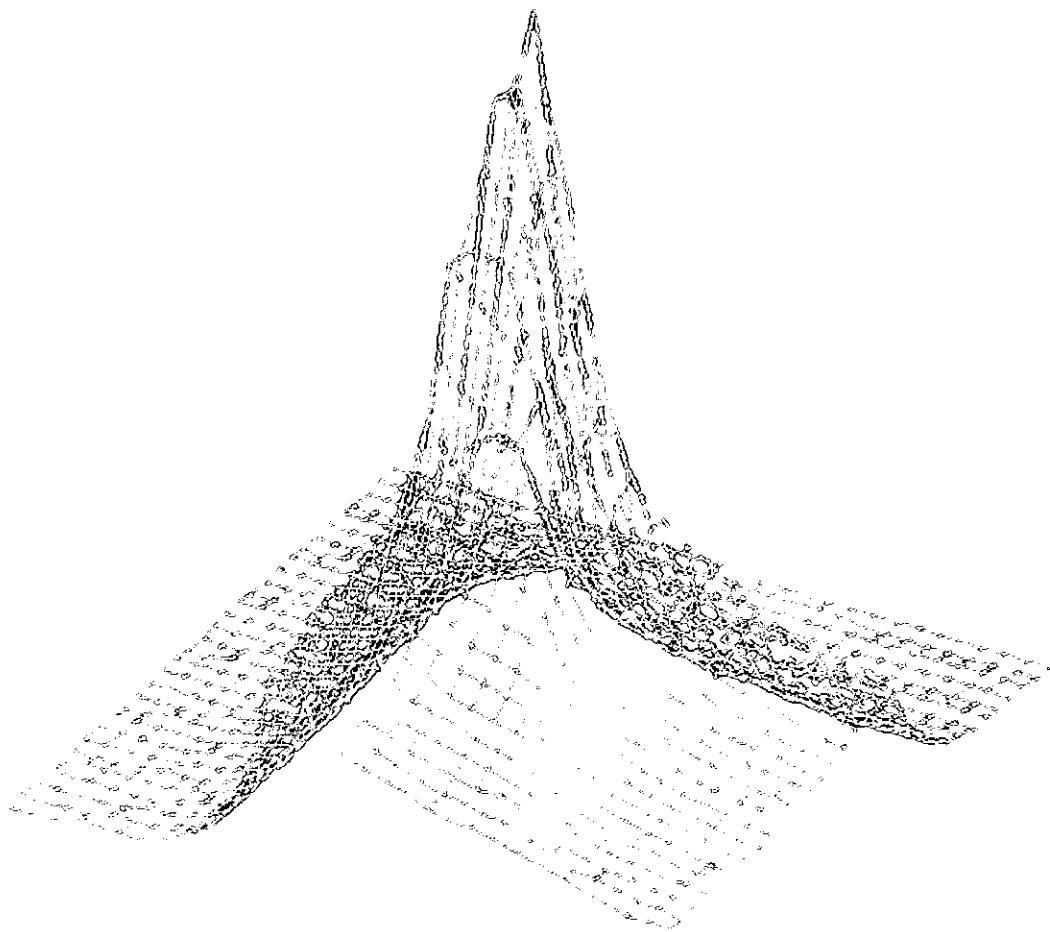


Cuthbert

Broadband Direct-Coupled and Matching RF Networks



Thomas R. Cuthbert, Jr., Ph.D.

This practical guide is for engineers and technicians who design RF networks that filter and match impedances over wide bands or match at a just one frequency. The networks may consist of L's, C's, open- and short-circuited stubs and cascade transmission lines, and transformers. New techniques are described clearly and at a level between seminars and graduate-level instruction.

Direct-coupled filters consist of parallel (or all series) resonators coupled by reactive subnetworks and are found in many forms in all frequency bands. Simple couplings are combinations of L's and/or C's to provide all-pole or elliptic response shapes over any band width. Until now, direct-coupled filters were limited to narrow passband widths. Now broadband design is easy in terms of the loaded Q's of resonators. A wide range of positive element values is always available, with automatic adjustment of design parameters to useful criteria simplified by spreadsheet optimizers.

The grid approach to broadband impedance matching (GRABIM) maximizes or shapes power transfer between source and load described only by discrete-frequency impedance data. It reliably locates the neighborhood of the likely global solution by an efficient grid search based on knowing each benign reflection function versus element parameters. Then, a minimax-constrained gradient optimization precisely locates the solution while pruning any unnecessary elements from candidate networks. GRABIM replaces sophisticated polynomial mathematics by optimization with assured outcome.

Many equations, illustrations, algorithms and 100 references support programming and further study. What little software is required is available. The most important feature is the concise explanation of these useful techniques.

Broadband Direct-Coupled
and Matching RF Networks

J. F.
Seiders

Of related interest...

Circuit Design Using Personal Computers

Thomas R. Cuthbert, Jr., Ph.D.

This practical guide to designing RF circuits in all frequency bands makes it easy to implement both classical and sophisticated design techniques. It is intended for practicing electrical engineers and for upper-level undergraduates. Its topics will also interest engineers, who design circuits derived in terms of complex variables and functions, to provide impedance matching, filtering, and linear amplification.

The numerical methods include solution of complex linear equations, integration, curve fitting by rational functions, nonlinear optimization, and operations on complex polynomials. Several direct-design methods for filters are described, and both single-frequency and broadband impedance matching techniques and limitations are explained. The methods are supported by 45 programs in BASIC source code for use with BASIC®, QBASIC®, GWBASIC®, or QuickBASIC® and are available on floppy disk from the author.

494 pp., John Wiley & Sons 1983. Republished by Krieger Publishing 1994, and by Author 1996.

About the Author

Thomas R. Cuthbert, Jr., Ph.D., PE, is a consultant and teacher. He was the Director of Advanced Technology at Rockwell International and Manager of Microwave Technology at Texas Instruments. He studied at M.I.T., Georgia Tech, and S.M.U. His two other books were originally published by John Wiley: *Circuit Design Using Personal Computers* (1983) and *Optimization Using Personal Computers* (1987).



Optimization Using Personal Computers With Applications to Electrical Networks

Thomas R. Cuthbert, Jr., Ph.D.

This practical guide to optimization, or nonlinear programming, describes the theory and application of methods that automatically adjust design variables. It is intended for engineers, upper-level undergraduates, graduate students, and scientists who use personal computers. Gradient-based search techniques are explained, as opposed to heuristic direct-search, random, or genetic search methods. This book emphasizes design objectives, especially for electrical networks and their analogs.

The material encourages interaction between the user and the computer by offering hands-on experience with the mathematics and computations of optimization. It shows how to produce useful answers quickly, while developing a feel for fundamental concepts in matrix algebra, calculus, and nonlinear programming. Thirty-three BASIC computer programs are included for use with BASIC®, QBASIC®, GWBASIC®, or QuickBASIC® and are available on floppy disk from the author.

474 pp., John Wiley & Sons 1987. Republished by Author 1998.

ISBN 0-9669220-0-X



Thomas R. Cuthbert, Jr., Ph.D.

975 Marymont Drive
Greenwood, AR 72936
Phone (501) 996-5713
FAX (501) 996-5618
E-Mail TRCPEP@AOL.COM

9 780966 922004

Broadband
Direct-Coupled and Matching
RF Networks

Broadband Direct-Coupled and Matching RF Networks

**Thomas R. Cuthbert, Jr.
Professional Engineer and Pilot
975 Marymont Drive
Greenwood, Arkansas 72936**

A TRCPEP Publication

Thomas R. Cuthbert, Jr.

Greenwood, Arkansas

Copyright © 1999 by Thomas R. Cuthbert, Jr.

All Rights Reserved.

This book reports original research and development and provides references to highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Reproduction, transmission, or translation of this book or any part in any form or by any means, including photocopying, microfilming, recording, or by information storage and retrieval system, without permission of the copyright owner is unlawful.

Direct all inquiries to:

**Thomas R. Cuthbert, Jr.,
975 Marymont Drive,
Greenwood, AR 72936.**

ISBN 0-9669220-0-X

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*To my brother,
Dr. Jerry W. Cuthbert,
who is always there to help Buddy
with all his projects,
including this one*

Curving Line

*In a crevice, in a corner
Of the parametric space,
Live conception's few exceptions
To the unimodal case.*

*O my darlin' O my darlin',
O my darlin' Curving Line,
You're not always unimodal;
Dreadful sorry, Curving Line!*

J. W. Cuthbert

Preface

This book is for engineers and technicians who want to use new and useful techniques to design RF networks that filter and match impedances over a wide band and those that match only at one frequency.

Direct-coupled filters provide a bandpass frequency response by using reactive structures to couple cascaded resonators from one to another and to source and load impedances. The resonators can be all series LC or all parallel LC pairs and the coupling can be any combination of L's or C's or parallel LC traps for stopband enhancement. Direct-coupled networks can be terminated by resistances at one or both ends to function as singly- or doubly-terminated filters, respectively. Direct-coupled filters occur in a many physical forms for use in frequency bands from VLF to K-band; their common design basis is the LC model.

When load and/or source impedances are described by data measured at a set of discrete frequencies over a band, then the broadband matching problem is to find a network that minimizes the power loss over all those frequencies. Since 1977, the highly mathematical real-frequency technique has been employed to solve that problem. This book describes the grid approach to broadband matching, GRABIM, a much superior method that is simple, reliable, and very likely optimal.

My two previous books treat direct-coupled filters, broadband impedance matching, and optimization in considerable detail. For the past four years I have been able to devote most of my time to researching, teaching, and consulting on these and related subjects. My more important discoveries have been new and useful methods for design of broadband direct-coupled and matching RF networks.

These new techniques are presented at a level between the valuable one-to-one contact in my seminars and the practical but graduate-level treatment in my two prior books. Much more detail is included here than is possible to present in my seminars, and more than 100 very specific references are cited.

In addition to direct-coupled and matching networks, there is considerable material included on comprehensive equal-ripple filters and special optimization topics. The former motivates the latter: the most straightforward and reliable way to obtain optimal impedance-matching results is by a grid search followed by minimax-constrained optimization.

The articles by Herbert Carlin, John Orchard, Virginia Torczon, and Mike Powell that inspired GRABIM are acknowledged. I also greatly appreciate the reviews of this manuscript and many valuable suggestions by Stephen Sussman-Fort, Bruce Murdock, and Jerry Cuthbert.

Thomas R. Cuthbert, Jr.

Greenwood, Arkansas

January 1999

Contents

1. INTRODUCTION.....	1
1.1 Purpose	1
1.2 Overview.....	2
1.3 Related Software	5
1.4 Revisions	7
2. FUNDAMENTALS	8
2.1 Power Transfer	8
2.1.1 Complex Source to Complex Load.....	8
2.1.2 Generalized Reflection Coefficient.....	9
2.1.3 Circle to Circle Mapping.....	12
2.1.4 Doubly-Terminated Networks.....	13
2.1.5 Singly-Terminated Networks.....	15
2.2 Major Response Shapes.....	17
2.2.1 Lowpass to Bandpass Transformation	17
2.2.2 Insertion Loss Behavior	19
2.2.3 Reflection Coefficient Behavior.....	20
2.2.4 Flat Loss.....	21
2.2.5 Stopband Ripple.....	22
2.2.6 Effects of Component Dissipation	23
2.3 Significance of Loaded Q.....	24
2.3.1 Series-Parallel Conversion	24
2.3.2 Resonator VA/W.....	25
2.3.3 Resonator Efficiency	26
2.4 Ladder Network Topologies.....	27
2.4.1 Lowpass Prototype Networks.....	27
2.4.2 Classical Bandpass Prototype Networks	28
2.4.3 Direct-Coupled Prototype Bandpass Networks	31
2.4.4 Transformations in Bandpass Networks	32
2.4.5 Duality	34

2.5 Network Component Values	35
2.5.1 Equal-Element Lowpass Prototype.....	35
2.5.2 Program ALLCHEBY.EXE	36
2.5.3 Elliptic Function Lowpass Prototype.....	38
2.5.4 Scaling.....	39
2.6 Network Analysis.....	40
2.6.1 ABCD Two-Port Parameters.....	40
2.6.2 Cascading Two-Port Subnetworks	41
2.6.3 Scattering Parameters	42
2.6.4 Special Relations for ABCD Parameters	43
2.6.5 Hilbert Transform.....	43
2.7 Summary of Fundamentals	45
3. DIRECT-COUPLED FILTERS	47
3.1 Prior Technology	47
3.1.1 Classical Filters	48
3.1.2 Coupled Resonator Concept	48
3.2 Properties.....	49
3.2.1 Topologies.....	49
3.2.2 Resonators.....	50
3.2.3 Inverters.....	51
3.2.4 Narrowband Choices in Parallel Resistance Space.....	53
3.3 Response Characteristics	55
3.3.1 Stopband Selectivity	55
3.3.2 Passband Width.....	57
3.3.3 Mismatched Terminations	60
3.3.4 Effects of Dissipation.....	61
3.3.5 Passband Distortion	62
3.4 Eliminating Passband Distortion	65
3.4.1 Stagger-Tuning Two Resonators.....	65
3.4.2 Stagger-Tuning Three Resonators.....	67
3.4.3 Exact Replacement of Resonators.....	68
3.4.4 All Possible All-Pole Topologies	71
3.4.5 Wideband Choices in Parallel Resistance Space	72
3.5 The Method of Choices	75
3.5.1 A Spreadsheet for Four Resonators	76
3.5.2 A Spreadsheet for Five Resonators.....	80
3.5.3 A Spreadsheet for a Four Resonator Elliptic Filter	82

