.

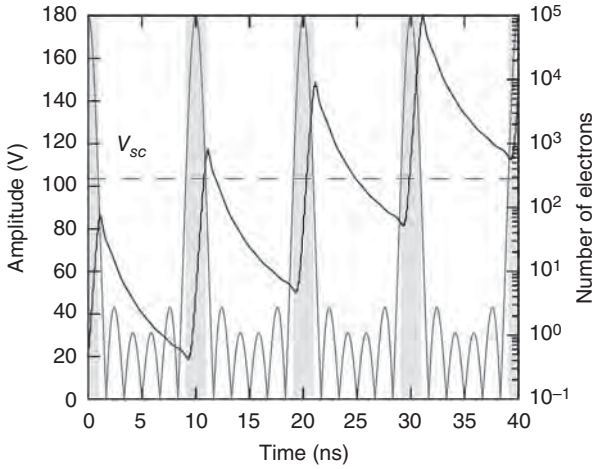


Figure 20.10 Evolution of electron population versus time for a six-carrier signal. The first carrier is at 3.57 GHz and the frequency spacing is 100 MHz. The geometry is a simple parallel-plate case with a critical gap of 0.43 mm.

Table 20.3 Examples of flight-qualified high-power equipment.

Satellites	Description	Design approach for multipactor prevention	Power handing (peak)
INTELSATs VI and VII	C-band harmonic filter	Dielectric filled Large gap separation	>1.5 kW
INTELSAT VII	Ku-band transmit reject filter	Large gap separation	4 kW
STC DBS	Output assembly (12 GHz)		
	Multiplexer	TE_{114} filters	>2 kW
	Harmonic filter	Stepped impedance	>2 kW
ERS-1	Output assembly (5 GHz)		
	Circulator	Dielectric filled	26 kW peak
	Harmonic filter	Large gaps	26 kW peak
	Mechanical switch	Large gaps	45 kW peak
Mobile Satcom	800 MHz diplexer	Dielectric coated probe Coax-coupled design	1 kW
SCS-1	Ku-band diplexer	Large gap separation	>4 kW
	Ku- and Ka-band isolators	Using ridges and posts	>4 kW
Aussat "B"	L-band diplexer	Coax-coupled design	400 W
Satcom "K"	12 GHz VPDs	Half-wave polarizers	>500 W
Olympus	18 GHz diplexer	Large gaps	180 W
	18 GHz isolator	Planar design	
NIMIQ 4	8-CH Ku-band MUX	TE113, large gaps	19.2 kW

Source: Courtesy: COM DEV, Cambridge, Ontario, Canada.

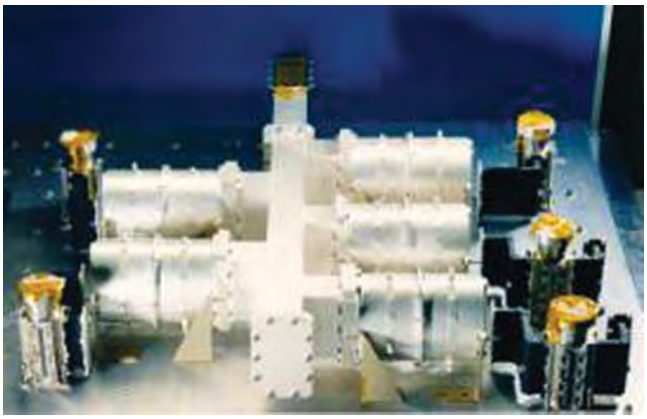


Figure 20.11 Five-channel high-power (100 W per channel) S-band multiplexer assembly. (*Source*: Courtesy: COM DEV.)

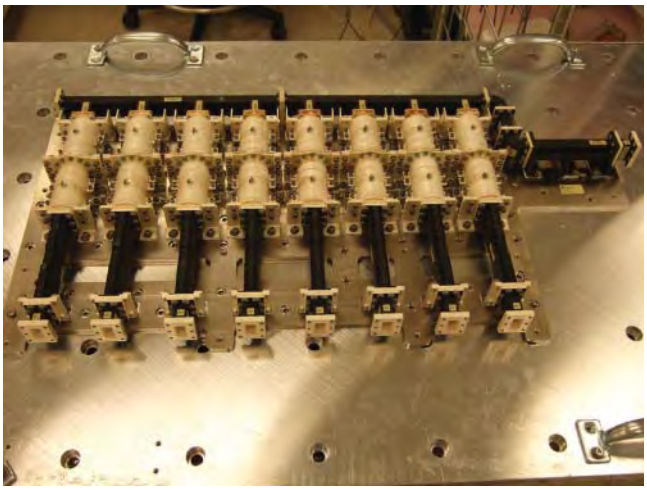


Figure 20.12 An eight-channel high-power (300 W per channel), low PIM (−140 dBm) Ku-band multiplexer. (*Source*: Courtesy: COM DEV.)

Figure 20.11 is the photograph of a high-power five-channel S-band multiplexer with a power rating of 100 W per channel. An eight-channel high-power multiplexer at Ku-band, rated at 300 W per channel is shown in Figure 20.12.

20.5.7 Detection and Prevention of Multipaction

Despite the many advances in multipactor theories, software tools, and simulations (see Section 20.5.5), there is still the need to perform final testing of this effect in order to reduce potential risks and satisfy the very stringent specifications from space industry. State-of-the-art multipactor test beds try to reproduce—as close as possible—the actual operating conditions of the device under test (DUT) for onboard satellite applications. This requires experimental facilities to be located inside clean room environment of at least 100,000 class, including high vacuum chambers, dedicated high-power RF test beds, and expensive instrumentation for RF signal generation, amplification, and measurements [28].

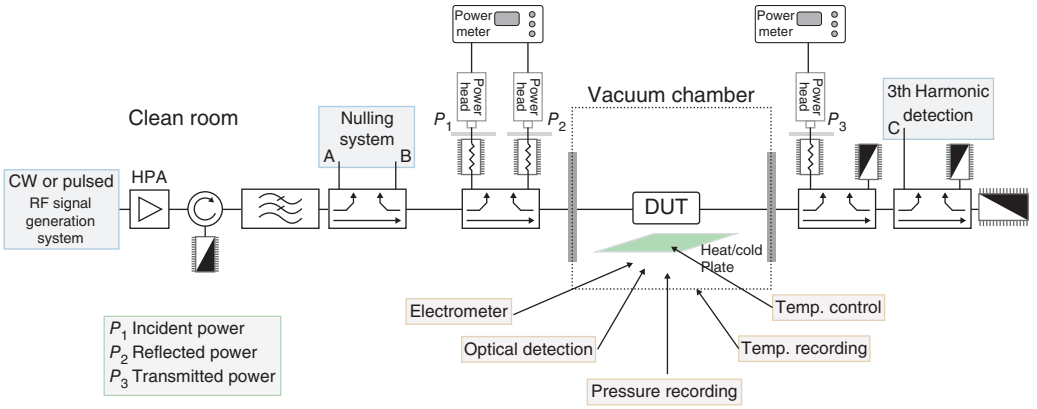


Figure 20.13 Test setup of for measuring RF breakdown effects (i.e., multipaction and gas discharge).

Classical setup for measuring RF breakdown effects (i.e., multipaction and gas discharge) is composed of three main parts: RF signal generation, electronic seeding, and breakdown detection methods as described in Figure 20.13. Most RF breakdown tests run under single-carrier mode, that is, a non-modulated signal tone (either pulsed or as CW) is generated, then amplified, filtered, and guided toward the DUT. Nevertheless, novel experimental setups considering either multi-carrier and/or digitally modulated signals are becoming available. Electron seeding (another critical aspect of these facilities) is usually being performed with Sr90 sources or with ultraviolet light (photoelectric effect). These electron sources are often located near the DUT region where multipaction is expected.

Breakdown detection methods can be classified into two categories: local and global. Local detection methods (such as electrometer, optical system, temperature, and pressure levels) are suitable in the vicinity close to the point of the actual discharge inside the DUT (see Figure 20.13). However, for system testing, global detection methods indicate that a discharge is present somewhere in the high-power assembly. Global detection methods such as close-to-carrier noise, phase and third harmonic noise, forward/reverse power, and microwave nulling (see Figure 20.13) are described in [17, 29].

Apart from being able to measure accurately undesired RF discharges under high power, it is more important to mitigate (or even prevent, whenever it is possible) their appearance. Earlier in this section, we learned that a number of constraints are essential to initiate and sustain multipaction discharge. The violation of any of those constraints should, therefore, provide the means to suppress multipaction. Possible ways to prevent the occurrence of this phenomenon are enumerated as follows:

- Control of the frequency–gap product
- Dielectric filling
- Pressurization
- Magnetic or dc bias
- Reduction of the surface potential by treatment of materials

The characteristics and relative advantages of these methods are summarized in Table 20.4.

20.5.8 Design Margins in Multipaction

Design margins are a sensitive issue for customers, prime contractors, and suppliers of high-power components and subsystems. Because of the inherent variability of multipaction

Table 20.4 Multipaction prevention methods.

Method	Description	Advantages	Disadvantages
Large gap size	The electron transit time is made nonresonant such that it requires much higher voltage to initiate and sustain multipaction	Simplest mechanical structure	Limitation on RF design and achievable performance
Gap size below cutoff	The electron transit time is made nonresonant. Electrons are scattered and absorbed, thus preventing the onset of multipaction	No power handling limitations exist due to multipaction breakdown	RF design unsuitable for most practical applications. Achievable values of Q are much lower
Dielectric filling	The mean free path of electrons is reduced by filling the empty spaces of the discharge region with foam or solid dielectric	Simple in theory	Lowers the effective Q of the devices. Difficult to apply above 12 GHz
Pressurization	The mean free path of electrons is rendered close to zero by pressurization using inert gas	Removal of the multipactor mechanism	A possible failure mode if leakage exists. Increase in system weight. System must use hermetically sealed connectors. A source of potential PIM
dc or magnetic bias	A dc bias voltage or applied magnetic field can alter the electron resonance condition to allow higher-power handling	Power handling margin improved by ~1 dB or more. Arching/discharge prevented across the gap	Surface erosion is still possible creating wideband noise and other performance degradation. Performance improvement not significant
Reduced surface potential	The secondary emission ratio of surfaces is spoiled by use of coating of low emission ratio	Simple to apply	Limited applicability due to higher loss

breakdown, its dependence on work quality standards, sensitivity to contaminants, complexity of component geometries, and change of personnel every few years, all contribute to the tendency by the customer to overspecify multipaction margins and test requirements. The prime contractor often has very little choice but to accept such specifications for competitive reasons. The supplier of high-power RF equipment is invariably faced with the difficult task of having to meet such specifications within tight schedules and cost constraints. The optimum choice for all concerned is to adopt certain guidelines for specifying design margins on the basis of theoretical calculations and experimental verifications of a repeatable nature.

Currently, there are two main standards dealing with multipaction: the European one, written by the ECSS [17], and the American one, written by Aerospace Corporation [30]. Besides margin definitions, the standards describe procedures on how to compute the multipactor breakdown predictions during the design phase and test procedures for adequate multipactor testing.

Although the given margins are not identical, both standards are similar since margins are defined in both cases depending on the following:

- *Type of component*: Components are classified based on the materials and topology.
- *Stage of design/test*: Margins are given differently if they are applied to predictions done during the design phase or if applied to a test campaign.
- *Type of signal*: Margins are given differently for single-carrier and multi-carrier operation.

Individual high-power components are designed using the multipaction preventing methods summarized in Table 20.4. For single-carrier signals, both standards fix a 3 dB test margin for the power handling capability. It is relatively easy to incorporate a 3 dB margin at the expense of increased volume and mass and other design features, all adding to extra complexity and cost of the RF hardware. Moreover, testing at these elevated power levels adds to the cost, especially at higher frequencies. However, the most complex and limiting operating condition is for RF equipment operating with multiple carriers.

With respect to multi-carrier signals, assuming a 3 dB margin over the peak power threshold implies that the specification for such equipment can be as high as $2N^2P_{\text{in}}$. If we assume six carriers, a power handling specification of 72 times the value of an individual carrier is implied. For ground-based equipment, we rarely require more than a few carriers at a satellite Earth station or a cellular site. Moreover, the mass and volume of the equipment is not so critical. However, for space applications with a relatively larger number of carriers, such a specification can be very costly to design and often beyond the range of high-power equipment to test it. In some cases, a specification of $2N^2P_{\text{in}}$ can just be met by placing equipment in hermetically sealed enclosures containing some inert gas. The problem becomes more complicated because mass, volume, and reliability are critical issues as well. Equipment has to operate for 10–15 years, trouble-free, in the hard vacuum of space. Hermetically sealed units are deemed too risky for such requirements because of the potential leakage of the inert gas over time. Besides, such an option is too costly. Consequently, one has to look critically at the realistic scenarios and design margins for multipaction breakdown in a multi-carrier environment.

20.6 High-Power Bandpass Filters

High-power equipment is designed to have low loss to conserve power and, at the same time, minimize thermal dissipation. The requirement of low-loss necessitates use of high-Q microwave structures for such equipment. For bandpass filters, the problem is further exacerbated owing to the requirement of relatively narrow bandwidths in most communication systems. Narrow bandwidths imply higher losses and higher electric fields inside the filter cavities. As a consequence, the power handling capability of high-Q, narrowband filters represents the most severe problem in the design of high-power equipment. Microwave filters are therefore limited in their power handling capacity by the high fields generated inside a filter, as well as by the heating of the cavities due to thermal dissipation. Cohn [31] was the first to discuss general design considerations for high-power filters. He provided a simple analytic expression for the peak electric fields inside the various cavities of a "narrowband" direct-coupled waveguide filter in terms of its lowpass prototype parameters. These formulas described the variation of electric field as a function of frequency in each cavity. The paper also describes how the variation of peak internal fields between the cavities can be minimized by enlarging certain cavities at the expense of an increased number of spurious resonances.

Young [32] introduced an alternative approach by analyzing the power ratios in a filter cavity in terms of the group delay of the filter. This analysis is based on the assumption that a

direct-coupled cavity filter resembles a "periodic structure." With that assumption, it can be readily shown that the stored energy in a filter cavity is proportional to its group delay. This analysis extended the computation of electric fields to direct-coupled "wideband" filters. This approach has been resurrected and extended in depth by Ernst [33]. Rigorous expressions for the time-averaged stored energy (TASE) in passive lossless two-port networks have been derived and shown to be proportional to the group delay for Butterworth, Chebyshev, and elliptic function filters. This allows use of group delay as an indicator of the power handling capability of a filter. This approach highlights the sensitivity of high power to a range of filter parameters, and specifically, to the selectivity of filters. Peak internal fields are very sensitive to the bandwidth and selectivity requirements of filters. The smaller the absolute bandwidth, the higher the group delay and electric fields inside it. Similarly, the higher the selectivity by virtue of having transmission zeros close to the passband, or a higher order of filter, the sharper the rise in group delays and peak internal fields. Stored energy in a resonator can be computed using circuit models derived from the lumped lowpass prototype filter. The stored energy in practical bandpass filters can be related to the original prototype filter by frequency transformation. For direct-coupled ladder filters, TASE of individual cavities can be readily computed in terms of the lowpass prototype parameters [33]. The next step is the computation of electrical fields in terms of the stored energy. For standard rectangular or circular waveguides, this can be done analytically. This approach simplifies the analysis of power handling capability of filters realized in standard topologies, such as ladder networks. For more complex structures involving discontinuities in the filter cavities, a more rigorous analysis is required.

A general procedure to estimate the voltage in filters based on the lumped element prototype without any restrictions has been introduced by [34–36]. It highlights the fact that it is possible to realize a variety of practical topologies (in theory, an infinite number if all cross-coupling are available) having the same filter response. The total stored energy remains the same in each topology; it simply gets redistributed among the various resonators. It leads to the possibility of a tradeoff between the topology of the filter and the peak electrical fields inside the resonators [33, 37]. In practice, the topology tradeoff must involve design simplicity and the sensitivity of the structure over the operating environment, and that puts a limiting constraint on the design options. This analysis allows computation of peak voltages in a variety of filter structures, including diplexers and multiplexers.

20.6.1 Bandpass Filters Limited by Thermal Dissipation

The power of ground-based transmitters is typically in the range of 100s of watts to kilowatts of RF power. Such high power generates lots of heat due to losses in the equipment and, especially, in the narrowband channel filters. Thermal dissipation is minimized by using high-Q structures. The Q of a cavity depends on its physical dimensions as well as the propagation mode. Use of larger cavities and low-loss propagation modes is typical in the design of high-power narrowband bandpass filters. However, lower loss is achieved at the expense of a larger number of spurious modes. Suppression of spurious modes is therefore an integral part of the design. Another design feature is to locate the hot spots in the filter and provide an efficient path for the conduction of heat to the thermal radiators, within the constraints of maintaining good thermal stability of the overall filter. This can be achieved by constructing the filter cavities out of thermally stable silver plated invar. The hot spots are centered at the coupling irises. Making irises of thick copper provides an efficient means to remove the heat away from the filter with a minimal impact on the overall thermal stability of the filter. The copper irises can be readily integrated with cooling fins. Such design features are typical in the design of high-power filters and multiplexers for satellite earth stations [38, 39].

For cellular base stations operating in the 800 MHz range, dielectric loaded cavities are preferred in order to reduce the size without too much of a sacrifice in the unloaded Q of the filter. This poses a challenge for the removal of heat under high power. Use of TM110 mode dielectric resonators and a monoblock construction have been employed to develop thermally stable high-power dielectric filters [40]. The monoblock resonator is made up of a high dielectric constant low-loss ceramic. It is encased in a surrounding ceramic that is silver plated on the outside providing a smooth contact with the metal housing. Such a construction provides a continuous thermal diffusion path from the resonators to the housing. At times it is necessary to use fans to conduct the heat away or even employ water-cooled systems. For space applications, the only means available to get rid of the heat is via conduction or radiation. However, the power requirement is typically much lower than that required for the ground-based equipment.

20.6.2 Bandpass Filters Limited by Voltage Breakdown

For space based applications, filters and multiplexers are subjected to the phenomenon of multipaction, described in the previous Section 20.5. It severely limits the breakdown voltages to much lower values if the equipment is to operate under conditions of hard vacuum. In addition, heat due to thermal dissipation must be carried away via conduction and radiation. As a consequence, evaluation of peak electric fields inside the filters, especially for the narrow bandwidth channel filters, is a critical requirement. This analysis is equally applicable for high-power ground equipment such as diplexers used in cellular base stations (see Section 18.5.1).

An exact analysis of the voltage and power distribution inside a filter involves the 3D analysis of the filter structure using numerical methods such as the FEM or the finite difference time domain (FDTD) method. However, such analysis is time-consuming and impractical if the filter in question is very narrowband, owing to the extreme sensitivity of the filter performance on dimensions. A good alternative is to compute voltages inside the filter based on the lumped element prototype network (coupling matrix representation) of a bandpass filter. This is an efficient procedure and often sufficient to evaluate the peak voltages. If necessary, the cavities with the highest energy can be analyzed more rigorously by using commercially available FEM software. It allows an accurate assessment of the hot spots owing to the specific geometry and coupling structures. An alternative method is proposed in [36]; the numerical analysis of single-cavity resonators is employed with lumped element prototype networks in order to compute the voltages inside a filter.

20.6.3 Filter Prototype Network

Bandpass filters of narrow to moderate bandwidth can be represented by a prototype network consisting of N resonators coupled through admittance inverters [8]. This technique to analyze and synthesize filter networks is described extensively in Chapter 8. Here, we provide an adaptation of the technique for the evaluation of voltage distribution inside filters and multiplexers [34, 41, 42].

Figure 20.14 shows the generic lowpass filter representation wherein there are N unit capacitances cross-coupled through frequency invariant admittance inverters of admittances M_{ij}. A bandpass filter with center f_0 and frequency bandwidth BW can be obtained by using the lowpass to bandpass transformation, namely, $s \leftarrow \frac{f_0}{\mathrm{BW}}\left(\frac{s}{\omega_0}+\frac{\omega_0}{s}\right)$ where $s=j\omega$ is the complex frequency variable and $\omega_0 = 2\pi f_0$. Thus, the bandpass filter network would consist of N lumped LC resonators with $N+2$ nodes wherein the nodes designated 0 and $N+1$ are the input and output, respectively. The frequency invariant shunt susceptance M_{ij} across each resonator represents the tuning state of the resonator. Losses can be introduced into the network by shunting

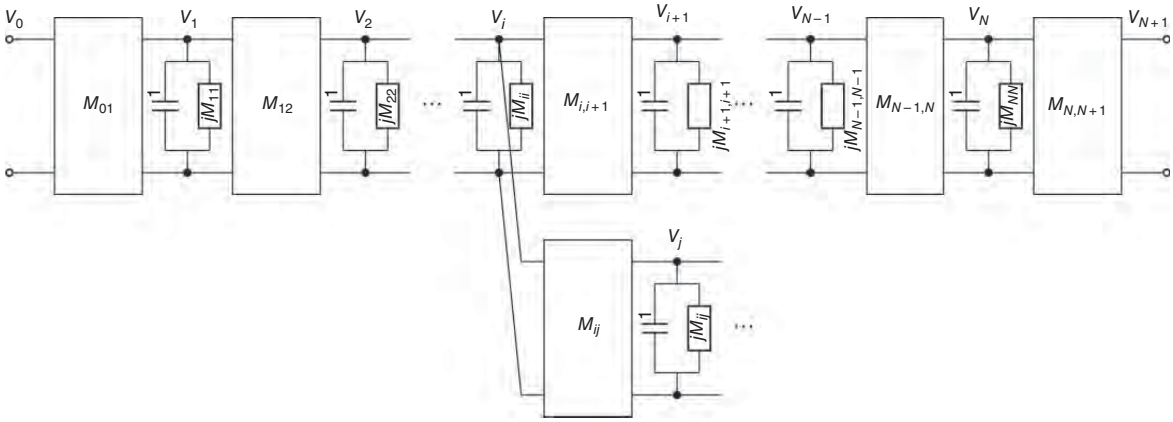


Figure 20.14 General admittance inverter cross-coupled lowpass filter prototype.

each resonator with a conductance $(f_0/\mathrm{BW})/Q_i$, where Q_i is the unloaded quality factor of the resonator. This network problem can be solved by writing down the nodal system of equations in the form

$$\mathbf{Y} \cdot \mathbf{v} = \mathbf{i} \tag{20.5}$$

where

$$\begin{aligned} Y_{\mathrm{pp}} &= s + jM_{\mathrm{pp}}, \quad p = 1, \ldots, N \\ Y_{\mathrm{pp}} &= 0, \quad p = 0, \quad N+1 \\ Y_{\mathrm{pq}} &= jM_{\mathrm{pq}}, \quad p \neq q \\ \mathbf{v} &= [V_0 \; V_1, \cdots, \; V_N \; V_{N+1}]^T \\ \mathbf{i} &= [I_0 \; 0 \; \cdots \; 0 \; \; I_{N+1}]^T \end{aligned}$$

where I_0 and I_{N+1} are the filter port excitation currents. The two-port S parameters of the filter can be computed from the two-port impedance matrix $\boldsymbol{Z}_f$ of the filter. This can be done by first evaluating the impedance matrix of $\mathbf{Z} = \mathbf{Y}^{-1}$ of the system, and then selecting the corner elements of the $\mathbf{Z}$ matrix to form the two-port impedance matrix $\boldsymbol{Z}_f$ of the filter:

$$\mathbf{Z}_f = \begin{bmatrix} Z_{0,0} & Z_{0,N+1} \\ Z_{N+1,0} & Z_{N+1,N+1} \end{bmatrix} \tag{20.6}$$

20.6.4 Lumped to Distributed Scaling

In practice, the resonators are realized as a waveguide or coaxial transmission-line cavity, and therefore an equivalence between a lumped and a distributed resonator has to be established for voltage computation. Here, we assume that the resonator is realized as an $n\lambda_{g0}/2$ long shorted cavity, where n is the number of half-cycle variations along the axis of the cavity and λ_{g0} is the guide wavelength at f_0. Although both the prototype LC resonator and the transmission-line cavity are resonant at f_0, their susceptance slopes are, respectively, (f_0/BW) and $n\pi(\lambda_{g0}/\lambda_0)^2/2$. Therefore, if the susceptance slope of the lumped resonator can be made equal to that of the cavity resonator, the lumped element prototype analysis becomes equivalent to that of a distributed transmission-line prototype. As a consequence, the unit capacitances in the lowpass prototype have to be scaled by a factor $T = n\frac{\pi}{2}\left(\frac{\mathrm{BW}}{f_0}\right)\left(\frac{\lambda_{g0}}{\lambda_0}\right)^2$. This scaling of the unit capacitance values in the prototype is carried out in such a way that the two-terminal filter response of the system in (20.5) is unchanged. This can be done, as indicated in [34], by multiplying all the rows and columns in the $\mathbf{Y}$ matrix except the first and last by $\sqrt{T}$. This scales the internal (resonator) voltages V_1 to V_N by $1/\sqrt{T}$. Therefore, the resonator voltages in a cavity filter can be obtained by first solving the lumped prototype problem from (20.5) and then by scaling the lumped resonator voltages V_1 to V_N by $1/\sqrt{T}$.

20.6.5 Resonator Voltages from Prototype Network

The general case of a filter excitation can be treated by considering a terminating impedance Z_L at the output port when the input is driven by a unit current. On treating the network as an $N+2$ port with an impedance matrix of $\mathbf{Z} = \mathbf{Y}^{-1}$, the voltage at the kth resonator is given by

$$V_k = I_0\left(Z_{k,0} - \frac{Z_{k,N+1}\; Z_{N+1,0}}{(Z_L + Z_{N+1,N+1})}\right), \quad k = 0, \ldots, N \tag{20.7}$$

Similarly V_0 may be determined as

$$V_0 = I_0 \left(Z_{0,0} - \frac{Z_{0,N+1}\ Z_{N+1,0}}{(Z_L + Z_{N+1,N+1})} \right) \tag{20.8}$$

thus allowing the ratio V_k/V_0 to be evaluated in terms of the elements of the impedance matrix (20.6). The generator voltage V_{gen} and the input voltage V_0 are related as follows [see equation (8.26a)]:

$$V_0 = \frac{V_{\text{gen}}}{2}(1 + S_{11}) \tag{20.9}$$

If the generator impedance $Z_S = Z_L = 1\ \Omega$, then at a frequency of perfect transmission through the lossless purely reactive network the source and load will be matched ($S_{11} = 0$), and the maximum power available from the generator is transferred to the load. From (20.9), in the special case where $S_{11} = 0$, $V_0 \to V^+_{0\,\max}$ giving $V^+_{0\,\max} = V_{\text{gen}}/2$. If V_{gen} is set to 2 V, then the incident voltage $V^+_{0\,\max} = 1\,\text{V}$, and the power transferred to the load is the maximum possible 1 W. The internal voltage V_k at any frequency, normalized to the voltage corresponding to an incident power of 1 W becomes

$$\frac{V_k}{V^+_{0\,\max}} = \frac{V_k}{V_0}(1 + S_{11}) \tag{20.10}$$

which becomes $\frac{V_k}{V_0}(1 + S_{11})\frac{1}{\sqrt{T}}$ when scaled to the transmission-line environment [14].

20.6.6 Example and Verification via FEM Simulation

In order to verify the proposed method, the voltage distribution inside a five-pole H-plane iris filter (see Figure 20.15) with Chebyshev response is computed using *Ansoft HFSS* FEM package. This response is compared with the computed performance using the prototype model in Figure 20.16. The voltage ratios of the five resonators are computed as a function of the frequency using the prototype filter and plotted in Figure 20.17. The voltage ratio corresponding to the E_z component of the electric field (E_z integrated between the broad walls) is computed along the center line of the filter at the band center (12,026 MHz) and at the 3 dB band edge (11,737 MHz) in Figure 20.18. The peak values in the voltage distribution are compared against the resonator voltages obtained from the prototype [equation (20.10)]. It can be seen that there is close correlation between the prototype values and the peak voltage values inside a filter.

It is interesting to note that (1) the middle resonator has the highest voltage at band center, (2) the second resonator has the highest voltage at band edge, and, (3) in the out-of-band region, the first resonator gets an appreciable amount of voltage due to the reflected power.

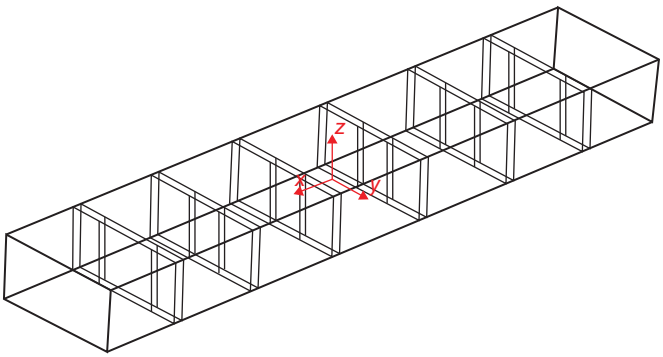


Figure 20.15 Five-pole H-plane iris filter (WR75).

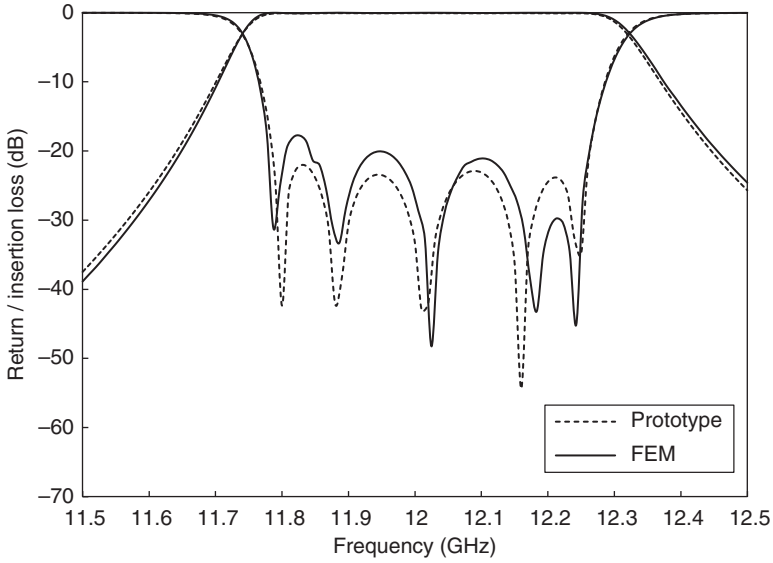


Figure 20.16 Five-pole H-plane iris filter response using FEM and prototype, $f_0 = 12026$ MHz, BW = 470 MHz. Couplings: $M_{01} = M_{56} = 1.078$, $M_{12} = M_{45} = 0.928$, $M_{23} = M_{34} = 0.662$.

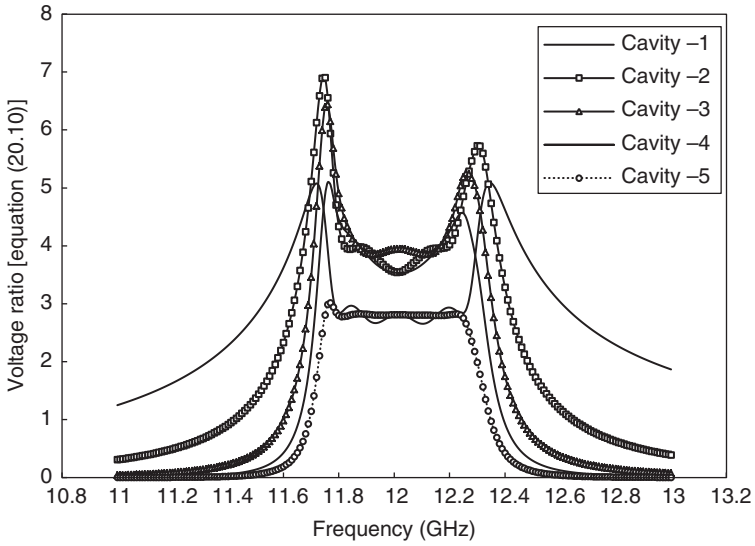


Figure 20.17 Resonator voltage ratios [equation (20.10)] as a function of frequency.

20.6.7 Example of High Voltages in a Multiplexer

Multiplexers for satellite applications typically comprise of channels realized as four-pole circular waveguide dual-mode filters with two transmission zeros. A typical example is a filter with a cavity diameter of 1.07 in. at 12,100 MHz with 36 MHz bandwidth operating in the TE113 mode. The voltage ratios in the four resonators at the three salient frequency points are given in Table 20.5. It can be seen that the peak voltage appears on the second mode at the band edge.

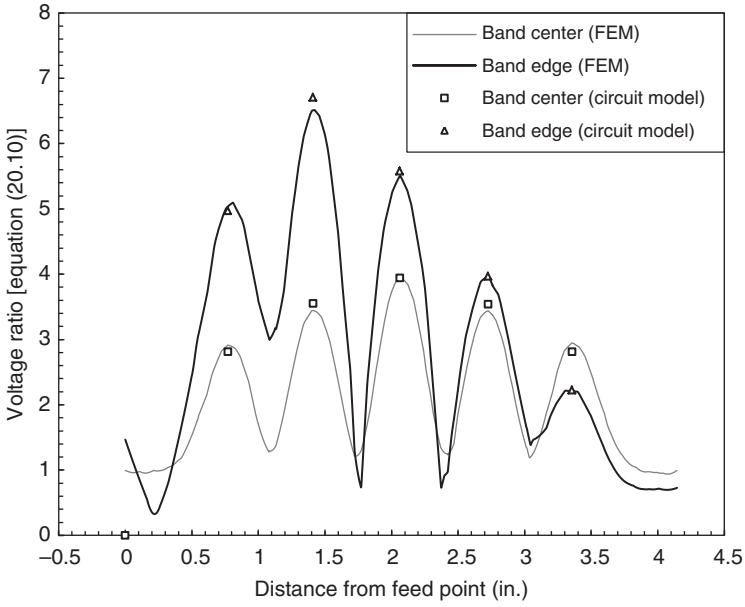


Figure 20.18 Voltage ratio distribution along the center line of the filter using FEM and comparison of the peak values to the values obtained using the prototype.

Table 20.5 Voltage ratios for the four resonances inside a typical four-pole dual-mode filter [34].

	Resonator voltage ratio			
Frequency (MHz)	**Resonator 1**	**Resonator 2**	**Resonator 3**	**Resonator 4**
12100, midband	6.16	7.18	6.48	6.39
12082, band edge	7.29	11.45	10.57	6.10
12020, out of band	3.29	0.67	0.09	0.150

The couplings are $M_{12} = M_{34} = 0.8745$, $M_{23} = 0.8024$, $M_{14} = -0.2363$, $M_{01} = M_{45} = 1.0429$, $f_0 = 12{,}100$ MHz, and BW = 36 MHz.
Source: Sivadas *et al*. 2000 [34]. Reproduced with the permission of IEEE.

Since power is proportional to the square of the voltage, the peak power in the second resonator can often reach up to 120 times the input power at the band edge.

20.6.8 Summary of Multipactor Breakdown Evaluation Methodologies

Multipactor breakdown evaluation is typically divided in two steps. First, it is necessary to evaluate the voltage/field in the filter structure. Then, once the fields are known, the multipactor analysis is performed. There are different techniques for solving both types of problems with different advantages and disadvantages. Some of them have been already presented in the previous sections. Here, a comparison between them with a brief description is outlined. Table 20.6 presents different methods for voltage or field evaluation in filter structures. As a general rule, any technique has a tradeoff between model complexity and accuracy. Simple theoretical approximations are much faster, but usually limited to simple geometries. Full 3D

Table 20.6 Summary of methods for voltage/field evaluation in filter structures.

Method	Description	Advantages	Disadvantages
LP Prototype model [34]	Ideal circuit model	Simple model Easy computation	Does not represent physical model Cannot identify hot spots
Group delay model [33]	Time averaged stored energy (TASE) in resonators, circuit models, homogeneous field resonators	Simple model Easy computation	Does not represent a physical model Cannot identify hot spots
EM techniques	3D model, arbitrary geometries, real EM fields	Represents physical models with arbitrary geometry Can identify hot spots Accurate predictions	Complex model Computationally intensive

numerical solvers involve heavy computations, but cover a wider range of cases with higher accuracy.

With respect to the multipactor analysis techniques, the same philosophy applies. Analytical methods are restricted to very simple geometries (PP case or similar), but provide fast computations. Full 3D numerical solvers are more versatile but imply a higher computational cost. Table 20.7 shows a summary of the different multipactor analysis techniques.

20.7 Passive Intermodulation (PIM) Consideration for High-Power Equipment

It is well known that active devices produce intermodulation (IM) products because of their inherent nonlinearity, as described in Section l.5.6. However, passive devices such as filters and antennas also generate lower-level IM products, referred to as PIM products [43]. In space, PIM was first observed in the late 1960s in the Lincoln Laboratories satellites LES-5 and LES-6.

Since then, PIM was discovered in other satellites, high-power Earth station equipment, and terrestrial microwave systems. Continual advances in spacecraft technology, requiring a higher number of transponders and power levels, have provided a strong motivation to focus on the problem of PIM and, in general, on the nonlinear behavior of passive components. In satellites, all the high-power RF channels are in close proximity with each other, as well as with the receive equipment, as described in Section 1.8.4. Moreover, the difference in power levels between the transmit portion and the receive section is of the order of 120 dB, implying that the PIM levels need to be 140–150 dB lower than the power of the transmit channels. This represents the most severe environment for PIM specifications. As a consequence, PIM is a factor in determining the frequency plans of satellite systems. As a minimum, the frequency plan must ensure that the dominant third-order PIM does not fall in the receive band. This helps the hardware providers in being able to meet the stringent PIM specification for the high-power equipment. Similar problems exist for high-power ground-based transmitters. However, for satellite earth stations, the problem is minimized by physically separating the transmit and the receive portions of the Earth station. For cellular stations, PIM is not a significant issue owing to the relatively smaller difference in power levels between the transmit and receive portions of the station.

Table 20.7 Summary of methods for multipactor breakdown analysis.

Method	Description	Advantages	Disadvantages
Classical multipactor theory [12–16][a)]	Parallel-plate (PP) geometry Based on simple two-surface electron resonance condition Assumes electrons are emitted with a fixed and deterministic velocity	Simple model, analytic computations	Does not represent a real physical model of filters Typically yields conservative values, by as much as 3–5 dB Inaccurate for multipaction order greater than one
Classical multipactor theory (fitted with experimental data) [12–16][b)]	Parallel-plate (PP) geometry Based on simple two-surface electron resonance condition Assumes electrons are emitted with a fixed and deterministic velocity Theoretical predictions have been fitted with experimental data for a limited set of materials	Simple model, analytic computations	Does not represent a real physical model of filters Typically yields conservative values, by as much as 3–5 dB Limited to certain materials
Nonstationary multipactor theory [18, 26]	Parallel-plate (PP) geometry Based on simple two-surface electron resonance condition Random distribution of energy and angles for electrons	Simple model, analytic computations Accurate for all multipactor orders Different number of carriers	Does not represent a real physical model of filters Typically yields conservative values, by as much as 3–5 dB
Numerical techniques [21, 22][c)]	3D Models with arbitrary geometries Numerical simulation of plasma Random distribution of energy and angles for electrons	Accurate analysis of complex structures Nonhomogeneous fields Different number of carriers and modulations	Complex analysis Computationally intensive

a) Used in Aerospace Corp Standard [30].
b) Used in ECSS Standard [17].
c) Numerical software tools such as Aurorasat SPARK3D®/FEST3D® or CST Particle Studio®.

The PIM sources are divided into two categories: contact nonlinearity and material nonlinearity. PIM results primarily from the formation of very thin oxide layers on metal surfaces, mechanical imperfections in the joints, or both. For metal surfaces separated by 10–40 Å thick oxide layers, nonlinear electron tunneling through the barrier occurs, giving rise to low level PIM. Nonlinearities also occur as a result of microcracks or voids in metal structures (bulk effects), or dirt or metal particles on the surfaces. The buildup of microvoltages at these locations produces microcurrents that are manifested as low level PIM. All interfaces and joints are susceptible to generate low level PIMs. The observed values of PIM represent the summation of all of these effects. Magnetic and dielectric materials have a degree of inherent nonlinearity and therefore are capable of producing significant levels of PIM. It should be noted that PIM reduction strategies can only mitigate, and not cure, the effects. Although experience has taught

us that some methods are better than others, PIM reduction must be carried out experimentally. Smooth and clean surfaces, free of dirt and oxide layers, are a good baseline.

20.7.1 PIM Measurement

PIM specifications for high-power communication systems, such as direct broadcast satellites or military satellite systems, can be as low as −140 dBm. The measuring equipment must therefore be at least 5–10 dB better than the specified level of PIM. This is a tall order for standard measurement instrumentation. Invariably, it requires custom measurement setups, suited for specific measurement accuracy and system design. Figure 20.19 describes a typical setup for PIM measurement (frequency band reflects a 6/4 GHz satellite system).

It consists of two high-power input signals at frequencies F_1 and F_2, the DUT and the low-power measuring equipment. Signals are generated using frequency synthesizers, amplified to the desired power levels and followed by narrowband filters to retain the purity of the signals. Power level can be adjusted via attenuators in the path prior to amplification of the signals. Directional couplers in the path of the signals allow measurement of the incident (FWD) and reflected (RVS) powers at the DUT interface.

DUT consists of a 3 dB hybrid and the device or component to be tested for PIM. The two input ports of the hybrid are connected to the signals F_1 and F_2. Each of the two output ports receives half the power of each carrier. One of the output ports is connected to the device to be tested and the other port to a load. Half of the power of the carriers is dissipated in the load, whereas the PIM power, generated by the component under test, is connected to a diplexer that separates the carrier signals from the PIM products. The remaining carrier power is dissipated

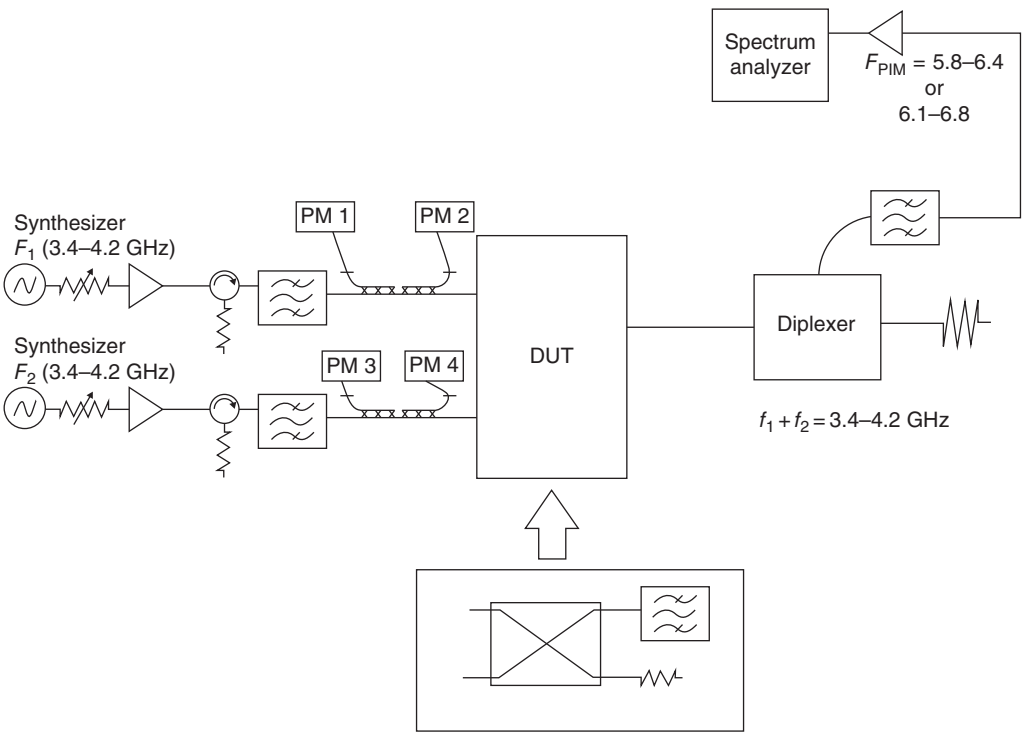


Figure 20.19 PIM measurement setup.

Table 20.8 PIM measurement systems.

Frequency (MHz)	Total power (W)	Third order PIM (dBm)	dBc
300	60	−120 to −130	−167 to −177
8,000	2,000	−150	−213
12,000	40	−140	−186

State of the Art.

in the load connected to the diplexer, whereas the PIM product is filtered, amplified by an LNA, and then fed to the spectrum analyzer. The measurement sequence involves calibration of the noise floor and any PIM generated by the measuring equipment. A high-power signal at the input of the diplexer (consisting of F_1 and F_2) is used to evaluate and minimize any PIM contribution from the load or the diplexer. A low-level signal (PIM frequency) at the input of the diplexer is used to calibrate the noise floor of the system.

From DUT onward, each component and interface in the measurement setup is capable of generating low-level PIM that must be maintained at power levels of 5–10 dB lower than the specification on the device. For a specification of −140 dBm, this implies PIM levels of −145 to −150 dBm. Identification of the sources of PIM and controlling their levels over time and the operating temperature range (i.e., repeatability) at such low power levels puts a severe constraint on the measuring equipment, especially at the various interfaces. The problem is at its worst when coaxial interfaces are involved. This is one area where experience plays a critical role in establishing precision setups for the measurement of low level PIMs. An alternative measurement setup based on reflected PIM is described in [44]. In this novel integrated solution, only one multiplexer [45, 46] takes care of combining all high-power input carriers and separating the desired PIM signal reflected by the DUT. Proceeding in this way, the number of interconnections (potential sources of undesired PIM) is minimized in the experimental setup.

In Table 20.8, the state of the art for low PIM measurement systems is described for several frequency ranges.

20.7.2 PIM Control Guidelines

The nature of PIM is such that it is not amenable to theoretical design. It depends on the quality of materials almost at the molecular level and the way they are joined together, surface conditions, work standards, and subsequent handling of the equipment. The most cost-effective approach to protection against or minimization of PIM is to enforce a set of guidelines to be followed at all stages of the program. Such guidelines are summarized as follows [17, 29]:

- Avoid the use of ferromagnetic material in high-power equipment.
- Use metals that tend to exhibit low levels of PIM, such as, gold, silver, copper, and brass. Aluminum and steel tend to generate higher levels of PIM.
- Reduce the regions of high current density by incorporating designs with larger dimensions of RF components.
- Pay special attention to the choice of materials, connectors, and interfaces. All contacts and joints must be composed of the same material and be rigidly mounted. Cleanliness and avoidance of surface contamination are critical to minimizing PIM.
- Minimize or eliminate the use of tuning screws.
- Take special care of electroplated surfaces and avoid nickel plating.
- Avoid sharp edges that can concentrate current flow.

Table 20.9 Third-order PIM in passive components with two input carriers at +45 dBm.

Component	Conventional units (dBm)	State-of-the-art units (dBm)
Diplexer	−70	−120
Multiplexer	−60	−120
Hybrid	−60	−140
Circulators	−70	−110
Coaxial connectors	−60 to −100	−140
Orthomode transducers	−1000	−140

State of the art.

Table 20.9 lists the state of the art in achievable performance for a range of components that require a low PIM performance.

Summary

High-power microwave equipment for terrestrial applications uses air or other gaseous materials inside the equipment. It is cost effective and requires a good understanding of the phenomenon of gaseous breakdown at microwave frequencies.

In space, enclosed structures are either vented to ensure a hard vacuum or hermetically sealed with an inert gas. The employment of hermetically sealed equipment is rare because of the very stringent requirement on the leakage of gas in the hard vacuum of space. Equipment operating in a vacuum environment is subjected to the phenomenon of multipaction breakdown under pressures of 10^{-3} Torr or less. In a geostationary orbit, the vacuum is of the order of 10^{-17} Torr. As a consequence, multipaction represents a major constraint for the design of high-power equipment for space applications.

This chapter provides an overview of the phenomena of microwave gaseous breakdown. It addresses the issue of voltage breakdown as a function of pressure, including the critical pressure. It highlights the importance of derating factors that can severely degrade the performance of high-power equipment for terrestrial applications.

The phenomenon of multipaction is described in some depth. It includes classical theories for the voltage breakdown based on parallel plate geometry and modeling of filter structures using lumped lowpass prototype parameters for the computation of electric fields. For most applications, this simple analysis provides conservative values for the voltage and power breakdown in space equipment. Computation of more accurate and realistic values of breakdown in microwave structures with arbitrary geometry is accomplished using a combination of (i) 3D modeling of physical hardware using EM techniques for the evaluation of electric fields and (ii) 3D modeling of structures using numeric techniques for the multipaction breakdown analysis. Such techniques, though computationally intensive, can analyze structures with arbitrary geometries, with different number of carriers, modulations, and power levels. A brief discussion of these techniques, including comparisons of breakdown voltages and powers using various methods for some practical filters, is included in the chapter.

Another phenomenon that becomes significant in designing high-power equipment is that of passive intermodulation (PIM), a significant issue in high-power satellite systems. PIM is

difficult to analyze and depends on the choice of materials and work quality standards. The chapter concludes with guidelines to measure and minimize PIM in the design of high-power equipment.

Acknowledgment

The authors wish to express a sincere and deep gratitude to their colleagues Dr. Carlos Vicente, Dr. Sergio Anza, and Dr. Jordi Gil and the staff of Aurora Software and Testing S.L. (AURORASAT) for their contributions and support in reviewing and augmenting this chapter.

References

1 MacDonald, A.D. (1966) *Microwave Breakdown in Gases*, Wiley, New York.
2 Kudsia, C., Cameron, R., and Tang, W.C. (1992) Innovations in microwave filters and multiplexing networks for communications satellite systems. *IEEE Transactions on Microwave Theory and Techniques*, **40**, 1133–1149.
3 Gordon, G.D. and Morgan, W.L. (1993) *Principles of Communications Satellites*, Wiley.
4 Strauss, R., Bretting, J., and Metivier, R. (1977) Traveling wave tubes for communication satellites. *Proceedings of the IEEE*, **65**, 387–400.
5 Anderson, D. *et al* (1999) Microwave breakdown in resonators and filters. *IEEE Transactions on Microwave Theory and Techniques*, **47** (12), 2547–2556.
6 Pinheiro-Ortega, T. *et al* (2010) Microwave corona breakdown prediction in arbitrarily-shaped waveguide based filters. *IEEE Microwave and Wireless Components Letters*, **20** (4), 214–216.
7 Jordan, U. *et al* (2007) Microwave corona breakdown around metal corners and wedges. *IEEE Transactions on Plasma Science*, **35** (3), 542–550.
8 Matthaei, G.L., Young, L., and Jones, E.M.T. (1964) *Microwave Filters, Impedance Matching Networks and Coupling Structures*, McGraw-Hill, New York.
9 Ciavolella, J. (1972) Take the hassle out of high power design. *Microwaves*, 60–62.
10 King, H.E. (1961) Rectangular waveguide theoretical CW average power rating. *IRE Transactions on Microwave Theory and Techniques*, **9** (4), 349–357.
11 Kishek, R.A., Lau, Y.Y., Ang, L.K. *et al.* (1998) Multipactor discharge on metals and dielectrics: historical review and recent theories. *Physics of Plasmas*, **5** (5), 2120–2126.
12 Gill, E.W.B. and von Engel, A. (1948) Starting potentials of high-frequency gas discharges at low pressure. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, **192** (1030), 446–463.
13 Vaughan, J.R.M. (1988) Multipactor. *IEEE Transactions on Electron Devices*, **35** (7), 1172–1180.
14 The study of multipactor breakdown in space electronic system, NASA CR-488, Goddard Space Flight Center, 1966.
15 Hatch, A.J. and Williams, H.B. (1954) The secondary electron resonance mechanism of low pressure-high frequency breakdown. *Journal of Applied Physics*, **25** (4), 417–423.
16 Vance, E.F. and Nanevicz, J.E. (1967) *Multipactor Discharge Experiments*, ARCRL-68-0063, SRI Project 5359, Stanford Research Institute, Palo Alto, CA.
17 Multipaction design and test, European Cooperation for Space Standardisation ESA-ESTEC Document, ECSS-E-20-0IA, May 2003.

18 Anza, S., Vicente, C., Gil, J. *et al.* (2010) Nonstationary statistical theory for multipactor. *Physics of Plasmas*, **17** (6), 062110.

19 Semenov, V.E., Rakova, E.I., Anderson, D. *et al.* (2007) Multipactor in rectangular waveguides. *Physics of Plasmas*, **14** (3), 033501.

20 Birdsall, C.K. and Langdon, A.B. (2005) *Plasma Physics Via Computer Simulation*, Taylor & Francis Group, London, UK.

21 Anza, S., Vicente, C., Raboso, D., Gil, J., Gimeno, B. and Boria, V.E. (2008) Enhanced prediction of multipaction breakdown in passive waveguide components including space charge effects. IEEE International Microwave Symposium, June 2008, Atlanta, USA, pp. 1095–1098.

22 Burt, G., Carter, R.G., Dexter, A.C., Hall, B., Smith, J.D.A. and Goudket, P. (2009) Benchmarking simulation of multipactor in rectangular waveguides using CST particle studio. Proceedings of SRF2009, Berlin, Germany, pp. 1–5.

23 Anza, S., Vicente, C., Gil, J. *et al.* (2012) Prediction of multipactor breakdown for multicarrier applications: the quasi-stationary method. *IEEE Transactions on Microwave Theory and Techniques*, **60** (7), 2093–2105.

24 Angevain, J.-C., Drioli, L.S., Delgado, P.S. and Mangenot, C. (2009) A boundary function for multicarrier multipaction analysis. 3rd European Conference on Antennas and Propagation, 2009. EuCAP2009, March 2009, pp. 2158–2161.

25 Wolk, D., Schmitt, D. and Schlipf, T. (2000) A novel approach for calculating the multipaction threshold in multicarrier operation. Proceedings of the 3rd International Workshop on Multipactor, RF and DC Corona and Passive Intermodulation in Space RF Hardware. ESTEC, September 4–6, Noordwijk, The Netherlands, pp. 85–91.

26 Anza, S., Mattes, M., Vicente, C. *et al.* (2011) Multipactor theory for multicarrier signals. *Physics of Plasmas*, **18** (3), 032105.

27 Anza, S., Vicente, C., Gil, J., Mattes, M., Wolk, D., Wochner, U., Boria, V.E., Gimeno, B. and Raboso, D. (2014) Multipactor prediction with multi-carrier signals: Experimental results and discussions on the 20-Gap-crossing rule. 8th European Conference on Antennas and Propagation (EuCAP), April 2014, pp. 1638–1642.

28 European High Power Radio-Frequency Space Laboratory, European Space Agency (ESA) and Val Space Consortium, see http://www.val-space.com/, Valencia, Spain, 2010.

29 Tang, W.C. and Kudsia, C.M. (1990) Multipactor breakdown and passive lntermodulation in microwave equipment for satellite application. IEEE Military Communication Conference Proceedings, Monterey, CA, October 1990.

30 Graves, T.P. (2014) Standard/Handbook for Radio Frequency (RF) Breakdown Prevention in Spacecraft Components, Report no. TOR-2014-02198, May 2014.

31 Cohn, S.B. (1959) Design considerations for high-power microwave filters. *IRE Transactions on Microwave Theory and Techniques*, **7** (1), 149–153.

32 Young, L. (1960) Peak internal fields in direct-coupled filters. *IRE Transactions on Microwave Theory and Techniques*, **8** (6), 612–616.

33 Ernst, C. (2000) Energy storage in microwave cavity filter networks. Ph.D. thesis. University of Leeds, School of Electronic and Electrical Engineering.

34 Sivadas, A., Yu, M., and Cameron, R. (2000) A simplified analysis for high power microwave bandpass filter structures. IEEE MTT-S Microwave Symposium Digest, June 2000.

35 Wang, C. and Zaki, K.A. (2001) Analysis of power handling capability of bandpass filters. IEEE MTT-S Microwave Symposium Digest, June 2001.

36 Yu, M. (2007) Power-handling capability for RF filter. *IEEE Microwave Magazine*, **8** (5), 88–97.

37 Senior, B.S., Hunter, I.C., Postoyalko, V. and Parry, R. (2001) Optimum network topologies for high power microwave filters. High Frequency Postgraduate Student Colloquium, 2001. 6th IEEE, Cardiff, pp. 53–58.

38 Atia, A.E. (1979), A 14-GHz high-power filter. Microwave Symposium Digest, 1979 IEEE MTT-S International, Orlando, FL, USA, pp. 261–264.

39 Kudsia, C.M., *et al*. (1979) A new type of low loss 14 GHz high power combining networks for satellite earth terminals. 9th European Microwave Conference, Brighton, England.

40 Nishikawa, T., Wakino, K., Hiratsuka, T. and Ishsikawa, Y. (1988) 800 MHz band high-power bandpass filter using TM_{110} mode dielectric resonators for cellular base stations. Microwave Symposium Digest, vol. 1. IEEE MTT-S International, New York, NY, USA, pp. 519–522.

41 Atia, A.E. and Williams, A.E. (1972) Narrow band waveguide filters. *IEEE Transactions on Microwave Theory and Techniques*, **MTT-20**, 258–265.

42 Cameron, R.J. (1982) General prototype network-synthesis methods for microwave filters. *ESA Journal*, **6**, 193–206.

43 Chapman, R.C., Rootsey, J.V., Poldi, I. and Davison, W. W. (1976) Hidden threat: multicarrier passive component IM generation. AIM/CASI 6th Communications Satellite Systems Conference, Paper 76-296, Montreal, April 5–8.

44 Soto, P., *et al*. (2015) CAD of multiplexers for PIM measurement set-ups. Proceedings of 6th International Workshop on Microwave Filters, Session 4: High-Power and Compact Filters and Multiplexers II, March 23–25, 2015, CNES-ESA, Toulouse, France, 5 pp.

45 Cogollos, S. *et al* (2015) Efficient design of waveguide manifold multiplexers based on low-order EM distributed models. *IEEE Transactions on Microwave Theory and Techniques*, **63** (8), 2540–2549.

46 Carceller, C. *et al* (2015) Design of compact wideband manifold-coupled multiplexers. *IEEE Transactions on Microwave Theory and Techniques*, **63** (10), 3398–3407.

21

Multiband Filters

21.1 Introduction

The majority of communication systems today are designed to support several frequency bands requiring the use of multiple filters for separating these bands. The number of filters can be reduced by employing multiband filters where one physical filter (with one input and one output) can serve two or three bands as shown in Figure 21.1. In wireless systems, this helps to consolidate more hardware into fewer units, which in turn leads to size and cost reduction of base stations. It also helps wireless operators to adapt to network demands by deploying other bands if needed. In satellite communication systems, a multiband filter can be used to transmit several noncontiguous channels to the same geographic region through one beam as depicted in Figure 21.2 [1]. In this case, only a single high-power amplifier needs to be used, which dramatically simplifies system architecture and reduces dc power consumption. Moreover, with the ever increasing demands to eliminate interference from other communication networks, the need for some form of in-band suppression has been growing, giving rise to the concept of multi-passband filtering characteristics that can be realized with the use of multiband filters, which have the potential to be employed in the design of diplexers and multiplexers, leading to a considerable size reduction.

There are four possible approaches for realizing multiband filters:

Approach I: The use of a wideband filter with a coupling matrix that allows insertion of in-band transmission zeros (TZs)
Approach II: The use of multimode resonators where each passband is associated with one mode
Approach III: The use of parallel connected filters
Approach IV: The use of a multi-notch reject filter in cascade with a wideband bandpass filter.

Each approach offers advantages and disadvantages. Approach I offers the flexibility of controlling the spacing and bandwidth of each band and makes it possible to realize a low insertion loss performance for each passband since the starting point is a wide bandwidth filter. This approach, however, does not offer any significant size reduction since it employs the same number of resonators needed if single-band filters are used. In addition, it requires the use of nontraditional filter synthesis techniques to generate the coupling matrix that allows the insertion of TZs inside the filter bandwidth. Section 21.7 gives details of synthesis procedures to design such type of multiband filters.

Approach II offers the smallest design size since the number of physical resonators is reduced significantly. Moreover, traditional single-band filter synthesis techniques can be used, since each passband, while combined with the other passbands in one physical structure, is treated

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

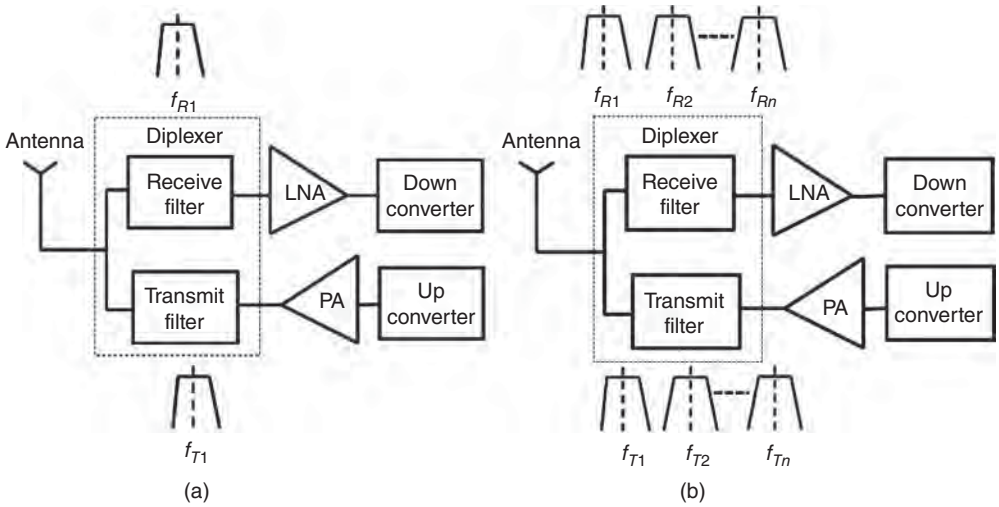


Figure 21.1 A diplexer employing (a) single-band filters and (b) dual-band filters.

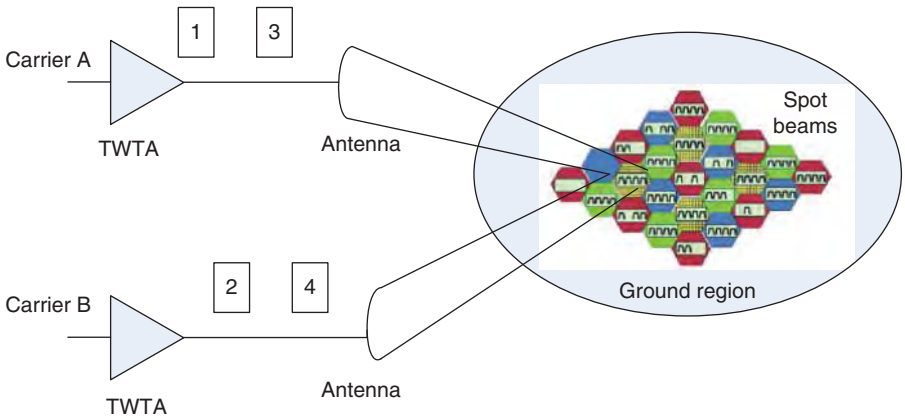


Figure 21.2 Multiple bandwidths transmitted by one beam through the use of multiband filters.

electrically as a separate filter with its own coupling matrix. This approach however requires to have a separate control of (1) the resonance frequencies of the multimode resonator, (2) the interresonator couplings of each passband, and (3) the amounts of input/output coupling to each passband. Usually, it is not easy to realize a high isolation between the passbands with this approach, particularly in the cases where the passbands are closely spaced to each other. In comparison with approaches I, III, and IV that can be implemented using conventional resonator configurations, approach II usually requires the use of nontraditional resonator configurations in order to be able to space the resonance frequencies of the multimode resonator to fit the desired center frequencies of the passbands. Thus, it offers filter researchers the opportunity to innovate and devise resonators structures and coupling configurations that allow full control of the performance of the multiband filter.

Approach III is the traditional approach that is widely used in realizing multiband filters where separate single-band filters are connected in parallel. Through the use of a manifold/junction at the input, the overall band is divided into separate passbands, which are then combined into one output through a similar manifold/junction. The advantage of this

technique is that it makes it easy to realize multiband filters with a very large number of passbands.

Approach IV is usually used in interference mitigation applications, where attenuation at specific frequency bands is realized by cascading a multi-notch filter with a bandpass filter. Such filters can be also viewed as multiband filters. While realizing a sharp skirt over a narrow reject band is a challenge, this approach is still much easier to implement than the other three approaches. Additionally, it offers some flexibility in tuning and controlling the bandwidth of each band through the use of a tunable reject filter.

Over the past decade, there have been several papers in literature on multiband filters. In particular, microstrip-based dual-band filters have the largest share of these publications. Because of low-Q limitation of microstrip resonators, microstrip multiband filters can only be employed in limited applications. The focus in this chapter is on high-Q multiband filters, addressing the state-of-the-art results reported to date on the realization of high-Q coaxial, waveguide, and dual-band filters of a dielectric resonator (DR) [2–16].

21.2 Approach I: Multiband Filters Realized by Having Transmission Zeros Inside the Passband of a Bandpass Filter

In principle, this approach can be applied to any resonator configuration, whether a combline, waveguide, or DR operating either as a single mode or dual mode. The design starts with finding the coupling matrix that allows having TZs inside the filter band and outside the band if needed. Once the coupling matrix is determined, the resonators and interresonator couplings are arranged in the same way as any bandpass filter. Thus, synthesizing the proper coupling matrix is a key design step in implementing this approach. Details of the synthesis procedures of this type of coupling matrices are given in Section 21.7.

Two configurations for a dual-band filter employing combline resonators were reported in [2], having equal and different passband widths. The filter's physical structure, the coupling matrix topology, and the measured results of the two filters are shown in Figures 21.3 and 21.4. It can be seen from the filter structures that the design can be easily implemented once the coupling matrix is identified. A dual-band filter formed by using dual-mode cylindrical cavities was proposed in [3]. The modes used in these cavities are the typical TE_{113} degenerate dual modes. Figure 21.5 shows the coupling matrix, the structure, and the simulated results of an 11-pole dual-band filter with four TZs.

A six-pole compact wideband dual-band filter realized in air-filled metallic ridge waveguides is reported in [4]. Figure 21.6 shows the filter structure and its measured results. The resonators are realized using sections of propagating ridge waveguides, whereas the interresonator coupling is realized using either evanescent rectangular or evanescent ridge waveguide sections. The dual-band filter was designed [4] with a fractional bandwidth close to 60% at 2.8 GHz with the two passbands having equal widths of 200 MHz and a guard band close to 1.25 GHz. This ridge waveguide configuration demonstrates the feasibility of realizing dual-band filters with passbands that are separated by a rather wide guard band.

A 16-pole dual-band filter with 10 TZs designed to meet the typical stringent requirements of wireless systems was reported in [5]. A schematic diagram of the coupling matrix of the dual-band filter is given in Figure 21.7. A very compact realization of folded ridge waveguide was implemented in an E-plane-like configuration using ridge waveguide resonators. The filter consists of two halves bolted together with a metal plate having the interresonator coupling irises. Positive and negative couplings were realized by controlling the size and location of the irises with respect to the two ridge resonators as shown in Figure 21.8. Figure 21.9 shows the

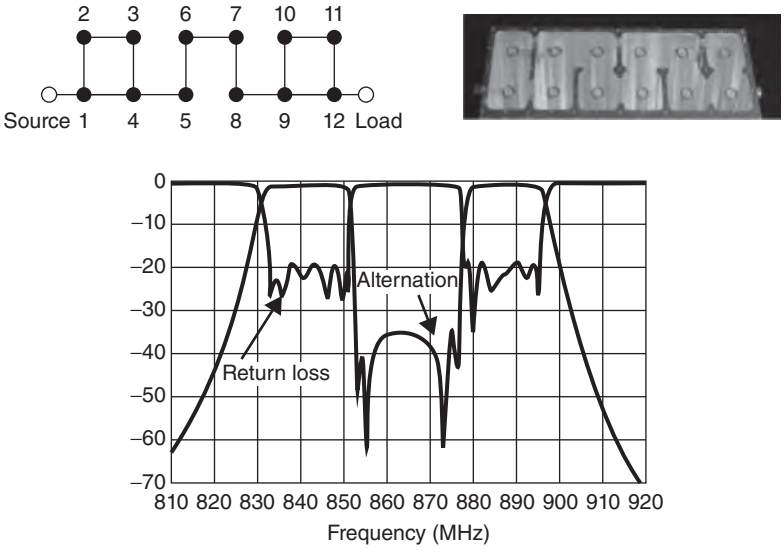


Figure 21.3 Filter structure, coupling matrix, and associated simulation results of a 12-pole dual-band filter having two passbands with equal widths. *Source*: Macchiarella and Tamiazzo 2005 [2]. Reproduced with the permission of IEEE.

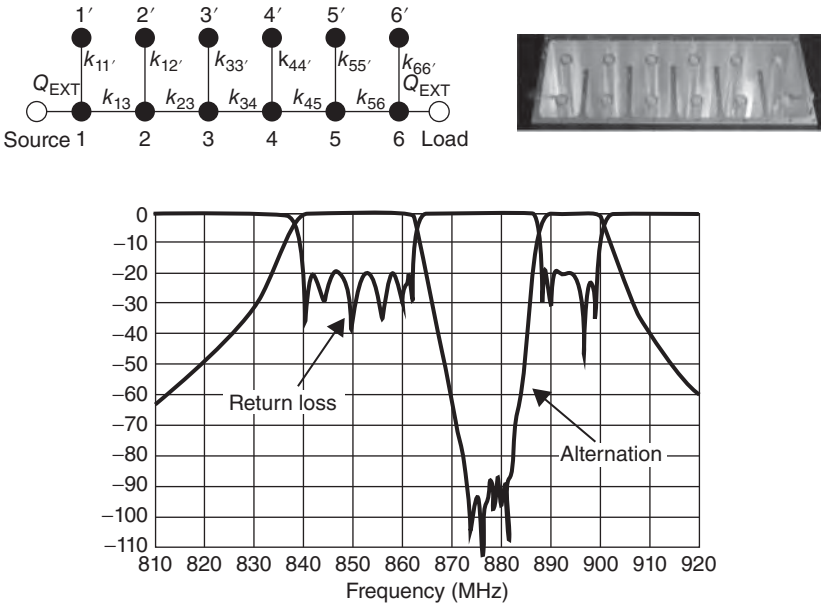


Figure 21.4 Coupling matrix and simulation results of a 12-pole dual-band filter having two passbands with nonequal widths. *Source*: Macchiarella and Tamiazzo 2005 [2]. Reproduced with the permission of IEEE.

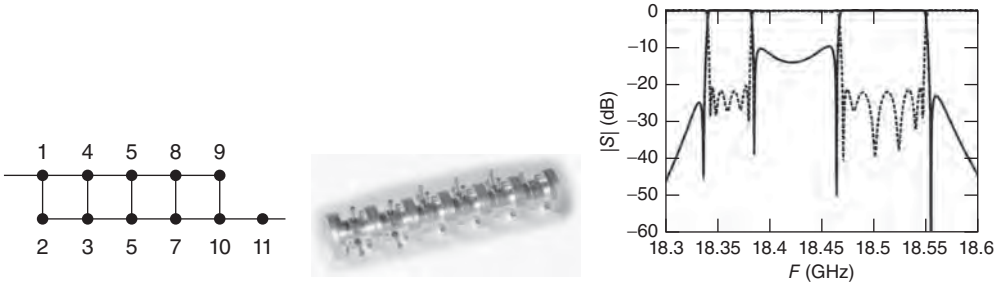


Figure 21.5 Coupling matrix, filter structure, and simulation results of a dual-band filter employing cylindrical dual-mode cavities. *Source*: Lenoir *et al*. 2006 [3]. Reproduced with the permission of IEEE.

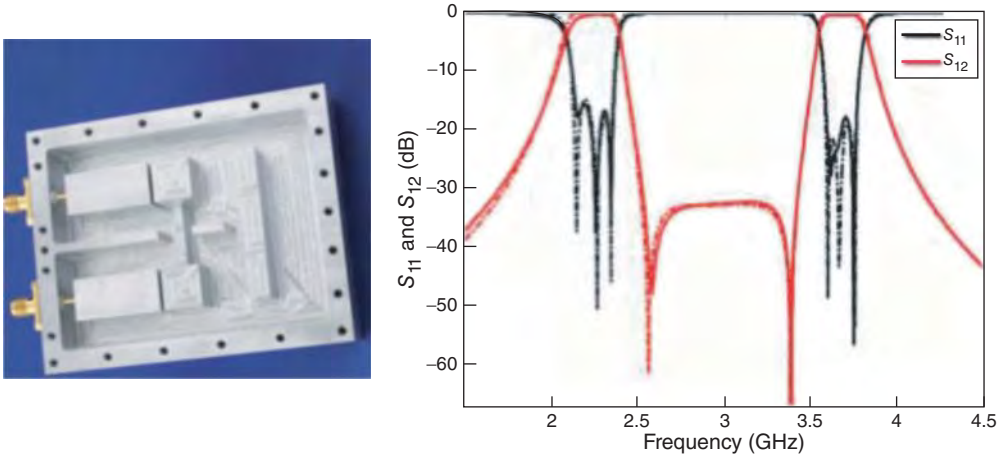


Figure 21.6 A six-pole dual-band wideband filter implemented in ridge waveguide technology. (a) Filter structure and (b) measured results. *Source*: Fahmi *et al*. 2010 [4]. Reproduced with the permission of IEEE.

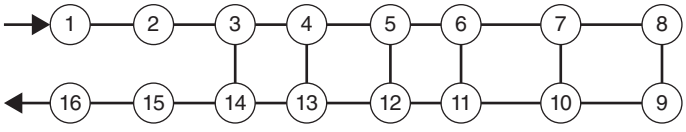


Figure 21.7 A schematic of the coupling matrix of the 16-pole dual-band filter.

measured results obtained for the 16-pole dual-band filter. A good insertion loss is realized due to the relatively large fractional bandwidth of the dual-band filter.

21.3 Approach II: Multiband Filters Employing Multimode Resonators

Multimode resonators can be used for designing multi-passband filters where each mode is used as the operating mode for one of the passbands. The advantages here are significant size

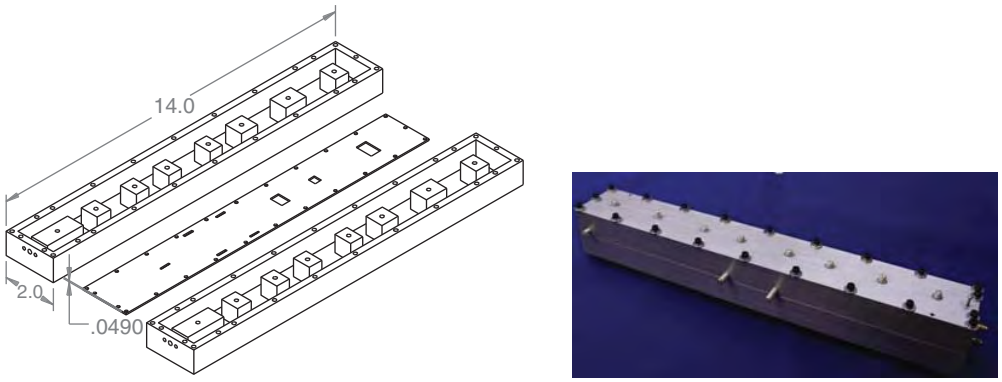


Figure 21.8 Mechanical assembly of the ridge waveguide realization of the filter: two identical parts with a metal plate in between. *Source*: Fahmi *et al*. 2016 [5]. Reproduced with the permission of John Wiley & Sons.

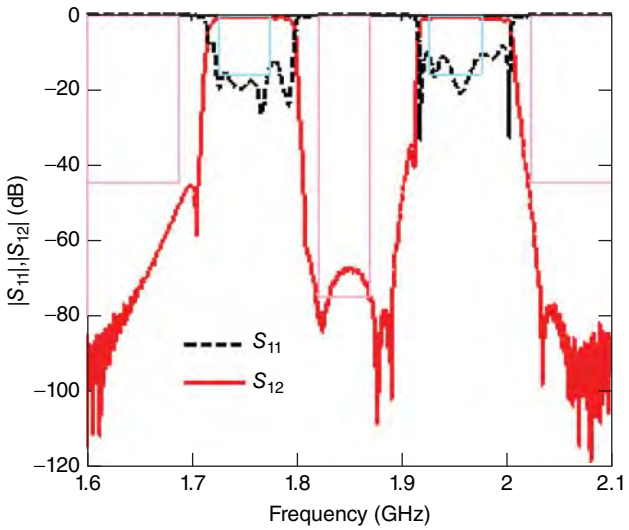


Figure 21.9 The measured response of the fabricated filter shown in Figure 21.8. *Source*: Fahmi *et al*. 2016 [5]. Reproduced with the permission of John Wiley & Sons.

reduction since one physical resonator can serve several passbands. There are, however, several challenges associated with this design approach:

a) For given specific center frequencies of the passbands, the designer must be able to have a full control of the spacing between the resonant modes to align with the center frequencies of the passbands. This requires the choice of a nontraditional resonator configuration or changing the configuration of a conventional resonator to allocate the resonant modes as required.
b) Often multiband filters have passbands with different widths. Thus, the same physical input/output coupling must be able to provide a different coupling to each of the resonant modes that exist in one cavity. This usually necessitates the choice of a nontraditional input/output coupling configuration.

c) Similarly, the interresonator coupling configuration must be able to provide the coupling needed between modes in adjacent and nonadjacent cavities, which can vary from one passband to another.
d) If tuning screws are to be added for fine-tuning, it is important to select the location of these tuning screws to minimize coupling between various modes, since this can limit the attainable rejection between the passbands.

The following subsections describe some of the dual-band designs reported over the past years based on this approach using combline, waveguide, and DRs.

21.3.1 Combline Multiband Filters Employing Multimode Resonators

Figure 21.10 illustrates the configuration of the triple-conductor combline resonator [6, 7]. It is made of three metallic conductors: inner post or rod, intermediate conductor, and enclosure. The inner post and the intermediate conductor are both short-circuited with the enclosure at one end and open-ended at the other end. The resonator carries two modes whose resonant frequencies are varied by controlling the heights of the inner conductor and the intermediate conductor. Table 21.1 gives resonant frequency and Q comparison between a conventional standard combline resonator and the triple-conductor resonator, assuming that the two resonators have the same outer dimensions. It can be seen that the triple-conductor cavity exhibits a lower Q value in comparison with that of a single combline cavity with the same outer dimensions. However the size reduction advantage offered by having two passbands from one physical multiband filter offers a good tradeoff.

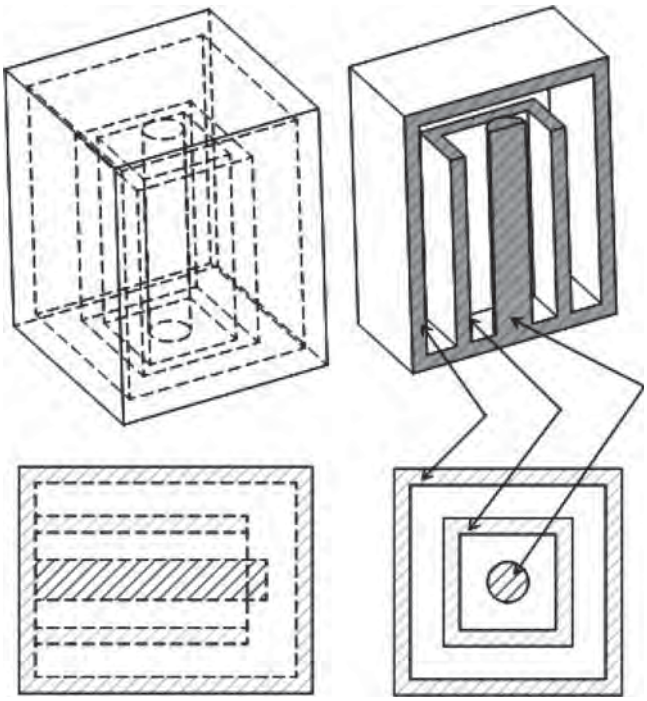


Figure 21.10 The triple-conductor combline resonator. *Source*: Ruiz-Cruz *et al*. 2012 [6]. Reproduced with the permission of IEEE.

Table 21.1 A comparison of a conventional combline resonator and triple-conductor combline resonator.

Combline resonator		f_{res} (GHz)	Q_u (Al)	Q_u/V_{cav} (1/mm^3)	$f_{first\ spurious}$ (GHz)
Standard		1.74	3560	0.097	5.1
Triple-conductor	Mode #1	1.70	2150	0.071	4.1
	Mode #2	1.94	2020	0.067	

Figure 21.11 shows the electric field pattern inside the triple-conductor resonator as obtained by a high frequency structure simulator (HFSS) eigenvalue analysis, for the first and second resonant modes. In Figure 21.11a the inner conductor is shorter in height than the intermediate conductor, whereas in Figure 21.11b we consider the case where the inner conductor is longer in height. We define two regions: region A (inside the intermediate conductor) and region B (outside the intermediate conductor). In both cases, the field of mode 1 points in the same direction in region A and in region B, whereas the field of mode 2 in region A points in an opposite direction to that in region B. We can see that the electromagnetic (EM) field of mode 2 is confined more in region A when the inner conductor is shorter than the intermediate conductors (Figure 21.11a), whereas it is confined more in region B when the inner conductor is longer than the intermediate conductor (Figure 21.11b). Thus, the relative length of the conductors provides a mechanism for controlling the distribution of the field in the outer and internal regions. This is crucial for controlling the coupling between adjacent cavities for dual-band filters having passbands with different widths.

Figure 21.12 shows a three-dimensional cross section of the filter and schematic illustrating a technique to manufacture the filter from three separate parts. The input/output probe is tapped in to the intermediate conductor and can provide enough degrees of freedom to realize different variations of input coupling values to both resonant modes [6]. The intercavity coupling is provided by windows opened at the intermediate conductors of each cavity and by controlling the spacing between the two cavities. Figure 21.13 shows prototype hardware units for a three-pole and a five-pole versions developed in [6] for such type of dual-band filters. The measured results given in Figure 21.14 for the five-pole dual-band filter show that a very good rejection can be achieved at the guard band, reaching 100 dB, due to the TZs inherent in the structure. The dual-band filer exhibits also an out-of-band spurious performance (at $3f_o$) similar to what is typically obtained by conventional combline filters. The skewed behavior that is shown below and above the two passbands follows the skewed dispersion effects observed in standard combline filters. This type of filter has the advantages of the low-cost compact combline filter technology, with further reduction in size. It promises to be useful in emerging dual-band wireless base station filter applications; more details on this structure are given in [6, 7].

21.3.2 Waveguide Multiband Filters Employing Multimode Resonators

Multiband filters can be realized by using waveguide cavities carrying multiple modes, as long as the resonant frequencies of these modes can be independently controlled to fit the required passbands. It is also important that these resonant modes do not couple to allow each filter band to be independently designed. In [8] elliptical cavities are employed to realize dual-band filters. The operating modes in the elliptical cavity are still based on TE_{11n} modes degeneracy

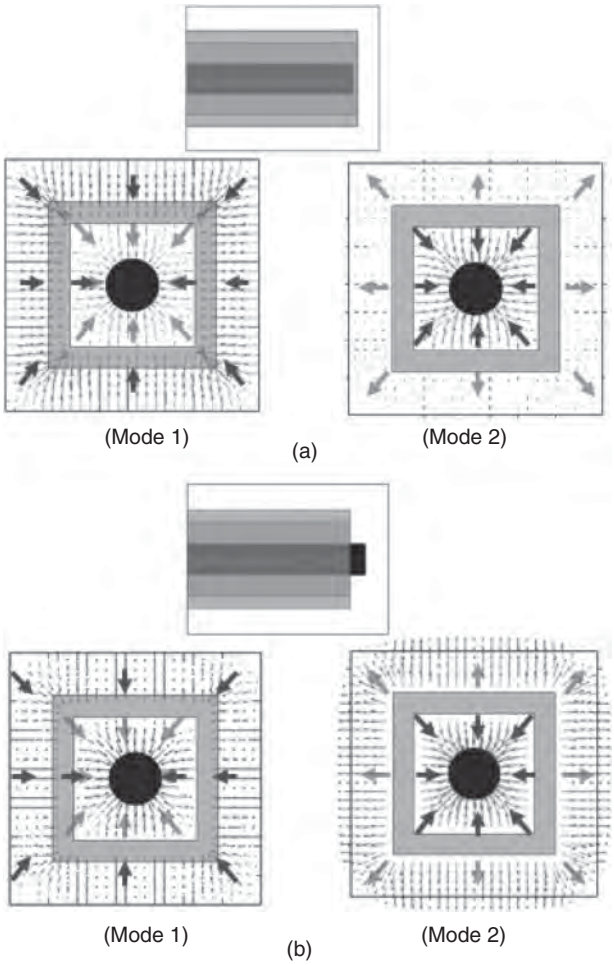


Figure 21.11 Field distribution in triple-conductor combline cavity. (a) Inner conductor is shorter in height than the intermediate conductor. (b) Inner conductor is longer in height than the intermediate conductor.

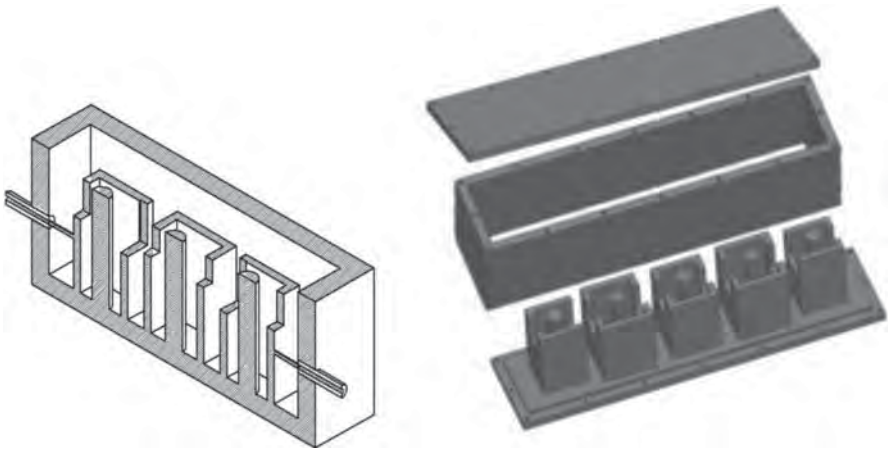


Figure 21.12 A three-dimensional cross section of the filter and schematic illustrating a technique to manufacture it from three separate parts. *Source*: Ruiz-Cruz *et al*. 2012 [6]. Reproduced with the permission of IEEE.

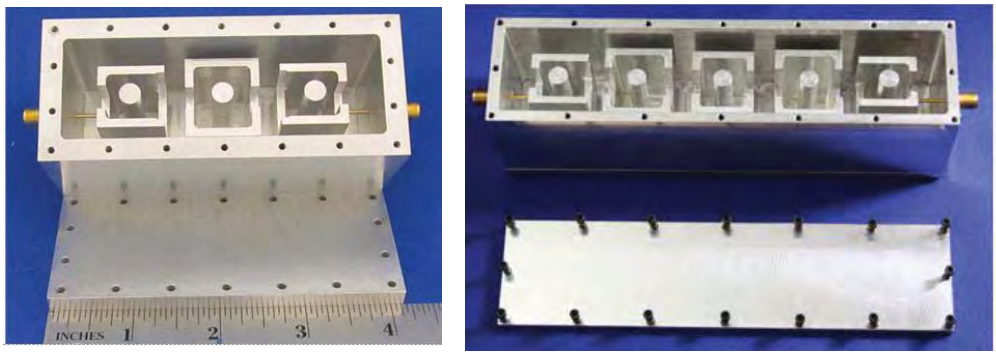
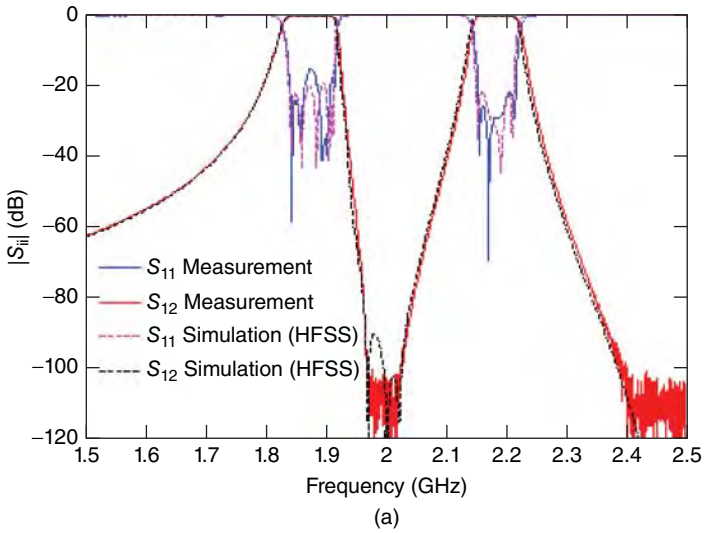


Figure 21.13 Two prototype units for three-pole and five-pole dual-band filters built using triple-conductor combline resonators. *Source*: Ruiz-Cruz *et al*. 2012 [6]. Reproduced with the permission of IEEE.



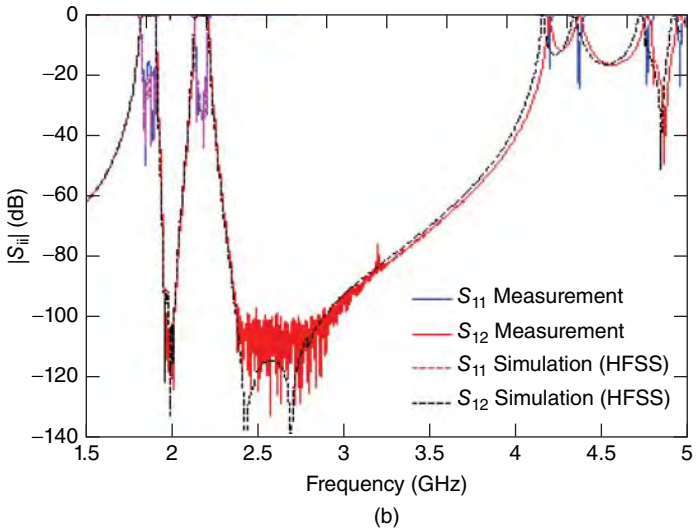


Figure 21.14 The measured results of the five-pole dual band shown in Figure 21.13. *Source*: Ruiz-Cruz *et al*. 2012 [6]. Reproduced with the permission of IEEE.