3.6 Tuning.....	84
3.6.1 Alternating Open- and Short-Circuit Method	84
3.6.2 Reactive Input Reflection Function	85
3.6.3 Narrow-Band Reflection Poles and Zeros	86
3.6.4 Wideband Networks Having Exact Responses.....	87
3.7 Summary of Direct-Coupled Filters.....	87
4. COMPREHENSIVE EQUAL-RIPPLE FILTERS	90
4.1 Purpose.....	90
4.2 Response Continuum.....	90
4.2.1 Transducer and Characteristic Functions.....	90
4.2.2 Transmission Zeros.....	92
4.2.3 Passband Selectivity.....	94
4.2.4 Stopband Selectivity.....	95
4.3 Challenges of Polynomial Synthesis	96
4.3.1 Underlying Concepts	96
4.3.2 Mathematical Operations and Sensitivities.....	96
4.3.3 The Approximation Problem	97
4.3.4 Realization of Element Values.....	98
4.3.5 Road Map for Topologies	98
4.4 Element Responses at Discrete Frequencies	99
4.4.1 Filters	100
4.4.2 Single-Match Broadband Networks.....	102
4.4.3 Double-Match Broadband Networks	104
4.4.4 Lessons Learned	105
4.5 Synthesis by Iterated Analysis	105
4.5.1 Zeros and Poles of the Characteristic Function	106
4.5.2 Characteristic Zeros of Ladder Filters.....	106
4.5.3 Balancing Variables and Constraints.....	107
4.5.4 Efficient Network Analysis	108
4.5.5 Efficient Optimization	109
4.6 Summary of Comprehensive Equal-Ripple Filters	110
5. MATCHING NETWORKS	112
5.1 Single-Frequency Matching.....	113
5.1.1 Zero Reflectance	113
5.1.2 El Sections Matching Resistances	114
5.1.3 El Sections Matching Impedances	115

5.1.4 Pi and T Sections	116
5.1.5 Cascade Transmission Lines.....	117
5.2 Analytic Gain-Bandwidth Theory.....	118
5.2.1 Gain-Bandwidth Limitation.....	119
5.2.2 The Single-Match Minimization Problem	120
5.2.3 Single-Match Optimal Results.....	121
5.2.4 Chebyshev Network Element Values	122
5.2.5 Other Terminal Impedances	123
5.2.6 Measured Loaded Q	124
5.3 Real-Frequency Technique	125
5.3.1 Non-Analytic Real-Frequency Data.....	126
5.3.2 Approximating the Network Resistance Function	127
5.3.3 Synthesis of a Resistance Function	129
5.3.4 Double-Matching Using the RFT.....	129
5.3.5 Double-Matching Using Brune Functions.....	130
5.3.6 Double-Matching Active Devices	131
5.4 Introduction to GRABIM	131
5.4.1 Thesis	132
5.4.2 Overview	133
5.4.3 Branch Parameter and Reactance Effects	136
5.4.4 The Response Surface	143
5.5 Algorithms for GRABIM.....	148
5.5.1 Efficient Network Selection and Analysis	148
5.5.2 Grid Searches.....	152
5.5.3 Constrained Optimization for Element Removal	157
5.6 Examples Using GRABIM	162
5.6.1 Example of a Non-Analytic Bandpass Problem	162
5.6.2 Example of a Distributed Interstage Network.....	164
5.6.3 Example of Neighborhood Matching.....	165
5.6.4 Example of Topological Simplification and Sampling	167
5.7 Summary of Matching Networks	167
6. GRABIM IN DETAIL	171
6.1 Formulation	171
6.1.1 The General Problem.....	171
6.1.2 The General Solution.....	171
6.1.3 The Specific Problem	172

6.2 Network Analysis.....	173
6.2.1 Transducer Function and Its Derivatives	173
6.2.2 Derivatives with Respect to the Variable Space	174
6.2.3 Lossless Ladder Network Analysis Equations	175
6.2.4 Lossless Ladder Network Analysis Algorithm	177
6.3 Grid Search.....	178
6.3.1 Minimax Objective Function.....	178
6.3.2 Pattern Searches.....	178
6.3.3 Grid Geometry	179
6.3.4 Grid Algorithm.....	181
6.3.5 Other Factorial Search Algorithms	183
6.4 Method of Multipliers.....	185
6.4.1 The Problem	185
6.4.2 Quadratic Penalty Functions	186
6.4.3 Adjusting the Multipliers	187
6.4.4 Gauss-Newton Unconstrained Minimizer	189
6.4.5 Alternative Constrained Optimization Methods.....	190
6.4.6 Frequency Sampling Strategy.....	192
6.5 Summary of GRABIM in Detail	193
APPENDIX A — CIRCLE TO CIRCLE MAPPING	195
A.1 Bilinear Mapping	195
A.1.1 Right-Half Plane to Circle Mappings	195
A.1.2 Circle to Circle Mapping	197
A.2 Interior Circular Images	197
A.2.1 Concentric Circle in a Unit Circle	197
A.2.2 Eccentric Circle in a Unit Circle.....	197
A.2.3 Neighborhood Parameters	198
A.3 Planar Lines to Circles.....	201
A.3.1 Constant Resistance.....	201
A.3.2 Constant Q.....	202
A.3.3 Eccentric Vector Magnitude	203
APPENDIX B — ABBREVIATIONS AND SYMBOLS.....	204
REFERENCES.....	205
AUTHOR INDEX.....	213
SUBJECT INDEX	215

1. Introduction

1.1 Purpose

This book describes how to design direct-coupled filters and impedance-matching networks consisting of L's, C's, open- or short-circuited transmission line stubs, and cascade transmission lines. The distinguishing feature of these methods is that they can be designed to perform over a broad band of frequencies. This material extends and consolidates some important subjects in my two previous books [Cuthbert,1983,1987]¹.

Ordinarily, direct-coupled networks are designed with assumptions that limit their application to narrow frequency bands, i.e., less than 20 per cent. It has only recently been discovered that their desirable topologies can be realized simply without the passband distortion previously accepted, and the method enables choice of a wide range of positive element values for any band width. Direct-coupled network theory underlies various types of microwave filters and has been the basis for many other filter applications. Some of these variations have been described [Cuthbert,1987:Chap.8]; this book takes a more elementary approach in order to maintain clarity. Because direct-coupled filters consist of coupled resonators (all series or all parallel), the elementary broadband impedance matching design technique based on the loaded Q of one or both terminating resonators is included.

The concept of loaded Q as the ratio of reactive to real power in a resonator is well known, especially for single-frequency impedance matching by the "1+Q²" method. The fundamental broadband matching (or gain-bandwidth) limitation can be expressed in terms of an output termination's loaded Q normalized to Q_{BW} , the ratio of passband center frequency to passband width. The limitation not only shows a theoretical maximum of power transfer over a frequency band, but also shows the futility of utilizing more than just a few network branches. This latter property leads to the concept of trying promising network topologies in what is superficially an exhaustive enumeration of the few element values in an attempt to match a discrete set of impedances versus frequency at either or both ends of the candidate network. Automation of that concept is the grid approach to broadband impedance matching (GRABIM) previously described [Cuthbert,1994,1997]. This book extends the rationale for such direct searches as recently developed by [Torczon, 1997].

The grid approach to a likely global solution for a candidate matching network does not reliably eliminate unnecessary network

¹References for the entire book are listed after Appendix B.

branches. That elimination takes place in a much more highly-convergent constrained optimization that utilizes the Lagrange multiplier concept from classical mathematics. This book shows how that is accomplished numerically, once the neighborhood of the solution is obtained by grid search. Thus, the broadband matching problem is described as a constrained optimization problem, and the basics of related optimization concepts are included for completeness.

1.2 Overview

Chapter Two presents six fundamental filter and matching concepts that are essential to understanding major topics in the following chapters. Foremost is the concise measure of available power that can be obtained from complex (having resistive and reactive components) or ideal sources and delivered to a complex load. In the case where some resistance occurs at both ends of a lossless two-port network, there are important relationships available at either port or at any other plane cutting the network. These relationships involve the generalized reflection coefficient, which is easily related to the conventional Smith chart. Transfer functions versus frequency for networks with resistance only at one end and for networks composed of some dissipative components are also described.

Important network response shapes versus frequency are described for filters and matching networks. Prototype lowpass and related bandpass network topologies are described, including direct-coupled bandpass configurations that may include trap couplings for stopband ripple (elliptic) responses. Conventional normalization of impedance and frequency units is described, and the necessary transformation of element values to unscaled levels is discussed. Several transformations for sections of bandpass networks are described, especially those that do not affect the network frequency response. Component values for the equal-ripple (Chebyshev) and equal-element (minimum-loss) responses are provided in tables or by program ALLCHEBY.EXE. References to tables of element values that produce many other response shapes are also given.

Chapter Two includes a brief but crucial description of the "1+Q²" technique to convert between series and parallel impedance forms. That subject acquaints the reader with the concept of loaded Q, a unifying parameter in both direct-coupled filters and broadband impedance matching. Methods for efficient analysis of ladder networks using the chain parameters are discussed, because they are essential for the grid approach to broadband matching and subsequent gradient optimization. The Hilbert transform that relates the terminal resistance to its related reactance of ladder networks is mentioned because of its crucial role in a well known but complicated broadband matching technique.

Chapter Three deals with lumped LC direct-coupled filters which provide bandpass responses through reactive structures that couple a cascade of all-series or all-parallel resonators from one to another and to source and load impedances. Prior technology is reviewed, because similar narrow-band filters have been designed for about 60 years. The narrow-band inverter concept was introduced about 40 years ago, and is described as the simplifying concept for connecting resonators in cascade. This book does not deal with the endless variation of inverter and resonator realizations, e.g., waveguide apertures connecting resonant cells. Rather, the fundamental LC networks underlying the various physical realizations are treated here using the unifying resonator loaded Q parameter. Asymptotes of stopband selectivity of direct-coupled filters are described as affected by the balance of inductive (or magnetic) and capacitive inverters and terminal couplings to source and load. The effect of source and load mismatch is related to response ripple peaks and/or flat-loss (dB offset) at band-center frequency (at dc in the related lowpass prototype network).

Chapter Three mentions the passband distortion inherent in narrow-band inverters as well as the added distortion due to dissipative components. It is shown how passband distortion easily can be eliminated in lossless direct-coupled filters. The concept is presented both as a stagger tuning of specified resonators by a simply-determined amount and as the direct conversion of all series (parallel) resonators to equivalent coupled parallel (series) resonators using simple equations. This new and useful design technique eliminates passband distortion over a passband of any width in direct-coupled filters having equal numbers of L and C couplings. Also, effective and convenient methods are described to avoid negative elements while identifying wide ranges of possible positive element values. Numerous examples are provided, including an elliptic filter that absorbs a resonant load over a broad frequency band.

Chapter Four bridges the gap between restricted direct-coupled network topologies and equal-ripple lowpass and bandpass filters having any useful topology. It is remarkably easy to state the constraints on locations of transmission zeros in the Laplace frequency plane for all such filters. The situation described includes compact expressions for both passband and stopband responses and passband frequencies where the response has peaks and valleys. Because element values for these general networks can be obtained only by polynomial synthesis, the complexity and limitations of that discipline are reviewed.

A second purpose of Chapter Four is to show how the reflection response of equal-ripple filters and matching networks behaves versus each network branch value. These cross sections of reflection magnitude versus L or C values show characteristics that are vital in several new techniques that follow. The iterated analysis method that obtains more

accurate element values than polynomial synthesis is described as one way to take advantage of the cross-section behavior of element values. In turn, that sets up the later introduction of the even more general GRABIM technique, the grid approach to broadband impedance matching.

Chapter Five begins with design of lumped-element networks that match resistances at a single frequency using the loaded Q parameter. These "el", T, and Pi examples clarify both loaded Q and the concept of parallel resistance levels. Then, analytic (classical) gain-bandwidth impedance-matching theory is reviewed, especially the interaction of reciprocal fractional bandwidth or Q bandwidth (Q_{BW}) and the loaded Q of a single LCR resonator load. Concise classical broadband matching results are presented for the three possible source cases: resistive, purely reactive, and single LCR resonator. The ALLCHEBY.EXE program that calculates all these matching cases as well as all other Chebyshev filter cases is mentioned again.

The real frequency broadband matching method introduced in 1977 and extended since then is very briefly described, in order that the reader can appreciate its mathematical and procedural complexity and limitations. Then, the grid approach to broadband impedance matching (GRABIM) is introduced, beginning with the initial process of locating an optimal solution for a chosen type network topology by a highly efficient direct-search technique. The underlying impedance mapping phenomenon that ensures optimal results is then introduced, followed by a view of the grid process as discrete line searches over all network branch/parameter values in log space. An extension is made to transmission-line elements in a matching network by showing its close relationship to the lumped LC case. Details of the grid search are provided in Chapter Six.

Chapter Five ends with the crucial last step in the GRABIM technique, a highly precise solution to the broadband matching problem in the context of a constrained optimization problem strongly related to classical Lagrange multipliers. An overview of the gradient optimization technique that must start at the approximate solution from the grid search is provided with details given in Chapter Six. Many examples show how the grid search finds the neighborhood of the global solution and how the gradient-based second step eliminates unneeded network branches by finding the precise solution. Included in the examples are applications of the matching of impedance neighborhoods that result from uncertainty of load and/or source data as with closely-coupled antenna elements where the impedance varies in a neighborhood about a nominal value at each frequency.

Chapter Six provides the mathematical and algorithmic detail required for programming. This most reliable broadband matching method, GRABIM, is a structured optimization process, beginning with a

direct grid search followed by an augmented Lagrangian method to accomplish a gradient-based constrained optimization. A brief overview of the general problem and solution is related to the specific broadband impedance matching problem. The first part of Chapter Six extends the efficient RF network analysis of Chapter Two to include means to calculate exact partial derivatives that are essential to gradient-based optimization.