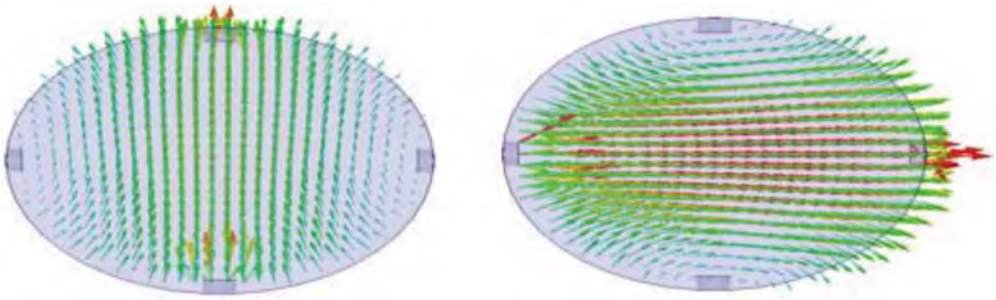


Figure 21.15 The field distribution of TE_{11n} modes in elliptical waveguide cavity.

whose field distribution is given in Figure 21.15. The center frequencies of these two modes can be adjusted by varying the elliptical axes, leading to two orthogonally polarized modes that are intrinsically isolated and can be employed to carry the two passband signals. The design of the dual-band filter in this case is similar to that of a dual-mode filter employing cylindrical TE_{11n} cavities with the exception that the input/output coupling in the dual-band filter must be designed to provide the proper coupling to the two modes. Having the proper coupling to the two modes can be realized by both controlling the rotation angle of the feeding waveguide with respect to the two axes and controlling the iris dimensions. The input coupling can be obtained by calculating the group delay of the reflection coefficient seen at the input of the feeding waveguide. Figure 21.16 shows the input group delay versus the rotation angle ϕ of the feeding waveguide. It clearly shows that the angle of rotation not only can control the input coupling to the two modes but also can cause a different loading on the two resonant nodes. Thus, a slight readjustment of the axis dimensions may be necessary to get both the center frequencies and the input/output coupling needed. A procedure similar to that given in Chapter 14 can then be used to calculate the coupling to both modes.

For coupling between adjacent cavities, cross irises are used as shown in Figure 21.17. The use of narrow-width cross slots helps to isolate the two passbands and makes it possible to design the dual-band filter as two independent filters. An eighth-order Ku-band dual-band filter using this concept was demonstrated in [8]. The filter configuration is given in Figure 21.18. It consists of six TE_{112}-mode and two TE_{113}-mode elliptical cavities. The two TE_{113} cavities were employed to improve the overall filter Q and its spurious-free window. Step-transformer portions are added between the interface waveguide and the input/output iris to increase the coupling to both resonant modes for a relatively wide fractional BW (about 2%). Tuning screws were added for fine-tuning. The measured performance compared with the simulation is also shown in Figure 21.18.

The same concept can be potentially employed in rectangular cavities where the dimensions of the cavities are selected to control the frequency spacing between modes TE_{101} and TE_{011}. The use of rectangular cavities makes it easy to realize dual-band filters where each passband has TZs. Figure 21.19 shows a coupling configuration for dual-band filter employing a trisection for each band. A trisection can be physically realized in dual-band filters employing rectangular waveguides as illustrated in Figure 21.20 by positioning two single-band cavity resonators between two adjacent dual-band cavities. This configuration generates two TZs, one on either side of each passband. If a filter contains N dual-band resonators, a maximum of $N - 1$ TZs can be potentially achieved [8].

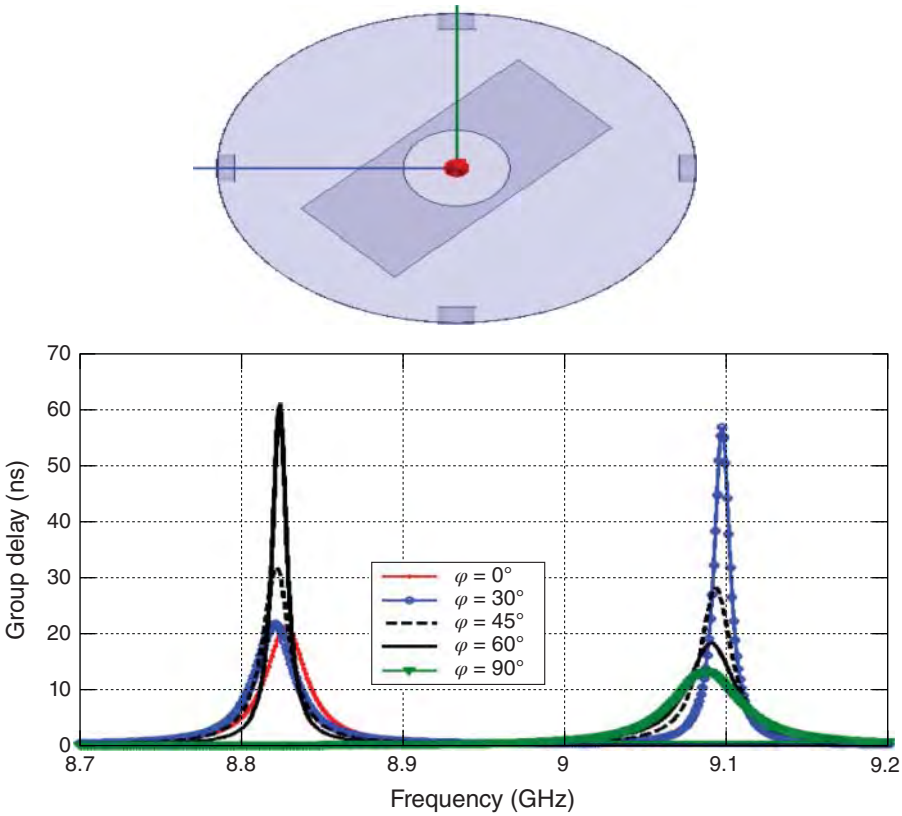


Figure 21.16 (a) Input round iris coupling with the input waveguide rotated by angle ϕ. (b) Group delay response with different angle ϕ.

Figure 21.17 Configuration of intercavity coupling with cross-slot iris.

A triple-band filter employing waveguide filters was proposed in [9]. Three modes in a rectangular waveguide cavity, TE_{101}, TE_{011}, and TM_{110}, are used where each mode is associated with one of the three filter passbands. For example, having a cavity with dimensions $2.2'' \times 1.9'' \times 2.4''$ yields a resonance frequency and Q value for the three modes as given in Table 21.2. Figure 21.21 shows a schematic of the coupling configuration and the filter structure for a fourth-order triple-band filter realized using rectangular waveguide cavities. The modes in the high, middle,

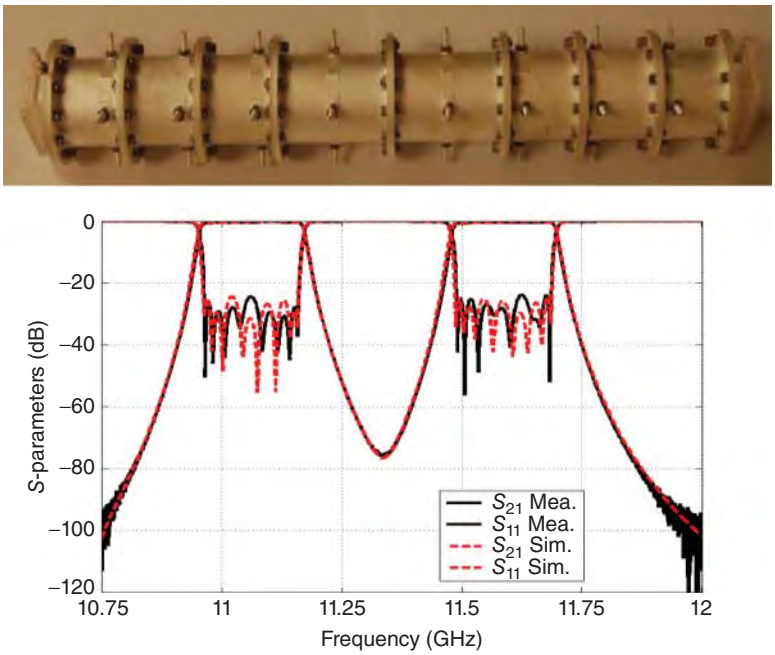


Figure 21.18 (a) Dual-band filter employing elliptic cavities. (b) Measured and simulated *S*-parameters. *Source*: Zhu *et al*. 2017 [8]. Reproduced with the permission of IEEE.

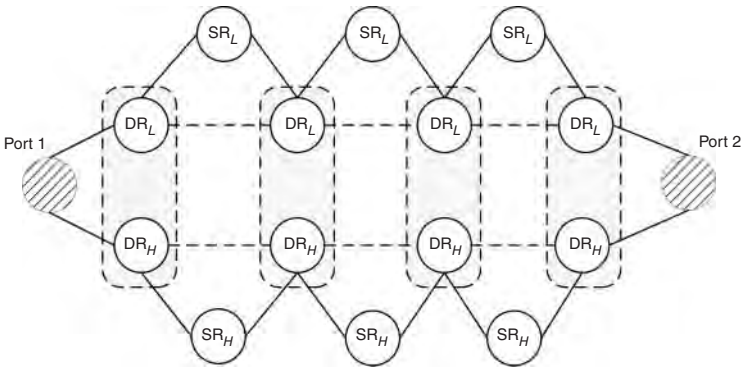


Figure 21.19 A coupling configuration for dual-band filter employing trisections for each band.

and low branches of the proposed triple-band filter do not couple, which allows each path to be synthesized and designed separately. The proposed structure has the ability to fully control all self-coupling and sequential coupling parameters.

The interresonator coupling is realized using the iris/probe configuration shown in Figure 21.22. The coupling between cavities is implemented using a hybrid coupling structure consisting of a cross slot and a probe. The slots/probe dimensions can be extracted using eigenvalue simulation by placing electric and magnetic walls along the symmetry plane as described in Chapter 14. A unique input/output coupling probe was used to provide the

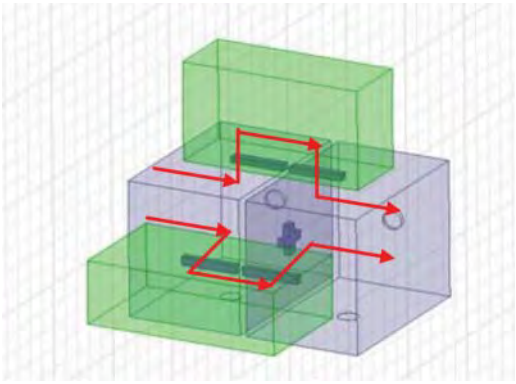


Figure 21.20 Physical realization of trisections in dual-band filters employing rectangular waveguides.

Table 21.2 The resonance frequencies and the *Q* for the three modes employed in triple-band rectangular waveguide filter.

Dimensions (in.)	Mode	Frequency (GHz)	Unloaded *Q* (copper)
$a = 2.2$	TE_{101}	3.62	16,716
$b = 1.9$	TM_{110}	3.83	17,104
$d = 2.4$	TE_{011}	4.04	17,566

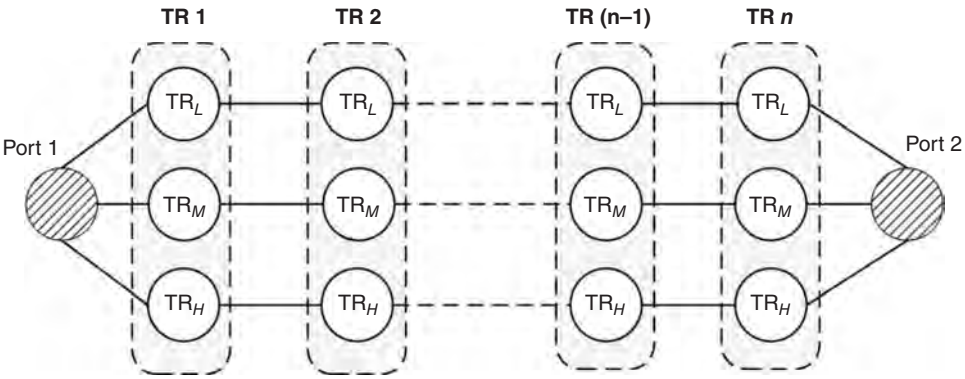


Figure 21.21 A schematic of the coupling configuration for a fourth-order triple-band filter.

required coupling to each of the three modes in a single cavity. Figure 21.23 shows an L-shaped probe configuration proposed in [9] to achieve this objective. The probe has four parameters, L_1, L_2, ΔL, and the angle θ, that can be independently varied to adjust the input/output coupling to the three modes. The coupling to mode TM_{110} is mainly determined by length L_1, whereas L_2, ΔL, and θ determine the coupling to the TE_{101} and TE_{011} modes [9]. The input coupling coefficients to three modes can be calculated from the maxima of three peaks of the group delay of the reflection coefficient seen at the probe. Figure 21.24 illustrates the manufactured triple-band filter and a schematic of its internal configuration. Three screws are added to each cavity to fine-tune filter performance. The measured results are also shown in the same figure. An isolation of more than 50 dB between the passbands was achieved.

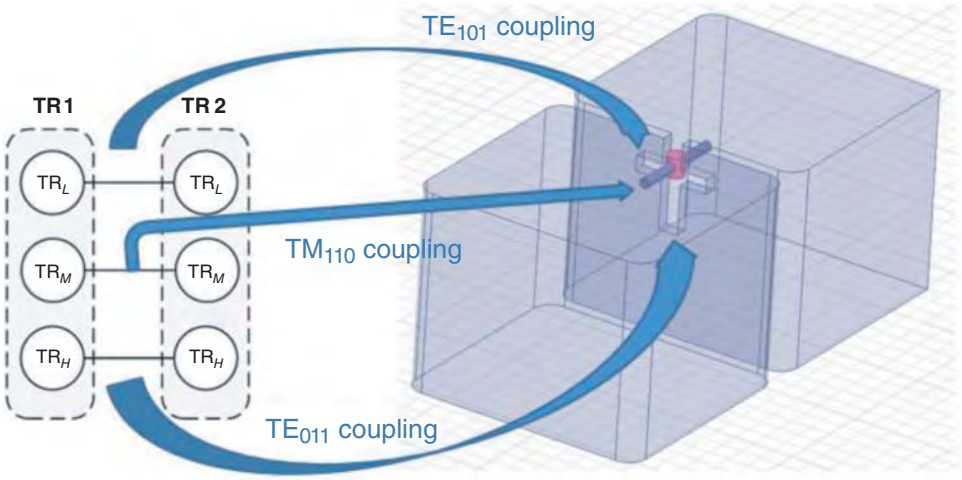


Figure 21.22 Interresonator coupling in triple-band waveguide filter.

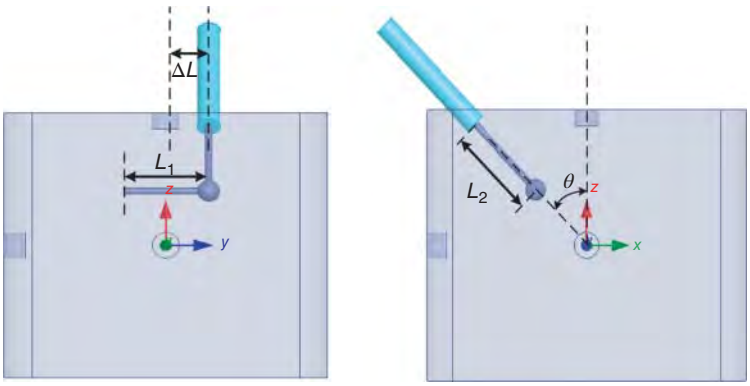
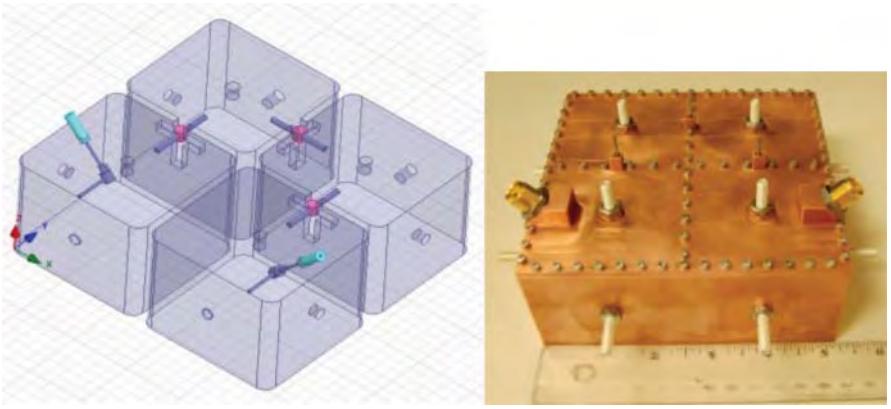


Figure 21.23 Configuration of the input/output coupling probes in the triple-band waveguide filter.

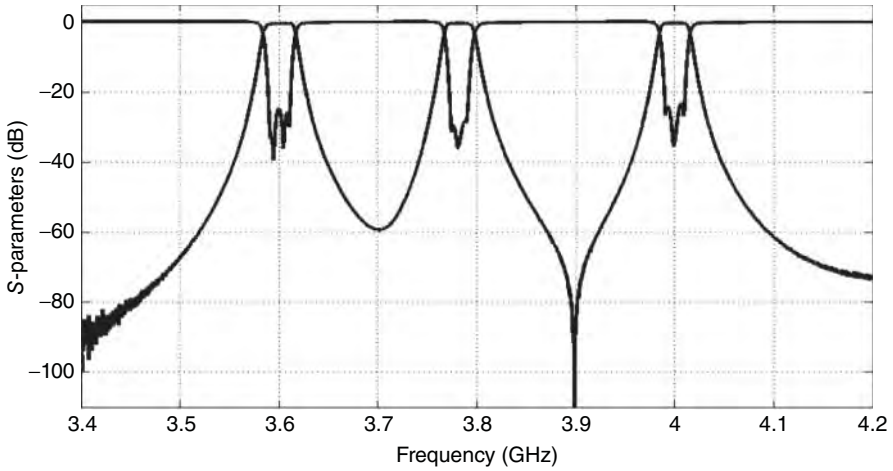
21.3.3 Dielectric Resonator Multiband Filters Employing Multimode Resonators

As the case with any microwave resonator, a DR exhibits an infinite number of resonant nodes. These nodes are closely spaced, making DRs appear attractive for use in the realization of multiband filters. However, since the conventional cylindrical-shaped DR has only two-dimensional parameters to adjust resonator diameter and resonator height, usually it is not easy to align the resonant frequencies exactly with the passband center frequencies required for the multiband filter. A DR having a nontraditional cylindrical shape needs to be used to simultaneously select the exact resonant frequencies needed.

Figure 21.25a shows a dual-band DR configuration that is machined from a single piece of the ceramic dielectric substrate with a high dielectric constant of 45 [10]. The resonator consists of two rectangular ceramic sections, R1 and R2, whose dimensions can be selected to create two resonant modes with specific center frequencies. Figure 21.25b and c illustrates the electric field distributions of the first two resonant modes as obtained by the HFSS eigenvalue solver. The field distribution of each of these two modes resembles the field distribution of the TEH_{01} modes that would exist in sections R1 and R2 if separated from each other. The field distribution



(a)



(b)

Figure 21.24 (a) Configuration of the triple-band waveguide filter. (b) The measured results. *Source*: Zhu *et al*. 2016 [9]. Reproduced with the permission of IEEE.

seen in Figure 21.25a and b makes it easy to associate each mode with a specific passband. Similar to the traditional DRs that operate in TEH_{01} modes, slot holes were introduced in the middle of sections R1 and R2 to shift the spurious modes away from the fundamental mode TEH_{01}.

To allow independent coupling of modes between adjacent cavities, the two coupling iris configurations shown Figure 21.26 are proposed in [10]. The coupling structure shown in Figure 21.26b was selected since it produces larger coupling values with a reasonable distance separation between the resonators. For input/output coupling, a coaxial probe is placed codirectional with the electric field of the first and second modes as shown in Figure 21.27. The desired input coupling to the two modes is realized by varying the length L, the horizontal location C, and the vertical location H of the probe.

Figure 21.28 shows a dual-band filter of order 2 implemented using such type of resonators. The DRs are supported inside the cavity using a dielectric support made of out of Teflon with a low dielectric constant of 2.1. It is cut and machined to the desired shape with a groove to allow mounting the dual-mode DRs parallel to each other. The filter housing is made of

	S	1	2	3	4	5	6	7	8	*L*
S		1.152				1.152				
1	1.152	−10.15	1.041							
2		1.041	−10	0.772						
3			0.772	−10	1.041					
4				1.041	−10	0				**−1.152**
5	1.152				0	10	1.041			
6						1.041	10	0.772		
7							0.772	10	1.041	
8								1.041	10.15	**1.152**
L					**−1.152**				**1.152**	

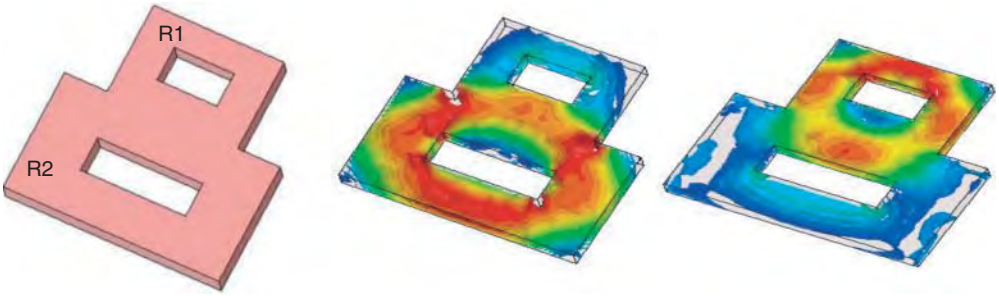


Figure 21.25 A dual-band dielectric resonator configuration that is machined from a single piece of the ceramic dielectric substrate and the field distributions of the first two modes.

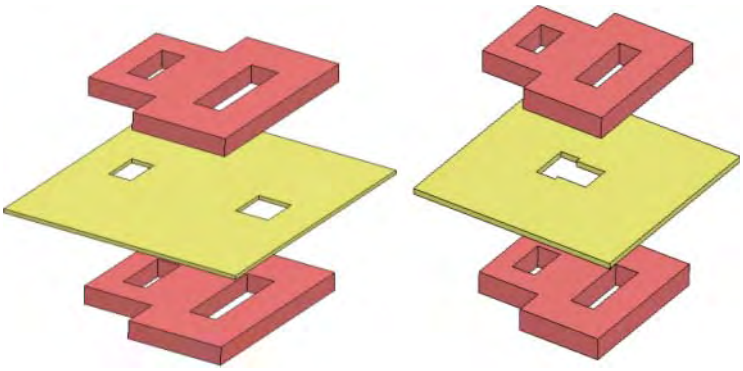


Figure 21.26 Two possible configurations for coupling modes in adjacent cavities.

two identical cavities that are bolted together with an iris in between. Two tuning screws are inserted in each cavity, one in the top wall and one in the bottom wall (not shown), to adjust the center frequencies of the two passbands of the filter to account for variation of dielectric constant from batch to batch. A third-order dual-band filter of this type was also demonstrated in [10].

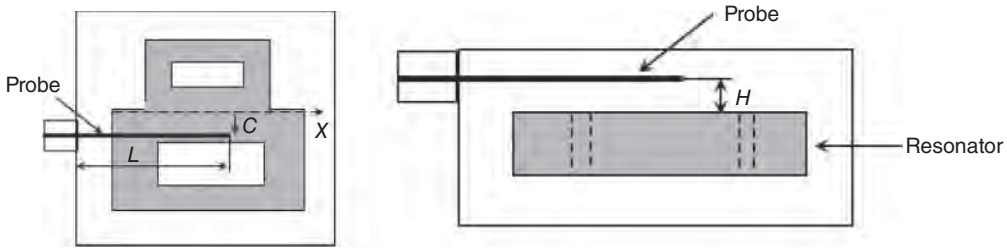


Figure 21.27 Input/output probe coupling.

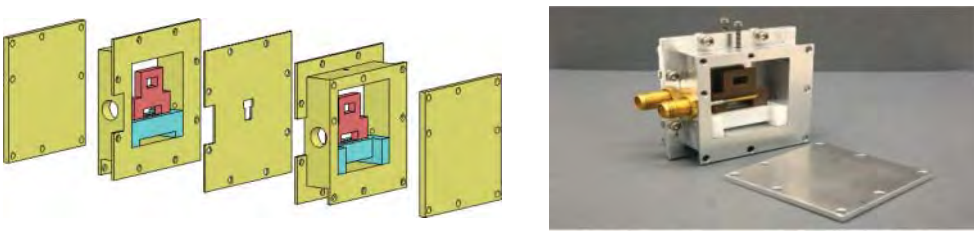


Figure 21.28 A second-order dual-band dielectric resonator filter. *Source*: Zhang and Mansour 2009 [10]. Reproduced with the permission of IEEE.

The half-cut DR structure was introduced in [11] as the building block of a new class of dual-mode DR filters [12, 13]. The half-cut structure is made from cutting the typical cylindrical DR into half along its axis, as shown in Figure 21.29. The DR is planar mounted on a support of low ε_r inside a metal enclosure. $^1/_2\text{HEE}_{11}$ and $^1/_2\text{HEH}_{11}$ are one-half of a single component of modes from the full cylinder resonator. By proper sizing of the resonator's diameter (D) and height (L), these modes can be made to resonate at an equal frequency as demonstrated in [11] or to resonate at different frequencies, where each resonance associates with a passband as demonstrated in [12].

The interresonator coupling is adjusted by using polarization discriminant irises, as shown in Figure 21.30, that allow for independent control of interresonator coupling in each band. The horizontal and vertical slots enable the almost independent coupling of the $^1/_2\text{HEE}_{11}$ and $^1/_2\text{HEH}_{11}$ modes, respectively. Aside from using two orthogonal irises, one slanted iris may also be used to couple both modes simultaneously. The input and output coupling of the filter can be realized using one probe. The location of the probe as well as its length determines the amount of coupling to the first and last resonators of each band. The spurious performance of the filter is also of importance and can be improved by adding a slot in the half-cut resonator [11].

Figure 21.31 shows the structure for a three-cavity dual-band filter employing half-cut resonators [12]. Its simulation results are also shown in the same figure. It can be seen that there exists an interband TZ between the two bands. The presence of this TZ depends on the order of the dual-band filter. In each cavity, the two modes are 180° out of phase for frequencies between the two bands. Therefore, for odd-order dual-band filters, the number of total phase reversals in each cavity is an odd multiple of 180°, and hence the two signal paths can subtract at the output and create a TZ. For even-order filters however, the same separate inline design does not introduce an interband TZ because the two interband signals coming from the two paths go through an even number of 180° phase reversals and always add up constructively. In order to introduce a TZ, one of the coupling values in the two bands must have a different sign.

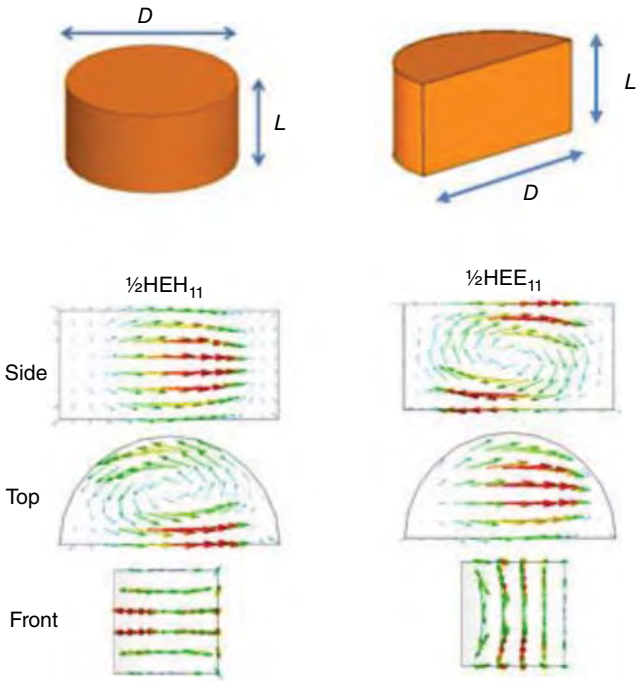


Figure 21.29 (a) A half-cut dielectric resonator generated by cutting a traditional DR resonator into two parts. (b) The field distribution of the first two modes of the half-cut DR resonator.

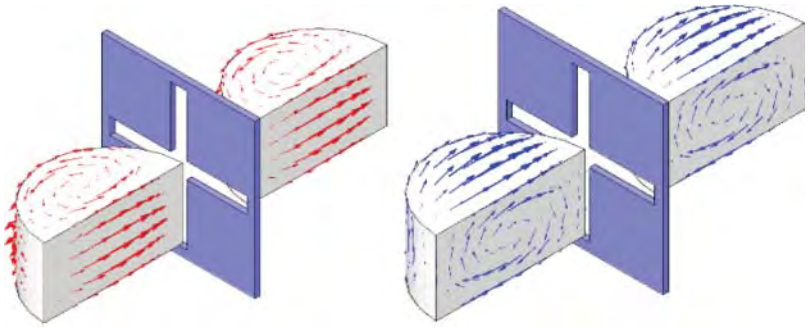


Figure 21.30 Polarization discriminant irises for coupling of the orthogonal modes between cavities: (a) coupling of $\frac{1}{2}HEE_{11}$ through horizontal iris and (b) coupling of $\frac{1}{2}HEH_{11}$ through vertical iris.

For example, consider the fourth-order dual-band DR filter shown in Figure 21.32. The coupling matrix of this filter is given in equation (3.1). The two passbands are designed separately. Using a positive sign for M_{8L}, yields no TZ between the two bands. However, choosing a negative sign yields an interband TZ, which translates to having a negative output coupling to the last resonator for only one band. Physically, this can be realized by placing the probe according to the arrangement shown in Figure 21.32b where the output probe is moved close to the other edge of the resonator, causing output coupling for the $\frac{1}{2}HEE_{11}$ mode to reverse in polarity but keeping the same polarity for the $\frac{1}{2}HEH_{11}$ mode. This in turn leads to the presence of a TZ between the two bands as shown in Figure 21.32c.

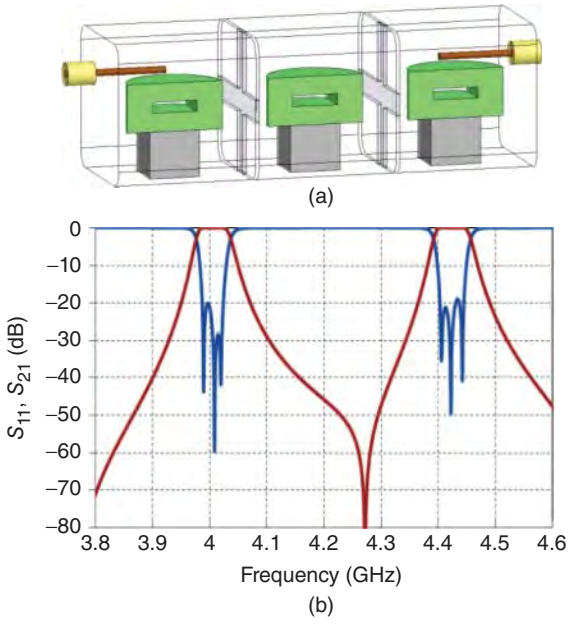


Figure 21.31 (a) The filter structure of a three-pole dual-band DR filter employing half-cut resonators. (b) Simulation results. *Source*: Memarinan and Mansour 2009 [12]. Reproduced with the permission of IEEE.

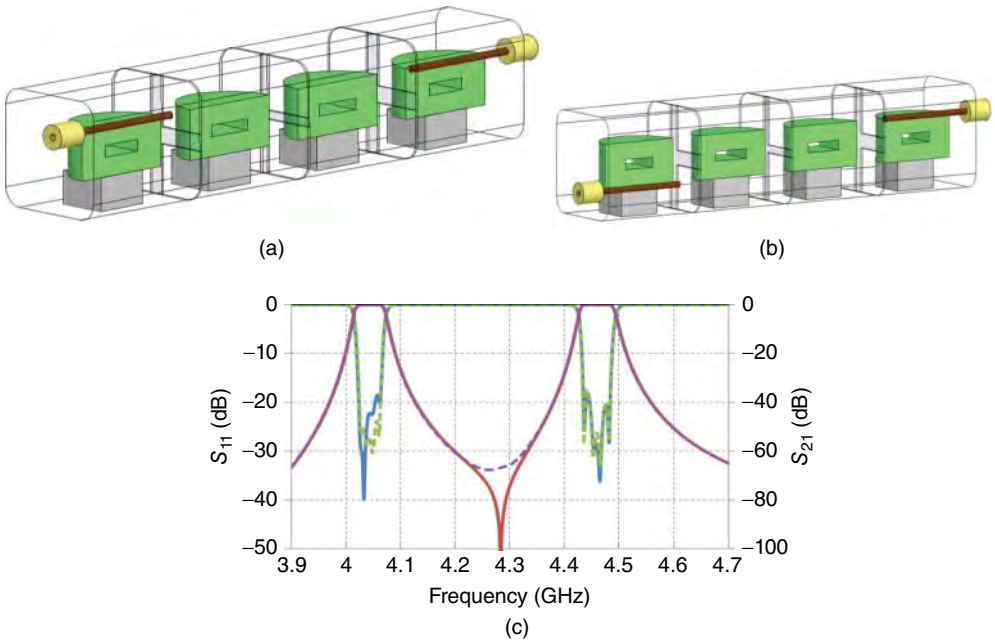


Figure 21.32 A fourth-order dual-band filter illustrating the impact of manipulating the locations of the input/output ports to insert a transmission zero between the two passbands. *Source*: Memarinan and Mansour 2009 [12]. Reproduced with the permission of IEEE.

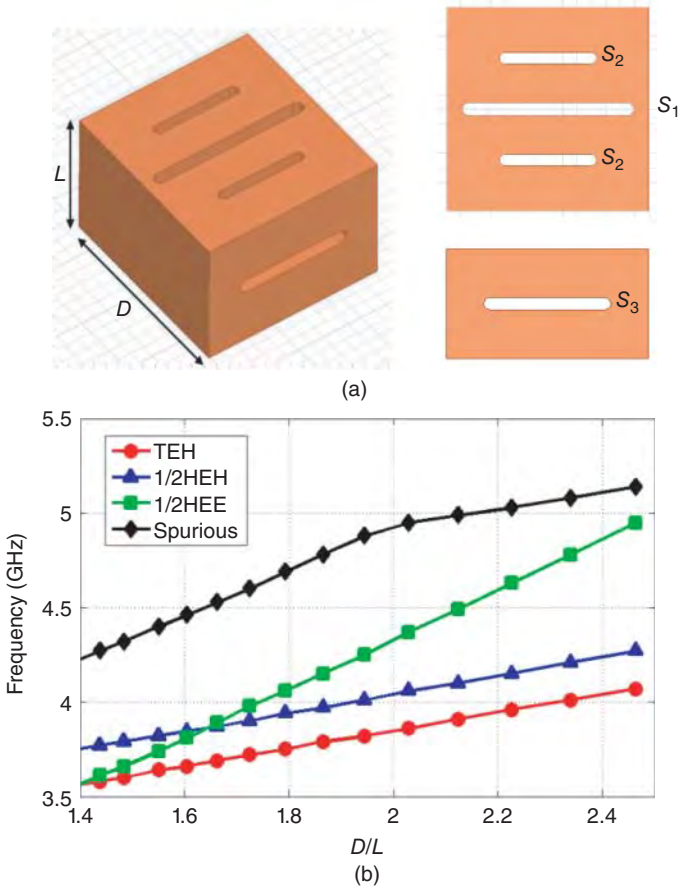


Figure 21.33 (a) Resonator configuration. (b) Mode chart of a triple-mode dielectric resonator with $D = 0.69$ in, $\varepsilon_r = 35$, and $1 \times 1 \times 0.9$ in.^{3} cavity. *Source*: Zhu *et al*. 2017 [14]. Reproduced with the permission of IEEE.

A unique triple-band DR filter was proposed in [14]. The filter employs DRs that support three resonant modes. The DR is in the form of a rectangular dielectric block integrated with slots to allow for a full control of the resonant frequency of the three modes and to facilitate interresonator coupling. Figure 21.33a shows the resonator configuration, while Figure 21.33b shows its mode chart. The first three modes are TEH_{01}, $½HEE_{11}$, and $½HEH_{11}$. Each mode is associated with one passband of the triple-band filter. The use of slots eliminates the degenerated counterparts of the operating $½HEE$ and $½HEH$ modes. The slots provide flexibility to fine-tune resonant frequencies of each mode, that is, S_1 to control TEH_{01} mode, S_2 to control $½HEH_{11}$ mode, and S_3 to control $½HEE_{11}$ mode [14]. Figure 21.34 shows a schematic of a third dual-band filter based on this concept. The interresonator coupling structure is a combination of a horizontal iris, a vertical iris, and an electrical probe, coupling the TEH_{01}, $½HEE_{11}$, and $½HEH_{11}$ modes, respectively. The electrical probe crossing at the opening of the vertical iris is flexible for tuning adjustment. A picture of the fabricated unit and the experimental results are shown in Figure 21.35. The filter demonstrates a loaded Q of close to 7000 at C-band. TZs were observed between the passbands due to the 180° phase cancelation described above. More design details on this filter can be found in [14].

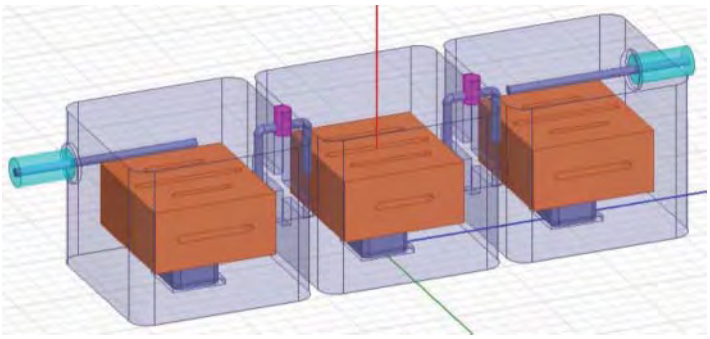


Figure 21.34 A schematic of a third-order dielectric triple-band filter.

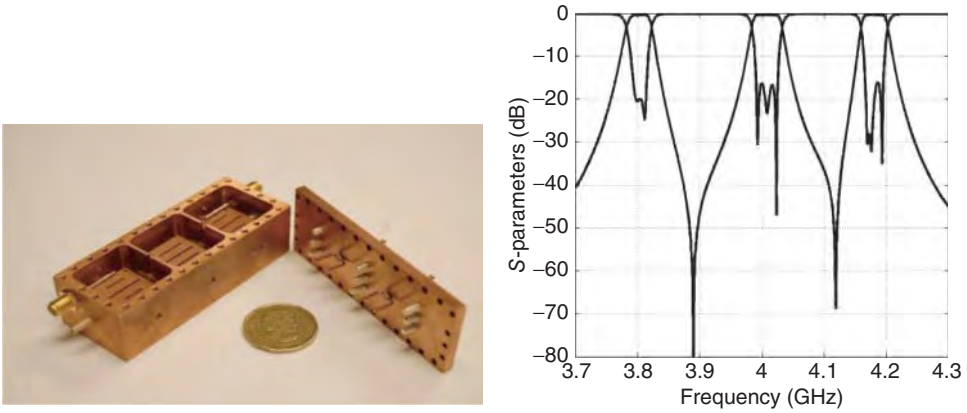


Figure 21.35 (a) A third-order triple-band dielectric resonator filter. (b) Measured results. *Source*: Zhu *et al*. 2017 [14]. Reproduced with the permission of IEEE.

21.4 Approach III: Multiband Filters Using Parallel Connected Filters

This approach can be implemented by having single-band filters connected in parallel using two manifold junctions. This is the simplest conventional approach to realize multiband filters. The challenge is in synthesizing the coupling matrix of filters that are terminated at both ends in manifold junctions. A synthesis procedure was reported in [15] for parallel-connected filters. It should be noted however that for filters that have physical symmetry such as Chebyshev filters, the use of conventional manifold-coupled multiplexer synthesis techniques with some optimization can yield a good solution. The single-band filters could be of any type, microstrip, combline, waveguide, or DRs.

A triple-band filter is implemented in [16, 17] using high-temperature superconductor (HTS) filters. Figure 21.36a shows the layout of the triple band. It consists of three HTS filters, each of order 8, connected in parallel using two microstrip manifolds made also of HTS films. In order to reduce the parasitic coupling effects between the resonators of the three filters, the filters and manifolds are assembled in a housing built with metal barriers as shown in Figure 21.36b. Wire bonds were used to connect the three filters to the two manifolds. The size of the overall triple-band HTS filter package is $1.68'' \times 1.68''$. The packaged HTS triple-band

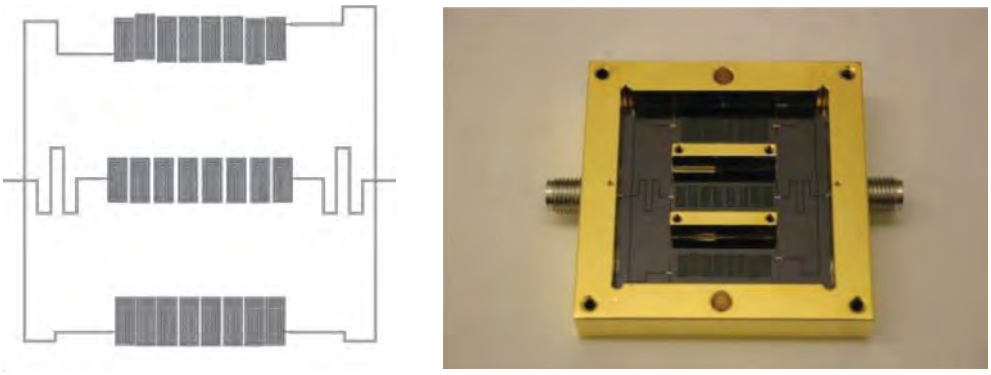


Figure 21.36 (a) A triple-band HTS filter employing parallel connected filters. (b) An image of the packaged triple-band HTS filter [16].

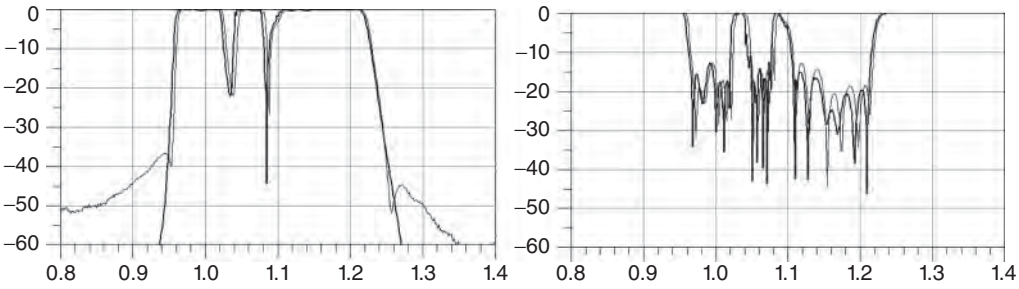
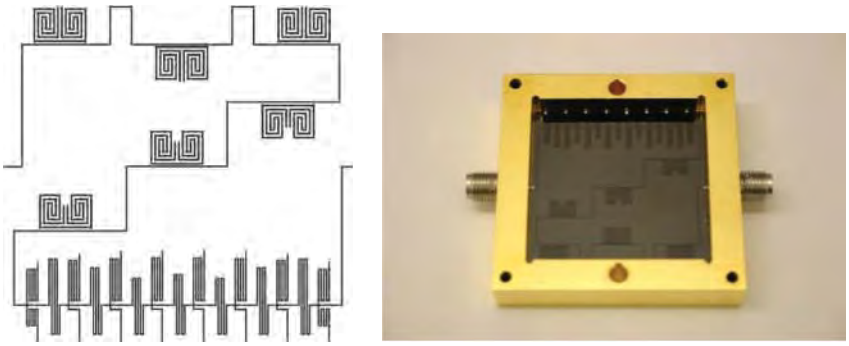


Figure 21.37 A comparison between the EM simulation results and the measured results of the HTS triple-band filter shown in Figure 21.36.

filter was then cooled down to a temperature of 77 K and measured over the frequency range of 800–1400 MHz. Figure 21.37 shows a comparison between the sonnet EM simulation and the measured results. It should be noted that in order to get such close correlation between EM simulation and measurement when dealing with superconducting circuits, the kinetic inductance of the HTS material must be taken into consideration in the sonnet simulation.

21.5 Approach IV: Multiband Filter Implemented Using Notch Filters Connected in Cascade with a Wideband Bandpass

There are many techniques available for designing a notch filter (see Chapter 10). A multi-notch filter can be implemented by simply cascading several of these single-notch filters. The multi-notch filter can be made tunable by adding tuning elements to the resonators, potentially allowing the realization of multiband filters with tunable widths for the passbands. Figure 21.38a shows the layout of an integrated triple-band HTS filter implemented by cascading a dual-notch HTS filter with a wideband HTS filter [16, 17]. The dual-notch filter consists of two sets of folded resonators coupled to a microstrip transmission line. The spacing between the resonators is properly selected to realize the required rejection bandwidth and to achieve enough high rejection. The wideband bandpass filter has a percentage bandwidth of approximately 24%. The design used for this wideband filter consists of HTS resonators in



(a) (b)

Figure 21.38 (a) A triple-band filter formed by integrating dual-notch filter with a wide bandpass filter. (b) A picture of the packaged triple-band filter. *Source*: Mansour and Laforge 2016 [17]. Reproduced with the permission of IEEE.

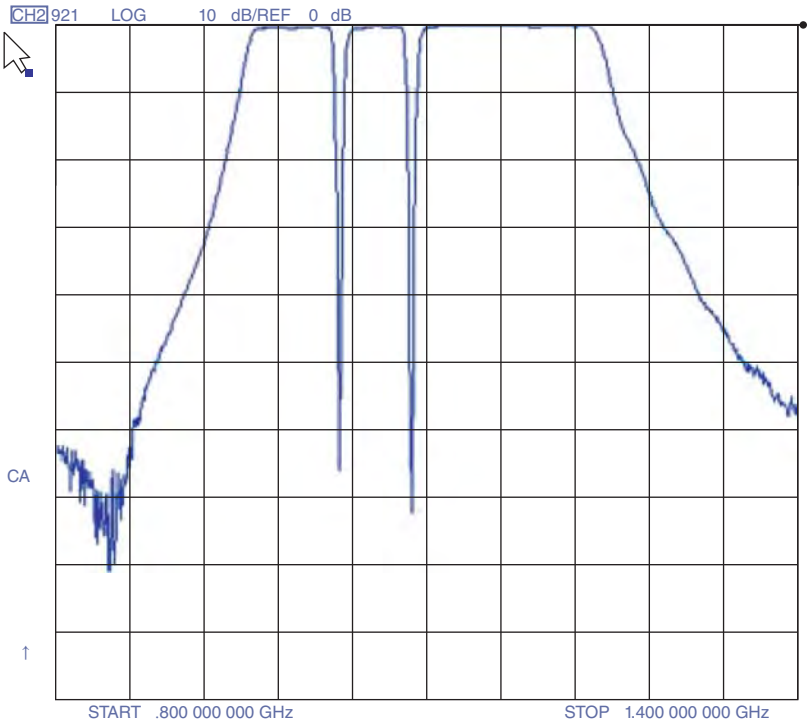


Figure 21.39 The measurement results of the HTS triple-band filter shown in Figure 21.38. *Source*: Mansour and Laforge 2016 [17]. Reproduced with the permission of IEEE.

the form of folded stubs spaced by $\lambda g/4$ microstrip lines. The $\lambda g/4$ inverters are also folded to reduce the size of the L-band eight-pole HTS filter to $1.2'' \times 0.35$. The length of the microstrip line that connects the two filters is carefully selected to improve the match between the two filters at the frequency band of interest. The integrated unit is assembled in the package shown in Figure 21.38b for radio frequency (RF) testing at cryogenic temperatures. Figure 21.39 shows the measured results. It can be seen that the overall band of the bandpass filter is divided into three bands whose widths are determined by the location of the notches introduced by the dual-notch filter.

21.6 Use of Dual-Band Filters in Diplexer and Multiplexer Applications

A dual-band filter can be potentially used as a diplexer, leading to a significant reduction in mass and size of the diplexer [13]. Figure 21.40 shows two possible schematic configurations for generating a diplexer from a dual-band filter. In the first approach, two output ports are connected to the last cavity where each port couples to one of the modes in the cavity. The second configuration uses two single-mode cavities connected to the last cavity, which are in turn coupled to the output ports. The advantage of the second approach is that it offers a better isolation performance between the two diplexer channels since the output ports are not coupled to the same cavity. Figure 21.41a shows a diplexer employing a fourth-order dual-band filter based on the approach described in Section 21.3.2 where four elliptical cavities are used to realize the dual-band filter. The only difference in this case is that the two passbands (channels) are separated through the use of two output waveguide ports that are connected perpendicularly to

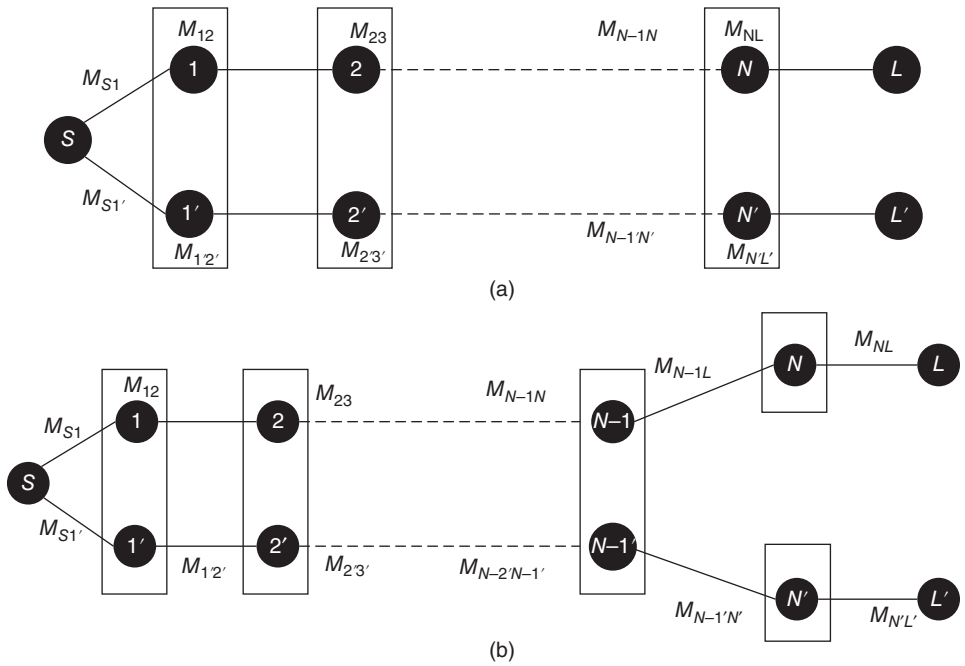
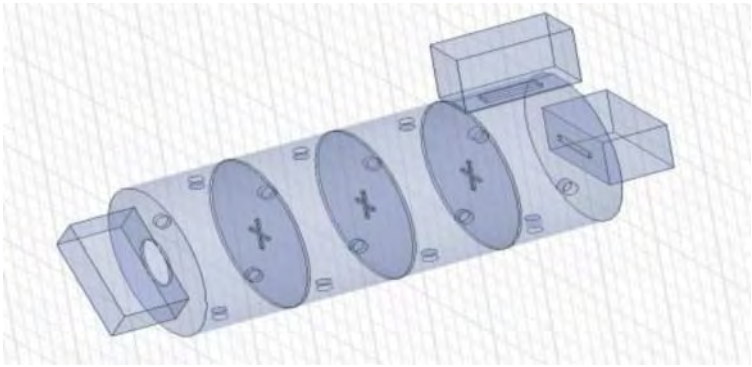
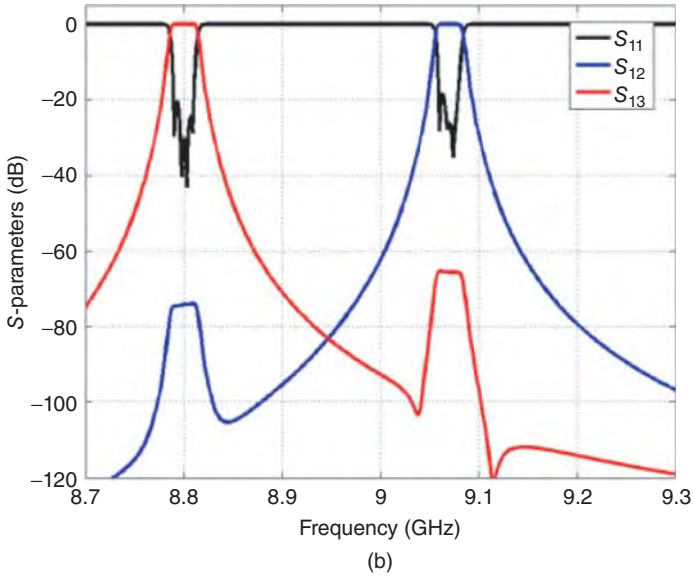


Figure 21.40 Two possible schematic configurations for generating a diplexer from a dual-band filter.



(a)



(b)

Figure 21.41 (a) A diplexer employing fourth-order dual-band filter. (b) Simulations.

couple to the two orthogonal modes in the last cavity. The simulation results of this diplexer are shown in Figure 21.41b. A hump close to 60 dB is observed in the S_{21} response of each channel at the band of the other channel. The hump is intrinsic from this type of diplexer because the two output ports share the same output dual-mode resonator [8]. The hump level can be reduced by using the approach shown in Figure 21.40b through the use of two single-mode rectangular waveguide cavities connected to the last cavity. In this case each of the channels will be of order 5 rather than order 4, yet the diplexer size will be still smaller than the traditional diplexer configurations that use manifold/coupling junctions to couple the two channels.

A diplexer employing a dual-band rectangular waveguide filter was proposed [8]. The diplexer configuration is given in Figure 21.42; it employs two output waveguides placed perpendicularly to the last cavity of a rectangular dual-band filter to form a diplexer. The dual-band filter has three cavities, each operating in TE_{101} and TE_{011} with resonant frequencies that correspond to the two passbands. A trisection in each band is formed using a rectangular cavity operating in TE_{102} mode coupled to the two first dual-band filter cavities as shown in Figure 21.42. Thus, each channel has an order of 4. It is interesting to note that this trisection coupling configuration

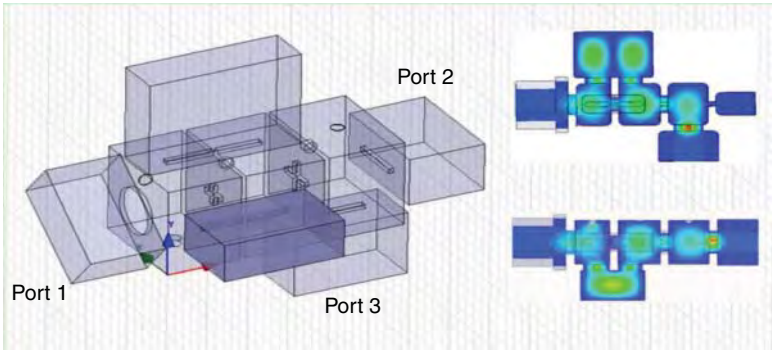


Figure 21.42 (a) A schematic of a diplexer employing a dual-band rectangular filter. (b) Simulations.

generates a TZ on the upper side band of one channel and on the lower side band of the other channel, a feature needed in diplexer applications to improve the isolation between the two channels.

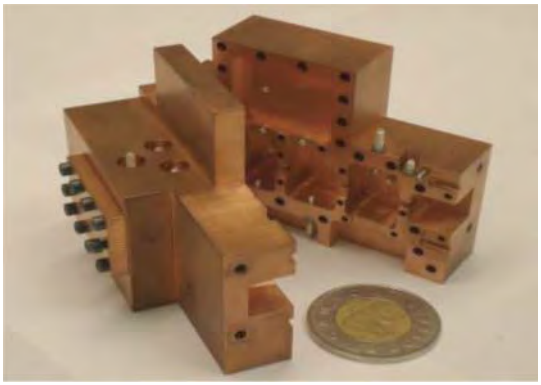
The designed diplexer was manufactured in two halves from copper as shown in Figure 21.43. Tuning screws are added for independent control of frequency and coupling parameters. Figure 21.43a shows the manufactured diplexer unit, while Figure 21.43b gives a comparison between the measured and simulation results. Two humps are also observed in the performance, representing the interaction between the two channels, since the output waveguides are connected orthogonally to the third dual-band cavity. As the case with the previous example, the hump level can be potentially reduced by adding two single-band cavities to the last cavity of the dual-band filter. Figure 21.44 shows a comparison between this diplexer and a conventional diplexer employing a waveguide coupling T-junction and two standard single-band waveguide filters [8]. The two diplexers were designed to meet the same requirements. Waveguide cavities operating in TE_{102} modes were used in both diplexers to form a trisection in each channel. It can be seen that a diplexer that employs a dual-band filter offers a footprint reduction.

The same concept can be applied in multiplexer applications to reduce size and mass. For example, consider the traditional circulator-coupled multiplexer shown in Figure 21.45a, where conventional single-band filters are used. The size can be reduced by close to 40% if the dual-band filters are employed as shown in Figure 21.45b where two of the single-band filters can be replaced by one dual-band filter. The channels can be arranged in a way to maximize the frequency spacing between passbands when combined in a dual-band configuration.

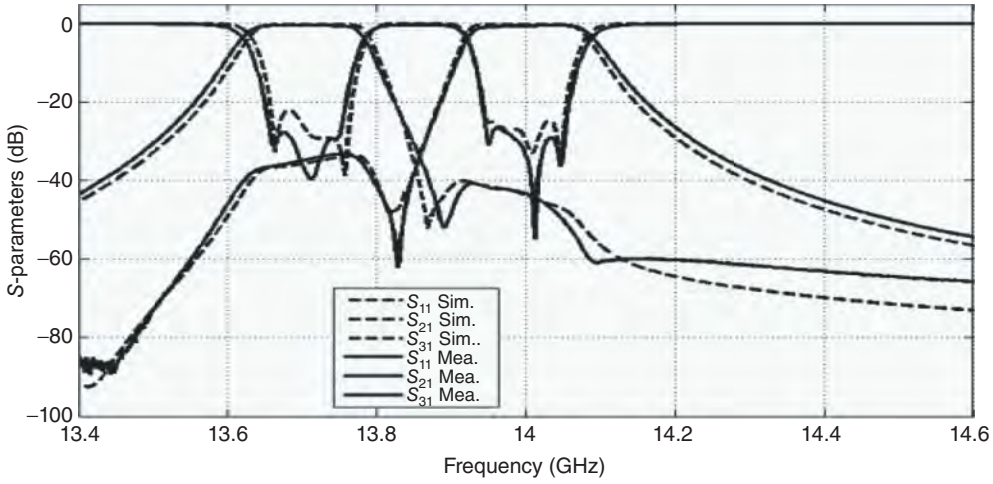
21.7 Synthesis of Multiband Filters

21.7.1 Introduction

In recent years several papers have introduced methods for the design of the prototype polynomials for symmetric and asymmetric multiband filtering functions [2, 3, 18–32]. In these papers it is noted that although sub-band bandwidths and TZ positions may be prescribed, the equiripple return loss (RL) levels within these bands are not necessarily predetermined. With fixed TZ positions, one has to "take what comes" in terms of RL and rejection lobe levels, where the designer would prefer to be able to prescribe these levels in order to address critical system specifications. The aim of this section is to build upon the previous work that has been done to



(a)



(b)

Figure 21.43 (a) A picture of the manufactured diplexer. (b) A comparison between simulation and measured results. *Source*: Zhu *et al*. 2017 [8]. Reproduced with the permission of IEEE.

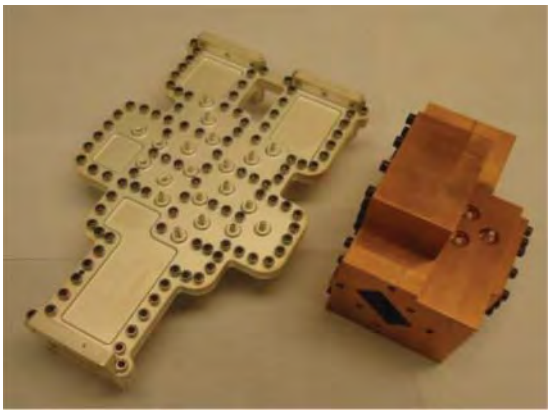


Figure 21.44 A size comparison between a diplexer employing dual-band waveguide filter and a conventional diplexer employing two single-band waveguide filters. *Source*: Zhu *et al*. 2017 [8]. Reproduced with the permission of IEEE.

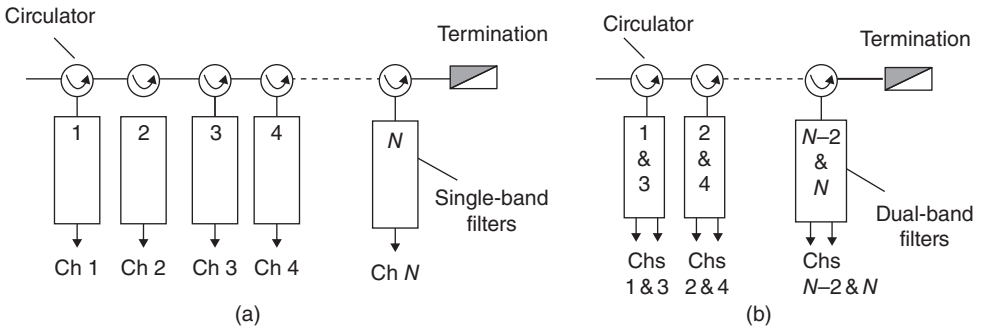


Figure 21.45 A comparison between a multiplexer employing single-band filters and a multiplexer employing dual-band filters.

formalize the design process to a certain extent and enable the prescription of in-band RL and rejection lobe levels.

21.7.2 Multi-passband Characteristics: Prototype Design

In one of the more recent papers on multiband filtering function design [22], Zhang *et al*. introduced an effective algorithm for the generation of the polynomials $E(s)$, $F(s)$, and $P(s)$ for the transfer and reflection functions $S_{21}(s)$ and $S_{11}(s)$ for multiband filters with equiripple passbands and prescribed TZ positions:

$$S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)}, \quad S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)} \tag{21.1}$$

The method is valid for any number of separate passbands (sub-bands) within the outer edges of the included bandwidth ($s = \pm j$), and the inner band-edges of each sub-band may also be prescribed. The method for the generation of the prototype polynomials is well explained in Ref. [22] and may be briefly summarized as follows:

1) Specify the number of sub-bands and their band-edge frequencies. The lower band-edge of the lowest frequency band is set at $s = -j$, while the upper band-edge of the highest frequency band is set at $s = +j$. Specify the number of reflection zeros within each band. The total number of reflection zeros over all the sub-bands gives the degree of the filter.
2) Specify the number of TZs and their positions to form the polynomial $P(s)$.
3) Set initial frequency positions for the reflection zeros within each band and form the polynomial $F(s)$.
4) Differentiate the characteristic function $C(s) = F(s)/P(s)$ and root it to find its turning points.
5) Evaluate $C(s)$ at the in-band turning point frequencies c_i and the band-edge frequencies for each sub-band, which will alternate with the zeros f_i of $F(s)$ (see Figure 21.46). The turning point and band-edge values of $C(s)$ will alternate between plus and minus values (see Figure 21.46).
6) With these values for $C(s)$ at the turning points c_i, apply the updating formula given in Ref. [22] to move f_i to new positions, reform the polynomial $F(s)$, and with the new $F(s)$ and the original $P(s)$ recalculate the new turning points of $C(s)$. The differences in the magnitudes of $C(s)$ over each individual band at the new turning points will be reduced. Steps 4–6 are repeated until the differences in the absolute values of $C(s)$ evaluated at the in-band turning points c_i and the edge-of-band frequencies of each sub-band are close to zero within a prescribed error.

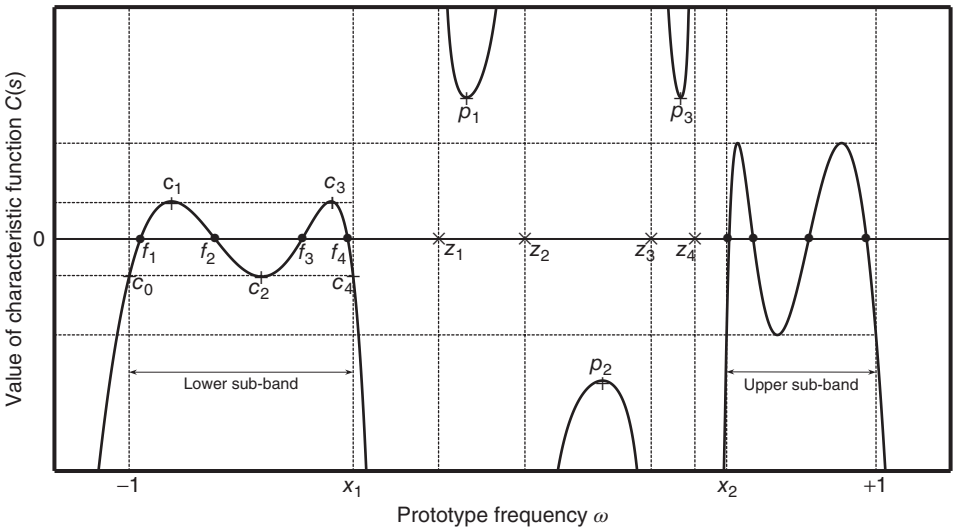


Figure 21.46 Asymmetric 8-4 dual-band example with internal band-edges at frequencies x_1 (=−0.4) and x_2 (=+0.6)—typical form of characteristic function $C(s)$ showing the critical frequencies in the lower band. Note that in this case the equiripple levels within the two bands have been set to be unequal.

As an example, the typical form of the characteristic function $C(s)$ for an asymmetric eighth-degree dual-passband filter with four TZs between the bands is given in Figure 21.46. f_i are the frequencies of the RL zeros, and c_j are the turning points and band-edge frequencies in the lower band, while p_k are the out-of-band rejection turning point frequencies. z_l are the positions of the prescribed TZs, and x_1 (=−0.4) and x_2 (=+0.6) are the inner band-edge frequencies of the two bands.

The method described above will allow for the prescription of the number of sub-bands and their internal band-edge frequencies and return loss levels. The lower band-edge of the lowest frequency band and the upper band-edge of the highest frequency band are fixed at $\omega = -1$ and +1, respectively. Also the out-of-band and between-band TZ frequencies and the number of reflection zeros within each sub-band may be user specified.

However the method will only allow for the equiripple RL level within one of the bands to be prescribed (here it is assumed that the RL in the uppermost band is prescribed). Except in the symmetric dual-band case, the equiripple level of the other bands depends on the numbers and positions of the TZs, the numbers of reflection zeros that have been allocated to each sub-band, and the bandwidths of the sub-bands that have been prescribed, as well as the equiripple level that has been set for the upper band. This will mean that for multiband filters with asymmetric characteristics (e.g., different sub-band bandwidths, asymmetric rejection lobe levels, etc.), the RL response within each of the sub-bands, although equiripple, will not necessarily be at levels equal to each other or to the equiripple level prescribed for the upper sub-band.

The situation is demonstrated in Figure 21.47, which shows the rejection and RL characteristics of an asymmetric eighth-degree dual-band filter, with five reflection zeros in the lower band and three in the upper band and no TZs. The band-edge frequencies for the lower band are set at $\omega = -1$ and −0.1, while for the upper band at $\omega = +0.8$ and +1. It may be seen from Figure 21.47 that after the application of the algorithm to create the equiripple bands, the RL level in the upper band is at the prescribed 22 dB, but the equiripple level in the lower band has come out at 16.7 dB. For other values of sub-band bandwidths and prescribed upper-band RL levels, the RL level that emerges for the lower sub-band may be impractically low or unrealistically high.

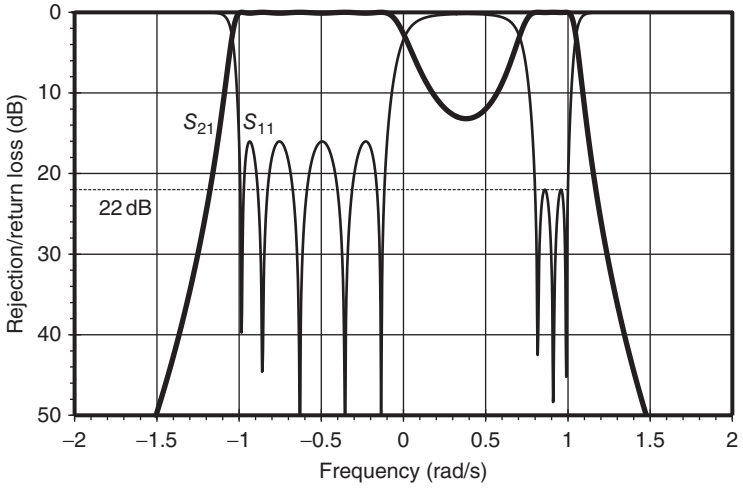


Figure 21.47 8-0 Asymmetric dual-band filter with no transmission zeros.

If it is required to prescribe the equiripple level in the lower band as well as the upper band, then it is necessary to assign at least one of the filter parameters to become an optimization variable and allow its value to be varied by an optimization routine with a prescribed equiripple level in the lower band as its target. One method of doing this is to allow the band-edge frequencies to vary (including the outer band-edges) [23], but in general this is not preferred because in practice the bandwidths of the sub-bands need to be fixed in order to meet system in-band and rejection specifications.

A more convenient method is to introduce a TZ into the filtering function and optimize its position to allow the second band equiripple level to be prescribed as well as the first. The effect of doing this is illustrated in Figure 21.48, where a single TZ has been introduced between the two bands of the 8-0 example of Figure 21.47. The position of the TZ has been optimized to

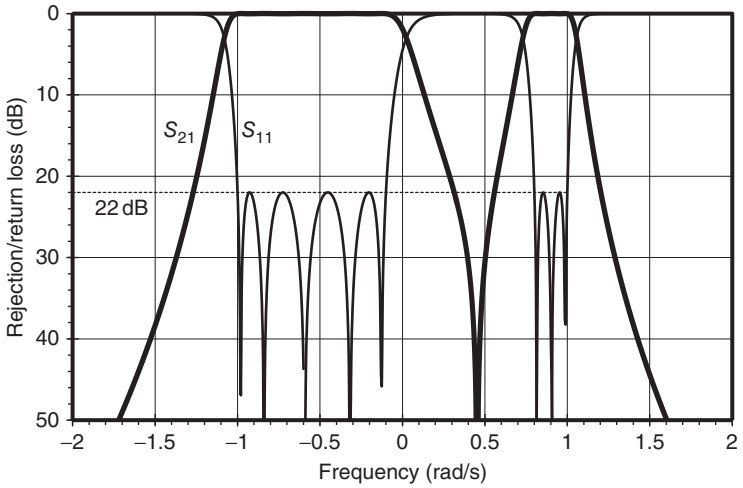


Figure 21.48 8-1 Asymmetric dual-band filter with one transmission zero. The equiripple level in the lower band has been set to be the same as for the upper band (22 dB).

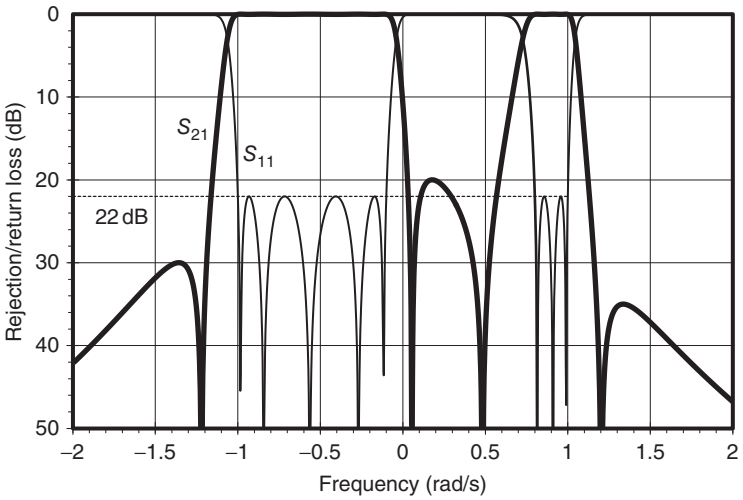


Figure 21.49 8-4 Asymmetric dual-band filter with four transmission zeros. The equiripple level in the lower band has been set to be the same as for the upper band (22 dB).

achieve an equiripple level of 22 dB for the second band, which in this case is the same level as the first band but does not necessarily have to be so.

Placing one TZ between the two bands like this does not produce a rejection lobe, and its position may be used as the optimization variable to prescribe the equiripple level of the lower band. Adding another TZ between the bands will produce a single rejection lobe, and now the two optimization variables available may be used to optimize the equiripple level and the rejection lobe level simultaneously to prescribed levels. Likewise any TZs placed on the outer sides of the two bands will also produce one rejection lobe each, which may be optimized to prescribed levels together with the interband TZs and the equiripple levels.

Figure 21.49 shows the eighth-degree case used above with two TZs between the bands producing a rejection lobe and two further TZs on either side of the two bands, each producing a rejection lobe. Now there are four TZs in total whose positions are available as optimization variables and four prescribed levels to be optimized: one in-band equiripple level in the lower band (set here at 22 dB, the same as in the upper band) and three lobe levels (set in this example at 30, 20, and 35 dB, respectively). Since the positions of the lobes come out as a by-product of the original process to create the equiripple in-band performance (the *out-of-band* turning point frequencies p_i; see the example in Figure 21.46), it is convenient to calculate the rejection levels at these points and include them as part of the overall optimization process.

This approach for enabling the prescription of different equiripple levels within each sub-band and asymmetric rejection lobe levels may be extended for any number of bands and lobes, provided a simple rule is observed:

- There must be at least one TZ to be used as an optimization variable between each sub-band of the multiband filter.

Of course the total number of TZs n_{fz} should not exceed N, the degree of the overall filter. Figure 21.50 shows the case of a 12th-degree filter with seven TZs and three bands each with four RL zeros. The seven available optimization parameters have been used to set seven levels:

- The RL levels in the two lower bands at 22 dB each, the same as the uppermost band
- A single TZ on the lower side producing a rejection lobe of level 30 dB

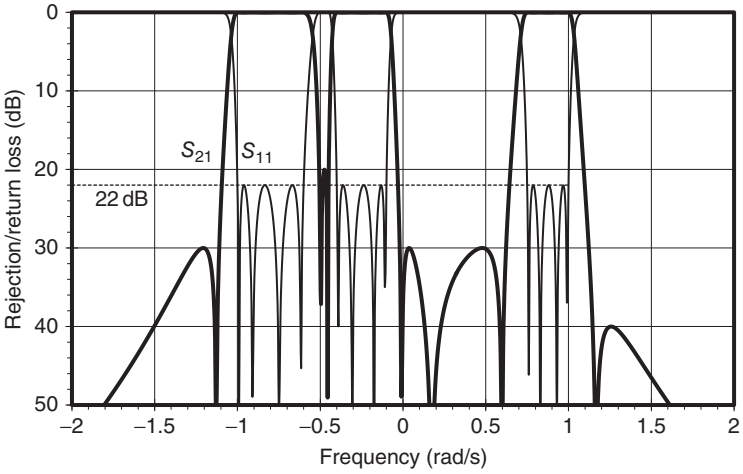


Figure 21.50 12-7 Asymmetric triple-band filter with seven transmission zeros. The equiripple level in the lower two bands has been set to be the same as for the upper band (22 dB).

- Two TZs between the two lower bands producing a single rejection lobe of level 20 dB
- Three TZs between the two upper bands producing two rejection lobes of level 30 dB each
- A single TZ on the upper side producing a rejection lobe of level 40 dB

The RL in the upper sub-band has already been set at 22 dB as part of the original algorithm for producing the equiripple RL levels within the bands.

21.7.3 Limitations and Restrictions

For a noncanonical single-band bandpass filter, each finite-position TZ that is built into the filtering function will produce one rejection lobe. To set the rejection levels of those lobes against prescribed target levels, the TZ frequency positions may be adjusted to be almost anywhere from frequencies close to the passband edges up to $s = \pm j\infty$. This freedom allows for the existence of an almost infinite range of possibilities for prescribing rejection lobe levels for single-band filters.

Although the out-of-band TZs of a multiband filter have the same freedom to range from close to band edge to infinite frequency, the interband zeros are restricted to the frequency bands *between* the passbands. This restriction on movement will exclude certain combinations of target specifications—sub-band bandwidths, RL levels, numbers of reflection zeros within each band, and rejection lobe levels. This means that the optimizer will fail for these combinations, as it drives the TZ positions into passband regions in an attempt to achieve the targets, and a solution will not be found. At present there is no known method to determine *a priori* which combinations have solutions or which do not, but experience has shown that in most realistic cases, a solution is usually forthcoming after some adjustment to bandwidths, interband rejection, and/or RL levels and increasing or decreasing the numbers of RL zeros within the sub-bands or TZs between the bands. Some recent work to address this problem and giving a method to determine optimal solutions may be found in Ref. [30].

Although it is possible to introduce complex out-of-band zeros to increase the numbers of optimization variables, this is generally not preferred because complex zeros do not contribute as strongly to the rejection performance as imaginary axis zeros do, but cross-couplings will still have to be provided to realize them in the filter structure.

21.7.4 Coupling Matrix Synthesis

Once the $E(s)$, $F(s)$, and $P(s)$ polynomials and the constants ε and ε_R have been determined, coupling matrix synthesis and reconfiguration of the multiband filter proceeds in exactly the same way as for a conventional single-band filter of the same degree. To demonstrate the process, the coupling matrix for the asymmetric 8-4 dual-band filter with the response as shown in Figure 21.49 is used as an example. The lower passband of this example was set to extend between $\omega = -1$ and -0.1, while the upper passband between $\omega = +0.8$ and $+1$. The target rejection levels are 30, 25, and 35 dB, respectively, and the target RL level for the lower band is 22 dB, the same as for the upper band.

With initial values for the four TZs, the transmission numerator polynomial $P(s)$ may be formed, and using the algorithm outlined above, the dual-band reflection numerator polynomial $F(s)$ may be found. Having prescribed the RL level for the upper band at 22 dB, the common denominator polynomial $E(s)$ may be then be formed using, for example, the alternating pole method of Section 6.2. Using the out-of-band rejection turning point frequencies, the levels of the rejection lobes may be computed and, together with the RL level evaluated at one of the edges of the lower band and the in-band target levels, a cost function may be calculated. Now the whole process is repeated with adjusted positions for the TZs until the cost function is arbitrarily close to zero. The final values for the polynomials are given in Table 21.3.

Using the transverse matrix method (see Section 8.4 and [26]), the $N + 2$-folded coupling matrix may now be synthesized. Here the matrix has been reconfigured as two cascaded quartets using the methods of Section 10.3, as shown in Figure 21.51. The first quartet 1234 realizes the outer TZ pair (#1 and #4 in the coupling matrix shown below), while the second quartet 5678 realizes the inner pair (#2 and #3 in Table 21.3). This combination seems to give the smallest values for the diagonal cross-couplings within the quartets. The coupling values and frequency offsets all seem to be well within realizable ranges, with the input/output couplings a little smaller than what is usual for single-band filters.

21.7.5 Dual-Band Filters with Transmission Zeros at Zero Frequency

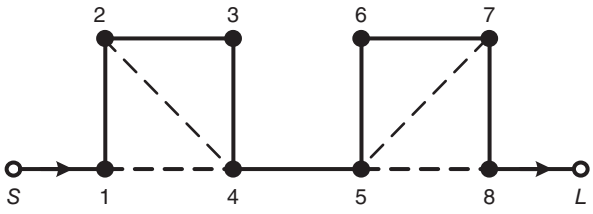
In 2005, Macchiarella and Tamiazzo demonstrated that a simplified configuration possibility arises with symmetric dual-band filters, where one or more of the zeros are at zero frequency [2]. An example is a sixth-degree symmetric dual-band characteristic where the lower band lies between $\omega = -1.0$ and -0.5 and the upper band lies between $\omega = +0.5$ and $+1.0$. There are

Table 21.3 Optimized zero positions for the $E(s)$, $F(s)$, and $P(s)$ polynomials for the 8-4 dual-band filtering function.

	Zeros of $P(s)$	Zeros of $F(s)$	Zeros of $E(s)$	
1	$-j1.2203$	$-j0.9836$	-0.0604	$-j1.0544$
2	$j0.0558$	$-j0.8430$	-0.2301	$-j0.9319$
3	$j0.4832$	$-j0.5637$	-0.3667	$-j0.5795$
4	$j1.2030$	$-j0.2689$	-0.2648	$-j0.1673$
5		$-j0.1168$	-0.0561	$-j0.0427$
6		$j0.8155$	-0.0767	$j0.7503$
7		$j0.9100$	-0.1330	$j0.9287$
8		$j0.9893$	-0.0446	$j1.0354$
	$\varepsilon = 12.1182$	$\varepsilon_R = 1.0$		

	S	1	2	3	4	5	6	7	8	L
S	0	0.7850	0	0	0	0	0	0	0	0
1	0.7850	0.2516	0.6857	0	–0.2453	0	0	0	0	0
2	0	0.6857	–0.2194	0.7399	–0.0206	0	0	0	0	0
3	0	0	0.7399	0.2943	0.4057	0	0	0	0	0
4	0	–0.2453	–0.0206	0.4057	–0.1568	0.6922	0	0	0	0
5	0	0	0	0	0.6922	0.0572	0.2529	0.0785	0.3943	0
6	0	0	0	0	0	0.2529	–0.0105	0.3267	0	0
7	0	0	0	0	0	0.0785	0.3267	–0.4067	0.6123	0
8	0	0	0	0	0	0.3943	0	0.6123	0.2516	0.7850
L	0	0	0	0	0	0	0	0	0.7850	0

(a)



(b)

Figure 21.51 Asymmetric 8-4 dual-band filter: (a) coupling matrix and (b) coupling/routing diagram.

three TZs between the two bands, one at zero frequency and two at finite frequencies, together producing two rejection lobe levels of 30 dB between the two bands. The in-band RL levels in the two bands have been set to 22 dB—see Figure 21.52a.

Again using the methods of Section 10.3, the network may be synthesized as a series of three trisections, the first two realizing the two finite frequency zeros, while the third realizes the TZ at zero frequency (Figure 21.52b). At this point the trisection 5-6-*L* realizing the TZ at zero frequency will automatically have a value of zero for one of its mainline couplings, M_{6L} in this case.

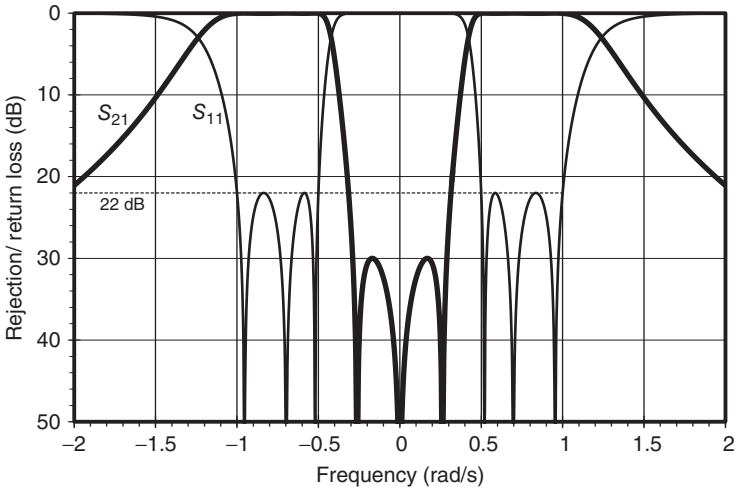
Rotations may now be applied to merge the trisections 1-2-3 and 3-4-5, which realize the two symmetric finite frequency zeros, to form a symmetric quartet 1-2-3-4. Figure 21.52c gives the coupling and routing diagram showing the quartet and the "extracted zero" resonators 5 and 6 on the right-hand side of the network, which could be conveniently realized with a dual-mode resonator. The corresponding coupling matrix is given in Figure 21.52d

Figure 21.53a shows a symmetric 10th-degree example with two pairs of zeros and one at zero frequency. Figure 21.53b gives one possible topology where the extracted zero pair is between the quartets, realizing the TZ pairs, but there is flexibility to place it elsewhere if more convenient.

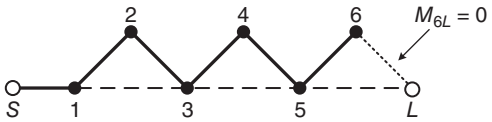
21.7.5.1 Multiple Transmission Zeros at Zero Frequency

This simplified topology also applies to symmetric dual-band characteristics with multiple TZs at zero frequency [2]. The limit is reached with an Nth-degree characteristic (N even) with $N/2$ TZs at zero frequency, whose characteristic is sometimes needed to produce a large rejection level between two adjacent bands. For these cases the whole network will comprise a series of $N/2$-coupled extracted zero pairs.

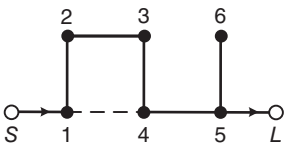
An example is taken of an eighth-degree symmetric dual-band design with four TZs at $\omega = 0$. The equiripple lower band has been set to be between $\omega = -1.0$ and -0.6, while the upper band



(a)



(b)



(c)

	S	1	2	3	4	5	6	L
S	0	0.8044	0	0	0	0	0	0
1	0.8044	0	0.6389	0	0.6387	0	0	0
2	0	0.6389	0	–0.0906	0	0	0	0
3	0	0	–0.0906	0	0.6673	0	0	0
4	0	0.6387	0	0.6673	0	0.5591	0	0
5	0	0	0	0	0.5591	0	0.7152	0.8044
6	0	0	0	0	0	0.7152	0	0
L	0	0	0	0	0	0.8044	0	0

(d)

Figure 21.52 Symmetric 6-3 dual-band filter with one transmission zero at zero frequency: (a) transfer and return loss characteristics, (b) initial circuit synthesis with trisections,(c) after reconfiguration into a quartet and an "extracted zero" pair, and (d) corresponding coupling matrix.

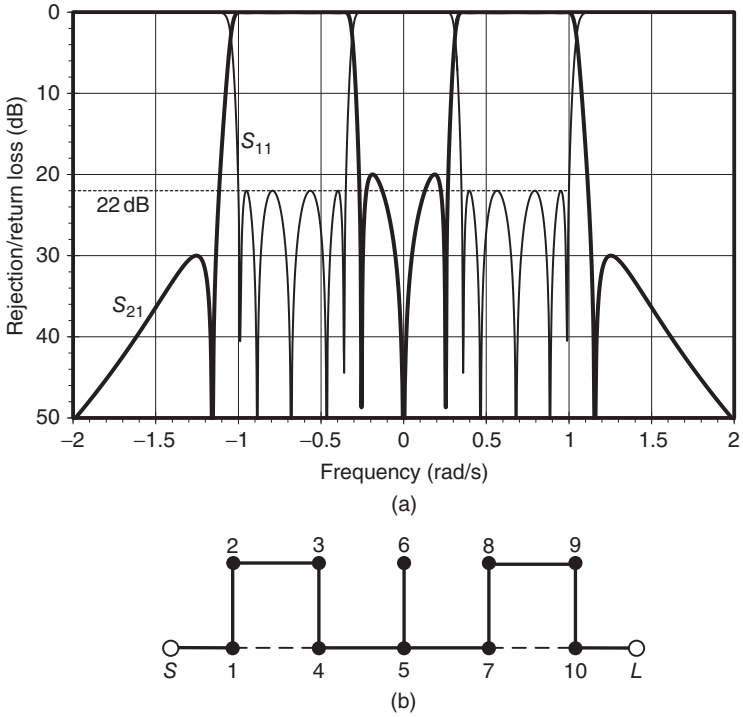


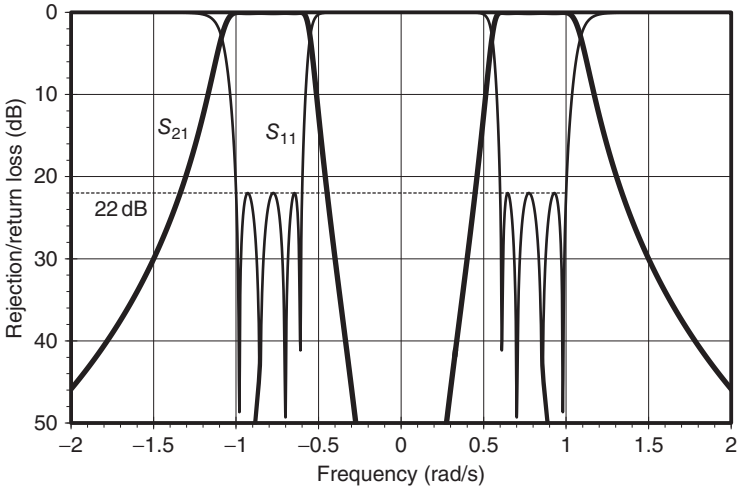
Figure 21.53 Symmetric 10-5 dual-band filter with one transmission zero at zero frequency and two pairs of zeros: (a) transfer and return loss characteristics and (b) coupling and routing diagram showing two quartets and an "extracted zero" pair.

between $\omega = +0.6$ and $+1.0$. RL in both bands is 22 dB. The transfer and rejection characteristics are shown in Figure 21.54a.

Following the same procedure used to synthesize the coupling matrix for the 8-4 example in Section 10.3, four cascaded trisections are synthesized, each producing a TZ at zero frequency. The coupling matrix and topology for this example is shown in Figure 21.54b and c, where it may be seen that each of the trisections automatically has one mainline coupling of zero value: M_{23}, M_{45}, M_{67}, and M_{8L} in this case. Another feature is that the couplings to the pole resonators M_{12}, M_{34}, M_{56}, and M_{78} all have the same value. Again it may be seen that this topology is convenient for realization with a dual-mode resonator technology.

With this class of configuration, it is possible to offset the resonant frequencies of all the pole resonators (resonators 2, 4, 6, and 8 in the above example) by the same amount without damaging the equiripple RL performance. This will change the relative widths of the two passbands, creating an asymmetric appearance to the transfer characteristics, but with the same topology as with the symmetric original.

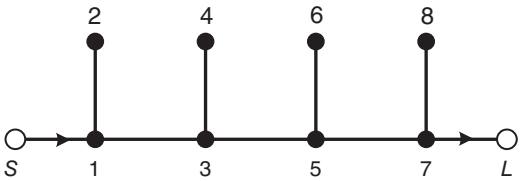
To guarantee the specified RF bandwidths, a method presented in [2] may be used to determine the position of the TZs. Here the specifications for the band edges of the two bands in the RF domain are used to calculate the positions of the $N/2$ TZs and the band edges of the lowpass prototype. Adopting the notation of [2] where ω_{L1} and ω_{H1} are, respectively, the lower and upper frequencies of the lower passband (passband 1) of the dual-band filter in the RF domain, ω_{L2} and ω_{H2} are the lower and upper frequencies of the upper passband (passband 2), the resonant frequencies of the mainline and pole resonators ω_{01} and ω_{02}, respectively, and two



(a)

	S	1	2	3	4	5	6	7	8	L
S	0	0.6844	0	0	0	0	0	0	0	0
1	0.6844	0	0.7746	0.3840	0	0	0	0	0	0
2	0	0.7746	0	0	0	0	0	0	0	0
3	0	0.3840	0	0	0.7746	0.2907	0	0	0	0
4	0	0	0	0.7746	0	0	0	0	0	0
5	0	0	0	0.2907	0	0	0.7746	0.3840	0	0
6	0	0	0	0	0	0.7746	0	0	0	0
7	0	0	0	0	0	0.3840	0	0	0.7746	0.6844
8	0	0	0	0	0	0	0	0.7746	0	0
L	0	0	0	0	0	0	0	0.6844	0	0

(b)



(c)

Figure 21.54 Symmetric 8-4 dual-band filter with four transmission zeros at zero frequency: (a) transfer and return loss characteristics, (b) synthesized coupling matrix, and (c) corresponding coupling/routing diagram.

bandwidth scaling factors, b_1 and b_2, may be calculated:

$$\omega_{01} = \sqrt{\frac{n_0 n_3}{n_1}}, \quad \omega_{02} = \sqrt{\frac{n_1}{n_3}} \tag{21.2a}$$

$$b_1 = \frac{\omega_{01}}{n_3}, \quad b_2 = \frac{-n_3\omega_{02}}{n_2 + \omega_{01}^2 + \omega_{02}^2} \tag{21.2b}$$

where

$$n_0 = \omega_{L1}\omega_{H1}\omega_{L2}\omega_{H2}$$
$$n_1 = \omega_{H1}\omega_{H2}(\omega_{L1} + \omega_{L2}) - \omega_{L1}\omega_{L2}(\omega_{H1} + \omega_{H2})$$

$$n_2 = \omega_{H1}\omega_{H2} + \omega_{L1}\omega_{L2} - (\omega_{L1} + \omega_{L2})(\omega_{H1} + \omega_{H2})$$
$$n_3 = (\omega_{H1} - \omega_{L1}) + (\omega_{H2} - \omega_{L2})$$

The corresponding prototype frequencies may be derived from the RF frequencies by using the bandpass-to-lowpass transformation [see equation (8.1)]:

$$\Omega = \text{BWR}\left(\frac{\omega}{\omega_{0B}} - \frac{\omega_{0B}}{\omega}\right) \tag{21.3}$$

where $\omega_{0B} = \sqrt{\omega_{L1}\omega_{H2}}$, the bandwidth ratio $\text{BWR} = \omega_{0B}/(\omega_{H2} - \omega_{L1})$, ω is the RF frequency, and Ω is the corresponding lowpass prototype frequency. The scaling factors may be transformed to the lowpass prototype equivalents by dividing by the bandwidth ratio:

$$b_1' = b_1/\text{BWR}, \quad b_2' = b_2/\text{BWR}$$

An example is taken of the RF frequency plan for the dual-band filter used to demonstrate "design method B" in Ref. [2]:

- Equiripple passband 1: $\omega_{L1} = 840$ MHz, $\omega_{H1} = 862$ MHz
- Equiripple passband 2: $\omega_{L2} = 888$ MHz, $\omega_{H2} = 899$ MHz.

For this case the lower passband is double the width of the upper band, 22 and 11 MHz, respectively. Applying equations (21.2) and (21.3) to this frequency plan yields $\Omega_{01} = -0.1553$, $\Omega_{02} = 0.3547$, $b_1' = 1.7785$ and $b_2' = 1.4695$. Ω_{01} and Ω_{02} are the prototype frequency offsets of the mainline and pole resonators, respectively. That is, in the dual-band coupling matrix, the self-couplings $M_{ii} = -\Omega_{01}$ for the mainline resonators and $M_{ii} = -\Omega_{02}$ for the pole resonators. For the mainline interresonator couplings, the prototype coupling values of an all-pole filtering characteristic of *half* the degree of the dual-band filter of degree N (N even) are scaled as follows (see Figure 21.54):

For the input and output couplings,

$$M_{S1} = M_{S1}'/\sqrt{b_1'}, \quad M_{N,L} = M_{N/2,L}'/\sqrt{b_1'} \tag{21.4a}$$

and for the mainline couplings,

$$M_{j,j+2} = M_{i,i+1}'/b_1' \tag{21.4b}$$

The couplings to the pole resonators all have the same value:

$$M_{k,k+1} = 1/\sqrt{b_1'b_2'} \tag{21.4c}$$

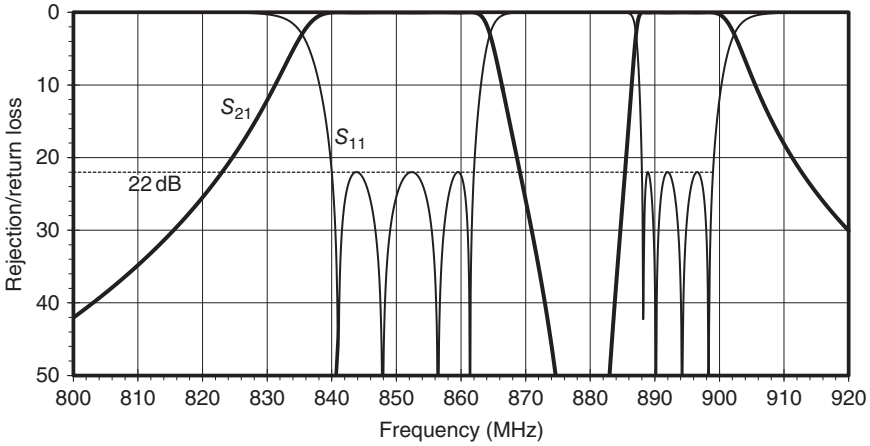
In equations (21.4) the dashed coupling values refer to the half-degree prototype.

For the eighth-degree 22 dB RL Chebyshev example used above, the $N/2$ = fourth-degree Chebyshev coupling values are $M_{S1}' = M_{4L}' = 1.0822$, $M_{12}' = M_{34}' = 0.9600$, and $M_{23}' = 0.7268$. Scaling these values according to equations (21.4a) and (21.4b) yields $M_{S1} = M_{8L} = 0.8115$, $M_{13} = M_{57} = 0.5398$, and $M_{35} = 0.4086$, and using equation (21.4c) to scale the pole resonator couplings yields $M_{12} = M_{34} = M_{56} = M_{78} = 0.6186$. The completed coupling matrix is shown in Figure 21.55a and plots of its RF performance in Figure 21.55b.

From the coupling matrix shown in Figure 21.55a, it may be seen that the simplified topology for this asymmetric dual-band filter has been retained—see Figure 21.54c. This feature should prove useful for meeting asymmetric rejection specifications while helping to ease the manufacturing and alignment processes. This bandwidth setting procedure needs only the frequency plan of the dual-band filter and may be used with any appropriate filter prototype.

	S	1	2	3	4	5	6	7	8	L
S	0	0.8115	0	0	0	0	0	0	0	0
1	0.8115	0.1553	0.6186	0.5398	0	0	0	0	0	0
2	0	0.6186	–0.3547	0	0	0	0	0	0	0
3	0	0.5398	0	0.1553	0.6186	0.4086	0	0	0	0
4	0	0	0	0.6186	–0.3547	0	0	0	0	0
5	0	0	0	0.4086	0	0.1553	0.6186	0.5398	0	0
6	0	0	0	0	0	0.6186	–0.3547	0	0	0
7	0	0	0	0	0	0.5398	0	0.1553	0.6186	0.8115
8	0	0	0	0	0	0	0	0.6186	–0.3547	0
L	0	0	0	0	0	0	0	0.8115	0	0

(a)



(b)

Figure 21.55 Asymmetric 8-4 dual-band filter with four transmission zeros: (a) synthesized coupling matrix and (b) RF transfer and return loss performance.

21.7.6 Synthesis of Multiband Coupled-Resonator Filters Using Reactance Transformations

The technique described in Section 21.7.5 may be extended for multibands using a reactance transformation technique due to Brand *et al.* [31, 32]. This nonoptimization method involves developing a frequency mapping function that enables the synthesis of compound resonators that replace the single resonators of an original single-band prototype that is synthesized separately. The prototype may be symmetric or asymmetric and have cross-couplings to realize TZs, and the multiband prototype will support an arbitrary number of sub-bands with individually prescribed bandwidths while retaining a rather simple coupling topology. The bands of the multiband filter become bandwidth-scaled and frequency-translated versions of the original prototype, which implies a restriction that different filtering degrees, RL, and rejection lobe levels, if any, cannot be individually prescribed for each sub-band.

The frequency mapping function takes the form of a rational polynomial equation where the degree of the numerator polynomial is of degree N_b and the denominator is degree $N_b - 1$, N_b being the number of sub-bands to be realized:

$$\omega'(\omega_i) = \frac{p_{Nb}\omega_i^{N_b} + p_{Nb-1}\omega_i^{N_b-1} + p_{Nb-2}\omega_i^{N_b-2} + \cdots + p_2\omega_i^2 + p_1\omega_i + p_0}{q_{Nb-1}\omega_i^{N_b-1} + q_{Nb-2}\omega_i^{N_b-2} + \cdots + q_2\omega_i^2 + q_1\omega_i + 1} = \frac{P(\omega)}{Q(\omega)} \qquad (21.5)$$

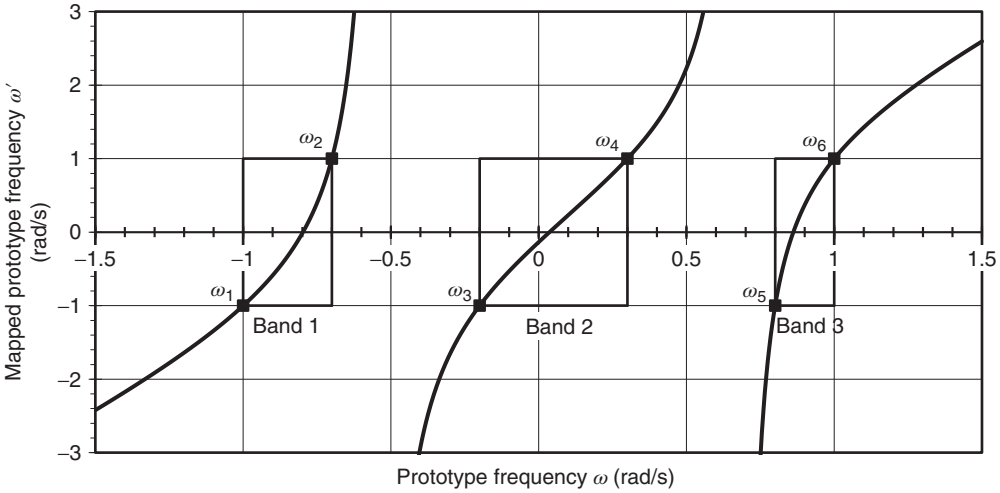


Figure 21.56 Mapping function for a triple-band example: band 1, $\omega = -1.0$ to -0.7; band 2, $\omega = -0.2$ to $+0.3$; band 3, $\omega = +0.8$ to $+1.0$.

where ω is the frequency in the multiband domain and $\omega'(\omega)$ is the corresponding frequency in the single-band domain. This equation has N_b zeros and $N_b - 1$ poles and is required to have poles and zeros that interleave along the ω axis, and $\omega'(\omega) = \pm 1$ when $\omega = \omega_1, \omega_2, \ldots, \omega_{2N_b}$; see Figure 21.56 for a three-band example. The mapping function acts so that when ω varies between -1 and $+1$, the mapped frequency $\omega'(\omega)$ alternates between -1 and $+1$ N_b times, reaching these values at the band edges of the sub-bands. The $N_b - 1$ poles of the mapping function ($\omega'(\omega) = \pm\infty$) occur between the sub-bands.

The mapping function is required to pass through the sub-band-edge frequencies with value -1.0 at the lower sub-band edge and $+1$ at the upper sub-band edge, as illustrated in Figure 21.56. Assuming that $\omega_1 = -1$ and $\omega_{2N_b} = +1$, it becomes possible to construct the following set of $2N_b$ coordinate pairs that must be located on the mapping function:

$$\{\omega_i, \omega'(\omega_i)\} = \{(-1,-1), (\omega_2, +1), (\omega_3, -1), \cdots, (\omega_{2N_b-1}, -1), (+1, +1)\} \tag{21.6}$$

To find the $2N_b$ unknown coefficients of polynomials $P(\omega)$ and $Q(\omega)$ in equation (21.5) that satisfy these conditions, we start by reorganizing equation (21.5):

$$P(\omega_i) - \omega'(\omega_i)Q(\omega_i) = 0, \quad i = 1,2,\ldots,2N_b \tag{21.7}$$

This set of $2N_b$ linear equations may be expressed in the matrix form $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A}$ is a $2N_b \times 2N_b$ square matrix, $\mathbf{X}$ is a column matrix containing the $2N_b$ unknown coefficients of polynomials $P(\omega)$ and $Q(\omega)$, and $\mathbf{B}$ is a column matrix with the $2N_b$ values of $\omega'(\omega_i)$ at the prescribed edges of the sub-bands where $\omega = \omega_1, \omega_2, \ldots, \omega_{2N_b}$:

$$\mathbf{A} = \begin{bmatrix} \overbrace{\omega_1^{Nb} \quad \omega_1^{Nb-1} \quad \cdots \quad \omega_1 \quad 1}^{\triangleleft \cdots P(\omega_i) \cdots \triangleright} & \overbrace{-\omega'(\omega_1)\omega_1^{Nb-1} \quad -\omega'(\omega_1)\omega_1^{Nb-2} \quad \cdots \quad -\omega'(\omega_1)\omega_1}^{\triangleleft \cdots -\omega'(\omega_i)Q(\omega_i) \cdots \triangleright} \\ \omega_2^{Nb} \quad \omega_2^{Nb-1} \quad \cdots \quad \omega_2 \quad 1 & -\omega'(\omega_2)\omega_2^{Nb-1} \quad -\omega'(\omega_2)\omega_2^{Nb-2} \quad \cdots \quad -\omega'(\omega_2)\omega_2 \\ \vdots \quad \vdots \quad \cdots \quad \vdots \quad \vdots & \vdots \quad \vdots \quad \cdots \quad \vdots \\ \omega_{2N_b}^{Nb} \quad \omega_{2N_b}^{Nb-1} \quad \cdots \quad \omega_{2N_b} \quad 1 & -\omega'(\omega_{2N_b})\omega_{2N_b}^{Nb-1} \quad -\omega'(\omega_{2N_b})\omega_{2N_b}^{Nb-2} \quad \cdots \quad -\omega'(\omega_{2N_b})\omega_{2N_b} \end{bmatrix}$$

$$\mathbf{X} = \text{transpose}\,[p_{Nb}, p_{Nb-1}, p_{Nb-2}, \cdots, p_1, p_0, q_{Nb-1}, q_{Nb-2}, \cdots, q_2, q_1]$$
$$\mathbf{B} = \text{transpose}\,[\omega'(\omega_1), \omega'(\omega_2), \omega'(\omega_3), \cdots, \omega'(\omega_{2N_b})] \tag{21.8}$$

Because the rank of matrix **A** and that of the augmented matrix **A**|**B** are both $2N_b$, a solution for the matrix equation exists, and arbitrary (real) values may be chosen for **B**. Choosing the $2N_b$ values of $\omega'(\omega_i)$, $i = 1, 2, \ldots, 2N_b$, at the sub-band edges [equation (21.6)] gives

$$\mathbf{B} = \text{transpose}\,[-1, +1, -1, \ldots, -1, +1]$$

Should the band-edge frequencies be specified at RF, bandpass-to-lowpass mapping functions such as equation (21.3) for lumped-element filters may be used.

Now the matrix equation may be solved for **X** to determine the values of the coefficients of polynomials $P(\omega)$ and $Q(\omega)$:

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B} \tag{21.9}$$

Analyzing the mapping function [equation (21.5)] with these values for the coefficients of $P(\omega)$ and $Q(\omega)$ will give a response such as that shown in Figure 21.56. With these coefficients, other prototype frequencies, for example, TZ or lobe-level positions, may also be mapped to the multiband domain by finding the N_b roots of equation (21.7).

21.7.6.1 Synthesis of the Compound Resonator Network

Because the mapping function is a ratio of two finite-degree polynomials, the denominator one degree less than the numerator, and whose poles and zeros are real and interlaced on the frequency (ω) axis, the mapping function $\omega'(\omega_i)$ is equivalent to an admittance function and may be synthesized as a one-port ladder network of capacitors, frequency-invariant reactances (FIRs), and inverters using the classical methods as outlined in Chapter 7.

Figure 21.57a shows the circuit for a typical "lowpass resonator" for a single-band filter. The resonator comprises a shunt capacitor C in parallel with an FIR, which may be positive, negative, or zero in value. In practice the FIR is realized as a frequency offset from the filter center frequency. The admittance of this single resonator is $Y(\omega) = sC + jB$. Figure 21.57b shows the one-port network for the compound resonator of a three-band filter. It comprises three lowpass resonators interconnected by two inverters. This network will replace each of the single resonators of the single-band filter to achieve the triple-band response.

To illustrate the procedure, an example of a triple-band characteristic will be used. Here band 1 lies between $\omega = -1.0$ and -0.7, band 2 between $\omega = -0.2$ and $+0.3$, and band 3 between $\omega = +0.8$ and $+1.0$ (see Figure 21.56). With these prescribed sub-band-edge frequencies, the

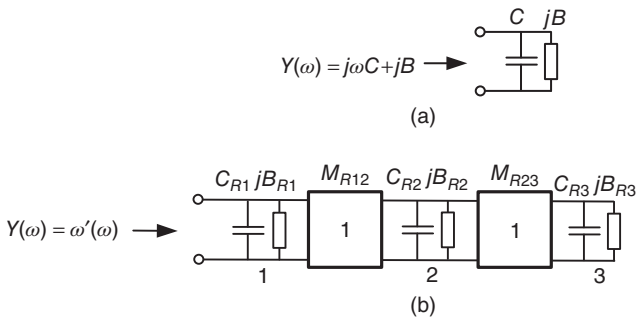


Figure 21.57 Single-band and multiband compound resonator prototypes. (a) Single-band "lowpass resonator" with frequency-invariant reactance (FIR). (b) Triple-band compound resonator.

6×6 matrix $\mathbf{A}$ may be constructed [equation (21.8)]:

$$\mathbf{A} = \begin{bmatrix} -1.0000 & 1.0000 & -1.0000 & 1.0000 & 1.0000 & -1.0000 \\ -0.3430 & 0.4900 & -0.7000 & 1.0000 & -0.4900 & 0.7000 \\ -0.0080 & 0.0400 & -0.2000 & 1.0000 & 0.0400 & -0.2000 \\ 0.0270 & 0.0900 & 0.3000 & 1.0000 & -0.0900 & -0.3000 \\ 0.5120 & 0.6400 & 0.8000 & 1.0000 & 0.6400 & 0.8000 \\ 1.0000 & 1.0000 & 1.0000 & 1.0000 & -1.0000 & -1.0000 \end{bmatrix}$$

$$\mathbf{B} = \text{transpose } [-1, +1, -1, +1, -1, +1]$$

The column matrix $\mathbf{X}$ may now be evaluated:

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B} = \text{transpose } [-5.4054, 0.5405, 3.7027, -0.1351, -2.7027, 0.4054]$$

Knowing the coefficients of $P(\omega)$ and $Q(\omega)$, the rational mapped admittance function $\omega'(\omega)$ may now be constructed as follows:

$$\omega'(\omega) = \frac{P(\omega)}{Q(\omega)} = \frac{-5.4054\omega^3 + 0.5405\omega^2 + 3.7027\omega - 0.1351}{-2.7027\omega^2 + 0.4054\omega + 1.0} \tag{21.10}$$

Using classical circuit synthesis techniques (Chapter 7), the values of the components of the one-port ladder network of Figure 21.57b may now be determined:

$$\begin{array}{lll} C_{R1} = 2.0000 & B_{R1} = 0.1000 & M_{R12} = 1.0 \\ C_{R2} = 1.6260 & B_{R2} = -0.0139 & M_{R23} = 1.0 \\ C_{R3} = 1.6566 & B_{R3} = -0.2344 & \end{array}$$

The compound resonators may be used to replace the single resonators of any prototype network, and for this example a symmetric 4-2 quasielliptic prototype with 22 dB RL and 20 dB rejection lobe levels has been selected. The mainline couplings values are $M_{S1} = M_{1L} = 1.0602$, $M_{12} = M_{34} = 0.8589$, and $M_{23} = 0.8527$, and the cross-coupling $M_{14} = -0.3519$. All the self-couplings M_{11}, M_{22}, M_{33}, and M_{44} are zero. The coupling and routing diagram is shown in Figure 21.58a, and the rejection and RL performance is given in Figure 21.58b.

To obtain the triple-band performance, each resonator of the single-band prototype will be replaced by the one-port compound resonator. This results in a $4 \times 3 = 12$th-degree network with a coupling and routing diagram as shown in Figure 21.59.

In order to construct the equivalent coupling matrix, all the capacitors at the nodes have to be scaled to unity. Firstly scaling the components of the compound resonator networks, $M_{R12} = 1/\sqrt{C_{R1}C_{R2}}$ and $M_{R23} = 1/\sqrt{C_{R2}C_{R3}}$. If the core network resonators that the compound resonators are to replace include an FIR B_i (as in Figure 21.57a), this is added to B_{R1} and scaled as $B_i + B_{R1} \rightarrow (B_i + B_{R1})/C_{R1}$.The rest of the FIRs in the compound resonator networks are scaled as $B_{R2} \rightarrow B_{R2}/C_{R2}$, $B_{R3} \rightarrow B_{R3}/C_{R3}$. For this example, the FIRs will appear in the coupling matrix in groups of 3: M_{11}, M_{22}, and M_{33} for the first resonator; M_{44}, M_{55}, and M_{66} for the second; and so on. Next the internal couplings (including cross-couplings) of the core network are divided by C_{R1} and the input/output couplings by $\sqrt{C_{R1}}$ to scale the capacitors of the core network back to unity. Finally the three compound resonator capacitors C_{R1}, C_{R2}, and C_{R3} are also set to unity. Using the node numbering scheme of Figure 21.59, the $N + 2$ coupling matrix may now be constructed. The nonzero values of the coupling elements are tabulated in Table 21.4, and the rejection and RL performances are given in Figure 21.60.

The original prototype had a pair of TZs at $\omega' = \pm 1.5858$, which through equation (21.7) map into $\omega = -1.1812$, -0.2914, and $+0.7797$ for the lower TZ and $\omega = -0.6680$, $+0.4165$, and

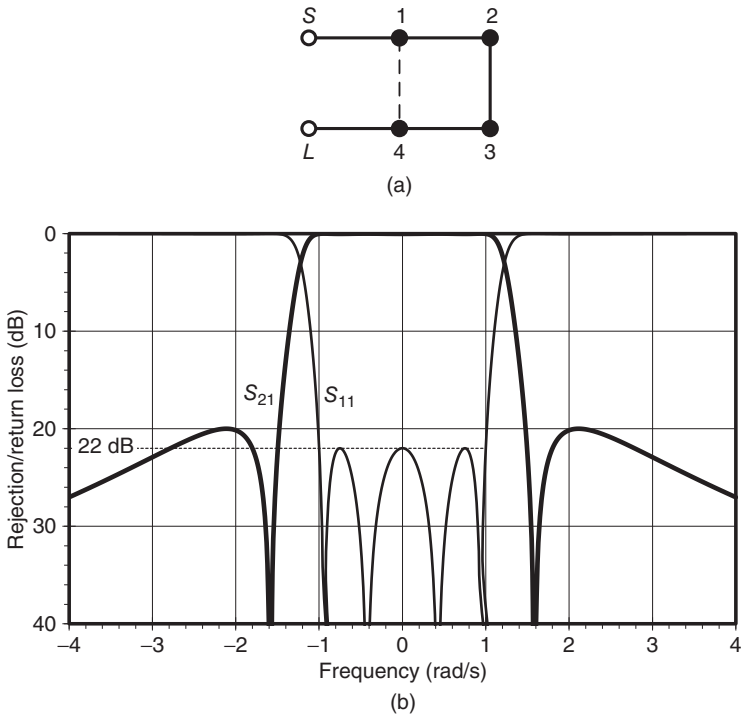


Figure 21.58 4-2 Symmetric quasielliptic single-band filter prototype: (a) folded network coupling and routing diagram and (b) rejection and return loss performance.

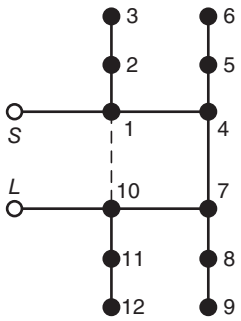


Figure 21.59 Coupling/routing diagram of a 4-2 quasielliptic triple-band filter. Each resonator of the single-band prototype is replaced with a triple-band compound resonator as shown in Figure 21.57b.

+1.1444 for the upper TZ. The interband reactance poles (the $N_b - 1$ roots of $Q(\omega)$) map to $\omega = -0.5379$ and $+0.6879$, corresponding to the single-band prototype frequencies $\omega' = \pm\infty$.

In the coupling matrix, it may be seen that all the compound resonators have the same coupling values and frequency offsets (self-couplings) and there are no negative couplings except for the one negative cross-coupling of the core prototype. In general, the design of a compound multiband resonator only requires the specification of the number of sub-bands and their band-edge frequencies. Once designed, the same compound resonator may be applied to any prototype, asymmetric or symmetric, to give bandwidth-scaled and frequency-translated versions of the original prototype within each sub-band.

21.7.6.2 Multiband Bandstop Filters

It was mentioned above that a multiband response may be created simply by replacing the single-band resonators of any prototype with the compound resonator. The sub-bands of the multiband response become frequency-translated and bandwidth-scaled versions of the original prototype. This also applies to direct-coupled bandstop prototypes, where the same simple coupling topology as for the bandpass types is retained. To demonstrate the possibility, an example is taken of a cul-de-sac bandstop filter, as introduced in Section 10.2.1.

Table 21.4 Coupling values for 4-2 triple-band filter (12th degree).

Core filter couplings		Compound resonator: self-couplings		Compound resonator: interresonator couplings	
$M_{S1}, M_{10,L}$	0.7497	$M_{11}, M_{44}, M_{77}, M_{10,10}$	0.0500	$M_{12}, M_{45}, M_{78}, M_{10,11}$	0.5545
$M_{14}, M_{7,10}$	0.4295	$M_{22}, M_{55}, M_{88}, M_{11,11}$	−0.0085	$M_{23}, M_{56}, M_{89}, M_{11,12}$	0.6093
M_{47}	0.4263	$M_{33}, M_{66}, M_{99}, M_{12,12}$	−0.1415		
Cross-coupling $M_{1,10}$	−0.1759				

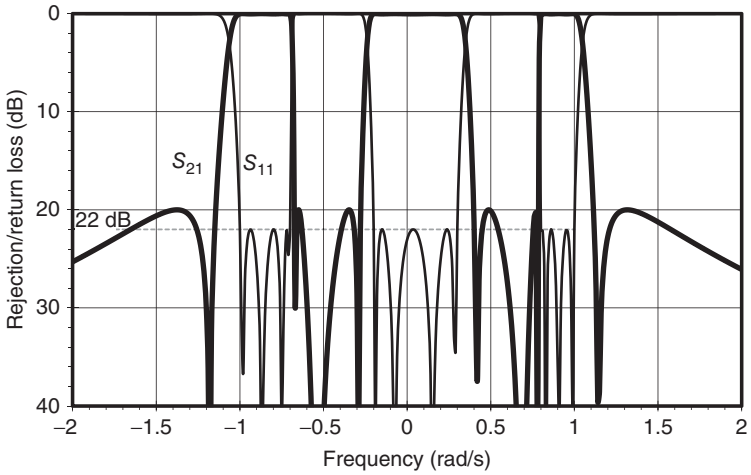


Figure 21.60 Rejection and return loss performance of 12th-degree triple-band prototype filter. Each sub-band is a frequency-mapped and bandwidth-scaled version of the original single-band 4-2 prototype.

In this case a 4-2 symmetric lowpass prototype has been chosen with a rejection lobe level of 20 dB and an RL level of 26 dB. In the process of designing the bandstop prototype, the transfer and RL characteristics exchange, and synthesizing the cul-de-sac network produces the following coupling values (see coupling and routing diagram in Figure 21.61a):

$$M_{S1} = M_{4L} = 1.6259 \qquad M_{SL} = 1.0 \qquad M_{11} \equiv B_1 = 0.3776$$
$$M_{12} = M_{34} = 0.9576 \qquad M_{14} = 1.3218 \qquad M_{22} \equiv B_2 = -0.9052$$
$$M_{23} = 0.0 \qquad\qquad M_{33} \equiv B_3 = 0.9052$$
$$M_{44} \equiv B_4 = -0.3776$$

The same prototype bandwidths as the triple-band bandpass filter example above will be used for the three stopbands, making the design of the compound resonator the same as before. Adding the compound resonators to the prototype bandstop network results in the 12th-degree network as shown in Figure 21.61b. After scaling, the coupling values of the elements in the coupling matrix for the compound resonators stay as in Table 21.4, but the values for the core network change as follows:

$$M_{S1} = M_{10,L} = 1.1497 \qquad M_{\mathrm{SL}} = 1.0 \qquad M_{11} = (B_1 + B_{R1})/C_{R1} = 0.2388$$
$$M_{14} = M_{7,10} = 0.4788 \qquad M_{1,10} = 0.6609 \qquad M_{44} = (B_2 + B_{R1})/C_{R1} = -0.4030$$
$$M_{47} = 0.0 \qquad\qquad M_{77} = (B_3 + B_{R1})/C_{R1} = 0.5030$$
$$M_{10,10} = (B_4 + B_{R1})/C_{R1} = -0.1388$$

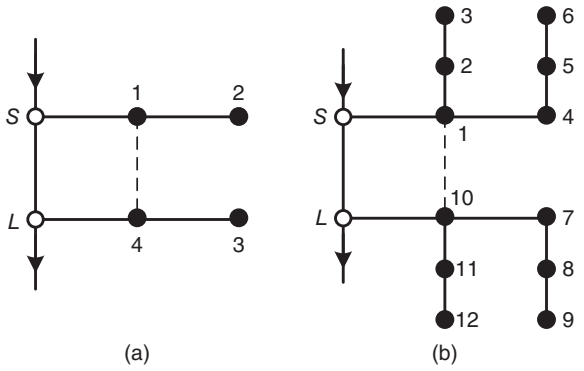


Figure 21.61 Direct-coupled triple-band bandstop filter—cul-de-sac configuration (a) prototype network (b) single resonators replaced with triple-band compound resonators.

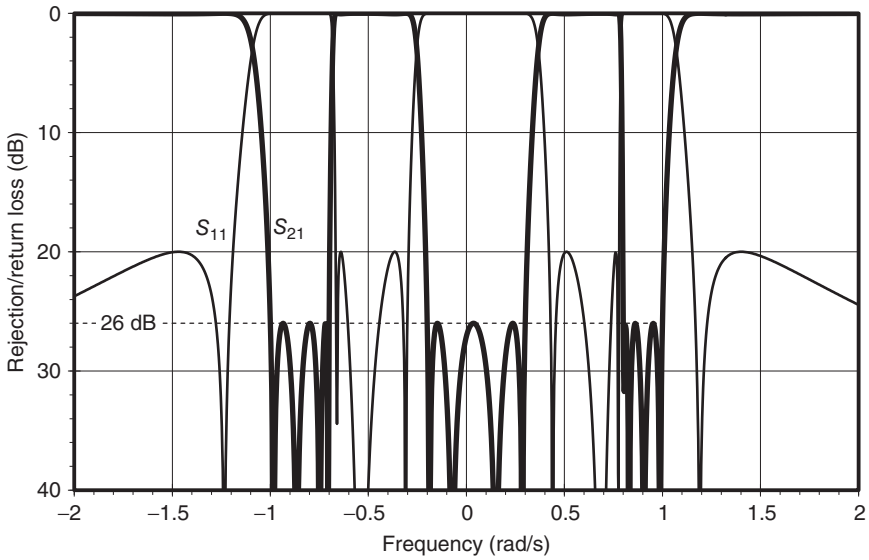


Figure 21.62 Triple-band bandstop filter—return loss and insertion loss performance.

Note that the direct source-load coupling M_{SL} is unchanged by the scaling process and remains equal to unity.

The results of analyzing this coupling matrix are shown in Figure 21.62, where it may be seen that the original prototype RL level of 26 dB has now become the equiripple insertion loss level within the three stopbands and the original 20 dB rejection lobe level has now become the RL characteristic outside the stopbands.

21.7.6.3 Alternative Configurations for the Compound Resonator

The compound resonator may be synthesized in "star" topology using the method of residues similar to that employed for the synthesis of the transverse matrix for two-port networks as described in Chapter 8. The star topology of an N_b—band compound resonator consists basically of the core resonator as synthesized for the "inline" configuration, with $N_b - 1$ one-port resonators connected to it. Figure 21.63 shows the topology for the basic three-, four-, and five-band compound resonators. Figure 21.64 shows the fourth-degree three-band prototype of Figure 21.61b where the second and third resonators (resonators [456] and [789]) have been

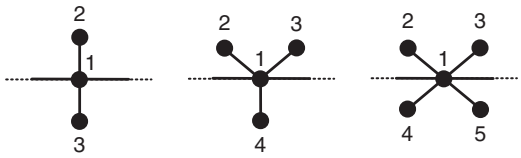


Figure 21.63 Star configurations for three-, four-, and five-band compound resonators.

configured in star topology. From the star configuration, other topologies may be derived using matrix similarity transforms.

The synthesis procedure for a star-configured N_b band compound resonator begins with the extraction of a capacitor/FIR pair from the rational admittance function exactly as for the inline configuration. Then the $N_b - 1$ one-port resonators are synthesized from the residues and poles of the remainder admittance function of degree $N_b - 1$, expressed as a partial fraction expansion.

Figure 21.65 shows the configuration of one branch of a star-configured compound resonator. An N_b band compound resonator will have $N_b - 1$ such one-port networks connected in parallel to the central core resonator.

The admittance of the ith one-port element, $i = 2, 3, \ldots, N_b$, of the star resonator is given by equation (21.11a):

$$Y_i(\omega) = -jY_i(s) = \frac{-j}{j\omega C_{Ri} + jB_{Ri}} = \frac{-1/C_{Ri}}{(\omega + B_{Ri}/C_{Ri})} \qquad (21.11a)$$

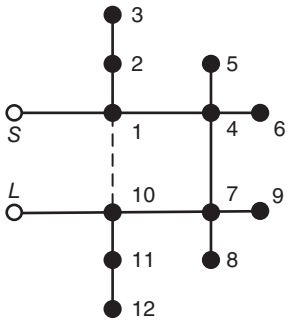


Figure 21.64 Triple-band 4-2 bandpass prototype. Compound resonators [456] and [789] are configured in star topology.

and the total admittance $Y'(\omega)$ of the $N_b - 1$ one-port elements connected in parallel to the first resonator is the sum of the individual one-port admittances:

$$Y'(\omega) = -\sum_{i=2}^{N_b} \frac{1/C_{Ri}}{(\omega + B_{Ri}/C_{Ri})} \qquad (21.11b)$$

Following the extraction of the capacitor/FIR pair for the first resonator of the compound resonator (the "core" resonator), the rational admittance polynomials of equation (21.10) reduce to

$$\omega'(\omega) = \frac{P(\omega)}{Q(\omega)} = \frac{p_{Nb-2}\omega^{N_b-2} + \cdots + p_2\omega^2 + p_1\omega + p_0}{q_{Nb-1}\omega^{N_b-1} + q_{Nb-2}\omega^{N_b-2} + \cdots + q_2\omega^2 + q_1\omega + 1} \qquad (21.12)$$

where the numerator polynomial has reduced in degree by 2. This rational polynomial equation may be expressed as a partial fraction expansion:

$$Y'(\omega) = \omega'(\omega) = \sum_{i=2}^{N_b} \frac{r_i}{(\omega - \omega_{0i})} \qquad (21.13)$$

where r_i are the $N_b - 1$ residues and ω_{0i} are the $N_b - 1$ zeros of the denominator polynomial $Q(\omega)$. The admittance equation (21.13) in partial fraction form may now be equated to the admittance equation (21.11b) to find the Cs and Bs of the outlying resonators of the star:

$$C_{Ri} = -1/r_i, \quad B_{Ri} = -\omega_{0i}C_{Ri}, \quad i = 2,3, \ldots, N_b \qquad (21.14)$$

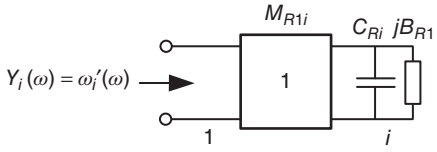


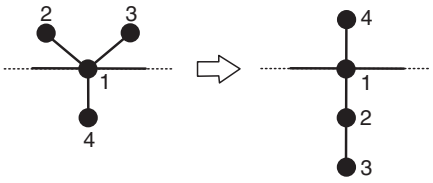
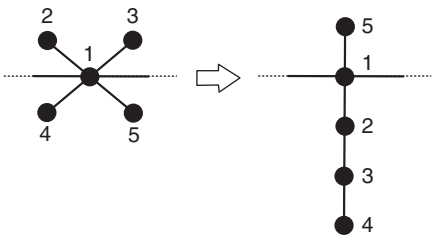
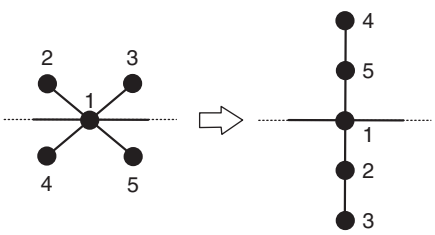
Figure 21.65 Single one-port element of the star-configured compound resonator.

Scaling between the core resonator and the outlying resonators of the star finally yields the values for each inverter coupling to the core resonator and its associated capacitor (=1.0) and FIR:

$$M_{R1i} = 1/\sqrt{C_{R1}C_{Ri}} = \sqrt{-r_i/C_{R1}}, \quad B_{Ri} \to B_{Ri}/C_{Ri} = -\omega_{0i}, \quad C_{Ri} \to 1, \quad i = 2,3,\ldots,N_b \tag{21.15}$$

To illustrate the procedure, the compound resonator that was synthesized for the three-band 4-2 bandpass prototype in Section 21.7.6.1 is used as an example. The rational admittance

Table 21.5 Alternative configurations: rotations for Four- and Five-band compound resonators.

Rotation number, *r*	Pivot [*i*, *j*]	Element to be annihilated	$\theta_r = \tan^{-1}(cM_{kl}/M_{mn})$				
			k	*l*	*m*	*n*	*c*
		Four band					
1	[2, 3]	M_{13}	1	3	1	2	−1
							
		Five band (1)					
1	[3, 4]	M_{14}	1	4	1	3	−1
2	[2, 3]	M_{13}	1	3	1	2	−1
3	[3, 4]	M_{24}	2	4	2	3	−1
							
		Five band (2)					
1	[4, 5]	M_{15}	1	5	1	4	−1
2	[2, 3]	M_{13}	1	3	1	2	−1
3	[4, 5]	—			90°		
							

function is given in equation (21.10), from which the first capacitor and FIR are first extracted: $C_{R1} = 2.0$, $B_{R1} = 0.1$. The remainder admittance function is given by

$$Y'(\omega) = \frac{1.6622\omega - 0.2351}{-2.7027\omega^2 + 0.4054\omega + 1.0} \tag{21.16a}$$

which may be expressed in partial fraction form:

$$Y'(\omega) = -\left[\frac{0.3408}{(\omega + 0.5379)} + \frac{0.2742}{(\omega - 0.6879)}\right] \tag{21.16b}$$

Equating equations (21.16b) and (21.11b) yields the values for the outlying resonators:

$$r_2 = -0.3408, \quad \omega_{02} = -0.5379, \quad r_3 = -0.2742, \quad \omega_{03} = 0.6879:$$

$$\therefore \quad C_{R2} = -1/r_2 = 2.9343, \quad B_{R2} = -\omega_{02} C_{R2} = 0.5379 C_{R2}$$

$$C_{R3} = -1/r_3 = 3.6470, \quad B_{R3} = -\omega_{03} C_{R3} = -0.6879 C_{R3}$$

Now scaling inverters to set C_{R2} and C_{R3} to unity,

$$M_{R12} = 1/\sqrt{C_{R1} C_{R2}} = 0.4128, \quad M_{R13} = 1/\sqrt{C_{R1} C_{R3}} = 0.3702$$

$$B_{R2} \to B_{R2}/C_{R2} = 0.5379, \quad B_{R3} \to B_{R3}/C_{R3} = -0.6879$$

$$C_{R2} \to 1.0, \quad C_{R3} \to 1.0$$

Once the coupling values for the star-configured compound resonators have been entered into the matrix for the overall multiband prototype filter, similarity transforms may be applied to produce further topologies for each compound resonator. Some examples are given in Table 21.5 for four- and five-band resonators, which may be mixed with other topologies for the other compound resonators of the overall multiband filter. The node numbering in Table 21.5 assumes that the first compound resonator in the network is being worked upon, but any or all of the resonators may be reconfigured by suitably indexing the resonator numbering.

A disadvantage with the star configuration or its derivatives is that the core resonator is more heavily loaded, with at least four branches instead of three as for the "inline" resonator. However it may be that layout problems are eased for certain technologies and that the loading problem becomes acceptable.

Summary

With today's ever-increasing demands on the available RF spectrum, the need for microwave bandpass filters with some form of in-band interference suppression has been growing, giving rise to the concept of 'multi-passband' filtering characteristics. Multiband filtering characteristics differ from the well-known single-band bandpass filtering function in that transmission zeros are built into the in-band regions, so splitting a single passband into two or more passbands separated by stopbands. Such filter characteristics are now finding application within radio-astronomy, military and wireless systems, suppressing narrowband jamming or interfering signals that might occur within the wider usable passband width.

This chapter presented four different techniques for realizing multiband filters illustrating the advantage and disadvantage of each technique and proving examples of high-Q multiband filters reported recently in literature. The chapter also provides a detailed synthesis procedure for multi-band filters.

References

1 Holme, S. (2002), Multiple passband filters for satellite applications. Proceedings of the 20th AIAA International Communications Satellite Systems Conference and Exhibit.
2 Macchiarella, G. and Tamiazzo, S. (2005) Design techniques for dual-passband filters. *IEEE Transactions on Microwave Theory and Techniques*, **53** (11), 3265–3271.
3 Lenoir, P., Bila, S., Seyfert, F. *et al.* (2006) Synthesis and design of asymmetrical dual-band bandpass filters based on equivalent network simplification. *IEEE Transactions on Microwave Theory and Techniques*, **54** (7), 3090–3097.
4 Fahmi, M., Ruiz-Cruz, J.A., Mansour, R.R., and Zaki, K.A. (2010) Compact wide-band ridge waveguide dual-band filter. IEEE International Microwave Symposium (IMS).
5 Fahmi, M., Ruiz-Cruz, J.A., and Mansour, R.R. (2016) Dual-band ridge waveguide filters for high selectivity wireless base station applications. *International Journal of RF and Microwave Computer-Aided Engineering*, **26** (8), 703–712.
6 Ruiz-Cruz, J., Fahmi, M.M., and Mansour, R.R. (2012) Triple-conductor combline resonators for dual-band filters with enhanced guard-band selectivity. *IEEE Transactions on Microwave Theory and Techniques*, **60** (12), 3969–3979.
7 Ruiz-Cruz, J.A., Fahmi, M.M., and Mansour, R.R. (2014) Method of operation and construction of filters and multiplexers multi-conductor multi-dielectric combline resonators. US Patent 20140347148, Nov.
8 Zhu, L., Mansour, R.R., and Yu, M. (2017) *Compact waveguide dual-band filter and diplexers*, vol. **65**, IEEE Trans. Microwave Theory Tech., pp. 1525–1533.
9 Zhu, L., Mansour, R.R., and Yu, M. (2016) Compact triple-band bandpass filter using rectangular waveguide resonators. IEEE International Microwave Symposium (IMS).
10 Zhang, R. and Mansour, R.R. (2009) Dual-band dielectric-resonator filters. *IEEE Transactions on Microwave Theory and Techniques*, **57** (7), 1760–1766.
11 Memarian, M. and Mansour, R.R. (2009) Quad-mode and dual-mode dielectric resonator filters. *IEEE Transactions on Microwave Theory and Techniques*, **57** (12), 3418–3426.
12 Memarian, M. and Mansour, R.R. (2009) Dual-band half-cut dielectric resonator filters. European Microwave Symposium (EuMC), pp. 555–558.
13 Memarian, M. and Mansour, R.R. (2012) Methods of operation and construction of dual-mode filters, dual-band filters and diplexers/multiplexers using half-cut and full dielectric resonators. US Patent 8,111,115, issued, Feb., 2012.
14 Zhu, L., Mansour, R.R., and Yu, M. (2017) Triple-band dielectric resonator bandpass filters. IEEE International Microwave Symposium (IMS), June 2017.
15 Macchiarella, G. and Tamiazzo, S. (2007) Dual-band filters for base station multi-band combiners. IEEE International Microwave Symposium (IMS).
16 Laforge, P. (2010) Tunable superconducting microwave filters. PhD thesis. University of Waterloo, Waterloo, Ontario, Canada.
17 Mansour, R.R. and Laforge, P. (2016) Multi-band superconductor filters. IEEE International Microwave Symposium (IMS).
18 Lee, J., Uhm, M.S., and Yom, I.-B. (2004) A dual-passband filter of canonical structure for satellite applications. *IEEE Microwave and Wireless Components Letters*, **14** (6), 271–273.
19 Di, H., Wu, B., Lai, X., and Liang, C. (2010) Synthesis of cross-coupled triple-passband filters based on frequency transformation. *IEEE Microwave and Wireless Components Letters*, **20** (8), 432–434.
20 Mokhtaari, M., Bornemann, J., Rambabu, K., and Amari, S. (2006) Coupling-matrix design of dual and triple passband filters. *IEEE Transactions on Microwave Theory and Techniques*, **54** (11), 3940–3946.

21 Lee, J. and Sarabandi, K. (2007) A synthesis method for dual-passband microwave filters. *IEEE Transactions on Microwave Theory and Techniques*, **55** (6), 1163–1170.

22 Zhang, Y., Zaki, K.A., Ruiz-Cruz, J.A., and Atia, A.E. (2007) Analytical synthesis of generalized multi-band microwave filters. 2007 IEEE MTT-S International Microwave Symposium Digest, June 2007, pp. 1273–1276.

23 Deslandes, D. and Boone, F. (2007) Iterative design techniques for all-pole dual-bandpass filters. *IEEE Wireless and Components Letters*, **17** (11), 775–777.

24 Guan, X., Ma, Z., Cai, P. *et al.* (2006) Synthesis of dual-band bandpass filters using successive frequency transformations and circuit conversions. *IEEE Microwave and Wireless Components Letters*, **16** (3), 110–112.

25 Macchiarella, G. (2013) Equi-ripple synthesis of multiband filters using Remez-like algorithm. *IEEE Microwave and Wireless Components Letters*, **23** (5), 231–233.

26 Cameron, R.J. (2003) Advanced coupling matrix synthesis techniques for microwave filters. *IEEE Transactions on Microwave Theory and Techniques*, **51** (1), 1–10.

27 Cameron, R.J., Yu, M., and Wang, Y. (2005) Direct-coupled microwave filters with single and dual stopbands. *IEEE Transactions on Microwave Theory and Techniques*, **53** (11), 3288–3279.

28 Uchida, H. *et al.* (2004) Dual band-rejection filter for distortion reduction in RF transmitters. *IEEE Transactions on Microwave Theory and Techniques*, **52** (11), 2550–2556.

29 Bila, S., Cameron, R.J., Lenoir, P., Lunot, V., and Seyfert, F. (2006) Chebyshev synthesis for multi-band microwave filters. 2006 IEEE MTT-S International Microwave Symposium Digest, June 2006, pp. 1221–1224.

30 Lunot, V., Seyfert, F., Bila, S., and Nasser, A. (2008) Certified computation of optimal multiband filtering functions. *IEEE Transactions on Microwave Theory and Techniques*, **56** (1), 105–112.

31 Brand, T.G., Meyer, P., and Geschke, R.H. (2015) Designing multiband coupled-resonator filters using reactance transformations. *International Journal of RF and Microwave Computer-Aided Engineering*, **25** (1), 81–92.

32 Brand, G., Geschke, R.H., and Meyer, P. (2015) in *Advances in Multi-Band Microstrip Filters*, EuMA High Frequency Technologies Series (ed. V. Crnojević-Bengin), Cambridge University Press.

22

Tunable Filters

22.1 Introduction

High performance RF tunable filters are needed in reconfigurable systems to facilitate efficient utilization of the available frequency spectrum. They are in demand in front-end receivers for suppression of interfering signals and for relaxation of oscillator phase noise and dynamic range requirements. Tunable filters are also used to replace large filter banks in advanced systems concepts that self-adapt to environmental requirements. Tunable filters have been proposed as well for high-power applications. The advantages in this case are suppression of harmonics generated from the power amplifiers. Over the past two decades, several publications have been reported on tunable filters [1–66]. The majority of papers published on tunable filters are on microstrip tunable filters. To start with, microstrip filters have very low Q values (100–200); moreover the achievable Q degrades further once the filter is integrated with tuning elements. As a result this type of microstrip tunable filter has limited practical applications. In this chapter, we focus on high-Q tunable three-dimensional (3D) filters such as combline, dielectric resonator, and waveguide filters with Q values that can meet the stringent performance requirements of wireless base station and satellite applications. We address filters that are tunable in both center frequency and bandwidth.

The need for tunability and configurability in wireless systems arises when deploying different systems that coexist geographically. Such deployments take place regularly when an operator has already installed a network and needs to add a new generation network, for example, adding a fourth-generation (4G) network to an existing third-generation (3G) network or a 5G network to an existing 4G network. The availability of tunable/reconfigurable hardware also provides the network operator the means to efficiently manage hardware resources while accommodating multistandard requirements and achieving network traffic/capacity optimization. Wireless systems can also benefit from tunable filter technologies in other areas. Installing wireless infrastructure equipment, such as a remote radio unit (RRU) on top of a 15-story high communication tower, is a very costly task. By using tunable filters, one installation can serve many years. If there is a need to change the frequency or bandwidth, it can be done through remote electronic tuning rather than installing a new filter. Additionally, in urban areas, there is a very limited space for wireless service providers to install their base stations due to expensive real estate and/or maximum weight loading constraints on certain installation locations such as light poles or power lines. Therefore, once an installation site is acquired, it is natural for wireless service providers to use tunable filters in order to pack many functions such as multistandards and multibands into one site.

Communication satellite operators are beginning to use more and more reconfigurable payloads to adapt to changing markets over the 15 year lifetime of satellites. This is being made

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

possible by the advances in digital technology (channelizers, reprogrammable processors, FPGAs), steerable multi spot beams, reconfigurable phased array antennas and flexible traveling wave tube amplifier technology. Another element of flexibility is the ability to choose RF channel bandwidths to satisfy a particular set of customer requirements over the lifetime of the satellite. This is being made possible by the advancers in frequency and bandwidth controllable tunable filter networks. Tunable filters offer the potential of added flexibility in terms of frequency selection and bandwidth allocations to the various transponders. Use of tunable filters would also allow sharing of hardware resulting in reduced mass for the payload. Software defined radio technology on-board communication satellites is envisioned as the promising and likely answer to the compelling need for reconfigurable flexible payloads over ever longer satellite lifetimes.

The availability of high-Q tunable filters may also have a significant impact on production cost and delivery schedules in some communication systems. Such systems use multiple filters that are usually identical with the exception of center frequency and bandwidth. The production cost can be significantly reduced by building standard filter units that can be easily reconfigured during the production phase to fit the required frequency plan. In wireless and satellite applications, the delivery schedule has become a key factor in winning or losing a contact. Tunable devices can be built ahead of time and used to offer a competitive delivery schedule.

The ideal tunable filter must exhibit the following features: high loaded Q value, wide tuning range, high tuning speed, good linearity, high power handling capability, small size and mass, and high reliability. The loaded Q value of the tunable filter is determined by the Q value of the filter structure itself and by the inherent loss of the tuning element. The type of the tuning element used also impacts the tuning speed, tuning range, and dc power consumption. There are several different technologies available to realize the tuning elements. The list includes semiconductors [67–69], ferroelectric materials (barium strontium titanate (BST)) [70, 71], ferromagnetics (YIG and ferrite) [72, 73], and mechanical systems (miniature motors, piezoelectric, MEMS) [74–77]. More recently phase change materials (PCM) have been proposed to realize switches [78–82]. The materials exhibit a reversible transition from insulator to metal that can be induced by temperature changes, optical excitation, or current injection. A comparison between these tuning elements for insertion loss, linearity, dc power consumption, and size is given in Table 22.1.

The filter structures that can provide reasonably high Q values are 3D cavities such as coaxial, dielectric resonator, or waveguide filters. These are also the filter structures that are being currently employed in wireless and satellite systems as fixed filters. If reconfigurability is ever employed in wireless and satellite systems, 3D tunable filters stand to be the optimum solution. However, there are several challenges involved in the development of 3D tunable filters. This

Table 22.1 Comparison between tuning elements.

Parameter	Mechanical and piezoelectric	Semiconductor	BST	Ferromagnetic	MEMS	PCM
Unloaded Q	High	Low	Medium	High	High	High
Tuning range	Medium	Low	Medium	Wide	Medium	Medium
Power handling	High	Low	Medium	Medium	High	Medium
Tuning speed	Slow	Fast	Fast	Fast	Moderate	Moderate
Size	Large	Small	Small	Very large	Small	Small
Linearity	Excellent	Poor	Good	Good	Excellent	Very good
Frequency range (GHz)	<20	<30	<3	<20	<40	<100

includes realization of constant bandwidth filters over a wide tuning range, maintaining Q over the tuning range, integration of tuning elements with the high-Q 3D resonators, linearity performance, and power handling capability. These challenges for high-Q tunable filters are discussed in the following section, followed by the current state of the art of high-Q 3D tunable filters.

22.2 Major Challenges in Realizing High-*Q* 3D Tunable Filters

22.2.1 Maintaining Constant Bandwidth and a Reasonable Return Loss Over a Wide Tuning Range

In order to minimize the number of tuning elements and to improve the loss performance of the tunable filter, it is preferable to use tuning elements only to tune the resonator center frequencies. However, the variation of interresonator coupling with frequency is different from that of the input/output coupling. This in turn results in deterioration of the filter return loss and changes in the filter absolute bandwidth over the tuning range. Of course the simplest solution is to add tuning elements to control the interresonator coupling and the input/output coupling as well. In many cases, this solution may not be feasible because of size limitation, design complexity, and the inherent difficulty with tuning sequential and cross interresonator coupling. Therefore, it is preferable to use only tuning elements for the resonators to tune their frequency and rely on other means to maintain constant interresonator and input/output coupling or to compensate for their variations somehow over the tuning range. One simple approach is to use nonsynchronous tuning of the resonators, where the resonator frequencies are not varied by the same shift during the tuning process, which may help to compensate for changes in interresonator coupling. However, this approach can help to a certain extent but fails to yield acceptable results in wide tuning ranges (over 5%). In Section 22.3.1 we present details of a design technique for realizing tunable filters with a constant bandwidth.

22.2.2 Maintaining a Constant High *Q* Value Over a Wide Tuning Range

Most 3D tunable filters start with a relatively high Q value and exhibit a significant Q degradation over the tuning range. This problem can be easily explained by considering the case where we have a screw or a disk attached to a moving mechanism and is penetrating inside a 3D cavity resonator to tune its resonance frequency. As the resonator is tuned over a wide tuning range, the screw/disk needs to have more penetration inside the cavity, which in turn reduces Q significantly. A similar effect takes place when using MEMS, PCM, BST, or semiconductor tuning elements to tune resonators. The loading caused by the capacitance can significantly reduce the resonator Q value. In Section 22.3.2 we present an analysis showing the impact of capacitive loading on Q degradation.

22.2.3 Integration of Tuning Elements with 3D Filters

The integration of a tuning element in the form of a chip (MEMS, PCM, BST, semiconductor) with 3D high-Q resonators is a major challenge. Such tuning elements have a very small size, while 3D cavity resonators (waveguide, coaxial, dielectric resonators) have much larger dimensions. The chip-scale tuning elements need to be integrated in a way to interact with the electromagnetic field inside the resonator and accordingly tune the resonator over a wide tuning range without major Q degradation. Another major challenge in the case of using tuning in the form of switch such as MEMS, PCM, or semiconductor switches is to achieve a large number of tuning states, which requires the development of a large switched capacitor bank of order 5 or 6 (i.e., to achieve 32 or 64 states). Usually such large-bit capacitive banks exhibit a relatively low self-resonance frequency limiting their use to low frequency applications.

22.2.4 Linearity and Power Handling Capability

Many communication systems today have very stringent requirements on linearity performance. Mechanical tuning elements such as motors and MEMS are the optimum choice in such applications. These tuning elements also have the potential of realizing tunable filters capable of handling power levels comparable with their fixed counterpart filters.

In the following sections we present a sample of techniques reported in recent years for realizing combline, dielectric resonator, and waveguide tunable filters.

22.3 Combline Tunable Filters

22.3.1 Combline Tunable Filters with a Constant Absolute Bandwidth

Figure 22.1 shows a conventional cavity combline resonator. The resonator can be tuned by adjusting the gap between the metallic post and a tuning disk. This can be achieved by using a driving mechanism such as motors. The use of conventional mechanical stepper motors can result in fine-tuning steps; however, these motors are usually very expensive and bulky, increasing the overall size of the tunable filter. Alternatively, piezo motors [20] can be used to replace traditional motors. Piezo motors have high resolution and are available in various sizes and can be easily integrated with the tunable filter. A motor-based cavity combline tunable filter has been demonstrated in [12] for WiMAX applications. The filter is designed to operate continuously over the frequency range of 2550–2650 MHz, with a bandwidth of 30 MHz. The basic building block of the filter is the cavity shown in Figure 22.2. It has a size of 30 mm × 30 mm × 30 mm. The post has a height of 21 mm and a diameter of 12 mm. The cavity is integrated with a piezo motor [20] that is mounted to the lid while its shaft is attached to a tuning disk. As a result the disk can move up and down over the metallic post. The measured tuning performance of the cavity is shown in Figure 22.3. By changing the gap in 20 μm steps, the cavity was tuned from 2.565 to 2.645 GHz, demonstrating a measured Q from 2252 to 2914. The cavity in this case is made of copper with a tuning disk made of a gold plated aluminum. It is noted that almost a 25% degradation in Q value was observed over a 4% tuning range. Certainly, the degradation can be potentially reduced by optimizing the dimensions of the cavity and tuning disk, yet this example illustrates that Q can degrade considerably over the tuning range.

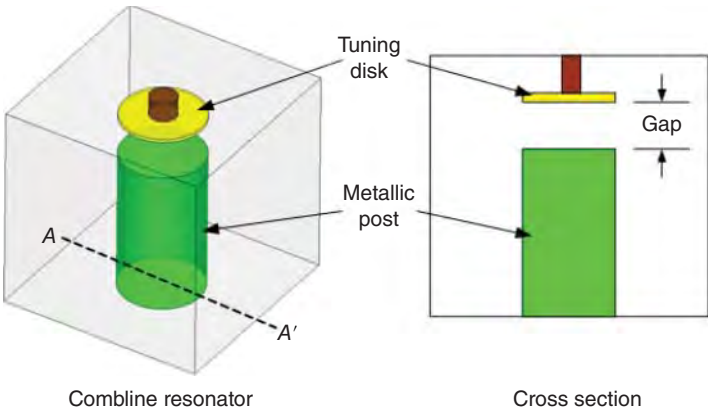


Figure 22.1 A conventional combline resonator.

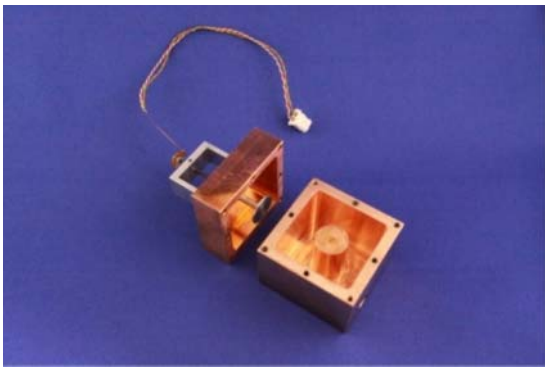


Figure 22.2 A combline resonator integrated with a piezo motor. (*Source*: Fouladi *et al*. 2013 [12]. Reproduced with the permission of IEEE.)

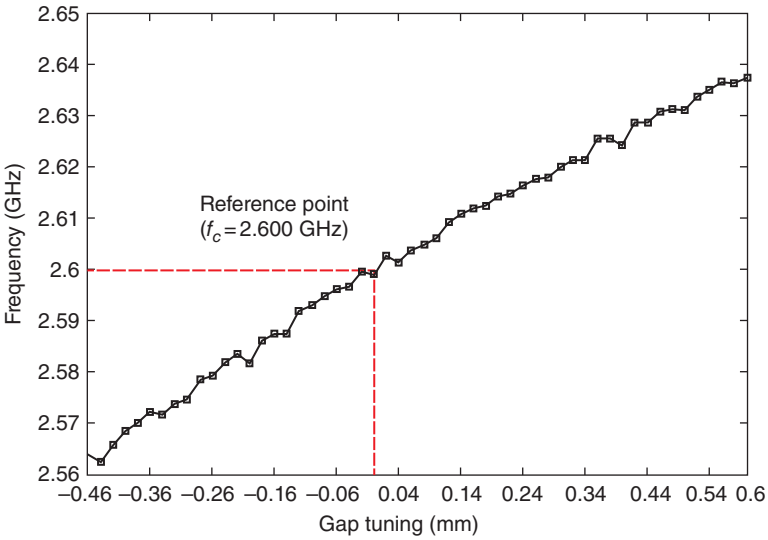


Figure 22.3 The measured tuning performance of the cavity shown in Figure 22.2. (*Source*: Fouladi *et al*. 2013 [12]. Reproduced with the permission of IEEE.)

A 6-pole filter was selected to meet the WiMAX requirements [12]. Figure 22.4 shows the filter configuration assembled with piezo motors. It is a 6-pole filter designed to have two transmission zeros, realized using a probe inserted between resonators 2 and 5 to provide a negative coupling. The piezo motors were used only to tune the resonance frequency of the resonators. The measured performance of the WiMAX tunable filter is shown in Figure 22.5 for three frequencies. The filter is capable of meeting all WiMAX requirements demonstrating a constant absolute bandwidth over the desired tuning range. The variation of interresonator and input/output coupling over the tuning range could be circumvented in this case by the use of nonsynchronous tuning since the tuning range is only 4%. To maintain a constant bandwidth over a wider tuning range, other techniques need to be implemented. In the following we outline a procedure for realizing tunable filters with a constant bandwidth where only tuning elements are used to tune the resonator center frequencies.

As in the case of fixed filters, the performance of tunable filter is determined by the elements of the coupling matrix. It should be noted that the elements of the coupling matrix are dimensionless as they are expressed in terms of the lowpass prototype. For a filter to maintain the same

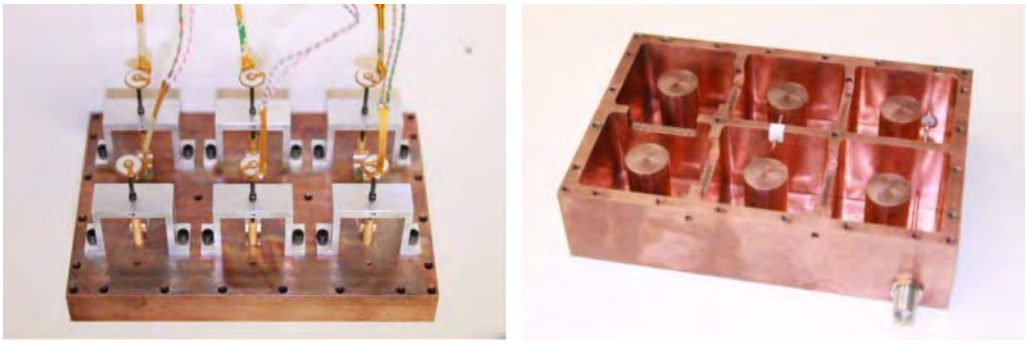


Figure 22.4 A picture of a 6-pole WiMAX tunable filter. (*Source*: Fouladi *et al*. 2013 [12]. Reproduced with the permission of IEEE.)

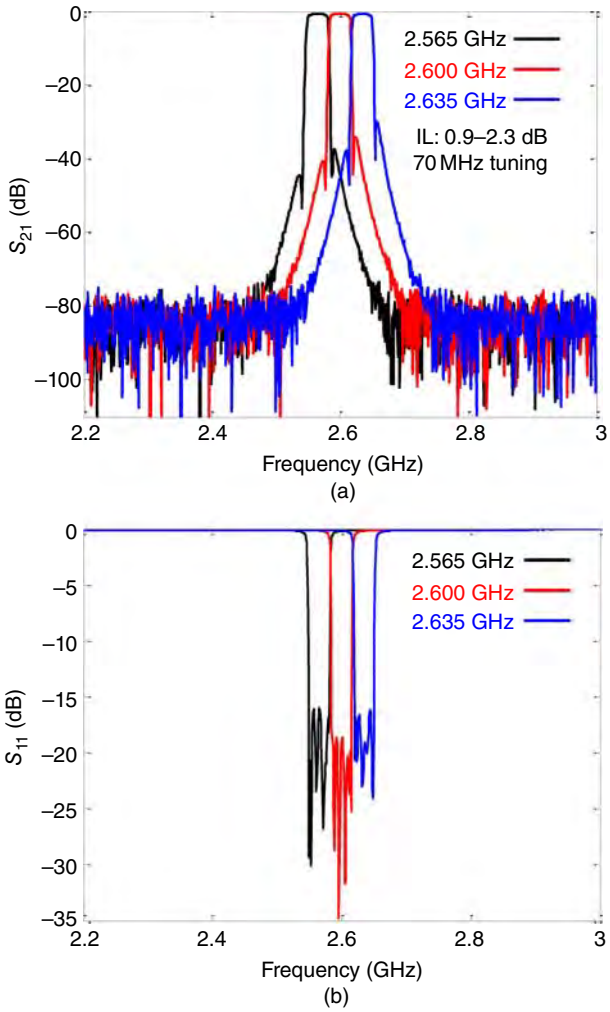


Figure 22.5 The measured results of the WiMAX tunable filter shown in Figure 22.4. (*Source*: Fouladi *et al*. 2013 [12]. Reproduced with the permission of IEEE.)

shape as it is tuned in center frequency, it should have a coupling matrix with identical normalized elements at all center frequencies. However, the normalized coupling element is related to the physical coupling that varies with frequency as the filter center frequency is tuned.

The normalized interresonator coupling value M_{ij} relates to the physical coupling value K_{ij} as

$$M_{ij} = \frac{\omega_0}{\Delta\omega} \cdot K_{ij} \tag{22.1}$$

where ω_0 is the center frequency and $\Delta\omega$ is the bandwidth. The normalized input resistance R can be calculated from the group delay of reflection coefficient seen at the first resonator, when all other resonators are removed or detuned far away from the filter center frequency (see Chapter 14):

$$R = \frac{4}{\Delta\omega \cdot \tau(\omega_0)} \tag{22.2}$$

where $\tau(\omega_0)$ is the group delay value of the input reflection coefficient at the center frequency ω_0.

Equations (22.1) and (22.2) suggest that for a tunable filter to maintain a constant bandwidth and an acceptable return loss over the tuning range, the input/output coupling configuration must be selected in a way to maintain a constant peak group delay as the resonance frequency is varied. The interresonator coupling configuration must be also selected such that the product $\omega_0 . K_{ij}$ remains constant as the resonance frequency is changed.

To illustrate the concept consider the design of the 4-pole combline tunable filter. The combline resonator consists of a cylindrical post with a diameter of WD = 12 mm and a height of HD = 21 mm inside a cavity resonator with dimensions of $H \times C \times D$ = 30 mm × 30 mm × 30 mm. The tuning element consists of a disk with a diameter of 12 mm and a threaded rod connecting the tuning disk to the cavity lid or ground. The resonator can be simply tuned by adjusting the gap between the tuning disk and the cylindrical post (gap). The center frequency can be tuned from 2.24 to 2.71 GHz (470 MHz tuning) by varying the gap from 3 to 8 mm.

22.3.1.1 Coupling Iris Design for Constant Bandwidth

Figure 22.6 shows the schematic diagram of the inter-resonator coupling slot that is used in the constant bandwidth tunable filter reported in [17, 18] for interresonator coupling values. Using such coupling structure and through optimization in high frequency structure simulator (HFSS), it is possible to obtain a balanced coupling that maintains a constant absolute bandwidth as the center frequency is varied. The coupled pair of resonators exhibit even and odd resonances f_e and f_m that are obtained by setting the appropriate boundary conditions for

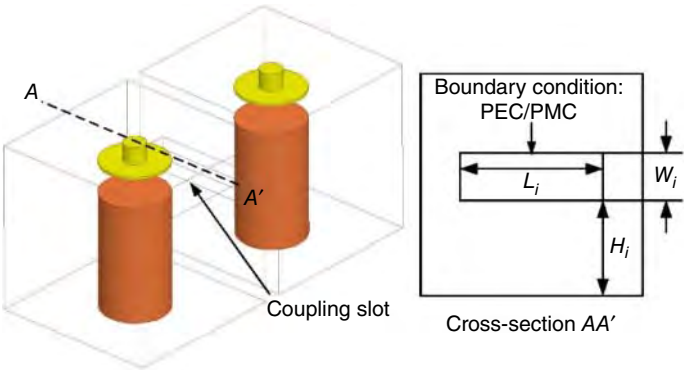


Figure 22.6 Schematic diagram of interresonator coupling.

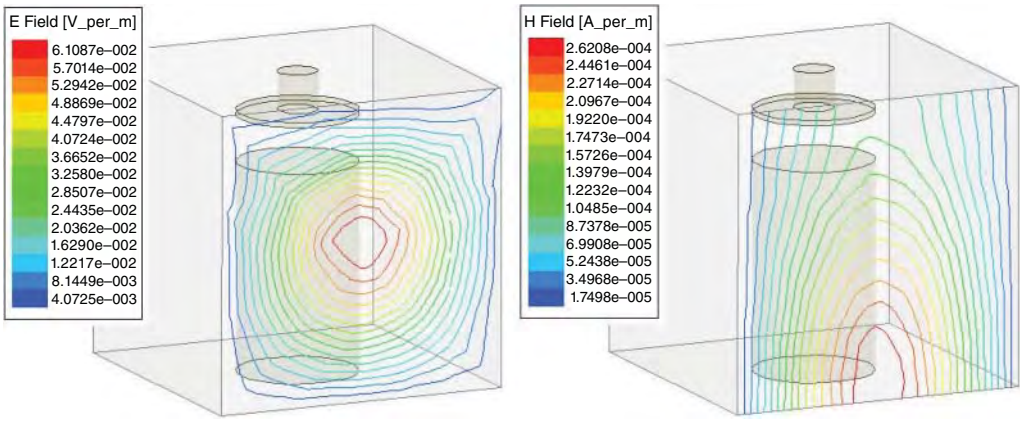


Figure 22.7 Magnitude of (a) electric and (b) magnetic fields on the sidewall of the cavity.

the horizontal slot, electrical wall (PEC) for f_e, and magnetic wall (PMC) for f_m. The physical coupling coefficient can be obtained as

$$K_{ij} = \frac{f_e^2 - f_m^2}{f_e^2 + f_m^2} \tag{22.3}$$

Figure 22.7 shows the simulated electric and magnetic fields on the cavity side wall where the coupling slot is located. As seen in this figure, the maximum electric field is occurring close to the center of the sidewall, while the magnetic field is maximized near the bottom of the cavity sidewall. The magnitude of the electric coupling and magnetic coupling can be adjusted by the slot vertical position H_i. Therefore, by optimizing the height of the horizontal slot, it is possible to obtain a balanced interresonator coupling to maintain the normalized coupling value constant as the center frequency is tuned.

The physical coupling value K_{ij} should be inversely proportional to the center frequency f_o. EM simulation results for different height values of the horizontal coupling slot are presented in Figure 22.8. Dimensions of the coupling slot are $W_i \times L_i = 4\text{ mm} \times 20\text{ mm}$. For a coupling slot near the bottom edge of the cavity ($H_i = 4$ mm), the normalized coupling value M_{ij} increases as the center frequency is tuned to higher values. On the other hand, for a coupling slot far away from the bottom of the cavity ($H_i = 24$ mm), M_{ij} decreases as the center frequency is increased. An optimum value exists between these two values of H_i. Simulation results for an optimum coupling slot height of $H_i = 17.2$ mm are presented in Figure 22.9; a constant normalized coupling value M_{ij} between the resonators is achieved with this height.

22.3.1.2 Input/Output Coupling Design for Constant Bandwidth

From equation (22.2), for a tunable filter to maintain a constant bandwidth during the tuning process, its normalized input impedance R should be a constant value over the entire tuning range. R is determined by the group delay, which can be obtained by using EM simulation of the reflection coefficient of a probe feeding the first resonator with the second cavity detuned as shown in Figure 22.10. The group delay simulation results for two probe length values of

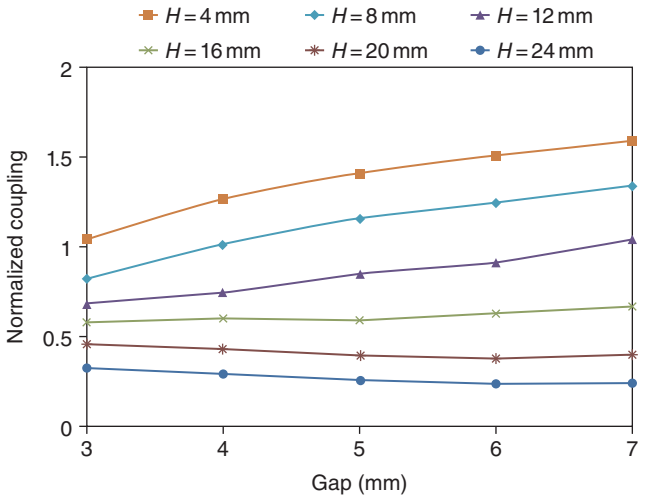


Figure 22.8 Variations in the normalized coupling value M_{ij} for different H_i values.

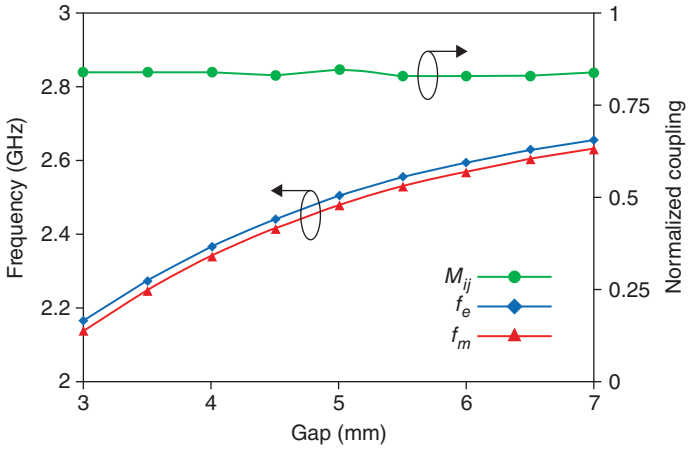


Figure 22.9 Normalized coupling value and even- and odd-mode resonance frequencies for $H_i = 17.2$ mm.

$L_p = 20$ mm and 25 mm are given in Figure 22.11 as the center frequency is tuned by varying the gap between the tuning disk and the metallic post. It can be seen that, for a probe length of $L_p = 20$ mm, the peak group delay varies from 25.4 to 9.74 ns (a ratio of 2.6) as the center frequency is tuned from 2.31 to 2.62 GHz. The change in the peak group delay for the case of $L_p = 25$ mm over the same frequency range is from 14.5 to 6.9 ns (ratio of 2.1) for the same tuning range. This indicates the trend in this specific case that a longer probe length yields is a lower variation in the peak group delay and correspondingly in the input impedance value R. A probe length of 29.3 mm was selected as the optimum length for this case. Figure 22.12 shows that the peak group delay value is relatively constant over the 500 MHz tuning range for $L_p = 29.3$ mm.

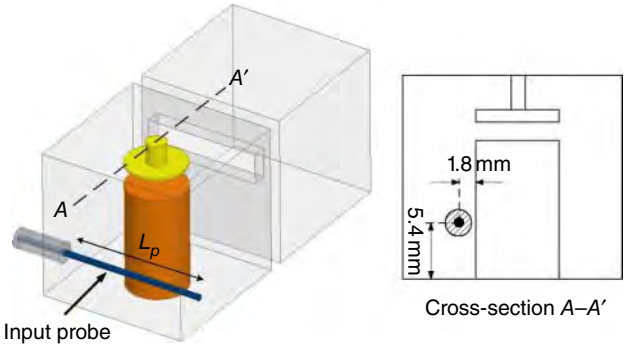


Figure 22.10 EM model to simulate group delay for different L_p values.

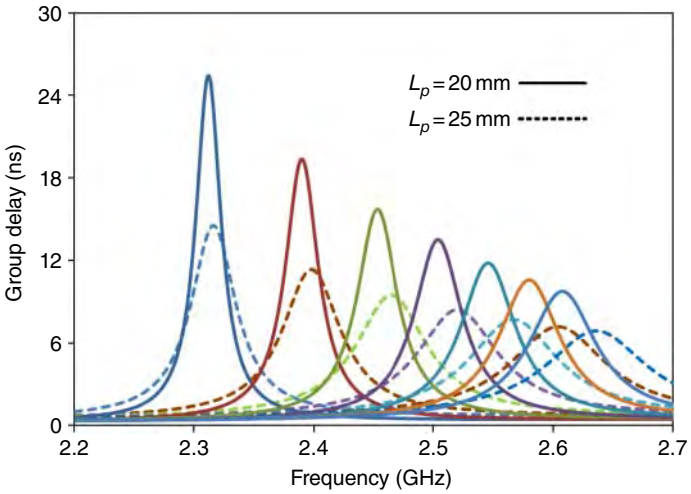


Figure 22.11 Variation in the group delay when the center frequency is tuned for two different values of L_p.

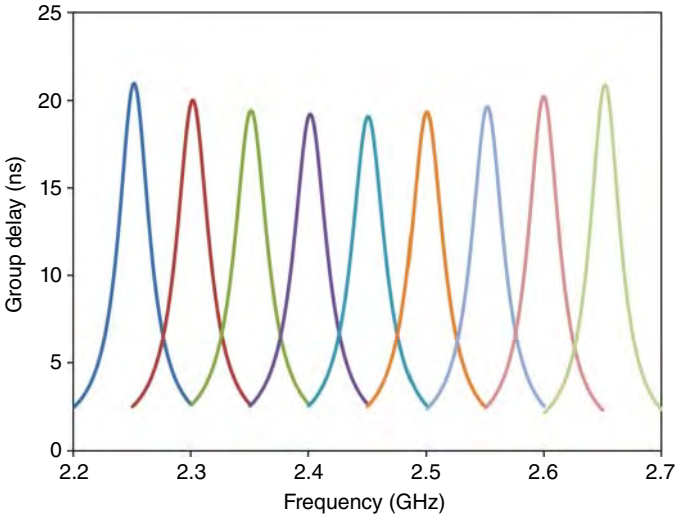


Figure 22.12 Variation in the group delay when the center frequency is tuned for $L_p = 29.3$ mm.

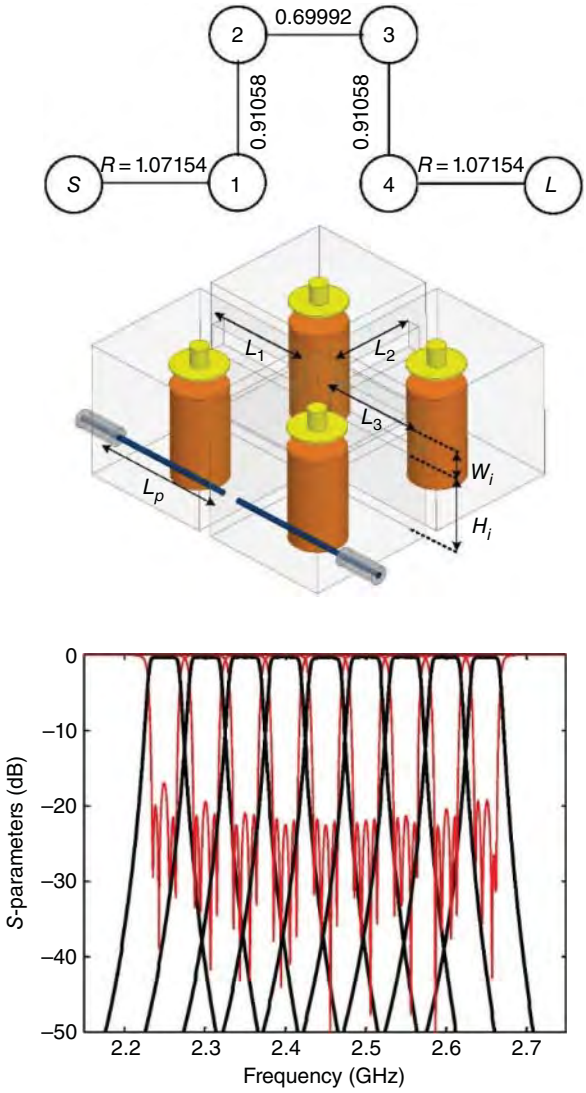


Figure 22.13 Simulated tuning performance of the 4-pole filter.

Using the previous analysis, a 4-pole tunable filter with a constant absolute bandwidth is designed. Figure 22.13 shows the coupling structure and 3D EM model of the designed filter. For this filter all the coupling slots are designed to have a height of $H_i = 17.2$ mm. Dimensions of the slots are $L_p = 20$ mm $L_p = 25$ mm $W_i = 6$ mm and $L_1 = L_3 = 23.125$ mm, $L_2 = 20.024$ mm. Long input probes ($L_p = 29.3$ mm) are used to minimize variation in the input/output impedance values (R). The EM simulation results are shown in Figure 22.13. The measured results of a prototype 4-pole tunable filter tuned manually from 2.25 to 2.65 GHz (400 MHz tuning range) are shown in Figure 22.14. The in-band return loss is better than 15 dB for all the tuning states. The measured maximum insertion loss varies from 1.04 to 0.53 dB over the tuning range, and the variation in the bandwidth is from 31.1 to 28.9, that is, a ±1.1 MHz (less than ±3.7% chance in bandwidth). The extracted Q values at different frequencies are presented in Figure 22.15. The measured unloaded Q values are better than 3000 over the tuning range. The filter lid is replaced

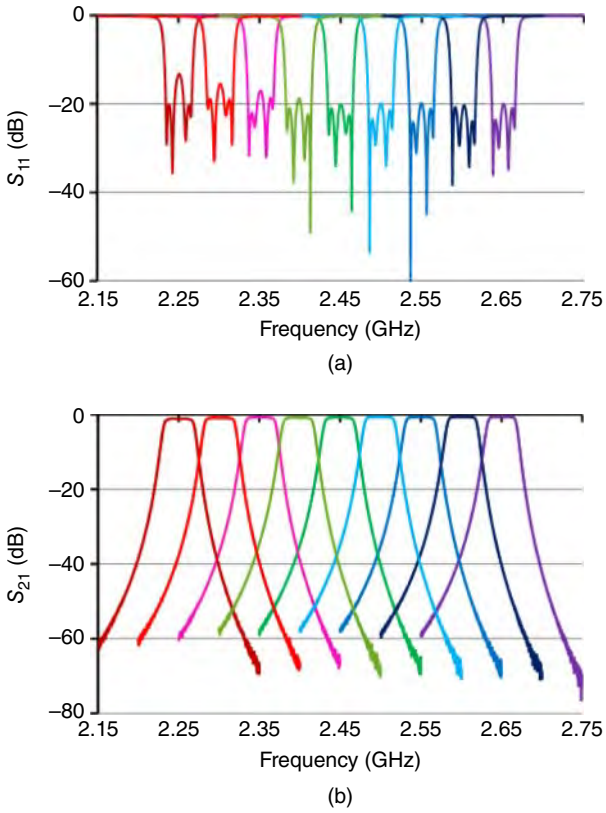


Figure 22.14 Measured return loss and insertion loss over tuning range (manual tuning). (*Source*: Courtesy: F. Huang [17].)

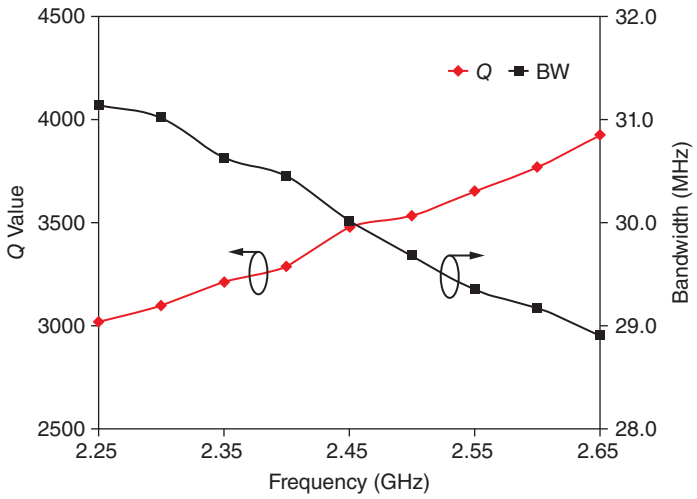


Figure 22.15 Measured *Q* value over tuning range.

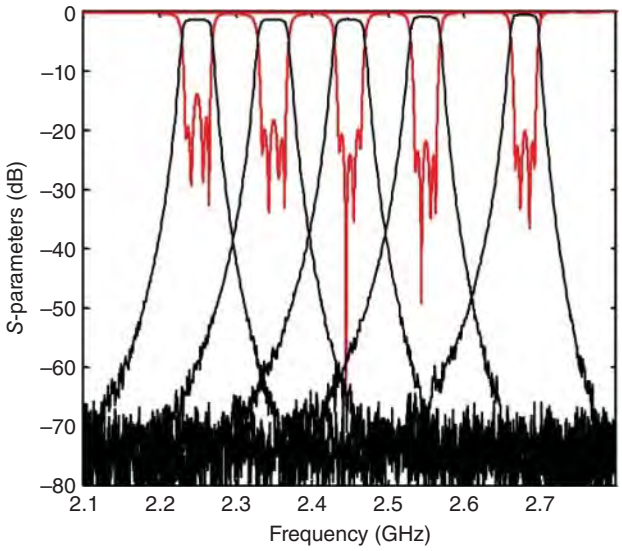


Figure 22.16 Measured *S*-parameters of the tunable filter over the tuning range (piezo motor tuning). (*Source*: Courtesy: F. Huang [17].)

with another lid integrated with piezo motors. The measured results obtained by piezo motors are shown in Figure 22.16.

Using the same approach, a 5-pole tunable filter at 5 GHz (for WiFi applications) with a tuning range of 1 GHz and a 7-pole 7.5 GHz tunable filter with a tuning range close to 1 GHz were developed in [18]. Figure 22.17 shows the measured results of 5 GHz WiFi tunable filter, whereas Figure 22.18 shows the filter configuration of the 7.5 GHz tunable filter and its measured results. These filters were tuned by piezo motors. An absolute constant bandwidth with an acceptable return loss performance was achieved. It should be noted that the filters shown in Figures 22.17 and 22.18 were built using aluminum housing. A much better insertion loss can be realized with the use of filter housings made of copper or silver plated aluminum.