Then, direct search methods are described, like the grid search, that do not require derivatives or functions that are smooth. The basic investigation of pattern search algorithms that include the grid search is cited for credible formulation of the first search phase of GRABIM. The constrained optimization problem is stated as in mathematical literature and as related to broadband matching. Common penalty functions such as the exterior quadratic penalty are extended to deal with the minimax (ideal equal-ripple) requirement.

An important part of Chapter Six shows how the sequence of unconstrained optimizations technique (SUMT) can be applied to obtain precise numerical solutions to constrained nonlinear optimization problems such as that required for broadband impedance matching. These so-called augmented Lagrangian methods simply extend the least-squared errors concept by adding variable goals or targets to the error residuals (differences). This extension of ordinary penalty function optimization techniques, known as the method of multipliers, completes the explanation of why the final step in the GRABIM method works so well.

The Gauss-Newton unconstrained optimization algorithm is the inner of two nested optimization loops in the method of multipliers. It is shown to solve the nonlinear least squares problem very efficiently and is easily adapted to the method of multipliers. Its basic step in variable space is given, and the exact formulation of the necessary first partial derivatives is clearly described.

In addition to numerous examples throughout these chapters, many tables and figures are provided. Appendix A describes essential properties of the bilinear circle-to-circle mapping functions that explain the benign nature of lossless network reflectance as a function of element variables.

References for further explanation and suggested development follow Appendix B, which is a collection of abbreviations and symbols utilized in this book.

1.3 Related Software

Executable programs for PC/DOS computers are available from the author at the address on the copyright page. Revisions are planned as improvements become available. The programs have been written in QuickBASIC® version 4.5. Generally, this book describes design methods

that may be superior to those currently programmed. Improvements in software to incorporate the superior techniques described in this book will be made available as soon as possible. The objective is to produce engineering design data in the most direct form for both the programmer and the user. Programs currently available are:

- **CONETOPM.** COnstrained NETwork OPtimizer for Matching. Includes ladder analysis versus frequency for resistive or sampled termination impedances. A wide range of responses, voltage-current, and sensitivity data are generated and can be saved to disk file (e.g., for use with spreadsheets, graphs, etc.). Constrained and bounded optimization of ladder networks by a Gauss-Newton optimizer is included. A complete GRABIM (GRid Approach to Broadband Impedance Matching) capability is also provided in the optimization menu, including cross-section reflectance versus all possible element values.
- **S11TOZ.** Converts a list of S parameter reflection data pairs (S_{11} or S_{22}) to resistance and reactance values normalized to one ohm. S parameter data must be in numerical magnitude (not dB) and angle in degrees and stored in an ASCII file. The impedance data can be converted to admittance data to observe real-part trends for model recognition. Converted data may be stored on disk files.
- **ALLCHEBY.** Designs all Chebyshev filter and matching prototype lowpass and bandpass networks by providing g_i elements and loaded Q values. Optimal gain-bandwidth matching is obtained for a single RLC load resonator and a source that is purely resistive, purely reactive, or an RLC resonator. The best possible result for infinite networks is provided as well as that for a specified number of elements or resonators. Passband and 20-dB stopband edge frequencies are provided, and attenuation at any requested normalized frequency is given. Estimates are provided for midband dissipative loss given a uniform unloaded Q value.
- **DENORM.** Denormalizes prototype element values and vice versa. Receives typed entries of normalized ohms reactance or susceptance and converts those to inductance and capacitance, respectively. Conversion is based on initial data entry of units of frequency, inductance, and capacitance as well as a specific frequency and an impedance scale factor, if not unity.
- **RIPFREQS.** This program implements Daniel's technique for predicting the Chebyshev passband peak and valley frequencies and the exact transducer loss function for all frequencies. The required data are the numbers of zeros of transmission at dc and infinity as well as those at arbitrary stopband frequencies. Orchard's filter

design by iterated analysis as well as pole-placer algorithms depend on this capability

- o EXCEL CHOICES Spreadsheets. EXCEL® version 5.0 spreadsheets are provided to design specific wideband direct-coupled filters. Subroutines for replacing series resonators by coupled parallel resonators and elliptic resonators by trap-coupled parallel resonators are included to simplify user construction of all possible coupling combinations. The goal is to aid design of direct-coupled networks that meet all the user's particular requirements, perhaps obtained by the built-in optimizer.

1.4 Revisions

It is hoped that this book can be revised occasionally as improvements, added scope, errata, and new research results become available.

2. Fundamentals

There are just a few concepts that the reader should have in mind to benefit from the following chapters. It is important to know how power is transferred from a source to a load through a two-port network, especially when those terminating impedances are complex, i.e., have both resistive and reactive components. One essential concept is the generalized reflection coefficient, which can be plotted on an ordinary Smith chart. The effects of impedance levels, resistance and reactance relationship, and dissipation enter these considerations. The topology of prototype ladder networks, prominent response shapes versus frequency, how to obtain the sets of element values that produce those responses, and how to analyze the network to calculate a response should be clearly in mind. Finally, the important parameter that unifies all of these effects is loaded Q, the reactive power relative to the real power at significant places in the ladder network.

2.1 Power Transfer

The load impedance of a two-port network must include a resistance to receive power. The case where the source also includes a resistance makes that a doubly-terminated network, and the maximum power that such a source can deliver has a finite limit. This case is best analyzed using generalized reflection coefficients related to Smith charts that map impedances into a unit circle. There is also use for mapping between Smith charts normalized to different impedances. Unlimited power can be delivered by a source with no resistance; then the network is said to be singly-terminated. The power that is delivered by such an ideal source is determined solely by the network's input resistance or conductance.

2.1.1 Complex Source to Complex Load

Power is delivered at a single frequency from a source having voltage E_s (rms) and internal impedance $Z_s = R_s + jX_s$ as shown in Figure 2.1.1. The *maximum possible power* that can be delivered to load

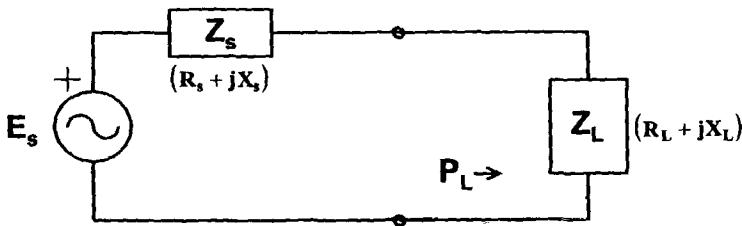


Figure 2.1.1. Power transfer from complex source to complex load.

impedance $Z_L=R_L+jX_L$ occurs when Z_L is equal to Z_s except that $X_L=-X_s$ (a conjugate match); that power is

$$P_{as} = \frac{|E_s|^2}{4R_s}. \quad (2.1.1)$$

Commonly, the load power relative to P_{as} is expressed as a complicated algebraic equation. It is significantly better to express this ratio compactly as the *transducer power gain*:

$$T = \frac{P_L}{P_{as}} = 1 - |\alpha|^2, \quad (2.1.2)$$

where the complex variable α is called the *generalized reflection coefficient* and is defined by

$$\alpha \equiv \frac{Z_L - Z_s^*}{Z_L + Z_s}. \quad (2.1.3)$$

The asterisk (*) superscript indicates conjugation, which reverses the sign of the imaginary part of the quantity.

When $Z_L=Z_s^*$, the numerator in (2.1.3) is zero, making $\alpha=0$ and $P_L=P_{as}$ according to (2.1.2). Besides being compact, these equations introduce the generalized reflection coefficient, α , which occurs in impedance mapping and other important areas of RF network design.

2.1.2 Generalized Reflection Coefficient

The Smith chart is a unit circle centered at the origin of a Cartesian plane; the abscissa represents the real part and the ordinate the imaginary (j) part of a reflection coefficient ρ :

$$\rho = \frac{Z - Z_c}{Z + Z_c^*} = \frac{(R - R_c) + j(X - X_c)}{(R + R_c) + j(X - X_c)}. \quad (2.1.4)$$

To denote Z_c as the impedance at chart center, the conjugate, Z_c^* , is placed in the denominator without loss of generality. For many decades since its introduction, the familiar transmission-line application of the Smith chart assigned $X_c=0$; beyond that, the most common case further assigned $Z_c=50+j0$, giving ρ as the reflection coefficient with respect to a 50-ohm resistance. In any event, the impedance level is often normalized to resistance R_c , so that

$$\rho = \frac{\bar{Z} - \bar{Z}_c}{\bar{Z} + \bar{Z}_c^*} = \frac{(\bar{R} - 1) + j(\bar{X} - \bar{X}_c)}{(\bar{R} + 1) + j(\bar{X} - \bar{X}_c)}, \quad (2.1.5)$$

where

$$\bar{R} \equiv \frac{R}{R_c}, \text{ and } (\bar{X} - \bar{X}_c) \equiv \frac{(X - X_c)}{R_c}. \quad (2.1.6)$$

In every sense of the word, (2.1.4) *maps* the Z plane into the ρ plane as shown in Figure 2.1.2. In fact, it maps the right-half Z plane

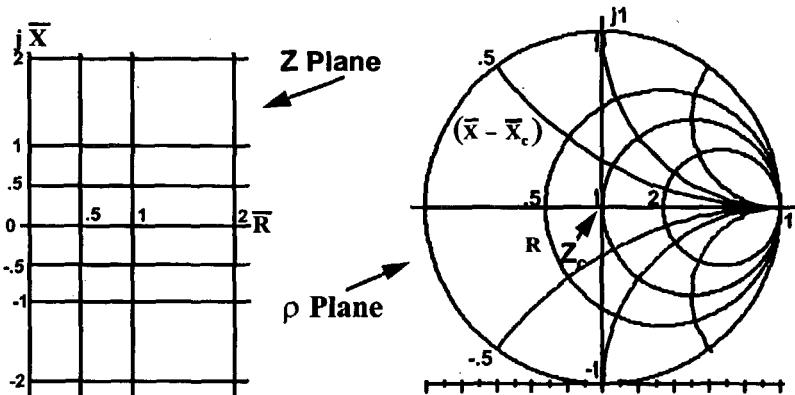


Figure 2.1.2. Generalized Smith Chart: normalized impedance representations.

into a unit circle about the origin of the ρ plane. (The left-half Z plane where $R < 0$ is mapped into that part of the ρ plane outside the unit circle.)

Again, the numerator of (2.1.4) is zero when $Z=Z_c$, which is why the center of the Smith chart, $\rho=0$, is labeled Z_c for Z center. Comparing (2.1.4) to (2.1.3) shows that the conjugation in (2.1.3) occurs in the denominator; that is simply an arbitrary definition to indicate the Smith chart center. For example, if one considers $Z_S=40+j30$ ohms in (2.1.3), then the number to use in (2.1.4) is $Z_c=40-j30$ when solving the power transfer equation (2.1.2).

The circles in the Smith chart in Figure 2.1.2 are loci of constant resistance R/R_c , and the circular arcs are loci of constant $(X-X_c)/R_c$, where $X_c=0$ in less general applications. Also, the arcs in the upper half of the Smith chart represent positive normalized reactance while those in the lower half represent negative reactances. This generalization of the Smith chart requires only that the user consider $(X-X_c)$ instead of just X . It is a little trickier; e.g., the Smith chart real axis (abscissa) represents not $X=0$ but $X-X_c=0$ or $X=X_c$. The power of this concept turns out to be well worth the bother.

The generalized reflection coefficient and the traveling wave on a transmission line having a complex Z_0 can be compared. Contrary to generalized reflection coefficient (2.1.4), the traveling wave reflection coefficient is $(Z_L-Z_0)/(Z_L+Z_0)$, where Z_0 may be either real or complex. When Z_0 is real, the reflection coefficient applied in (2.1.2) does give the relative power. When Z_0 is complex, the traveling wave's reflection magnitude is not directly related to power. The maximum power transfer takes place when $Z_L=Z_0^*$, and it is only when there is a particular traveling-wave reflection that maximum power is transferred from the transmission line to the load. Therefore, the traveling wave may be more convenient for expressing the properties of a port irrespective of the load

impedance, Z_L , but the power waves based on (2.1.4) "give a clearer and more straightforward understanding of the power relations between circuit elements connected through a multiport network" [Kurokawa].

Example 2.1.1. Consider a complex source connected to a complex load as in Figure 2.1.1. Suppose that $Z_S=25-j50$ ohms and Z_L can take on those impedances that cause a 2:1 *standing-wave ratio* (SWR) with respect to 50 ohms. *Problem:* Find the range of power delivered to the load. *Solution:* The SWR is a scalar mapping of the magnitude of a reflection coefficient:

$$SWR \equiv \frac{1 + |\rho|}{1 - |\rho|}. \quad (2.1.7)$$

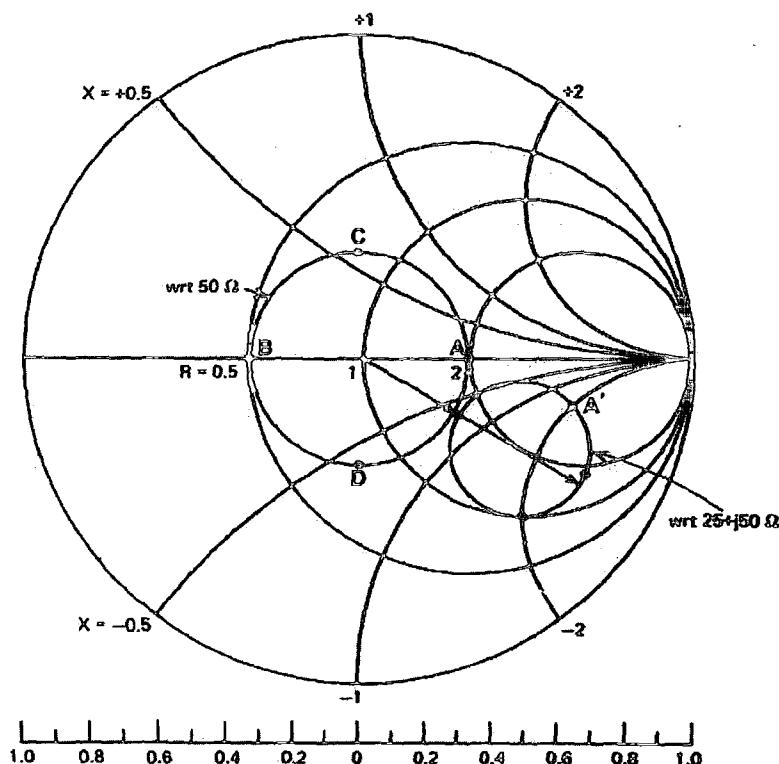


Figure 2.1.3. A 2:1 SWR circle normalized to two different impedances.