While the obvious approach to mechanically tune a combline resonator is to use a screw or disk that is moving up and down over the metallic post, such techniques yield a noticeable Q degradation over the tuning range. An alternative approach is to use a disk that moves radially over the metallic post. This will help to have less Q variation over a relatively wide tuning range. More importantly, it makes it possible to use simple motor designs where the motor shaft rotates radially, eliminating the need to use any complicated mounting jigs that are often required to get vertical displacement of the motor shaft. To illustrate the difference between vertical tuning and angular tuning, a cavity combline resonator was tuned by both techniques in [15]. Figure 22.19a shows the resonance frequency and Q versus gap of a resonator employing vertical tuning, whereas Figure 22.19b shows the Q and resonance frequency of a similar combline resonator employing angular tuning. The starting point for the resonance frequency, Q, and tuning range was kept the same in both cavities. It can be seen that the use of angular tuning approach yields almost a constant Q over a 15% tuning range. The vertical tuning approach yields close to 30% Q degradation over the same tuning range. It should be noted however that by optimizing the resonator configuration, it is possible to achieve a minimum Q degradation even with vertical tuning.

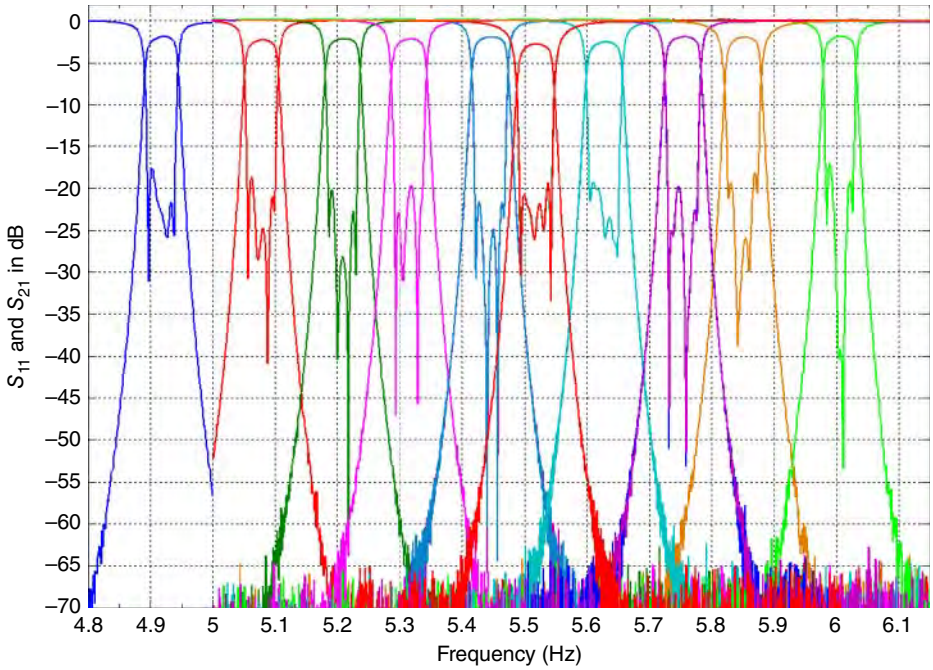
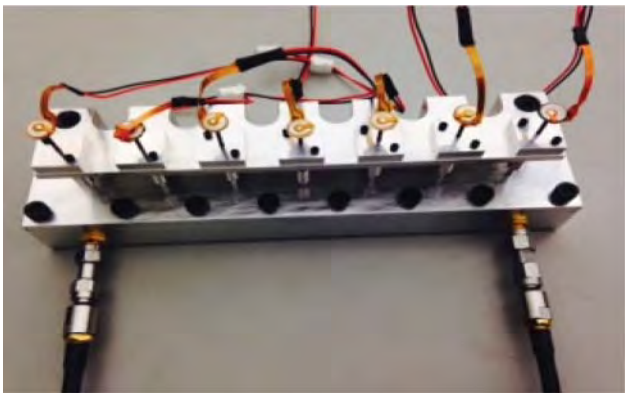


Figure 22.17 The measured results of 5-pole 5 GHz WiFi cavity combline tunable filter, the filter housing is made of aluminum. (*Source*: Courtesy: M. Nasresfahani [18].)

22.3.2 MEMS-Based Combline Tunable Filters

Cavity combline filters can be also tuned using MEMS, PCM, BST, or semiconductor tuning elements. The idea is to use an isolated disk located over the metallic post. The disk and the metallic post will form a floating capacitance that is not connected to ground. This capacitance can be connected to a MEMS-based switched capacitor bank or semiconductor varactor that is shorted to ground at the other end. Thus, the total capacitance that is loading the metallic post consists of the capacitance that is formed by the gap between the disk and the metallic post and



(a)

Figure 22.18 (a) A 7-pole tunable combline filter, tuned by piezo motors, and (b) its measured results; the filter housing is made up of aluminum. (*Source*: Courtesy: M. Nasresfahani [18].)

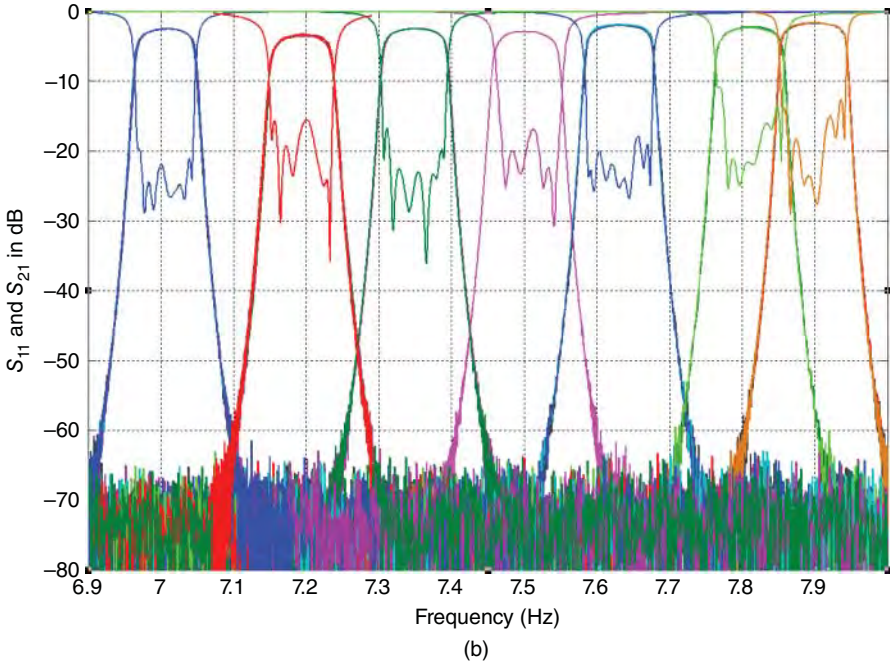


Figure 22.18 (*Continued*)

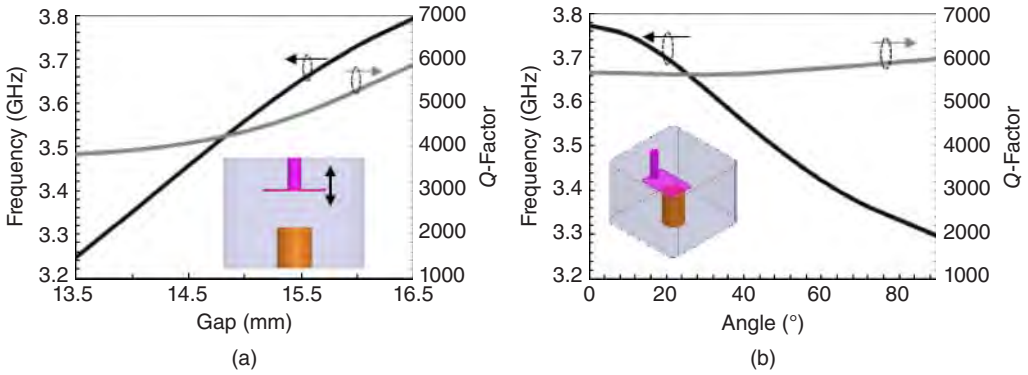


Figure 22.19 A comparison between a resonator tuned using (a) vertical tuning and (b) angular tuning. (*Source*: Iskander *et al*. 2014 [15]. Reproduced with the permission of IEEE.)

the capacitance of the tuning element. Figure 22.20 shows a tunable combline resonator tuned by RF MEMS switched capacitor bank [12]. The switched capacitor bank is assembled on a PCB board and is mounted on the top wall of the cavity. One end of the switched capacitor bank is attached to the tuning disk that is isolated from the cavity wall with a Teflon spacer, and the other end is attached to the ground with the use of via hole. The variable loading effect of the capacitor bank on the tuning disk is used to tune the resonator. The switched capacitor bank consists of high-Q capacitors in series with RF MEMS contact-type switches. A simplified schematic view and the equivalent circuit diagram of the RF MEMS tuning circuit is also presented in Figure 22.20c. By turning the MEMS switches on and off, using a dc actuation voltage, it is possible to change the value of loading capacitor and adjust the resonant frequency of the cavity.

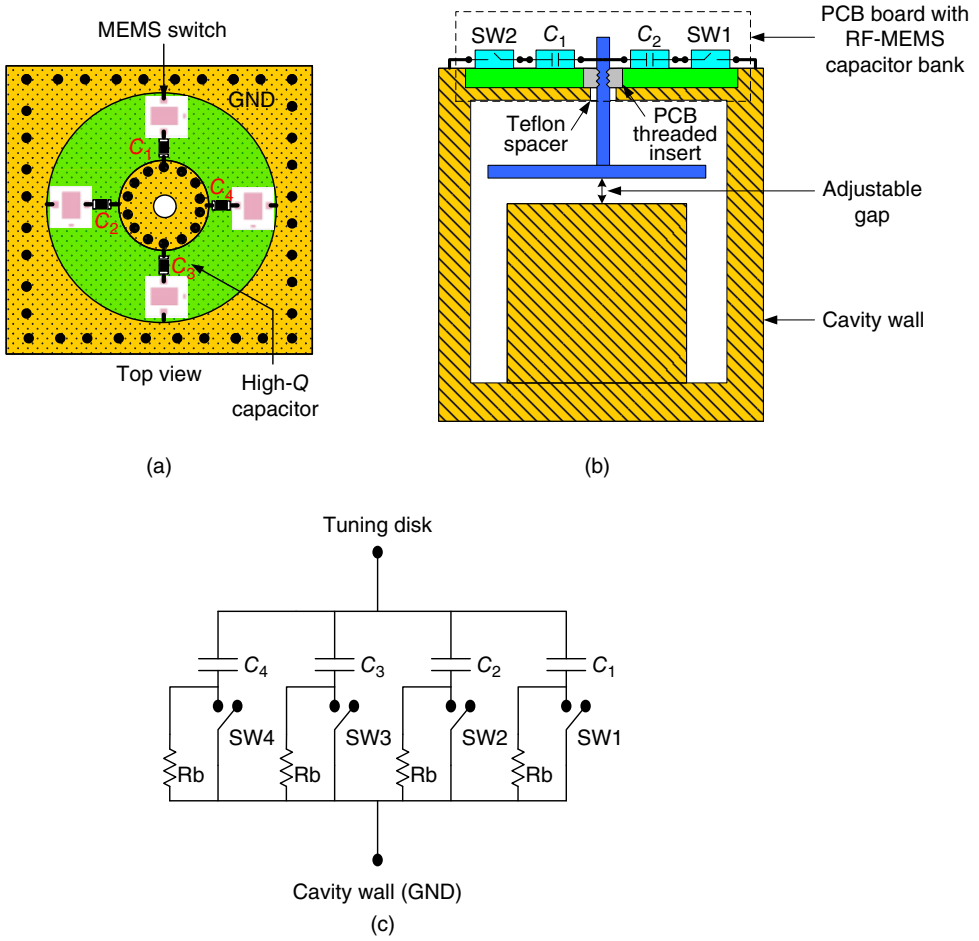


Figure 22.20 (a) Top view of a tunable combline resonator tuned by RF MEMS switched capacitor bank, (b) side view of the resonator, and (c) the equivalent circuit of the tuning circuit. (*Source*: Fouladi *et al*. 2013 [12]. Reproduced with the permission of IEEE.)

The switched capacitor bank is assembled using MEMS SPST switches connected to ceramic capacitors. Each MEMS switch is actuated separately. The resonator is tuned from 2.503 to 2.393 GHz (110 MHz), while the measured Q ranges from 1300 to 374 over the same tuning range. It should be noted that the values listed in [12] for Q and tuning range were obtained for a specific cavity structure. A better Q value and a larger tuning range can be potentially achieved by optimizing the resonator structure and the PCB board design.

A 6-pole WiMAX tunable combline filter employing this concept is demonstrated in [12]. Figure 22.21 shows the filter configuration. It has six PCB boards with the RF MEMS tuning elements assembled on the top cover of the filter. It is very important to establish a good ground contact between the MEMS capacitor banks and the cavity wall. For this reason, grounding gaskets on the lid and also sufficient number of via holes between the bottom ground plane of the PCB and the top ground plane of the MEMS capacitor bank are required [12]. The measured performance of the MEMS-based tunable filter is shown in Figure 22.22. A small tuning range is achieved due to the fact that this particular filter was tuned using synchronous tuning. The use of nonsynchronous tuning should yield a wider tuning range. This, however, will require

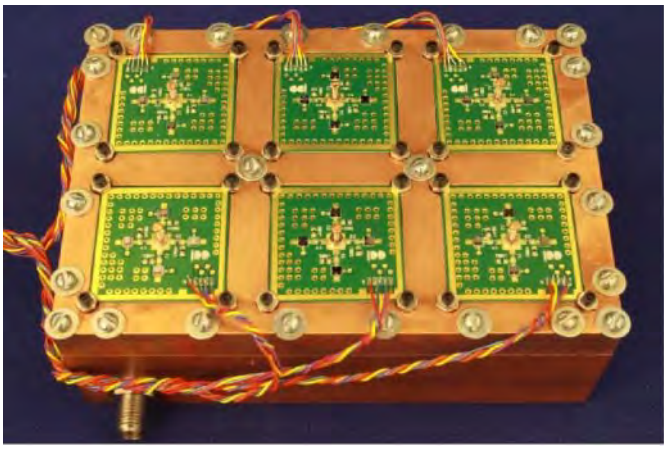


Figure 22.21 A picture of the MEMS-based tunable combline filter. (*Source*: Fouladi *et al*. 2013 [12]. Reproduced with the permission of IEEE.)

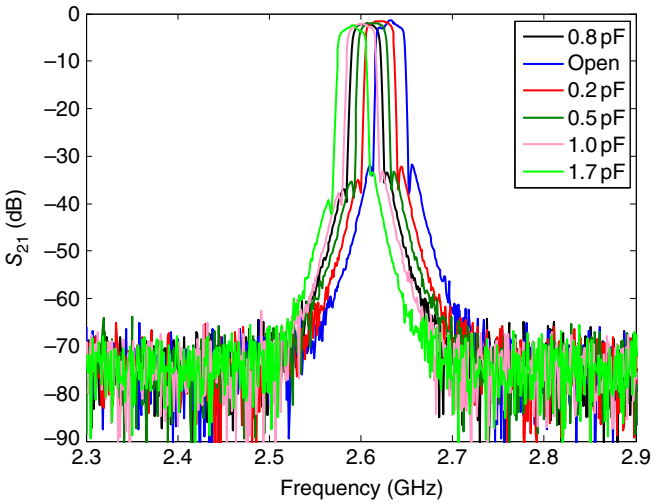


Figure 22.22 Measured results of the combline filter shown in Figure 22.21. (*Source*: Fouladi *et al*. 2013 [12]. Reproduced with the permission of IEEE.)

the use of a continuously variable capacitor or the use of a switched capacitor bank with a large number of states to allow selection of the appropriate capacitance value. An alternative solution is to apply the technique that is described earlier to achieve an absolute constant bandwidth over a wide tuning range.

The Q of this tunable filter is determined by the Q of the cavity, the resistance associated with the MEMS switch, the resistance of the chip ceramic capacitor, and the resistance associated with the PCB board and the via holes needed to connect the tuning elements to the filter housing (ground). It should be noted that capacitive loading also plays a major role in determining the overall loaded Q over the tuning range.

In order to study the effect of the capacitive tuning element on the Q of combline tunable resonator, an EM-based equivalent circuit model was developed in [19]. The model clarifies how a resonator performs when loaded with tuning element in the form of a variable capacitor.

Furthermore, it helps the designer to gain more insight into improving the tunable resonator performance. Figure 22.23 shows a 3D model for the basic combline cavity with five nodes annotated on it. These five nodes are used as a basis for introducing the ADS circuit model shown in Figure 22.24. Both coaxial transmission line circuit models and lumped components are used to model the sections between each node. For instance, between nodes 1 and 2, we have a coaxial line with known inner and outer radii, length, and dielectric, which can be easily modeled in ADS. Nodes 2–3 are modeled as an inductance to model the inner conductor extension and

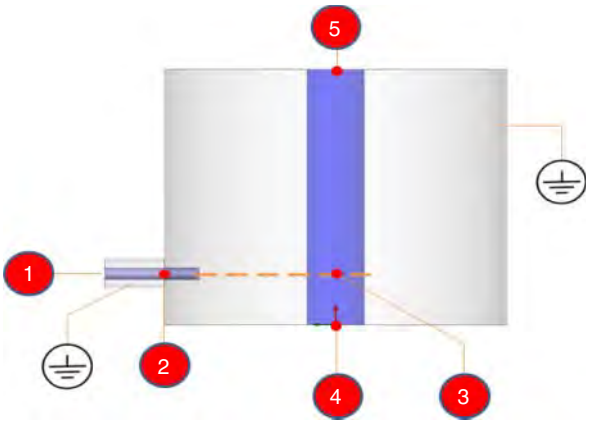


Figure 22.23 Coaxial cavity short-circuited from both sides with all nodes identified.

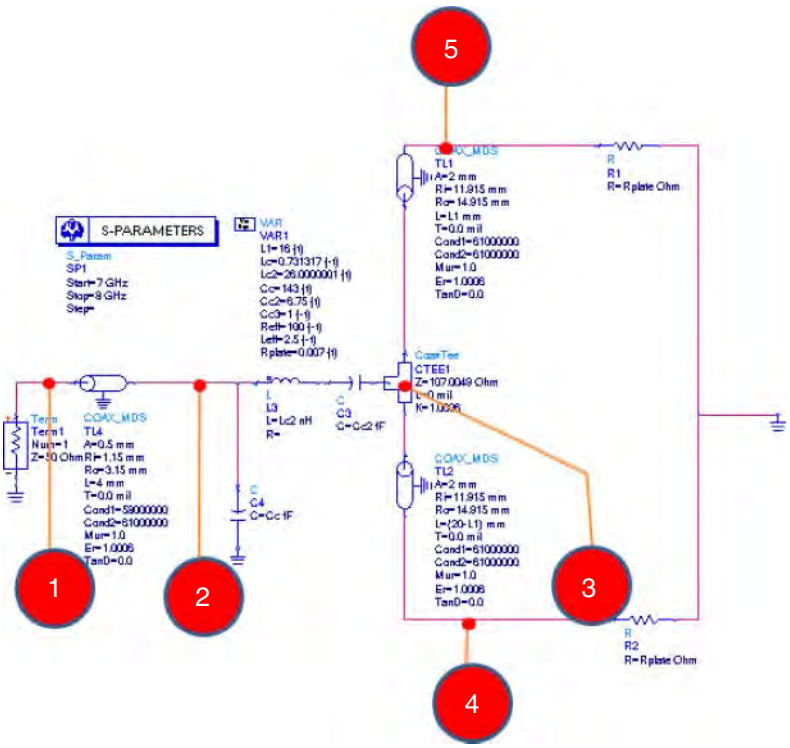


Figure 22.24 Equivalent circuit model following the nodes shown in Figure 22.23. (*Source*: Courtesy: M. Iskander [19].)

Table 22.2 A comparison between the EM model and circuit model.

	f_o (GHz)	Q_o
EM model	7.489	4941.7
Circuit model	7.447	4864.2

then a capacitance to model the air gap between the tip of the inner conductor and the vertical post and a capacitance to ground. At node 3, it is assumed that we have a T junction with three ports connected to the input coupling lumped circuit and the coaxial lines representing the resonator. Conductive losses in the transmission lines were included in the ADS models. Resistors were also included to represent the losses of the bottom and top metallic cavity walls. The values of the lumped components are obtained by fitting the performance of the ADS circuit model to that of the HFSS EM model. Table 22.2 shows that there is a very good match between the two models in terms of center frequency and Q factor.

Thus far, the equivalent circuit model is providing similar results to that of the EM model. The goal now is to obtain an equivalent circuit model of a coaxial resonator that is short-circuited at one end while the other end is connected to the tuning element represented here as an ideal varactor. The model consists of a varactor that is connected to a substrate and couples to the coaxial resonator by two vias, one of which is connected to the post and the other of which is connected to the resonator walls (ground). Figure 22.25 shows the 3D EM model for the coaxial resonator, while Figure 22.26 shows the equivalent circuit including the varactor. The substrate is presented by a capacitance and its dielectric losses are represented by a capacitance in series with the substrate capacitance. The via is represented by an inductance in series with a resistor.

Figure 22.27 shows the match between the EM model and the circuit model as the varactor capacitor is varied, that is, over a wide frequency range. The results obtained give confidence in the ADS circuit model. This model provides an understanding of the circuit components and the possible ways to improve the performance. For example, the designer can now have a very good idea about the effect of the losses of the substrate, how far it affects the performance, and

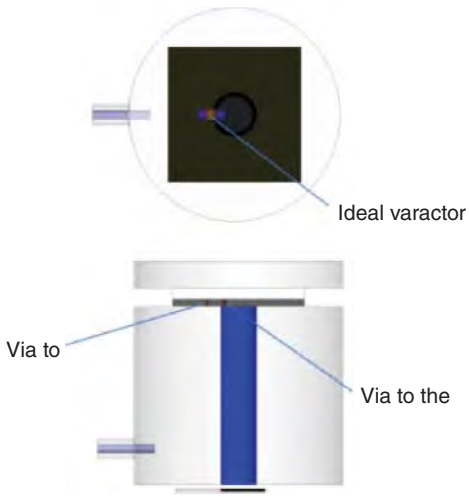


Figure 22.25 Different views for the EM model of a coaxial resonator connected to an ideal varactor at the top end.

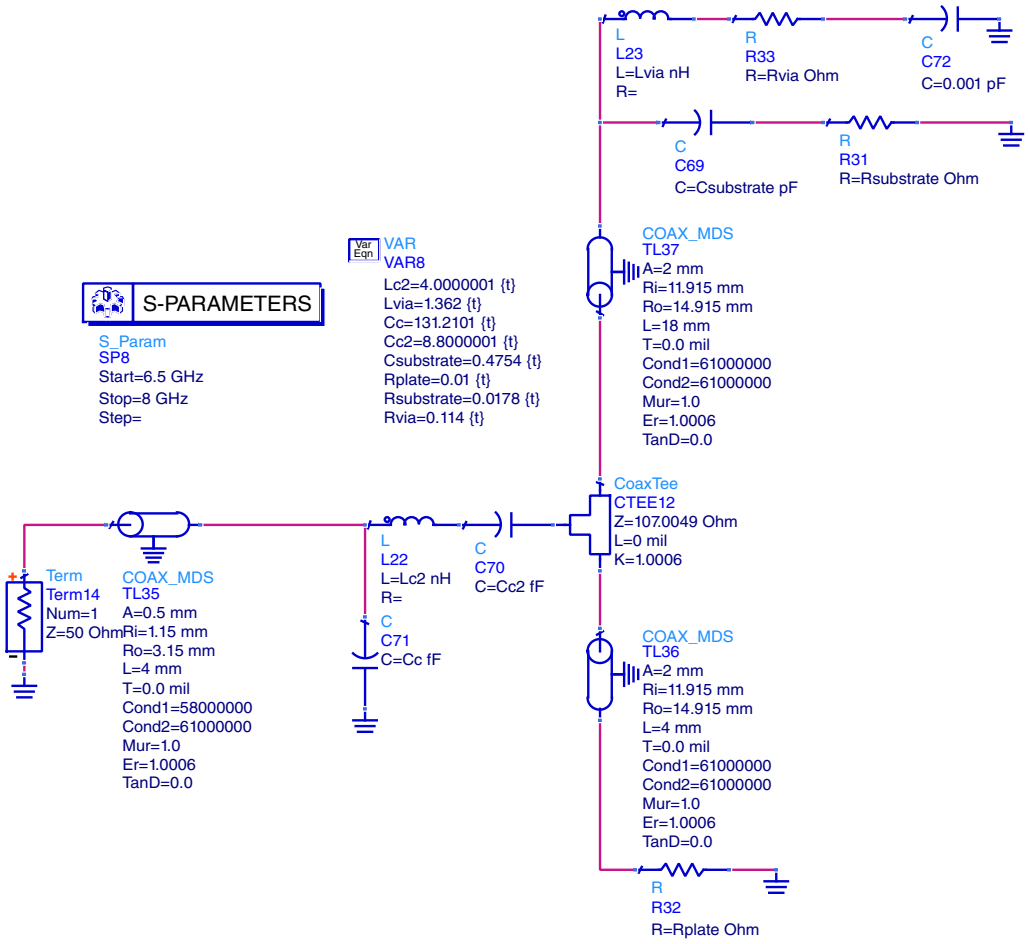


Figure 22.26 Final equivalent circuit model of a coaxial resonator connected to an ideal varactor at the top end, mounted on top of an RT/duroid 5880 substrate. (*Source*: Courtesy: M. Iskander [19].)

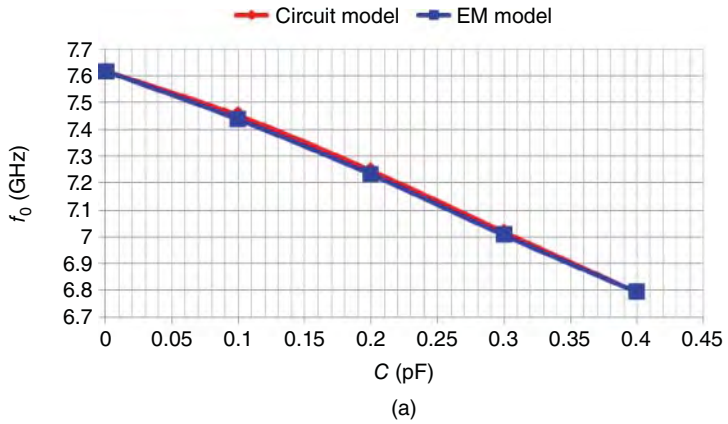


Figure 22.27 Comparison between circuit model and EM model performance: (a) center frequency versus capacitance and (b) *Q* versus capacitance. (*Source*: Courtesy: M. Iskander [19].)

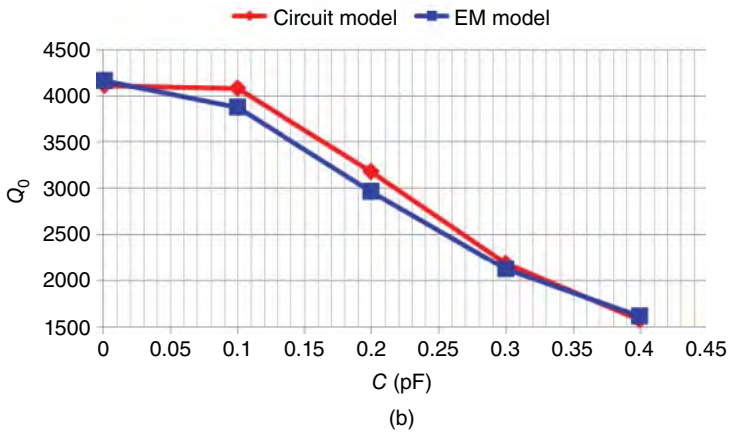


Figure 22.27 (*Continued*)

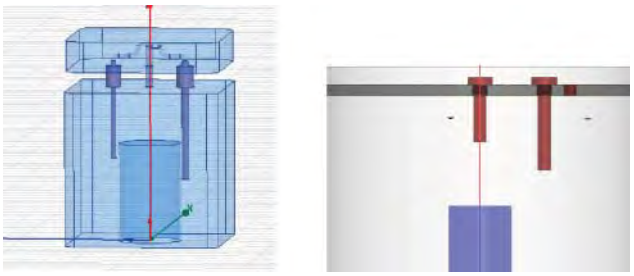


Figure 22.28 A schematic illustrating the concept of switched rod tuning (patent pending).

how indispensable the usage of low-loss substrates is for high-Q applications. It is interesting to observe that the loaded Q degrades significantly as the varactor capacitance increases even if we have an ideal lossless tuning element. This in turns highlights the tradeoff between the tuning range (capacitance loading) and Q degradation.

To circumvent this issue of capacitance loading, a novel concept was proposed in [13]. Figure 22.28 shows a schematic a 2-pole filter illustrating the tuning concept. It consists of cavity resonators having metallic tuning rods isolated from the cavity housing. The rods are connected or disconnected to housing through the use of MEMS switches or any other type of switches such as PCM switches or semiconductor switches. The rods can have different dimensions or be asymmetrically disposed within the cavity to get multiple states. For example, having three rods in the cavity can potentially yield 8 tuning states, and 4 rods can lead to 16 states.

Figure 22.29 shows a combline tunable filter based on this concept. The tuning elements are in the form of strips printed on a substrate that are connected or disconnected to the housing through the use of MEMS switches. Figure 22.30 shows the EM simulation results for the case the MEMS switches are presented by ideal switches with zero resistance. It can be seen the filter maintains a high Q value over the tuning states. The experimental results for the case where actual MEMS switches are used are given in [13].

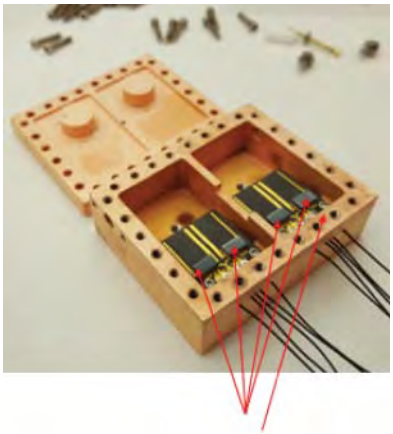


Figure 22.29 A tunable combline resonator employing the concept of switched rod tuning. (*Source*: Jiang and Mansour 2016 [13]. Reproduced with the permission of IEEE.)

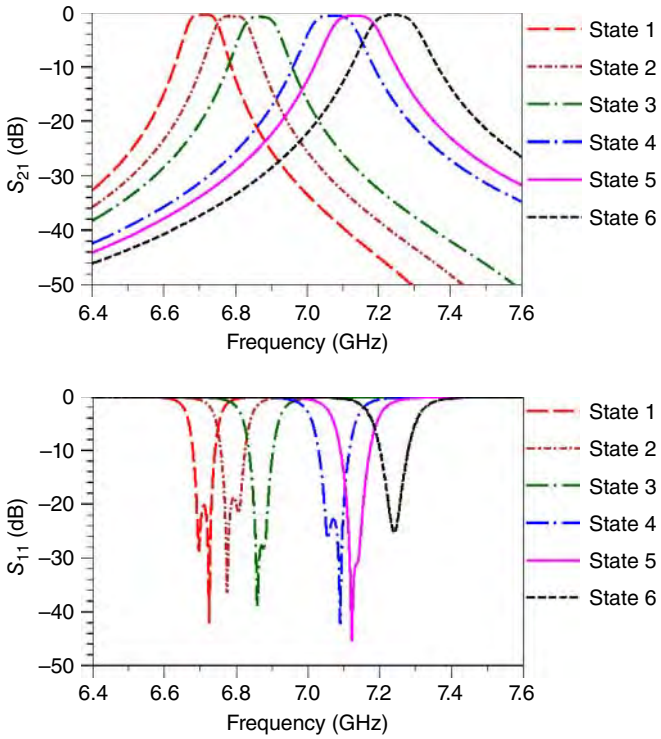


Figure 22.30 Simulation results for the 4-bit tuning structure using MEMS switches.

22.4 Tunable Dielectric Resonator Filters

22.4.1 Tunable Dielectric Resonator Filters Employing Mechanical or Piezoelectric Actuators

Dielectric resonator filters are emerging as the current baseline designs for the majority of RF filters used in wireless and satellite applications. They offer high Q value with a relatively

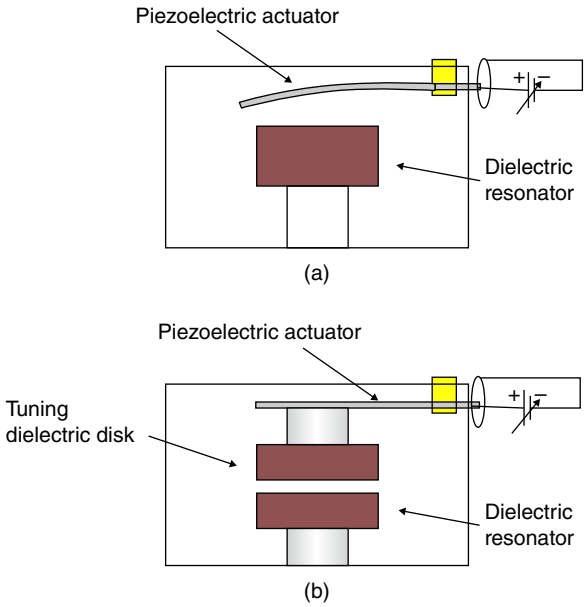


Figure 22.31 Tunable dielectric resonators disclosed in a 1987 patent by Wakino. (a) The tuning element is a piezoelectric actuator. (b) The tuning element is a piezoelectric attached to a tuning dielectric disk.

high Q/volume ratio in comparison with any other known filter technology. The first report on mechanically tuned dielectric resonator filters was by Wakino in 1987 [49]. Figure 22.31 illustrates two tunable dielectric resonator configurations disclosed in Wakino's patent [49]. In Figure 22.31a, a piezoelectric actuator is inserted on the top of a dielectric resonator, and the piezoelectric actuator acts as a tuning plate whose distance from the resonator is controlled by a dc applied voltage. Wakino reported a tuning range of 8% as the gap between the resonators is varied from 1 to 5 mm. In Figure 22.31b, a second dielectric disk attached to the piezoelectric actuator is brought in proximity of the dielectric resonator for tuning. Wakino [49] reported a 12 % tuning range with that approach over a 4 mm change in the gap between the dielectric resonator and the tuning dielectric disk. It should be mentioned that the same tuning range can be obtained with a smaller change in gap by placing the actuator closer to the resonator. This, however, would have an impact on the resonator Q. With the availability of EM simulation tools such as HFSS, tunable filter designers can easily optimize the resonator/actuator position for optimal tuning range and Q value.

Over the past years, several publications have addressed the tuning range of dielectric resonator filters. The concept of using a double resonator to improve the tuning range was implemented in [50, 51]. Since the field is mainly concentrated inside the dielectric resonator, the use of dielectric materials of the same dielectric constant or higher is an effective way to perturb the field in the original resonator and hence change its resonant frequency. In [50] a dielectric resonator is brought in proximity to the main dielectric resonator for tuning, whereas in [51] a dielectric resonator of a smaller diameter is inserted in the main dielectric resonator as shown in Figure 22.32. The concepts in [50, 51] were demonstrated using manual tuning with no integration of mechanical tuning elements, presenting experimental results for dielectric resonator filters with a 4–5% tuning range.

A cone-shaped dielectric resonator [52] was used to realize a tunable dielectric resonator filter that tunes in both center frequency and bandwidth. An experimental filter was demonstrated where the bandwidth was tuned from 5 to 20 MHz and the center frequency was tuned from 1930 to 1960 MHz with an achievable Q value of 16,000. A 200 MHz tuning range (1965–2165 MHz) was also demonstrated. Figure 22.33 shows a 7-pole filter version employing

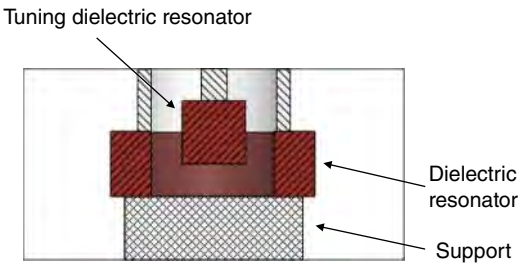
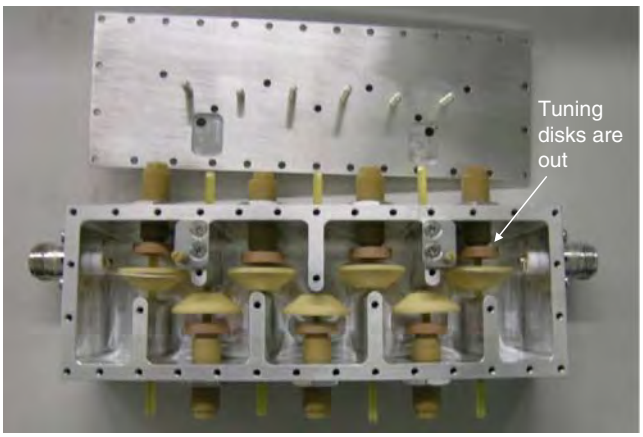
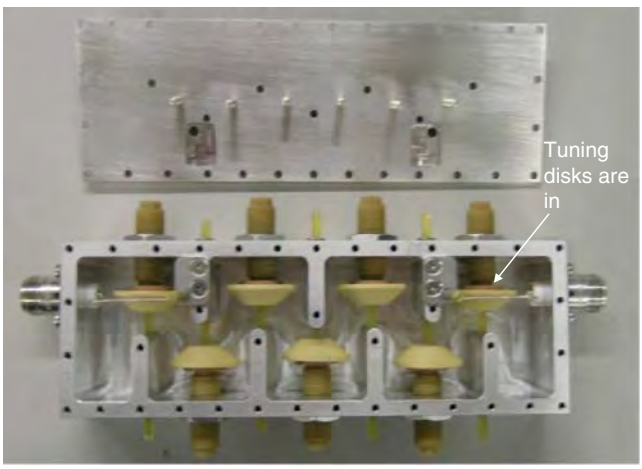


Figure 22.32 The use of a second smaller dielectric resonator to improve the tuning range.



(a)



(b)

Figure 22.33 A cone-shaped dielectric resonator filter tuned manually using dielectric plugs (other dielectric resonators). (a) The dielectric plug is outside the shaped cone resonator. (b) The dielectric plug is inside the cone-shaped resonator. (*Source*: Courtesy: L. Pance [53].)

the concept [53]. Dielectric plugs (basically, dielectric resonators of smaller dimensions) are used to tune the resonant frequency of the cone-shaped hollow resonator. The dielectric plugs are brought in and out of the cone-shaped resonator by screws. Figure 22.33a and b shows the filter with the dielectric plugs out and in, respectively. As the dielectric plug is moved inside the cone-shaped resonator, it shifts its frequency down. The coupling between the resonators is varied through the use of tuning screws located between the cone-shaped resonators. The reason for using a cone-shaped resonator rather than a cylindrical-shaped resonator [52] is to improve the filter's spurious performance. Figure 22.34 illustrates the experimental results obtained. Excellent experimental results were obtained using this concept. The results show a filter having a 20 MHz bandwidth tuned by 50 MHz in center frequency. The filter is retuned to have a bandwidth of 5 MHz and is then tuned in center frequency by 50 MHz. The structure allows manual tuning of both bandwidth and center frequency. The filter is useful in reducing manufacturing cost, by having one design that can be configured to any desired frequency around 2 GHz or any bandwidth when needed by manual tuning. With tuning screws replaced with motors, the filter can be made electronically tunable.

Mechanically tunable dielectric resonator filters using piezoelectric actuators were reported in [54–57]. Figure 22.35 illustrates the 4-pole filter structure reported in [56]. The filter consists of two dielectric resonators operating in dual mode coupled by an iris. Tuning of the resonator is achieved by having a metallic disk suspended above the resonator through a metal rod connected to a piezoelectric actuator located outside the cavity. A 2-pole version of the filter was demonstrated in [55] using a single dielectric resonator operating in a dual mode as shown Figure 22.36a. The experimental results obtained are shown in Figure 22.36b. The filter tunes over 2.25–2.45 GHz, demonstrating nearly 8% tuning range.

A piezoelectric tunable dielectric resonator filter operating in a TM mode is demonstrated in [58]. This TM mode is also referred as TME mode. The resonator is quite small in size and is configured to enhance tunability. Figure 22.37 illustrates the dielectric resonator configuration [58]. The resonator is mounted directly on the bottom cavity wall, forcing the TME mode to be the dominant mode. As a result, the resonant frequency shifts down significantly while maintaining a relatively acceptable high Q value (>2000), even when the gap between the dielectric resonator and the cavity top cover is very small. Higher Q values can be achieved with the use

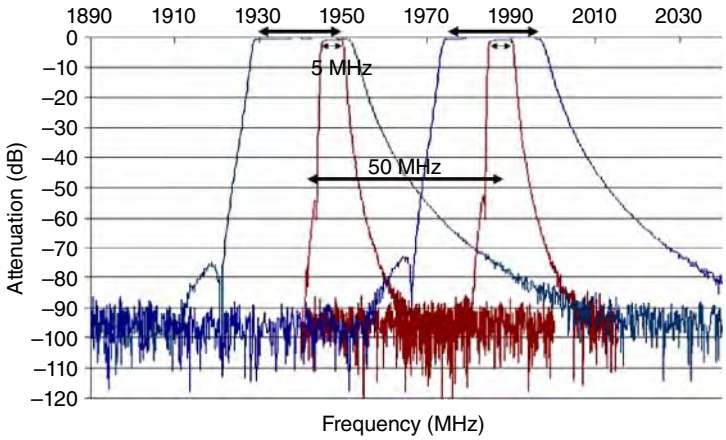


Figure 22.34 The experimental results of the cone-shaped resonator filter. (*Source*: Courtesy: L. Pance [53].)

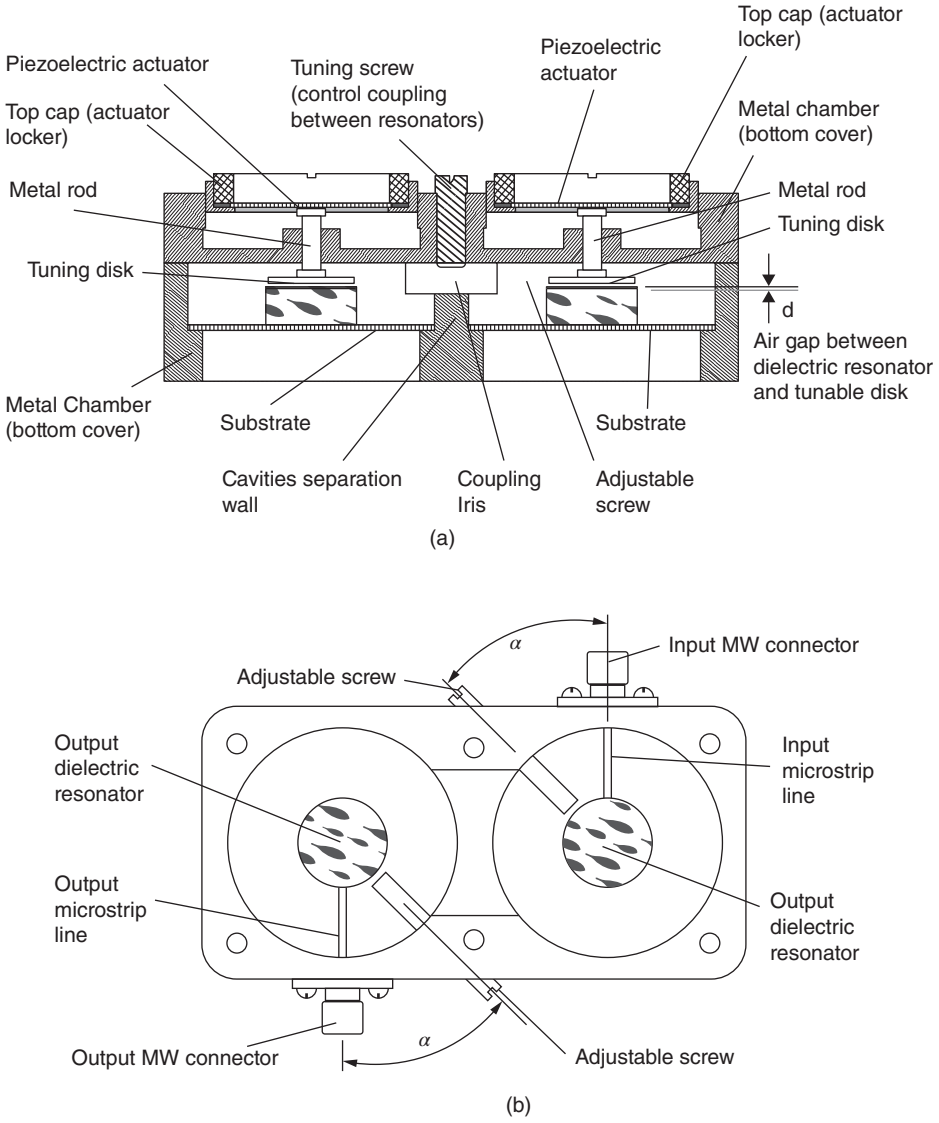


Figure 22.35 A dual-mode tunable dielectric resonator filter, tuned by piezoelectric actuators. (a) Side view. (b) Top view. The screw at the α angle is used to couple the dual modes. (*Source*: Alford *et al*. 2005 [56]. Reproduced with the permission of Taylor & Francis.)

of larger cavities. Figure 22.38a and b shows the HFSS simulated results for the tuning range and the Q values of the resonator as the gap between the top cover and the resonator is varied from 0.1 to 2 mm. It can be seen that more than a 70% theoretical tuning range can be achieved with this type of resonator [58].

Tuning of the dielectric resonator of [58] is achieved using a piezoelectric concept, where a thin metallic copper sheet is used to cover the top of the cavity that houses the TM dielectric resonator. The piezoelectric actuator is attached to the filter housing and is placed in contact with the copper sheet, as shown in Figure 22.39. With application of dc voltage to the piezoelectric actuators, the copper film, which covers the top of the cavity, is deflected by the force

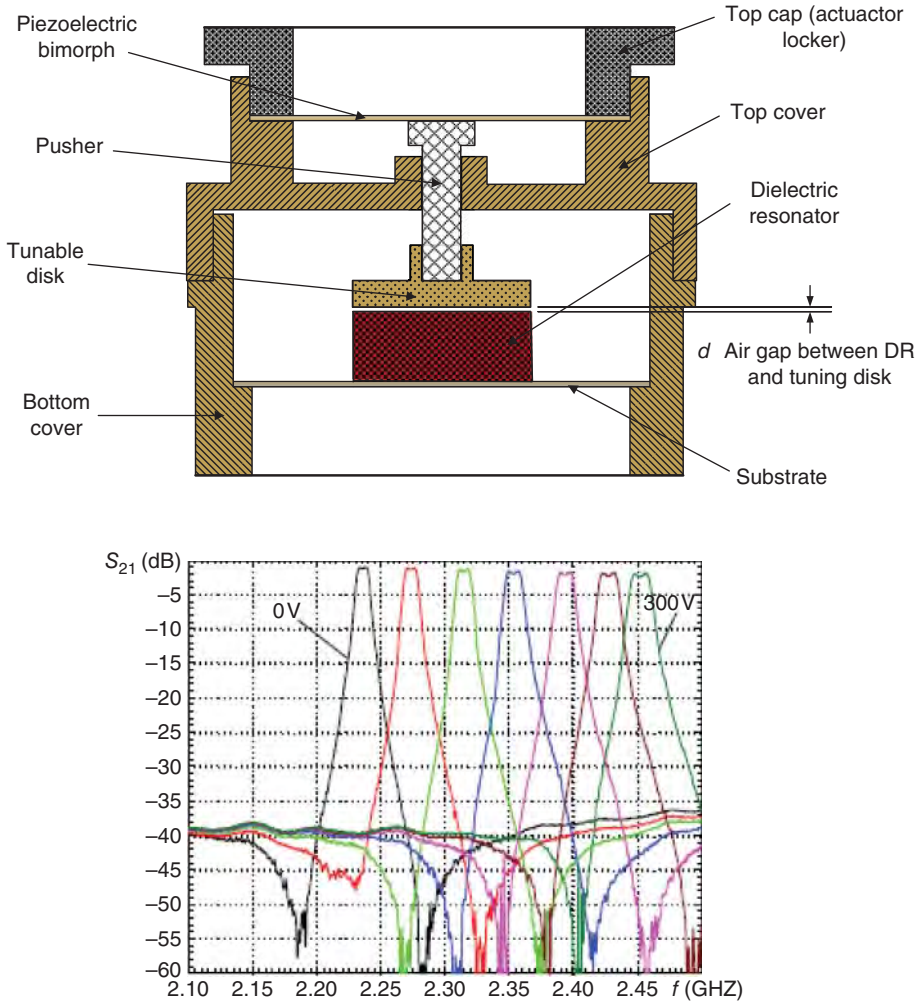
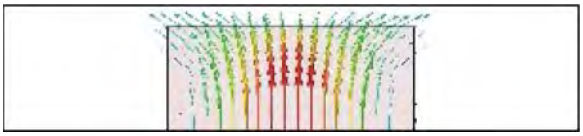


Figure 22.36 A 2-pole dual-mode dielectric resonator filter implemented using piezoelectric actuators. (*Source*: Petrov *et al*. 2004 [55]. Reproduced with the permission of Taylor & Francis.)

Figure 22.37 A dielectric resonator operating in the TM mode.



applied on its surface. Figure 22.39a illustrates a side view of the deflected cover of the cavity. The piezoelectric used only exerts a small force on the thin metallic copper sheet, causing it to deflect by 100–200 μm. From Figure 22.38, it can be seen that such a small gap change can cause a relatively large change in frequency when the starting gap between the top cover and the dielectric resonator is small; however, the Q drops sharply when operating at that range.

A 4-pole tunable dielectric resonator filter was built using this concept [58]. It consists of coupled four dielectric resonators; each is integrated with a thin film piezoelectric transducer. The resonator is cut from a low-loss high K-dielectric substrate of thickness 2.54 mm, $\varepsilon_r = 45$, and a loss tangent of 10^{-5}. Figure 22.40 shows the HFSS simulation results of the 4-pole tunable

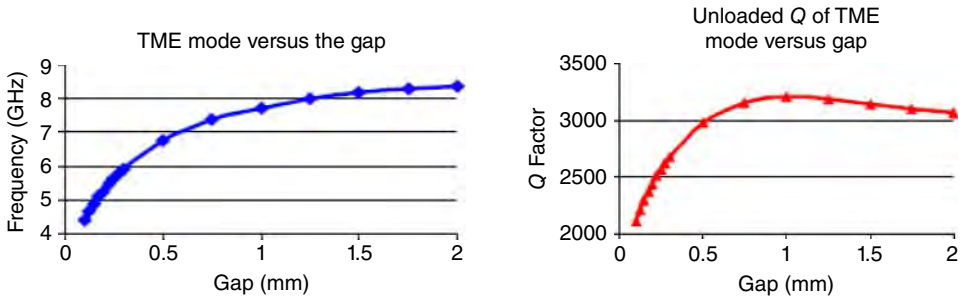
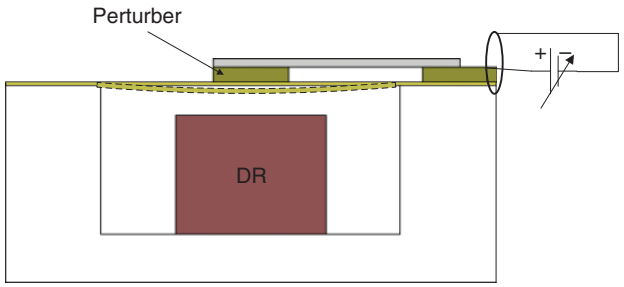


Figure 22.38 The tuning range and the unloaded *Q* versus gap for the dielectric resonator shown in Figure 22.37.



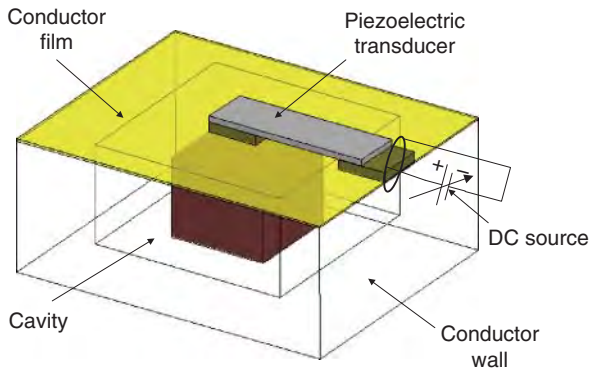


Figure 22.39 Piezoelectric tuning mechanisms. (*Source*: Huang and Mansour 2009 [58]. Reproduced with the permission of IEEE.)

filter when the gap between the top cover and the resonator is changed by only 20 μm. With no tuning and an initial gap of 160 μm, the filter has a center frequency of 5 GHz and a bandwidth of 50 MHz. It exhibits an insertion loss of 1.14 dB and a return loss of 17.5 dB. When the gap is reduced to 140 μm, the filter's theoretical insertion loss has increased to 1.75 dB with. The center frequency has shifted to 4.83 GHz, a 170 MHz shift in frequency. It should be noted that the filter is capable of achieving a wider tuning range; however, since no tuning was used for the coupling between resonators, the return loss degrades over a wider tuning range. Therefore the tuning range quoted in [58] is the range that can provide reasonable return loss.

Figure 22.41 shows a picture of the fabricated filter. The filter is first tested with a regular top cover to evaluate its RF performance. Figure 22.42 shows the experimental results of the filter

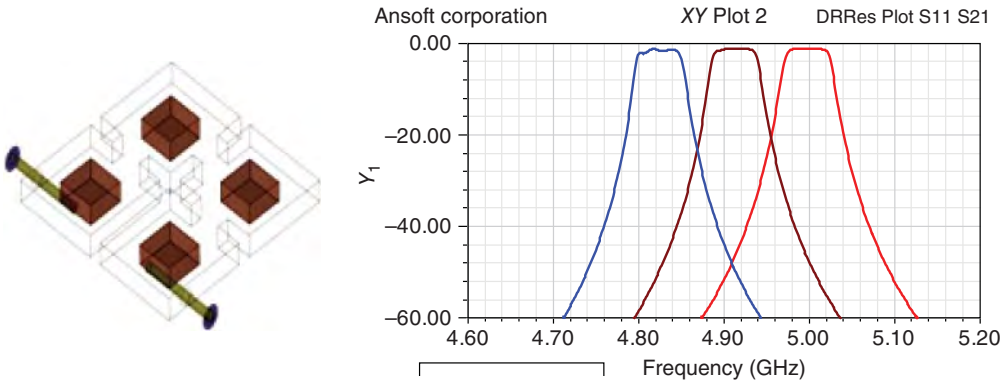


Figure 22.40 Filter schematic and HFSS simulation results.

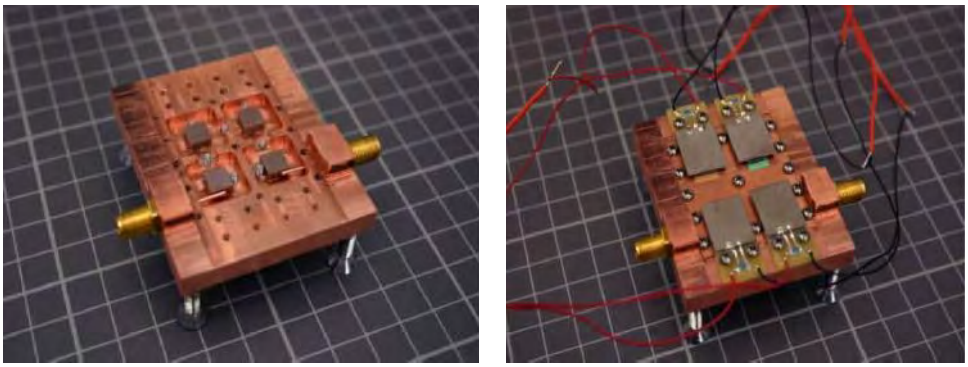


Figure 22.41 Picture of dielectric resonator filter tuned by piezoelectric actuators. (*Source*: Huang and Mansour 2009 [58]. Reproduced with the permission of IEEE.)

using a solid top cover (i.e., with no piezoelectric tuning). The filter exhibits an insertion loss of 2.7 dB. The tuning elements are then assembled and the filter was tested under various dc actuation voltages. While the theoretical Q is close to 2000, the experimental demonstrate a Q of 550. The relatively low Q obtained in this prototype unit is attributed to many factors including the use of high-loss adhesive materials to support the dielectric resonator in the cavity and the oxidation of the copper cavity and the input/output probes as well as the use of un-plated tuning screws.

In order to compare the performance of tunable dielectric resonator filters operating in the TM mode against evanescent-mode cavity filters having the same dimensions, Figure 22.43 illustrates a comparison between the Q factor of the two resonators as the gap between the top of the resonator and the cover is varied. It can be seen that for a small gap, the TME dielectric resonator offers a much high Q than that achieved by the evanescent-mode cavity resonator.

22.4.2 MEMS-Based Tunable Dielectric Resonator Filters

A MEMS-based tunable dielectric filter was reported in [10]. The basic building block consists of a dielectric resonator operating in a TE_{01} mode, a tuning disk, and multiple MEMS actuator. The MEMS actuators along with the disk are employed to replace the tuning screws typically used in conventional dielectric resonator filters. All the components are contained in a metallic cavity as shown in Figure 22.44. The tunability is achieved by moving the tuning disk along the

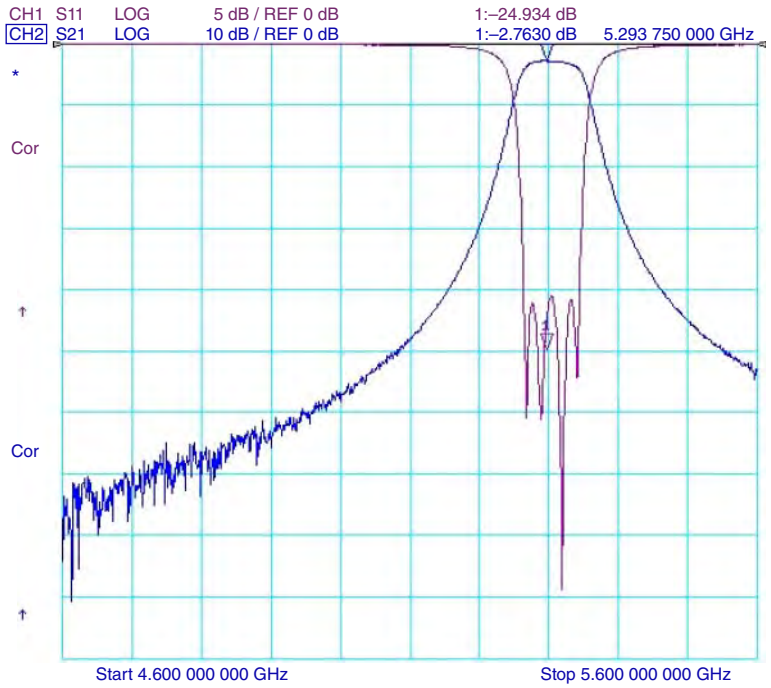


Figure 22.42 The measured dielectric resonator filter shown in Figure 22.41 with no tuning.

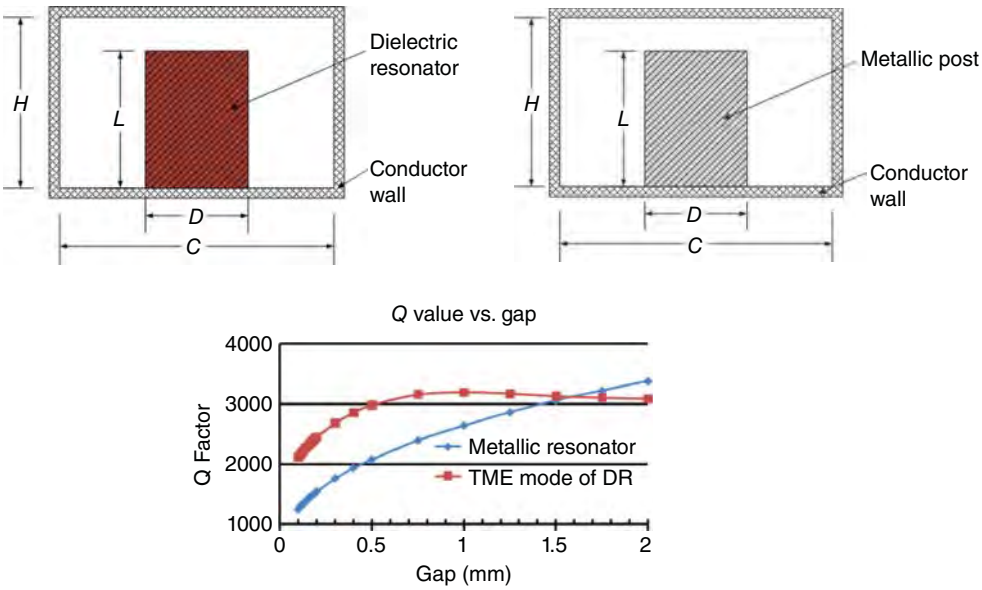


Figure 22.43 *Q* of a dielectric resonator operating in TME versus *Q* of an evanescent-mode resonator (metallic resonator) having the same dimensions.

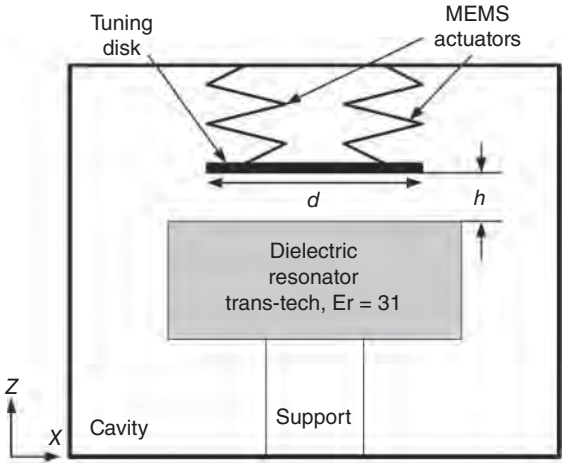


Figure 22.44 Schematic of the proposed tuning structure [10].

z axis with the use of the MEMS actuator. With no dc bias applied to the MEMS actuator, the tuning disk is at the closest position to the dielectric resonator, that is, the value of tuning gap h is at a minimum. This state corresponds to the highest resonant frequency. When a dc voltage is applied to the MEMS actuator, the tuning disk is pulled away from the dielectric resonator, and the resonant frequency begins to decrease.

EM (HFSS) and MEMS (CoventorWare) simulation tools were employed in [10] to determine the optimum tuning disk size for maximum tuning range and maximum quality factor while taking into account the structure's mechanical characteristics. The most important design parameters of this structure are tuning the gap h and the diameter of the tuning disk d, since they have the most effect on the tuning range and quality factor of the dielectric resonator. Two sets of simulations were performed to study the effects of these two parameters. Figures 22.45 and 22.46 illustrate the EM simulated variation in the frequency shift and quality factor as a function of the tuning gap and disk diameter, respectively. While a larger diameter disk will provide a wider tuning range, large tuning disks are subject to other design issues such as mechanical stability and warpage that will constrain the actuator design. Figure 22.47 shows the HFSS simulation results of a 3-pole tunable dielectric resonator filter. Circular tuning disks leaving a 3 mm diameter are employed in the simulation model. With a deflection of 0.7 mm provided

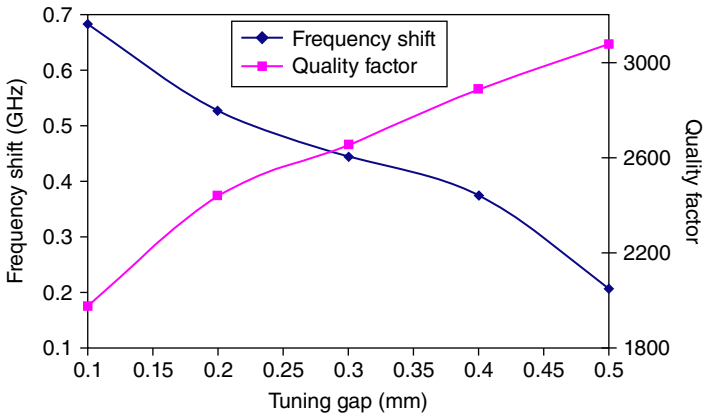


Figure 22.45 Frequency shift and quality factor as a function of the tuning gap when the tuning disk diameter is fixed.

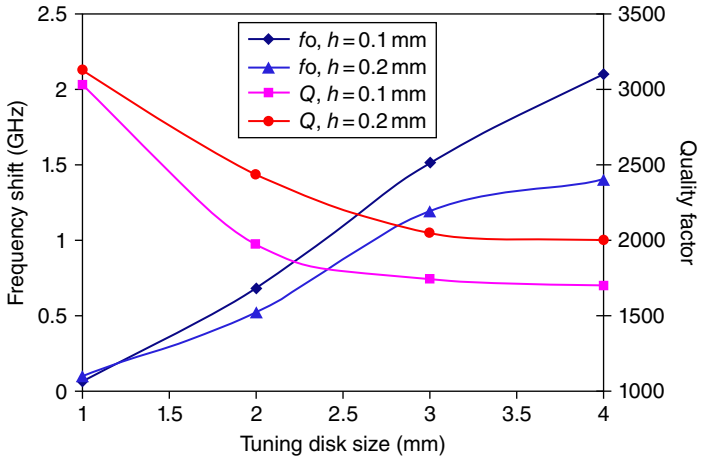


Figure 22.46 Frequency shift and quality factor as a function of the tuning disk diameter when the tuning gap is 0.1 and 0.2 mm.

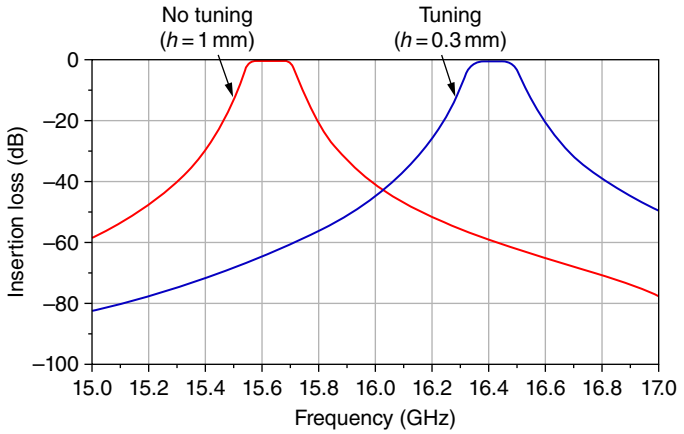


Figure 22.47 HFSS simulation results of a 3-pole dielectric resonator filter tuned to different frequencies (without loss).

by the MEMS tuning elements, the filter can be continuously tuned from 15.65 GHz to as high as 16.45 GHz, that is, a tuning range of 800 MHz.

In order to achieve a wide tuning range, it is essential to design a MEMS actuator that can generate enough force to lift a relatively large tuning disk as well as to generate a large deflection. In addition, a low actuation voltage and fast tuning speed are desirable. The concept of thermal plastic deformation assembly was adopted in [16] to design the MEMS tuning element with a vertical displacement for this filter. It is demonstrated in [16] that thermal actuator can lift a disk of size 3000 μm × 3000 μm × 2 μm plate more than 1 mm in the z direction. To eliminate warpage a hexagonal plate rather than a circular plate was used [10] as shown in Figure 22.48.

Figure 22.49 illustrates the tunable dielectric resonator filter. It is manufactured into two detached parts: the top cover in Figure 22.49a and the body in Figure 22.49b [10]. These two parts are gold plated to reduce the conductive losses from the metal cavity. Dielectric resonators

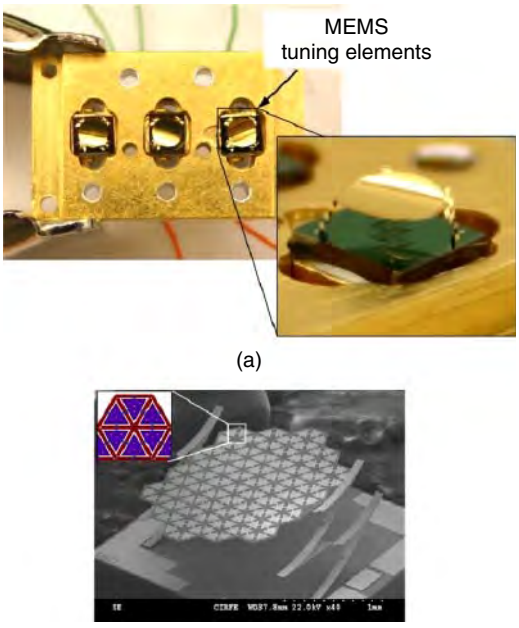


Figure 22.48 MEMS tuning elements: (a) a solid circular tuning disk with warpage and (b) a hexagonal tuning disk without warpage. (*Source*: Yan and Mansour 2007 [10]. Reproduced with the permission of IEEE.)

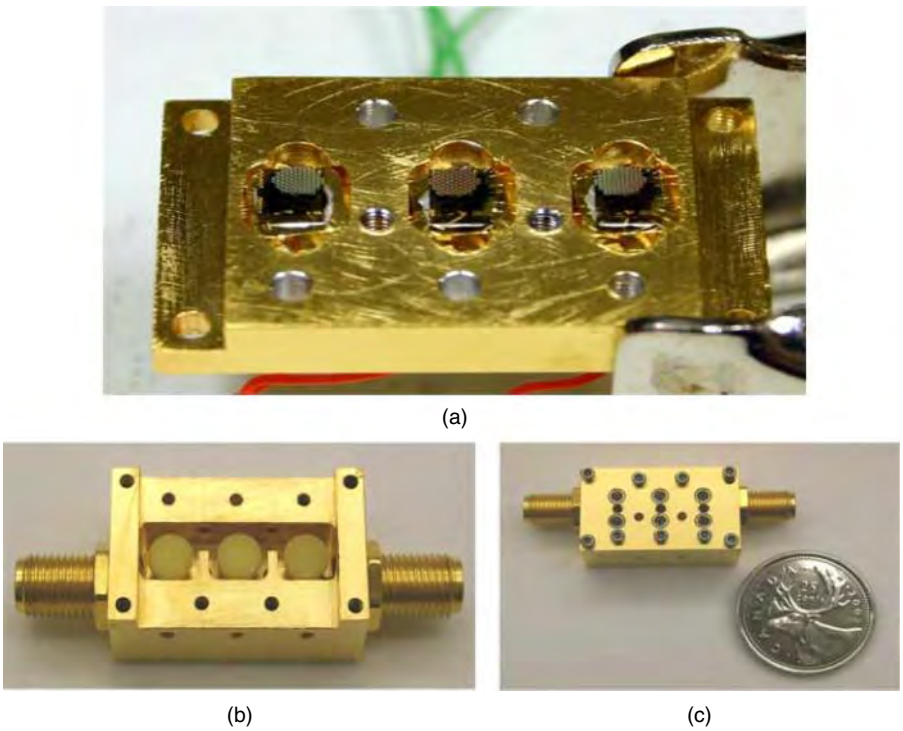


Figure 22.49 Proposed 3-pole tunable dielectric resonator filter: (a) the top cover with MEMS tuning elements, (b) the filter body, and (c) the assembled filter. (*Source*: Yan and Mansour 2007 [10]. Reproduced with the permission of IEEE.)

are assembled with the body and the MEMS tuning elements are integrated on the top cover. Each MEMS tuning element is connected to two dc feedthrough pins for applying the control voltage. This configuration allows each tuning element to be independently controlled. A few mechanical tuning screws are also included in the tunable filter design to compensate for the machining tolerance and variation of material properties.

The measured results for the insertion loss and the return loss response of the 3-pole tunable dielectric resonator bandpass filter are presented in Figure 22.50. The filter's center frequency is synchronously tuned from 15.6 to 16.0 GHz, demonstrating an increase in insertion loss from 1.5 to 4.5 dB. This high insertion loss is attributed to the assembly procedure and lossy materials (epoxy, stainless tuning screws, etc.) used to construct this prototype filter. Obviously, much better insertion loss performance can be achieved with further design optimization.

Another tunable dielectric resonator concept was reported in [57, 62]. The concept was demonstrated using piezoelectric material but can be implemented as well using MEMS technology. The tuning mechanism proposed consists of three main components: a dielectric resonator operating in the TE_{01} mode, a metal disk comprising a certain number

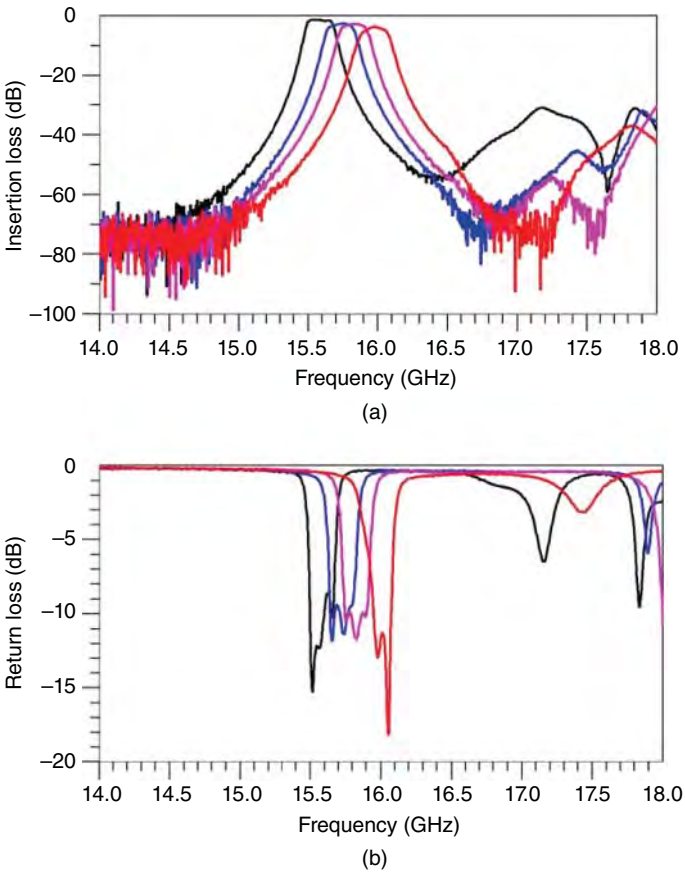


Figure 22.50 Comparison of the measured results of the proposed tunable dielectric resonator filter for different tuning states.

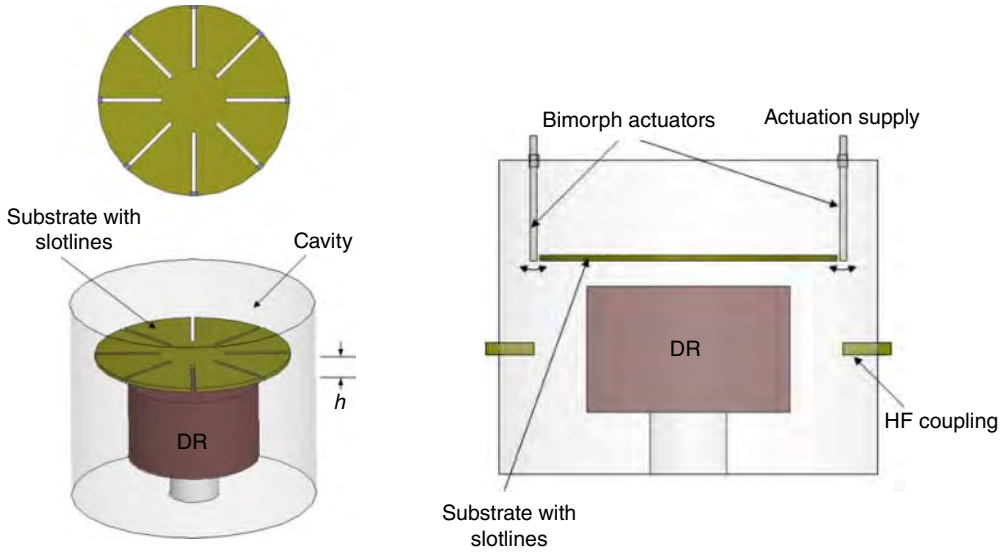


Figure 22.51 A tunable dielectric resonator using a disk integrated with switches. (*Source*: Panaitov *et al*. 2005 [62]. Reproduced with the permission of IEEE.)

of radial-arranged quarter-wave slotline resonators, and a switch for each slotline resonator, which then later is used to control the coupling between the TE_{01} mode and the slotline mode. The concept is illustrated in Figure 22.51. A substrate with radial slotlines is symmetrically arranged above a cylindrical dielectric resonator. Eight switches could be placed on the edge of the slots. With the switches in the off state, the eight slotlines represent capacitively loaded quarter-wave slotline resonators that possess a strong magnetic field component in the radial direction, providing strong coupling to the resonant fields of the mode. Due to the interaction between the dielectric resonator mode and the slotline resonator mode, the stored electromagnetic energy is distributed between the dielectric resonator and the slotlines. This leads to an increase of the effective size of the resonator and, as a consequence, to a decrease of the resonant frequency. The concept was demonstrated experimentally using the cavity shown in Figure 22.51b. The slotlines were switched ON and OFF using bimorph piezoelectric actuators. A tuning range of 5 MHz was achieved with a Q value of 12,000 at 2 GHz.

This concept was combined with the planar dielectric resonator filters to construct a tunable dielectric resonator filter. Figure 22.52 shows a 2-pole filter designed, fabricated, and tested [63]. The filter consists of two dielectric resonators formed from a high-K substrate. Two metal disks, attached to the top cover each with four slots, are placed in proximity to the resonators. In order to study the effects of the disk on the resonator Q, Table 22.3 compares the theoretical resonant frequency and Q of the resonator without the disk and with the disk with switches in the ON and OFF state. The switches here are simulated in HFSS using a metallic wire. The presence of the disk in proximity of the resonator reduces its Q by around 40%. The Q is further reduced when the switches are in the ON state. The fabricated unit is shown in Figure 22.53. The HFSS simulation results, as the edges of the slots are switched ON and OFF, are shown in

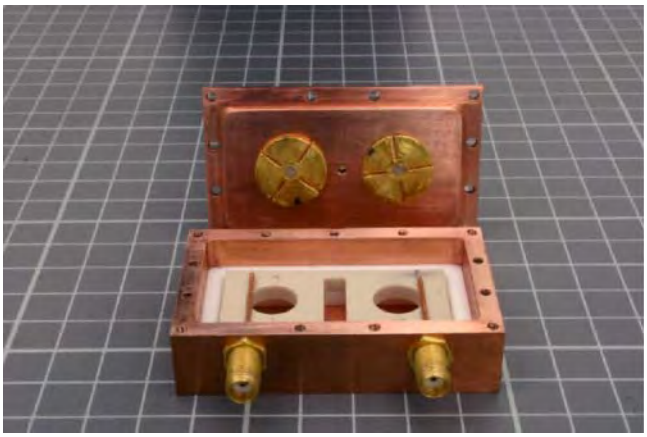


Figure 22.52 A 2-pole filter dielectric resonator filter with two disks integrated in the cover and with integrated switches. (*Source*: Mansour 2009 [63]. Reproduced with the permission of IEEE.)

Table 22.3 Comparison between the resonance frequency and *Q* of a resonator loaded with a disk.

	No Disk	With Disk switch is OFF	With Disk switch is ON
Frequency (GHz)	4.008	4.167	4.227
Q	6000	3730	2760

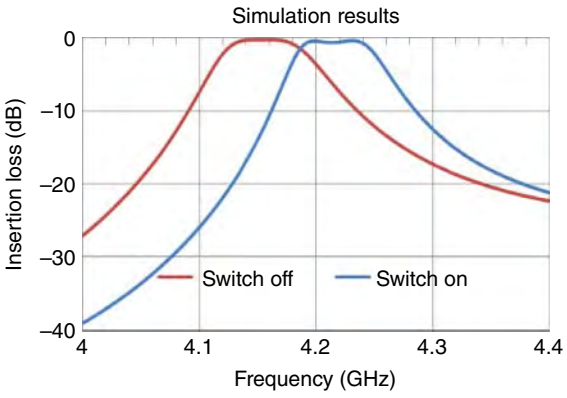


Figure 22.53 The HFSS simulation results of the filter shown in Figure 22.52.

Figure 22.53; a 60 MHz tuning range has been achieved. A larger tuning range can be achieved by bringing the disk closer to the resonator, but this will lower the Q value.

The realization of tunable dielectric resonators that are tuned by MEMS switches and semiconductor tuning elements requires integrating the tuning element with the DR resonators in a way to interact and perturb the EM field. The two common single modes for DR resonators widely used in filter applications are TME and TEH modes. While TME modes offer lower Q values in comparison with THE modes, the field distribution of TME makes it easy to integrate MEMS or the semiconductor tuning element with the DR resonator. In [11] a conductive strip loaded with MEMS switches is placed inside a dielectric resonator operating in the TME mode as shown in Figure 22.54a. The tunable resonator consists of a dielectric resonator with hole in the center mounted on a Teflon support and is inserted inside a metal enclosure. Dielectric resonator with holes in the middle are traditionally used to improve the spurious performance of DR resonators operating in the TEH modes. In this case the hole is used to provide the room to place the tuning circuit inside the resonator. The idea of tuning here is to have a metallic strip whose surface current interacts strongly with the field distribution of the TME mode. The tuning circuit, as shown in Figure 22.54b, consists of three conductive strips made of gold on a 625μm alumina substrate. The three strips have different lengths (L_1, L_2, and L_3) separated by a 100 μm gap from each other. Two sets of contact-type RF MEMS switches are utilized over the 100 μm gap between these conductive strips. Figure 22.55 shows an SEM picture of the MEMS switches used. The switches were fabricated at the University of Waterloo using the UW-MEMS process [83].

Each set of MEMS switches has a separate bias line connected to bias pads outside the cavity and can be actuated independently. When the bias voltage on the MEMS switches is zero, the effective length of the conductive strip is L_1. When the first set of switches are actuated to the down state, the second segment of conductive strip is electrically connected to the first strip, and the total length of the strip is increased to $L_1 + L_2$, causing EM field perturbation and hence tuning the resonance frequency of the resonator. A third state can be achieved by

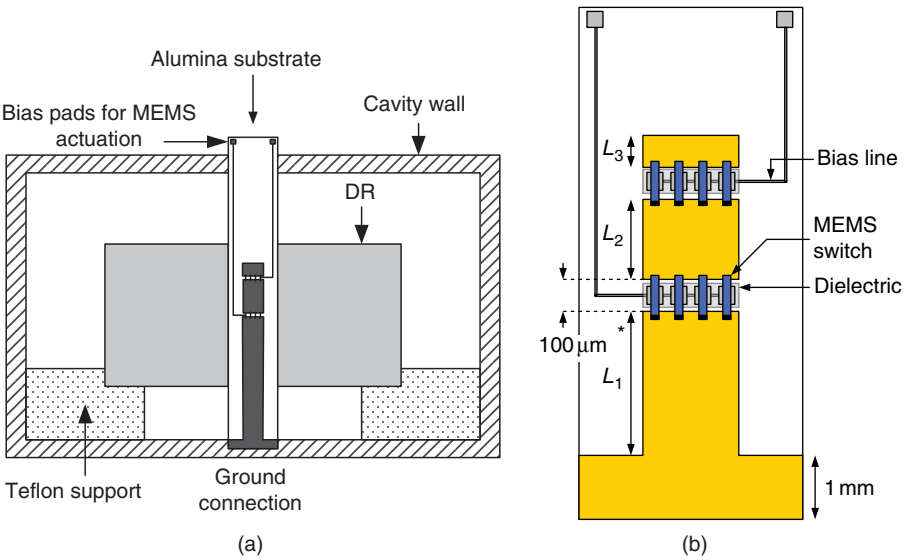


Figure 22.54 (a) A dielectric resonator integrated with a tuning strip loaded with MEMS switches and (b) details of the tuning circuit. (*Source*: Huang *et al*. 2011 [11]. Reproduced with the permission of IEEE.)

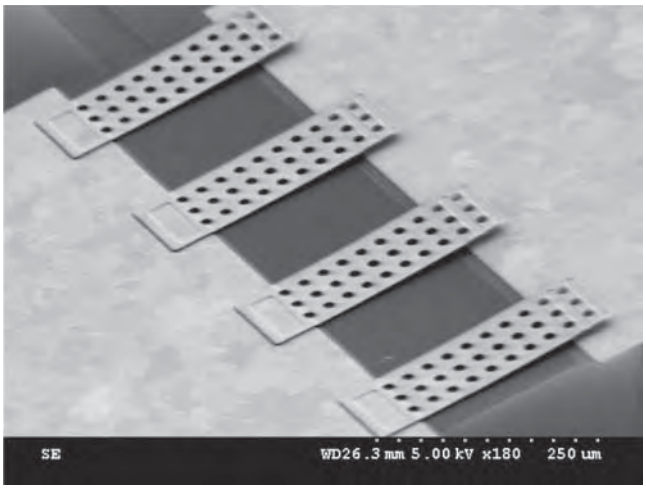


Figure 22.55 A SEM picture of the MEMS switches used in Figure 22.54. (*Source*: Huang *et al*. 2011 [11]. Reproduced with the permission of IEEE.)

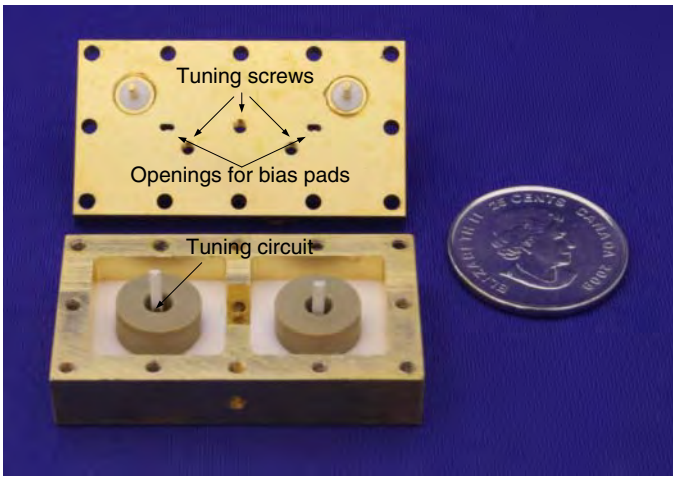


Figure 22.56 A picture of a 2-pole MEMS-based tunable filter. (*Source*: Huang *et al*. 2011 [11]. Reproduced with the permission of IEEE.)

simultaneously actuating both sets of the MEMS switches. A 2-pole filter operating at around 4.8 GHz with a 20 MHz bandwidth was demonstrated in [11] using this concept. Figure 22.56 shows the fabricated tunable filter unit, whereas Figure 22.57 shows its measured results in the three states. The return loss is better than 20 dB over a tuning range of 160 MHz from 4.64 to 4.8 GHz. The Q ranged from 1220 to 510. It should be noted that the DR resonators used for that structure were from old dielectric resonator stock with a relatively low $Q \times f = 40{,}000$. The use of dielectric resonator materials with $Q \times f = 100{,}000$ should provide higher Q values than that reported in [11].

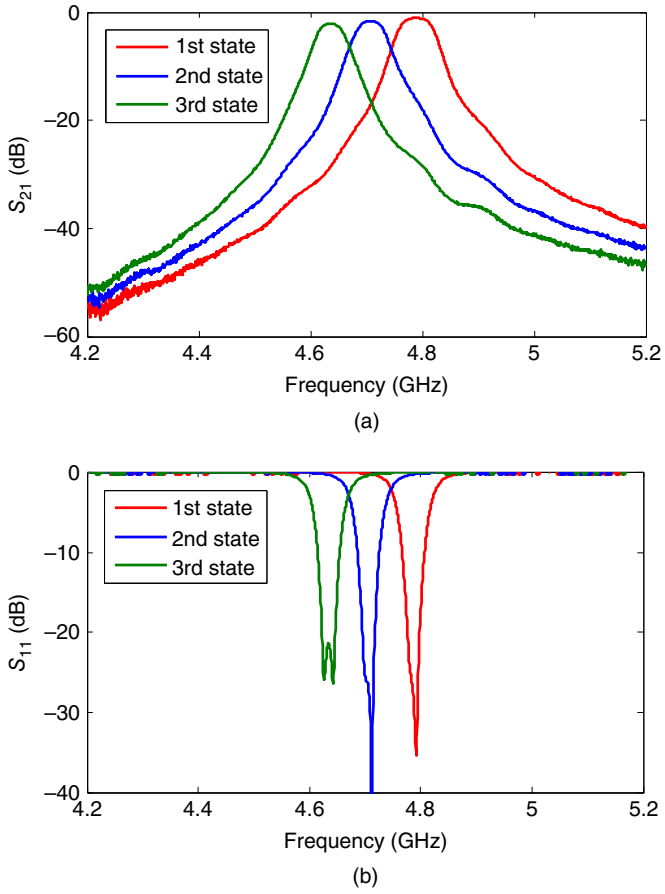


Figure 22.57 The measured results of the MEMS-based tunable dielectric resonator shown in Figure 22.56. (*Source*: Huang *et al*. 2011 [11]. Reproduced with the permission of IEEE.)

To obtain continuous tuning, semiconductor varactors can be used to replace the RF MEMS switches. Figure 22.58 shows the schematic diagram of the tuning circuit with the varactor elements [11]. It consists of two gold strips of different lengths on an alumina substrate connected by a semiconductor varactor. A commercial GaAs varactor with a capacitance ranging from 1.1 to 0.32 pF with a reverse bias voltage varying from 0 to 10 V is used to connect the two conductive segments. The first conductive strip is placed inside the hole within the DR, whereas the other conductive strip makes contact with the cavity lid and establishes a ground connection. The measured results achieved are shown in Figure 22.59. The filter demonstrated an insertion loss of 1.17 and 4.6 dB for a bias voltage of $V_{\text{bias}} = 10$ V and $V_{\text{bias}} = 0$ V, respectively. The high insertion loss is attributed to the lower Q value of the varactor.

In order to improve the loss performance of the tunable DR filter while achieving a continuous tuning (or more tuning states), a tuning circuit utilizing 4-bit RF MEMS switched capacitor bank is also employed in [11]. Figure 22.60 shows a schematic diagram of the tuning circuit. It has two conductive strips: one connected to the cavity wall and the other placed inside the hole

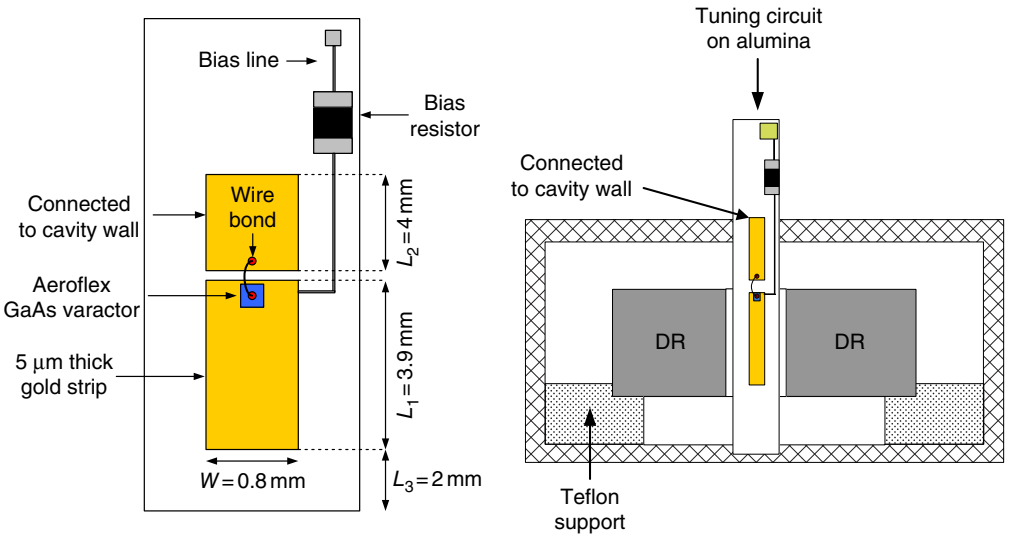


Figure 22.58 A dielectric resonator tuned by a varactor. (*Source*: Huang *et al*. 2011 [11]. Reproduced with the permission of IEEE.)

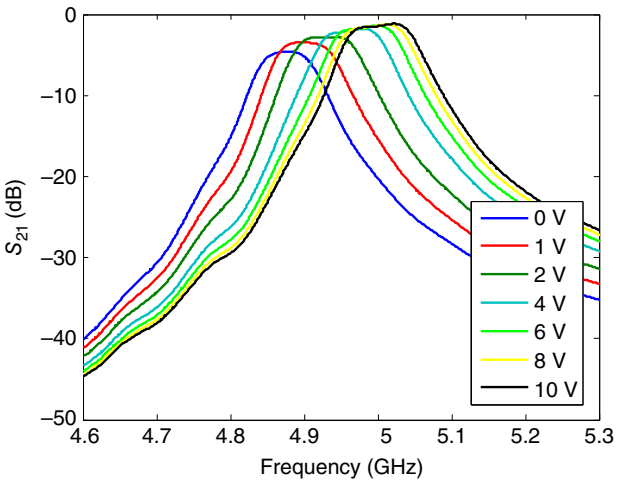


Figure 22.59 The measured results of a 2-pole varactor-based tunable dielectric resonator filter.

within the DR. An SEM image of the 4-bit RF MEMS capacitor bank and its equivalent circuit are given in Figure 22.61. It consists of four metal–insulator–metal (MIM) capacitors C1, C2, C3, and C4 and four capacitive RF MEMS switches with up-state and down-state capacitance values of Cu and Cd, respectively. This 4-bit RF MEMS circuit was fabricated at the University of Waterloo using the UW-MEMS process [83]. The capacitance value changes from 347 to 744 fF. The fabricated filter unit is shown in Figure 22.62, whereas the measured results for the 16 tuning states are given in Figure 22.63. A tuning range from 5.20 to 5.02 GHz (180 MHz) is achieved, with a return loss close to 15 dB. The maximum insertion loss is better than 2.8 dB.

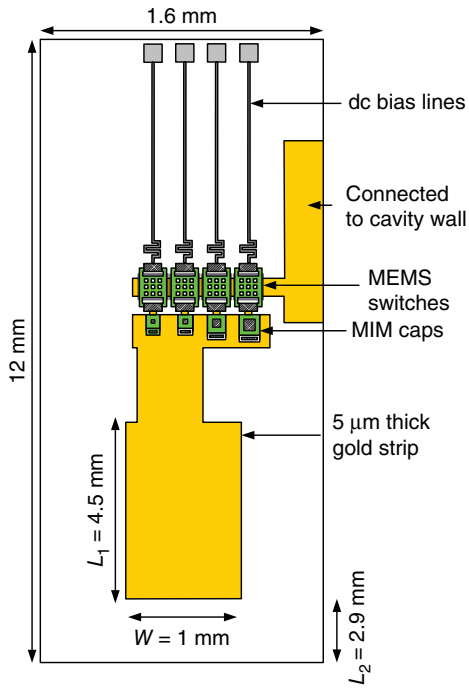


Figure 22.60 A circuit employing 4-bit RF MEMS switched capacitor for tuning a dielectric resonator filter. (*Source*: Huang *et al*. 2011 [11]. Reproduced with the permission of IEEE.)

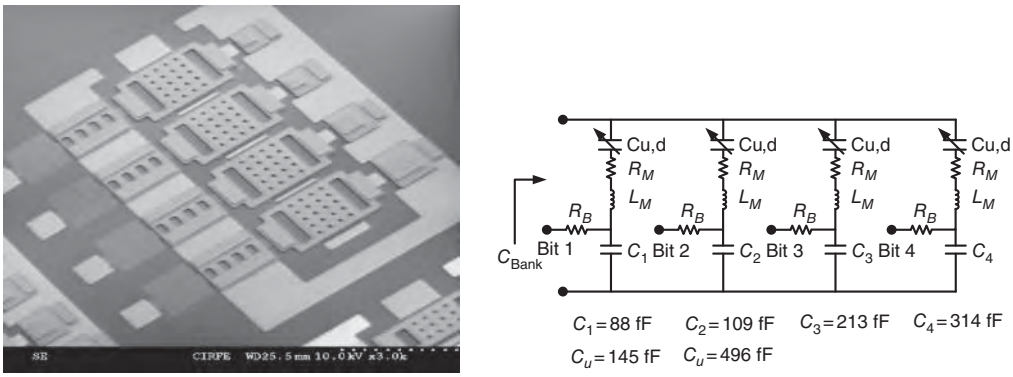


Figure 22.61 A SEM picture of the 4-bit RF MEMS switched capacitor bank and its equivalent circuit. (*Source*: Mansour 2009 [1]. Reference [83]. Reproduced with the permission of IEEE. Reproduced with the permission of University of Waterloo.)

The measured Q value changes from 800, when all of the MEMS switches are up, to 550, when all of the MEMS switches are actuated to the down-state position.

A concept similar to what used in [12] for combline filters is proposed in [14] to realize tunable dielectric resonator filters. Figure 22.64 shows the 3D view of the tunable dielectric resonator. It consists of a dielectric resonator placed in the center of a metal cavity and is supported by a thin Teflon disk. The metal disk at the bottom of the cavity is used to align the dielectric resonator during the assembly process. The cavity is covered by a patch-printed PCB board. A disk patch

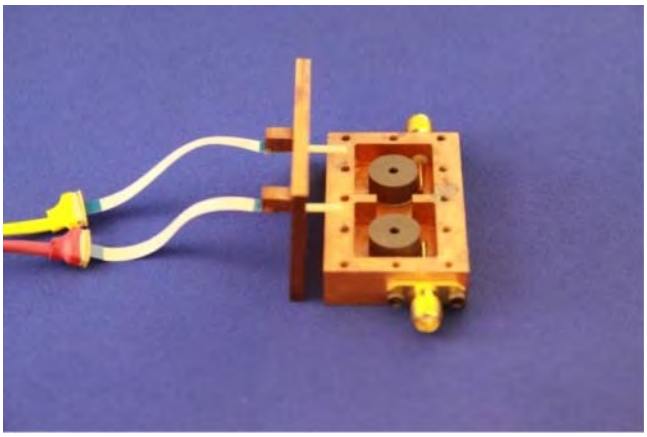


Figure 22.62 A picture of a 2-pole tunable MEMS-based tunable dielectric filter. (*Source*: Huang *et al*. 2011 [11]. Reproduced with the permission of IEEE.)

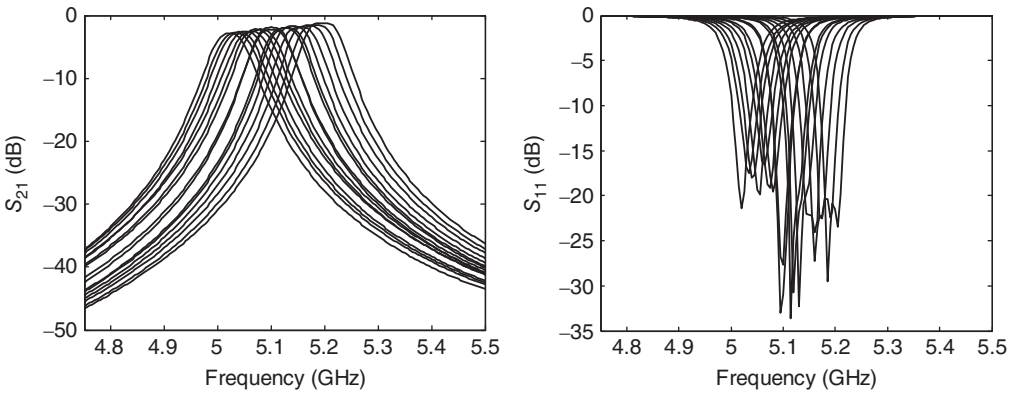


Figure 22.63 The measured results of the 16 tuning states of the MEMS-based tunable dielectric resonator shown in Figure 22.62. (*Source*: Huang *et al*. 2011 [11]. Reproduced with the permission of IEEE.)

is printed on the bottom surface of the PCB board and linked to the top patch of the PCB board through a via hole. A slot is made around the via hole to isolate it from the grounded top patch. A varactor is placed over the slot, and a high value dc bias resistor is used to avoid RF leakage while applying dc voltage to the varactor. A 2-pole 5 GHz tunable filter was demonstrated using this concept, exhibiting a 300 MHz tuning range [14].

22.5 Waveguide Tunable Filters

A waveguide MEMS-based tunable filter was reported in [42]. The filter uses RF MEMS switches placed between conductive lines printed on a substrate, which is mounted over the sidewall of a rectangular waveguide in a way such that the conductive lines are in contact with the waveguide broad walls. Figure 22.65 illustrates schematics for the concept. Opening and closing the RF MEMS switches effectively realizes a waveguide with an equivalent movable sidewall, which in turns varies the resonant frequency of the waveguide cavity. A tunable

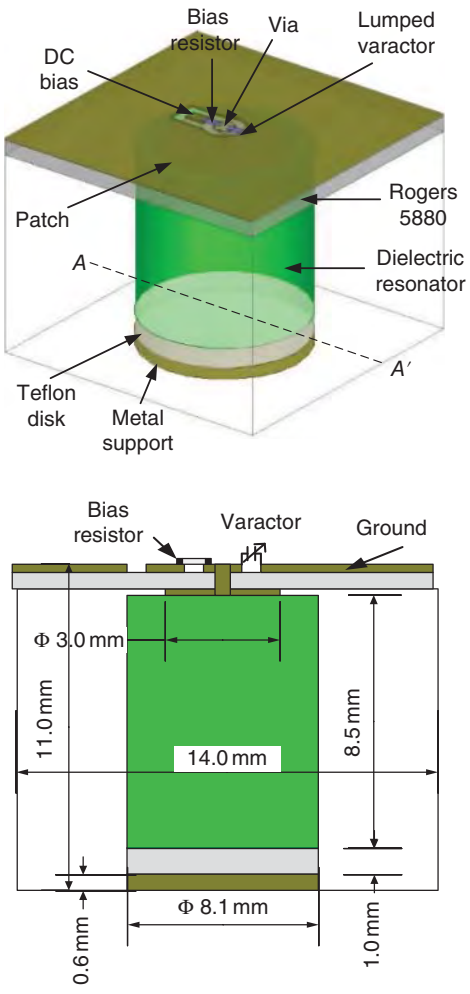


Figure 22.64 A dielectric resonator cavity tuned by a varactor. (*Source*: Huang and Mansour 2013 [14]. Reproduced with permission of IEEE.)

E-plane waveguide filter operating at 23 GHz was reported in [42], demonstrating Q values between 750 and 1450 and a tuning range of 3.1%.

MEMS switches were also employed in [43] to realize a switchable bandstop waveguide filter. The concept is based on MEMS switched planar resonators placed on the sidewall of a rectangular waveguide. Figure 22.66 shows a waveguide loaded with five planar resonators. It also shows a schematic of the planar resonator. With the MEMS switches in the OFF state, the planar resonators resonate at a frequency outside the waveguide band and have a little impact on the waveguide S_{21} response. On the other hand, having MEMS switches in the ON state brings the resonance frequency of the planar resonators within the waveguide band, which results in a stopband in the waveguide S_{21} response around the resonance of the planar resonators. The design allows the realization of multiple stopbands, which can be switched in and out of the waveguide band. Each stopband is controlled by a set of planar resonators designed to operate at that specific band. Figure 22.66b shows the measured results obtained for a Ku-band bandstop filter employing five planar resonators as the planar resonators switched to resonance within the waveguide band.

Mechanically tunable waveguide filters operating in the TE_{011} modes were reported in [45, 46]. The tuning is realized using movable piston with a diameter that is very close to the

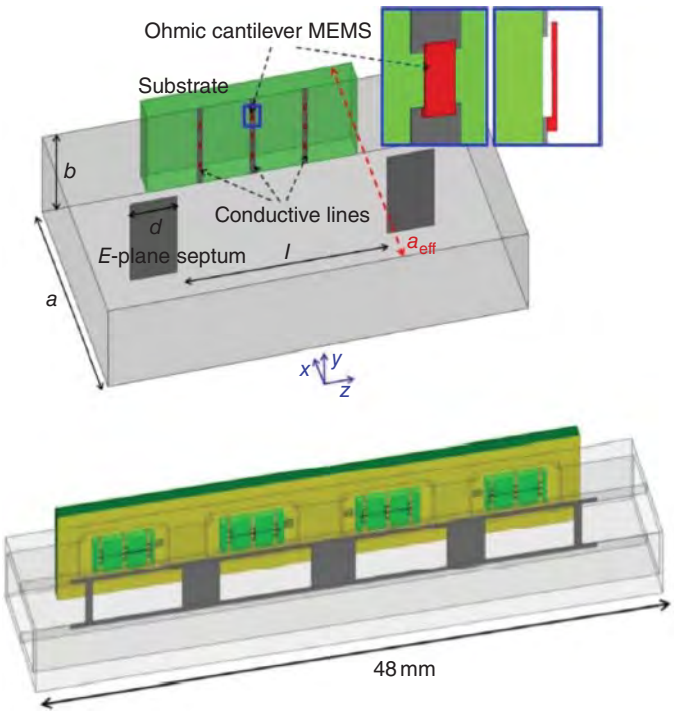


Figure 22.65 MEMS-based tuning concept proposed in [38] is applied to a waveguide rectangular cavity coupled by E-plane conductive septa. (*Source*: Pelliccia *et al*. 2016 [42]. Reproduced with the permission of IEEE.)

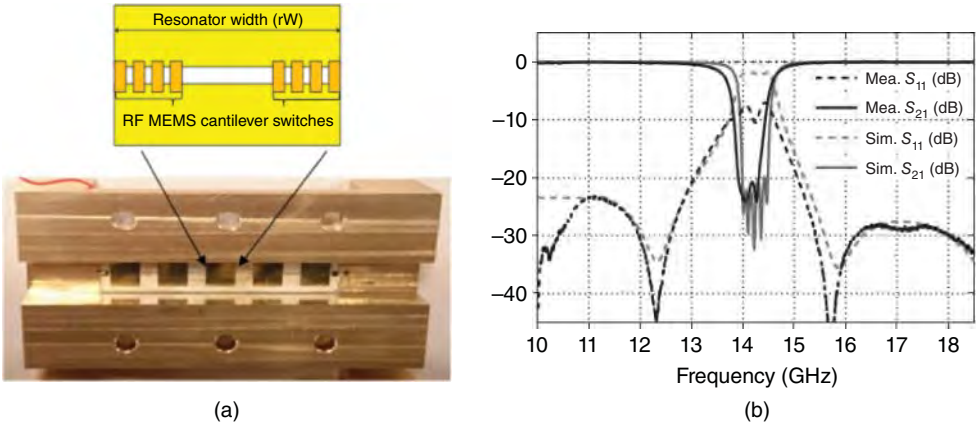


Figure 22.66 (a) Pictures of the fabricated WR62 rectangular waveguides with the RF MEMS switchable resonators and (b) the measured results when the MEMS switches are in ON state. (*Source*: Chan *et al*. 2016 [43]. Reproduced with the permission of IEEE.)

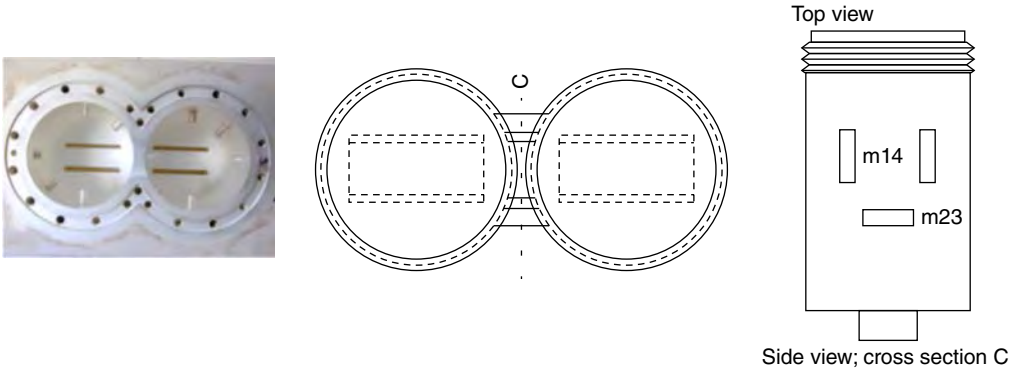


Figure 22.67 Iris configurations for bellows-mounted tunable filter design. (*Source*: Yassini *et al*. 2009 [25]. Reproduced with the permission of IEEE.)

cavity diameter. The field distribution of the TE_{011} mode allows for a non-contacting piston because the wall currents are purely circumferential. The main disadvantage however of operating in TE_{011} modes is the narrow spurious band, which in turn limits the filter tuning range. Tunable filters based on cavities operating in the dual-mode TE_{113} were reported in [47], where tuning is performed by mechanically moving the end wall of each cavity. Measured results were presented for a 26 GHz tunable filter with a bandwidth of 32 MHz operating over a tuning range of 900 MHz and demonstrating an effective Q of 11,000.

A Ku-band high-Q tunable elliptic filter operating in TE_{113} modes employing a unique iris configuration to maintain a constant bandwidth over the tuning range was reported in [25]. Tuning is achieved through the use of flexible copper bellows mounted on the top of the cavity resonators, allowing the cavity length to be varied by pushing the bellow in and out of the cavity. Both the input/output and interresonator irises were designed with a set of coupling slots that are strategically placed to minimize variation of coupling over frequency. Figure 22.67 shows the filter configuration, while Figure 22.68 shows the results obtained. A tuning range of 3.3% was achieved.

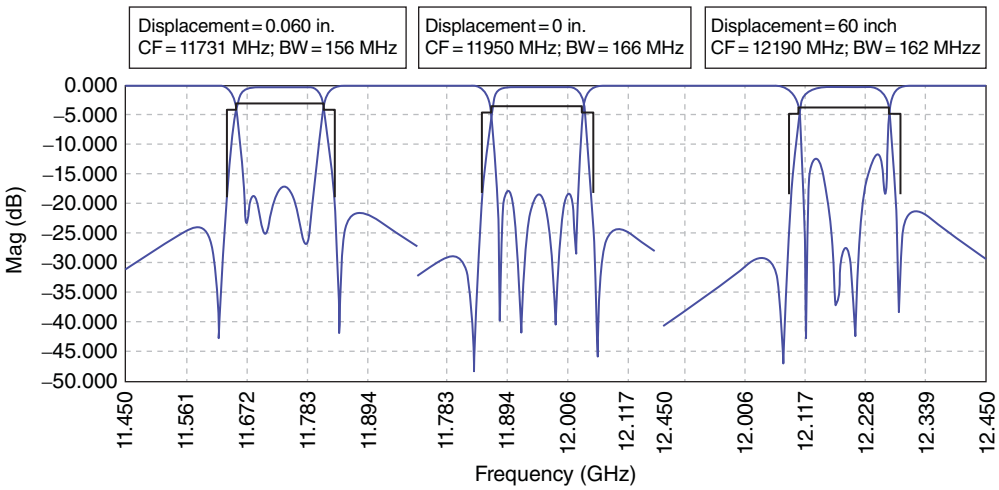


Figure 22.68 Measured tuning performance, provided by a 60 mil bellows displacement. (*Source*: Yassini *et al*. 2009 [25]. Reproduced with the permission of IEEE.)

22.6 Filters with Tunable Bandwidth

One approach to design filters that are tunable in both center frequency and bandwidth is to cascade two filters that are tunable in center frequencies as shown in Figure 22.69. These two filters could be a tunable lowpass filter and a tunable highpass filter, or simply two tunable bandpass filters. The overlapping between the passbands of the two filters can lead to a filter that is tunable in center frequency and bandwidth. An isolator is typically used between the two filters to reduce interaction between them. Two 6-pole combline tunable bandpass filters of the type presented in Section 22.3 were used in [18] to realize a tunable filter that is tunable in center frequency and bandwidth. The two tunable combline bandpass filters were designed to have a constant bandwidth over the tuning range. Figure 22.70 shows the measured results obtained. The resultant filter can have a bandwidth that can be tuned from 10 to 100 MHz with a center frequency that can be tuned from 1.185 to 2.2 GHz, that is, a bandwidth of 34%. In [22] two 6-pole tunable filters employing TE_{011} cavities were connected in cascade with an isolator in between to realize a tunable filter that is tunable in center frequency and bandwidth. The filters are tuned mechanically with plugs that are pushed in and out of the TE_{011} cavities. Figure 22.71 shows the results obtained [22].

While the use of an isolator helps to get good results for S_{21}, it can be seen that the input return of the tunable bandwidth filter is basically the return loss of the first filter in the cascade configuration. This results in a tunable bandwidth filter that is nonreciprocal with a wideband return loss that does not relate to the filter bandwidth. Having a nonreciprocal response precludes their use in multiplexer applications.

Investigations were carried out in [18] on replacing the isolator with a transmission line of fixed length and re-optimizing the resonance frequencies of two cascaded filters accordingly. The measured results obtained are shown in Figure 22.72. While such an approach yields a reciprocal response, it results in some degradation in isolation performance.

Another approach that only uses resonator tuning was reported in [36] for realizing combline filters with a tunable bandwidth. The idea is based on using a dual-band filter structure where resonator tuning is optimized to tune the first band of the dual band. This approach however has a limited tuning range.

The obvious approach to realize filters that are tunable in center frequency and bandwidth is to use tuning elements for the resonators and input/output and interresonator couplings. However, it is usually intricate to tune input/output and interresonator couplings, particularly

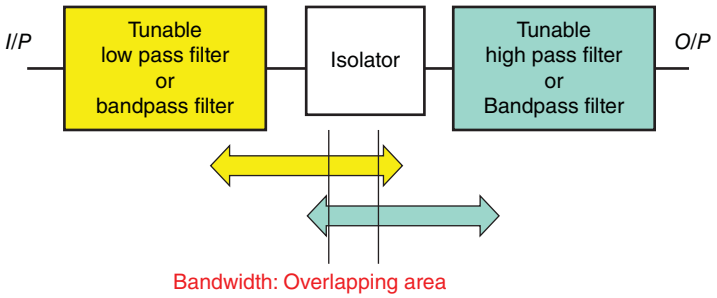


Figure 22.69 A configuration to realize filters that are tunable in both center frequency and bandwidth.

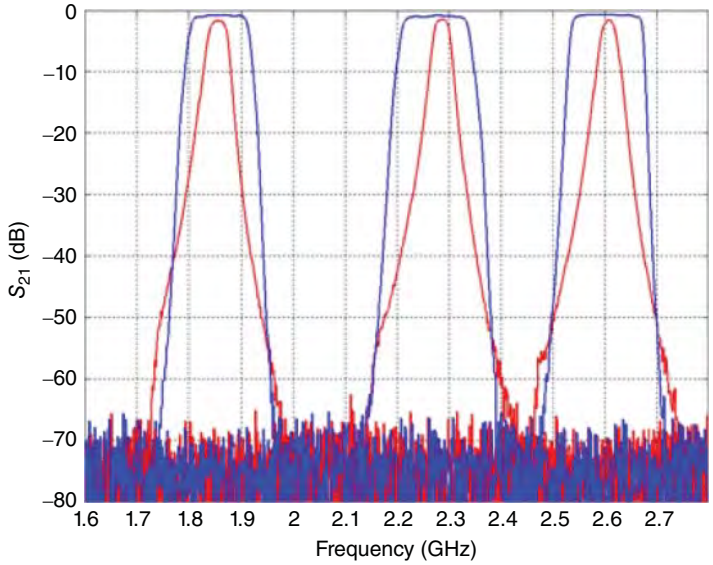


Figure 22.70 Measured results of a combline filter that is tunable in center frequency and bandwidth using an isolator between the two filters. (*Source*: Courtesy: M. Nasresfahani [18].)

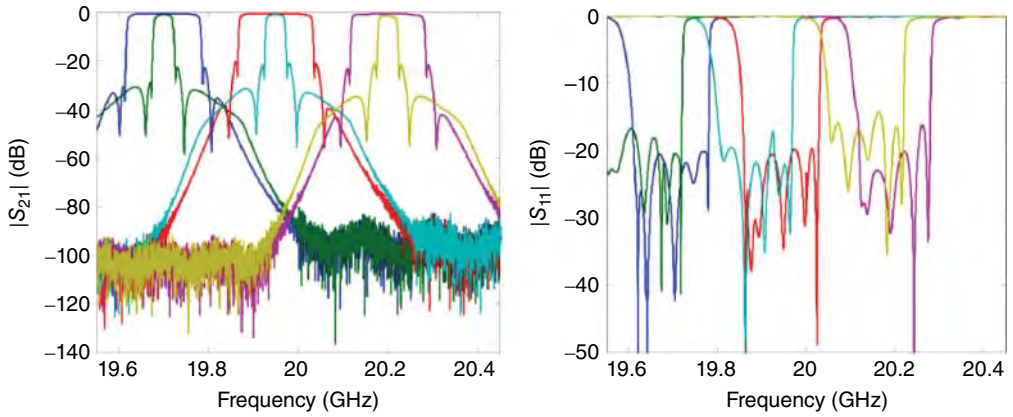


Figure 22.71 Tunable filter measured performance. Shown is 40–160 MHz bandwidth tuning over more than 500 MHz of center frequency tuning range. (*Source*: Yassini *et al*. 2012 [22]. Reproduced with the permission of IEEE.)

if mechanical tuning elements are employed. In [21] the concept of using off-resonance resonators to realize reactive coupling was used to realize a waveguide filter that is tunable in center frequency and bandwidth. Figure 22.73 shows the filter configuration, where the coupling between resonators are realized using resonators having the same configuration as the main resonators but are of smaller size. Both the filter main resonators and the off-resonance resonators (couplings) can then be tuned using similar mechanical tuning elements.

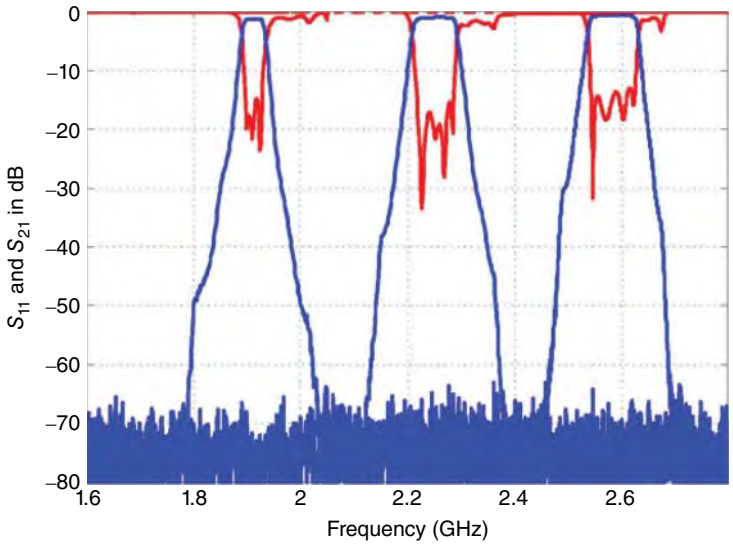


Figure 22.72 Measured results of a combline filter that is tunable in center frequency and bandwidth using a transmission line between the two filters. (*Source*: Courtesy: M. Nasresfahani [18].)

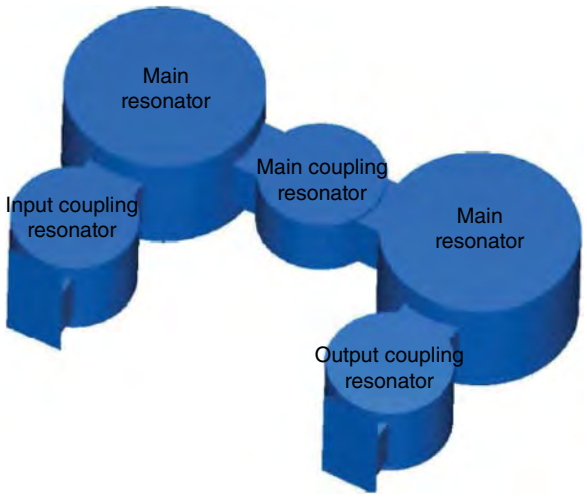


Figure 22.73 A 3D model for 2-pole Chebyshev filter employing off-resonance cavities to realize input/output and interresonator couplings. (*Source*: Arnold *et al*. 2014 [21]. Reproduced with the permission of IEEE.)

Summary

Tunable 3D coaxial, waveguide, and dielectric resonator filters can potentially address wireless base station and satellite applications that require high Q values (3000 and up). Yet, there are several challenges in developing such high-Q tunable filters for real system applications. The major challenge is to maintain an absolute constant bandwidth, a reasonable return loss, and a high Q value over a wide tuning range. While 3D filters can be readily tuned by motors, it is not easy to integrate miniature tuning elements such as MEMS or semiconductor varactors with 3D resonators. The tuning element must be integrated in a way to create enough perturbation to the electromagnetic field inside the 3D resonator in order to achieve a wide tuning range. Certainly, more research efforts are needed to explore the potential for realizing 3D high-Q

tunable filters with an RF performance that can approach the performance of the existing fixed filters used in wireless base station and satellite applications.

References

1 Mansour, R.R. (2009) High-Q tunable dielectric resonator filters. *IEEE Microwave Magazine*, **10** (6), 84–98.
2 Wong, P.W. and Hunter, I. (2009) Electronically tunable filters. *IEEE Microwave Magazine*, **10** (6), 46–54.
3 Rebeiz, G.M. *et al* (2009) Tuning in RF MEMS. *IEEE Microwave Magazine*, **10** (6), 46–54.
4 Hong, J.S. (2009) Reconfigurable planar filters. *IEEE Microwave Magazine*, **10** (6), 46–54.
5 Mansour, R.R., Huang, F., Fouladi, S. *et al.* (2014) High-Q tunable filters: challenges and potential. *IEEE Microwave Magazine*, **15** (5), 70–82.
6 Peroulis, D., Naglich, E., Sinani, M., and Hickle, M. (2014) Transfer-function-adaptive filters in evanescent-mode cavity-resonator technology. *IEEE Microwave Magazine*, **15** (5), 55–69.
7 Yu, M., Tassini, B., Keats, B., and Wang, Y. (2014) High-performance tunable filters based on air-cavity resonators. *IEEE Microwave Magazine*, **15** (5), 83–93.
8 Guyette, A. (2014) Controlled agility. *IEEE Microwave Magazine*, **15** (5), 32–42.
9 Gomez-Garcia, R., Sanchez-Soriano, M., Tam, K., and Xue, Q. (2014) Flexible filters. *IEEE Microwave Magazine*, **15** (5), 43–54.
10 Yan, W.D. and Mansour, R.R. (2007) Tunable dielectric resonator bandpass filter with embedded MEMS tuning elements. *IEEE Transactions on Microwave Theory and Techniques*, **55**, 154–160.
11 Huang, F., Fouladi, S., and Mansour, R.R. (2011) High-Q tunable dielectric resonator filters using MEMS technology. *IEEE Transactions on Microwave Theory and Techniques*, **59**, 3401–3409.
12 Fouladi, S., Huang, F., Yan, W.D., and Mansour, R.R. (2013) High-Q narrowband tunable combline bandpass filters using MEMS capacitor banks and piezomotors. *IEEE Transactions on Microwave Theory and Techniques*, **61**, 393–402.
13 Jiang, J. and Mansour, R.R. (2016) High-Q tunable filter with a novel tuning structure. European Microwave Symposium, October 2016.
14 Huang, F. and Mansour, R.R. (2013) A novel varactor tuned dielectric resonator filter. IEEE-MTT-S International Microwave Symposium (IMS), June 2013.
15 Iskander, M., Nasr, M. and Mansour, R.R. (2014) High-Q tunable combline bandpass filter using angular tuning technique. European Microwave Symposium, EUMC.
16 Yan, W.D. (2007) Microwave and millimeter-wave MEMS tunable filters. Ph.D. thesis. University of Waterloo, Waterloo, Ontario, Canada.
17 Huang, F. (2012) High-Q tunable filters. Ph.D. thesis, University of Waterloo, Waterloo, Ontario, Canada.
18 Nasresfahani, M. (2014) Design and fabrication of centre frequency and bandwidth tunable cavity filters. M.Sc. thesis. University of Waterloo, Waterloo, Ontario, Canada.
19 Iskander, M. (2014) RF tunable resonators and filters. M.Sc. thesis. University of Waterloo, Waterloo, Ontario, Canada.
20 http://www.piezo.com
21 Arnold, C., Parleba, J., and Zwick, T. (2014) Reconfigurable waveguide filter with variable bandwidth and center frequency. *IEEE Transactions on Microwave Theory and Techniques*, **62** (8), 1663–1670.

22 Yassini, B., Yu, M., and Keats, B. (2012) A Ka-Band fully tunable cavity filter. *IEEE Transactions on Microwave Theory and Techniques*, **60**, 4002–4012.

23 Snyder, R.V. (2000) A wideband tunable filter technique based on double-diplexing and low-Q tuning elements. Proceedings IEEE International Microwave Symposium Digest, vol. 3, pp. 1759–1762.

24 Rhodes, J.D., Mobbs, C.I. and Ibbetson, D. (2011) Tunable bandpass filters. US patent 7,915,977, Mar. 29, 2011.

25 Yassini, B., Yu, M., and Kellett, S. (2009) A Ku-band high-tunable filter with stable tuning response. *IEEE Transactions on Microwave Theory and Techniques*, **57** (12), 2948–2957.

26 Kurudere, S. and Ertürk, V.B. (2013) Novel microstrip fed mechanically tunable combline cavity filter. *IEEE Microwave and Wireless Components Letters*, **23**, 570–580.

27 Liu, X., Katehi, L.P.B., Chappell, W.J., and Peroulis, D. (2010) High tunable microwave cavity resonators and filters using SOI-based RF MEMS tuners. *Journal of Microelectromechanical Systems*, **19** (4), 774–784.

28 Naglich, E.J., Peroulis, D. and Chappell, W.J. (2013) Wide spurious free range positive-to-negative inter-resonator coupling structure for reconfigurable filters. International Microwave Symposium (IMS).

29 Arif, M.S. and Peroulis, D. (2014) All-silicon technology for high-Q evanescent mode cavity tunable resonators and filters. *Journal of Microelectromechanical Systems*, **23** (3), 727–739.

30 Sirci, S., Martinez, J.D., Taroncher, M. and Boria, V.E. (2012) Analog tuning of compact varactor-loaded combline filters in substrate integrated waveguide. European Microwave Symposium, pp. 257–260.

31 Mira, F., Mateu, J., and Collado, C. (2012) Mechanical tuning of substrate integrated waveguide resonators. *IEEE Microwave and Wireless Components Letters*, **22** (9), 447–449.

32 Armendariz, M., Sekar, V. and Entesari, K. (2010) Tunable SIW bandpass filters with PIN diodes. European Microwave Conference, pp. 830–833.

33 Hyunseong, K., Sam, S., Ik-Jae, H., Chang-Wook Baek, B. and Sungjoon Lim, L. (2013) Silicon-based substrate-integrated waveguide-based tunable band-pass filter using interdigital MEMS capacitor. Asia Pacific Microwave Conference Proceedings (APMC), pp. 456–458.

34 Da-Peng, W., Wen-Quan, C. and Russer, P. (2008) Tunable substrate-integrated waveguide (SIW) dual-mode square cavity filter with metal cylinders. Proceedings IEEE MTT-S International Microwave Workshop Series on Art of Miniaturizing RF and Microwave Passive Components, pp. 128–131.

35 Joshi, H., Sigmarsson, H.H., Moon, S. *et al.* (2009) High fully reconfigurable tunable band pass filters. *IEEE Transactions on Microwave Theory and Techniques*, **57** (12), 3525–3533.

36 Abunjaileh, A.I. and Hunter, I.C. (2010) Tunable bandpass and bandstop filters based on dual-band combline structures. *IEEE Transactions on Microwave Theory and Techniques*, **58**, 3710–3719.

37 Abunjaileh, A.I. and Hunter, I.C. (2009) Tunable combline bandstop filter with constant bandwidth. Proceedings IEEE International Microwave Symposium Digest, pp. 1349–1352.

38 Matthaei, G.L. (2003) Narrow-band, fixed-tuned, and tunable bandpass filters with zig–zag hairpin-comb resonators. *IEEE Transactions on Microwave Theory and Techniques*, **51**, 1214–1219.

39 Zahirovic, N., Fouladi, S., Mansour, R.R. and Yu, M. (2011) Tunable suspended substrate stripline filters with constant bandwidth, IEEE MTT-S International Microwave on Symposium Digest.

40 Psychogiou, D. and Peroulis, D. (2014) Tunable VHF miniaturized helical filters, *IEEE Transactions on Microwave Theory and Techniques*, **62**, 282–289 2014.

41 Jiang, J. and Mansour, R.R. (2016) High-Q tunable filter with a novel tuning structure. European Microwave Symposium.

42 Pelliccia, L., Cacciamani, F., Farinelli, P., and Sorrentino, R. (2016) High-tunable waveguide filters using Ohmic RF MEMS switches. *IEEE Transactions on Microwave Theory and Techniques*, **63** (10), 3381–3390.

43 Chan, K.Y., Ramer, R., Mansour, R.R., and Sorrentino, R. (2016) Design of waveguide switches using switchable planar bandstop filters. *IEEE Microwave and Wireless Components Letters*, **26** (10), 798–800.

44 Chan, K.Y., Ramer, R., and Mansour, R.R. (2016) Switchable iris bandpass filter using RF MEMS switchable planar resonators. *IEEE Microwave and Wireless Components Letters*, **26** (10), 34–36.

45 Rosenberg, U., Rosowsky, D., Rummer, W. and Wolk, D. (1988) Tuneable manifold multiplexers – a new possibility for satellite redundancy philosophy. European Microwave Conference EuMC88, Proceedings, Stockholm, Sweden, pp. 681–686.

46 Kunes, M.A. and Connor, G.G. (1989) A digitally controlled tunable high power output filter for space applications. European Microwave Conference EuMC89, Proceedings, London, UK, pp. 870–875.

47 Rosenberg, U. and Knipp, M. (2005) Novel tunable high Q filter design for branching networks with extreme narrowband channels at mm-wave frequencies. EuMC2005 Conference Proceedings (CD), Paris, France, October 2005.

48 Arnold, C., Parleba, J. and Zwick, T. (2016) Reconfigurable doublet dual-mode cavity filter designs providing remote controlled center frequency and bandwidth re-allocation. 46th European Microwave Symposium, pp. 532–536.

49 Wakino, K., Tamura, H. and Isikawa, Y. (1987) Dielectric resonator device. US Patent, 4,692,712, Sept. 1987.

50 Chen, S., Zaki, K.A., and West, R.G. (1990) Tunable temperature-compensated dielectric resonators and filters. *IEEE Transactions on Microwave Theory and Techniques*, **38**, 1046–1051.

51 Wang, C. and Blair, W.D. (2002) Tunable high-Q dielectric loaded resonator and filter. IEEE-IMS, June 2002.

52 Pance, L. and Rochford, G. (2008) Multiple band and multiple frequency dielectric resonators tunable filters for base stations. Proceedings of the 38th European Microwave Conference, Amsterdam, October 2008, pp. 488–491.

53 L. Pance, Private Communication, October 2008.

54 Furman, E., Lanagan, M., Golubeva, I. and Poplayko, Y. Piezo-controlled microwave frequency agile dielectric devices. IEEE, International Ultrasonic Ferroelectric and Frequency Control, August 2004, pp. 266–271.

55 Petrov, P.K. *et al.* (2004) Tuneable two pole one dielectric resonator filter with elliptic characteristics. *Integrated Ferroelectrics*, **66**, 261–266.

56 Alford, N. *et al*. (2005) Tunable 4-pole piezoelectric filter based on two dielectric resonators. *Integrated Ferroelectrics*, **77**, 123–128.

57 Panaitov, G., *et al*. (2004) Tuning of high-q dielectric resonators in IEEE-IMS workshop on tunability for highly selective microwave filters. IEEE-IMS, June 2004.

58 Huang, F. and Mansour, R.R. (2009) Tunable compact dielectric resonator filters. European Microwave Symposium EuMC, pp. 559–562.

59 Yan, W.D. and Mansour, R.R. (2006) Micromachined millimeter-wave ridge waveguide filter with embedded MEMS tuning elements. IEEE-MTT-IMS, June 2006.

60 Joshi, H., *et al*. (2007) Highly loaded evanescent cavities from widely tunable high-Q filters. IEEE-IMS.

61 Amadjikpe, A. and Papapolymerou, J. (2008) A high-Q electronically tunable evanescent-mode double ridged rectangular waveguide resonator. IEEE-IMS, June 2008.

62 Panaitov, G., Ott, R., and Klein, N. (2005) Dielectric resonator with discrete electromechanical frequency tuning. *IEEE Transactions on Microwave Theory and Techniques*, **53**, 3371–3377.

63 Mansour, R.R. (2009) IEEE-IMS workshop on emerging applications of RF MEMS. IEEE-IMS.

64 Faar, A.N., Blackie, G.N. and Williams, D. (1983) Novel techniques for electronic tuning of dielectric resonators. Proceedings 13th European Microwave Conference, Nürnberg, Germany, September 1983, pp. 791–796.

65 Krupka, J., Abramowicz, A., and Derzakowski, K. (2006) Magnetically tunable filters for cellular communication terminals. *IEEE Transactions on Microwave Theory and Techniques*, **54**, 2329–3335.

66 Virdee, B.S., Virdeet, A. and Trinogga, L.A. (2003) Novel invasive electronic tuning of dielectric resonators. IEEE-IMS.

67 Gevorgian, S. (2009) Agile microwave devices. *IEEE Microwave Magazine*, **10** (5), 93–98.

68 Chandler, S.R., Hunter, I.C., and Gardiner, J.G. (1993) Active varactor tunable bandpass filter. *IEEE Microwave and Guided Wave Letters*, **3** (3), 70–71.

69 Torregrosa-Penalva, G., Lopez-Risueno, G., and Alonso, J.I. (2002) A simple method to design wide-band electronically tunable combline filters. *IEEE Transactions on Microwave Theory and Techniques*, **50** (1), 172–177.

70 Nath, J. *et al* (2005) An electronically tunable microstrip bandpass filter using thin-film barium-strontium-titanate BST varactors. *IEEE Transactions on Microwave Theory and Techniques*, **53** (9), 2707–2712.

71 Tombak, A. *et al* (2003) Voltage-controlled RF filters employing thin-film barium-strontium-titanate tunable capacitors. *IEEE Transactions on Microwave Theory and Techniques*, **51** (2, part 1), 462–467.

72 Murakami, Y. *et al* (1987) A 0.5–4.0-GHz Tunable Bandpass Filter Using YIG Film Grown by LPE. *IEEE Transactions on Microwave Theory and Techniques*, **35**, 1192–1198.

73 Tsutsumi, M. and Okubo, K. (1992) On the YIG film filters. *IEEE MTT-S International Microwave Symposium Digest*, **3**, 1397–1400.

74 Rebeiz, G.M. (2003) *RF MEMS: Theory, Design, and Technology*, Wiley, Hoboken, NJ.

75 Yan, W. and Mansour, R.R. (2007) Compact tunable bandstop filter integrated with large deflected actuators. IEEE-IMS, June 2007.

76 Rebeiz, G.M. *et al* (2009) Tuning in RF MEMSRF MEMS. *IEEE Microwave Magazine*, **10** (6), 55–72.

77 Mansour, R.R. (2013) RF MEMS-CMOS device integration: an overview of the potential for RF researchers. *IEEE Microwave Magazine*, **14**, 39–56.

78 Wang, M. and Rais-Zadeh, M. (2014) Directly heated four-terminal phase-change switches. IEEE International Microwave Symposium (IMS).

79 El-Hinnawy, N., *et al*. (2013) A 7.3 THz cut-off frequency, inline, chalcogenide phase-change RF switch using an independent resistive heater for thermal actuation. Compound Semiconductor IC Symposium (CSICS), p. 14.

80 Moon, J.S., *et al*. (2015) 10.6 THz figure-of-merit phase-change RF switches with embedded micro-heater. 2015 IEEE 15th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems, pp. 73–75.

81 Jiang, J., Chugunov, G. and Mansour, R.R. (2017) Fabrication and characterization of VO2-based series and parallel RF switches. IEEE International Microwave Symposium (IMS), June 2017.
82 Jiang, J. and Mansour, R.R. (2017) A VO2-based 30 GHz variable attenuator. IEEE International Microwave Symposium (IMS), June 2017.
83 University of Waterloo, UW-MEMS Process, https://uwaterloo.ca/centre-integrated-rf-engineering/mems-process-design-handbook.

23

Practical Considerations and Design Examples

Chandra M. Kudsia[1], *Vicente E. Boria*[2], *and Santiago Cogollos*[2]

[1] *University of Waterloo, former Chief Scientist, COM DEV International, Canada*
[2] *Universidad Politécnica de Valencia, Valencia, Spain*

This chapter is devoted to a series of examples to add clarity in understanding the material described in various chapters of this book. Another aim of this chapter is to bridge the gap between theory and practical realization of microwave filters and multiplexing networks. The chapter is divided into five sections:

Section 23.1 provides examples that cover system considerations and tradeoffs in the development of specifications for filter networks. It includes simulation of channel characteristics using pole–zero configuration of filter functions. The constraints imposed by the manufacturing tolerances and the operating environment on filter specifications are dealt with by invoking the concept of equivalent linear frequency drift (ELFD). It covers material in Chapters 1–4.

Section 23.2 provides examples that deal with filter design and synthesis techniques to enhance the understanding of the material described in Chapters 5–15. It exemplifies the use of electromagnetic (EM) tools in the computation of physical dimensions of filters. It clarifies the methodology in developing filter topologies for a given filter specification.

Section 23.3 describes simple examples to help understand the design methodology of the two most widely used multiplexers, the circulator-coupled configuration and the manifold-coupled design. Examples highlight the tradeoff between channel performance and design complexity in enhancing the understanding of the optimization process for the manifold-coupled multiplexers.

Section 23.4 is devoted to examples related to high-power requirements for filters and multiplexers in space and terrestrial environment. Examples highlight field distributions in different practical microwave structures (waveguide and coaxial) for realizing high-power filters to guard against phenomenon of multipaction in space and corona in terrestrial environments.

Section 23.5 is devoted to a simple analysis of manufacturing tolerances using EM techniques.

23.1 System Considerations for Filter Specifications in Communication Systems

Examples in this section cover system considerations and tradeoffs in the development of filter networks. Simulation of ideal channel characteristics are carried out using the pole–zero configurations of filter functions. Pole–zero location is readily determined by closed-form formulas for the classical filters. Computer-aided techniques as described in Chapter 4 may be used to

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

determine the poles and zeros for filters with arbitrary response. Practical constraint of finite electrical conductivity in materials used in the physical realization of filters is incorporated by assuming uniform dissipation in resonators. This assumption represents a close approximation of losses in practical filters. Effect of uniform dissipation can be readily included by shifting the pole–zero pattern of the filter function by an amount related to the unloaded Q of resonators as described in Chapter 3. The effect of the operating temperature environment and manufacturing tolerances is taken into account by invoking the concept of ELFD—related to the thermal properties of materials—described in depth in this section. This methodology allows accurate computation of channel characteristics, knowing just the pole–zero configuration of filter functions and basic material properties of the physical structure. It does not require knowledge of filter synthesis techniques. It represents a simple yet powerful tool for system engineers as well as hardware designers in developing the tradeoffs and cost-effective specifications for filter networks. This section exemplifies the material described in Chapters 1–4.

23.1.1 Equivalent Linear Frequency Drift (ELFD)

The term *ELFD* is employed to evaluate the performance tradeoffs of filters over the specified operating environment in communication payload equipment. It includes the effect of operating temperature range and allowances for tolerances in the manufacture and alignment of filter networks. ELFD is therefore composed of two components:

1) Frequency shift due to temperature (see Appendix 23A):
   - $f_0 \propto$ length of the resonator
   - $\dfrac{\Delta f}{f_0} = -\dfrac{\Delta l}{l_0}\Delta T$
   - $\Delta f = -\alpha \Delta T f_0$
     Δf: Total linear frequency drift ($f_0 \pm \frac{|\Delta f|}{2}$)
     α: Coefficient of thermal expansion
     ΔT: Operating temperature range.

   For a 4 GHz filter operating between 0 and 50 °C,

$$\text{Frequency drift} = \pm 2.24 \text{ MHz using aluminum } (\alpha = 22.4 \text{ ppm})$$
$$= \pm 0.16 \text{ MHz using invar } (\alpha = 1.6 \text{ ppm})$$

2) Equivalent frequency shift due to manufacturing tolerances and misalignment.

In ambient environment, tuning elements are used to align the filter close to its ideal response, effectively compensating for manufacturing tolerances. However, when a filter is subjected to changes in temperature, especially large temperature excursions (>20 °C), its response changes, both in terms of shift of its center frequency and response shape. It is the latter change that is difficult to predict. It arises due to complexity of filter structure, presence of tuning elements and irises, thermal expansion of materials, and joints and mounting structure. Physical expansion and contraction of the filter structure changes not only the lengths of the resonators but also the coupling values. Change in frequency response no longer conforms to a lateral shift of the ambient response with no change in its near-ideal equiripple shape. In order to capture this uncertainty, as well as an allowance in the initial alignment of filters in ambient environment in a well-defined measure, the industry developed the concept of ELFD [1] as follows:

> Equivalent Linear Frequency Drift is the total frequency excursion of the nominal center frequency of an ideal filter over a specified temperature range as determined by a given set of passband, and out-of-band performance constraints.

Passband constraint is normally chosen to be the group delay at the band edges of the specified passband bandwidth (BW). The out-of-band criterion is typically chosen as the stopband BW corresponding to the isolation in the 20–30 dB range. The choice is determined by the parameters that are most sensitive to temperature changes. Using such a criterion, ELFD is evaluated as follows:

- The ideal filter characteristics at ambient temperature are used to input a computer program that simulates the filter group delay and isolation characteristics with incremental upward and downward shifts in nominal center frequency.
- The group delay at band edges and isolation at, say, 20 dB is measured over the maximum specified operating temperatures.
- Measured values over temperature are compared with the simulated filter responses. The maximum simulated shift in center frequency that provides a group delay equal to or less than the measured value and an isolation equal to or greater than the measured value determines the "ELFD" over the specified temperature range.

23.1.1.1 Material Selection Tradeoffs and ELFD

The choice of materials and technology for the practical realization of microwave filters and multiplexers is determined by the frequency of operation, desired BW, insertion loss, operating environment, size, and mass. For high-power applications, power handling capability and specifications for passive intermodulation are often critical in choosing the design configuration. Cost is invariably an overarching constraint.

As described in the introduction of Section 23.1, the starting point for the filter designers is to conduct performance tradeoffs using the system constraints as the guide. For practical filters, an assessment of the unloaded Q (as determined by the specifications of insertion loss and its variation across the passband) of resonators and the operating temperature range are essential in the conduct of these tradeoffs.

The most widely used materials for microwave filters have been aluminum (light weight, ease of manufacture) and Invar when thermal stability is a key requirement. However, since the 1980s, there have been spectacular advances in the quality of dielectric materials. As described in Chapter 16, high-Q dielectric materials with dielectric constants ranging from 20 to 90, and temperature drift from −6 to + 6 ppm/°C, are commercially available. The Q value depends upon the value of dielectric constant; typical values are specified as quality factor times frequency in GHz ($Q \times f$) product; it ranges between 90,000 for $\varepsilon_r = 29$ and 50,000 for $\varepsilon_r = 45$. Another useful feature is the use of aluminum structures, be it planar, coax, or waveguide, to house dielectric resonators without any adverse effect on thermal stability. Using temperature compensation techniques (easier to implement in aluminum housings), it is possible to realize effective thermal drifts within ±1 ppm/°C range. Aluminum structures represent lower cost, mass, and smaller size. For these reasons, dielectric-loaded filters are widely used for low and moderate power (up to tens of watts, depending on the frequency band) applications. Practical Qs achieved for dielectric filters are similar to those achieved for the equivalent waveguide filters.

The next step in conducting performance tradeoffs is to choose an appropriate value of ELFD. As described earlier, it has two components: one due to the operating temperature range and the other related to design complexity, ease of manufacturing, and alignment tolerances. The first one is straightforward to evaluate, depending upon the choice of material. Its contribution represents the lower limit for ELFD. The second part reflects design complexity, workmanship standards, and most importantly the cost. The allowance for manufacturing tolerance is sensitive to cost; tighter allowance implies better performance, but the cost of implementation

Table 23.1 Typical values of ELFD.

Operating temperature range: 0–50 °C

Allowance for misalignment and manufacturing tolerances

ELFD: ±0.1 MHz at 1 GHz, ±0.15 MHz at 2 GHz, ±0.25 MHz at 4 GHz, and ±0.4 MHz at 12 GHz

			Frequency band (GHz)			
	α (ppm/°C)	Effective α (ppm/°C)	1	2	4	12
			Total ELFD (±MHz)			
Aluminum	22.4	25	0.80	1.40	2.75	8.0
Invar[a)]	1.6	2–2.5	0.15	0.25	0.45	1.0
Dielectric-loaded structures[a)]	2–4	2–6	0.15	0.25	0.45	1.0

a) Assumed value of α is 2 ppm/°C for Invar and dielectric-loaded structures.

would be higher and vice versa. At the system level, performance versus ELFD tradeoffs is an important criterion in budgeting and developing specifications of filter and multiplexing networks. Industry experience plays a key role in the choice of overall ELFD. Table 23.1 describes typical values of ELFD employed in specifying filter networks for some of the commonly used frequency bands and materials used to realize microwave filter networks. Values in this table are based on the following assumptions:

- Operating temperature range: 0–50 °C.
- α Represents the coefficient of thermal expansion of raw material in parts per million per degree celsius or ppm/°C.
- Effective α represents the assumed value of coefficient of thermal expansion for the resonant structures. It includes an allowance for the degradation observed due to manufacturing processes and complexity of the physical structures. Computed values in the table are based on $\alpha = 25$ ppm/°C for Al and 2 ppm/°C for Invar and dielectric-loaded structures.
- Allowance for misalignment and manufacturing tolerances: These values are arbitrary and reflect industry experience. Presence of tuning elements and discontinuities in the microwave structure produces undesired couplings (often referred to as second-order effects) when the filter is subjected to the operating temperature environment. This allocation is assumed to be independent of the technology used for the filters.

23.1.2 Synchronous versus Asynchronous Resonator Filters

For bandpass and bandstop filters exhibiting symmetrical response, it is possible to use synchronously tuned or asynchronously tuned resonators. For filters that require asymmetric response, it is necessary to use one or more asynchronously tuned resonators.

23.1.2.1 Synchronously Tuned Filters

Most practical systems require RF channels to exhibit symmetrical response about the center frequency. Synchronously tuned filters are ideally suited to realize symmetrical electrical response. Cohn's 1957 paper laid the foundation for the synthesis technique for direct coupled resonator filters using synchronously tuned resonators. Ever since, it has been the mainstay of nearly all commercial microwave filters.

In the 1970s, dual-mode filter technology and CM (Coupling Matrix) synthesis techniques were developed. Dual-mode technology is ideally suited to accommodate synchronous resonators by choosing square or circular waveguides. Similarly, planar structures with square or circular configuration for the resonators can readily accommodate two orthogonal modes in a

single physical resonator. A coupling element, such as a tuning screw or a discontinuity, is used to couple the two modes in the resonators. Dual mode structure allows easy implementation of coupling between some non-adjacent resonators. Such couplings allow realization of elliptic and quasielliptic filter functions in a variety of filter configurations. Another consequence is that the available couplings often exceed the degrees of freedom (i.e., order of the filter) to realize a given filter function, leading to an infinity of possible solutions. Practical dual-mode structures are derived by minimizing the number of couplings. For practical applications, dual-mode topology represents the most compact and cost-effective filter structure.

23.1.2.2 Asynchronously Tuned Filters

For applications where asymmetrical response is required, it is necessary to use at least one or more asynchronously tuned resonators. In the 2000s, CM techniques were extended to include the design of filters with asymmetric response. Chapters 9, 10, and 13 include a number of examples and topologies using asynchronous resonators. These include cul-de-sac, extracted pole, and extended box configurations.

Asynchronously tuned resonators are a consequence of designing filters with asymmetric response. However, it is possible to realize symmetric response using asynchronously tuned resonators at the expense of extra couplings; for example, employing two trisections in the filter topology. Use of asynchronously tuned resonators to achieve symmetrical response implies that the physical dimensions of resonators cannot be identical. This restricts the use of dual-mode structure for designs where the two resonant frequencies in the same cavity are not separated by a large amount. Use of two resonant frequencies in the same cavity also implies that one of the tuning screws has to penetrate the cavity by a larger amount to account for the difference in resonance frequency, thereby incurring degradation of Q_0. For these reasons, asynchronously tuned resonator design to realize symmetrical response is not suited for practical filter networks. However, there are applications where asynchronously tuned resonators are required to achieve symmetrical response. One such example is the design of filters corresponding to the end channels in an RF channel combiner or multiplexer. Such filters require asymmetrical frequency response to compensate for the lack of a surrounding channel on one side of the filter. This is accomplished by incorporating asynchronous resonators in the filter design. Another example is duplexers required for cellular stations. In a two-channel multiplexer, both filters need to be designed with asymmetrical response in order to realize the optimum (symmetrical) response for each channel.

23.1.3 Filter Tradeoffs in the Characterization of RF Channels

Examples in this section demonstrate how one can perform tradeoffs and characterize RF channels in a communication system based on using Chebyshev and quasielliptic function filters. Information on such filter functions is widely available or readily computed. The procedure is general and applicable to any class of filter function. Assuming uniform dissipation in the physical realization of filters (as is typically the case), it is possible to characterize channel performance both in amplitude and phase/group delay quite accurately. Invoking the concept of ELFD, it is possible to evaluate the performance tradeoffs over the operating environment and acceptable manufacturing tolerances.

Example 23.1 ***Practical Filter Simulation Using Poles and Zeros***

The intent of this example is to demonstrate the computation of amplitude and group delay response of practical filters using poles and zeros of the filter function. Another feature of this example is to highlight the performance tradeoffs over the operating temperature environment and sensitivity to manufacturing tolerances.

For this example we choose a six-pole quasielliptic filter with a single pair of transmission zeros and a characteristic factor of $F = 60$ dB (see Chapter 3). Other assumed parameters are RL = 22 dB, center frequency $f_0 = 4$ GHz, specified passband BW = 24 MHz, and unloaded $Q = 4000$, representative of a coax structure. The design BW of a filter must exceed the specified passband BW to allow for performance degradation over the operating environment and due to manufacturing tolerances. For this example, we choose a design BW of 26 MHz.

Solution:
The critical frequencies of the prototype filter are provided in Table 3A.3 or deduced thereof:

$$\text{Transmission zeros: } s_{tz} = \{\pm j1.4869\}$$
$$\text{Reflection zeros: } s_r = \{\pm j0.2818, \pm j0.7415, \pm j0.9723\}$$
$$\text{Poles: } s_p = \{-0.1068 \pm j1.0908, -0.3687 \pm j0.8775, -0.5980 \pm j0.3463\}$$

The polynomials obtained from the critical frequencies and the values of ϵ and ϵ_R are shown in Table 23.2.

The dissipation factor δ for this bandpass filter is computed as:

$$\delta = \frac{f_0}{\Delta f}\frac{1}{Q_0} = 0.03846$$

The effect of uniform dissipation is incorporated by subtracting δ from the real part of the poles and zeros.

As described in Section 23.1.1, we invoke the concept of ELFD to demonstrate the performance tradeoffs of practical filters under typical operating environment. For this example, let us assume that the filter is subjected to operate over a temperature range of 0–50 °C. The minimum ELFD required for the temperature range is ±0.2 MHz, compatible with using Invar or equivalent material for the realization of the filter structure. An allocation of ELFD = ±0.3 MHz is assumed for the manufacturing tolerances for a total ELFD = ±0.5 MHz.

Figure 23.1 represents the nominal amplitude response of the filter. For most applications, the near-out-of-band isolation in the 20–30 dB range is the constraining specification. Such a constraint arises when considering noise budget allocation in a transponder or channel of a communication system as described in Chapter 1.

For this example, we choose stopband BW of ±18 MHz and isolation of 30 dB or greater as the constraining specifications. As seen in this figure, attenuation slope at the 30 dB level is very

Table 23.2 Polynomial coefficients for the normalized polynomials $E(S)$, $F(S)$, and $P(S)$.

	$E(S)$	$F(S)$	$P(S)$
s^6	1.0000	1.0000	—
s^5	2.1471	—	—
s^4	3.8797	1.5747	—
s^3	4.2421	—	—
s^2	3.4607	0.6386	$j1.0000$
s^1	1.8171	—	—
s^0	0.5197	0.0413	$j2.2109$
	$\epsilon_R = 1.0000$		$\epsilon = 4.2673$

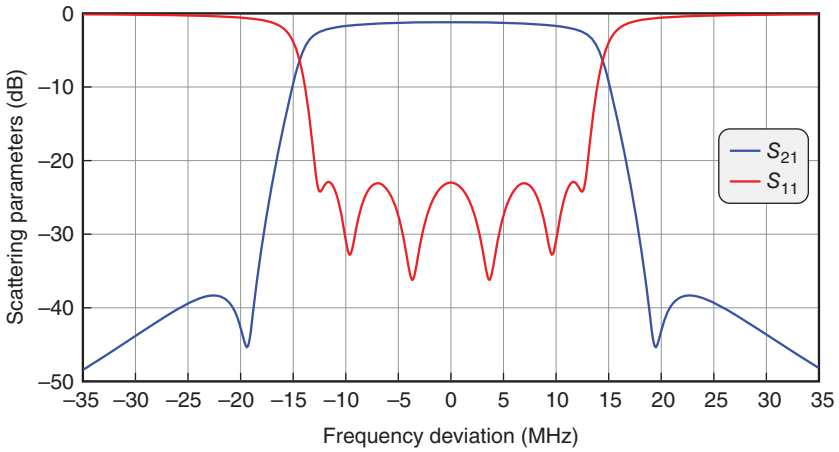


Figure 23.1 Frequency response for the 6-2 filter of Example 23.1.

steep, implying large change in performance with respect to ELFD. To highlight this tradeoff, Figure 23.2a shows the isolation around 30 dB point on an expanded scale.

For the passband, the most sensitive parameters are the relative group delay and insertion loss at the band edges of the passband. Using the specified passband BW as the pivot points, Figure 23.2b and c shows the sensitivity of group delay and insertion loss variation at the band edges. It is possible to improve passband performance by using group delay equalizers or linear phase filters, involving a new set of tradeoffs.

Table 23.3 provides the computed values of the nominal performance and the worst-case values for an ELFD of ±0.5 MHz. This table represents just a sample of the performance for a specific filter function and specification constraints. Following this methodology, it is possible to generate performance tradeoffs using different values for the near-out-of-band isolation and stopband BWs as the constraining specifications. In practice, it is usual to consider a variety of filter functions to generate the tradeoffs before making the final selection.

This simple example demonstrates the methodology in developing tradeoffs for practical filter networks for any application and operating environment, just knowing the pole–zero configuration of the filter function. The operating temperature range defines the minimum value of ELFD, given the thermal stability of the material used for filter realization. ELFD beyond the minimum value provides a good measure of the performance versus the impact on manufacturing tolerances and cost. The choice of Q_0 determines the technology tradeoffs. It depends upon the filter topology, physical structure, propagation mode, and workmanship standards. Q_0 is an important parameter in establishing the size and mass of the filter network. The methodology described in this example provides a powerful tool for the system engineers as well as filter designers to analyze and optimize filtering requirements in communication systems.

Example 23.2 *Channelization Filter Tradeoffs for a 6/4 GHz Satellite System*

Intent

In the previous example, we showed how it is possible to compute performance of practical filter networks knowing the critical frequencies (pole–zero configuration) of the filter functions. In this example, we extend this technique to evaluate system and design parameters for the channelization of the available frequency spectrum for a 6/4 GHz satellite system. Narrowband

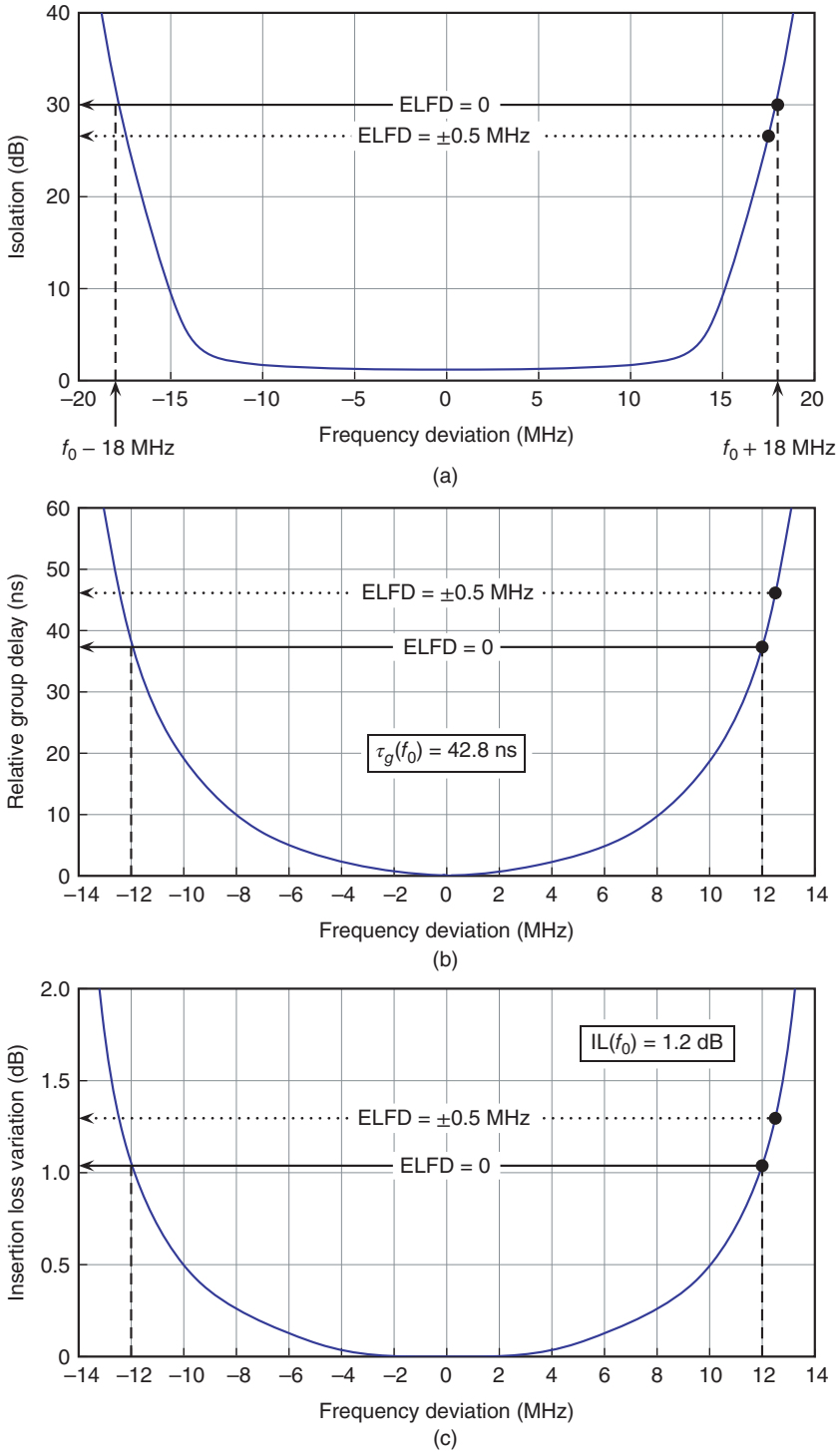


Figure 23.2 Demonstration of filter tradeoffs with respect to ELFD for the given filter function. (a) Out-of-band attenuation. (b) Passband relative group delay. (c) Passband insertion loss variation.

Table 23.3 Performance tradeoff summary.

	ELFD (MHz)		
Parameter (worst case)	**0**	**±0.2**	**±0.5**
Attenuation at ±18 MHz	30.0 dB	29.2 dB	26.6 dB
Relative group delay[a)] at ±12 MHz	37.3 ns	40.6 ns	46.4 ns
Relative insertion loss[b)] at ±12 MHz	1.0 dB	1.1 dB	1.3 dB

a) Absolute group delay at $f_0 = 42.8$ ns.
b) Absolute insertion loss at $f_0 = 1.2$ dB.

microwave filters required for channelization have a major impact on the characteristics of channels or transponders. In effect, performance of such filters determines the usable BWs of transponders. More BW implies higher communication capacity. Maximization of achievable BWs requires the evaluation of higher-order complex filter functions as described in this example. The effect of the operating environment and technologies used in the implementation of filter networks can be readily included in the behavior of the system level tradeoffs. Comparison of computed performance and respective EM models (see Example 23.7) provides validation of this methodology. It is applicable in assessing filter tradeoffs for any communication system.

System Constraints

As described in Section 1.8, frequency spectrum for the downlink of a 6/4 GHz satellite system extends from 3.7 to 4.2 GHz. The 500 MHz BW is typically divided into 12 channels; each channel is assigned a BW of 40 MHz. A guard band of 4 MHz is assumed between each channel to allow a margin for the practical implementation of channelizing filters; this results in a specification of 36 MHz as the usable BW for each channel. Allocation of 4 MHz as guard band is arbitrary and is governed by the achievable performance of practical filter networks. As described in Section 1.5, it is desirable to keep the interfering noise levels from adjacent and other copolarized channels to less than 20–30 dB below the desired signals. This implies that the filters must exhibit very sharp rise in attenuation, ideally 20 dB at the band edge of the adjacent channel. To achieve such sharp rise of attenuation at the band edge of adjacent channels (i.e., $f_0 \pm 22$ MHz and beyond) tends to distort the passband performance. As a compromise, typical specifications require isolation constraints of greater than 10 dB at $f_0 \pm 22$ MHz and greater than 35 dB at $f_0 \pm 25$ MHz, and beyond within 3.7–4.2 GHz frequency band.

Filter Tradeoffs

Filter tradeoffs are carried out using three alternative designs:

1) An 8-4 quasielliptic filter with a 10 dB difference in the attenuation minima (see Appendix 4A). Transmission zeros in the lowpass prototype are located at $s_{tz} = \{\pm j1.1601, \pm j1.4321\}$; RL = 22 dB and center frequency $f_0 = 4$ GHz are the standard features chosen for this example. Tradeoffs require maximization of passband BW within the constraints of achieving greater than 10 dB isolation 22 MHz away from band center. This tradeoff yields a BW of 39.5 MHz and provides greater than 15 dB nominal isolation at $f_0 \pm 22$ MHz. The amplitude response is shown in Figure 23.3.

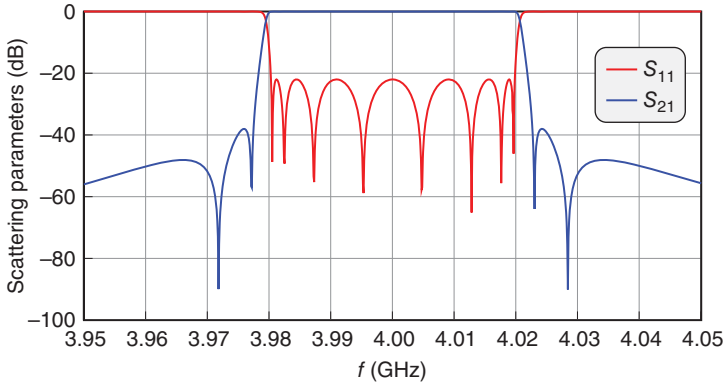


Figure 23.3 Ideal amplitude response of the 8-4 quasielliptic filter with a 10 dB difference in the attenuation minima.

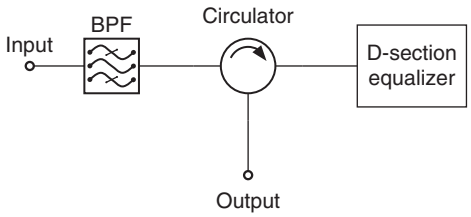


Figure 23.4 Bandpass filter equalized using a D-section reflection-type equalizer.

2) The previous 8-4 filter in conjunction with a D-section equalizer (see Figure 23.4), designed using an optimization technique as described in Chapter 17. The goal here is to optimize the group delay variation to within a 2 ns ripple over the middle portion of the passband. The circulator used for the equalizer is assumed to have (1) 0.2 dB loss in the forward direction, (2) reverse isolation of 30 dB, and (3) ideal performance over the operating temperature range.[1]
3) A 10-2-2 linear phase filter. Transmission zeros of the lowpass prototype are located at $s_{tz} = \{\pm j1.2015, \pm 0.719\}$. The real transmission zeros in the s plane have been chosen to allow a maximum variation of 2 ns in the group delay. RL = 22 dB, center frequency $f_0 = 4$ GHz, and BW = 39.5 MHz are the features chosen for this example. The magnitude response is shown in Figure 23.5.

The three responses of the group delay for these filters are shown in the same plot in Figure 23.6 for comparison purposes. The 8-4 filter in conjunction with the equalizer can achieve a flat group delay over a wider BW than the self-equalized 10-2-2 filter.

Assuming either waveguide or dielectric-loaded resonator technology, an unloaded Q factor of 10,000 is assumed for all filters. A conservative value of ELFD = ±0.5 MHz is used to allow for the operating temperature environment and manufacturing tolerances.

Table 23.4 summarizes the filter tradeoffs for the three alternative designs. This table highlights the practical performance tradeoffs for channelizing filters in a communication system using externally equalized and linear phase filters. For a given set of amplitude constraints, the unequalized filter has the worst passband performance; however, it offers the simplest configuration and a minimum of mass and volume. The externally equalized filter offers the best

1 Practical circulators do exhibit performance variation over temperature. Their performance can be optimized over narrow bandwidths.

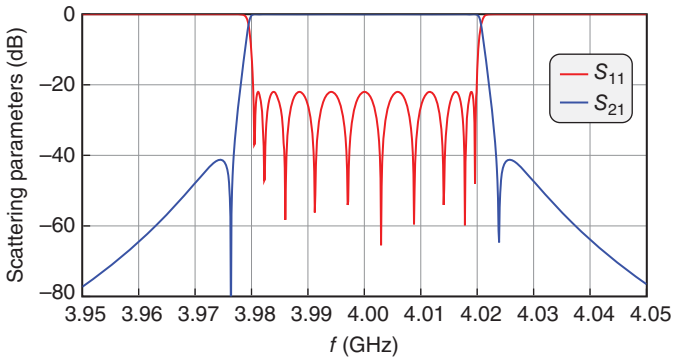


Figure 23.5 Ideal amplitude response of the 10-2-2 quasielliptic filter.

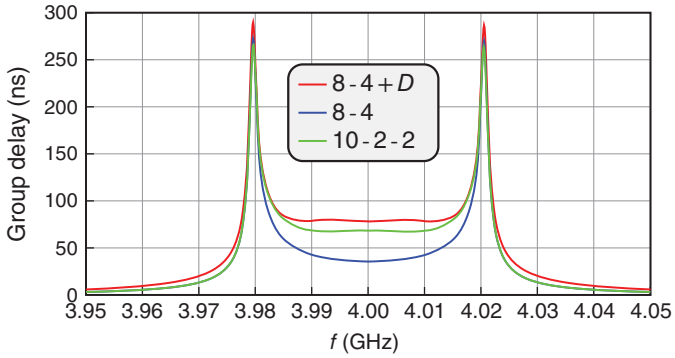


Figure 23.6 Group delay comparison of the three alternative designs.

Table 23.4 Tradeoff summary for the channelizing filters in a 6/4 GHz satellite system.

	Isolation (dB)		Rel. group delay (ns)		Loss variation (dB)	
	$f_0 \pm 22$	$f_0 \pm 25$[a)]	$f_0 \pm 15$	$f_0 \pm 18$	$f_0 \pm 15$	$f_0 \pm 18$
8-4 dual mode[b)]	14.8	38.2	26.3	68.2	0.3	0.7
8-4 + D dual mode[c)]	16.2	39.1	5.4	42	0.3	0.4
10-2-2[d)]	14.8	41.9	14.9	52.5	0.2	0.6

ELFD = ±0.5 MHz is assumed.

a) Spec. values imply isolation at $f_0 \pm 25$ MHz and beyond within the range 3.7–4.2 GHz.

b) The absolute group delay and insertion loss at band center are 35.5 ns and 0.4 dB, respectively.

c) The absolute group delay and insertion loss at band center are 79.5 ns and 1.3 dB, respectively.

d) The absolute group delay and insertion loss at band center are 68.5 ns and 0.8 dB, respectively.

performance at the expense of increased complexity and mass and higher susceptibility to performance degradation over the operating temperature range. The linear phase filter shows good compromise in terms of performance; however, linear phase structures are more sensitive to manufacturing tolerances.

23.2 Filter Synthesis Techniques and Topologies

This section deals with some methodologies to design filters starting from the ideal models (lumped and/or distributed) to the step-by-step computation of the physical dimensions using EM simulators.

This section includes a number of examples to demonstrate the use of CM to synthesize circuit models for a variety of filter topologies. These circuit models are then used to compute the physical dimensions using EM-based techniques and full-wave analysis. Physical realization includes waveguide, coaxial, and dielectric-loaded structures. Filter topologies are chosen to highlight demanding specifications and in clarifying the advantages and drawbacks of the different structures. Judicious use of EM techniques to compute physical dimensions is critical in applying this powerful tool for filter design. Ample description is included in examples highlighting the development of the initial model, its selective augmentation by distributed elements leading to the full-scale EM model, and rationale for choosing a particular EM technique.

Example 23.3 ***EM Modeling of a Single-Mode Filter***
In this example, we consider a coupled-cavity Chebyshev bandpass filter in waveguide technology with the following specifications:

- Order $N = 3$
- RL = 20 dB
- Center frequency $f_0 = 4$ GHz
- BW = 40 MHz
- Structure: rectangular waveguide, full-inductive couplings.

The filter is all pole and therefore no transmission zeros are present in this design. The polynomials involved can be directly obtained with classical formulas. The starting point is the expression of the scattering parameters in the lowpass domain:

$$S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)}, \quad S_{11}(s) = \frac{F(s)/\varepsilon_r}{E(s)}$$

Since it is an all-pole design and $N - n_{tz} = 3$, we set $P(s) = 1$, $\varepsilon_r = 1$. The polynomial $E(s)$ is obtained from its roots (poles of the filter)

$$p_k = j\cosh(\eta + j\theta_k)$$
$$\eta = \frac{1}{\eta}\operatorname{asinh}(\varepsilon_1), \quad \varepsilon_1 = \sqrt{10^{\mathrm{RL}/10} - 1}$$
$$\theta_k = \frac{2k-1}{2N}\pi, \quad k = 1, 2, 3$$

giving three poles:

$$p_1 = -1.17172, \quad p_{2,3} = -0.58586 \pm j1.33405$$

The obtention of the reflection zeros [i.e., zeros of $F(s)$] is straightforward following the classical formulas

$$z_k = j\cos(\theta_k), \quad \theta_k = \frac{2k-1}{2N}\pi, \quad k = 1, 2, 3$$

giving three reflection zeros:

$$z_1 = 0, \quad z_{2,3} = \pm j0.8660$$

The polynomials have the following expressions:

$$\begin{aligned} E(s) &= (s-p_1)(s-p_2)(s-p_3) \\ &= 2.4875 + 3.4958s + 2.3434s^2 + s^3 \\ F(s) &= (s-z_1)(s-z_2)(s-z_3) \\ &= 0.75s + s^3 \end{aligned}$$

Finally, the constant ε is obtained enforcing the condition at the band edge. The Chebyshev filter has (exactly) the prescribed RL at $\omega = \pm 1$:

$$\varepsilon = \frac{1}{\varepsilon_1}\left|\frac{P(s)}{F(s)}\right|_{s=\pm j} = 0.40202$$

Once the polynomials are obtained, the mathematical expression of the filter has to be converted into a circuit. The simplest standard form of this circuit is a lumped ladder network that gives exactly the same response provided by the mathematical function. For Chebyshev filters, Belevitch [2] derived analytical expressions for the circuit element values (see Section 14.1), thus obviating the need for synthesis via polynomials. The values of the lumped elements are

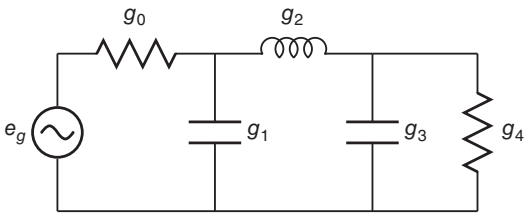
$$\begin{aligned} g_0 &= g_4 = 1\ \Omega \\ g_1 &= g_3 = 0.85345\ \text{F} \\ g_2 &= 1.10387\ \text{H} \end{aligned}$$

and the lowpass circuit is shown in Figure 23.7. If the formulas for the lumped elements are not known, the lumped circuit has to be obtained from the polynomials through a suitable synthesis procedure. The dual circuit can also be used if required, since the same response is obtained from the dual network. In the physical realization of microwave filters, it is preferable to deal with the same distributed element (be it an inductor or capacitor) throughout the filter structure (see Appendix E), it makes the structure simpler and more robust. This can be accomplished by the use of impedance or admittance inverters. It is better to obtain a network with all elements normalized to unity. For example, the lowpass network with all elements being shunt normalized capacitors is sketched in Figure 23.8.

The inverter values as derived in Appendix E and also reported in Section 14.1 are as follows:

$$J_{i,i+1} = \frac{1}{\sqrt{g_i g_{i+1}}} \Rightarrow \begin{cases} J_{01} = 1.08246 = J_{34} \\ J_{12} = 1.03027 = J_{23} \end{cases}$$

Figure 23.7 Lumped lowpass ladder network for the filter.



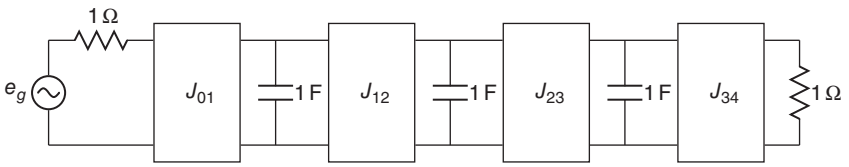


Figure 23.8 Normalized prototype model of the filter with *J*-inverters.

and therefore the $N+2$ CM is straightforward:

$$\mathbf{M} = \begin{pmatrix} 0 & J_{01} & 0 & 0 & 0 \\ J_{01} & 0 & J_{12} & 0 & 0 \\ 0 & J_{12} & 0 & J_{23} & 0 \\ 0 & 0 & J_{23} & 0 & J_{34} \\ 0 & 0 & 0 & J_{34} & 0 \end{pmatrix}$$

The next step is to convert the lowpass prototype filter into its bandpass configuration. This is accomplished by applying the lowpass to bandpass transformation to the capacitors as explained in Section 3.8. The shunt capacitors will be transformed in parallel resonators (all identical) with the formulas

$$C_p = \frac{1}{2\pi \text{ BW}} = 3.978874 \text{ nF}$$

$$L_p = \frac{1}{4\pi^2 f_0^2 C_p} = 0.3978874 \text{ pH}$$

and the inverters remain unchanged since they are frequency independent. The bandpass network is shown in Figure 23.9, and its electrical response is depicted in Figure 23.10. The lumped model and the equivalent CM give us a good starting point to obtain the distributed model composed of transmission lines and inverters. The next step is the choice of technology for the physical realization of the filter. In this example, we use the waveguide technology; therefore the transmission-line response has to take into account the dispersion

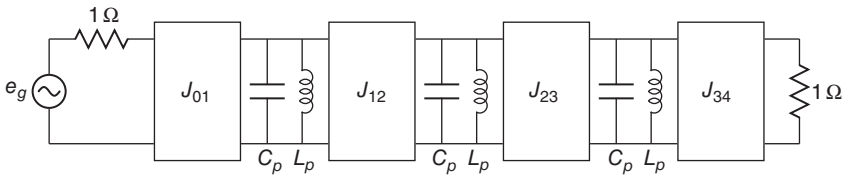


Figure 23.9 Equivalent lumped bandpass filter.

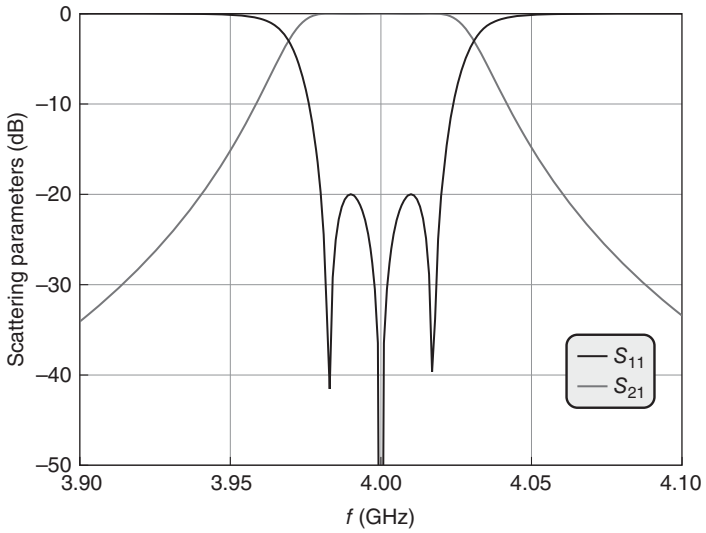


Figure 23.10 Scattering parameters of the lumped bandpass filter.

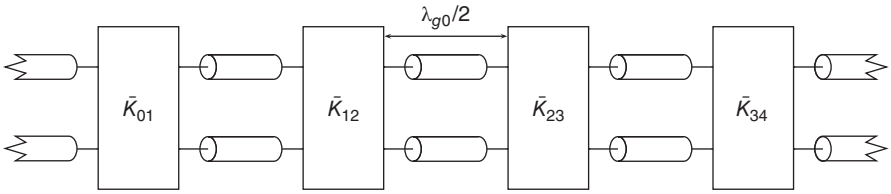


Figure 23.11 Distributed model of the filter.

characteristics of the propagation mode (TE_{10} in this case). Since the final realization is based on inductive windows and $\lambda_{g0}/2$ waveguide resonators, the suitable distributed model[2] is represented with K-inverters in Figure 23.11. The waveguide chosen for this example is WR-229 (58.116 × 29.083 mm). The values of the normalized K-inverters are computed as follows (see Appendix E):

$$\overline{K}_{01} = \overline{K}_{34} = M_{S1}\sqrt{\frac{n\pi}{2}\mathcal{W}_\lambda} = 0.17749$$

$$\overline{K}_{12} = \overline{K}_{23} = M_{12}\frac{n\pi}{2}\mathcal{W}_\lambda = 0.02770$$

where

$$\lambda_{g0} = \frac{\lambda_{g1} + \lambda_{g2}}{2}, \quad \mathcal{W}_\lambda = \frac{\lambda_{g1} - \lambda_{g2}}{\lambda_{g0}}$$

and

- n number of half wavelengths for the resonant mode (in this case $n = 1$)
- λ_{g1} wavelength at lower band-edge frequency
- λ_{g2} wavelength at upper band-edge frequency.

For narrow BWs, the frequency response using the distributed model is virtually the same as obtained using the lumped model as shown in Figure 23.10.

Now, it is time to obtain the physical dimensions of the filter using EM techniques (this is called the EM model). The translation of transmission lines to waveguide lengths is straightforward, but the inverters have to be translated to coupling windows of suitable dimensions. The model of an inductive window is described in Figure 23.12 (see Table 14.2). The model is suitable for thick irises, and in this case the thickness t of all windows is fixed at $t = 2$ mm. The phase ϕ of the model is negative, while X_s and X_p (normalized with regard to Z_0, characteristic impedance of the waveguide fundamental mode) are positive.