Along the real axis of the Smith chart, $SWR=R$ or $SWR=1/R$, since $SWR \geq 1$. See Figure 2.1.3, where the concentric 2:1 SWR circle with respect to a normalized 50-ohm chart center locates all those load impedances to be considered.

The range of load power relative to the maximum available from the source is found from (2.1.2); in this case it can be found graphically by simply taking several numbers from the SWR impedance circle and plotting them on this same Smith chart with respect to $Z_s^*=25+j50$, i.e. as α in (2.1.3). Pick four arbitrary points on the SWR circle: $A=100+j0$, $B=25+j0$, $C=42.5+j32.5$, and $D=42.5-j32.5$ ohms. The smaller circle in Figure 2.1.3 is thus found by the corresponding four values of α using (2.1.3). For example, consider how point A is replotted wrt $Z_c=25+j50$. For $A=100+j0$, (2.1.6) uses $R/R_c = 4$ and $(X-X_c)/R_c = -2$. Then, (2.1.5) shows that that $\rho=(3-j2)/(5-j2)$, so $\rho=0.67$ at angle -11.9 degrees. That is plotted as point A' in Figure 2.1.3.

Graphically, the magnitude of α (distance from chart center to points on the smaller circle) varies from about 0.36 to 0.80. Then (2.1.2) shows that the load power varies from 0.36 to 0.87 of the maximum available power from the source (2.1.1). A general technique for an analytic solution is described next.

2.1.3 Circle to Circle Mapping

A more general view of what is illustrated in Example 2.1.1 and Figure 2.1.3 is shown in Figure 2.1.4, which shows the mapping from the right-half Z plane into the unit circle in an f plane and also into a second unit circle in a g plane. In terms of the problem in Example 2.1.1, $Z_f=50+j0$ so that the 2:1 SWR circle in the Z plane is centered on the $Re Z$ axis as well as being concentric in the f plane in Figure 2.1.4. Z_f is also the center of the f plane circle, so that is the Smith chart showing the 2:1 SWR with respect to $50+j0$ at its center. Mapping function $F(Z)$, shown in Figure 2.1.4, is (2.1.4) with $Z_c=50+j0$. The g-plane unit circle can be considered a generalized Smith chart according to mapping function $G(Z)$ in (2.1.3) so that chart's center represents Z_s^* . Then the maximum and minimum points in the g plane of Figure 2.1.4 represent the extreme values of $|\alpha|$ on the locus. (When the g-plane origin is encircled, as in Figure 2.1.4, the minimum within the locus is $|\alpha|=0$.) There are compact expressions for the radius to those points in the two cases where the SWR image encircles the g plane origin and when it does not. The details are in Appendix A, Section A.2.2.

The important conclusion is that mapping from the right-half Z plane into a generalized Smith chart is possible, as well as mapping between Smith charts having different normalizing impedances (chart centers; see mapping function $H(f)$ in Figure 2.1.4). The interpretations vary, depending on the problem at hand. The SWR expression in (2.1.7) in terms of any generalized reflection coefficient also has applications in situations where there is no transmission line involved.

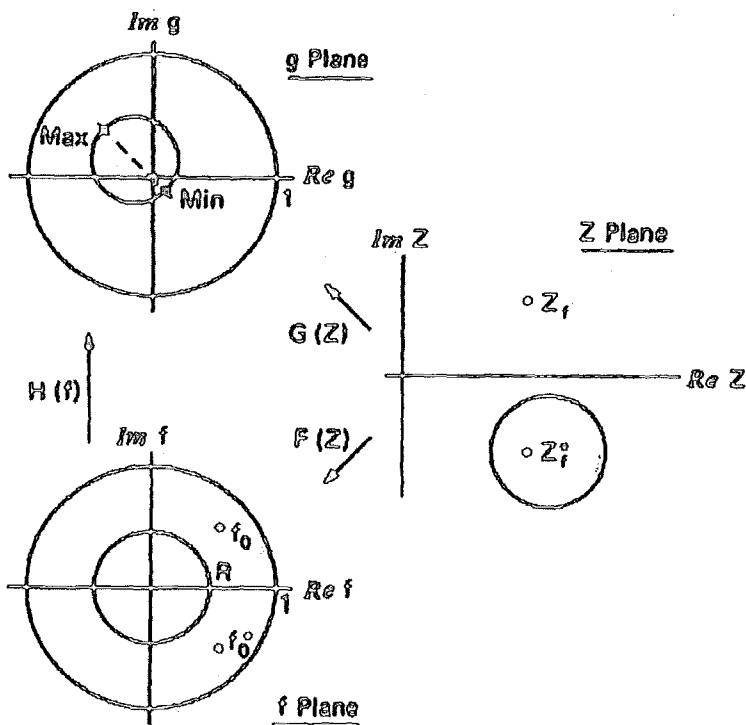


Figure 2.1.4. Bilinear transformations between the Z plane and unit circles.

2.1.4 Doubly-Terminated Networks

Figure 2.1.5 shows that doubly-terminated two-port networks have

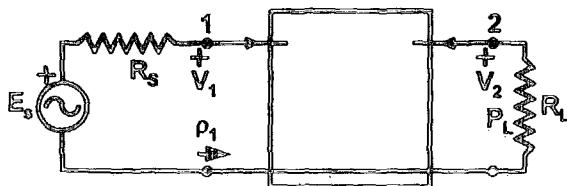


Figure 2.1.5. Doubly-terminated filters have resistances at both ports.

resistances at both ports, and there can also be a reactance with one or both resistances. With those terminations, the frequency selective response of power transfer is determined by the frequency behavior of the input reflection coefficient, ρ_{in} .

It is often useful to assume that the two-port network is lossless, so that the power delivered by the source all reaches the load, as shown in Figure 2.1.6. The power delivered by the source, P_{in} , is governed by (2.1.2), as shown in the left-hand fraction in (2.1.8):

$$|\alpha| = \left| \frac{Z_{in} - Z_s^*}{Z_{in} + Z_s} \right| = \left| \frac{Z_f - Z_T^*}{Z_f + Z_T} \right| = \left| \frac{Z_L - Z_b^*}{Z_L + Z_b} \right|. \quad (2.1.8)$$

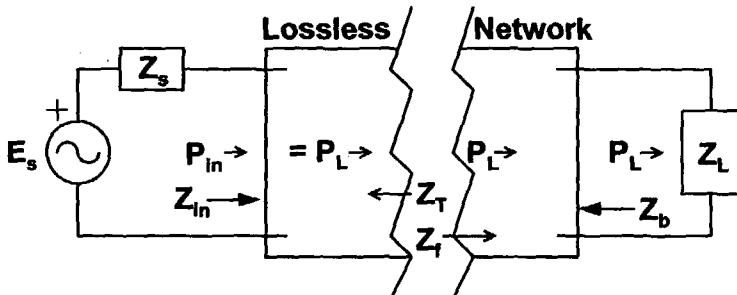


Figure 2.1.6. Power conservation and impedances in a lossless network.

At the output port, the right-hand fraction in (2.1.8) is the pertinent generalized reflection coefficient, where Z_b is the Thevenin equivalent source impedance at that interface. (Z_b is the impedance seen looking into port two when $E_s=0$.) Also, at any interface in the lossless network, a forward impedance, Z_f , and a Thevenin equivalent source impedance, Z_T , can be found, so that the middle fraction in (2.1.8) is defined. The magnitudes of all three of those fractions (reflectance) must be equal because the power is the same at any point in the lossless network. Another important conclusion is that a conjugate match at any point implies a conjugate match everywhere in the network, i.e. $|\alpha|=0$.

A useful fact for designers is that the voltage or current at any point in a conjugately-matched lossless network can only increase by the square root of SWR in the presence of a reflection mismatch. This fact is well known for the conventional SWR of transmission lines relative to voltages and currents in a "flat" line [ITT:24-9]. Remarkably, it is also true for lossless two-port networks of any kind, using generalized reflection coefficients in (2.1.7). The explanation is seen in the Z plane in Figure 2.1.4, where there is an "SWR" circle with the same extremes of $Re\ Z$, i.e., resistance or conductance, whether or not there is an X_c offset as in (2.1.4). Therefore, SWR in (2.1.7) is a quantity just as important in the generalized reflection case, with physical significance regarding standing waves on transmission lines.

Physical networks have dissipative components and are thus lossy to some extent, of course. It is convenient to mention several common performance parameters for lossy doubly-terminated networks such as shown in Figure 2.1.7. *Return loss* (RL) is the ratio P_{ref}/P_{as} in dB:

$$RL = -20 \log_{10} |\Gamma| \text{ dB}, \quad (2.1.9)$$

where Γ is a simple reflection coefficient:

$$\Gamma = \frac{Z_{in} - R_1}{Z_{in} + R_1}. \quad (2.1.10)$$

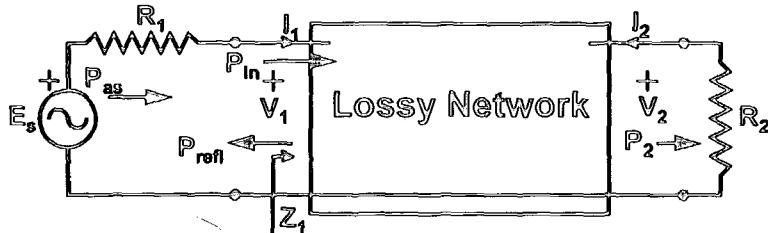


Figure 2.1.7. Power and reflection in a lossy, doubly-terminated network.

Mismatch loss, like SWR, is simply another scalar mapping of a reflection coefficient magnitude; mismatch loss (ML) is the ratio P_{in}/P_{as} in dB:

$$ML = -10 \log_{10} (1 - |\Gamma|^2) \text{ dB}. \quad (2.1.11)$$

Efficiency (η) is the ratio P_2/P_{in} in dB:

$$\eta = -10 \log_{10} (P_{in}/P_2) \text{ dB}. \quad (2.1.12)$$

It follows that the *insertion loss*, P_2/P_{as} , is equal to the sum of mismatch loss and efficiency in dB.

2.1.5 Singly-Terminated Networks

Figure 2.1.8 shows that singly-terminated two-port networks have a resistance at only one port, and there can also be a reactance associated with that resistance. The other port of a singly-terminated network is terminated by an ideal voltage or current source. As shown in Figure 2.1.8, an ideal voltage source must be adjacent to a series network element (E). If the adjacent branch were in parallel, then it would be useless; the branch voltage would be fixed by the ideal voltage source.

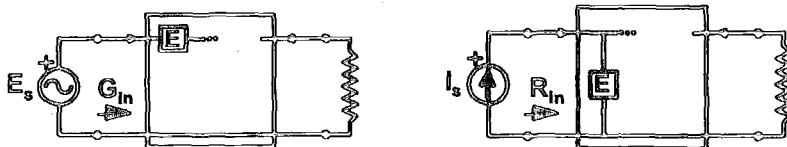


Figure 2.1.8. Singly-terminated filters have a resistance at only one port.

The similar situation with an ideal current source is also shown in Figure 2.1.8, where the adjacent network element (E) must be in parallel.

The power available from either kind of ideal source is infinite, so the power delivered to the load cannot be normalized to P_{as} like the doubly-terminated case. Figure 2.1.9 shows a singly-terminated network designed so that the input impedance is simply resistance R_4 at some reference frequency, say ω_0 . It is easy to show that R_4 establishes the

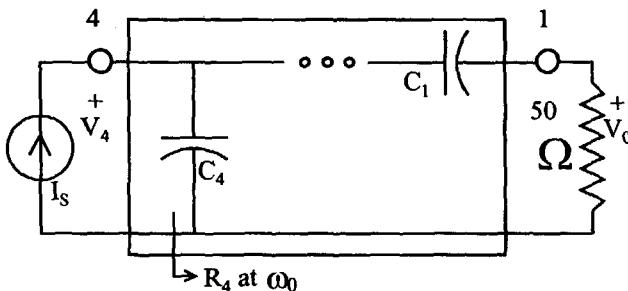


Figure 2.1.9. A singly-terminated network having input impedance R_4+j0 at ω_0 .

magnitude of transfer response V_4/V_0 at ω_0 . The power delivered to the network by the source is $|Is|^2 \times R_4$, and the power reaching the load is $|V_0|^2/R_0$. If the network is lossless, those two power quantities must be equal. Therefore, noting that $V_4=Is \times R_4$,

$$\left| \frac{V_4}{V_0} \right| = \sqrt{\frac{R_4}{R_0}}. \quad (2.1.13)$$

The dual case with an ideal voltage source requires that the input admittance be pure real (a conductance) at a reference frequency in order to establish a 0 dB response level.