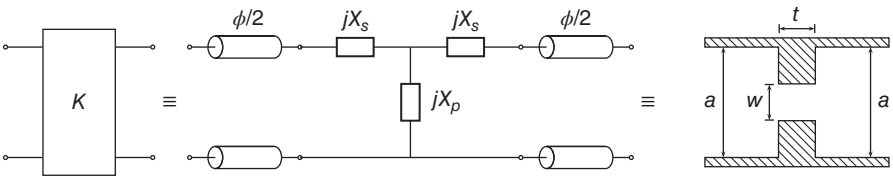


Figure 23.12 Inverter model (center) for an inductive window in waveguide technology. The top view of the inductive window (right) shows two main design parameters: the width (*w*) and the thickness (*t*).

2 The physical behavior of the filter is like almost-closed (short-circuit) half-wavelength resonators modeled with series resonators and inverters acting as shunt high reactances. This is the dual circuit of the model shown in Figure 23.9.

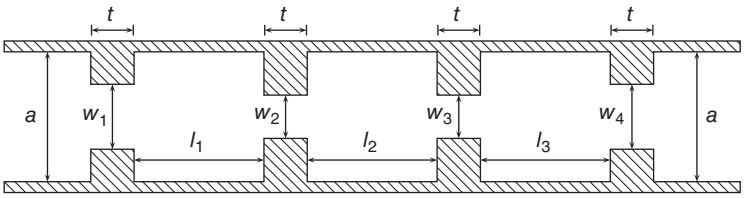


Figure 23.13 Layout of the filter with inductive irises.

The algorithm for the symmetrical iris design follows these steps:

1) Obtain an initial estimation for the iris width, using, for example, models from [3] or some EM data of previous designs.
2) Perform an EM simulation at the center frequency of the iris to obtain S_{11} and S_{12}.
3) Obtain the values of the model with the formulas of Section 14.5:

$$jX_s = \frac{1 - S_{12} + S_{11}}{1 - S_{11} + S_{12}}$$

$$jX_p = \frac{2S_{12}}{(1 - S_{11})^2 - S_{12}^2}$$

4) Obtain the value for the real inverter of the first iteration $K^{(1)}$ that is given by

$$K = \left| \tan\left[\frac{\phi}{2} + \arctan(X_s)\right]\right|$$

$$\phi = -\arctan(2X_p + X_s) - \arctan(X_s)$$

5) Compare $K^{(1)}$ with the desired K. If $K^{(1)}$ is lower than K, then increase the coupling (open slightly the inductive window); otherwise decrease the coupling (close slightly the window).
6) Repeat steps 2–5 n times until the difference $K - K^{(n)}$ becomes negligible.
7) With the final value of ϕ for all the inverters, obtain the modified lengths of the adjacent resonators. Resonator r will be modified by inverters $r-1$ and r (the inverters at each end of the resonator r):

$$l_r = \frac{\lambda_{g0}}{2\pi}\left[\pi + \frac{1}{2}(\phi_{r-1} + \phi_r)\right], \quad r = 1, \ldots, N$$

The dimensions (all in mm) obtained following this procedure are

$$\begin{aligned} a &= 58.115, \quad w_1 = w_4 = 20.937, \quad l_1 = l_3 = 45.012 \\ b &= 29.083, \quad w_2 = w_3 = 10.663, \quad l_2 = 47.624 \\ t &= 2, \quad \text{for all irises} \end{aligned} \tag{23.1}$$

The final layout is depicted in Figure 23.13. The physical dimensions given by the algorithm can be plugged in an EM full-wave simulator, giving the response shown in Figure 23.14 without further optimization.

Example 23.4 ***EM Model of a Dual-Mode Filter***
In this example, we consider a quasielliptic filter in waveguide technology using dual-mode configuration and with the following specifications:

- Order $N = 4$
- RL = 22

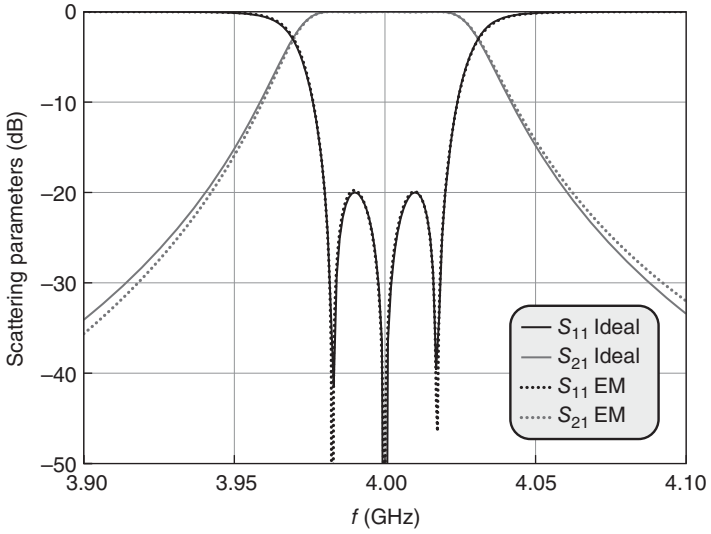


Figure 23.14 Full-wave simulation of the filter response. The solid line is the reference curve of the ideal response.

- Out-of-band rejection greater than 40 dB
- Center frequency $f_0 = 12$ GHz
- BW = 76 MHz
- Unloaded quality factor $Q_u > 10{,}000$
- Structure: circular waveguide, TE_{113} propagation mode in dual-mode configuration.

In order to achieve the required out-of-band rejection, the location of the transmission zero is determined to be at $s = \pm j2.6207$ for the lowpass prototype filter. With the order N, the RL, and the location of the transmission zeros, the recursive technique proposed in Chapter 6 can be applied to obtain the polynomials of the scattering parameters

$$S_{21}(s) = \frac{P(s)/\varepsilon}{E(s)}, \quad S_{11}(s) = \frac{F(s)/\varepsilon_R}{E(s)}$$

where $\varepsilon = 4.0532$ and $\varepsilon_r = 1$. The location of the poles is

$$p_{1,2} = -0.2973 \pm j1.2310$$
$$p_{3,4} = -0.8565 \pm j0.5713$$

The polynomial coefficients are shown in Table 23.5. The filter is dual mode, and therefore the routing diagram is a single quartet as shown in Figure 23.15.

The CM configuration of a single quartet is in fact a folded form of a (4-2) filter, and therefore the corresponding $N + 2$ CM is obtained following the algorithm given in Chapter 8:

$$\mathbf{M} = \begin{pmatrix} 0 & 1.0741 & 0 & 0 & 0 & 0 \\ 1.0741 & 0 & 0.9359 & 0 & -0.1069 & 0 \\ 0 & 0.9359 & 0 & 0.7666 & 0 & 0 \\ 0 & 0 & 0.7666 & 0 & 0.9359 & 0 \\ 0 & -0.1069 & 0 & 0.9359 & 0 & 1.0741 \\ 0 & 0 & 0 & 0 & 1.0741 & 0 \end{pmatrix}$$

Table 23.5 Polynomial coefficients for the normalized polynomials $E(s)$, $F(s)$, and $P(s)$ of Example 23.4.

	$E(s)$	$P(s)$	$F(s)$
s^4	1.0000	—	1.0000
s^3	2.3076	—	—
s^2	3.6821	$j1.0000$	1.0197
s^1	3.3774	—	—
s^0	1.6998	$j6.8681$	0.1350
	$\varepsilon_R = 1.0000$		$\varepsilon = 4.0532$

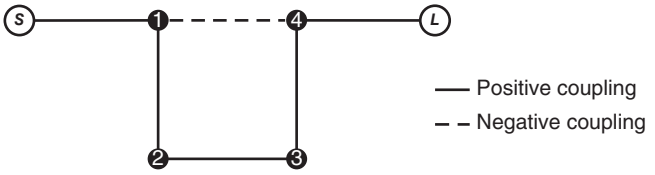


Figure 23.15 Routing diagram for the filter of Example 23.4.

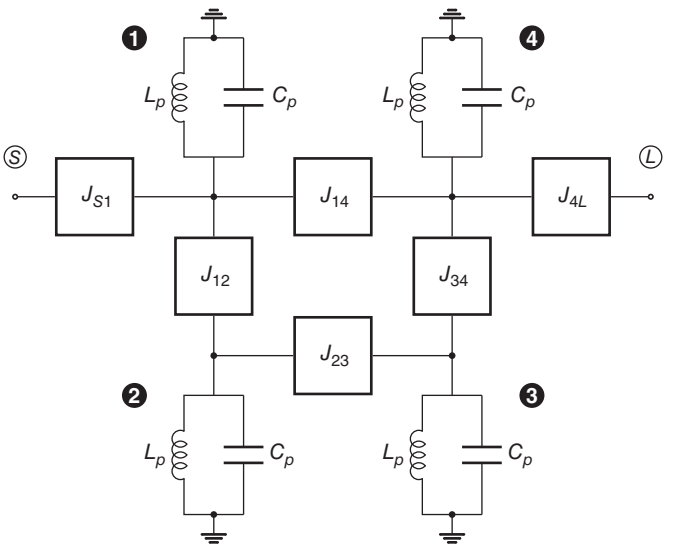


Figure 23.16 Lumped model of the dual-mode filter.

The lowpass prototype circuit can be viewed as shunt capacitors with $C = 1$ F in the locations of the resonating nodes, joined through J-inverters given by the CM. The corresponding bandpass lumped model is obtained substituting the shunt capacitors with parallel resonators as shown in Figure 23.16. All resonators are exactly equal, and the element values are given by the lowpass to bandpass transformation applied to a shunt capacitor of value $C = 1$ F:

$$C_p = \frac{1}{2\pi\ \mathrm{BW}} = 2.094\ \mathrm{nF}$$

$$L_p = \frac{1}{4\pi^2 f_0^2 C_p} = 0.084\ \mathrm{pH}$$

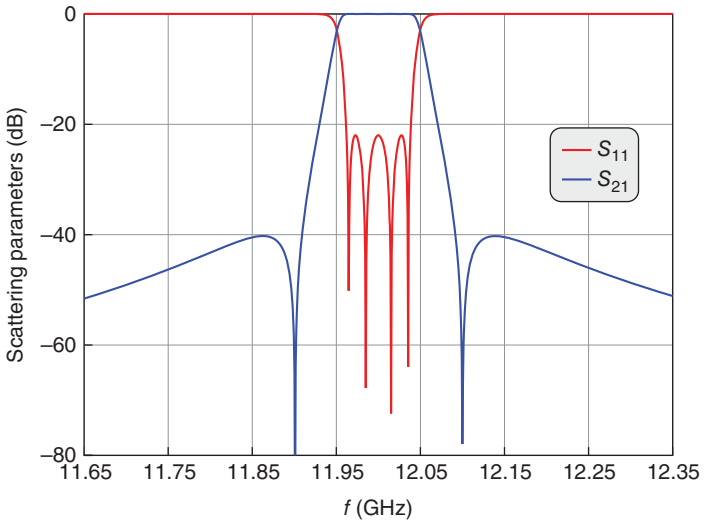


Figure 23.17 Scattering parameters of the lumped model for the dual-mode filter.

The electrical response of the lumped model meeting the specifications is shown in Figure 23.17.

Following the same steps as in Example 23.3, a distributed model would be the natural step forward. For this kind of filters, a simple distributed model has been given in [4]. Using the most common configuration, that is, the TE_{113} mode of the circular cavities, the model has the schematic depicted in Figure 23.18. Using the TE_{113} mode means that all resonators have an electrical length of $3\lambda_g/2$. The only parameter to be fixed in this model is the radius of the circular waveguide. The choice of this value is determined by two competing factors:

- The filter structure should not allow the propagation of any modes except TE_{113} within the allocated frequency band 11.7–12.2 GHz.
- Maximization of unloaded Q.

It is essential to meet the first criterion, and that sets the limit of achievable Q_0. However, it is not a significant drawback since TE_{113} mode is chosen as the propagation mode precisely for the reason to achieve a higher value of quality factor as compared with the dominant TE_{111} mode, at the expense of higher volume and mass of the filter structure.

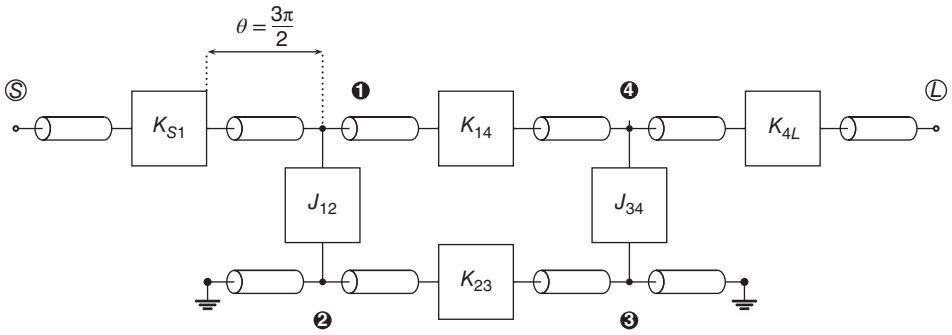


Figure 23.18 Distributed model for the dual-mode filter. The line lengths are all equal, assuming a propagating TE_{113} mode.

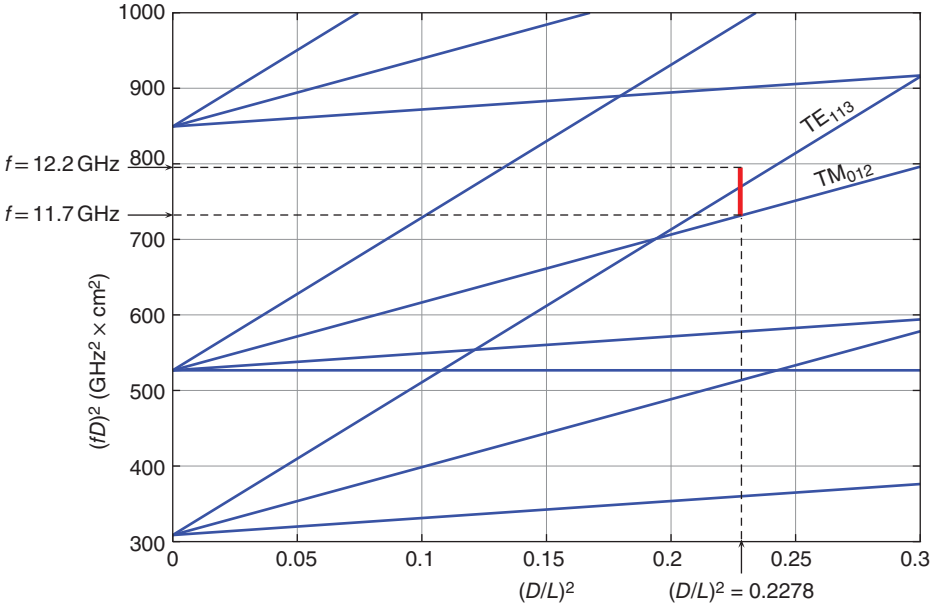


Figure 23.19 Representation of $(fD)^2$ versus $(D/L)^2$ for a circular waveguide resonator. For the shape factor $(D/L)^2 = 0.2278$ chosen for Example 23.4.

The common procedure to choose the resonator dimensions is through a graph representing the $(fD)^2$ product for each resonant mode (f stands for center frequency and D stands for resonator diameter) in terms of the shape factor of the resonator D/L (diameter D divided by resonator length L). This kind of graph is composed of straight lines as shown in Figure 23.19, where the line corresponding to the TE_{113} mode is crossed by a vertical line representing a fixed shape factor in the frequency span covering 11.7–12.2 GHz. Additionally, the vertical line does not cross the line corresponding to the TM_{012} mode close to the lower frequency limit, thus meaning that ideally no spurious resonance will appear in the band of interest.

The usual procedure is to choose, using mere visual inspection, a shape factor producing a point $(fD)^2$ as far away as possible from the other lines corresponding to spurious resonances. For this example, the criterion has been to have a spurious-free window between 11.7 and 12.2 GHz and, additionally, having the maximum range free of spurious in the upper side of the passband. In this case, we obtain the following physical dimensions for the TE_{113} resonance at $f_0 = 12$ GHz ($\lambda_g = 32.29$ mm):

$$(D/L)^2 = 0.2278 \Rightarrow \begin{cases} \text{Waveguide radius} \;\; a = D/2 = 11.557 \text{ mm} \\ \text{Resonator length} \;\; L = 48.43 \text{ mm} \end{cases}$$

We are now ready to obtain the values of the inverters shown in Figure 23.18:

$$\overline{K}_{S1} = M_{S1}\sqrt{\frac{3\pi}{2}\mathcal{W}} = 0.2398$$

$$\bar{J}_{12} = M_{12}\frac{3\pi}{2}\mathcal{W} = 0.0467$$

$$\overline{K}_{14} = M_{14}\frac{3\pi}{2}\mathcal{W} = -0.0053$$
$$\overline{K}_{23} = M_{23}\frac{3\pi}{2}\mathcal{W} = 0.0382$$
$$\bar{J}_{34} = M_{34}\frac{3\pi}{2}\mathcal{W} = 0.0467$$
$$\overline{K}_{4L} = M_{4L}\sqrt{\frac{3\pi}{2}}\mathcal{W} = 0.2398$$

where $\mathcal{W} = \frac{\lambda_{g1}-\lambda_{g2}}{\lambda_{g0}}$. The distributed model is fully determined, and its electrical response, as shown in Figure 23.20, is very close to the response given by the lumped model in Figure 23.17.

The next step in the design process is to transform the quasielliptic distributed model into an EM model based on physical dimensions. The physical structure for a dual-mode filter composed of a single quartet is shown in Figure 13.11. In this model there are four steps to fulfill:

1) Design of the cruciform iris
2) Design of the input/output iris
3) Design of the tuning screws
4) Design of the coupling screws.

For the first step, the method applied in Example 23.3 for obtaining the iris dimensions can be carried out. The two arms of the cross are nearly independent, since both arms are narrow (see Figure 23.21), and therefore they can be designed independently.

The second step is a slight variation of the first one, since the input/output irises are asymmetrical. The input waveguide is a standard rectangular waveguide, whereas the input/output cavities are circular waveguides (see Figure 23.22). Therefore, the formulas given in Chapter 14 have to be extended to the nonsymmetrical iris. The input/output inverter can be modeled with

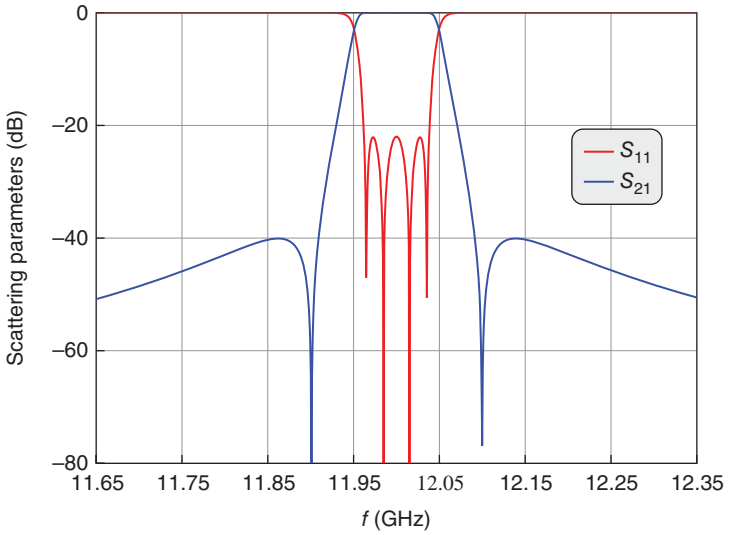


Figure 23.20 Scattering parameters of the distributed model for the dual-mode filter.

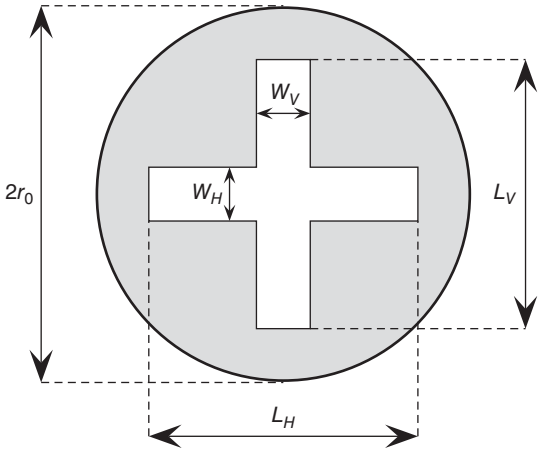


Figure 23.21 Cruciform iris allocated between the two dual-mode cavities. Dimensions of both arms are designed independently assuming $L_H \gg W_H$ and $L_V \gg W_V$.

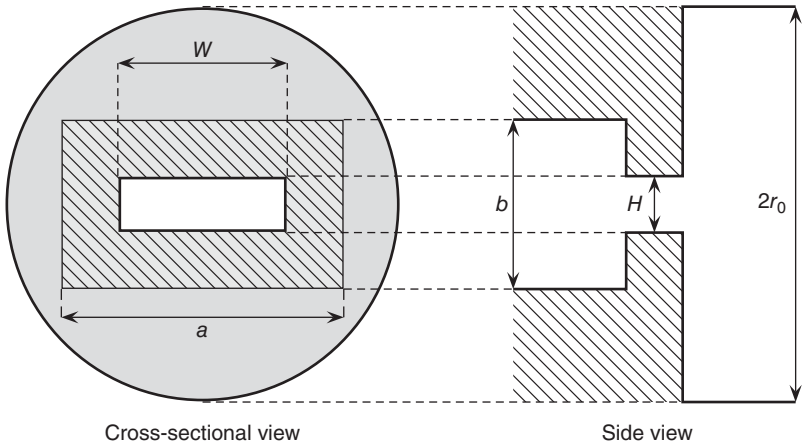


Figure 23.22 Input/output irises are rectangular to circular waveguide junctions.

the equivalent circuit shown in Figure 23.23. The values of the circuit parameters are obtained as [4]

$$jX_{s1} = \frac{(1+S_{11})(1-S_{22}) + S_{21}^2 - 2S_{21}}{(1-S_{11})(1-S_{22}) - S_{21}^2}$$

$$jX_{s2} = \frac{(1-S_{11})(1+S_{22}) + S_{21}^2 - 2S_{21}}{(1-S_{11})(1-S_{22}) - S_{21}^2}$$

$$jX_p = \frac{2S_{21}}{(1-S_{11})(1-S_{22}) - S_{21}^2}$$

$$\phi_1 = -\arctan\left(\frac{\Sigma}{\Pi_-}\right) - \arctan\left(\frac{\Delta}{\Pi_+}\right)$$

$$\phi_2 = -\arctan\left(\frac{\Sigma}{\Pi_-}\right) + \arctan\left(\frac{\Delta}{\Pi_+}\right)$$

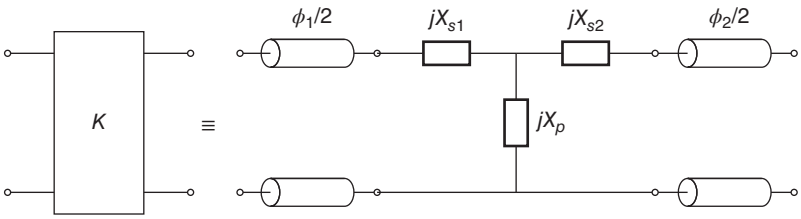


Figure 23.23 Equivalent circuit for the input/output inverters.

where

$$\Sigma = X_{s1} + X_{s2} + 2X_p$$
$$\Delta = X_{s1} - X_{s2}$$
$$\Pi_+ = 1 + X_{s1}X_{s2} + X_p(X_{s1} + X_{s2})$$
$$\Pi_- = 1 - X_{s1}X_{s2} - X_p(X_{s1} + X_{s2})$$

The inverter value can be obtained as

$$K = \sqrt{\left|\frac{1 + \Gamma e^{-j\phi_1}}{1 - \Gamma e^{-j\phi_1}}\right|}$$
$$\Gamma = \frac{j\Delta - \Pi_+}{j\Sigma + \Pi_-}$$

The algorithm for finding the iris dimensions shown in Example 23.3 can be completely followed. With regard to step 3, the penetration of the tuning screws is designed to achieve a resonance at f_0. Using an EM simulation of the group delay τ_g of the S_{11} parameter of the circular cavity (with the designed irises at both ends), the penetration is changed until the peak value of τ_g is located at f_0. In a circular resonator two orthogonal polarizations coexist, and therefore, two different screw penetrations per cavity have to be designed, but the procedure is exactly the same for the orthogonal mode. The tuning screws are labeled T_i in Figure 23.24 and are placed orthogonally to tune both polarizations independently.

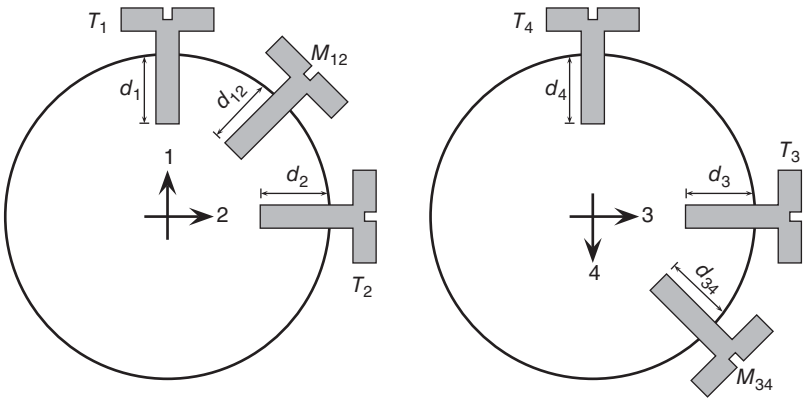


Figure 23.24 Location of the tuning T_i and coupling M_i screws for both cavities. The position for the coupling screws in the adjacent cavities is different because a virtual negative coupling is created.

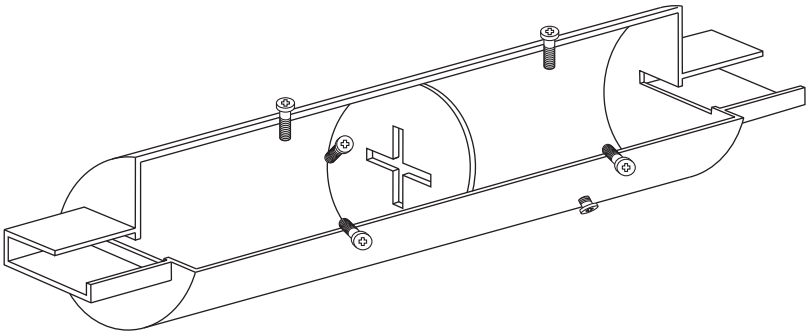


Figure 23.25 3D model of the dual-mode filter.

Finally, step 4 (penetration of the coupling screws) is carried out using curve fitting. The response of the distributed model for the first cavity is compared with the EM response of the corresponding structure as explained in [4]. Each process (one per coupling screw) is quite straightforward, because there is only one parameter per process to be obtained (penetration d_{12} or d_{34}). Fine-tuning can be performed by slightly moving the corresponding tuning screws.

Following the algorithm, the final dimensions (all given in mm) are computed as follows:

- Access ports: WR-75 ($a = 19.05$, $b = 9.525$)
- Iris thickness (i/o irises and cruciform irises) = 1
- Input iris: $W = 10.1172$, $H = 2$
- Cruciform iris:
  - $W_H = 1$, $L_H = 5.0786$
  - $W_V = 1$, $L_V = 8.4025$
- Cavity lengths: $L_1 = 47.2720$, $L_2 = 47.3156$
- Screws:
  - Thickness = 2
  - $d_1 = 1.5$, $d_2 = 2.907$, $d_3 = 2.8883$, $d_4 = 1.5$
  - $d_{12} = 2.6307$, $d_{34} = 2.4153$.

The 3D layout of the filter is shown in Figure 23.25. Its electrical response without performing an overall optimization is shown in Figure 23.26. This performance is close to the ideal response and demonstrates the efficacy of the distributed model. An optimization procedure can then be applied to the dimensions that would bring the filter even closer to the ideal response. The next example includes such an optimization procedure, which is general and can be applied to any kind of design.

Example 23.5 ***Combline Filter***

The assumed filter parameters are Chebyshev function, order $N = 4$, RL = 20 dB, and an inline topology. Frequency band chosen is 1805–1880 MHz, this band is widely used for wireless communication services.

Solution:
The filter BW = 75 MHz (1805–1880 MHz) is centered at

$$f_0 = \sqrt{1805 \times 1880} \approx 1842.12 \text{ MHz}$$

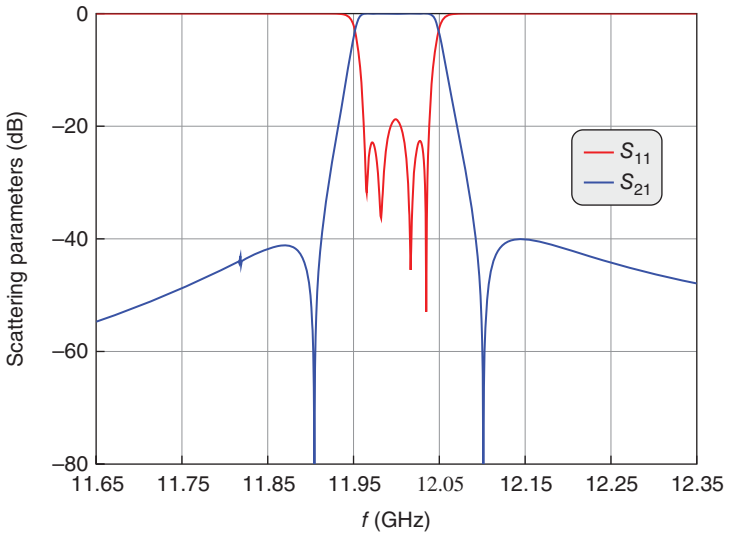


Figure 23.26 Scattering parameters obtained with the full-wave EM simulation of the dual-mode filter without optimization.

The filter has a relative BW of $\frac{\text{BW}}{f_0} \times 100 \approx 4\%$. The inverter values for this filter are directly obtained using (14.1)

$$M_{S1} = M_{4L} = 1.03515$$
$$M_{12} = M_{34} = 0.91058$$
$$M_{23} = 0.69992$$

The filter is synchronously tuned (all resonators have the same frequency f_0) and therefore $M_{ii} = 0$ for all i.

The cavity configuration is chosen to achieve high Q_0 and a wide spurious-free window reflecting typical requirements for practical applications. The geometry of the cavity is depicted in Figure 23.27.

The cavity dimensions have been set to $a = 5$ cm, $b = 5$ cm, and $c = 4$ cm to obtain the parallelepiped $a \times b \times c$. The radius of the cylindrical metallic post is set to $R = 8$ mm. Then h/c (ratio between the height of the post h and the height of the cavity c) is computed to obtain the required resonance at f_0. In this case $h/c = 0.7277$.

In order to obtain the coupling values given by the M_{ij} elements, windows with fixed height $w_h = 15$ mm, fixed thickness $t = 3$ mm, and suitable width are chosen. The geometry of the window is depicted in Figure 23.28.

The next step is to obtain the iris dimensions for the different interresonator couplings. The procedure is analogous to the method used in Chapter 14. We use the electric and magnetic wall symmetry analysis to obtain the f_e and f_m frequencies. An EM simulator is used to compute the resonant frequencies of the even and odd modes. The structure to be simulated is composed of two resonators with $h/c = 0.7277$ coupled through a coupling window. If symmetry is applied to the EM simulation, only the structure shown in Figure 23.28 has to be analyzed with an electric or magnetic wall placed in the symmetry plane (the plane of the iris). Otherwise, two exactly equal resonators coupled through an iris window have to be analyzed, and both resonant frequencies need to be identified.

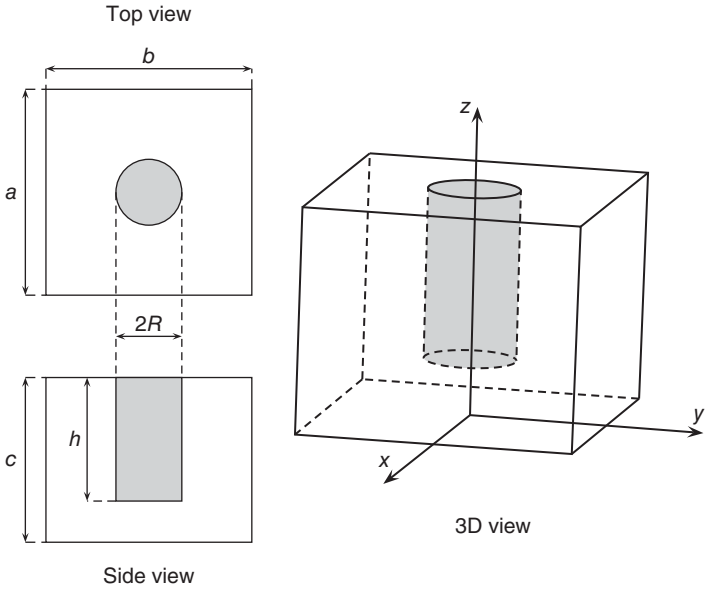


Figure 23.27 Geometry for the cavities of the combline filter. Top left: Top view (*x*–*y* plane). Bottom left: Side view (*y*–*z* plane). Right: 3D view with the associated coordinate system.

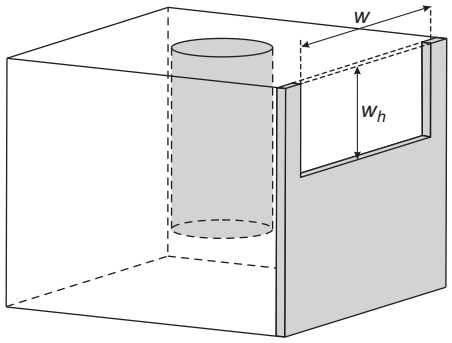


Figure 23.28 Cavity with the coupling window. $w_h = 15$ mm in this example. The parameter used to control the coupling value is the width of the window *w*. The thickness of the coupling window is fixed at 3 mm.

The results of several simulations are collected in Table 23.6. If some value of M is required, and the table does not provide the suitable width, then interpolation methods are usually applied. It is important to remark that every row of Table 23.6 requires an EM simulation with high precision, thus requiring a significant computational effort.

By interpolating data of Table 23.6, a smooth curve can be obtained (see Figure 23.29). Using this curve, the corresponding values of the irises placed between resonators are found to be

$$M_{12} = M_{34} = 0.9106 \Rightarrow w_{12} = w_{34} = 3.50932 \text{ cm}$$
$$M_{23} = 0.6999 \Rightarrow w_{23} = 3.06847 \text{ cm}$$

Once the interresonator irises are calculated, a better refinement has to be applied. In the previous step, two resonators are simulated with a coupling window placed between them. There are no other apertures for the resonators in such analysis. Indeed, this structure is too ideal, since each resonator within an inline filter is actually connected to two resonators (except the first and last ones connected to the input/output). The loading effect in the cavity

Table 23.6 Computed values of f_e, f_m, and the corresponding coupling matrix value M for some values of iris width w.

w (cm)	f_e (GHz)	f_m (GHz)	M
2.4	1.836641	1.806892	0.400
2.6	1.837332	1.801157	0.488
2.8	1.836419	1.794826	0.562
3.0	1.836686	1.788311	0.654
3.1	1.836286	1.784791	0.698
3.2	1.836376	1.775877	0.824
3.4	1.836089	1.774696	0.835
3.6	1.836110	1.766641	0.947
3.8	1.836344	1.759624	1.047
4.0	1.836040	1.753646	1.120
4.2	1.835843	1.748760	1.192
4.4	1.835910	1.743730	1.264

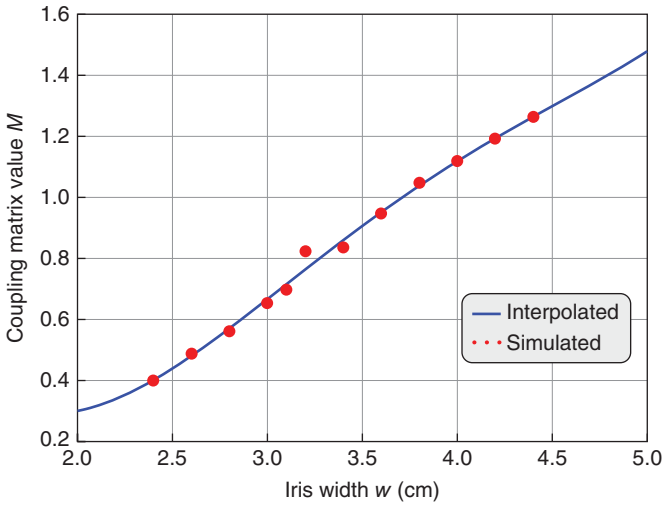


Figure 23.29 Interpolated curve for interresonator coupling versus iris dimensions.

resonant frequency is too important to be neglected, and a good improvement is achieved after simulating a single resonator with its two corresponding irises. The loading effect is simulated using detuned adjacent resonators (see Figure 23.30) as it was proposed in Section 14.4.2. Once the simulation is performed, a slight deviation from the resonant frequency is observed. This deviation is corrected by changing the height of the metallic post to tune the resonator again at the filter center frequency f_0. The new ratio for resonator 2, and by symmetry for resonator 3, is $h/c = 0.6968$.

The input/output resonators are designed to achieve the external quality factor given by the CM with the resonator kept tuned at f_0. The iris loading effects are also included. The coupling with the source/load is obtained through a coaxial probe whose position gives the required Q_{ext}.

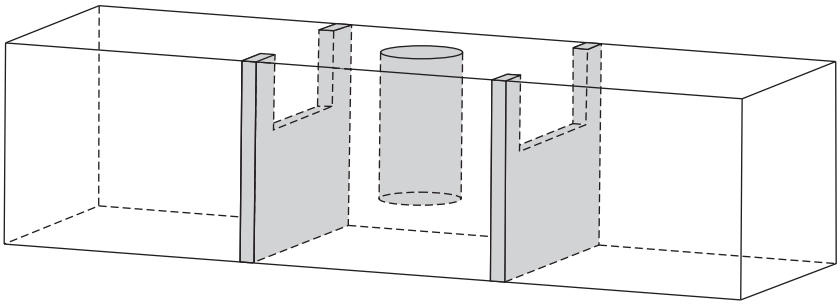


Figure 23.30 Resonator with irises coupled to adjacent detuned resonators.

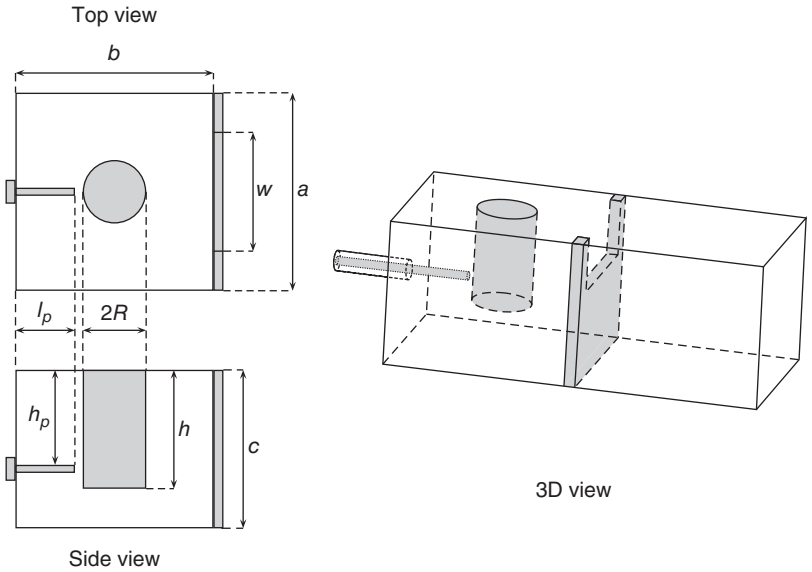


Figure 23.31 First/last resonators with the input/output probe. Left: Top and side view. Right: 3D view of the structure showing the coupling iris with the second cavity that has been detuned (removing the metallic post) to design the first cavity with an approximate loading.

In this case

$$Q_{\text{ext}} = \frac{f_0}{\text{BW R}_1} = \frac{f_0}{\text{BW M}_{S1}^2} = 22.92$$

The method used for designing the Q_{ext} is the same one explained in Section 14.3.2, where the group delay of the S_{11} (at f_0) is used to optimize the parameters of the input cavity. The structure to be optimized with an EM simulator is shown in Figure 23.31. In order to achieve the required Q_{ext}, as well as to center the resonant frequency of the input cavity, the parameters to be optimized are as follows:

- The length of the probe l_p that is inserted inside the cavity. After the optimization process, the optimum value has been found to be $l_p = 16.4$ mm.
- The position of the probe referred to the top wall of the cavity (h_p in Figure 23.31). In the other dimension, the probe is centered as shown in the top view (see top left of Figure 23.31). The optimum value is $h_p = 20.256$ mm.

- The ratio h/c (length of the post related to the height of the cavity). The optimum value is $h/c = 0.651676$.

The optimization procedure with the first/last resonators is very similar to the one explained in Section 14.4.3. The parameters R, w_h, a, b, and c are the same ones used for the rest of the filter. The probe has the values of inner radius $a_p = 1.1176$ mm and outer radius $b_p = 2.54$ mm, producing an impedance value near 50 Ω.

After the calculation of the input/output coupling, all dimensions are obtained and a first complete EM simulation can be performed. The result of the simulation compared with the ideal response is shown in Figure 23.32. The physical dimensions of all parameters of the filter are collected in Table 23.7, and the layout is sketched in Figure 23.33.

If the response satisfies a given set of requirements, there is nothing more to do. In this example, we demonstrate the next step, mostly required for practical filters: the optimization of

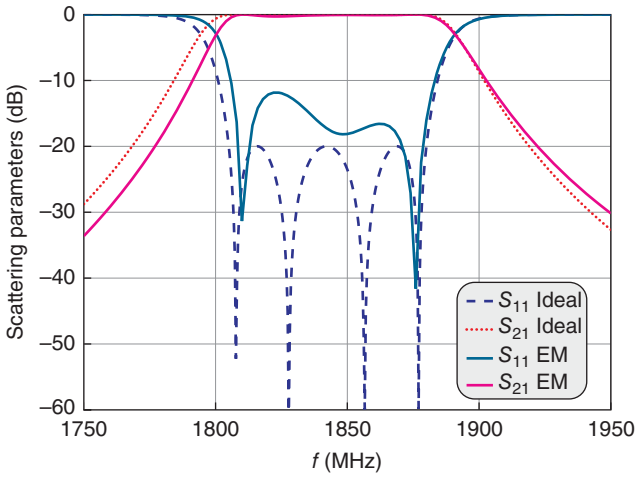


Figure 23.32 EM simulation of the filter compared with the ideal response.

Table 23.7 Dimensions of the designed four-pole combline filter waveguide.

Parameters	Dimensions (mm)
Cavity dimensions	$a = b = 50$ $c = 40$
Post size	$R = 8$ $h_1 = h_4 = 26.067$ $h_2 = h_3 = 27.872$
Irises	$w_{12} = w_{34} = 35.0932$ $w_{23} = 30.6847$ $w_h = 15$ $t = 3$
Input/output probe dimensions	$h_p = 20.256$ $l_p = 16.4$ $a_p = 1.1176$ $b_p = 2.54$

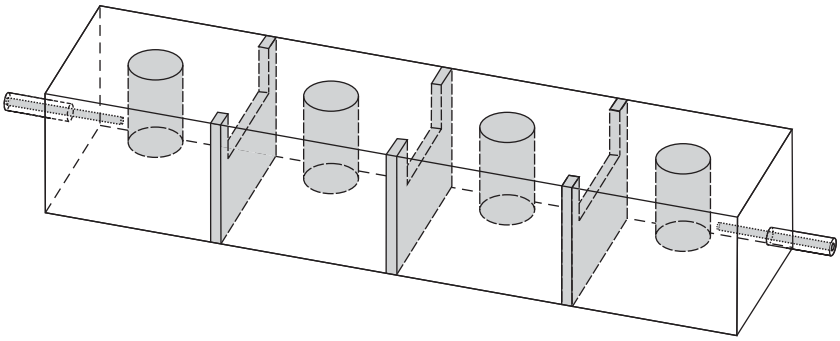


Figure 23.33 Structure of the four-pole combline waveguide filter.

the overall structure. For this type of design, the aggressive space mapping (ASM) technique [5] and the calibrated coarse model (CCM) technique [6] are usually the best choices (see Chapter 15). Here, we are going to apply a simplified version of the CCM for demonstration purposes.

The CCM algorithm requires two models (as ASM does) named coarse (usually fast but somewhat inaccurate) and fine (usually slow but very accurate). In this design the coarse model is the segmented EM model (i.e., the physical dimensions built with the procedure explained above starting from the CM). The fine model will be the EM model built as a whole (i.e., with all full-wave interactions taken into account) that uses physical dimensions as parameters. Exploiting the symmetry of the structure, the parameters chosen for the optimization in both models are w_{12} (coupling window between resonators 1 and 2), w_{23} (coupling window between resonators 2 and 3), h_1/c (ratio between post height and cavity height in resonator 1), h_2/c (ratio between post height and cavity height in resonator 2), and l_p (penetration of the probe).

Hence

$$\mathbf{x}_c = \mathbf{x}_f = \begin{pmatrix} l_p \\ w_{12} \\ w_{23} \\ h_1/c \\ h_2/c \end{pmatrix}$$

The following steps are the simplified version of the CCM technique employed for this particular example:

Step 0. We start with the ideal CM $\mathbf{M}^*$. We enforce our target CM $\mathbf{M}_{\text{obj}}^{(0)} = \mathbf{M}^*$. Number of iterations is set to $i = 0$.

Step 1. Extract $\mathbf{x}_c^{(i)}$ to recover the matrix $\mathbf{M}_{\text{obj}}^{(i)}$. When $i = 0$ this procedure is the one we followed before to obtain the physical dimensions for the first time. For $i \neq 0$ this procedure involves designing the filter as it has been done up to now but starting from $\mathbf{M}_{\text{obj}}^{(i)}$ instead of the ideal CM.

Step 2. Set $\mathbf{x}_f^{(i)} = \mathbf{x}_c^{(i)}$ and simulate to obtain the response of the fine model $R(\mathbf{x}_f^{(i)})$. Stop if $R(\mathbf{x}_f^{(i)})$ meets the required performance.

Step 3. Obtain a CM $\mathbf{M}_f^{(i)}$ using the response just obtained $R(\mathbf{x}_f^{(i)})$. Usually, this step is done through optimization, using a circuit model, and is called parameter extraction.

Step 4. Compute $\Delta\mathbf{M}^{(i)} = \mathbf{M}_f^{(i)} - \mathbf{M}^*$. We define a sufficiently small η, and if $\| \Delta\mathbf{M}^{(i)} \| < \eta$, we stop the process.

Step 5. The new target CM is computed as $\mathbf{M}_{\text{obj}}^{(i+1)} = \mathbf{M}_{\text{obj}}^{(i)} - \Delta\mathbf{M}^{(i)}$.

Step 6. $i = i + 1$ and go to step 1.

In this example, only three iterations of the CCM algorithm have been required. The evolution of the algorithm is shown in Figure 23.34. The parameter evolution in the fine model is collected in Table 23.8. The parameter extraction procedure provides the CM $\mathbf{M}_f^{(i)}$ producing the response in the fine model $R(\mathbf{x}_f^{(i)})$. Table 23.9 shows the evolution of the extracted coupling matrices toward the optimal $\mathbf{M}^*$.

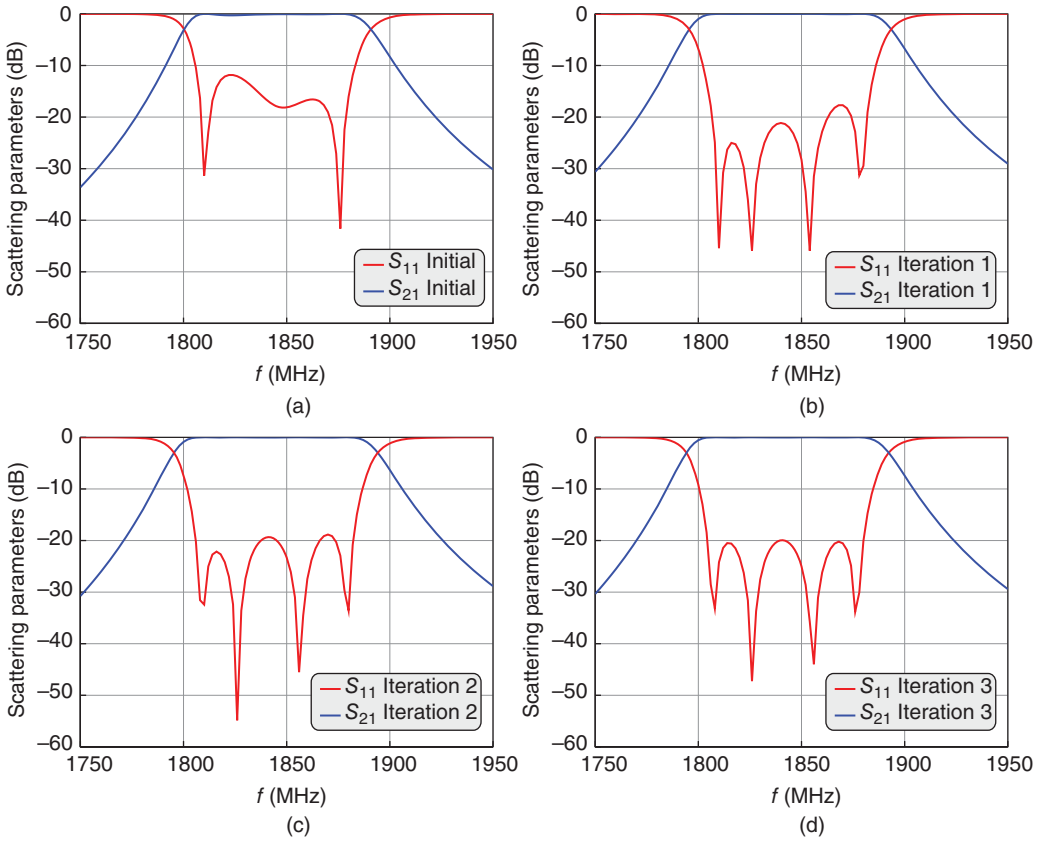


Figure 23.34 Evolution of the CCM algorithm for the combline four-pole filter. (a) Response of the initial point before optimization $R(\mathbf{x}_f^{(0)})$. (b) Response for the first iteration $R(\mathbf{x}_f^{(1)})$. (c) Response for the second iteration $R(\mathbf{x}_f^{(2)})$. (d) Response for the third iteration $R(\mathbf{x}_f^{(3)})$.

Table 23.8 Evolution of the optimization algorithm for the physical dimensions of the combline filter.

$\mathbf{x}_f$	$\mathbf{x}_f^{(0)}$	$\mathbf{x}_f^{(1)}$	$\mathbf{x}_f^{(2)}$	$\mathbf{x}_f^{(3)}$
l_p	16.400	16.379	16.368	16.368
w_{12}	35.093	36.930	36.930	36.930
w_{23}	30.685	30.890	30.890	30.890
h_1/c	0.6517	0.6539	0.6546	0.6546
h_2/c	0.6968	0.6954	0.6955	0.6961

Table 23.9 Evolution of the extracted coupling matrix.

$\mathbf{M}_f$	$\mathbf{M}_f^{(0)}$	$\mathbf{M}_f^{(1)}$	$\mathbf{M}_f^{(2)}$	$\mathbf{M}_f^{(3)} \approx \mathbf{M}^*$
M_{S1}	1.0583	1.0425	1.0351	1.0352
M_{12}	0.8284	0.9106	0.9106	0.9106
M_{23}	0.6941	0.6999	0.6999	0.6999
M_{11}	−0.120	0.0065	0.0000	0.0000
M_{22}	−0.053	−0.039	−0.0237	0.0000

The last column shows the ideal coupling matrix that coincides with the last extraction.

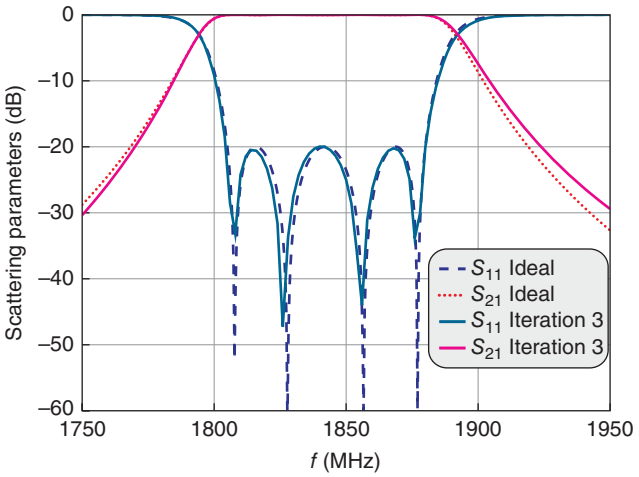


Figure 23.35 Final response of the filter after optimization.

In Figure 23.35, the final response $R(\mathbf{x}_f^{(3)})$ is compared with the ideal response, showing a very good agreement between both curves.

Example 23.6 ***EM Modeling of a Dielectric-Loaded Filter***

In this example we consider a dielectric resonator filter centered at $f_0 = 4$ GHz and having 1% BW. The assumed filter parameters are Chebyshev function, order $N = 4$, RL = 20 dB, and an inline waveguide topology.

The filter BW is 40 MHz centered at $f_0 = 4$ GHz. The inverter values for this filter are directly obtained using (14.1)

$$M_{S1} = M_{4L} = 1.03515$$
$$M_{12} = M_{34} = 0.91058$$
$$M_{23} = 0.69992$$

This filter has exactly the same order and RL as the filter designed in Example 23.5. As the center frequency and BW is irrelevant in the lowpass lumped model, the inverter values are exactly the same ones obtained before. The same criteria for the cavity configuration and constraints apply here.

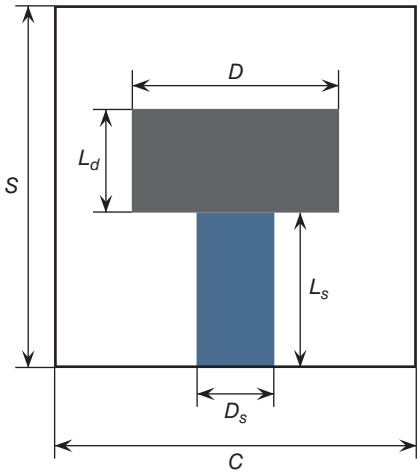


Figure 23.36 Side view of an isolated cavity for the dielectric filter.

Similar basic configuration for the cavity resonators proposed in Chapter 14 is considered here. The side view of a cavity with a dielectric resonator mounted on a dielectric electrical support is shown in Figure 23.36. The rectangular cavity dimensions are $(S \times C \times C)$ 24.8 mm × 24.5 mm × 24.5 mm. The dielectric resonator operates in $TE_{01\delta}$ mode with $\varepsilon_r = 34$, diameter $D = 14.4$ mm, and height $L_d = 5.88$ mm. The dielectric support has a dielectric constant $\varepsilon_r = 10$ and dimensions $L_s = 9.8$ mm and diameter $D_s = 6.86$ mm. These parameters have been designed to achieve a resonant frequency of 4 GHz.

The steps followed in this example are the same ones explained before in Example 23.5. They are summarized as follows:

- Design the isolated cavity to achieve a resonant frequency equal to the center frequency f_0.
- Design the coupling windows. The procedure is the same one followed in the dielectric filter of Chapter 14 and in Example 23.5. Simulate two cavities with a coupling window between them. Change the window size to obtain a lookup table, where the window size is related with the coupling value obtained using the magnetic and electric resonant frequencies. After interpolation of the simulated data, the coupling windows have the following aperture dimensions (constant thickness of 3.83 mm is assumed for all windows):

$$M_{12} = M_{34} = 0.9106 \Rightarrow w_{12} = w_{34} = 2.9651 \text{ mm}$$
$$M_{23} = 0.6999 \Rightarrow w_{23} = 1.93419 \text{ mm}$$

- Refine the cavity structure of inner resonators by simulating a resonator with its windows coupled to detuned resonators. In this example, the diameter of the dielectric resonators 2 and 3 (due to symmetry) have changed to $D = 14.242$ mm.
- Design the input/output cavities (see Figure 23.37). Use the group delay to obtain the right Q_e of the input/output resonators. Change the suitable parameters related with the probe considering the coupling window connecting the input/output resonator with the adjacent resonator. After some optimization the penetration of the probe has a value of $H = 12.7909$ mm. The diameter of the dielectric resonator is changed to $D = 14.3117$ mm in order to maintain the resonant frequency of the input/output resonator to $f_0 = 4$ GHz.
- Simulate the complete structure and optimize if the specifications are not met. In this example, the CCM algorithm has been used again.

The final structure is sketched in Figure 23.38. The electrical performance is shown in Figure 23.39 compared with the ideal response.

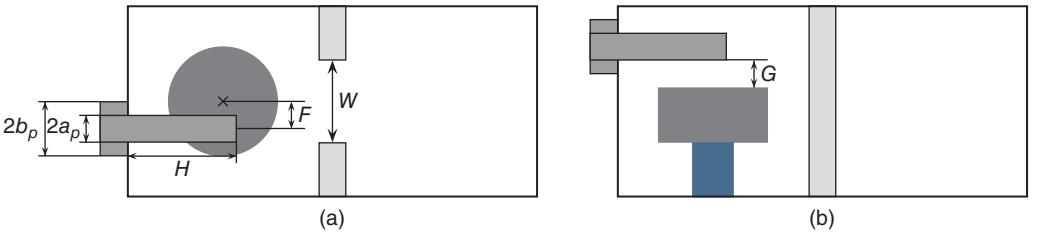


Figure 23.37 Input/output cavity showing the physical dimensions of the probes and the relative position to the dielectric resonator. $a_p = 1.1176$ mm, $b_p = 2.54$ mm, $F = 7.47$ mm, $G = 0.254$ mm. H is used as a design parameter. (a) Top view. (b) Side view.

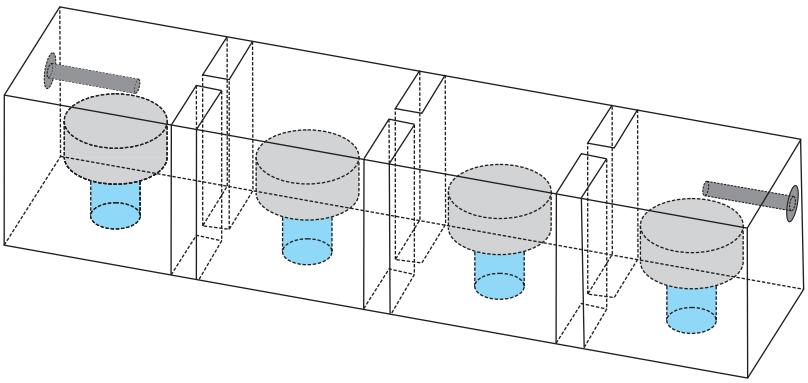


Figure 23.38 Complete structure of the dielectric resonator filter.

The optimization procedure is almost the same one followed in Example 23.5. At this stage, the reader should be familiar with the CCM optimization procedure. However, some data of the procedure is given below. That way the reader can trace the whole procedure.

The parameters to be optimized in this example are the coupling window widths w_{12} and w_{23}, the diameter of the resonators D_1 and D_2, and the penetration H of the probe inside the cavity. Hence

$$\mathbf{x}_c = \mathbf{x}_f = \begin{pmatrix} w_{12} \\ w_{23} \\ D_1 \\ D_2 \\ H \end{pmatrix}$$

which means that the parameters in the coarse and in the fine models are exactly identical. The ideal matrix $\mathbf{M}^*$ is the same one as in Example 23.5. Therefore the procedure is exactly the same, since there are five parameters as well. The only difference is the number of iterations until the stop criterion is reached. In this case there are five iterations (the starting point in this filter is worse as Figure 23.39 reveals, probably due to stronger interactions between the structure parts), and the evolution of the algorithm is shown in Figure 23.40.

The evolution of the physical dimensions is shown in Table 23.10, and the evolution of the CM is shown in Table 23.11.

Finally, Figure 23.41 shows the good agreement between the final performance of the filter and the ideal response taken as the design goal.

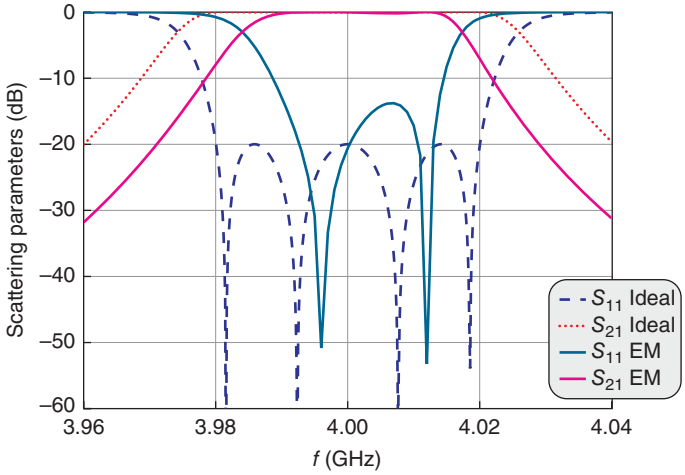


Figure 23.39 Frequency response of the dielectric resonator filter before optimization (solid lines). The ideal performance is shown in dotted (S_{21}) and dashed lines (S_{11}).

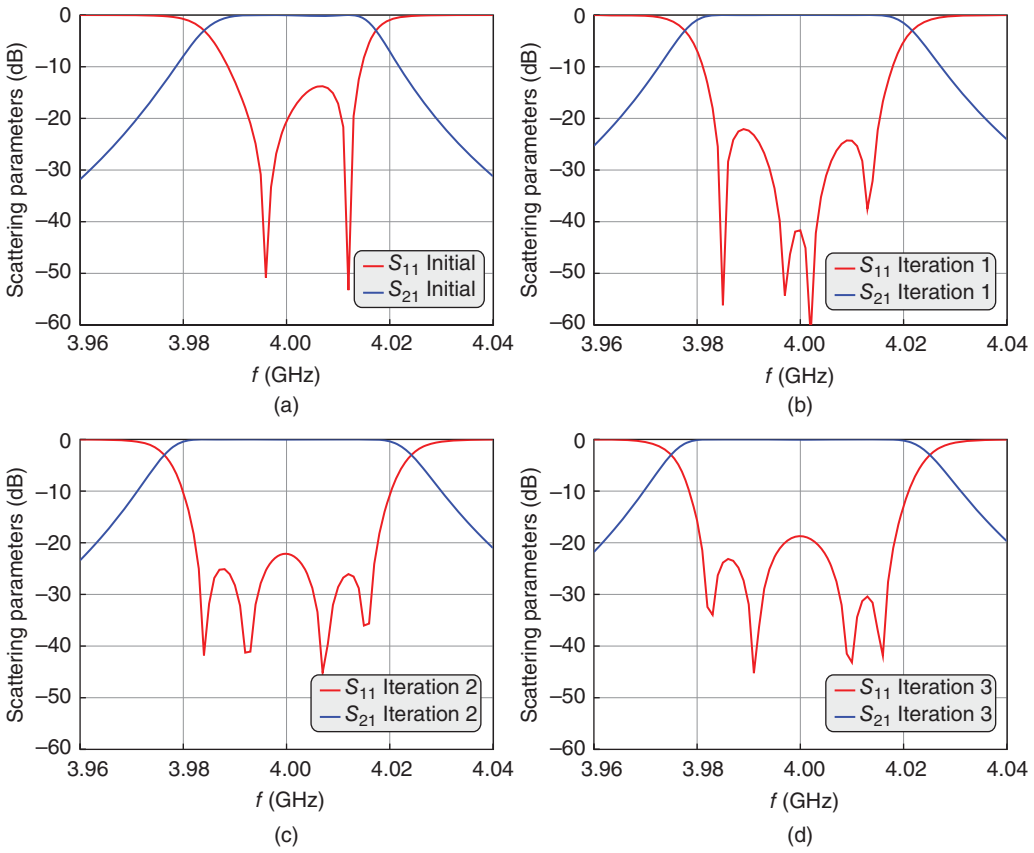


Figure 23.40 Evolution of the CCM algorithm for the dielectric four-pole filter. (a) Response of the initial point before optimization $R(\mathbf{x}_f^{(0)})$. (b) Response for the first iteration $R(\mathbf{x}_f^{(1)})$. (c) Response for the second iteration $R(\mathbf{x}_f^{(2)})$. (d) Response for the third iteration $R(\mathbf{x}_f^{(3)})$. (e) Response for the fourth iteration $R(\mathbf{x}_f^{(4)})$. (f) Response for the fifth iteration $R(\mathbf{x}_f^{(5)})$.

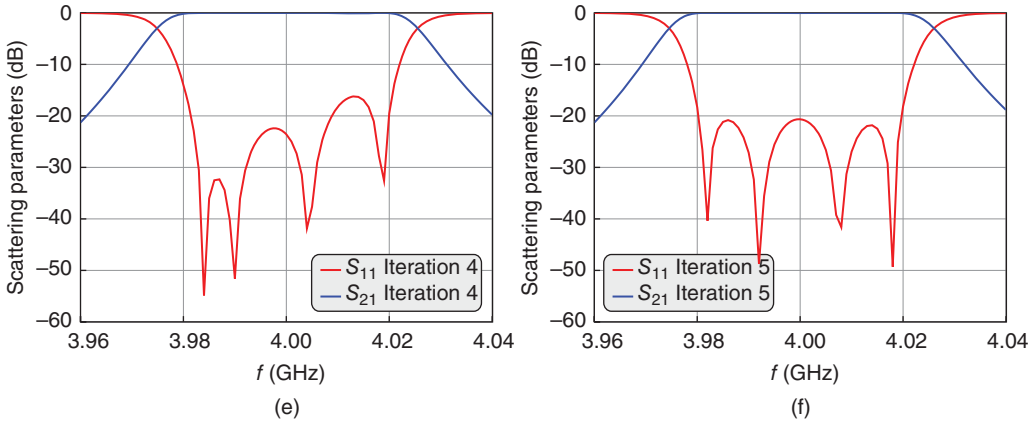


Figure 23.40 (*Continued*)

Table 23.10 Evolution of the optimization algorithm for the physical dimensions of the dielectric filter.

$\mathbf{x}_f$	$\mathbf{x}_f^{(0)}$	$\mathbf{x}_f^{(1)}$	$\mathbf{x}_f^{(2)}$	$\mathbf{x}_f^{(3)}$	$\mathbf{x}_f^{(4)}$	$\mathbf{x}_f^{(5)}$
w_{12}	2.960	4.130	4.780	5.170	5.200	5.289
w_{23}	1.930	3.080	3.550	3.850	4.044	4.131
D_1	14.312	14.271	14.256	14.248	14.255	14.245
D_2	14.242	14.212	14.195	14.189	14.182	14.181
H	12.791	12.791	12.788	12.785	12.7914	12.785

Table 23.11 Evolution of the extracted coupling matrix.

$\mathbf{M}_f$	$\mathbf{M}_f^{(0)}$	$\mathbf{M}_f^{(1)}$	$\mathbf{M}_f^{(2)}$	$\mathbf{M}_f^{(3)}$	$\mathbf{M}_f^{(4)}$	$\mathbf{M}_f^{(5)} \approx \mathbf{M}^*$
M_{S1}	1.0344	1.0392	1.0405	1.0424	1.0427	1.0352
M_{12}	0.6740	0.8100	0.8637	0.8977	0.8937	0.9106
M_{23}	0.4656	0.6017	0.6379	0.6626	0.6833	0.6999
M_{11}	0.1400	0.0180	−0.0010	−0.0162	0.0759	0.0000
M_{22}	−0.0240	0.0270	0.0000	0.0065	0.00000	0.0000

The last column shows the ideal coupling matrix that coincides with the last extraction.

Example 23.7 ***EM Modeling of Higher-Order Filters***

This example extends the power of efficient EM modeling techniques in evaluating the exact performance of high-order filters. The filter chosen represents an advanced higher-order linear phase filter suitable for channelizing the frequency band in 6/4 GHz satellite systems. This example provides a benchmark in the application of EM techniques for microwave filter design.

The specifications for this particular example are taken from Example 23.2:

- $f_0 = 4$ GHz and BW = 39.5 MHz
- $s_{tz} = \{\pm j1.2015, \pm 0.719\}$
- RL = 22 dB

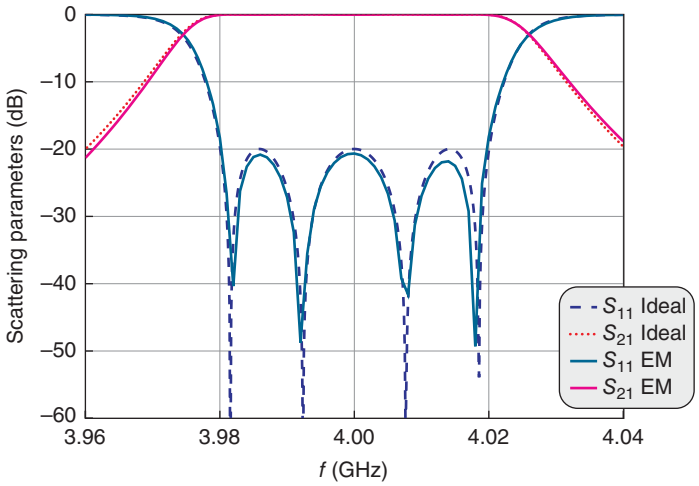


Figure 23.41 Final EM response after the optimization procedure compared with the ideal response.

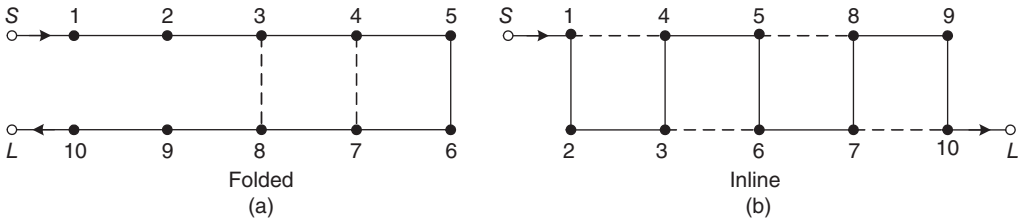


Figure 23.42 Folded configuration (a) and inline configuration (b) of the 10-2-2 filter of Example 23.7.

The symmetrical realization of the filter is a 10th-degree inline topology that has been obtained from the folded configuration as it was explained in Chapter 9. Both routing diagrams are shown in Figure 23.42.

This inline topology is best suited for filters in dual-mode configuration. For this case we choose a circular waveguide with radius $R = 35$ mm. The input/output waveguide is the standard WR-229 ($a = 58.166$ mm and $b = 29.083$ mm).

The notation chosen for the design of this filter is as follows: t is the thickness of all irises, W is the aperture length, and H is its height, as shown in Figure 23.22. If the iris is vertical, W will be the longest dimension and H the shortest. The iris thickness is usually fixed to a reasonable value and the ratio W/H is optimized until the required coupling is achieved.

Following the procedure of Example 23.4, the physical dimensions are readily obtained. The amplitude response before a final optimization is shown in Figure 23.43. This initial response is remarkably good considering the complexity and high order of the filter, thus confirming the efficacy of the initial distributed model of the filter. The next step is the final optimization of the filter response. The optimized amplitude and group delay responses are depicted in Figures 23.44 and 23.45, respectively. The corresponding dimensions are described in Table 23.12.

The lengths of the cavities as well as the penetration of the screws are shown in Table 23.13.

Finally, the position of the screws inside each cavity is reported in Table 23.14.

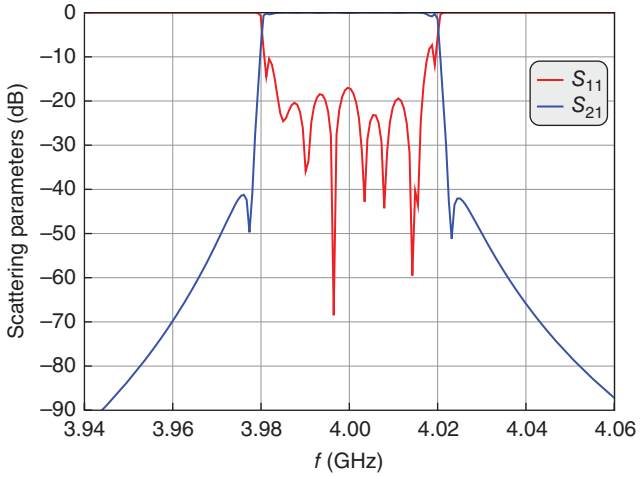


Figure 23.43 Full-wave EM response of the 10-2-2 filter before the final optimization.

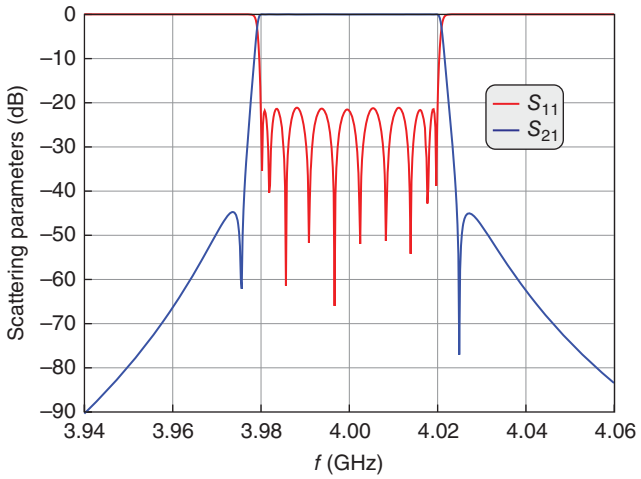


Figure 23.44 Full-wave EM response of the 10-2-2 filter.

Example 23.8 *High-Order Lowpass Filter*

This example shows the EM techniques applied in conjunction with the distributed model for the design of a high-order lowpass filter in waveguide technology. The example chosen is a harmonic filter for Ku-band satellite systems. The primary system constraint is the requirement of high isolation at second and third harmonics. Here is a case where a distributed model cannot predict such far-out-of-band response, and thus the use of EM techniques is deemed essential.

The specifications are summarized as follows:

- Passband: 11.7–12.2 GHz
- Return loss: RL > 26 dB
- Isolation:
  - $I > 55$ dB at 14 GHz
  - $I > 70$ dB at 14.5 GHz

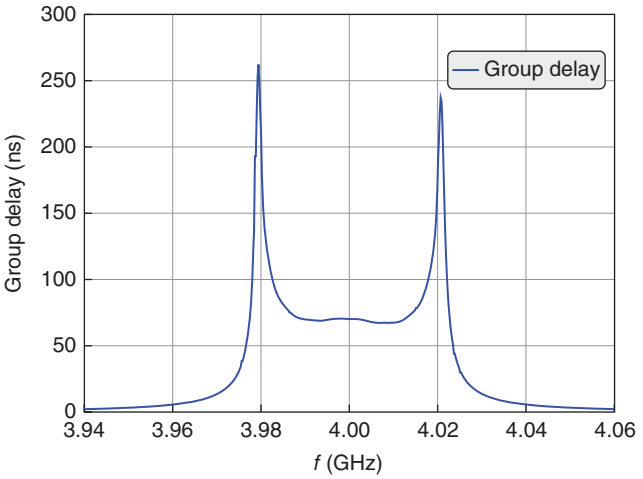


Figure 23.45 Full-wave EM simulation of the group delay for the 10-2-2 filter.

Table 23.12 Iris dimensions for the 10-2-2 filter.

Iris	*t* (mm)	*H* (mm)	*W* (mm)
Input/output iris	3.000	6.000	28.696
Iris 1: horizontal arm	3.000	3.000	17.988
Iris 1: vertical arm	3.000	3.000	22.167
Iris 2: horizontal arm	3.000	3.000	18.164
Iris 2: vertical arm	3.000	3.000	14.797
Iris 3: horizontal arm	3.000	3.000	14.797
Iris 3: vertical arm	3.000	3.000	18.164
Iris 4: horizontal arm	3.000	3.000	22.167
Iris 4: vertical arm	3.000	3.000	17.988

Iris i means iris between cavity i and $i+1$.

Table 23.13 Cavity lengths and screw penetrations for the 10-2-2 filter.

Cavity number	Cavity length penetration (mm)	Vertical screw penetration (mm)	Horizontal screw penetration (mm)	Oblique screw penetration (mm)
1 and 5	45.926	2.487	8.286	5.822
2 and 4	47.538	3.589	2.676	3.866
3	47.538	4.179	4.179	4.646

All the screws have been simulated as having a square cross section of 5×5 mm^2.

Table 23.14 Position of the screws for each cavity of the 10-2-2 filter.

Cavity number	Vertical screwposition (°)	Horizontal screw position (°)	Oblique screw position (°)
1	90	0	45
2	90	180	135
3	90	180	135
4	90	180	135
5	90	0	45

  - $I > 70$ dB in the range 23.4–24.4 GHz
  - $I > 60$ dB in the range 35.1–36.6 GHz
- Insertion loss: IL <0.25 dB
- Waveguide: WR-75 ($a = 19.05$ mm and $b = 9.525$ mm).

The starting point is the election of the filter order. The procedure due to Levy [7] and explained in Chapter 12 gives a first guess to start the design. We have to choose a frequency where the maximum given attenuation occurs. Knowing that starting with the ideal distributed model would require some degree of optimization, it is best to assume a conservative value, say, 120 dB of attenuation at 25 GHz. That will produce a valley between the lowpass response and the first replica that will meet the specifications of the attenuation mask (see Figure 23.46). This choice gives the length of the commensurate design, since the maximum attenuation occurs at the electrical length of $\phi = \pi/2$. Therefore the sections have $\phi = \pi/2$ at 25 GHz. The wavelength in the WR-75 at such frequency is $\lambda_g(f = 25\ \text{GHz}) = 12.63$ mm. The electrical length gives the physical length since $\phi = \beta l = (2\pi/\lambda)l$; l is evaluated as $l = 3.158$ mm for each section.

Therefore, at the band edge we have

$$\lambda_g|_{f=12.2\ \text{GHz}} = 32.155\ \text{mm}$$
$$\phi|_{f=12.2\ \text{GHz}} = 35.36°$$

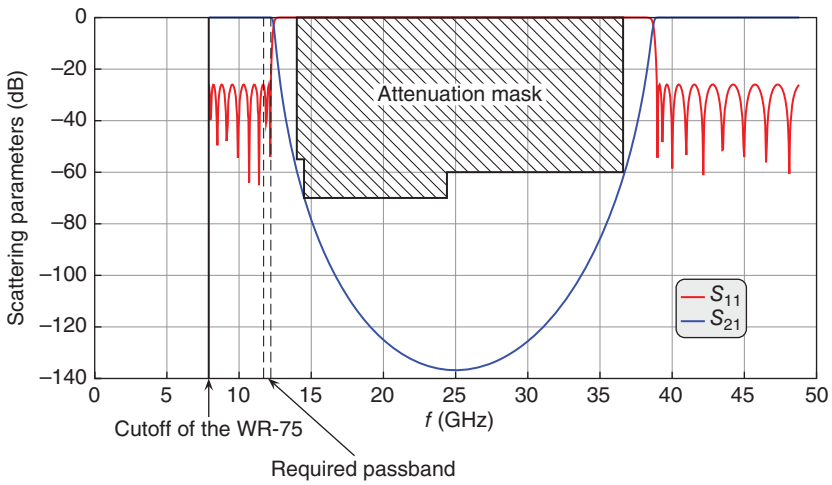


Figure 23.46 Ideal response of the distributed model for the lowpass filter of Example 23.8.

The so-called commensurate-line length is thus $\theta_c = 35.36°$. The maximum attenuation has the following expression:

$$A_{\max} = 10 \log_{10}[1 + \varepsilon^2 \cosh^2(N \cosh^{-1}\omega_c)]$$

where $\omega_c = 1/\sin\theta_c$ and $\varepsilon = 1/\sqrt{10^{\mathrm{RL}/10} - 1} = 0.0502$. In order to reach 120 dB, the filter order has to be at least $N = 17$. We can proceed with the extraction of the characteristic impedance of each section. The normalized impedances are given as follows (see Section 12.3):

$$Z_{\text{in}} = Z_{\text{out}} = 1$$
$$Z_1 = Z_{17} = 1.6020$$
$$Z_2 = Z_{16} = 0.4733$$
$$Z_3 = Z_{15} = 2.8029$$
$$Z_4 = Z_{14} = 0.3674$$
$$Z_5 = Z_{13} = 3.1018$$
$$Z_6 = Z_{12} = 0.3512$$
$$Z_7 = Z_{11} = 3.1724$$
$$Z_8 = Z_{10} = 0.3472$$
$$Z_9 = 3.1883$$

The response of the ideal distributed circuit (cascade connection of ideal TE_{10} lines) is shown in Figure 23.46. The next step is to transform the distributed model into an EM model. The reasons to carry out this procedure are summarized as follows:

- The ideal distributed circuit does not account for the EM response of the steps between adjacent waveguide sections.
- Limitation of the distributed model due to its single-mode nature. Modes beyond TE_{10} are not taken into account in the model. In this design, the TE_{20} mode appears at 15.73 GHz, which is quite close to the passband, and TE_{30} appears at 23.61 GHz.
- The impedance of the waveguide section can be scaled since it depends on the ratio b/a and a never changes. If we decrease b, the EM design becomes better because undesired responses are pushed up in frequency. However, we have to fix the minimum b to handle enough power and to avoid discharge phenomena. We have enforced the minimum to be $b = 0.7$ mm. With this minimum b, the maximum value of b required to achieve the impedance level is less than 9.525 mm; therefore the need for a transformer becomes apparent.

The final performance of the filter is shown in Figure 23.47. It shows an isolation greater than 120 dB over the second harmonic (23.4–24.4 GHz band) and greater than 80 dB over the third harmonic (35.1–36.6 GHz band), while the passband extends from 11.7 to 12.2 GHz. This is deemed as excellent performance.

The specifications are met in the passband, and in the attenuation band the response is shown beyond the first replica of the filter appearing around 40 GHz. The physical dimensions of the filter are shown in Table 23.15. The side view of the final design is shown in Figure 23.48. A 3D view with the relative strength of the electric field is shown in Figure 23.49. The concentration of the field in the minimum gap confirms the need of a careful design in terms of power handling capabilities. For this reason, this example is chosen for high-power analysis in Section 23.4.

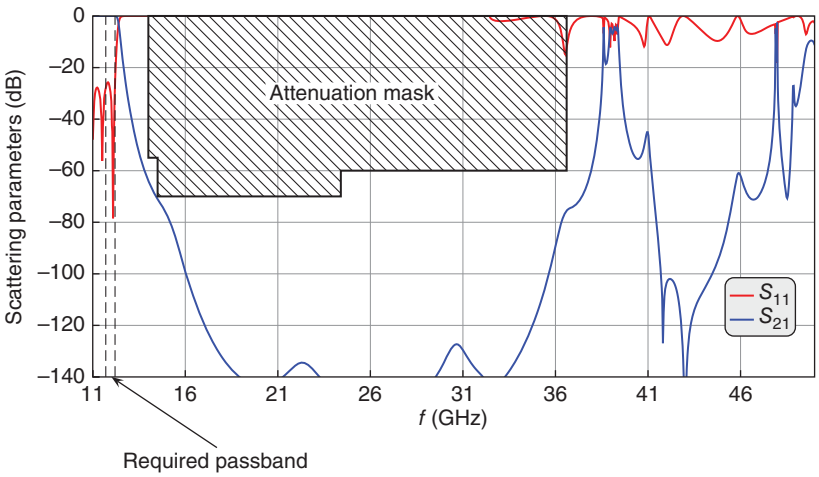


Figure 23.47 Full-wave EM response of the lowpass filter.

Table 23.15 Dimensions of the lowpass filter after EM optimization.

Section	Width *a* (mm)	Height *b* (mm)	Length *l* (mm)
1, 25 (transformer)	19.05	9.525	3.000
2, 24 (transformer)	19.05	6.389	8.516
3, 23 (transformer)	19.05	3.006	8.072
4, 22 (transformer)	19.05	2.016	5.119
5, 21 (filter)	19.05	3.230	4.061
6, 20 (filter)	19.05	0.954	2.526
7, 19 (filter)	19.05	5.652	3.317
8, 18 (filter)	19.05	0.741	2.280
9, 17 (filter)	19.05	6.254	3.412
10, 16 (filter)	19.05	0.708	2.267
11, 15 (filter)	19.05	6.397	3.052
12, 14 (filter)	19.05	0.700	2.562
13 (filter)	19.05	6.429	3.002

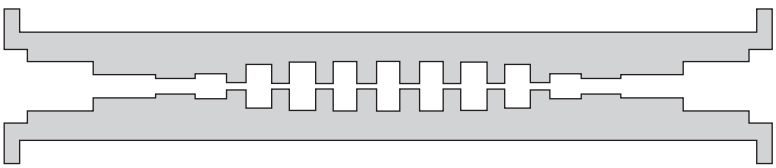


Figure 23.48 Side view for the 17th-degree lowpass filter with the transformers at the ends. The filter has been drawn to scale.

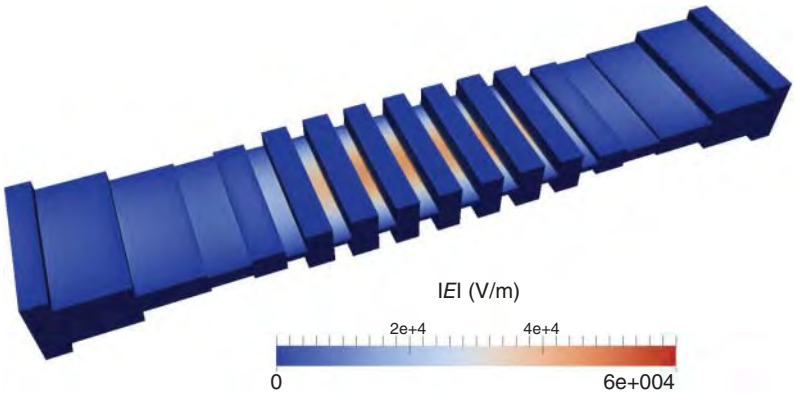


Figure 23.49 3D view of the lowpass filter with a sketch of the electric field.

23.3 Multiplexers

This section describes simple examples to help understand the design methodology of the two most widely used multiplexers, the circulator-coupled configuration and the manifold-coupled design. Examples highlight the tradeoffs between channel performance and design complexity in enhancing the understanding of the optimization process for the manifold-coupled multiplexers.

Example 23.9 ***Circulator-Coupled Multiplexer***

This channelizing scheme makes use of the nonreciprocal ferrite circulator coupled to a filter to extract individual channels as described in Sections 1.8.2 and 18.3.2.

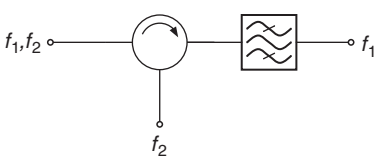
Channel filter serves two functions; it extracts the channel and reflects off all the remaining channels for extraction by subsequent channel filters in the chain. It is therefore important to understand how the reflection response affects the characteristics of the channels in the multiplexed configuration.

Reflection Response of Channel Filters: En Passant Distortion

A typical frequency plan at 4 GHz has channels 40 MHz apart. Let us examine the reflection response of a channel filter centered at $f_1 = 3940$ MHz over the BW of the adjacent channel f_2 centered at 3980 MHz as described in Figure 23.50. The channel filter chosen is 10-2-2 from Example 23.2, and the usable BW of each channel is 36 MHz. The computed reflection response is described in Figure 23.51.

The reflected group delay shows large variation near the lower band edge of channel f_2, decreasing sharply, and is nearly flat from 3965 MHz onward. This group delay variation, referred to as *en passant distortion* in Section 18.3.2, gets added on to channel 2 performance, adversely affecting its usable BW (shaded area in Figure 23.51). Amplitude variation follows a similar response. Needless to say, such a degradation of usable BW is not acceptable. Note that the rejection provided by the channel filter centered at f_1 at the band edge (3962 MHz) of channel 2 is about 14 dB, this value rising sharply, and from 3965 onward it is greater than

Figure 23.50 Circulator-coupled module. Channel 1 is extracted and the filter reflects off the remaining channel.