Another consideration that is especially important for singly-terminated networks is the *Reciprocity Theorem*. As in Figures 2.1.9 and 2.1.10, the reciprocity theorem says that the voltmeter that reads V_0 can be swapped with the current source Is , and the voltage transfer ratio will not change. Similarly, an ideal voltage source swapped with an ammeter will not change the transfer ratio. Reciprocity requires a network designed for open-circuit voltage response to have the same short-circuit current response, and vice versa. For example, the network designed as

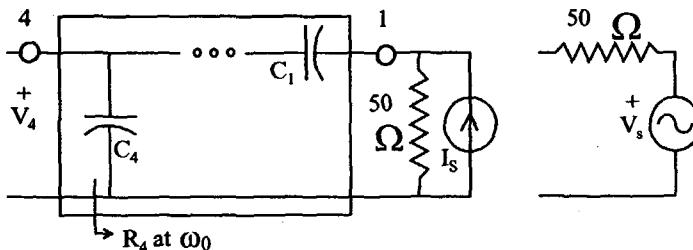


Figure 2.1.10. Voltage response equal to the current response in Figure 2.1.9.

in Figure 2.1.9 can be used as in Figure 2.1.10, where the resistive source can be either the Norton equivalent current source or the Thevenin equivalent voltage source shown to the right. An application for the network in Figure 2.1.10 could be as a preselector for a voltage-controlled operational amplifier.

Note that applying the reciprocity theorem to Figure 2.1.5 shows that doubly-terminated networks can always be turned end-for-end without any effect on the frequency response.

2.2 Major Response Shapes

The response shapes or frequency selectivity characteristics for many of the filters and matching networks in this book are derived from the defining lowpass response starting at dc. Then the lowpass response is translated, scaled and reflected to create a comparable bandpass response characteristic. The process is more simple than it sounds, and familiarity with the few variations of passband and stopband shapes, including those on a Smith chart, clarifies choices that the RF designer commonly encounters.

2.2.1 Lowpass to Bandpass Transformation

Figure 2.2.1 shows a lowpass response shape normalized to 1

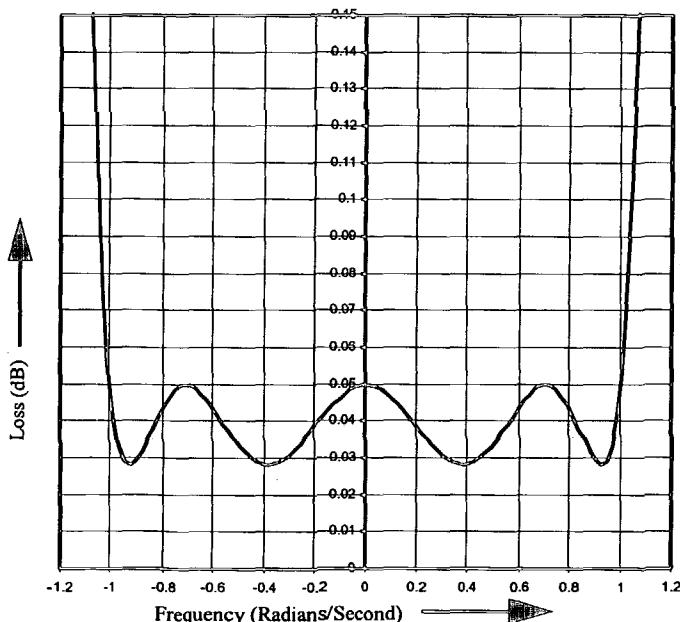


Figure 2.2.1. Normalized lowpass response with flat loss and ripple loss at dc.

radian per second (rad/s) at band edge. The negative frequencies from -1 rad/s to 0 (dc) are not always shown, because of the *arithmetic symmetry* of the lowpass response about dc. Of course, there can be any shape from dc to 1 rad/s; Figure 2.2.1 illustrates the equal ripple (Chebyshev) shape that has insertion loss at dc in addition to flat loss across the band.

Figure 2.2.2 shows a bandpass response shape derived from the

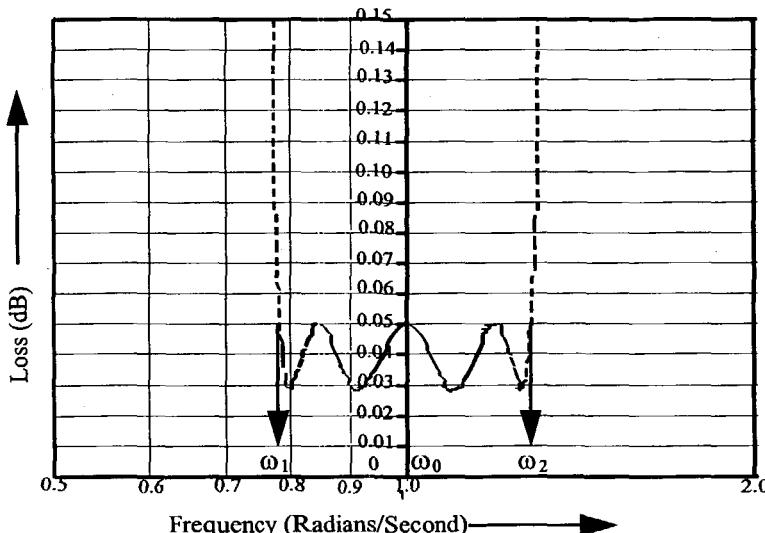


Figure 2.2.2. A bandpass response with geometric symmetry on a log abcissa.

lowpass shape in Figure 2.2.1. The bandpass shape is identical to the lowpass shape except that:

- The bandpass shape is normalized to 1 rad/s at band center.
- Band center frequency ω_0 is the geometric mean of band edges ω_1 and ω_2 , hence visual symmetry is obtained by giving the abscissa a logarithmic scale.
- The passband from ω_1 to ω_2 can be scaled to any width.

The geometric properties of the bandpass case require

$$\omega_0 = \sqrt{\omega_1 \omega_2}. \quad (2.2.1)$$

Also, it is both meaningful and convenient to define a *bandwidth quality factor*:

$$Q_{B\gamma} \equiv \frac{\omega_0}{\omega_2 - \omega_1} = \frac{100}{\%BW}. \quad (2.2.2)$$

When considering lowpass networks, $Q_{B\gamma} \approx 1$. A common calculation is the location of the passband edge given the % band width or $Q_{B\gamma}$:

$$\frac{\omega_2}{\omega_0} = \left[1 + \sqrt{(2Q_{BW})^2 + 1} \right] / (2Q_{BW}). \quad (2.2.3)$$

Generally, $\omega_0=1$ rad/s is assumed, which makes $\omega_1=1/\omega_2$ according to (2.2.1).

The LP-BP mapping described in this section is a standard *reactance transformation*; there are other reactance transformations. See [Daniels, 1974:Ch.6], [Cuthbert, 1983:Sect.6.6]. A more general method to obtain any equal-ripple bandpass response shape is described in Section 4.2.

2.2.2 Insertion Loss Behavior

The four passband response shapes shown in Figure 2.2.3 are often employed. The most common are the equal ripple (over-coupled) or

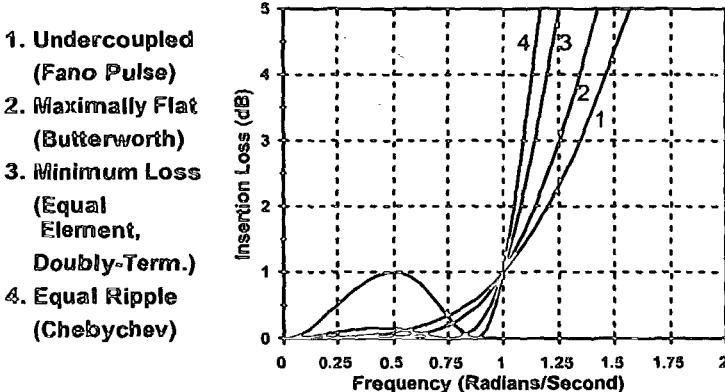
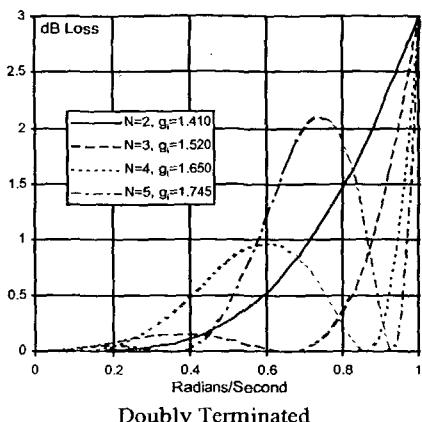


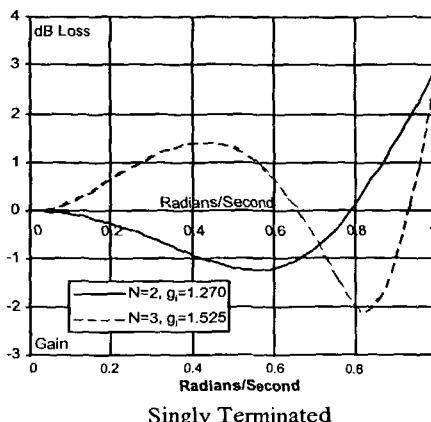
Figure 2.2.3. Four main passband response shapes.

Chebyshev shape and the maximally-flat or Butterworth shape. The undercoupled shape is less well known [Cuthbert, 1983:310-314], but it has a minimal transient overshoot characteristic. The undercoupled shape's transient characteristic is similar to the better known Bessel response, which could be used instead [Zverev]. All the stopband responses depicted in Figure 2.2.3 are all pole, i.e. monotonic without any ripples or zeros of transmission.

Many filter designers are familiar with the equal-element or minimum-loss shape, which results from making all prototype network element values equal. That also produces a network having minimum sensitivity as well as minimum loss at the reference frequency in the presence of dissipative network elements. The three-element ($N=3$) minimum-loss shape is shown in Figure 2.2.3. The useful shapes for doubly- and singly-terminated minimum-loss filters are shown in Figure 2.2.4. Unfortunately, there are only a few such shapes with acceptable passband ripple (less than 3 dB). The defining prototype g_i values shown in Figure 2.2.4 are discussed in Section 2.5.1.

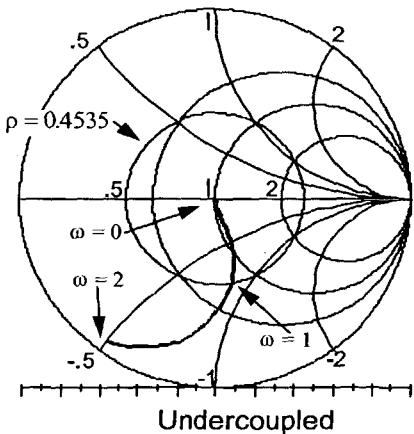


Doubly Terminated

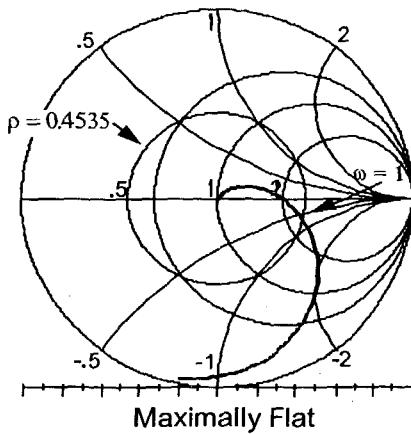


Singly Terminated

Figure 2.2.4. Equal-element LP filter responses normalized to 3-dB at 1 rad/s.



Undercoupled



Maximally Flat

Figure 2.2.5. Reflection coefficients for undercoupled and Butterworth shapes.

2.2.3 Reflection Coefficient Behavior

Clearly, the reflection coefficient in (2.1.2) controls the power transfer in lossless, doubly-terminated networks. Smith charts of the reflection behavior versus frequency provide a better understanding of how the filter produces its response. Figure 2.2.5 shows the input reflection coefficients of lowpass networks having undercoupled and maximally-flat responses. At dc, the filter is transparent to the equal source and load resistances, so the reflection is zero, i.e. the center of the Smith chart. Being normalized for a passband edge at 1 rad/s, these responses pass through a reflection magnitude circle of radius 0.4535 at that frequency, corresponding to $P_L = 0.7943 \times P_{as}$ (1.0 dB), according to (2.1.2). As frequency increases, the reflection coefficient locus goes to the

short-circuit side of the Smith chart, indicating that the input element of this particular lowpass filter is a shunt C as opposed to a series L.

Figure 2.2.6 shows the added flair for shapes with passband ripples. Starting from the chart center at dc again, the equal-ripple locus

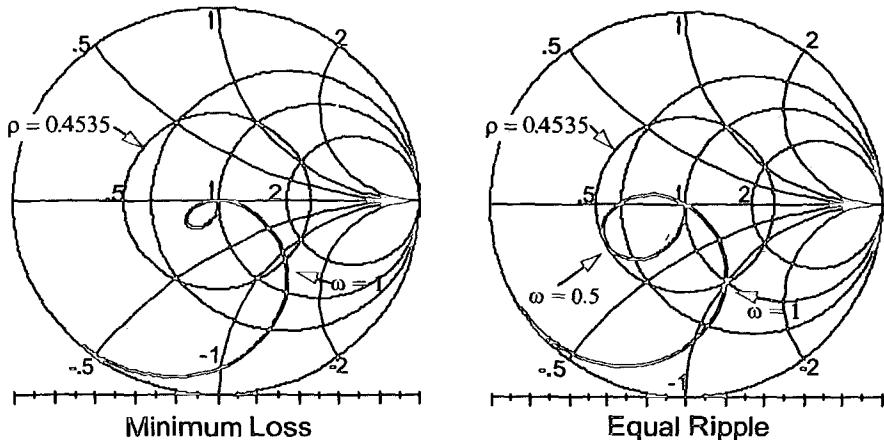


Figure 2.2.6. Reflection coefficients for minimum-loss and equal-ripple shapes.

loops out to the 1-dB reflection (0.4535 radius) and then goes back through chart center (0 dB) before again passing through the 0.4535 circle at 1 rad/s. Note the correspondence with the insertion loss behavior in Figure 2.2.3. Also, the equal ripple locus in Figure 2.2.6 starts at the center because an odd degree ($N=3$) Chebyshev filter was employed. Even-degree Chebyshev lowpass filters have the ripple value at dc (e.g. see Figure 2.2.1 for $N=4$), and thus have a reflection coefficient at dc that is on the real axis to one side of the Smith chart's center.