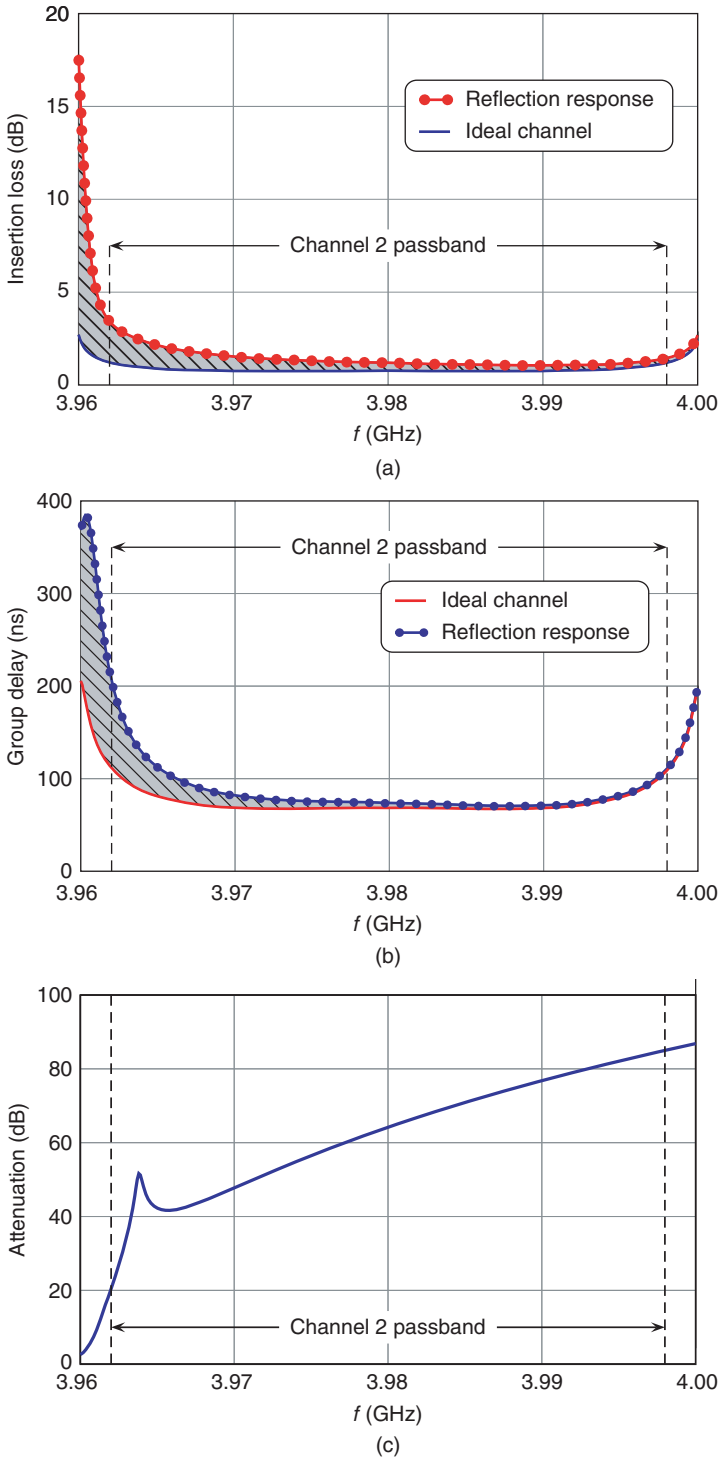


Figure 23.51 En passant distortion. (a) Contribution to in-band loss variation over channel 2. (b) Contribution to group delay variation over channel 2. (c) Isolation response of channel 1 filter over channel 2 frequency band.

40 dB. This shows that the contribution to variations in group delay and insertion loss goes down with increasing value of rejection provided by the channel 1 filter. Two conclusions are drawn from this analysis:

1) This type of multiplexer is not suitable to channelize contiguous channels.
2) This multiplexer design is suitable to channelize noncontiguous channels. It may also be used for quasi-contiguous channels, provided the rejection by the preceding channel over the pass BWs of subsequent channels is greater than 60 dB.

Circulator-Coupled Multiplexer with minimal En Passant Distortion

A six-channel multiplexer using the 10-2-2 filter is described in Figure 23.52. Channels chosen on either branch of the hybrid (designed to cover 500 MHz BW) are noncontiguous and use the same filter design with a BW = 39.5 MHz and RL = 22 dB. The channels are 40 MHz apart and centered at 3940, 3980, 4020, 4060, 4100, and 4140 MHz. Note that the use of hybrid allows extraction of noncontiguous channels from each branch of the hybrid. The penalty incurred is the additional 3 dB loss due to the hybrid. From the system perspective, this additional loss is quite acceptable as explained in Chapter 1. Such multiplexers are widely employed for the channelization of broadband signals in satellite systems.

The circulators are assumed to have losses of 0.2 dB per trip and isolation of greater than 25 dB. The transmission parameters of the multiplexer are shown in Figure 23.53; it includes the combined insertion loss of the circulators and the hybrid estimated to be 4 dB. Impact of en passant distortion is minimal on channel response.

The common port RL is shown in Figure 23.54. The filters are identical except for the center frequencies. However, different frequencies cross through different paths with a different number of circulators. This produces a high variation in the CPRL and slightly different insertion losses for each channel.

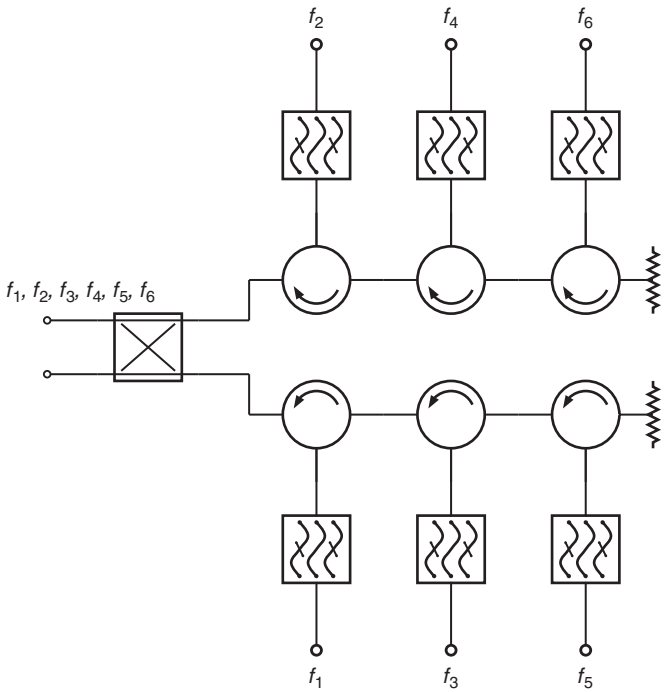


Figure 23.52 Circulator-coupled configuration of the input multiplexer (channelizer) with six channels.

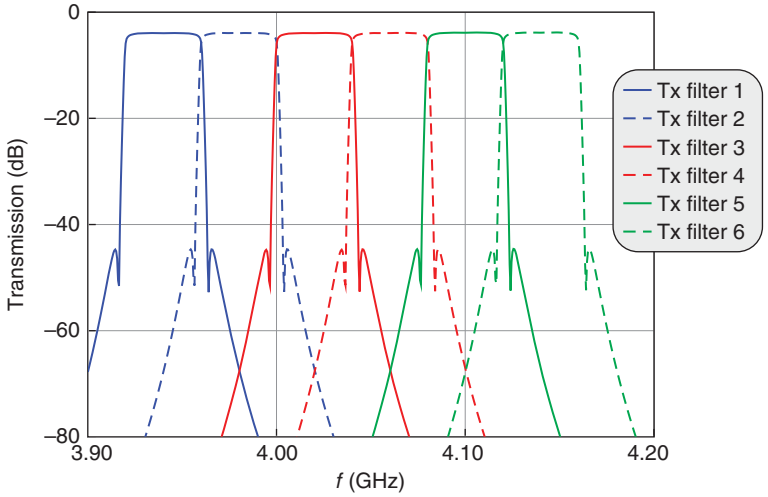


Figure 23.53 Transmission parameters of the input multiplexer from common port to the output of each filter. Odd channels are plotted with solid lines and even channels with dashed lines.

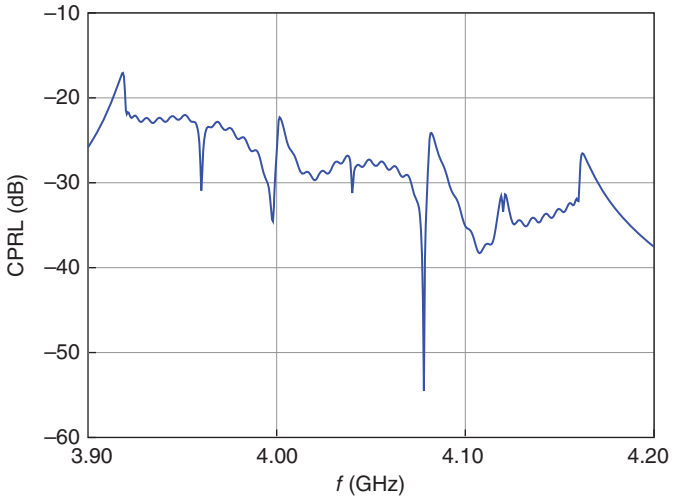


Figure 23.54 Common port return loss for the input multiplexer.

Example 23.10 ***Manifold-Coupled Noncontiguous Multiplexer***

In this example, we consider a five-channel noncontiguous manifold-coupled multiplexer design in the 4 GHz frequency band. The frequency plan is described in Figure 23.55. This plan, showing only alternate channels specific to a noncontiguous multiplexer configuration, is a subset of the overall frequency plan of a 6/4 GHz satellite system. Usable BW of each channel is 36 MHz, and filter BW must be less than 40 MHz to minimize noise in adjacent channels, as explained in Section 1.8.4.

Noncontiguous multiplexers are characterized by wide separation between channels. This implies that one can design lower-order channel filters to provide high rejection in other channel frequencies. This minimizes the interaction between channels when filters are mounted on a common manifold. High rejection between channels also allows the use of doubly terminated filter models as the starting point for the multiplexer design.

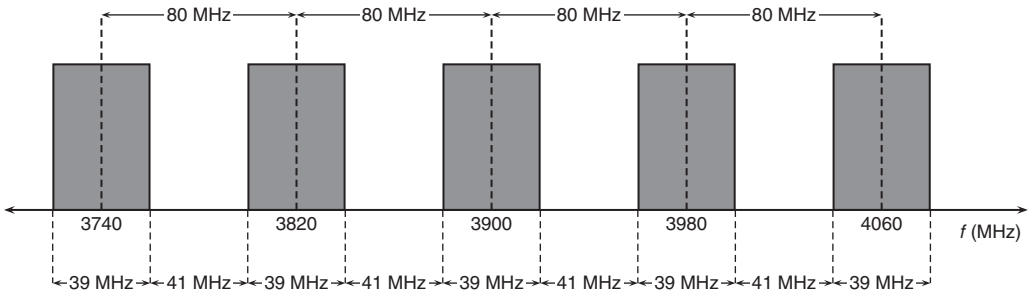


Figure 23.55 Frequency plan for the noncontiguous manifold multiplexer of Example 23.10. This frequency plan is referred to odd/even channels since the remaining even/odd channels are located in the guard bands, and they are dealt with in another multiplexer connected to this one to achieve a real frequency plan with even and odd channels properly interweaved.

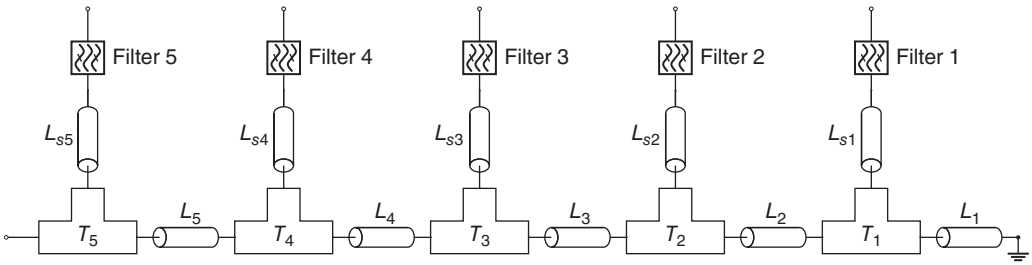


Figure 23.56 Schematic of the manifold multiplexer of Example 23.10. Numbers increase from the short circuit. T_i stands for the ith E-plane T junction. L_i stands for the spacing between the ith and the $(i+1)$-th T junctions. L_{si} stands for the spacing between the ith T junction and the ith filter.

The classification of a multiplexer as being noncontiguous is somewhat arbitrary. Based on experience, if the guard band, defined as the ratio of channel separation to channel BW, is greater than 25%, the multiplexer is classified as noncontiguous. Such multiplexers can be designed to achieve channel performance that resembles closely the ideal performance of an individual filter. For the present example, the guard band is $41/39 \approx 105\%$, well above the threshold of 25% for noncontiguous multiplexers. All the filters have the same features: BW = 39 MHz, order $N = 4$, RL = 22 dB, and transmission zeros $n_{tz} = 2$ and out-of-band rejection of 40 dB or greater. The CM is the same used for Example 23.4. The filters are dual-mode, realized in circular waveguide using the dominant TE_{111} propagation mode.

The multiplexer model is composed of (i) a waveguide manifold incorporating E-plane T junctions and, (ii) dual-mode filters connected to the T junctions with lengths of waveguide (shorter than $\lambda_g/2$) for the initial model. A schematic with the nomenclature of the lines and filters is shown in Figure 23.56.

The starting point for the MUX design is to set all lengths (L_i and L_{si}) to $\lambda_g/2$ (λ_g at band center of each channel). The next step is to perform piecewise optimization of the multiplexer performance as described in Chapter 18.

1) Draw in the simulator the junction T_1, lines L_1, L_{s1}, filter 1, and the short.
2) Tune L_1 and L_{s1} to achieve a CPRL as close as possible to the specification. It can be done manually or through the optimization of the two variables involved with a linear simulator, since T_1 and the filter are never touched at this stage.
3) Add the second stage: T_2, L_2, L_{s2}, and filter 2.

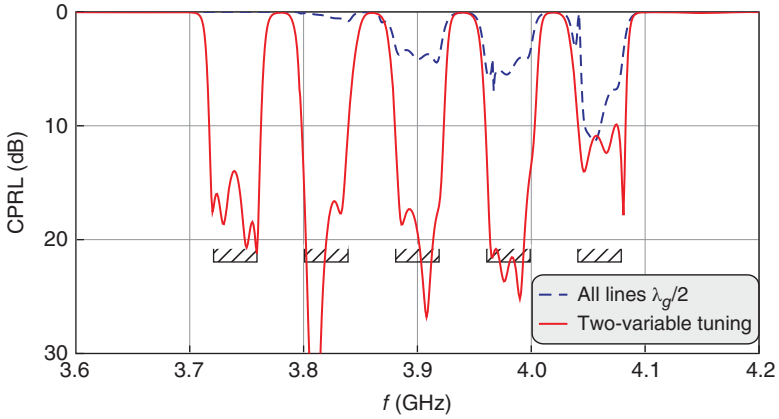


Figure 23.57 Initial CPRL of the multiplexer with (1) all lines set at $\lambda_g/2$ and (2) piecewise optimization using two variables.

4) Tune only L_2 and L_{s2} to achieve a CPRL as close as possible to the specification. Again, this can be done manually with a linear simulator while keeping the values of L_{S1} and L_1 the same.
5) Perform the same procedure adding stage i and tuning only L_i and L_{si}.

Figure 23.57 describes the comparison between the starting point and the result after the two-variable piecewise optimization [8]. The next step is the refinement process using the piecewise strategy described in Chapter 18, as follows:

1) Wideband optimization: Optimize all L_i simultaneously.
2) Narrowband optimization: Optimize each L_{si} and the first parameters of the ith filter. In this case only M_{s1}, M_{11}, M_{12}, M_{22}, and M_{14} have been optimized.

The final result for the CPRL is shown in Figure 23.58 and the transmission response of the channels is shown in Figure 23.59.

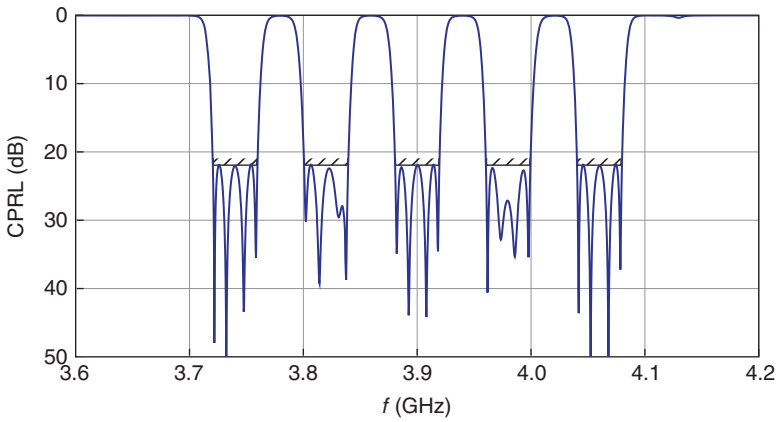


Figure 23.58 CPRL response of the multiplexer after the optimization procedure.

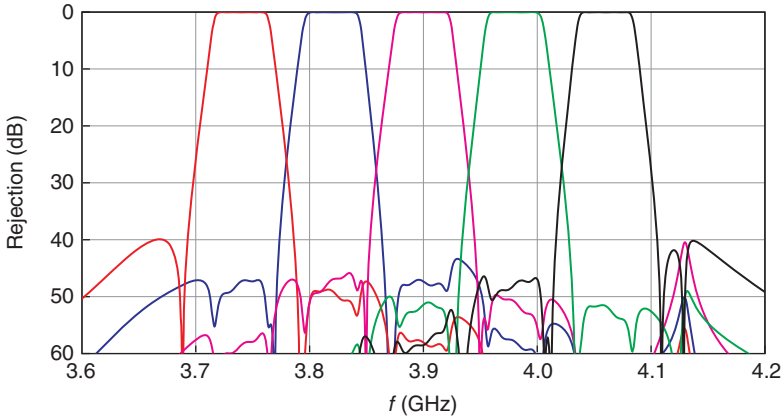


Figure 23.59 Final response for the transmission performance of the multiplexer.

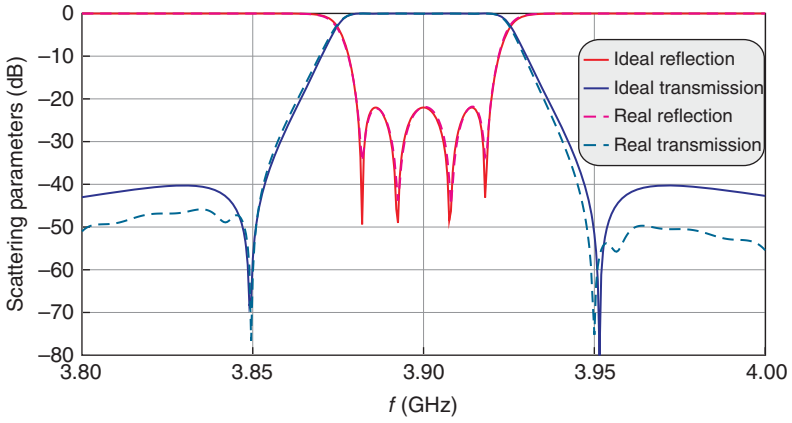


Figure 23.60 Final response for the transmission performance of the middle channel compared with the ideal response.

As can be seen, enabled by the lack of interaction among various channels, the optimized response of the overall noncontiguous multiplexer appears close to ideal performance. However, the ultimate test is a comparison of transmission response of individual channels in the multiplexer as compared with the performance of an equivalent individual filter on its own. This comparison is shown in Figure 23.60. This clearly shows that noncontiguous multiplexers provide close to ideal performance for combining (or separating) channels in a manifold with a minimum of insertion loss. Such multiplexers are best suited for the transmit section of communication systems where minimum insertion loss is a key requirement.

With regard to the input impedance/admittance, a single isolated channel of this multiplexer has the performance shown in Figure 23.61. As we can see the real parts are even as expected and move around 1 Ω within the passband. The imaginary parts are odd and they move around zero. In fact, within the passband, when the imaginary part of the immittance equals zero, the real part is 1, and there is a reflection zero. The whole multiplexer has to show a similar performance for each channel from the common port, as reported in Figure 23.62. Namely, in the passband of each channel, the real part of the immittance should be around 1 and the imaginary part around zero.

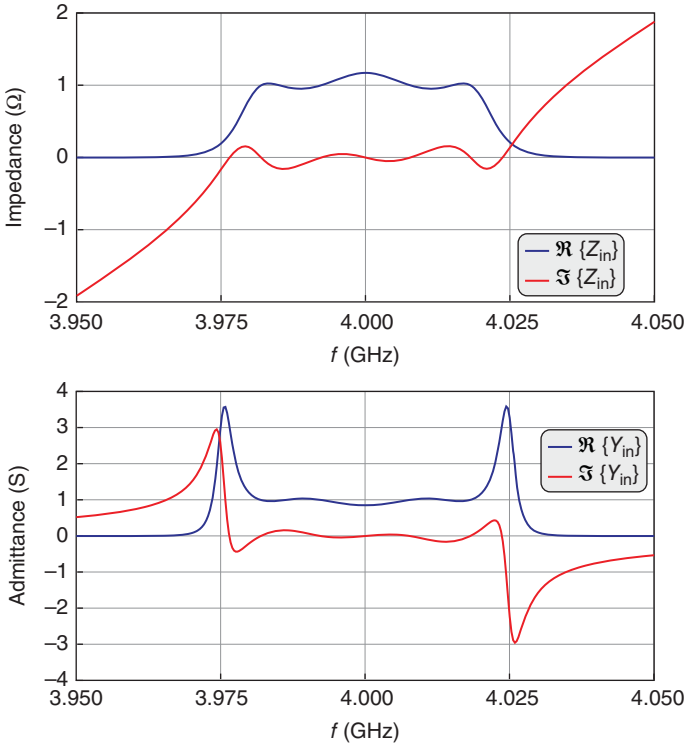


Figure 23.61 Impedance/admittance response of an isolated channel.

Example 23.11 ***Manifold-Coupled Contiguous Band Multiplexer***

In this example, we consider a five-channel contiguous multiplexer in the 12 GHz frequency band. The frequency plan is shown in Figure 23.63. The guard band for this example is 4/76 or nearly 5%, much less than the threshold of 25% for noncontiguous multiplexers. When guard bands are small, as is the case with contiguous multiplexers, it is preferable to use singly terminated prototype filters for the initial design of the multiplexer as explained in Chapter 18.

The lowpass prototype filter function (4-2) chosen for all the channel filters is the same as in Example 23.4. However, the CM is not the same since filter models used for the initial design are based on singly terminated filters. The bandpass configuration of all filters has the same features: BW = 76 MHz, order $N = 4$, RL = 22 dB, and two transmission zeros to achieve the minimum out-of-band rejection of 40 dB. Filters are dual-mode realized in circular waveguide using the dominant TE_{111} propagation mode. Following the procedure of Example 23.4 and using singly terminated filters, the $N + 2$ CM is computed as

$$\mathbf{M} = \begin{pmatrix} 0 & 0.9167 & 0 & 0 & 0 & 0 \\ 0.9167 & 0 & 0.7698 & 0 & -0.1772 & 0 \\ 0 & 0.7698 & 0 & 0.9333 & 0 & 0 \\ 0 & & 0.9333 & 0 & 1.4789 & 0 \\ 0 & -0.1772 & 0 & 1.4789 & 0 & 1.5191 \\ 0 & 0 & 0 & 0 & 1.5191 & 0 \end{pmatrix} \tag{23.2}$$

The multiplexer model is composed of E-plane T junctions connected with waveguides forming the manifold. The schematic is exactly the same one shown in Figure 23.56.

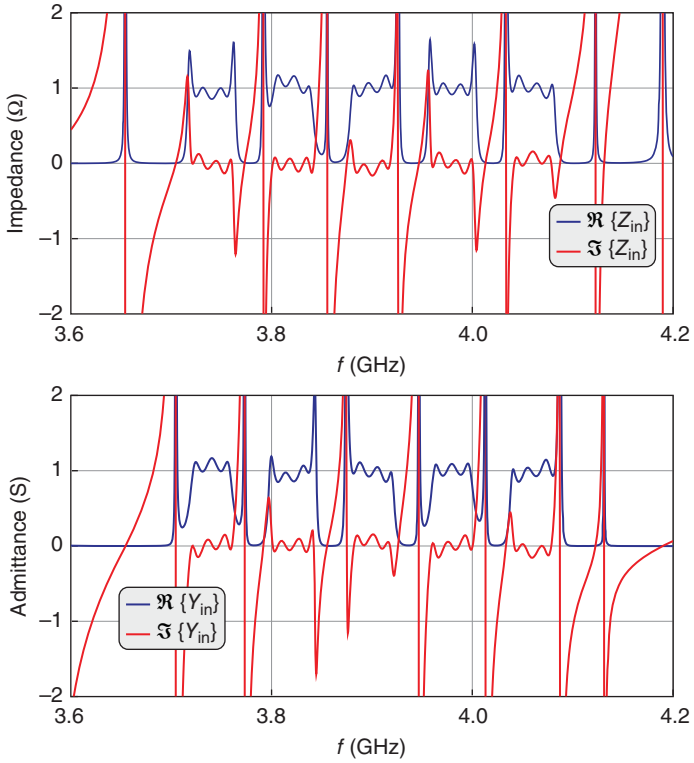


Figure 23.62 Impedance/admittance profiles of the whole multiplexer.

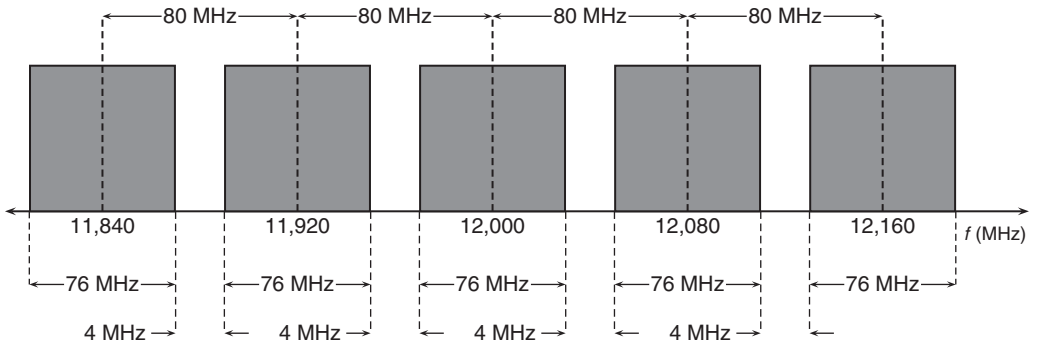


Figure 23.63 Frequency plan for the manifold multiplexer of Example 23.11.

The strategy for setting the initial point of optimization is the same as used for the noncontiguous multiplexer, that is, separation of channel filters by $\lambda_g/2$ along the manifold.

After applying the two-variable piecewise optimization, there is significant improvement in the common port RL response as shown in Figure 23.64. The next step is to consider additional parameters for optimization. The parameters chosen are

- The six couplings: M_{S1}, M_{12}, M_{23}, M_{34}, M_{4L}, and M_{14}
- The four frequency shifts of the resonators: M_{11}, M_{22}, M_{33}, and M_{44}.

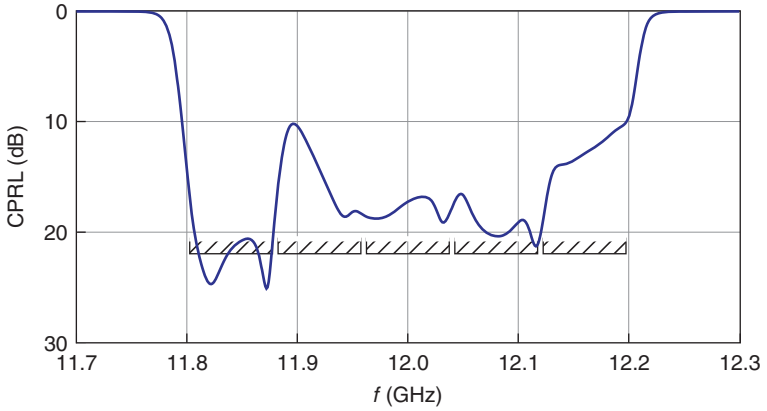


Figure 23.64 Initial CPRL of the multiplexer of Example 23.11.

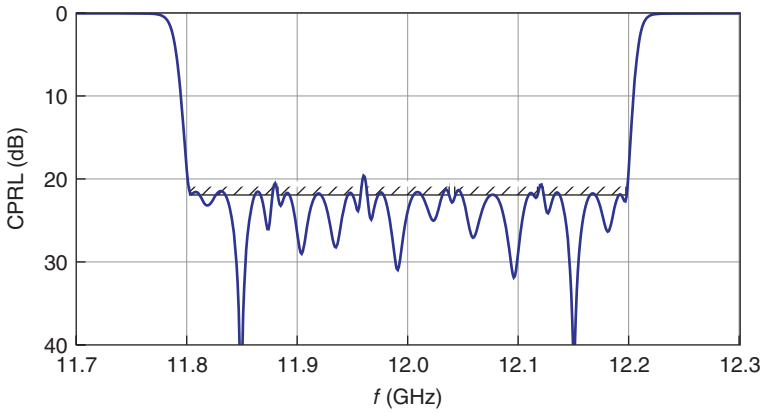


Figure 23.65 CPRL of the multiplexer of Example 23.11 after seven cycles of optimization.

Refinement process is carried out by optimizing the ten parameters, one channel at a time. After seven cycles of optimization, the best solution found has the CPRL shown in Figure 23.65. The amplitude response of the overall multiplexer is depicted in Figure 23.66.

It is interesting to compare the CPRL for the noncontiguous and contiguous band multiplexers as described in Figures 23.58 and 23.65, respectively. Unlike the CPRL of the noncontiguous multiplexer that fluctuates between 22 dB over desired bands and nearly zero over the guard bands, the CPRL for the contiguous multiplexer remains above 20 dB over the entire BW including the guard bands due to the extremely narrow guard band used for this contiguous multiplexer. Such a narrow guard band incurs large channel interactions, as is evident in the transmission response of the overall multiplexer depicted in Figure 23.66. The comparison between transmission performance of a typical channel in the multiplexer and the performance of an equivalent individual filter on its own is described in Figure 23.67. It highlights the performance tradeoffs of a channel in contiguous multiplexer: sharper attenuation response at the expense of relatively large group delay and insertion loss variations at the band edges of the channels. It highlights the tradeoff between the guard band and the achievable performance of a channel over the specified usable BW. Experience shows that a guard band of 10% provides a good tradeoff for most communication systems. Additional qualitative information on this

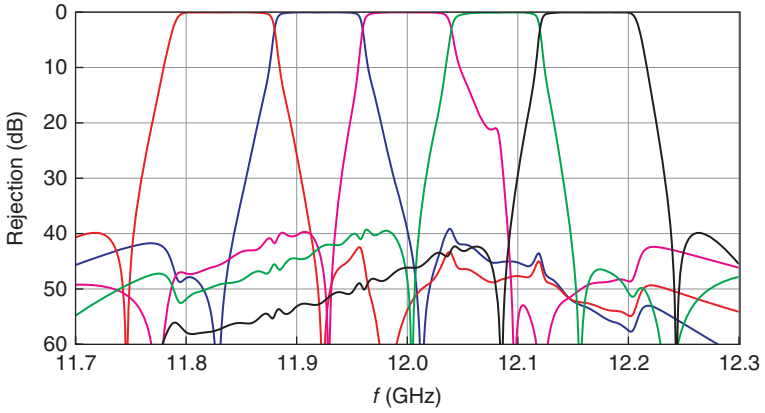


Figure 23.66 Amplitude response of the overall multiplexer after 7 cycles of optimization.

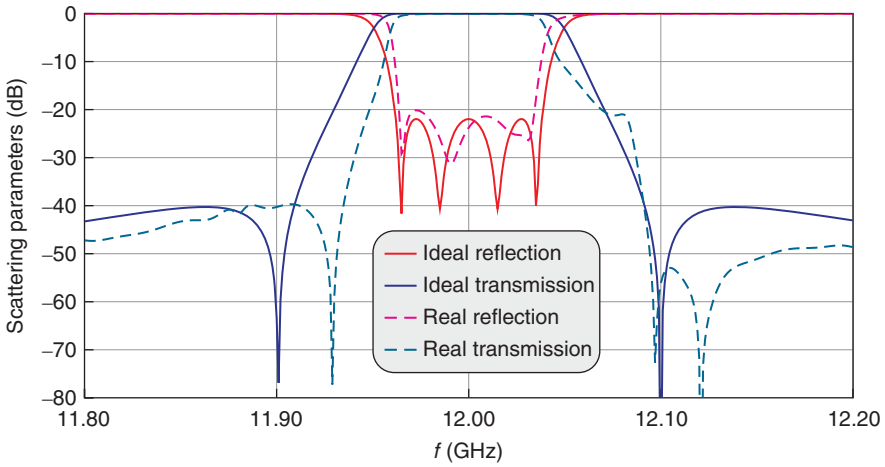


Figure 23.67 Comparison between transmission performance of a typical channel in the multiplexer and the ideal filter response.

topic is contained in Section 1.10, which gives an overview of tradeoffs involved for the frequency plan and achievable channel performance in communication systems. Complexity of contiguous multiplexing makes the design difficult and very sensitive to manufacturing tolerances.

Figure 23.67 shows a typical response after optimization using a singly terminated model. Due to the strong coupling between the last resonator and the load ($M_{4L} = 1.5191$ at the beginning of the process), the optimization gets trapped in a local minimum with high M_{4L} perhaps with a non-realizable value of the coupling. In order to avoid this high value, the optimization strategy can start with a filter model with lower RL to avoid a large M_{4L}. When the optimization starts to improve the overall CPRL, the couplings are changed accordingly, but the coupling values have to be controlled to avoid non-practical values. Using this strategy, the final performance obtained is shown in Figure 23.68. The CPRL is shown in Figure 23.69.

The input admittance/impedance for the singly terminated filter is shown in Figure 23.70. With regard to the admittance (left), the response of its real part follows the response of the

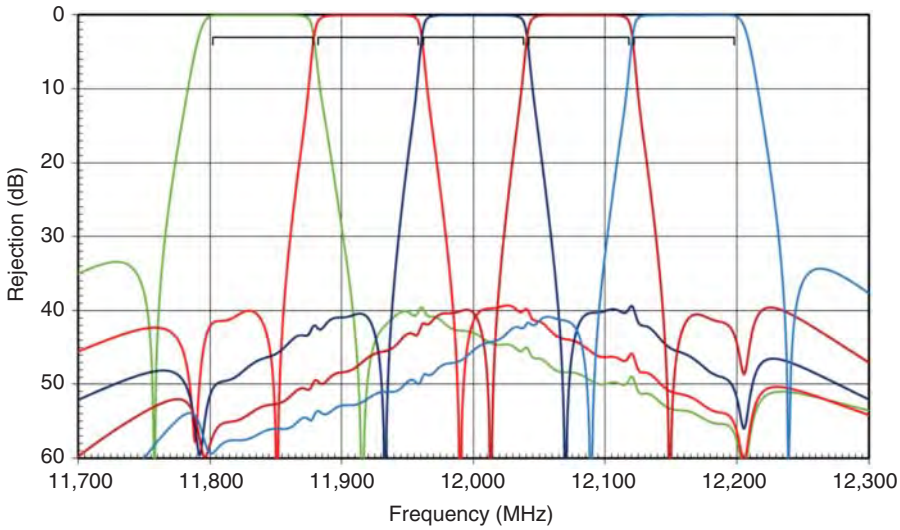


Figure 23.68 Final transmission amplitude response of the overall multiplexer.

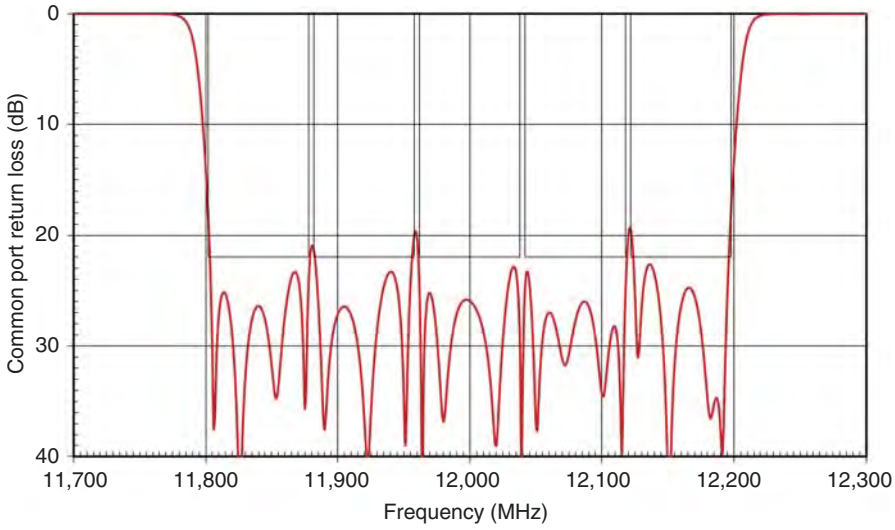


Figure 23.69 Final CPRL performance of the overall multiplexer.

filter (in linear units) since the model is taken with admittance inverters and parallel resonators. Its real part is close to 1 (and even), while the imaginary part is around zero (and odd).

Finally, the overall multiplexer has to show a similar performance for each channel from the common port, as reported in Figure 23.71. In the passband of each channel, the real part of the immittance should be around 1, and the imaginary part around zero. Unlike the noncontiguous case, the guard bands are so narrow that the whole multiplexer behaves as a single filter in terms of impedance. The whole BW of the multiplexer has an input immittance that resembles the immittance of the isolated filter shown in Figure 23.70.

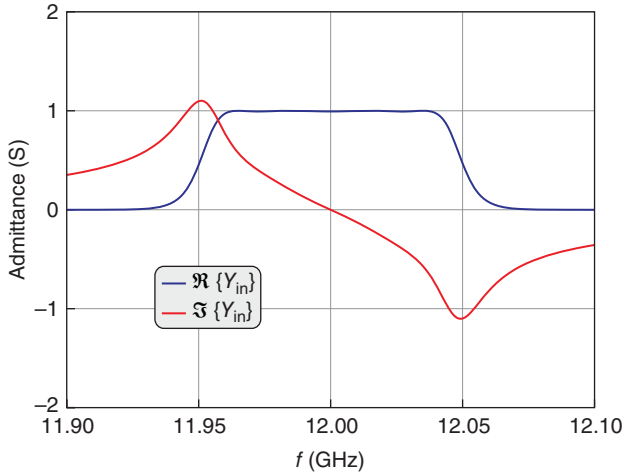


Figure 23.70 Admittance response of an isolated channel.

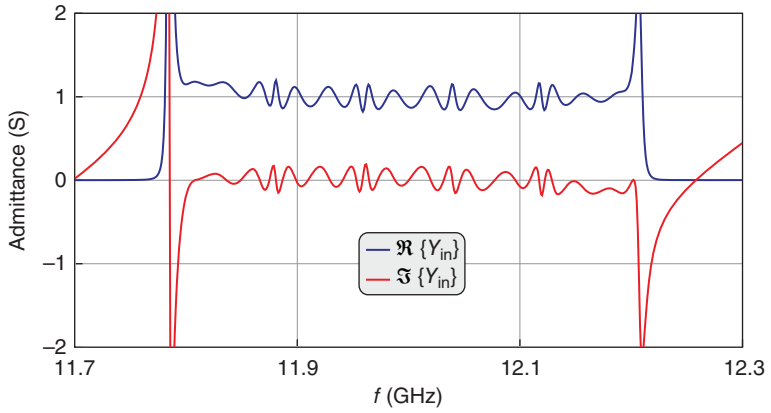


Figure 23.71 Admittance profile of the whole multiplexer.

23.4 High-Power Considerations

Following the theoretical aspects of high-power breakdown phenomenon in passive components as described in Chapter 20, examples in this section illustrate the field distributions and tradeoffs of voltage and power breakdown versus geometry in a variety of filter structures.

Example 23.12 ***High-Power Lowpass Filter***

As an example of high-power breakdown analysis, the lowpass filter designed in Example 23.8 is analyzed. The first step in this analysis is to determine the critical regions in which the discharge may occur with the lowest input power level. This depends on the geometry of the structure as well as the signal frequency. It is the combination of the field strength and the gap size that indicates the position of the critical regions as described in Figure 20.14.

The operating frequency band for this filter is 11.7–12.2 GHz. As an example of the field strength in this frequency band, the amplitude of the electric field at the center (x coordinate) along the direction of propagation has been computed numerically at 12.2 GHz, as shown in Figure 23.72. The four peaks of the electric field correspond to the four small gaps of size 0.954, 0.741, 0.708, and 0.700 mm as described in Table 23.15. This figure highlights the fact that the field is not quite symmetric (despite the symmetry of the physical structure) and the criticality of the small gaps where the field strength has the highest values. Table 23.16 shows the computed voltage at the middle point (x coordinate) between the top and bottom surfaces of each of the small gap sections over the entire frequency band for an input power level of 1 W.

It is worth mentioning that the trend of the field/voltage amplitude with the frequency is characteristic of lowpass filters: the closer to the cutoff frequency of the filter, the higher the field.

Considering silver as the surface material of the component and knowing the frequency-gap product, one can use the multipactor susceptibility charts (see, e.g., Figure 20.9) available in the literature to find the breakdown voltage level. In this case, the European Cooperation for Space Standardization (ECSS) Multipactor Standard[3] has been used to compute the breakdown voltages at different sections and frequencies, as shown in Table 23.17. Rows 1 and 4, as well as

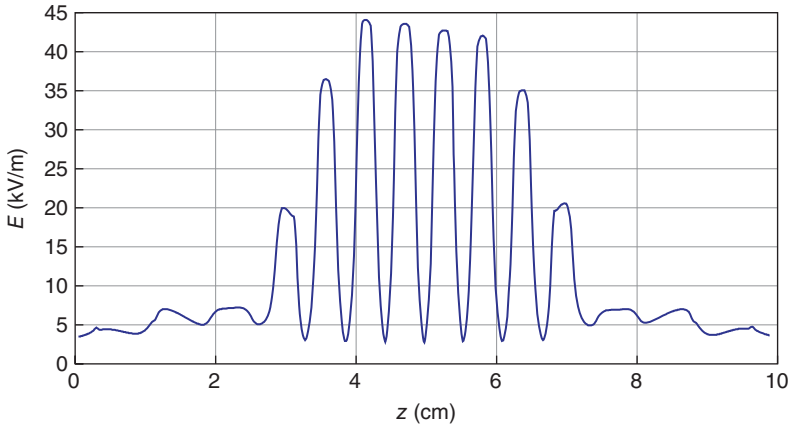


Figure 23.72 Electric field amplitude at the center (x coordinate) of the filter along the propagation axis at 12.2 GHz.

Table 23.16 Calculated voltage at 1 W input power in the four center sections for five different frequencies.

	Voltage (V)					
Section	**11.7 GHz**	**11.8 GHz**	**11.9 GHz**	**12.0 GHz**	**12.1 GHz**	**12.2 GHz**
1	22.75	23.86	24.99	26.02	27.41	31.23
2	22.31	22.97	23.65	24.60	26.51	30.54
3	21.64	22.15	22.95	24.3	26.51	29.93
4	22.19	22.98	24.06	25.52	27.42	29.78

The section numbering increases with z.

3 ECSS-E-20-01A, May 5, 2003. Space Engineering: Multipaction Design and Test.

Table 23.17 Breakdown voltage for the four inner sections using the ECSS Multipactor Standard, considering silver as surface material and the corresponding frequency-gap product.

	Breakdown voltage (V)					
Section	**11.7 GHz**	**11.8 GHz**	**11.9 GHz**	**12.0 GHz**	**12.1 GHz**	**12.2 GHz**
1	518.5	523.0	527.5	532.0	536.0	541.0
2	513.0	517.0	521.5	526.0	530.0	535.0
3	513.0	517.0	521.5	526.0	530.0	535.0
4	518.5	523.0	527.5	532.0	536.0	541.0

2 and 3, result in the same breakdown voltage because the filter is symmetric with respect to its center in the propagation direction (therefore the gap values of Sections 1 and 4 are identical, as well as the gap values of Sections 2 and 3).

In this table, a slight increase of the breakdown voltage with the frequency is noted due to the increase of the frequency-gap product. Since both the voltage at 1 W input power in the critical gaps and the breakdown voltage levels are known, the breakdown power level can be easily found by using

$$P_B = \left(\frac{V_B}{V_{@1\ \mathrm{W}}} \right)^2 \tag{23.3}$$

Table 23.18 shows the breakdown power levels in each section and for the whole frequency band. One could expect an increase of the breakdown power levels due to the increase of the breakdown voltages at higher frequencies (see Table 23.17). However, the results clearly show that the increase of the voltage amplitude in the critical gap at larger frequencies has more impact on the breakdown power level than the slight increase in breakdown voltages at higher frequencies. Therefore, in this particular structure, it is seen that the breakdown analysis has to be focused on the frequencies close to the cutoff of the filter. This can be extrapolated to almost any lowpass filter configuration unless the frequency band is too wide.

The calculations performed so far are based on analytical formulations that assume an infinite parallel-plate approach where the fields are considered homogeneous and where only the electric field is considered. This kind of approach represents the worst case and it may be very conservative. However, nowadays, it is possible to perform a full numerical multipactor analysis based on particle tracking (see, e.g., [9]). The idea in this case is to track electrons inside the device and determine if an increase of the electron population takes place because of the

Table 23.18 Breakdown power levels for the four inner sections using the data in Tables 23.16 and 23.17 and applying Equation (23.3).

	Breakdown power (W)					
Section	**11.7 GHz**	**11.8 GHz**	**11.9 GHz**	**12.0 GHz**	**12.1 GHz**	**12.2 GHz**
1	519.0	480.5	445.5	418.0	382.0	300.0
2	529.0	506.5	486.0	457.0	400.0	307.0
3	562.0	545.0	516.0	468.5	400.0	319.5
4	546.0	518.0	481.0	434.5	382.0	330.0

impacts of such electrons on the device walls. The advantages of the numerical approach are summarized as follows:

1) The model is more realistic than the analytical approach based on the parallel-plate geometry. It considers the entire structure and full-wave EM solution including magnetic fields.
2) The model incorporates complex forms of electron resonances, and it is not limited to double-surface multipactor processes only.
3) Breakdown is computed based on real input power and no voltage-to-power conversion is necessary. Therefore, inaccuracies caused by the calculation of the voltage in just one particular position are solved.
4) It is not limited to the materials for which charts are available. Materials with known secondary emission yield (SEY) parameters may be used as well.
5) Additional effects, such as the presence of external dc electric or magnetic fields, can also be added. For example, the effect of external magnets or coils can be studied.

In this example, a significant difference between the analytical and numerical approaches is not expected, because the irises have a small gap (and, therefore, an electric field quite homogeneous, at least around the center of the iris, is expected) and the geometry is similar to a parallel-plate situation. However, a small shift is expected since the ECSS Multipactor Standard adds some margin (of around 1 dB) to the actual values. In order to show this effect, numerical simulations have been performed in two structures:

a) A parallel-plate structure of 0.7 mm gap
b) The filter analyzed in this example.

The breakdown voltages of case (a) are given in Table 23.19. The silver surface properties tabulated in the ECSS Multipactor Standard have been used.

The numerical results for case (b) are initially provided in terms of input power, but they can be easily converted to voltage using the voltage at 1 W and equation (23.3). The voltage at 1 W for this case has been already given in Table 23.16. The numerical breakdown power and breakdown voltage are given in Tables 23.20 and 23.21.

Table 23.19 Numerical breakdown voltages for a parallel plate case with 0.7 mm gap, considering silver as surface material.

Frequency (GHz)	11.7	11.8	11.9	12.0	12.1	12.2
Breakdown voltage (V)	598	604	609	603	613	622

Table 23.20 Numerical breakdown power levels for the four inner sections, considering silver as surface material and the corresponding frequency-gap product.

	Breakdown power (W)					
Section	**11.7 GHz**	**11.8 GHz**	**11.9 GHz**	**12.0 GHz**	**12.1 GHz**	**12.2 GHz**
1	760	697	660	662	565	445
2	753	721	697	659	578	451
3	791	790	747	672	578	467
4	791	753	703	634	572	498

Table 23.21 Numerical breakdown voltages for the four inner sections.

	Breakdown voltage (V)					
Section	**11.7 GHz**	**11.8 GHz**	**11.9 GHz**	**12.0 GHz**	**12.1 GHz**	**12.2 GHz**
1	627	630	642	669	652	659
2	612	617	624	632	637	649
3	609	623	627	630	637	647
4	624	631	638	643	656	665

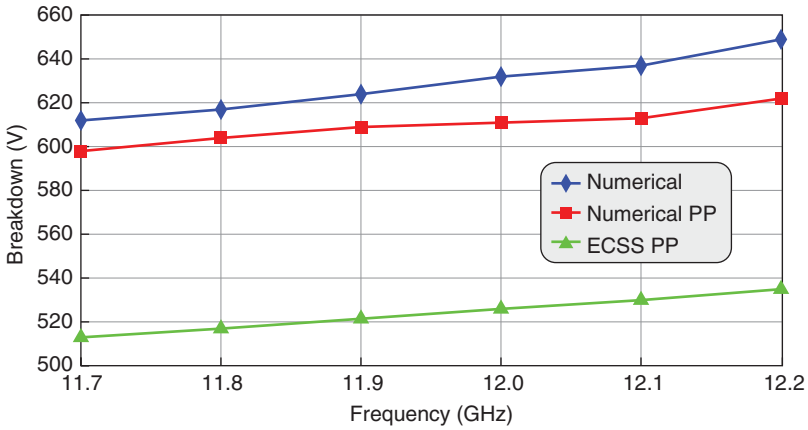


Figure 23.73 Breakdown voltages calculated in Section 3 of the filter structure with three different analysis methods.

Breakdown voltages for the inner Section 2 of the filter (yielding the lowest breakdown power) corresponding to the three analysis methods are represented in Figure 23.73. As expected, the three curves follow a similar trend, since the analyzed example presents a geometry similar to parallel plates. There is a shift of approximately 1 dB between ECSS prediction and the numerical parallel plates due to the 1 dB margin applied by the ECSS. The small difference between the numerical parallel-plate results and the numerical results of the actual filter is due to the fact that, in the real filter structure, the fields are not perfectly homogeneous in the whole gap area.

Example 23.13 *High-Power Interdigital Filter*

This example analyzes the power handling capability of a five-pole inline interdigital filter. A 3D view of the filter is shown in Figure 23.74. It consists of five resonators with coaxial input and output ports. The *S*-shaped wire connecting the outer resonators to the coaxial ports is used to avoid temperature stress at high-power levels. Tuning screws are located on top of the resonators in order to adjust the frequency response once the filter is manufactured.

The frequency response of the filter is shown in Figure 23.75. It is a 100 MHz BW filter centered at 1.75 GHz. The high-power analysis of bandpass filters is typically performed at the lowest passband frequency as it combines both the highest electric field and the minimum frequency-gap product.

The electric field at 1.75 GHz is shown in Figure 23.76. As expected, the highest electric field is between the end of each resonant post and the tuning screw.

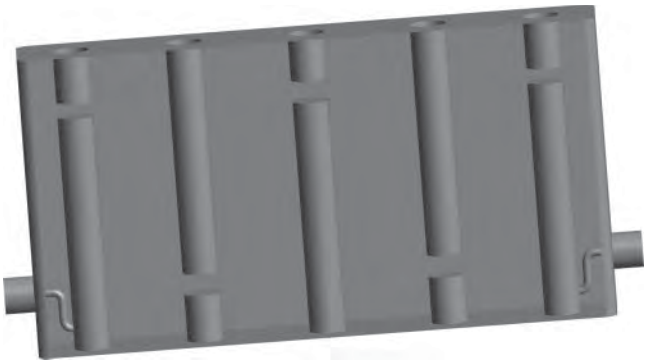


Figure 23.74 3D model of the five pole interdigital filter.

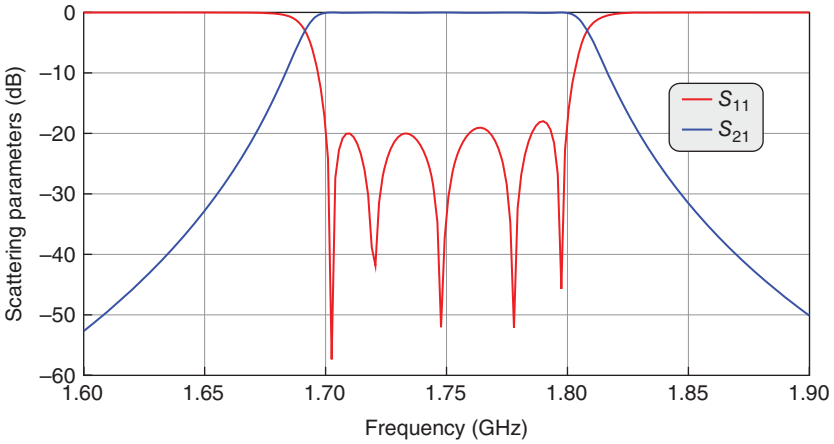


Figure 23.75 Frequency response of the filter in Figure 23.74.

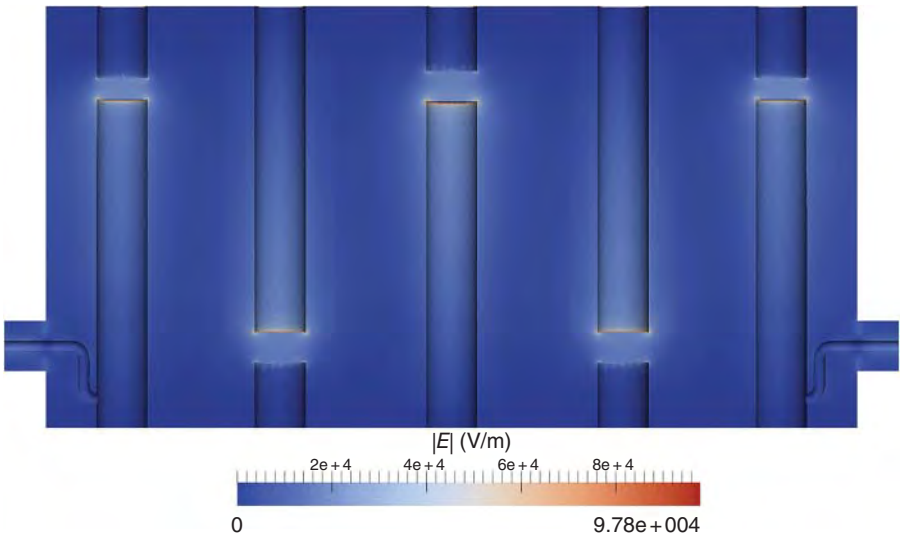


Figure 23.76 Electric field at 1.75 GHz (1 W input power).

In order to perform full analysis, the breakdown power level is computed in all resonators at 1.7, 1.75, and 1.8 GHz. To do so, first, the parallel-plate analysis based on the ECSS Multipactor Standard is done between the resonators and the tuning screw. Following the previous example, the voltage is calculated for each resonator using 1 W as the input power by integrating the field in a line connecting the resonator and the corresponding tuning screw, following their common axis, that is, at the center of the resonator. Table 23.22 shows the computed voltage of the resonators. As expected, at the center frequency of the filter, the field is symmetric in the direction of propagation due to the symmetry of the filter in that direction. Another expected result is that the maximum field takes place in the central resonator.

The next step is to determine the breakdown voltage and power. Following the procedure of the previous example, making use of the ECSS Multipactor Standard and Equation (23.3), Tables 23.22–23.24 summarize the results for the breakdown voltage and power, using silver as the surface material. The breakdown voltages of resonators 1 and 5 and 2 and 4 are identical owing to the symmetry of the structure. The results confirm that the lowest breakdown power level occurs at the lowest frequency. This behavior can be extrapolated to almost any bandpass filter configuration.

Prediction of the breakdown voltage is typically based on analytical multipactor charts, assuming a parallel-plate configuration with homogeneous fields. However, in this example, the geometry of the filter and the EM fields differ considerably from such an assumption. As a comparison, Table 23.25 describes the computed power breakdown with full 3D numerical tools [9]. It clearly demonstrates that analytical tools provide conservative values for the multipactor breakdown in practical filters.

Table 23.22 Computed voltage of the resonators using 1 W input power.

	Voltage (V)		
Resonator	**1.70 GHz**	**1.75 GHz**	**1.80 GHz**
1	66	58.8	63
2	109.7	69.7	110.2
3	115	80	115.3
4	98.5	69.7	97
5	60	58.8	56.3

Table 23.23 Computed breakdown voltage of resonators for the corresponding frequency-gap product.

		Breakdown voltage (V)		
Resonator	**Gap (mm)**	**1.70 GHz**	**1.75 GHz**	**1.80 GHz**
1	2.5	265	273.5	281
2	3.57	380	391	402
3	3.56	378.5	390	401
4	3.57	380	391	402
5	2.5	265	273.5	281

Table 23.24 Computed breakdown power levels for the four inner sections of the filter.

		Breakdown power (W)		
Resonator	**Gap (mm)**	**1.70 GHz**	**1.75 GHz**	**1.80 GHz**
1	2.5	16	21.5	20
2	3.57	12	31.5	13
3	3.56	10.8	23.5	12
4	3.57	15	31.5	17
5	2.5	19.5	21.5	25

Table 23.25 Breakdown power levels for the four inner sections using a full 3D numerical solver as [9].

		Breakdown power (W)		
Resonator	**Gap (mm)**	**1.70 GHz**	**1.75 GHz**	**1.80 GHz**
1	2.5	29	32	37
2	3.57	23	61	27
3	3.56	20	47	24
4	3.57	28	61	34
5	2.5	78	32	29

Table 23.26 Breakdown voltages calculated in resonator 3 with the three different analysis methods.

	Breakdown voltage (V)		
Prediction	**1.70 GHz**	**1.75 GHz**	**1.80 GHz**
Numerical	514	548	565
Numerical PP	441	456	474
ECSS PP	379	390	401

As expected, the breakdown power in the critical gap (resonator 3 at 1.7 GHz) is around 3 dB higher. In order to verify that this difference actually comes from the deviation from the parallel-plates geometry, a parallel-plate case with a gap of 3.56 mm has been also numerically simulated. Table 23.26 shows the breakdown voltage at resonator 3 for the three kinds of prediction methods.

Figure 23.77 shows graphically the results. In this case a noticeable difference between the actual structure and the parallel plates can be seen.

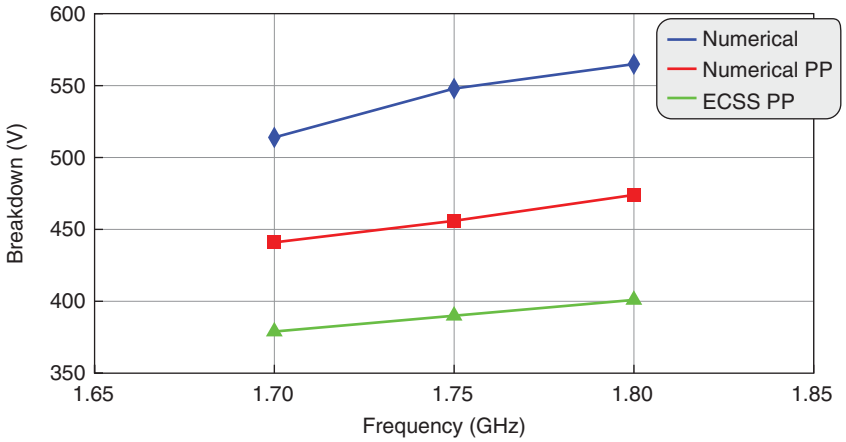


Figure 23.77 Breakdown voltages calculated in resonator 3 with the three different analysis methods.

Example 23.14 ***High-Power Capability of Dual-Mode Filters***

This example shows the high-power analysis of a four-pole dual-mode filter with two transmission zeros. A 3D view of the filter is shown in Figure 23.78. The filter consists of two cylindrical cavities (dual mode) connected through a cross-shaped iris. The input/output couplings to the standard rectangular waveguide access ports (WR-75) are achieved by means of rectangular irises. In addition to the diagonal coupling screws, tuning screws are located in the cylindrical cavities to adjust the frequency response once the filter is manufactured.

The in-band S-parameter response of this filter is shown in Figure 23.79. It is a 76 MHz BW bandpass response centered at 12 GHz (the filter designed in example 23.4). The high-power analysis of bandpass filters is typically performed in three points—lowest, center, and highest passband frequencies—in order to find the worst-case combination of both the electric field and $f \times d$ product. The highest breakdown risk is expected to occur either in the input/output irises or at the central cross-section iris. Cavities are not considered because of the huge gaps involved. Therefore, in order to perform a full analysis, the breakdown power level is calculated in the central cross-shaped and input/output rectangular irises at 11.962, 12, and 12.038 GHz.

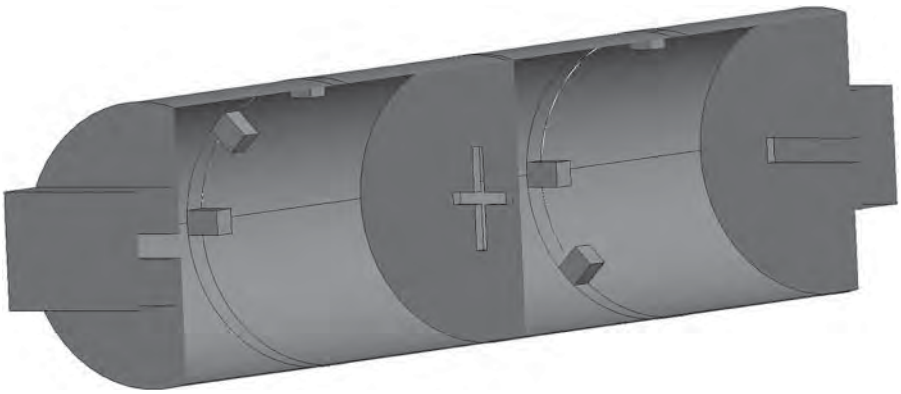


Figure 23.78 3D model of the four-pole dual-mode filter.

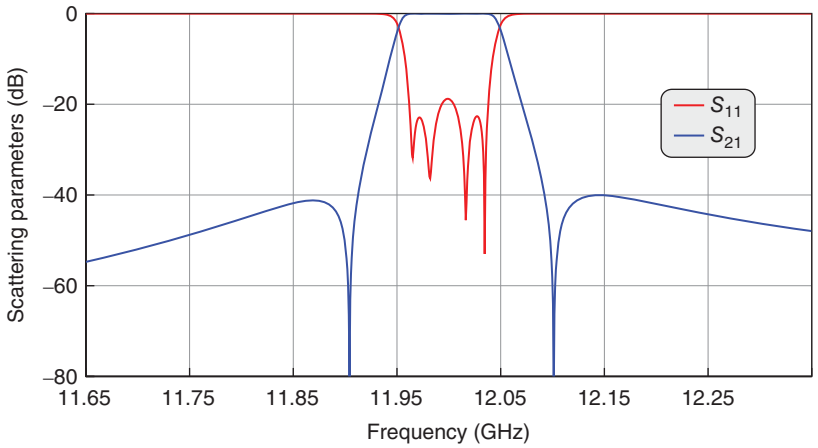


Figure 23.79 Frequency response of the filter in Figure 23.78.

As seen in the two previous examples, the breakdown power calculation using parallel-plate analysis involves field integration along the multipactor gap to find the voltage at 1 W. In the case of the input/output irises, it is quite straightforward to determine the integration path. Since they are rectangular waveguides, the electron resonance will take place between the two closest surfaces at the position of maximum field. In the case of the cross-shaped iris, however, it is not trivial to determine such path. The *E*-field in the middle of the iris (i.e., at a plane located at the middle point in the propagation direction) at 12 GHz is shown in Figure 23.80a. The integration path yielding the maximum voltage has been taken (white line) in order to carry out the parallel-plate analysis. The field along this path is represented in Figure 23.80b.

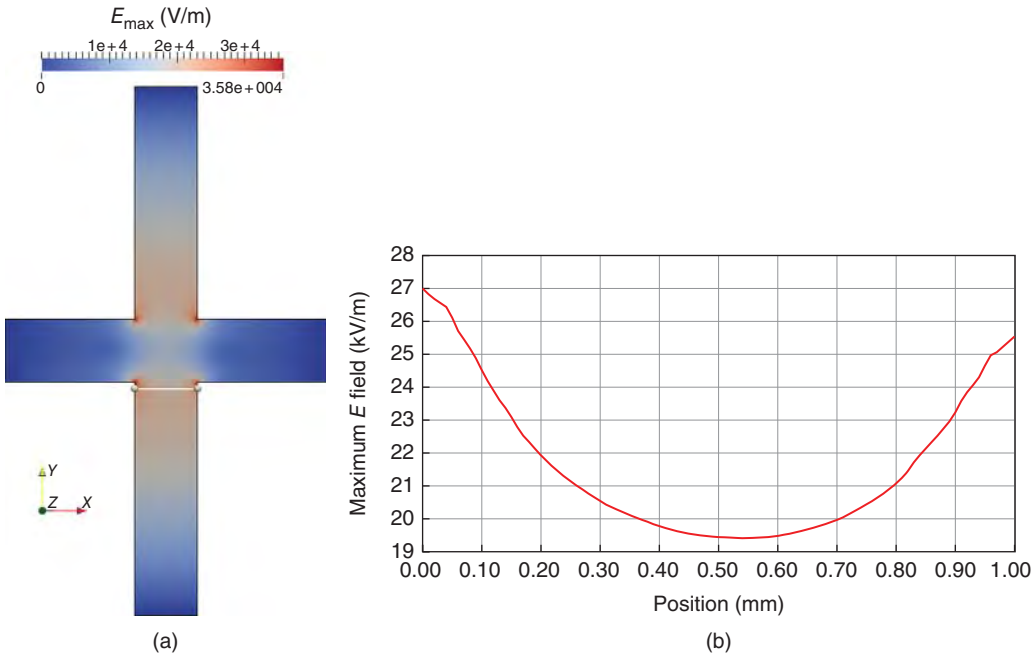


Figure 23.80 Electric field in cross-shaped iris of Example 23.4 at 12 GHz, (a) a pictorial view of the E field across the iris and (b) E field variation along the path highlighted in (a).

Table 23.27 Calculated voltage at 1 W input power in the three considered irises for three different frequencies.

	Voltage (V)		
Region	**11.962 GHz**	**12 GHz**	**12.038 GHz**
Input iris	51.5	43.4	53.1
Cross	41.7	21.6	13.1
Output iris	48	48	48

The voltage at 1 W input power has been computed for the three considered irises. Table 23.27 shows the results. It can be seen that, whereas the voltage at the input/output irises remains rather constant at all frequencies, the voltage at the cross-shaped iris dramatically increases at lower frequencies. This behavior cannot be extrapolated to any dual-mode filter configuration but indicates the need to carefully assess multipactor in all critical gaps in this component and, also, at different frequencies covering the passband of the filter.

Considering silver as the surface material of the component and knowing the $f \times d$ product, one can use the multipactor susceptibility charts available in the literature to find the breakdown voltage level as shown in Figure 20.9. Once again, the ECSS Multipactor Standard has been used to compute the breakdown voltages at the different sections and frequencies, as shown in Table 23.28. Breakdown voltages for input/output iris are identical, since the gap is 2 mm for both of them. The breakdown voltage at the central cross iris is lower since the gap is only 1 mm. Since this is a narrowband filter, the analyzed frequencies do not yield significant differences in the breakdown voltage levels (frequency-gap products are very similar in the whole frequency band).

Since both the voltage at 1 W input power in the critical gaps and the breakdown voltage levels are known, the breakdown power level can be easily found by using again equation (23.3). Table 23.29 clearly shows that for higher frequencies the lowest breakdown level occurs at the input/output irises. However, for the lowest frequency, the huge increase of the voltage at the cross-shaped iris has a large impact on the breakdown level, and, as a consequence, this iris becomes the critical area with respect to multipactor breakdown.

As already stated in the previous example, this kind of prediction, based on analytical multipactor charts, assumes a parallel-plate case with homogeneous fields. However, in this example, the geometry of the irises and the EM field distribution differ considerably from such assumption. One of the consequences of the iris shape is that electrons can easily escape from the gap, moving to large areas (cavities) where they are absorbed, thus inhibiting the discharge [10].

Table 23.28 Breakdown voltage values for the three irises, considering silver as surface material and the corresponding $f \times d$ product.

		Breakdown voltage (V)		
Region	**Gap (mm)**	**11.962 GHz**	**12 GHz**	**12.038 GHz**
Input iris	2	1501	1506	1510
Cross	1	749	752	754
Output iris	2	1501	1506	1510

Table 23.29 Breakdown power levels in each iris and for the whole frequency band.

		Breakdown power (W)		
Region	**Gap (mm)**	**11.962 GHz**	**12 GHz**	**12.038 GHz**
Input iris	2	849	1204	809
Cross	1	323	1211	3314
Output iris	2	978	984	990

Table 23.30 Breakdown power levels for the four inner sections using a 3D numerical solver [9].

		Breakdown power (W)		
Region	**Gap (mm)**	**11.962 GHz**	**12 GHz**	**12.038 GHz**
Input iris	2	2453	3281	2266
Cross	1	930	3406	8937
Output iris	2	2719	2719	2781

Table 23.31 Breakdown voltages calculated at the cross-section iris with the three different analysis methods.

	Breakdown voltage (V)		
Prediction	**11.962 GHz**	**12 GHz**	**12.038 GHz**
Numerical	1272	1261	1238
Numerical PP	883	884	889
ECSS PP	749	752	754

Therefore, it is expected that the simulation with 3D numerical tools [9] will yield significantly higher breakdown levels. Table 23.30 shows the breakdown power levels of this dual-mode filter computed with such a numerical tool.

As expected, the breakdown power in the critical gap (cross-shaped iris at 11.962 GHz) is almost 5 dB higher than the analytical prediction. In order to verify that this difference actually comes from the deviation from the parallel-plates geometry, a parallel-plate case with a gap of 1 mm has been also numerically simulated. Table 23.31 shows the breakdown voltage at the cross iris, for the three kinds of breakdown predictions.

Figure 23.81 shows graphically the results. In this case a noticeable difference between the actual structure and the parallel plates can be seen.

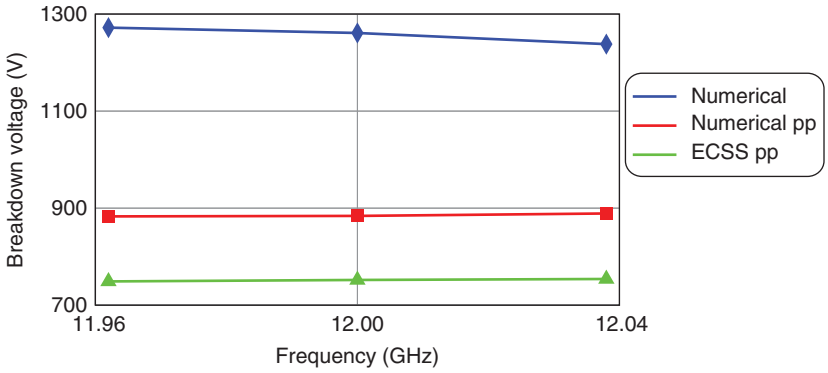


Figure 23.81 Breakdown voltages calculated in cross-section iris with the three different analysis methods.

23.5 Tolerance and Sensitivity Analysis in Filter Design

Advances in EM techniques have made a major impact in the implementation of practical filter networks. It has reduced, and even eliminated (depending upon the specifications), the need to construct engineering models prior to going into production. The overall impact is reduced cost, improved performance, and faster turnaround from the design stage to the manufacture of equipment.

23.5.1 Manufacturing Tolerance Analysis

Manufacturing tolerances and ease of alignment in meeting specifications are critical parameters to consider in the practical implementation of filter networks. This section presents a simple overview of manufacturing tolerances and sensitivity analysis.

The inductive filter of Example 23.3 is chosen to demonstrate the effect on filter response due to manufacturing tolerances. Let us assume that the machining tolerance can be controlled within ± 10 μm for any dimension in the filter. The two parameters that are of primary interest to the filter designer are as follows:

1) Worst-case performance deviations with respect to specification requirements.
2) Sensitivities of the different physical parameters. The sensitivities are defined in a classical way as a variation of the response with respect to a variation in a particular dimension. For example, $\partial S_{11}/\partial w_1$ is the variation of the reflection coefficient with respect to the first iris window.

Both problems are treated statistically. An EM solver can run several simulations with some principal types of variations: randomly chosen, uni- or bidirectional, maximum/minimum values, and so on. There are some options to choose randomly the dimensions following different distribution functions. The commonly used functions are as follows:

- Uniform distribution. The simplest although the less realistic. It is assumed that the error in a physical dimension can happen with the same probability in the whole range within the limits. The probability density function is a constant within a lower and an upper limit and zero outside. The cumulative distribution function is a ramp between 0 and 1.

- Normal distribution. More realistic, with a mean (around the exact value if the machine is not biased) and a standard deviation (e.g., the ±10 μm). The only drawback is that the range of this distribution is infinite. There are less likely events for the normal distribution that are impossible for the machine. The formulas for this distribution are

$$\phi(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{the probability density function}$$

$$\Phi(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} d\xi \quad \text{the cumulative distribution function}$$

- Truncated normal distribution. This distribution is more realistic. The formulas for this distribution are

$$f(x;\mu,\sigma,a,b) = \frac{\phi(x;\mu,\sigma)}{\Phi(b;\mu,\sigma) - \Phi(a;\mu,\sigma)} \quad \text{the probability density function}$$

$$F(x;\mu,\sigma,a,b) = \frac{\Phi(x;\mu,\sigma) - \Phi(a;\mu,\sigma)}{\Phi(b;\mu,\sigma) - \Phi(a;\mu,\sigma)} \quad \text{the cumulative distribution function}$$

where a and b are the limits and μ and σ are the mean and the standard deviation, respectively. Here $a < x < b$, and ϕ and Φ are the functions defined in the normal distribution function.

In order to perform a tolerance analysis, we have chosen this distribution. 100 simulations have been carried out with randomly chosen widths and lengths with mean μ, standard deviation $\sigma = 10$ μm for all variables, and ultimate limits $a = \mu - 15$ μm and $b = \mu + 15$ μm. The probability density function and the cumulative distribution function assuming a zero mean are plotted in Figure 23.82.

The next step in this methodology is to perform a Monte Carlo analysis varying the widths of the coupling windows and the lengths of the resonators. The plot is shown in Figure 23.83, where we can see that the insertion loss is almost unaffected and the worst-case variation in the RL is around 3 dB. The results for $\sigma = 25.4$ μm (=1 mil or 1000th of an inch) and a limit of ±38.1 μm (1.5 mils) are displayed in Figure 23.84.

The reader may raise the question as to which performance parameter is most sensitive to variations in dimensions. Alternatively, we need to determine which variables have the biggest impact on a specific performance parameter. This kind of sensitivity analysis is very useful to identify the most important parameters or physical dimensions. This knowledge will provide some insight of the structure, and based upon it the engineer can decide the strategy for a possible tuning process.

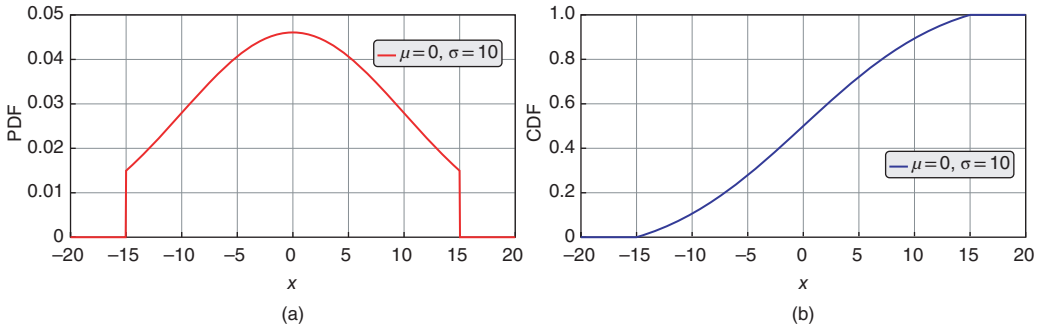


Figure 23.82 Probability density function (a) and cumulative distribution function (b) for the truncated normal distribution with $\mu = 0$, $\sigma = 10$, $a = -15$, and $b = 15$.

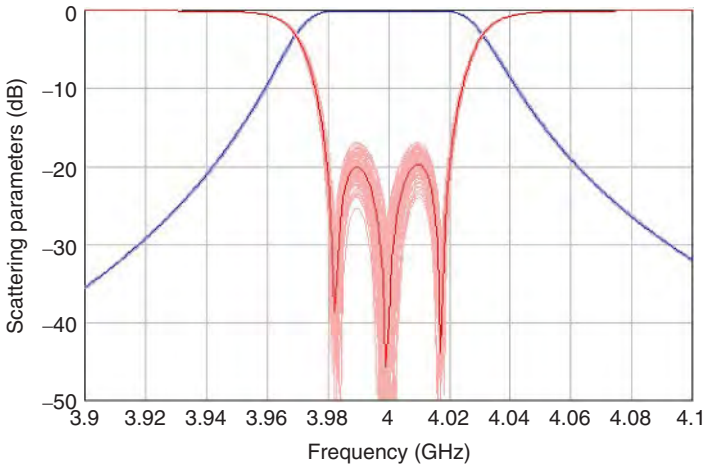


Figure 23.83 Tolerance analysis of the inductive filter of Example 23.3. Chosen variables: Window widths and resonator lengths. Standard deviation is $\sigma = 10$ μm and the maximum allowed deviation is ± 15 μm.

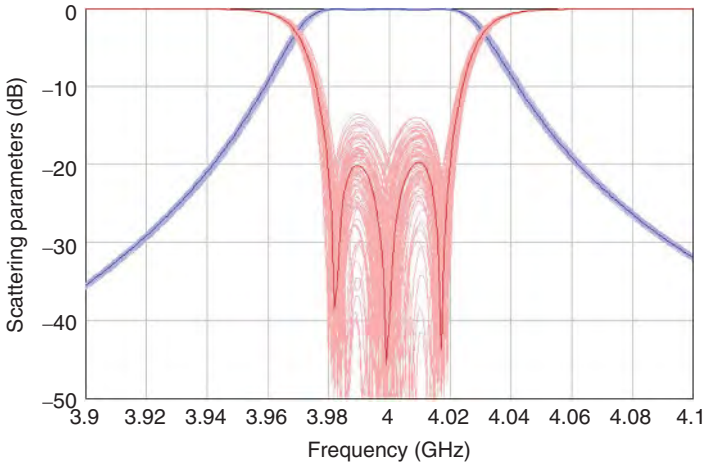


Figure 23.84 Tolerance analysis of the inductive filter of Example 23.3. Chosen variables: Windows widths and resonator lengths. Standard deviation is $\sigma = 1$ mil and the maximum allowed deviation is ± 1.5 mils.

If we represent with bars the variation of some function (in this case $|S_{11}|$) with respect to the variation of some parameter among many (in this case w_1, w_2, l_1, l_2), we obtain a Pareto chart. After a thousand EM simulations with statistical variations of the four parameters, we obtain the Pareto chart shown in Figure 23.85.

The analysis presented here is for a very simple filter, and it is intended as an introduction to sensitivity analysis. Practical filters of higher order, typically with transmission zeros, tend to be very sensitive to machining tolerances and invariably require tuning elements to capture the ideal performance.

23.5.2 Sensitivity Analysis of a Dielectric-Loaded Cavity

For dielectric-loaded filters, sensitivity analysis needs to include any variation in the dielectric constant (stated by the manufacturer) of the material used for resonators, as well as

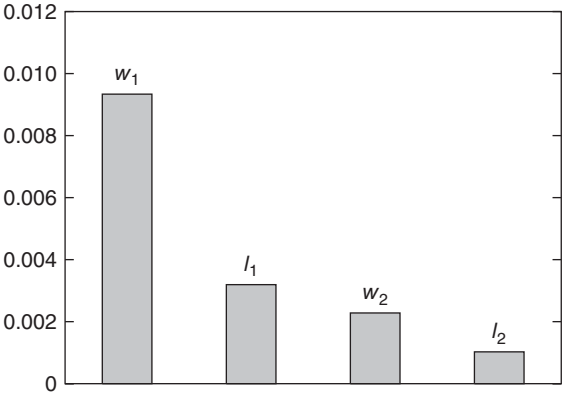


Figure 23.85 Pareto chart for the variation in S_{11} after 1000 EM simulations with random variation of the four parameters.

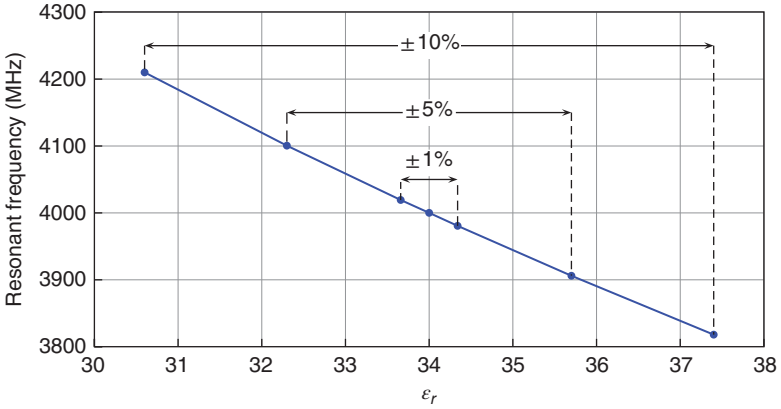


Figure 23.86 Variation of the resonant frequency with respect to variation of ε_r.

manufacturing tolerances in the fabrication of dielectric pucks and the housings required to support the pucks. Such an analysis is complex and very time-consuming. For reasons of simplicity, we analyze here a single isolated dielectric-loaded cavity of Example 23.6. This simplified analysis provides a good window in understanding dielectric-loaded components and an approximate measure of the sensitivities involved in dielectric-loaded structures.

Let us assume a single dielectric-loaded cavity from Example 23.6. The cavity and dielectric dimensions are reported in Figure 23.36. Performing an EM analysis, the resonant frequency is centered at 4 GHz. However, the dielectric material has a nominal value $\varepsilon_r = 34$ with some tolerance given by the manufacturer.

The dielectric material is machined according to the dimensions obtained for the diameter D and the height L_d. These two dimensions are subject to machining tolerances, and therefore a resonant frequency shift is expected. For the sake of simplicity, we assume that the cavity dimensions and the size and characteristics of the dielectric support are fixed to the nominal values. We are interested in the study of the variation of the resonant frequency due to the variations of the dielectric permittivity and its dimensions.

The variation of the resonant frequency with the dielectric constant is shown in Figure 23.86. The typical points for 1%, 5%, and 10% variations in ε_r are shown.

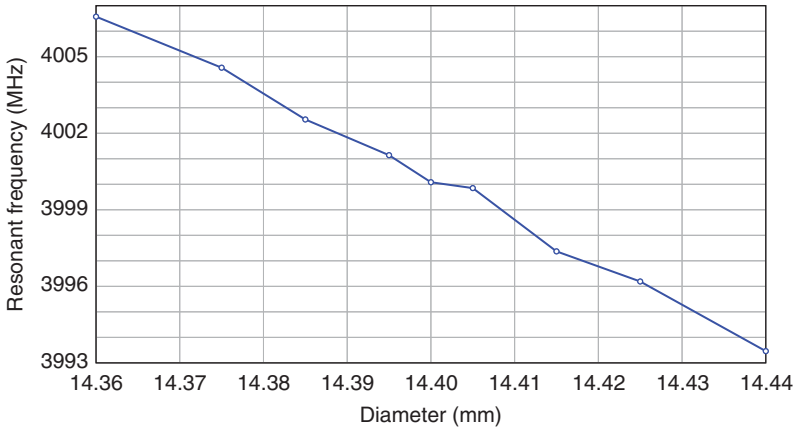


Figure 23.87 Variation of the resonant frequency with respect to variation of the diameter of the dielectric material *D*.

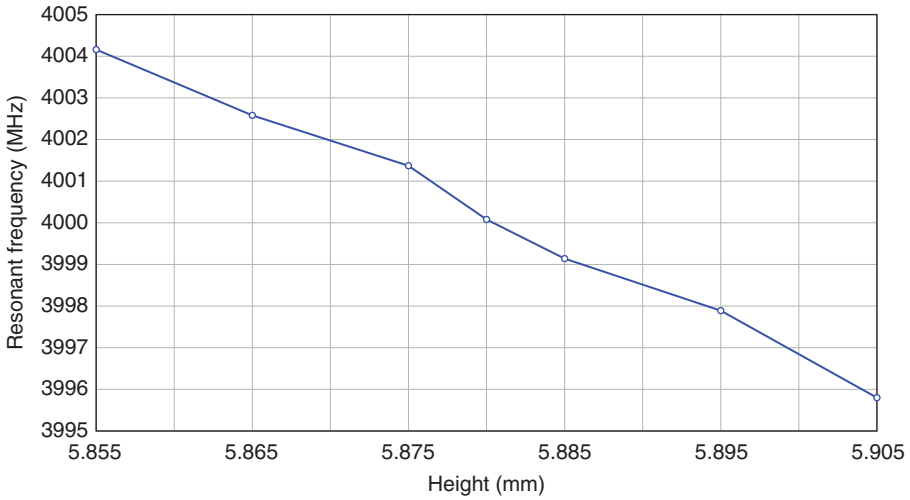


Figure 23.88 Variation of the resonant frequency with respect to variation of the height L_d of the dielectric puck.

The study of the variation of the resonant frequency with respect to machining of the dielectric piece is divided into two categories. In Figure 23.87 the variation with the diameter is shown. The points shown in the curve are reference point at ±5, ±15, ±25, and ±40 μm.

In Figure 23.88 the variation of the frequency is computed in terms of the variation of the height of the dielectric piece L_d. The points shown are located at ±5, ±15, and ±25 μm.

Sensitivity analysis based on a single dielectric-loaded cavity shows how critical the effect is on the resonant frequency with respect to any variation in the dielectric constant of the material or in the dimensions of the resonant pucks during the manufacturing process. For many applications, it is necessary to ensure that the dielectric pucks are resonating at the right frequency before installation in their housing.

23.5.3 Degradation Caused by Tuning Screws in Filter Design

23.5.3.1 Degradation of Quality Factor (Q_0)

Any discontinuity in a microwave structure tends to introduce losses and therefore adversely affect its quality factor or unloaded Q, and tuning elements are no exception. Tuning screws are extensively employed in filter networks. For dual-mode filters, they are essential to provide coupling between orthogonal modes. For high performance filters, tuning screws are used to mitigate the effects of manufacturing tolerances of resonators and coupling elements, and their use provides the most cost-effective approach in the production of practical filter networks. As an illustration, we consider here a single dual-mode waveguide cavity, specifically one of the cavities used in Example 23.4, to analyze the effect on Q_0 as a function of penetration depth of tuning screws.

The top and side views are shown in Figure 23.89. Cavity and screw dimensions are the same ones used for Example 23.4, namely, $R = 11.557$ mm, $L = 48.43$ mm, and $d = 2$ mm.

In practical filters, smoothness of the surface, quality of silver plating, and input and output coupling elements further tend to influence the quality factor. Realization of 60–70% of the theoretical Q is deemed to be the industry standard. Higher values of Q are achieved by using higher propagation modes (e.g., TE_{113} in Example 23.4) at the cost of larger size and mass.

Figure 23.90 shows the computed value of the unloaded Q as a function of the screw depth. The cavity is composed of copper (conductivity $\sigma = 5.88 \times 10^7$ S/m). The nominal Q of this cavity (without screw) is computed to be 17,461.

23.5.3.2 Degradation of Spurious-Free Window

Tuning elements not only degrade the quality factor but also introduce spurious modes. Such modes cause degradation in the out-of-band rejection response of filter networks. This can be problematic depending upon the system requirements to control interference to other communication systems. Such modes can also be a potential source of noise into the receive band of the system under consideration. As an illustration of this effect, we again consider the same dual-mode cavity used to analyze the degradation of the quality factor.

Figure 23.91 shows the mode chart of a single 12 GHz resonant cavity used in Example 23.14. It covers the transmit and receive bands (11.7–12.2 and 14.0–14.5 GHz, respectively)

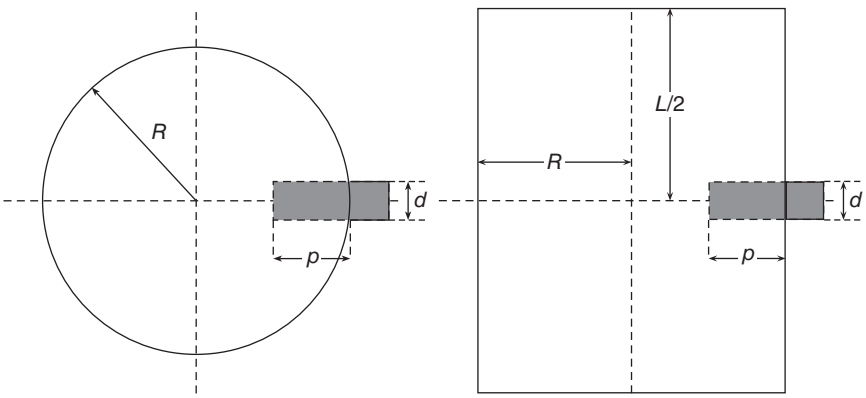


Figure 23.89 Top view (left) of a dual-mode cavity in circular waveguide with $R = 11.557$ mm. Side view (right) shows the placement of the screw in the middle of the cavity length $L = 48.43$ mm. The screw has circular cross-section with diameter $d = 2$ mm.

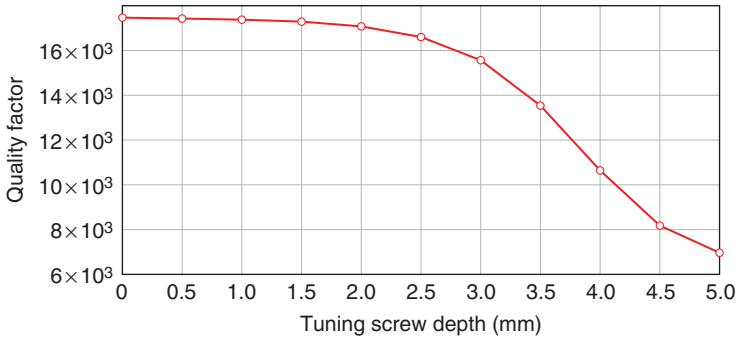


Figure 23.90 EM simulation of the degradation of *Q* as a function penetration depth of the tuning screw in a 12 GHz resonant cavity using TE_{113} mode.