2.2.4 Flat Loss

Figure 2.2.7 shows an odd-degree equal-ripple response with added flat loss. This situation is typical of the broadband matching networks in Chapter Five. In this and other cases where there is loss at dc, the reflection locus starts to one side of chart center and remains within an annular ring whose radii correspond to the ripple extreme values in the passband. Comparison of the rectangular insertion loss graph and the Smith chart in Figure 2.2.7 shows that the locus starts tangent to the inner circle, becomes tangent to the outer circle at $\omega=0.5$, tangent to the inner circle again at $\omega=0.86$, and passes through the outer circle at $\omega=1$.

All loss at dc in lowpass networks is due to unequal terminating resistances. That loss can result from flat loss and/or the even-order Chebyshev ripple magnitude.

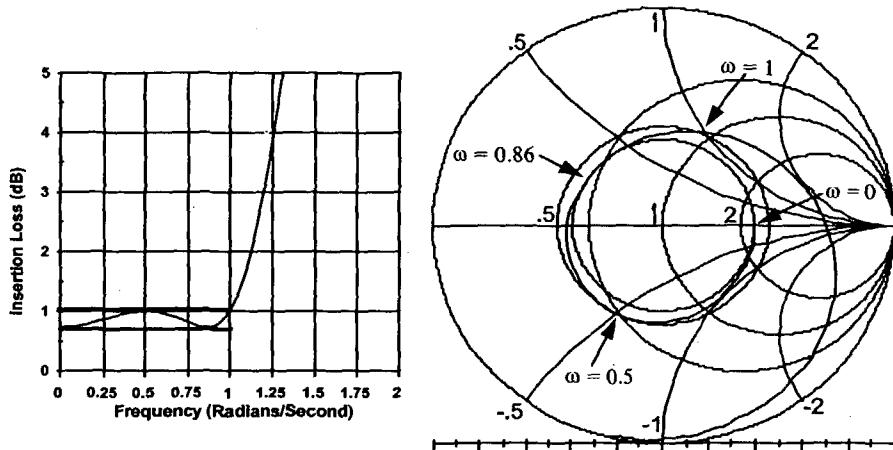


Figure 2.2.7. Broadband matching response behavior of reflection coefficient.

Example 2.2.1. Consider an input port connection as in Figure 2.1.7 where the input reflection coefficient, Γ , is defined by (2.1.10). Suppose that the two-port network is lossless and low pass, i.e. all shunt C's and series L's. **Problem:** Find the source and load resistances when $\Gamma=0.39+j0$ as in Figure 2.2.7. **Solution:** Solve (2.1.10) for Z_{in} :

$$Z_{in} = R_i \frac{1 + \Gamma}{1 - \Gamma} = R_i \left[\frac{2}{1 - \Gamma} - 1 \right]. \quad (2.2.4)$$

Using the given value of Γ in (2.2.4) shows that $Z_{in}=2.2787+j0$, i.e. when the source resistance is unity and the load resistance is 2.2787 ohms. That produces the 0.72 dB flat loss in Figure 2.2.7.

2.2.5 Stopband Ripple

Elliptic function responses are shown in Figure 2.2.8. The defining characteristic is the stopband ripple behavior which minimizes the transition between pass and stop bands. The most familiar shape provides ripples in the passband as well. However, a valuable passband shape is maximally flat while retaining the stopband ripples. That *inverse Chebyshev* shape provides docile transient response while retaining the benefits of stopband ripples. Unfortunately, there are only a few response nomographs available [Christian,1977:296]. The Cauer filter with ripples in both the pass band and stop band is very well documented [Zverev].

The presence of stopband ripple implies zeros of transmission (attenuation poles) at a finite number of stopband frequencies. These are

achieved by parallel LC "traps" connected in series or series LC traps connected in parallel in ladder networks, as described in Section 2.4.1.

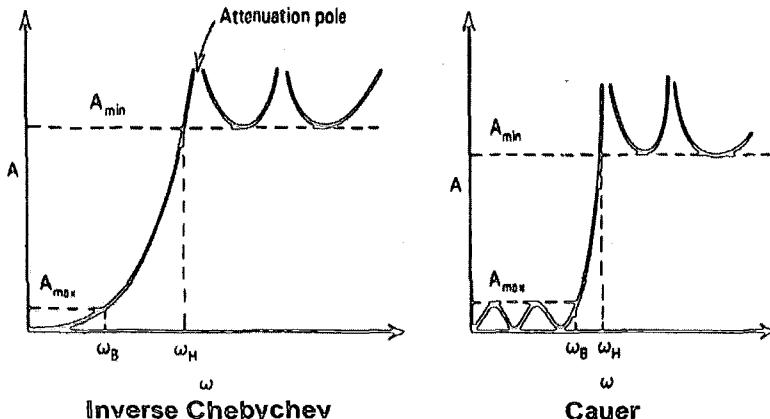


Figure 2.2.8. Inverse Chebyshev and Cauer elliptic filter responses.

2.2.6 Effects of Component Dissipation

The four most useful passband shapes, described in Section 2.2.2, are displayed in Figure 2.2.3 under the assumption of a lossless filter. When the elements are dissipative, as indicated by a finite element unloaded quality factor Q_u , the passband insertion loss increases, especially at passband edges. Figure 2.2.9 shows only the upper-half passband of a doubly-terminated minimum-loss (equal-element) filter of

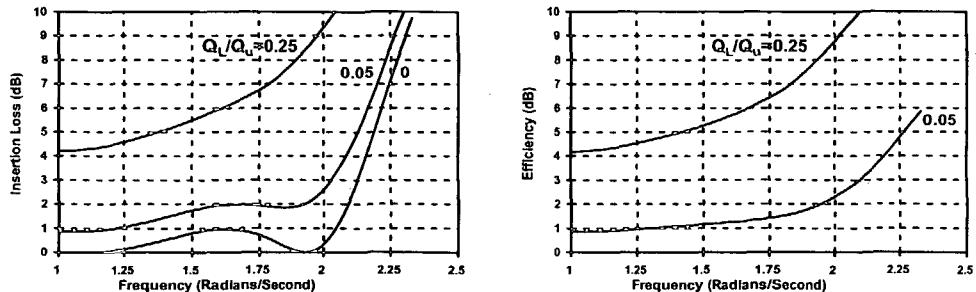


Figure 2.2.9. Upper pass band of an $N=4$ minimum-loss dissipative filter.

degree 4 and centered at 1 rad/s for $Q_u=\infty$, 20, and 4. The lossless filter passband edge is at $\omega=2.12$ rad/s corresponding to 3-dB loss; see Figure 2.2.4. The simple steps for designing this filter are presented in Section 2.4.2.

As noted in Section 2.1.4, the insertion loss, P_2/P_{as} , for doubly-terminated filters is the sum of input mismatch loss and efficiency, all in dB. For infinite Q_u , the response is entirely mismatch loss, because the efficiency loss is 0 dB. Reduced values of Q_u result in increased loss at midband; this is easily calculated as explained in Sections 2.3.3 and 3.3.4.

Figure 2.2.9 clearly shows worsening loss as the passband edge is approached, and the largest part of the insertion loss is seen to be in efficiency, not reflection. This dominance of efficiency continues into the stopband, but the insertion loss is nevertheless predicted by the lossless case with increasing accuracy when well removed from the pass band. The minimum-loss filter responses have been tabulated [Cuthbert, 1983:458], [Taub, 1963], [Taub, 1964] and show these trends clearly.

The minimum-loss or equal-element filter is important because it approximates an average of all other filters in terms of element values and performance. Therefore, the described dissipation effects apply to most filters and matching networks.

2.3 Significance of Loaded Q

The most obvious application of loaded Q (Q_L) is in the conversion between series and parallel forms of impedance. However, loaded Q can be found as a property of complex source and load terminations as well as at internal interfaces within ladder networks. Its physical meaning is the ratio of reactive (stored) to real power. In conjunction with Q_{BW} defined by (2.2.2), loaded Q is the main parameter in filter and matching network design. Certainly, loaded Q is the unifying parameter for this book and must be introduced before proceeding.

2.3.1 Series-Parallel Conversion

Figure 2.3.1 shows a resistance, R , associated with a reactance, X .

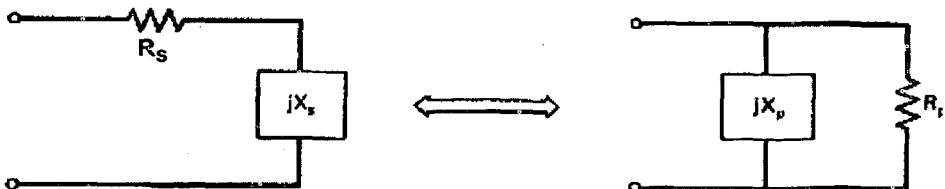


Figure 2.3.1. Series and parallel impedance forms equivalent at a frequency.

The reactance could be either inductive or capacitive and equal to ωL or $1/\omega C$, respectively. For purposes of loaded Q, the negative sign associated with capacitive reactance may be ignored. Subscripts indicate series (s) or parallel (p) connection. The application is to relate the resistances and reactances so that the same impedance at that particular frequency is obtained at both the series and the parallel terminals. For notation, it can be said that $R_p \parallel jX_p$, i.e. R_p is in parallel with jX_p .

For the equivalence to hold, the ratio of reactive to real power must be preserved in each of the two forms shown in Figure 2.3.1. The power ratio is defined as

$$Q_L \equiv \frac{VA}{W}, \quad (2.3.1)$$

where VA is the volt-amperes or reactive power, and W is the watts or real power. For current I entering the series circuit, the reactive power is $|I|^2 X_S$ and the real power is $|I|^2 R_S$. Similarly for voltage V across the parallel circuit, the reactive power is $|V|^2/X_P$ and the real power is $|V|^2/R_P$. Therefore, no matter what values of I or V exist at the terminals, (2.3.1) shows that

$$Q_L = \frac{X_S}{R_S} = \frac{R_P}{X_P}. \quad (2.3.2)$$

It is much more convenient to employ loaded Q in the conversion between forms in Figure 2.3.1 than the more fundamental relationship between admittance Y and impedance Zs:

$$Y = G + jB = \frac{1}{Z_S} = \frac{R_S}{R_S^2 + X_S^2} + j \frac{-X_S}{R_S^2 + X_S^2}. \quad (2.3.3)$$

It is conventional to use parallel resistance rather than conductance ($R_P=1/G$) and parallel reactance rather than susceptance ($X_P=1/|B|$). In those terms, (2.3.3) shows that

$$R_P = R_S(1 + Q^2), \quad (2.3.4)$$

and the same expression solved for Q is

$$Q = \sqrt{\frac{R_P}{R_S}} - 1. \quad (2.3.5)$$

It is important to note that (2.3.5) requires $R_P > R_S$ in every practical case. The main results of this section are contained in (2.3.2), (2.3.4), and (2.3.5). These important equations are used countless times and should be committed to memory.

Example 2.3.1. Convert the series impedance $20-j10$ ohms to parallel form. *Problem:* Find R_P and X_P . *Solution:* From Figure 2.3.1, $R_S=20$ and $X_S=-10$ ohms. By (2.3.2), $Q=0.5$; by (2.3.4) $R_P=25$; and by (2.3.2) again, $X_P=-50$ ohms. Note that both the series and parallel equivalents are capacitive, as denoted by prefixing the negative sign after the formulas were employed.

2.3.2 Resonator VA/W

Resonators are composed of an L and C, connected either in parallel or in series, and resonant at some desired frequency, ω_0 :

$$\omega_0 = \frac{1}{\sqrt{LC}}. \quad (2.3.6)$$

Figure 2.3.2 shows a parallel resonator terminated in a parallel resistance R_P . If the resonator is resonant at ω_0 , then the inductive and capacitive reactances are equal in magnitude at ω_0 :

$$X_P = \omega_0 L = \frac{1}{\omega_0 C}. \quad (2.3.7)$$

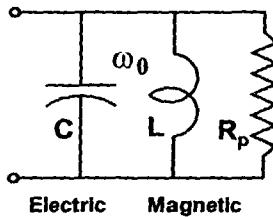


Figure 2.3.2. Loaded Q in a parallel resonator.

The stored energy (VA) oscillates between the inductor and the capacitor, like a balance wheel and hairspring in a watch or a mass bobbing up and down on a spring, and the real power is delivered to R_p .

The voltage across any of the three elements is the same, but by (2.3.2), $X_p = R_p/Q$. Therefore, there is Q times as much current through the L and through the C as there is through R_p . The stress magnification in high Q circuits often cannot be ignored. The same analysis applies to series RLC resonators, except that there is voltage magnification by the Q factor.

2.3.3 Resonator Efficiency

In addition to loaded Q , Q_L , which involves real power delivered outside a resonator, there is unloaded Q , Q_u , which involves real power lost in the L and C , mainly in dissipative inductances. Concerning the resonator in Figure 2.3.2, when R_p is disconnected, then Q_u is the ratio of volt-amperes reactive power to that real power dissipated within the mostly reactive elements.

There is a simple and fairly accurate formula for the power lost in mildly dissipative (lossy) resonators:

$$L_0 \approx 4.34 \frac{Q_L}{Q_u} \text{ dB.} \quad (2.3.8)$$

L_0 is the approximate power loss in dB per resonator at the resonance frequency ω_0 . It is significant that bandpass filter midband dissipation loss is inversely proportional to unloaded Q ; e.g., to halve the midband loss in dB, Q_u must be doubled.

Example 2.3.2. Suppose that four resonators like that in Figure 2.3.2 are coupled in cascade by a nearly lossless coupling arrangement. Also suppose that each resonator has a loaded Q , Q_L , of unity. *Problem:* Find the midband dissipative loss (i.e. efficiency) in dB when the loaded Q , Q_u , is either 20 or 4. *Solution:* Using 4 times the value found by (2.3.8) gives $L_0 \approx 0.87$ dB for $Q_u=20$ and $L_0 \approx 4.34$ dB for $Q_u=4$. These are the cases shown in the graphs in Figure 2.2.9, where $\omega_0=1$ rad/sec. There is no flat (reflection) loss at that band center frequency.

2.4 Ladder Network Topologies

The interconnection of elements or components in a filter or matching network is called its *topology*. Lowpass and various types of bandpass networks have characteristic patterns in their topologies. Topologies of bandpass networks can be altered by replacing certain subsets of elements by the same or increased number of elements so as not to affect the frequency response of the network. Those topologies and subsets are introduced in this section.

2.4.1 Lowpass Prototype Networks

Figure 2.4.1 shows the two possible lumped-element topologies for lowpass networks that have a monotonic stopband (as shown in Figure 2.2.1). These networks are duals of one another as described in Section

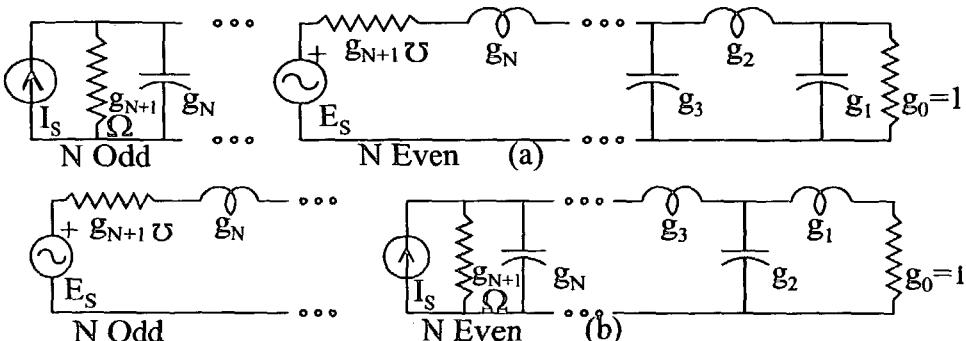


Figure 2.4.1. Dual all-pole lowpass prototype networks.

2.4.5. The distinguishing topological feature is with the element next to the load resistance; it is either in series or in parallel. The input element is also in series or parallel, depending on the degree of the filter, which is equal to the number of elements. All series elements are inductors with reactances that increase with increasing frequency and parallel capacitors with susceptances that increase with increasing frequency. Elements are numbered from load to source; the g_i values are in henrys and farads for L's and C's, respectively. The prototype load resistance, g_0 , is always unity.

The source immittance real part, g_{N+1} , is shown in parallel ohms or in series mhos and may not be unity. This convention may seem awkward, but it accommodates the common practice of assigning $g_{N+1}=\infty$ for singly-terminated networks. When g_{N+1} is finite, the Thevenin voltage source with series conductance and the Norton current source with parallel resistance are equivalent at the network's input terminals and can be interchanged at will.

Figure 2.4.2 shows dual Cauer or elliptic-function lowpass prototype networks. The topology conventions from Figure 2.4.1 also apply. The distinguishing feature is the inclusion of capacitors in parallel

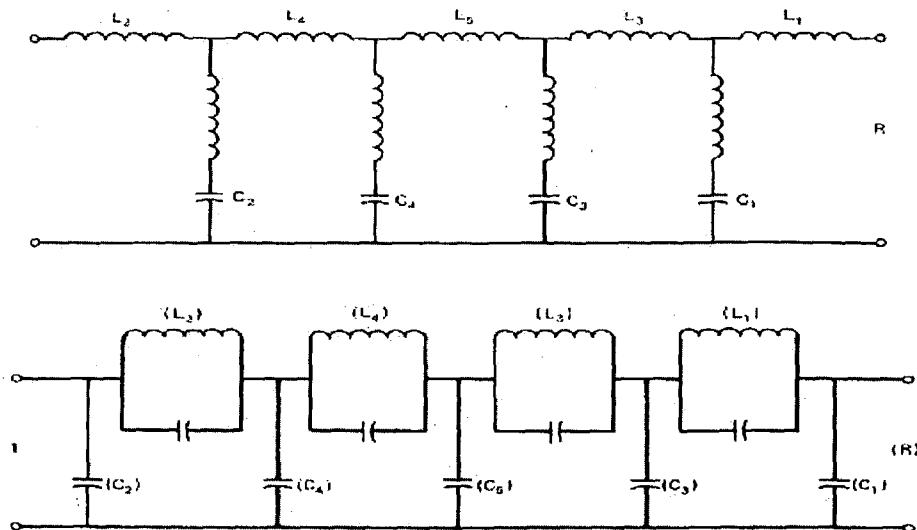


Figure 2.4.2. Dual Cauer (elliptic-function) lowpass prototype networks.

with some or all of the series inductances, or inductors in series with some or all of the parallel capacitors. These LC branches cause zeros of transmission at their resonance frequencies in the stop band, as illustrated in Figure 2.2.8.

For classical broadband impedance matching, it is important to recognize that the complex load impedance includes element g_1 as well as the resistance g_0 in Figure 2.4.1. According to (2.3.2), the loaded Q of a lowpass broadband matching load is always

$$Q_L = g_1 \times g_0. \quad (2.4.1)$$

When there is also a complex source impedance (the double match case), the source impedance includes element g_N as well as the resistance or conductance g_{N+1} . Again, (2.3.2) shows that the loaded Q of a lowpass broadband matching source is always

$$Q_S = g_N \times g_{N+1}. \quad (2.4.2)$$

As is described in Section 2.5.2, computer program ALLCHEBY accepts an assigned value for Q_L , and perhaps Q_S as well, and then find all element values, g_i , to obtain the least-possible insertion loss over the passband.

2.4.2 Classical Bandpass Prototype Networks

It is very easy to convert lowpass prototype networks into bandpass prototype networks. Figure 2.4.3 shows that every lowpass capacitor is replaced with a parallel LC resonator, and every lowpass inductor is replaced with a series LC resonator. The parallel resonators have $C=Q_{BW} \times g_i$ and $L=1/C$, where Q_{BW} is defined by (2.2.2). The series

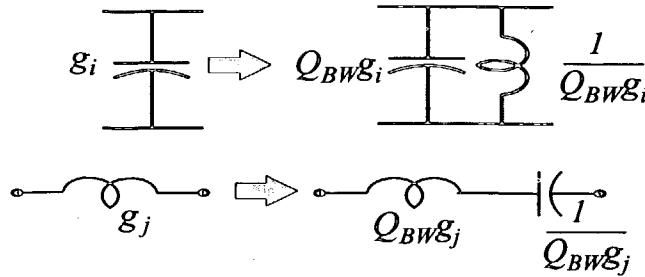


Figure 2.4.3. Reactance transformations: lowpass elements to bandpass resonators.

resonators have $L=Q_{BW}\times g_i$ and $C=1/L$. This tunes all resonators to $\omega_0=1$ rad/s.

It is worth remarking that the replacement of lowpass branches by bandpass resonators corresponds to a transformation of frequency scale. Referring to Figure 2.2.2, if the lowpass frequency variable is ω' and the bandpass frequency variable is ω , then

$$\omega' \leftarrow Q_{BW} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right), \quad (2.4.3)$$

i.e., ω' is replaced by the right-hand expression in (2.4.3). Multiplying both sides of (2.4.3) by g_i gives the reactances or susceptances of the related branches in the lowpass and bandpass domains; these are zero at their respective center frequencies. The bandpass group delay response differs considerably from the lowpass group delay because of the nonlinearity (versus frequency) of (2.4.3).

Figure 2.4.4 shows the result of setting $N=4$ and using the lowpass prototype in the lower right part of Figure 2.4.1, which has the g_1 inductor in series with the load resistance g_0 . The variable name Q_i is employed in Figure 2.4.4:

$$Q_i = Q_{BW} \times g_i, \quad i = 1 \text{ to } N. \quad (2.4.4)$$

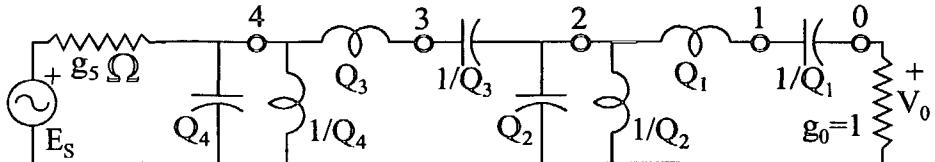


Figure 2.4.4. A four-resonator classical bandpass filter.

The significance of Q_i is that it is loaded Q . For example, the inductor labeled Q_1 in Figure 2.4.4 is in series with the load resistance $g_0=1$, and Q_1 henrys at 1 rad/s has a reactance of Q_1 ohms. By (2.3.2), Q_1 is the loaded Q of that series resonator at the band center frequency of 1 rad/s.

At $\omega_0=1$ rad/s, all series resonators are short circuits and all parallel resonators are open circuits. Therefore, consider the parallel resonator labeled Q_2 in Figure 2.4.4. Looking toward the load, it also sees the load resistance, $g_0=1$ ohm. By (2.3.2), Q_2 is the loaded Q of that parallel resonator at the band center frequency of 1 rad/s. Similar statements can be made concerning resonators Q_3 and Q_4 . They are *singly-loaded* Q's, i.e., any resistance loading a resonator toward the source is NOT considered. For example, in Figure 2.1.1 the singly-loaded Q of Z_L is X_L/R_L , whereas the doubly-loaded Q is $X_L/(R_L+R_S)$. Only singly-loaded Q's are employed in this book.

The L and C conversions from lowpass branches to bandpass branches in Figure 2.4.3 also apply to the Cauer topologies in Figure 2.4.2. For the case where capacitors have been added in parallel with series inductors, as in the lower part of Figure 2.4.2, the transformed bandpass branch appears as in Figure 2.4.5. For the case where inductors

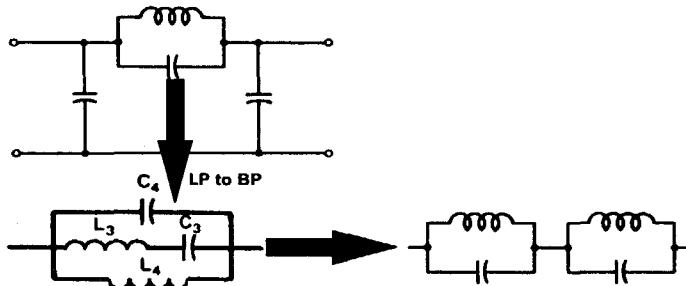


Figure 2.4.5. Conversion of series lowpass traps to bandpass branches.

have been added in series with parallel capacitors, as in the upper part of Figure 2.4.2, the related bandpass branch appears as in Figure 2.4.6.

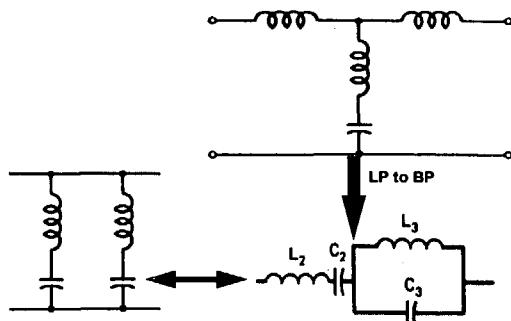


Figure 2.4.6. Conversion of parallel lowpass traps to bandpass branches.

There are various other equivalent topologies for these bandpass branches [Cuthbert, 1983:470-3], but they are not useful in direct-coupled filters as developed in Section 3.4.3.3.

There are high impedance levels in the middle of the series resonators at nodes 1 and 3 in Figure 2.4.4 when the Q_i 's are much greater than unity. Loaded Q's may exceed unity when the passband width is less than 100%, according to (2.2.2) and (2.4.4), because the g_i values vary about unity. At the midband frequency of 1 rad/s, (2.3.4) shows that the parallel resistance at node 1 in Figure 2.4.4 looking toward the load, for example, is $R_p=(1+Q_1^2)$ and $X_p=-(1+Q_1^2)/Q_1$. For $Q_1 \gg 1$, $R_p \approx Q_1^2$ and $X_p \approx -Q_1$, which can cause several problems:

- o Midband voltages to ground at series nodes 1 and 3 are $(1+Q^2)^{1/2} \times V_0$, where V_0 is the load voltage. These voltages appear across equivalent parallel resistances as described and scale with load voltage because of power conservation in this lossless network,
- o Stray capacitance to ground from nodes 1 and 3 is likely to exist in the physical network. If the normalized stray capacitance is not considerably less than $1/Q$ farads, then the filter impedance levels will be significantly different from what is required, and
- o The respective ratios of series to parallel L values are $Q^2:1$; the same is true of the ratio of extreme values of C as well.