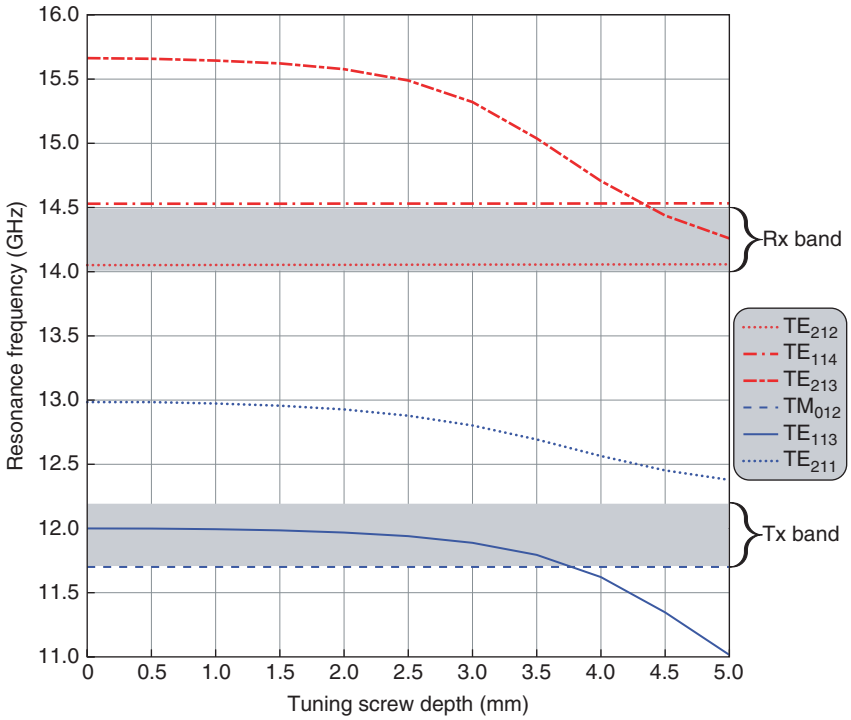


Figure 23.91 Variation of the resonance for the TE_{113} mode and the two adjacent modes to provide information about the spurious-free window.

of a Ku-band satellite system. The propagation mode is TE_{113}. Modes close to the transmit and receive bands are included in the figure. It shows that, with increasing depth of the tuning screw, mode TE_{213} tends to fall within the receive band, whereas mode TE_{211} at 13 GHz is tending toward the transmit band. Incidence of spurious modes in a high-power transmit multiplexer is much higher and more complex as compared with the simple analysis based on a single cavity. The intent here is to broaden the understanding of the impact of filter hardware on the overall system performance.

Summary

The primary aim of this chapter is to bridge the gap between theory and the practical realization of microwave filters and multiplexing networks. It addresses the constraints imposed by the operating environment, technology limitations, and manufacturing tolerances. These constraints are captured by invoking the concept of ELFD. It describes the rationale of this concept and its use in assessing tradeoffs between achievable performance under typical operating environment and manufacturing tolerances. Another aim of this chapter is to add clarity in the understanding of system parameters that impact filter specifications in various parts of a communication system. In conjunction with ELFD, this provides the understanding and the basis to develop optimum specifications of filter networks. The next step is the design methodology for the realization of practical filters. This is accomplished by a series of examples that address all aspects of filter design, including high-power handling capability. Starting with the basic pole–zero configuration of a filter function, a simple lumped-element-based circuit model of the filter is developed; this model is then augmented selectively (to keep it simple) by distributed elements. This quasi-distributed model is used as the initial design to start the process of computation of the physical dimensions using EM techniques, including full-wave analysis and optimization. A key feature in all examples is the detailed description of the step-by-step design process. The examples are related to practical communication systems and complement the theoretical aspects of filter and multiplexer design. The chapter concludes with a brief and simple overview of EM-based tolerance and sensitivity analysis in filter design.

Acknowledgments

The authors of this chapter would like to thank all members from Microwave Applications Group (GAM), Technical University of Valencia (Spain), and specially to our colleague Dr Pablo Soto, for their active contribution to the results included in several examples. Authors are also in debt to staff members of Aurora Software and Testing S.L. (AURORASAT), in particular with Dr Carlos Vicente, Dr Jordi Gil, and Dr Sergio Anza, for their help with the preparation of examples about high-power issues on filters.

Appendix 23A

Thermal Expansion

Most solid materials expand when heated. Suppose a rod or hollow cylinder or rectangular cavity has a length L_0 at some initial temperature T_0. When the temperature increases by an amount ΔT, the length increases by ΔL. Experiments show that if ΔT is not too large, ΔL is directly proportional to ΔT and L_0. Introducing a proportionality constant α (different for different materials), we have

$$\Delta L = \alpha L_0 \Delta T \tag{23A.1}$$

The constant α, which characterizes the thermal expansion properties of a particular material, is called the temperature coefficient of linear expansion. Two assumptions implied in this

Table 23A.1 Coefficient of linear thermal expansion.

Material	α (ppm/°C)
Aluminum	24
Brass	20
Copper	17
Invar	0.9
Quartz	0.4

formulation need to be addressed. For materials having no preferential directions, every linear dimension changes in accordance with equation (23A.1). However, we are mainly concerned with the change in length, which changes the resonant frequency of resonators, and not with the thickness (or volume), which has a very minor effect on the resonator characteristics. Moreover, the equation is approximately correct for small temperature changes as typically encountered in the operating environments of communication systems. Thus, we may define by the equation

$$\alpha = \frac{1}{L}\frac{dL}{dT} \tag{23A.2}$$

α can then be expressed as change in length per unit length per degree celsius. For practical applications, it is defined in terms of parts per million ($\times 10^{-6}$) per degree celsius or ppm/°C. Average values for a range of materials (raw) are described in Table 23A.1.

References

1 Kudsia, C., Ainsworth, K., and O'Donovan, M. (1980) Microwave filters and multiplexing networks for communications satellites in the 1980s. Proceedings of the AIAA 8th Communications Satellite Systems Conference, pp. 290–301, April 1980.

2 Belevitch, V. (1952) Chebyshev filters and amplifier networks. *Wireless Engineer*, **29**, 106–110.

3 Marcuvitz, N. (1986) *Waveguide Handbook*, Electromagnetic Waves Series, Peter Peregrinus Ltd, IEE, London.

4 Cogollos, S., Brumos, M., Boria, V.E. *et al.* (2012) A systematic design procedure of classical dual-mode circular waveguide filters using an equivalent distributed model. *IEEE Transactions on Microwave Theory and Techniques*, **60**, 1006–1017.

5 Bandler, J., Cheng, Q., Dakroury, S. *et al.* (2004) Space mapping: the state of the art. *IEEE Transactions on Microwave Theory and Techniques*, **52**, 337–361.

6 Peik, S. and Mansour, R. (2002) A novel design approach for microwave planar filters. Microwave Symposium Digest, 2002 IEEE MTT-S International, vol. 2, pp. 1109–1112.

7 Levy, R. (1965) Tables of element values for the distributed low-pass prototype filter. *IEEE Transactions on Microwave Theory and Techniques*, **13**, 514–536.

8 Morini, A., Rozzi, T., and Morelli, M. (1997) New formulae for the initial design in the optimization of T-junction manifold multiplexers. Microwave Symposium Digest, 1997, IEEE MTT-S International, vol. 2, pp. 1025–1028, vol.2.

9 Anza, S., Vicente, C., Raboso, D., Gil, J., Gimeno, B., and Boria, V.E. (2008) Enhanced prediction of multipaction breakdown in passive waveguide components including space charge effects. Microwave Symposium Digest, 2008 IEEE MTT-S International, pp. 1095–1098.
10 Anza, S., Esteve, R., Armendariz, J., Vicente, C., Gil, J., and Raboso, D. (2014) Study of high order modes and fringing fields in multipactor effect. Proceedings of the 8th International Workshop on Multipactor, RF and DC Corona and Passive Intermodulation in Space RF Hardware, pp. 1095–1098.

A

Physical Constants

Speed of light in free-space, $c = 2.998 \times 10^8$ m/s
Permittivity of free-space, $\varepsilon_0 = 8.854 \times 10^{-12}$ F/m
Permeability of free-space, $\mu_0 = 4\pi \times 10^{-7}$ H/m
Impedance of free-space, $\eta_0 = 376.7\ \Omega$
Charge of electron, $e = 1.602 \times 10^{-19}$ C
Mass of electron, $m = 9.107 \times 10^{-31}$ kg
Boltzmann's constant, $k = 1.380 \times 10^{-23}$ J/K

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

B

Conductivities of Metals

Material	Conductivity S/m (20 °C)	Material	Conductivity S/m (20 °C)
Aluminium	3.816×10^7	Mercury	1.04×10^6
Brass	2.564×10^7	Lead	4.56×10^6
Bronze	1.00×10^7	Nickel	1.449×10^7
Chromium	3.846×10^7	Platinum	9.52×10^6
Copper	5.813×10^7	Silver	6.173×10^7
Germanium	2.2×10^6	Steel (stainless)	1.1×10^6
Gold	4.098×10^7	Solder	7.0×10^6
Graphite	7.0×10^4	Tungsten	1.825×10^7
Iron	1.03×10^7	Zinc	1.67×10^7

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

C

Dielectric Constants and Loss Tangents of Some Materials

Material	Frequency	ϵ_r	tan δ
Alumina	10 GHz	9.7–10	0.0002
Fused quartz	10 GHz	3.78	0.0001
Gallium arsenide	10 GHz	13	0.0016
Glass (pyrex)	3 GHz	4.82	0.0054
Glazed ceramic	10 GHz	7.2	0.008
Plexiglass	3 GHz	2.60	0.0057
Polyethylene	10 GHz	2.25	0.0004
Polystyrene	10 GHz	2.54	0.00033
Porcelain	100 GHz	5.04	0.0078
Rexolite	3 GHz	2.54	0.00048
RT/duriod 5880	10 GHz	2.2	0.0009
RT/duriod 6002	10 GHz	2.94	0.0012
RT/duriod 6006	10 GHz	6.15	0.0019
RT/duriod 6010	10 GHz	10.8	0.0023
Silicon	10 GHz	11.9	0.004
Styrofoam	3 GHz	1.03	0.0001
Teflon	10 GHz	2.08	0.0004
Vaseline	10 GHz	2.16	0.001
Water (distilled)	3 GHz	76.7	0.157

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

D

Rectangular Waveguide Designation

Designation WR-XX	Recommended Frequency Range (GHz)	TE_{10} Cutoff Frequency (GHz)	Inside Dimensions Inches (cm)
WR-650	1.12–1.70	0.908	6.500 × 3.250 (16.51 × 8.255)
WR-430	1.70–2.60	1.372	4.300 × 2.150 (10.922 × 5.461)
WR-284	2.60–3.95	2.078	2.840 × 1.340 (7.214 × 3.404)
WR-187	3.95–5.85	3.152	1.872 × 0.872 (4.755 × 2.215)
WR-137	5.85–8.20	4.301	1.372 × 0.622 (3.485 × 1.580)
WR-112	7.05–10.0	5.259	1.122 × 0.497 (2.850 × 1.262)
WR-90	8.20–12.4	6.557	0.900 × 0.400 (2.286 × 1.016)
WR-62	12.4–18.0	9.486	0.622 × 0.311 (1.580 × 0.790)
WR-42	18.0–26.5	14.047	0.420 × 0.170 (1.07 × 0.43)
WR-28	26.5–40.0	21.081	0.280 × 0.140 (0.711 × 0.356)
WR-22	33.0–50.5	26.342	0.224 × 0.112 (0.57 × 0.28)

(Continued)

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

Designation WR-XX	Recommended Frequency Range (GHz)	TE_{10} Cutoff Frequency (GHz)	Inside Dimensions Inches (cm)
WR-19	40.0–60.0	31.357	0.188×0.094 (0.48×0.24)
WR-15	50.0–75.0	39.863	0.148×0.074 (0.38×0.19)
WR-12	60.0–90.0	48.350	0.122×0.061 (0.31×0.015)
WR-10	75.0–110.0	59.010	0.100×0.050 (0.254×0.127)
WR-8	90.0–140.0	73.840	0.080×0.040 (0.203×0.102)
WR-6	110.0–170.0	90.854	0.065×0.0325 (0.170×0.083)
WR-5	140.0–220.0	115.750	0.051×0.0255 (0.130×0.0648)

E

Impedance and Admittance Inverters

The lowpass prototype filter design comprises inductive and capacitive elements arranged alternately in series and parallel configuration in a ladder network as described in Figure E.1. In the physical realization of microwave filters, it is preferable to deal with the same distributed element (be it an inductor or capacitor) throughout the filter structure, which makes structure simpler and more robust. This can be accomplished by the use of impedance or admittance inverters.

Figure E.2a describes an impedance inverter. It is equivalent to an ideal quarter-wave transformer with a characteristic impedance $\sqrt{K}$ connected to an impedance Z_L. The quarter-wave line transforms the impedance given by the relationship

$$Z_{\text{in}} = \frac{K^2}{Z_L} \tag{E.1}$$

For $K = 1$, $Z_{\text{in}} = 1/Z_{L,}$ and $Z_{\text{in}} = 1$ for $K = \sqrt{Z_L}$. The load impedance connected at one end is inverted with respect to the inverter constant. In addition, inverters can be utilized to shift impedance or admittance levels by the choice of K parameter. Because of the inverting property, a series inductance with an inverter on each side appears as a shunt conductance from its exterior terminals. In a similar way, a shunt capacitance with an inverter on each side appears as a series inductance from its exterior terminals.

Similar relationship holds for admittance inverters as described in Figure E.2b:

$$Y_{\text{in}} = \frac{J^2}{Z_L} \tag{E.2}$$

E.1 Filter Realization with Series Elements

We use the simple two-pole prototype filter described in Figure E.3a to demonstrate the use of inverters in filter design. Any change in filter topology must preserve the filter transfer function. This implies that the input impedance or admittance must be maintained while manipulating changes to the filter topology. Keeping this in mind, the input impedance Z_{in} of the network in Figure E.3a is given by

$$Z_{\text{in}} = \frac{1}{g_1 s + \dfrac{1}{z_1}} = \frac{1}{g_1 s + \dfrac{1}{g_2 s + \dfrac{1}{g_3}}} \tag{E.3}$$

where $z_1 = g_2 s + 1/g_3$.

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

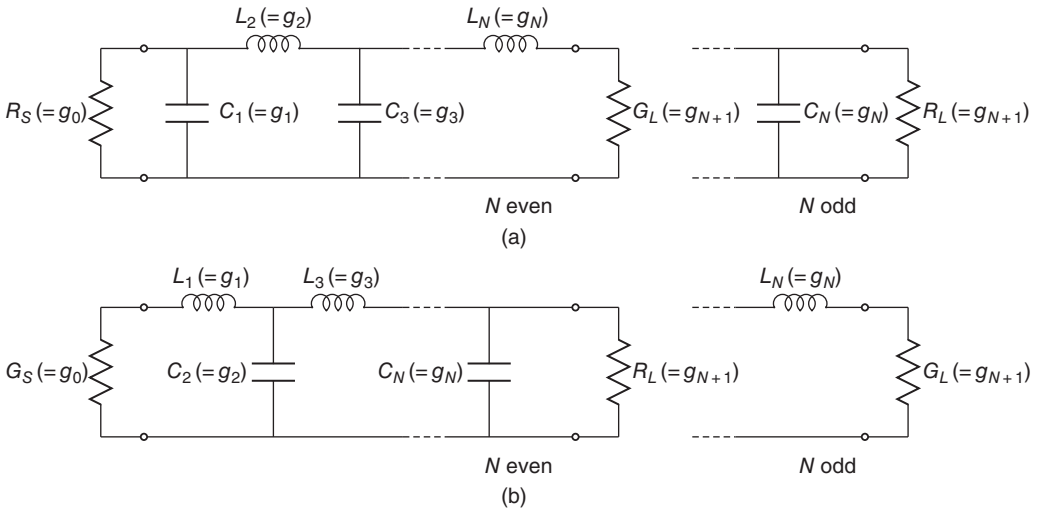


Figure E.1 Lowpass prototype ladder network, (a) where the leading component is a shunt capacitor and (b) the dual of the network.

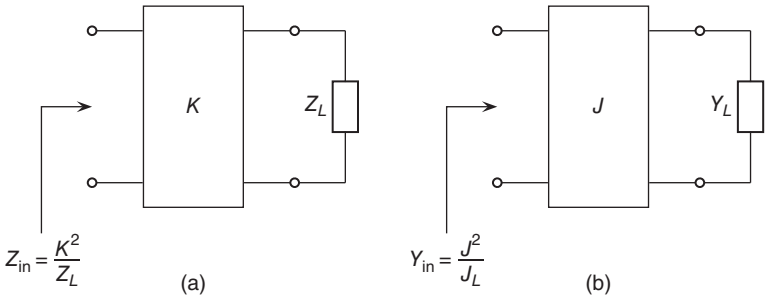


Figure E.2 Impedance inverter (a) and admittance inverter (b).

Figure E.3b shows conversion of the parallel capacitance into an equivalent series inductance by incorporating unity inverters ($K = 1$) on each side of the capacitance. The input impedance is evaluated as follows:

$$z_{1b} = g_2 s + \frac{1}{g_3} \tag{E.4a}$$

$$z_{2b} = \frac{K^2}{z_{1b}} = \frac{1}{g_2 s + \frac{1}{g_3}} \tag{E.4b}$$

$$z_{3b} = g_1 s + z_{2b} = g_1 s + \frac{1}{g_2 s + \dfrac{1}{g_3}} \tag{E.4c}$$

$$Z_{\text{in}} = \frac{K^2}{z_{3b}} = \frac{1}{g_1 s + \dfrac{1}{g_2 s + \dfrac{1}{g_3}}} \tag{E.4d}$$

This Z_{in} represents the same value as given by (E.3), thus confirming the preservation of the input impedance of the prototype filter.

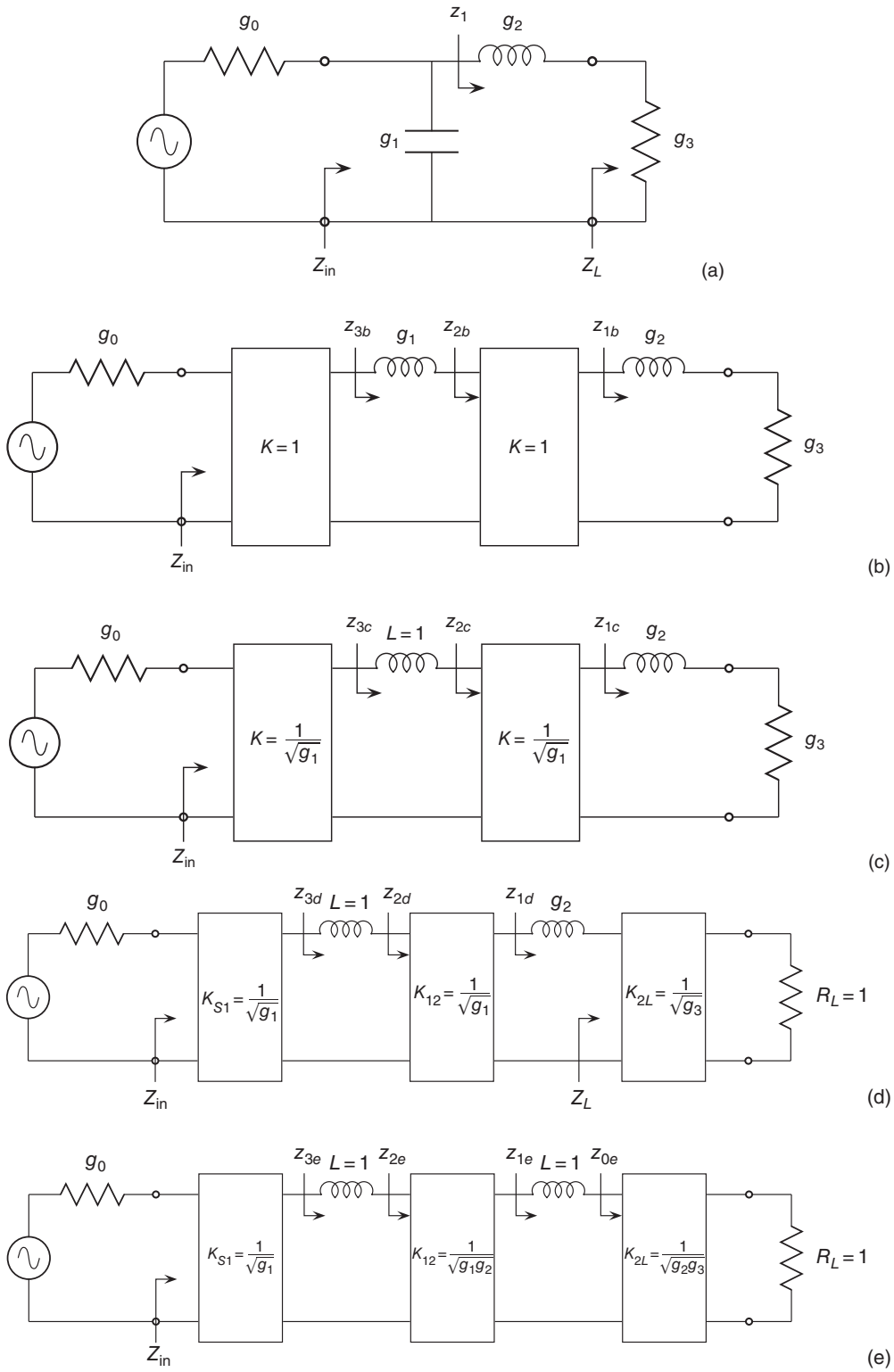


Figure E.3 (a–e) Steps to transform the lumped lowpass ladder network for the two-pole filter circuit into the lumped lowpass prototype with impedance inverters.

E.2 Normalization of the Element Values

In addition to making it possible for the filter network to be composed of only series elements (inductors) or parallel ones (capacitors), inverters can also be utilized to normalize the element values to be unity (or an arbitrary value) by choosing appropriate values for the characteristic impedance K.

Figure E.3c demonstrates the normalization of element value g_1 to unity. This is accomplished by choosing $K = 1/\sqrt{g_1}$ for inverters on each side of g_1. As before, the overall input admittance must be maintained:

$$z_{1c} = g_2 s + \frac{1}{g_3} \tag{E.5a}$$

$$z_{2c} = \frac{K^2}{z_{1c}} = \frac{1/g_1}{g_2 s + \dfrac{1}{g_3}} \tag{E.5b}$$

$$z_{3c} = s + z_{2c} = s + \frac{1/g_1}{g_2 s + \dfrac{1}{g_3}} \tag{E.5c}$$

$$Z_{\text{in}} = \frac{K^2}{z_{3c}} = \frac{1/g_1}{s + \dfrac{1/g_1}{g_2 s + \dfrac{1}{g_3}}} = \frac{1}{g_1 s + \dfrac{1}{g_2 s + \dfrac{1}{g_3}}} \tag{E.5d}$$

This Z_{in} represents the same value as given by (E.3), preserving the input impedance of the prototype filter.

Instead of choosing a normalization value of unity for g_1, let us consider an arbitrary value of inductance to be L_1. The value of K is then adjusted to be $\sqrt{L_1/g_1}$. Substituting this value of K in (E.5), we obtain

$$z_{1c} = g_2 s + \frac{1}{g_3} \tag{E.6a}$$

$$z_{2c} = \frac{K^2}{z_{1c}} = \frac{L_1/g_1}{g_2 s + \dfrac{1}{g_3}} \tag{E.6b}$$

$$z_{3c} = L_1 s + z_{2c} = L_1 s + \frac{L_1/g_1}{g_2 s + \dfrac{1}{g_3}} \tag{E.6c}$$

$$Z_{\text{in}} = \frac{K^2}{z_{3c}} = \frac{L_1/g_1}{L_1 s + \dfrac{L_1/g_1}{g_2 s + \dfrac{1}{g_3}}} = \frac{1}{g_1 s + \dfrac{1}{g_2 s + \dfrac{1}{g_3}}} \tag{E.6d}$$

Thus, we may choose an arbitrary element value in conjunction with a commensurate value of the characteristic impedance for the inverter and maintain the same input impedance as for the original prototype filter.

The next step is the normalization of g_2, constrained by the fact that it is the last element in the chain and terminates into the load impedance $1/g_3$. This implies that the inverter to be introduced after g_2 must ensure that the load impedance from its input port is preserved. This is done by choosing $K_{2L} = 1/\sqrt{g_3}$. Figure E.3d shows this configuration. The normalization of g_2 to unity can now be implemented by adjusting the values of impedance inverters on each side

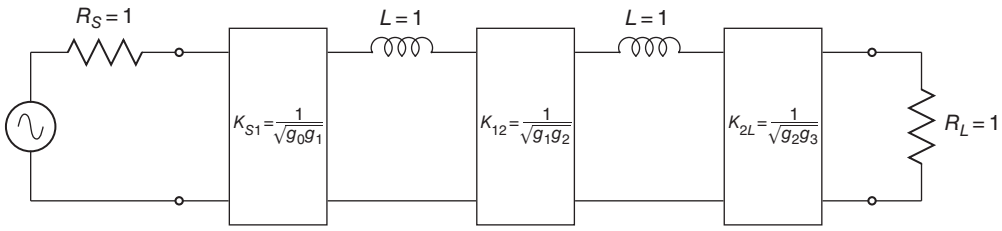


Figure E.4 Lumped lowpass ladder network for the two-pole filter circuit with impedance inverters after all elements are normalized to unity.

of g_2 to $K_{12} = 1/\sqrt{g_1 g_2}$ and $K_{2L} = 1/\sqrt{g_2 g_3}$, respectively, as shown in Figure E.3e. The input impedance is evaluated as follows:

$$z_{0e} = \frac{K_{2L}^2}{R_L} = \frac{1}{g_2 g_3} \tag{E.7a}$$

$$z_{1e} = s + z_{0e} = s + \frac{1}{g_2 g_3} \tag{E.7b}$$

$$z_{2e} = \frac{K_{12}^2}{z_{1e}} = \frac{\dfrac{1}{g_1 g_2}}{s + \dfrac{1}{g_2 g_3}} = \frac{1}{g_1 g_2 s + \dfrac{g_1}{g_3}} \tag{E.7c}$$

$$z_{3e} = s + z_{2e} = s + \frac{1}{g_1 g_2 s + \dfrac{g_1}{g_3}} \tag{E.7d}$$

$$Z_{\text{in}} = \frac{K_{S1}^2}{z_{3e}} = \frac{\dfrac{1}{g_1}}{s + \dfrac{1}{g_1 g_2 s + \dfrac{g_1}{g_3}}} = \frac{1}{g_1 s + \dfrac{1}{g_2 s + \dfrac{1}{g3}}} \tag{E.7e}$$

The last step is the normalization of the source impedance. This is done the same way as for the load impedance by adjusting the characteristic impedance of the first inverter to be $K_{S1} = \frac{1}{\sqrt{g_0 g_1}}$ and $R_S = 1$ as shown in Figure E.4. Note that the impedance from the output terminal of the first inverter looking toward the source in Figure E.4 is the same in all configurations described in Figure E.3.

E.3 General Lowpass Prototype Case

Following the procedure as outlined in (E.6), the K parameters for the general case with arbitrary inductances for the two-pole filter, as described in Figure E.5, are derived as

$$K_{S1} = \sqrt{\frac{R_S L_1}{g_0 g_1}}, \quad K_{12} = \sqrt{\frac{L_1 L_2}{g_1 g_2}}, \quad K_{2L} = \sqrt{\frac{R_L L_2}{g_2 g_3}} \tag{E.8}$$

If we make a further normalization by choosing the values of the inductances to be unity and the terminating impedances to be unity as well, the inverter values are given by

$$K_{S1} = \frac{1}{\sqrt{g_0 g_1}}, \quad K_{12} = \frac{1}{\sqrt{g_1 g_2}}, \quad K_{2L} = \frac{1}{\sqrt{g_2 g_3}} \tag{E.9}$$

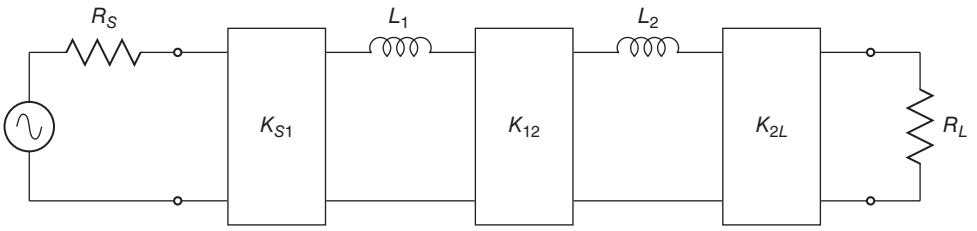


Figure E.5 Lowpass prototype with arbitrary values for impedances and inductances.

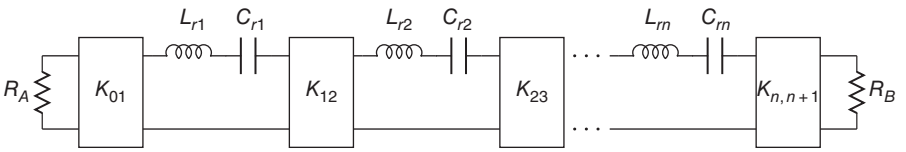


Figure E.6 Bandpass prototype using series resonators and impedance inverters.

By induction, or repeating the process, we can extend the aforementioned equations to the general case described in Figure E.1. The inverter value between two arbitrary inductances is given by

$$K_{i,i+1} = \sqrt{\frac{L_i L_{i+1}}{g_i g_{i+1}}}, \quad \text{for the general case}$$
$$K_{i,i+1} = \frac{1}{\sqrt{g_i g_{i+1}}}, \quad \text{when the inductance values are normalized to unity} \tag{E.10}$$

E.3.1 Coupling Coefficient: Lowpass Prototype

For lumped circuits, the coupling coefficient is defined by

$$k_{i,i+1} = \frac{K_{i,i+1}}{\sqrt{L_i L_{i+1}}} = \frac{1}{\sqrt{g_i g_{i+1}}} \tag{E.11}$$

The value of the coupling coefficient is determined by the g_i parameters that are fixed by the selection of the lowpass prototype filter. It implies that the values of $K_{i,i+1}$, L_i, and L_{i+1} may be chosen arbitrarily within the constraint of preserving the value of $k_{i,i+1}$. If the inductances are normalized to unity, $k_{i,i+1}$ equals $K_{i,i+1}$. From a physical point of view, K is analogous to a mutual reactance, and coupling coefficient represents the required energy transfer between the filter elements (Figure E.6).

E.4 Bandpass Prototype

The lumped bandpass prototype is derived using the lowpass to bandpass frequency transformation

$$\Omega = \frac{\omega_0}{\Delta\Omega}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \tag{E.12}$$

where Ω is the normalized frequency variable in the lowpass domain and ω is the bandpass variable.

The reactance of an inductance L_a in the lowpass and bandpass domain is related by

$$\Omega L_a = \frac{\omega_0}{\Delta\omega}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) L_a = \frac{L_a\omega}{\Delta\omega} - \frac{L_a\omega_0^2}{\omega\Delta\omega} = L\omega - \frac{1}{C\omega} \tag{E.13}$$

where

$$L = \frac{L_a}{\Delta\omega} \quad C = \frac{\Delta\omega}{L_a\omega_0^2}$$

Hence, an inductance in the lowpass prototype leads to a series resonator in the bandpass prototype with reactance

$$X(\omega) = L\omega - \frac{1}{C\omega}$$

E.4.1 Slope Parameter

The equivalence between lumped lowpass and bandpass prototype filters is established by the reactance slope parameter defined as

$$\mathcal{X} = \frac{\omega_0}{2}\frac{dX}{d\omega}\bigg|_{\omega_0} = L\omega_0 = \frac{1}{C\omega_0} \tag{E.14}$$

Therefore

$$L_a = \mathcal{X}\frac{\Delta\omega}{\omega_0} \tag{E.15}$$

The slope parameters provide a measure of reactance around the resonant frequency of the resonators for the bandpass prototype filter. Following (E.8), substituting values of L_a, the K values are computed as

$$K_{S1} = \sqrt{\frac{R_A\mathcal{X}_1\mathcal{W}}{g_0g_1}} \quad K_{12} = \mathcal{W}\sqrt{\frac{\mathcal{X}_1\mathcal{X}_2}{g_1g_2}} \quad K_{2L} = \sqrt{\frac{R_B\mathcal{X}_2\mathcal{W}}{g_2g_3}} \tag{E.16}$$

where the fractional bandwidth $\mathcal{W}$ is defined as

$$\mathcal{W} = \frac{\Delta\omega}{\omega_0} = \frac{\Delta f}{f_0}$$

Similar relationships can be derived for J-inverters.

E.4.2 Coupling Matrix Parameter *M*

For the bandpass prototype model used in the coupling matrix (CM) synthesis described in Chapter 8,

1) Resonators are synchronously tuned with $\omega_0 = 1$ rad/s.
2) Fractional bandwidth $\Delta\omega/\omega_0$ is normalized to unity.
3) Parameter M represents the mutual coupling or reactance between resonators.

In the corresponding lowpass domain (E.13), this implies that all the inductances (L_{a1}, L_{a2}) are normalized to unity.

Following (E.10), this normalization simply implies that M is given by the coupling coefficient as defined in (E.11). Coupling parameter M can therefore expressed as

$$\begin{aligned} M_{01} &= \frac{1}{\sqrt{g_0 g_1}} \\ M_{i,i+1} &= \frac{1}{\sqrt{g_i g_{i+1}}} \quad (i = 1, \ldots, N-1) \\ M_{N,N+1} &= \frac{1}{\sqrt{g_N g_{N+1}}} \end{aligned} \tag{E.17}$$

It is important to note that (E.17) is applicable for prototype ladder networks as described in Figure E.1. It implies that this relationship holds for Butterworth and Chebyshev filters for which analytic formulas are available for the g parameters, as well as other filter functions with monotonically increasing amplitude response for which g parameters can be readily derived using [$ABCD$] matrix approach of Chapter 7. For filters with transmission zeros or, for that matter, any general filter function, coupling parameters M are determined using polynomials of filter functions and CM synthesis techniques described in Chapters 3, 4, and 8–10.

E.4.3 Coupling Coefficient: Bandpass Prototype

Following (E.13), it can be seen that the slope parameter X for the bandpass prototype is analogous to self-inductance or the g parameters for the lowpass prototype filter. Making use of (E.13), (E.14)–(E.16), the coupling coefficient based on lumped resonators is given by

$$k_{i,i+1} = \frac{K_{i,i+1}}{\sqrt{\mathcal{X}_i \mathcal{X}_{i+1}}} = \mathcal{W} M_{i,i+1} \tag{E.18}$$

As expected, the coupling coefficient is related to the corresponding value in the lowpass domain via the fractional bandwidth of the bandpass prototype filter.

It is important to note that this coupling coefficient can be evaluated in terms of the physical dimensions of the coupling apertures. Cohn [1] and others developed approximate expressions for practical apertures based on Bethe's theory for small ideal (zero thickness) apertures. The aperture dimensions can also be evaluated experimentally as described in [2]. Such models can now be rendered much more accurate using EM techniques as described in Chapter 14.

Up until now, we have considered lumped elements in conjunction with ideal impedance inverters that are invariant with respect to frequency. Such inverters do not exist. The next step is to extend the prototype model based on lumped resonators to include transmission-line resonators.

E.4.4 Slope Parameter of Transmission-Line Resonators

From transmission-line theory, the input impedance of a lossless transmission line terminated in a load Z_T is given by

$$Z_{\text{in}} = Z_0 \frac{Z_T + jZ_0 \tan \beta l}{Z_0 + jZ_T \tan \beta l}$$

where Z_T is the terminating impedance, $\beta = 2\pi/\lambda = \omega/c$ is the phase constant, and l is the length of the transmission line.

For a series $\lambda/2$ resonator $Z_T = 0$, the input impedance is given by

$$Z_{\text{in}} = jZ_0 \tan \beta l = jX$$

Evaluating the slope parameter ($l = \lambda/2$ at $\omega = \omega_0$),

$$\mathcal{X} = \frac{\omega_0}{2}\frac{dX}{d\omega}\bigg|_{\omega_0} = \frac{\omega_0}{2}\frac{Z_0 l/c}{\cos^2\beta l}\bigg|_{\omega_0} = \frac{Z_0\pi}{2}$$

Substituting this value in (E.16), setting the values of $R_A = R_B = Z_0$, and extending it to the general case as in (E.10), the normalized inverter values are evaluated as

$$\begin{aligned}
\frac{K_{01}}{Z_0} &\approx \sqrt{\frac{\pi\mathcal{W}}{2g_0g_1}} \\
\frac{K_{i,i+1}}{Z_0} &\approx \frac{\pi\mathcal{W}}{2\sqrt{g_ig_{i+1}}} \quad (i = 1, \ldots, N-1) \\
\frac{K_{N,N+1}}{Z_0} &\approx \sqrt{\frac{\pi\mathcal{W}}{2g_Ng_{N+1}}}
\end{aligned} \tag{E.19}$$

E.4.5 Slope Parameter for Waveguide Resonators

For a half-wavelength resonator in a waveguide, the phase constant β is

$$\beta = \frac{2\pi}{\lambda_g}$$

where

$$\lambda_g = \frac{\lambda}{\sqrt{1-\left(\frac{f_c}{f}\right)^2}} = \frac{\frac{2\pi c}{\omega}}{\sqrt{1-\left(\frac{\omega_c}{\omega}\right)^2}} = \frac{2\pi c}{\sqrt{\omega^2-\omega_c^2}}$$

In the vicinity of the resonant frequency, the input impedance can be approximated as

$$Z_{\text{in}} = jZ_0\tan\beta l = jX$$

The slope parameter is given by

$$\mathcal{X} = \frac{\omega_0}{2}\frac{dX}{d\omega}\bigg|_{\omega_0} = \frac{\omega_0}{2}\frac{d}{d\omega}(Z_0\tan\beta l)\bigg|_{\omega_0} \approx \frac{Z_0\omega_0}{2}\frac{1}{\cos^2\beta l}2\pi l\frac{d}{d\omega}\left(\frac{1}{\lambda_g}\right)\bigg|_{\omega_0}$$

Since $l = \lambda_{g0}/2$, then $\cos^2\beta l = 1$ and

$$\mathcal{X} \approx \frac{\pi Z_0\omega_0}{2}\lambda_{g0}\frac{d}{d\omega}\left(\frac{1}{\lambda_g}\right)\bigg|_{\omega_0} = Z_0\frac{\pi}{2}\frac{\lambda_{g0}}{\lambda_0}\frac{d}{d\omega}\left(\sqrt{\omega^2-\omega_c^2}\right)\bigg|_{\omega_0} = Z_0\frac{\pi}{2}\left(\frac{\lambda_{g0}}{\lambda_0}\right)^2$$

Substituting this value of slope parameter in (E.16), the normalized inverter values are evaluated as

$$\begin{aligned}
\frac{K_{01}}{Z_0} &\approx \sqrt{\frac{\pi\mathcal{W}_\lambda}{2g_0g_1}} \\
\frac{K_{i,i+1}}{Z_0} &\approx \frac{\pi\mathcal{W}_\lambda}{2\sqrt{g_ig_{i+1}}} \quad (i = 1, \ldots, N-1) \\
\frac{K_{N,N+1}}{Z_0} &\approx \sqrt{\frac{\pi\mathcal{W}_\lambda}{2g_Ng_{N+1}}}
\end{aligned} \tag{E.20}$$

where

$$\mathcal{W}_\lambda = \frac{\lambda_{g1} - \lambda_{g2}}{\lambda_{g0}} \approx \mathcal{W}\left(\frac{\lambda_{g0}}{\lambda_0}\right)^2$$

is the guide-wavelength fractional bandwidth [1, 2].

Making use of (E.17), the inverter values can be expressed in the following form:

$$\frac{K_{01}}{Z_0} \approx M_{01}\sqrt{\frac{\pi\mathcal{W}}{2}} \quad \frac{K_{i,i+1}}{Z_0} \approx M_{i,i+1}\frac{\pi\mathcal{W}}{2} \quad \frac{K_{N,N+1}}{Z_0} \approx M_{N,N+1}\sqrt{\frac{\pi\mathcal{W}}{2}} \quad \text{(TEM)}$$

$$\frac{K_{01}}{Z_0} \approx M_{01}\sqrt{\frac{\pi\mathcal{W}_\lambda}{2}} \quad \frac{K_{i,i+1}}{Z_0} \approx M_{i,i+1}\frac{\pi\mathcal{W}_\lambda}{2} \quad \frac{K_{N,N+1}}{Z_0} \approx M_{N,N+1}\sqrt{\frac{\pi\mathcal{W}_\lambda}{2}} \quad \text{(TE)} \qquad \text{(E.21)}$$

E.4.6 Practical Impedance and Admittance Inverters

The impedance and admittance inverters used in the analysis are assumed to be ideal, exhibiting frequency independent characteristics. Such inverters do not exist. The simplest inverter is a quarter wavelength of transmission line. Needless to say, its inverting properties are suitable over very narrow bandwidths, and, in addition, using quarter-wave lines would make the filter structure very large. The more practical inverters consisting of certain types of discontinuities are well described in [1–3]. Such inverters serve the dual function of impedance inversion over wider bandwidths while providing a structure for realizing resonators, resulting in compact implementation of practical microwave filters.

References

1 Cohn, S.B. (1957) Direct-coupled-resonator filters. *Proceedings of the IRE*, **45**, 187–196.

2 Matthaei, G.L., Young, L., and Jones, E.M.T. (1980) *Microwave Filters, Impedance Matching Networks, and Coupling Structures*, Artech House, New Jersey.

3 Collin, R.E. (2000) *Foundations for Microwave Engineering*, 2nd edn, Wiley-IEEE Press, New York.

Index

a

[*ABCD*] matrix:
  buildup for third-degree network, circuit synthesis approach 217–218
  matrix representation, multi port microwave network analysis 149–150
[*ABCD*] polynomial synthesis:
  circuits with inverters, lowpass prototype circuits 222–227
  singly terminated filter prototype, lowpass prototype circuits 227–229
Achieser–Zolotarev functions, lowpass prototype transfer polynomials 400–401
Adjacent copolarized channel interference, communication systems noise 16
Adjacent cross-polarized channel interference, communication systems noise 17
Admittance function $[Y_N]$:
  circuit approach, transversal coupling matrix, $N+2$ coupling matrix 270–271
  transfer and reflection polynomials, transversal coupling matrix, $N+2$ coupling matrix 269–270
Admittance $[Y]$ matrices, impedance $[Z]$ and, multiport microwave network analysis 147–149
Advanced modulation schemes, digital transmission, noise 45–49
Aggressive space mapping (ASM), electromagnetic (EM) simulation technique 500–504
Air, high-power breakdown phenomena 643–645
Allpass phase and group delay equalizer networks 541–564
  characteristics of 541–543
  dissipation effect 562–563
    distributed networks 550
    lumped-element first-order all pass equalizer 542, 550
    second-order lumped equalizer 542, 550
  equalization tradeoffs 563
  lumped-element
    generally 543–544
    network realizations 545–547
    resistively terminated symmetric lattice networks 544–545
  microwave allpass networks 547–550
  optimization criteria 557–562
  overview 541
  physical realization 550–552
  reflection-type allpass networks:
    practical narrowband 554–557
      C-section in waveguide structure 555–556
      D-section in waveguide structure 556
      generally 554–555
      narrowband TEM reactance networks 556–557
    synthesis 553–554
All-pole prototype filter functions, characteristic polynomials response shapes compared 97
Alternating pole method, denominator polynomial $E_{(s)}$ determination 186–188
Amplifiers, high-power considerations 641
Amplitude modulation, noise 32–34
Amplitude optimization, single-and dual-mode resonator filters 429

Microwave Filters for Communication Systems: Fundamentals, Design, and Applications, Second Edition.
Richard J. Cameron, Chandra M. Kudsia, and Raafat R. Mansour.
© 2018 John Wiley & Sons, Inc. Published 2018 by John Wiley & Sons, Inc.

Amplitude response:
  lossless lowpass prototype filter function circuit configurations 111–113
  lowpass prototype characteristics 95
Amplitude shift keying (ASK), digital modulation-demodulation schemes 44
AM systems, FM systems compared, modulation-demodulation schemes 37–39
Analog angle-modulated signals, noise 35–37
Analog systems, systems classification 77
Angle-modulated signals, noise 35–37
Annihilation of matrix elements, similarity transformation and, coupling matrix synthesis 261–268
Antennas, communication channel and link budgets 11–15
Arbitrarily shaped resonators:
  microwave resonators, resonant frequency calculation 380–383
  unloaded *Q* factor, microwave resonators 386
Arbitrary response, asymmetric filters, computer-aided characteristic polynomial synthesis 140–142
Arrow coupling matrix, extracted pole and trisection elements, trisections 350–355
Asymmetric (4-2) filter network, network-circuit approach synthesis example 235–244
Asymmetric filters, computer-aided characteristic polynomial synthesis 138–142
  arbitrary response 140–142
  Chebyhev passband 140, 141
  generally 138–139
Asymmetric realizations, for symmetric characteristics, folded coupling matrix reconfiguration 300–301
Asymmetric response filters, lossless lowpass prototype filter functions 115–118
Asynchronous resonators/asynchronously tuned filters 787
Atmosphere, high-power breakdown phenomen 643–645
Attachment, high-power breakdown phenomena 643
Attenuation zeros, characteristic polynomials 95

b

Bandpass filter(s) (high-power considerations) 658–666
  FEM simulation 663–664
  filter prototype network 650
  generally 658–659
  lumped to distributed scaling 650
  multiplexer 654–655
  resonator voltages from prototype network 662–663
  thermal dissipation limitations 659–660
  voltage breakdown 660
Bandpass filter group delay, lossless lowpass prototype filter function circuit configurations 113
Bandpass prototype:
  coupling matrix 247
  lowpass prototype and, coupling matrix synthesis 248–250
Bandstop filter synthesis (extracted pole and trisection elements) 335–343
  direct-coupled 338–343
  generally 335–338
Bandwidth, baseband signal formation, noise, modulation-demodulation schemes 34–35
Bandwidth efficiency, advanced modulation schemes, digital transmission 46–49
Bandwidth signal, information destination 6
Barium strontium titanate (BST) 732, 733, 744
Baseband signal formation, noise 34–35
Baseband signal, information source 3
Binary transmission, error rates in, digital transmission noise 42–44
Boolean logic, fuzzy logic compared computer-aided diagnosis and tuning 632–634
Box section:
  extended, extracted pole and trisection elements 364–370
  extracted pole and trisection elements 361–370

C
Calibrated coarse model (CCM), electromagnetic (EM) simulation technique 504–510
Carrier-to-noise (C/N) ratio, RF link design, communication systems noise 29–30
Cascade, of two networks, multiport microwave network analysis 160–167
Cascaded networks:
  communication systems noise 23–25
  n-tuplet synthesis, extracted pole and trisection elements, trisections 356–361
Cascaded quartets (CQs):
  dual-mode filters, coupling matrix configurations 446, 448
  folded coupling matrix reconfiguration 304–306
Cascade trisections, extracted pole and trisection elements 350–356
Cauchy technique, multidimensional, EM-based, optimization techniques 495
Cavity configuration calculation, dielectric resonator filter 470
Cellular system design, radio frequency (RF) filter networks 62–66
Channel imperfection distortion, communication systems noise 26–29
Channelizer (input multiplexer, IMUX) section, satellite system design 53–56
Channel transfer analysis, common port return loss and, manifold multiplexer 585–588
Channel transfer characteristics, manifold multiplexer 588–589
Characteristic polynomials, lossless low pass prototype filter functions 93–95. *See also* Computer-aided characteristic polynomial synthesis; Polynomials
Chebyshev passband, asymmetric filters, computer-aided characteristic polynomial synthesis 140, 141
Chebyshev approximation filters:
  compared 107–108
  lossless lowpass prototype filter functions 99–101
  optimization criteria for 134, 135
Chebyshev bandpass filter circuit models, coupled resonator filters 459–463
Chebyshev filter functions (general polynomial synthesis method) 189–200
  polynomial forms for symmetric and asymmetric functions 196–198
  polynomial synthesis 189–193
  recursive technique 193–196
Chebyshev polynomials, of the second kind, lowpass prototype transfer polynomials 397–399
Circuit model parameter extraction computer-aided diagnosis and tuning 615–619
Circuit synthesis approach:
  extracted pole and trisection elements, trisections 344–349
  network-circuit approach synthesis 216–221
    [*ABCD*] matrix buildup for third-degree network 217–218
    generally 216–217
    network synthesis 218–221
Circuit theory approximation 75–86
  assumptions in 84–85
  frequency-domain exponential driving function network equation 80–83
  historical perspective 77–78
  linear systems 75–76
  steady-state response to sinusoidal excitations 83–84
  systems classification 76–77
  time domain network equation 78–80
Circular cavities, unloaded *Q* factor, microwave resonators 385–386
Circulator-coupled multiplexing networks (MUX) 567
  RF channel combiners 578
  RF channelizers (demultiplexers) 573–574
Classical filters. *See* Lossless lowpass prototype filter functions, classical filters
Coaxial low pass filters. *See* Waveguide and coaxial lowpass filters
Coaxial transmission lines, waveguides and, high-power breakdown phenomena, power ratings 645–646
Coefficient of thermal expansion 784, 786
Combline filter 806–814

Commensurate-line building elements, waveguide and coaxial lowpass filters 395–396
Common port return loss:
  channel transfer analysis and, manifold multiplexer 585–588
  computation of, manifold multiplexer 588
Communication, defined 1
Communication channel:
  communication system model 5
  link budgets and, RF filter networks 10–15
    antennas 11–15
    signal power 10–11
Communication link location, RF filter specifications 70
Communication systems noise 15–32. *See also* Noise; Radio frequency (RF) filter networks
  adjacent copolarized channel interference 16
  adjacent cross-polarized channel interference 16
  cascaded networks 23–25
  channel imperfection distortion 26–29
  intermodulation (IM) noise 25–26
  multipath interference 17
  RF link design 29–32
  thermal 17–23
Complex frequency variable 80–81
Complex transmission zeros, extracted pole and trisection elements 355–356
Compound resonator 718, 720–727
Computer-aided characteristic polynomial synthesis 129–146
  asymmetric filters 138–142
    arbitrary response 140–142
    Chebyhev passband 140, 141
    generally 138–139
  critical frequencies 143–144
    unconventional 8-pole filter 145–146
  linear phase filters 142–143
  novel classes of 136–138
  objective function:
    analytic gradients of 131–134
      equality constraint gradient 133–134
      generally 131
      inequality constraint gradient 133
      unconstrained 132
    symmetric lowpass prototype filter networks 129–131
  optimization criteria 134–136
  overview 129
Computer-aided diagnosis and tuning 609–640
  automated systems 637–639
  circuit model parameter extraction 615–619
  coupled resonator filters, sequential tuning 610–615
  fuzzy logic tuning 627–637
    applications 634–637
    Boolean logic compared 632–634
    generally 627–628
    systems description 628–629
    systems setup 629–632
  input reflection coefficient, poles and zeros of 619–622
  overview 609–610
  time-domain tuning 622–627
    generally 622–623
    golden filter response 627
    interresonator coupling 624–627
    resonator frequencies 623–624
Conductivities, of metals 859
Constant absolute bandwidth 734–744
Constant bandwidth 733, 737–744
Constant parameter systems, systems classification 76
Constants:
  dielectric, of materials 861
  physical 857
Continuous exponentials, signal representation by 82
Conventional transmission-line resonators, resonant frequency calculation 376–378
Corona 783
Coupled resonator filters 457–484
  Chebyshev bandpass filter circuit models 459–463
  computer-aided diagnosis and tuning sequential tuning 610–615
  dielectric resonator filter example 468–475
    cavity configuration calculation 470
    generally 468–470
    input/output calculation 472–475
    iris dimensions calculation 471–472

input/output coupling calculation 467–468
frequency domain method 467
group delay method 468
interresonator coupling calculation 471–473
electric wall and magnetic wall symmetry 463–465
S parameters 465–466
microstrip filter using *J*-admittance inverter model example 478–483
overview 457–458
waveguide iris filter using impedance inverter model example 475–478
Coupling matrix:
dual-mode filter configurations 444–447
cascade quartet 446, 448
extended box 447, 448
folded array 444–445
Pfitzenmaier configuration 445–446
propagating forms 446, 447
extracted pole and trisection elements, trisections 350–356
single-and dual-mode resonator filters 427–428 (*See also* Single-and dual-mode resonator filters)
Coupling matrix synthesis 247–293
bandpass prototype and lowpass prototypes 248–250
development of 247
direct analysis 257–258
direct synthesis 258–261
even-mode 282–292
lowpass prototype circuit elements 253–255
network analysis 255–257
$N + 2$ coupling matrix 268–282
example 276–278
transversal coupling matrix 269–275
$N \times N$ coupling matrix formation 250–253
odd-mode 282–292
reduction 261–268
Coupling sign correction, single-and dual-mode resonator filters 442–444
Critical pressure, high-power breakdown phenomena 644–645
Cryogenic dielectric resonator filters 534–536
C-section allpass equalizer, in waveguide structure 555–556
Cul-de-sac configuration (folded coupling matrix reconfiguration) 311–321
additional forms 312–319
generally 311–312
sensitivity considerations 319–321

d

Decoder, receiver 5
Demodulator, receiver 5
Demultiplexers. *See* RF channelizers (demultiplexers)
Denominator polynomial $E_{(s)}$ determination, alternating pole method 186–188
Derating factors, high-power breakdown phenomena 646
Diagonal cross-couplings 284–285
Dielectric constants, of materials 859
Dielectric resonator filter(s) 517–543
configurations:
novel 531–534
standard 524–528
coupled resonator filters 468–475
cavity configuration calculation 470
generally 468–470
input/output calculation 472–475
iris dimensions calculation 471–472
cryogenic 534–536
design considerations 528–531
power handling capability 531
Q value 528–529
spurious performance 529–530
temperature drift 530
hybrid dielectric/superconductor filters 536–538
overview 517
resonant frequency calculation 517–521
rigorous analyses 521–524
Diffusion, high-power breakdown phenomena 643
Digital systems, systems classification 77
Digital transmission noise 39–50. *See also* Noise
advanced schemes 45–49
error rates in binary transmission 42–44
generally 39–40
modulation and demodulation schemes 44–45
pulse-code-modulated (PCM) systems 41
quality of service and S/N ratio 50

Digital transmission noise (*contd.*)
quantization 41
quantization noise
pulse-code-modulated (PCM) systems 41–42
systems 41–42
sampling 40
Diplexers, transmit-receive, multiplexing networks 597–602
Direct-coupled bandstop filter, extracted pole and trisection elements 338–343
Direct-coupled extracted pole filter:
extracted pole and trisection elements 333–335
single-and dual-mode resonator filters 447–450
Directional filter combiner, RF channel combiners 582
Directional filter multiplexing networks configurations 567, 569
Dissipation effect:
lossless lowpass prototype filter circuit configurations, equivalent * for bandpass and bandstop filters 115
lossless lowpass prototype filter function circuit configurations 113–115
dissipation factor * and quality factor Q_0 relationship 114–115
generally 113–114
Distortionless transmission, ideal filter 87–88
Distributed networks:
all pass phase networks, dissipation effect 550
bandpass filters, high-power considerations 662
Distributed stepped impedance (SI) lowpass filter 395, 401–410
generally 401
prototype circuit synthesis 404–405
realization 405–410
transfer function S_{21}, 401–403
Doppler shift, radio propagation 6, 7
Downconverter, receiver 5
Driver amplifier (DAMP), satellite system design 53
Driving function, frequency-domain exponential network equation 80–83
D-section allpass equalizer, in waveguide structure, reflection-type allpass networks 556
Dual band filter 680–683, 686, 689, 691, 694–696, 698, 703–705, 707–710, 712–718
Dual mode filter 526, 537, 540
Dual mode resonator 540, 542
Dual-band 680–683, 685–699, 703–710, 712–718
Dual-band bandpass filter transformation, network filter synthesis 208–210
Dual-mode dielectric resonator 539–542
Dual-mode resonator filters. *See* Single-and dual-mode resonator filters
Dynamic and instantaneous systems, systems classification 77

e

Effective aperture, antenna gain, communication channel and link budgets 12–13
Effective input thermal noise, communication systems 20–21
Eigenmode analysis, arbitrarily shaped resonators 380–381
Eighth-degree filter, symmetric realizations for dual-mode filters 298–299
Electrical networks, transfer function of, frequency-domain exponential driving function network equation 82–83
Electric wall symmetry, magnetic wall symmetry and 463–465
Electromagnetic (EM) simulation technique 485–514
advanced design techniques 496–513
calibrated coarse model (CCM) 504–510
generalized calibrated coarse model (CCM) 510–513
generally 496
space mapping 496–504
optimization techniques 486–496
EM-based fuzzy logic 495–496
EM-based multidimensional Cauchy technique 495
EM-based neural network models 491–495

EM simulator 487–488
EM simulator with adaptive frequency sampling 490–491
generally 486–487
semi-EM simulator 488–490
overview 485
synthesis techniques 485
Electromagnetic (EM) simulator, with adaptive frequency sampling 490–491
Elliptic function filters:
compared 107–108
lossless lowpass prototype filter functions 101–104
optimization criteria for 135–136
Elliptical cavities 686, 689, 703
EM simulation. *See* Electromagnetic (EM) simulation technique
Encoder, transmitter 3
En passant distortion:
RF channel combiners 580–582
RF channelizers (demultiplexers) 574–577
Equalization tradeoffs, allpass phase and group delay equalizer networks 563
Equiripple passbands, stopbands and, computer-aided characteristic polynomial synthesis 137
Equivalent Linear Frequency Drift (ELFD) 784–786
Error rates, in binary transmission, digital transmission noise 42–44
European Cooperation Standardization Multipactor Standard (ECCS) 838
Even-mode coupling matrix synthesis technique 282–292
Even-mode coupling submatrices, odd-mode coupling submatrices and, folded coupling matrix reconfiguration 310
Even-order elliptic function filters, lossless lowpass prototype filter functions 105–106
Extended box:
dual-mode filters, coupling matrix configurations 447, 448
extracted pole and trisection elements 364–370
Extracted pole filters, phase-and direct-coupled, single-and dual-mode resonator filters 447–450
Extracted pole and trisection elements 323–371
bandstop filter synthesis 335–343
direct-coupled 338–343
generally 335–338
box section 361–370
extended box section 364–370
extracted pole filter synthesis 323–335
analysis 330–332
direct-coupled 333–335
example 327–330
generally 323–324
pole element synthesis 324–327
overview 323
trisections 343–361
circuit approach 344–349
coupling matrix approach 350–356
generally 343–344
synthesis of advanced circuits approach 356–361
Extracted zero/extracted zero pair 713–715
Extraction process, parallel-coupled inverters, ladder network synthesis 229–235

f

Faraday, Michael 77
Federal Communications Commission (FCC) 8
FEM simulation, bandpass filters high-power considerations 663–664
Ferroelectric 732
Ferromagnetics 732
Filter networks. *See* Network filter synthesis
Filter prototype network, bandpass filters, high-power considerations 660–662
Finite transmission zeros, characteristic polynomials response shapes compared 97–98
First-order lattice section, lumped-element allpass networks 545–547
FM systems, AM systems compared, modulation-demodulation schemes, noise 37–39
Folded array, dual-mode filters, coupling matrix configurations 444–445
Folded canonical form, $N + 2$ transversal matrix, circuit approach, transversal coupling matrix, $N + 2$ coupling matrix 275

Folded coupling matrix reconfiguration 295–321
asymmetric realizations for symmetric characteristics 300–301
cascaded quartets (CQs) 304–306
cul-de-sac configuration 311–321
additional forms 312–319
generally 311–312
sensitivity considerations 319–321
overview 295
parallel-connected two-port networks 306–310
even-mode and odd-mode coupling submatrices 310
generally 306–309
"Pfitzenmaier" configurations 301–303
symmetric realizations for dual-mode filters 295–300
eighth-degree filter 298–299
generally 295–297
sixth-degree filter 297–298
tenth-degree filter 298–299
twelth-degree filter 299–300
Frequency-domain exponential driving function network equation, circuit theory approximation 80–83
Frequency domain method, input/output coupling calculation, coupled resonator filters 467
Frequency plan, RF filter specifications 68–69
Frequency shift keying (FSK), digital modulation-demodulation schemes 44–45
"Full inductive" filter dual-mode filters 450–453
Fully canonical filter functions, circuit approach, transversal coupling matrix, $N+2$ coupling matrix 271–272
Fuzzy logic, EM-based, electromagnetic (EM) simulation optimization 495–496
Fuzzy logic tuning, computer-aided diagnosis and tuning 627–632
applications 634–637
Boolean logic compared 632–634
generally 627–628
systems description 628–629
systems setup 629–632
$F \times d$ product dependence, multipaction breakdown, high-power considerations 650

g

Gain, antenna, communication channel and link budgets 12–13
Gaseous breakdown, high-power breakdown 645
Golden filter response, time-domain tuning 627
Group delay, lossless lowpass prototype filter function circuit configurations 113
Group delay equalizer networks. *See* Allpass phase and group delay equalizer networks
Group delay method, input/output coupling calculation, coupled resonator filters 468
Group delay optimization, single-and dual-mode resonator filters 431–434
Group delay response, lossless lowpass prototype filter function circuit configurations 113

h

Half-cut dielectric resonator 539–542
Half-cut resonator 696, 698
Harmonic rejection filter, transmitter section architecture 60, 61
High-power amplifier (HPA):
RF filter specifications 69
satellite system design 56–58
High-power considerations 639–671
amplifiers 641
bandpass filters 658–666
FEM simulation 663–664
filter prototype network 650
generally 658–659
lumped to distributed scaling 650
multiplexer 654–655
resonator voltages from prototype network 662–663
thermal dissipation limitations 659–660
voltage breakdown 660
breakdown phenomena 658–666
atmosphere 643–645
attachment 643
critical pressure 644–645
derating factors 646
diffusion 643
gaseous 642
generally 641

mean free path 642
thermal dissipation impact 646–647
waveguides and coaxial transmission lines, power ratings 645–646
multipaction breakdown 647–648
design margins 656–658
detection and prevention strategies 655–656
$f \times d$ product dependence 650
generally 641
power level criteria 653–655
RF voltage dependence 650
surface conditions 649–650
vacuum environment 648–649
overview 639
passive intermodulation (PIM) 666–670
control guidelines 666–670
generally 666–668
measurement 668–669
requirements 639–641
High-power output circuits, passive intermodulation (PIM) requirements, transmitter section architecture 60–62
Homogeneous fields 652, 666, 843, 847
Hybrid branching network, multiplexing networks, demultiplexers 572–573
Hybrid-coupled filter combiner module (HCFM) multiplexer, RF channel combiners, multiplexing networks 578–582
Hybrid coupled multiplexing networks 567
Hybrid dielectric/superconductor resonator filters 536–538

i

Ideal filter, lossless lowpass prototype filter functions 87–88
Impedance inverter(s):
realization if, distributed stepped impedance lowpass filter 408–410
waveguide iris filter, coupled resonator filters 475–478
Impedance [Z], admittance [Y] matrices and, multiport microwave network analysis 147–149
Information concept, radio frequency (RF) filter networks 8–10
Information destination, communication system model 5–6
Information source, communication system model 3
Input/output calculation:
coupled resonator filters 467–468
frequency domain method 467
group delay method 468
dielectric resonator filter, coupled resonator filters 472–475
Input reflection coefficient, poles and zeros of, computer-aided diagnosis and tuning 619–622
Input switch matrices (ISM), satellite system design 52
Instantaneous and dynamic systems, systems classification 77
Institute of Radio Engineers (IRE) 21
Interdigital filter 841–845
Interference environment, RF filter specifications 69
Intermodulation (IM) distortion and noise:
communication systems noise 25–27
summary table 74
International Radio Consultative Committee (CCIR) 8
International Telecommunications Union (ITU) 8
Inter-resonator coupling 680, 681, 685, 691, 693, 696, 699, 717
Interresonator coupling, time-domain tuning, computer-aided diagnosis and tuning 624–627
Interresonator coupling calculation (coupled resonator filters) 471–473
electric wall and magnetic wall symmetry 463–465
S parameters 465–466
Inverters, [$ABCD$] polynomial synthesis, lowpass prototype circuits for coupled-resonator microwave bandpass filters 222–227
Iris dimensions calculation, dielectric resonator filter, coupled resonator filters 471–472

j

J-admittance inverter model, microstrip filter, coupled resonator filters 478–483

k

Kelvin 701
Kinetic inductance 701

l

Ladder network synthesis, network-circuit approach synthesis 229–235
Linear display of reflection coefficient, unloaded *Q* factor measurement 392–393
Linear phase filters:
  computer-aided characteristic polynomial synthesis 142–143
  lossless lowpass prototype filter functions 107
Linear systems (circuit theory approximation) 75–76
  frequency-domain exponential driving function network equation 80–83
  steady-state response to sinusoidal excitations 83–84
  time domain network equation 78–80
Linearity 734
LNA, receiver 5
Loaded *Q* factor measurement, unloaded *Q* factor measurement and 387–393
Lossless lowpass prototype filter functions 87–127. *See also* Computer-aided characteristic polynomial synthesis
  asymmetric response filters 115–118
  characteristic polynomials 93–95
    response shapes compared 96–98
  circuit configurations 109–113
    frequency response 111–113
    generally 109
    scaling 110–111
  classical filters 98–108
    Chebyshev approximation 99–101
    compared 107–108
    elliptic function filters 101–104
    even-order elliptic function filters 105–106
    linear phase filters 107
    maximally flat filters 98–99
    odd-order elliptic function filters 104–105
    transmission zeros and maximally flat passband filters 107
  dissipation effect 113–115
    dissipation factor * and quality factor Q_o relationship 114–115
    equivalent * for bandpass and bandstop filters 115
    generally 113–114
  ideal filter 87–88
  lowpass characteristics 95–96
  polynomial functions for doubly terminated 88–92
    generally 88–89
    normalization 92–93
    reflection and transmission coefficients 90–92
  unified design charts (UDC) 121–127
  unified design chart (UDC) relationships 108–109
Loss tangents, of materials 861
Lossy element noise, thermal noise 22
Lowpass prototype circuits. *See also* Lossless lowpass prototype filter functions
  bandpass prototype and, coupling matrix synthesis 248–250
  for coupled-resonator microwave bandpass filters 221–229
    [*ABCD*] polynomial synthesis for circuits with inverters 222–227
    [*ABCD*] polynomial synthesis for singly terminated filter prototype 227–229
    generally 221–222
  coupling matrix synthesis 253–255
Lowpass prototype transfer polynomials (waveguide and coaxial lowpass filters) 396–401
  Achieser–Zolotarev functions 400–401
  Chebyshev polynomials of the second kind 397–399
Lowpass resonators, extracted pole and trisection elements, trisections 343
Lumped distribution systems:
  bandpass filters, high-power considerations 662
  systems classification 76–77
Lumped-element allpass networks 543–547. *See also* Allpass phase and group delay equalizer networks
  generally 543–544
  network realizations 545–547

resistively terminated symmetric lattice networks 544–545
Lumped-element first-order all pass equalizer, dissipation effect 542, 550, 620

m

Magnetic wall symmetry, electric wall symmetry and, coupled resonator filters 463–465
Manifold-coupled multiplexer 583–584
Manifold-coupled multiplexing networks 570–571
Manifold multiplexer 582–592
  channel filter design 589–590
  channel transfer characteristics 588–589
  common port return loss and channel transfer analysis 585–588
  common port return loss computation 588
  design methodology 584–585
  optimization processes 590–593
  star junction 583
Manufacturing tolerances 784–789, 792, 793, 835, 849–851
Materials:
  dielectric constants and loss tangents of 861
  surface conditions, multipaction breakdown 655–656
Matrix representation (multiport microwave network analysis) 147–160. *See also* specific matrices
  [*ABCD*] matrix 149–150
  impedance [*Z*] and admittance [*Y*] matrices 147–149
  scattering [*S*] matrix 150–155
  transmission [*T*] matrix 155–157
  two-port network analysis 157–160
Maximally flat filters:
  compared 107–108
  lossless lowpass prototype filter functions 98–99
  passband filters, transmission zeros and, lossless lowpass prototype filter functions 107
Maximum power transfer, distortionless transmission 88
Maxwell, James Clerk 77
Mean free path, high-power breakdown 642–643
Mechanically tunable 755, 773
MEMS switches 745, 746, 767–774
Metals, conductivities of 859
Micro-electro-mechanical system (MEMS) 732–734, 744, 746, 759, 761–764, 778
Microstrip filter, *J*-admittance inverter model, coupled resonator filters 478–483
Microwave filter technology, limitations of 70–71
Microwave resonators 373–393
  configurations 373–376
  loaded and unloaded *Q* factor measurement 387–393
  overview 373
  resonant frequency calculation 376–383
    arbitrarily shaped resonators 380–383
    conventional transmission-line resonators 376–378
    transverse resonance technique 378–380
  unloaded *Q* factor 383–386
Misalignment tolerances 784, 786
Mixed lumped distributed lowpass filter 395, 411–425. *See also* Waveguide and coaxial lowpass filters
  generally 411, 413
  realization 418–425
  tapered-corrugated lowpass prototype circuit 415–418
  transfer and reflection polynomial formation 413–415
Mobile telephone switching office (MTSO), cellular system design 63
Modulation-demodulation schemes:
  digital transmission, noise 44–49
  noise 32–39
    amplitude modulation 32–34
    angle-modulated signals 35–37
    baseband signal formation 34–35
    FM and AM systems compared 37–39
Modulation schemes, RF filter specifications 69
Modulator, transmitter 3
Multi-band 727
Multiband domain 719, 720
Multiband filters 679–727
Multidimensional Cauchy technique, EM simulation optimization techniques 495
Multi-mode resonators 679, 683–700

Multi-notch filter 681, 701
Multipaction breakdown (high-power considerations) 647–658
  design margins 656–658
  detection and prevention strategies 655–656
  $f \times d$ product dependence 650
  generally 641
  power level criteria 653–655
  RF voltage dependence 650
  surface conditions 649–650
  vacuum environment 648–649
Multipaction breakdown/single carrier, multicarrier, single surface, two surface 647–658
Multipath interference, communication systems noise 17
Multiple RF link in tandem, communication systems noise 30–32
Multiplexer, bandpass filters, high-power considerations 664–665
Multiplexing networks 565–603
  background of 565–567
  configurations
    circulator-coupled approach 567
    directional filter approach 567, 569, 570
    hybrid coupled approach 567, 569
    manifold-coupled approach 570–571
  RF channel combiners 577–597
    circulator-coupled MUX 578
    directional filter combiner 582
    generally 577–578
    hybrid-coupled filter combiner module (HCFM) multiplexer 578–582
    manifold multiplexer 582–597
      channel filter design 589–590
      channel transfer characteristics 588–589
      common port return loss computation 588
      common port return loss and channel transfer analysis 585–588
      design methodology 584–585
      manifold-coupled 583–584
      optimization processes 590–597
      star junction 583
  RF channelizers (demultiplexers) 571–577
    circulator-coupled MUX 573–574
    en passant distortion 574–577
    generally 571–573
    hybrid branching network 572–573
  transmit-receive diplexers 597–602
Multiport microwave network analysis 147–175
  cascade of two networks 160–167
  matrix representation 147–160
    [*ABCD*] matrix 149–150
    impedance [Z] and admittance [*Y*] matrices 147–149
    scattering [*S*] matrix 150–155
    transmission [*T*] matrix 155–157
    two-port network analysis 157–160
  multiport networks 167–169
  multiport networks analysis 169–174
  overview 147
Multiport networks:
  analysis of, multiport microwave network analysis 169–174
  multiport microwave network analysis 167–169

n

Narrowband FM, analog angle-modulated signals, noise 35–36
Narrowband TEM reactance networks, reflection-type allpass networks 556–557
Network-circuit approach synthesis 215–245
  circuit synthesis approach 216–221
    [*ABCD*] matrix buildup for third-degree network 217–218
    generally 216–217
    network synthesis 218–221
  example, asymmetric (4-2) filter network 235–244
  ladder network synthesis 229–235
  lowpass prototype circuits for coupled-resonator microwave bandpass filters 221–229
    [*ABCD*] polynomial synthesis for circuits with inverters 222–227
    [*ABCD*] polynomial synthesis for singly terminated filter prototype 227–229
    generally 221–222
  overview 215–216
Network filter synthesis 177–213

Chebyshev filter functions, general polynomial synthesis method 189–200
polynomial forms for symmetric and asymmetric functions 196–198
polynomial synthesis 189–193
recursive technique 193–196
circuit synthesis approach, network-circuit approach synthesis 218–221
denominator polynomial $E_{(s)}$ determination, alternating pole method 186–188
dual-band bandpass filter transformation 208–210
polynomial forms of transfer and reflection parameters $S_{21}(s)$ and $S_{11}(s)$ for two-port network 177–186
predistorted filter characteristics 200–208
generally 200–205
synthesis 205–208
Neural network models, electromagnetic (EM) simulation technique, optimization techniques 491–495
Noise 15–50. *See also* Communication systems noise; Digital transmission noise
adjacent copolarized channel interference 16
adjacent cross-polarized channel interference 16
cascaded networks 23–25
channel imperfection distortion 26–29
defined 15
digital transmission 39–50
advanced schemes 45–49
error rates in binary transmission 42–44
generally 39–40
modulation and demodulation schemes 44–45
pulse-code-modulated (PCM) systems 41
quality of service and S/N ratio 50
quantization 41
quantization noise, pulse-code-modulated (PCM) systems 41–42
sampling 40
intermodulation (IM) noise 25–26
modulation-demodulation schemes 32–39
amplitude modulation 32–34
angle-modulated signals 35–37
baseband signal formation 34–35
FM and AM systems compared 37–39
multipath interference 17
RF link design 29–32
thermal 17–23
Noise figure, communication systems 21
Nonequiripple stopbands, equiripple passbands and, computer-aided characteristic polynomial synthesis 136–138
Nonhomogeneous fields 652, 667
Nonlinear high-power amplifier (HPA), intermodulation (IM) noise 25–27
Normalization, polynomial functions 92–93
$N+2$ transversal matrix:
circuit approach, transversal coupling matrix, $N+2$ coupling matrix, coupling matrix synthesis 273–275
reduction to folded canonical form, circuit approach, transversal coupling matrix, $N+2$ coupling matrix, coupling matrix synthesis 275
$N+2$ coupling matrix, coupling matrix synthesis 268–282
transversal coupling matrix 269–275
$N \times N$ coupling matrix formation, coupling matrix synthesis 250–253, 259–261
Notch filter 701–703

O

Odd-mode coupling matrix synthesis technique 282–292
Odd-mode coupling submatrices, even-mode coupling submatrices and 310
Odd-order elliptic function filters, lossless lowpass prototype filter functions 104–105
Operating environment, RF filter specifications 70
Optimization techniques. *See* Electromagnetic (EM) simulation technique
Original equipment manufacturers (OEMs) 8
Output multiplexer (OMUX):
satellite system design 52
transmitter section architecture, satellite system design 58–60
Output switch matrices (OSM), satellite system design 52

p

Parallel connected filters 679, 700–701
Parallel-connected two-port networks (folded coupling matrix reconfiguration) 306–310
  even-mode and odd-mode coupling submatrices 310
  generally 306–309
Parallel-coupled inverters, extraction process, ladder network synthesis 233–235
Particle-in-cell (PIC) codes 652
Passive intermodulation (PIM):
  high-power considerations 639, 666–670
    control guidelines 669–670
    generally 666–668
    measurement 668–669
  transmitter section architecture, satellite system design 60–62
Pfitzenmaier configuration:
  dual-mode filters, coupling matrix configurations 445–446
  folded coupling matrix reconfiguration 301–303
Pfitzenmaier, G., 301
Phase change materials (PCM) 732, 733, 744, 751
Phase-coupled extracted pole filters, single-and dual-mode resonator filters 447–450
Phase length removal, extracted pole filter synthesis 324–325
Phase linearity, lowpass prototype characteristics 96
Phase modulation (PM), angle-modulated signals, noise 35
Phase response, lowpass prototype characteristics 95–96
Phase shift keying (PSK), digital modulation-demodulation schemes 45
Physical constants 857
Piezoelectric 752–759
Piezomotors 734, 735, 743, 744
Polar display of reflection coefficient, unloaded *Q* factor measurement 392
Poles, zeros and, input reflection coefficient 619–622
Polynomial forms, Chebyshev filter functions/network filter synthesis 196–198
Polynomial functions, for doubly terminated lossless lowpass prototype filters 88–93
Polynomials, characteristic, lossless lowpass prototype filter functions 93–95. *See also* Computer-aided characteristic polynomial synthesis
Polynomial synthesis, Chebyshev filter functions, network filter synthesis, general polynomial synthesis method 189–193
Polynomial transfer forms, network filter synthesis 177–186
Power efficiency, advanced modulation schemes, digital transmission 48–49
Power flux density, communication channel and link budgets 13
Power handling capability, dielectric resonator filter design 531
Power rating(s):
  thermal dissipation impact, high-power breakdown phenomena 646–647
  waveguides and coaxial transmission lines, high-power breakdown phenomena 645–646
Power transmission, communication channel and link budgets 13–15
Predistorted filter characteristics, network filter synthesis 200–208
  generally 200–205
  synthesis 205–208
Probe coupling 696, 699
Proof-of-concept model, cul-de-sac configuration 314–319
Propagation, radio spectrum 6–7
Prototype circuit synthesis, distributed stepped impedance lowpass filter 404–405
Public switched telecommunications network (PSTN), cellular system design 63
Pulse-amplitude-modulated (PAM) system, noise, digital transmission 41
Pulse-code-modulated (PCM) systems:
  noise, digital transmission 41
  quantization noise, digital transmission 41–42

q

Quadruple mode dielectric resonator 538–539
Quadruple mode filter 538
Quadruple mode resonator 526
Quality factor, dissipation effect, lossless lowpass prototype filter function circuit configurations, dissipation factor * and 114–115
Quality of service, S/N ratio and, digital transmission, noise 50
Quantization noise:
  digital transmission 41
  pulse-code-modulated (PCM) systems 41–42
Quasi-dual mode resonator 532
Q value, dielectric resonator filter design 528–529

r

Radio Consultative Committee (CCIR) 8
Radio frequency (RF) filter networks 1–74. *See also* Communication systems; Digital transmission; Modulation-demodulation schemes; Noise; Systems design; specific RF entries
  communication channel and link budgets 10–15
    antennas 11–15
    signal power 10–11
  communication system model 1–6
  information concept 8–10
  noise 15–50
    communication systems 15–32
    digital transmission 39–50
    modulation-demodulation schemes 32–39
  overview 1
  radio spectrum 6–8
  systems design 50–72
    cellular systems 62–66
    RF filter specifications 68–71
    RF filter technology 72
    satellite system 50–62
Radio propagation, radio spectrum 6–8
Radio spectrum 6–8
  as natural resource 8
  radio propagation 6–7
Reactance transformations 718–727
Realization:
  distributed stepped impedance lowpass filter 405–410
  mixed lumped distributed lowpass filter 418–425
  single-and dual-mode resonator filters 434–440
Receive antenna:
  communication channel and link budgets 11–15
  receiver 5
Receiver, communication system model 3, 5
Receive station, satellite system design 53
Reconfigurable filter 731, 732
Rectangular waveguide designations 863–864
Rectangular waveguide resonator, unloaded *Q* factor 385–386
Recursive technique, Chebyshev filter functions, network filter synthesis 193–196
Reduction, coupling matrix synthesis 261–268
Reflection coefficients:
  transmission coefficients and, polynomial functions 90–92
  unloaded *Q* factor measurement 392–393
Reflection parameters $S_{21}(S)$ and $S_{11}(S)$, polynomial transfer forms and 177–186
Reflection polynomials:
  admittance function $[Y_N]$, transversal coupling matrix, $N+2$ coupling matrix 269–270
  formation, mixed lumped distributed lowpass filter 413–415
Reflection-type allpass networks
  described 552
  practical narrowband 553–554
    C-section in waveguide structure 555–556
    D-section in waveguide structure 556
    generally 554–555
    narrowband TEM reactance networks 556–557
  synthesis 553–554
Rejection lobe optimization, single-and dual-mode resonator filters 429–431

Reflection zeros, polynomial functions 88–89
Repeat passband, mixed lumped distributed lowpass filter 421–422
Resistively terminated symmetric lattice networks 544–545
Resonant frequency calculation:
  dielectric resonator filters 517–521
  microwave resonators 376–383
    arbitrarily shaped resonators 380–383
    conventional transmission-line resonators 376–378
    transverse resonance technique 378–380
Resonant pair removal, extracted pole filter synthesis 325–327
Resonator frequencies, time-domain tuning, computer-aided diagnosis and tuning 623–624
Resonator voltages, bandpass filters high-power considerations 662–663
Response shapes, characteristic polynomials compared, lossless lowpass prototype filter functions 96–98
RF amplifier, transmitter 3
RF channel combiners 577–597
  circulator-coupled MUX 578
  directional filter combiner 582
  generally 577–578
  hybrid-coupled filter combiner module (HCFM) multiplexer 578–582
  manifold multiplexer 582–597
    channel filter design 589–590
    channel transfer characteristics 588–589
    common port return loss and channel 585–588
    common port return loss computation 588
    design methodology 584–585
    manifold-coupled 583–584
    optimization processes 590–597
    star junction 583
RF channelizers (demultiplexers) 571–577
  circulator-coupled MUX 573–574
  en passant distortion 574–577
  generally 571–572
  hybrid branching network 572–573
RF filter specifications, systems design 68–71
RF filter technology, Systems design 72
RF link design, communication systems noise 29–32
RF multiplexer, transmitter 3
RF signal power, communication channel and link budgets 10–11
RF voltage dependence, multi paction breakdown, high-power considerations 650
Ridge waveguide 681, 683, 684
Ripple factor, UDC relationships, lossless lowpass prototype filter functions 109

S

Sampling, noise 40
Satellite system design 50–62
  channelizer (input multiplexer, IMUX) section 50–56
  generally 44–45
  high-power amplifier (HPA) 56–58
  receive station 53
  transmitter section architecture 58–62
Scattering parameter relations. *See* Network filter synthesis
Scattering [*S*] matrix, multiport microwave network analysis 150–155
Second-order lattice section, lumped-element allpass networks 546–547
Second-order lumped equalizer, dissipation effect 542, 550
Semiconductor varactor 744, 769
Semiconductors 732, 733, 744, 767
Sequential tuning, coupled resonator filters 610–615
Short-step transformers, waveguide and coaxial lowpass filters 410–412
Signal power, communication channel and link budgets 10–11
Signal representation, by continuous exponentials, frequency-domain exponential driving function network equation 82
Signal-to-noise ratio (SNR):
  FM and AM systems compared, modulation-demodulation schemes 37–39
  quality of service and, digital transmission, noise 50
  thermal noise, communication systems 19
Similarity transformation, annihilation of matrix elements and 261–268

Single-and dual-mode resonator filters 427–454
coupling sign correction 442–444
dual-mode filters 440–442
coupling matrix configurations 444–447
"full inductive" filter 450–453
generally 440–441
virtual negative couplings 441–442
extracted pole filters, phase-and direct-coupled 447–450
filter function design 428–434
amplitude optimization 429
group delay optimization 431–434
rejection lobe optimization 429–431
overview 427–428
realization and analysis 434–440
synthesis process 428
Single band domain 719
Singly terminated filter prototype, [*ABCD*] polynomial synthesis, lowpass prototype circuits for coupled-resonator microwave bandpass filters 227–229
Sinusoidal excitations, steady-state response to, circuit theory approximation 83–84
Sixth-degree filter, symmetric realizations for dual-mode filters 297–298
Space mapping, electromagnetic (EM) simulation technique 496–504
S parameters:
analysis of, arbitrarily shaped resonators 381–383
interresonator coupling calculation, coupled resonator filters 465–466
Spectrum. *See* Radio spectrum
Spurious performance, dielectric resonator filter design 529–530
Spurious response, mixed lumped distributed lowpass filter 421
Star junction manifold multiplexer 583
Steady-state response, to sinusoidal excitations 83–84
Stepped impedance (SI) filter. *See* Distributed stepped impedance (SI) lowpass filter; Waveguide and coaxial lowpass filters
Stopbands, equiripple passbands and 137
"Straight" cross-couplings 284
Superconductor filters, hybrid dielectric/superconductor resonator filters 536–538
Surface conditions, multipaction breakdown, high-power considerations 649–650
Switch matrices, satellite system design 52
Switched capacitive bank 733, 744–747
Symmetric lattice networks, resistively terminated, lumped-element allpass networks 543–547
Symmetric lowpass prototype filter networks 129–131
Symmetric realizations (dual-mode filters) 295–300
eighth-degree filter 298–299
generally 295–297
sixth-degree filter 297–298
tenth-degree filter 298–299
twelth-degree filter 299–300
Synchronous resonators/synchronously tuned filters 786–787
Synthesis. *See* Network filter synthesis
Synthesis of advanced circuits approach, extracted pole and trisection elements, trisections 356–361
Systems design 50–72
cellular systems 62–66
RF filter specifications 68–71
RF filter technology 72
satellite system 50–62
channelizer (input multiplexer, IMUX) section 53–56
generally 50–53
high-power amplifier (HPA) 56–58
receive station 53
transmitter section architecture 58–62

t

Tapered-corrugated lowpass prototype circuit, mixed lumped distributed lowpass filter 415–418
Temperature drift, dielectric resonator filter design 530
Tenth-degree filter, symmetric realizations for dual-mode filters 298–299
Thermal dissipation:
bandpass filters, high-power considerations 659–660

Thermal dissipation (*contd.*)
impact of, high-power breakdown phenomena 646–647
Thermal noise, communication systems 17–23
Third-degree network, [*ABCD*] matrix buildup for, circuit synthesis approach 217–218
Three dimensional (3D) filters 733
Three dimensional (3D) resonators 733
Time averaged stored energy (TASE) 659, 666
Time domain network equation, circuit theory approximation 78–80
Time-domain tuning (computer-aided) 622–627
generally 622–623
golden filter response 627
interresonator coupling 624–627
resonator frequencies 623–624
Time-invariant and time-variant systems, systems classification 76
Transfer function:
of electrical networks, frequency-domain exponential driving function network equation 82–83
frequency-domain exponential driving function network equation 81–82
Transfer function S_{21}, distributed stepped impedance lowpass filter 401–403
Transfer polynomial:
admittance function $[Y_N]$, transversal coupling matrix, N+ 2 coupling matrix 269–270
formation, mixed lumped distributed lowpass filter 413–415
Transmission coefficients, reflection coefficients and, polynomial functions 90–92
Transmission matrix, multiport microwave network analysis 155–157
Transmission-type equalizers, allpass phase and group delay equalizer networks 551–552
Transmission zeros, maximally flat passband filters and, lossless lowpass prototype filter functions 107
Transmit antenna:
communication channel and link budgets 11–15
transmitter 3
Transmit-receive diplexers, multiplexing networks 597–602
Transmitter section architecture, satellite system design 58–62
Transponder, satellite system design 52
Transversal coupling matrix, $N + 2$ coupling matrix 269–275
Transverse resonance technique, microwave resonators 378–380
Traveling-wave-tube amplifier (TWTA), high-power amplifier (HPA), satellite system design 56–57
Triple mode dielectric resonator 526, 527
Triple mode filter 528
Triple mode resonator 527
Triple-band 690–694, 699–702, 711, 719–725
Triple-conductor 685–688
Trisections (extracted pole and trisection elements) 343–361. *See also* Extracted pole and trisection elements
circuit approach 344–349
coupling matrix approach 350–356
generally 343–344
synthesis of advanced circuits approach 356–361
Trust region aggressive space mapping (TRASM), EM simulation technique 497
Tunable bandwidth 776–778
Tunable combline filter 744, 746, 747
Tunable dielectric resonator filter 752–772
Tunable filter 731–779
Tunable waveguide filter 773
Tuning element 733
Tuning range 733
Tuning speed 732, 762
Twelth-degree filter, symmetric realizations for dual-mode filters 299–300
20-gap crossing rule 652, 653
Two networks, cascade of, multiport microwave network analysis 160–167
Two-port admittance $[Y_N]$, circuit approach, transversal coupling matrix, $N + 2$ coupling matrix 272–273
Two-port networks:
analysis of, multiport microwave network analysis 157–160

maximum power transfer, distortionless transmission 88
polynomial transfer forms and reflection parameters $S_{21}(s)$ and $S_{11}(s)$ 177–186

u

Unified design charts (UDC):
lossless lowpass prototype filter functions 121–127
relationships among, loss less lowpass prototype filter functions 108–109
single-and dual-mode resonator filters, amplitude optimization 429
Unloaded *Q* factor:
measurement of, loaded *Q* factor measurement and 387–393
microwave resonators 383–386
Upconverter, transmitter 3

v

Vacuum environment, multipaction breakdown, high-power considerations 648–649
Vector network analyzer (VNA) computer-aided diagnosis and tuning 637–639
Virtual negative couplings, dual-mode filters 441–442
Voltage breakdown 660, 839
Voltage breakdown, bandpass filters, high-power considerations 660

w

Waveform, distortionless transmission 87–88
Waveguide(s):
coaxial transmission lines and, breakdown 645–646
designations, rectangular 863–864
Waveguide and coaxial lowpass filters 395–425
commensurate-line building elements 395–396
distributed stepped impedance lowpass filter 401–410
generally 401
prototype circuit synthesis 404–405
realization 405–410
transfer function S_{21}, 401–403
lowpass prototype transfer polynomials 396–401
Achieser–Zolotarev function 400–401
Chebyshev polynomials of the second kind 397–399
mixed lumped distributed lowpass filter 411–425
generally 411, 413
realization 418–425
tapered-corrugated lowpass prototype circuit 415–418
transfer and reflection polynomial formation 413–415
overview 395
short-step transformers 410–412
Waveguide iris filter, impedance inverter model, coupled resonator filters 475–478
Wideband FM, analog angle-modulated signals, noise 36–37
World Administrative Radio Conference (WARC) 8

z

Zeros, poles and, input reflection coefficient 628–629