Example 2.4.1. Suppose that the classical bandpass network topology for $N=4$ in Figure 2.4.4 is a doubly-terminated minimum-loss filter having 20% 3-dB bandwidth. **Problem:** Find the voltages to ground at nodes 1 and 3 relative to V_0 , and find the equivalent parallel reactance looking toward the load from those nodes, all at the midband frequency of 1 rad/s. **Solution:** From Figure 2.2.4, $g_i=1.650 \forall$ ("for all") i. By (2.2.2), 20% bandwidth implies $Q_{BW}=5$, and, by (2.4.4), $Q_i=8.25 \forall$ i. By (2.3.4), the equivalent parallel resistance seen toward the load from nodes 1 and 3 is 69.06 ohms. Voltages across parallel resistances are proportional to the square root of resistance, so the voltages to ground from nodes 1 and 3 are $\sqrt{69.06} \times V_0 = 8.31 \times V_0$.

2.4.3 Direct-Coupled Prototype Bandpass Networks

Figure 2.4.7 shows four direct-coupled parallel resonators; in general there can be any number of resonators. The " \wedge " symbol stands

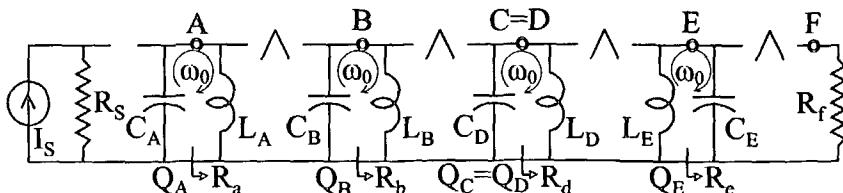


Figure 2.4.7. Direct-coupled network topology with only parallel resonators.

for an L or a C or a parallel-LC-trap coupling branch in series. A top coupling “^” between nodes E and F might also appear between the source and node A, or there might be no terminal top couplings at all. An even more specific example is shown in Figure 2.4.8. It is often unnecessary to have an equal number of L and C couplings in narrow-band direct-coupled filters; however, coupling constraints for the broadband case are described in Section 3.6. A top coupling L is also equivalent to using an RF transformer; see Section 2.4.4.

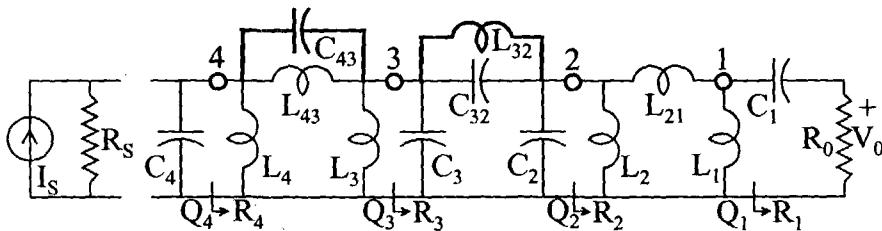


Figure 2.4.8. An $N=4$ direct-coupled elliptic filter with LCCLC couplings.

Corresponding to the parallel resonator topology in Figure 2.4.7 is a “dual” case that employs only series resonators and L, C, and series-LC-trap coupling branches in parallel. The topology using series resonators is often advantageous in low impedance environments. The simple method for obtaining dual networks is described in Section 2.4.5.

The direct-coupled network using only parallel resonators in Figure 2.4.7 can overcome the disadvantages just listed for the classical bandpass network in Figure 2.4.4 when passband widths are less than 100%:

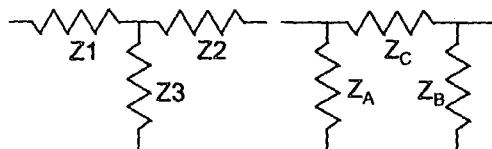
- Top coupling branches in Figure 2.4.7 enable control of the midband parallel resistances R_a , R_b , etc. and therefore control of node voltages to ground,
- Any stray capacitance to ground is absorbed into a resonator, and
- The range of extreme values of L's and of C's can be minimized because of the wide range of feasible parallel resistances that do not affect the network's selectivity.

2.4.4 Transformations in Bandpass Networks

Lowpass filter topologies, Figures 2.4.1 and 2.4.2, do not present two elements of like kind alone in adjacent branches, but that does occur in bandpass topologies. For example, in Figure 2.4.8, there are three L's alone in adjacent branches: L_1 , L_{12} , and L_2 . This occurrence allows substitution of an equivalent subsection that does not affect the

frequency response. The substituted subsection may offer a more desirable topology or more acceptable element values.

The T-to-Pi and Pi-to-T (also called star-delta) transformations involve three elements of like kind. Figure 2.4.9 shows the subsection topologies, where the Z's could be all L's or all C's. Figure 2.4.10 shows a transformation for three L's in a Pi (or an equivalent T) that is equivalent to an RF transformer with primary and secondary windings. A table of these transformations with equations is available [Zverev:529].



$$Z_T \equiv Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3,$$

$$Z_S \equiv Z_A + Z_B + Z_C,$$

$$Z_C = Z_T / Z_3, \quad Z_1 = Z_A Z_C / Z_S$$

$$Z_A = Z_T / Z_2, \quad Z_2 = Z_B Z_C / Z_S$$

$$Z_B = Z_T / Z_1, \quad Z_3 = Z_A Z_B / Z_S.$$

Figure 2.4.9. Equivalent T and Pi configurations.

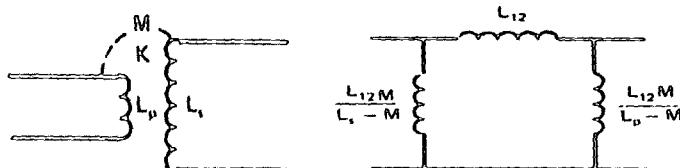


Figure 2.4.10. An RF transformer and its equivalent inductive Pi subsection.

The Norton transformation involves two L's or two C's alone in adjacent branches. Figure 2.4.11 shows that if an ideal transformer is inserted to one side of the two C's, then that is equivalent to three capacitors. A similar el section of inductors and a transformer has an

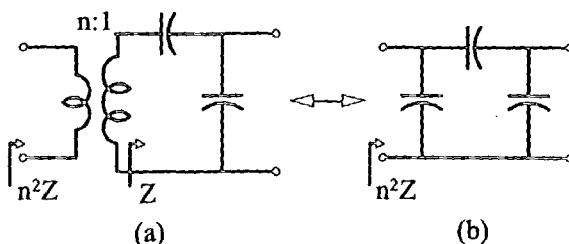


Figure 2.4.11. A capacitive Norton transformation.

equivalent Pi of inductors. All Norton transformations preserve the exact frequency response. The transformer turns ratio, n , depends on the values in the el section but is restricted to a limited range above or below unity for all elements in the equivalent Pi to be positive. Insertion of a transformer in a network requires impedance rescaling by n^2 of all networks elements to one side of the transformer.

For example, Norton transformations can be applied to the network in Figure 2.4.4. Note the el section composed of the capacitors having values $1/Q_3$ and Q_2 . An ideal transformer can be inserted to the left of that el section at node 3, requiring values of inductors Q_3 and $1/Q_4$ and resistance g_5 to be decreased and capacitor Q_4 increased by the square of the turns ratio, n^2 . Insertion of the transformer creates the Norton transformation between nodes 2 and 3 depicted in Figure 2.4.11(a). That section can be replaced by the Pi section shown in Figure 2.4.11(b). Additionally, a second Norton transformation could be applied to the two adjacent inductors in Figure 2.4.4 that originally had values $1/Q_4$ and Q_3 with further adjustments of impedance levels. Formulas for the three C's and three L's are straightforward [Borlez:84], but can be avoided entirely by the techniques in Section 3.4. Piecemeal applications of Norton transformations are always possible, but it is difficult to visualize potential benefits to ranges of element values and current or voltage levels in large networks [Zverev:530-533].

Also, four-element branches consisting of two L's and two C's occur in bandpass elliptic function filters to produce pairs of transmission zeros. Replacement of those branch topologies is discussed in Section 2.5.3.

2.4.5 Duality

Every LC network has a dual topology that has identical responses. For example, the two lowpass networks in Figure 2.4.1 are duals. There are a few simple rules to obtain a dual network:

- Change every parallel branch into a series branch and every series branch into a parallel branch,
- Change those elements that are in series within a branch to be in parallel and those elements that are in parallel within a branch to be in series (compare Figures 2.4.5 and 2.4.6),
- Change each inductor to a capacitor and vice versa while retaining its numerical value; i.e. 3 H becomes 3 F,
- Change each resistance to a conductance and vice versa while retaining its numerical value; e.g. 5 ohms becomes 5 mhos, and
- Change voltage sources into current sources and vice versa.

It is stated in Section 2.4.3 that nodal (parallel resonators) direct-coupled networks have a mesh (series resonators) dual. For example, the dual of the network in Figure 2.4.8 appears in Figure 2.4.12, where all the rules given above apply.

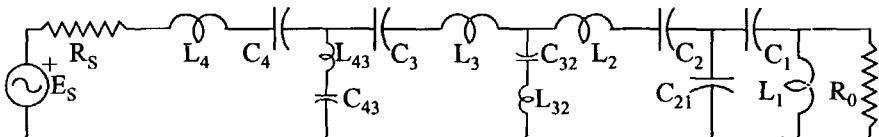


Figure 2.4.12. The mesh network with series resonators dual to Figure 2.4.8.

2.5 Network Component Values

The conventional way to design filters is to start with some lowpass prototype network element values, g_i , for use as a lowpass network or for conversion to a bandpass network. There are some advantages to just making all the g_i equal. But for greater variety of response shapes there are numerous tables of prototype lowpass element values, some being normalized to 1 rad/s where the ripple ends as in Figure 2.2.1 [Matthaei, 1964:100-102;109], and others to 3 dB no matter what the ripple and/or flat loss [Zverev, 1967:312-340], [Williams, 1991:Chap.11]. It is also possible to use an arbitrary $\{g_i\}$ set.

A convenient program is described that generates all cases of equal-ripple passband responses for all-pole (monotonic stopband) filters. Cauer filters and the similar inverse Chebyshev filters can be designed by tables and programs that provide lowpass prototype element values. Comprehensive equal-ripple filters having arbitrary zeros described in Chapter Four must be designed by polynomial synthesis. Examples are provided for the classical filters; design of direct-coupled and broadband matching networks is treated in Chapters Three and Five, respectively. In every case, frequency and impedance scaling are required.

In the case of narrow-band, direct-coupled networks, it is possible to start with the bandpass structure and assign somewhat arbitrary component values that are easy to calculate, as explained in Chapter Three.

2.5.1 Equal-Element Lowpass Prototype

It is mentioned in Section 2.2.2 that the equal-element or minimum-loss response shape resulted from making all the g_i elements in the lowpass prototype network in Figure 2.4.1 identical. Although sometimes used in the lowpass topology, more often that is a step toward design of bandpass filters, in either the classical topology, Figure 2.4.4, or the direct-coupled topology, Figure 2.4.7. In bandpass cases, equal-element designs mean equal loaded Q's and equal resonator efficiencies for uniform dissipation (Section 2.3.3).

For either doubly- or singly-terminated networks having equal elements, the response shapes can be calculated and displayed in an organized way [Cuthbert, 1983:314-319]. However, this class of filters is unique in that the frequency response shapes are easily displayed as functions of Q_L/Q_u , i.e. in the presence of even large dissipation quality factors [Taub, 1963, 1964]. As shown in Figure 2.2.4, only low-degree, lossless equal-element filters have a reasonable passband ripple value, especially for the singly-terminated filters.

Example 2.5.1. Given a lossless four-resonator bandpass filter like that in Figures 2.4.4 or 2.4.7, suppose that a passband width of 165% is required. *Problem:* Find the resonator loaded Q's for the doubly-terminated equal-element shape. *Solution.* According to (2.2.2), $Q_{BW}=0.6061$. Figure 2.2.4 shows that $g_i=1.650$, so that (2.4.4) yields $Q_i=1.000 \forall i$. The bandwidth was selected so that each element in the bandpass prototype network in Figure 2.4.4 is equal to unity, as are all the resonator loaded Q's in Figure 2.4.7.

2.5.2 Program ALLCHEBY.EXE

The equal-ripple response shape shown in Figures 2.2.1 and 2.2.2 is often required for both filter and broadband impedance matching problems. Either network category could be lowpass or bandpass and doubly or singly terminated, and might have flat loss as well as ripple. However, the broadband matching problem assigns a loaded-Q value for one end of the network and perhaps for the other end as well; see (2.4.1) and (2.4.2). In that case, there will be flat loss obtained by unequal terminating resistances.

All these various cases are solved by program ALLCHEBY.EXE as displayed by the flowchart in Figure 2.5.1. The program starts with the option of asking for two passband edge frequencies in any units; then the %BW and normalized ($\omega_0=1$) bandpass edge frequencies are returned. The problem is either filtering or broadband matching. The various two-port termination combinations are illustrated in Figure 2.5.1. In particular, infinite Q_s is a singly-terminated case that matches a resistance or loaded resonator to an ideal source over a frequency band. In all doubly-terminated cases, the ripple and flat loss are determined in terms of dB insertion loss, reflection return loss, and SWR. For broadband matching cases, the least-possible insertion loss when using a matching network having infinite complexity ($N \rightarrow \infty$) is given to add significance to the result for the selected value of N , $N < 15$.

All g_i and Q_i values are computed assuming load resistance $g_0=1$. Also, the geometric mean loaded Q is reported as an average stored-energy indicator approximating an equal-element network. (The significance of loaded-Q product is described in Section 3.3.1.)

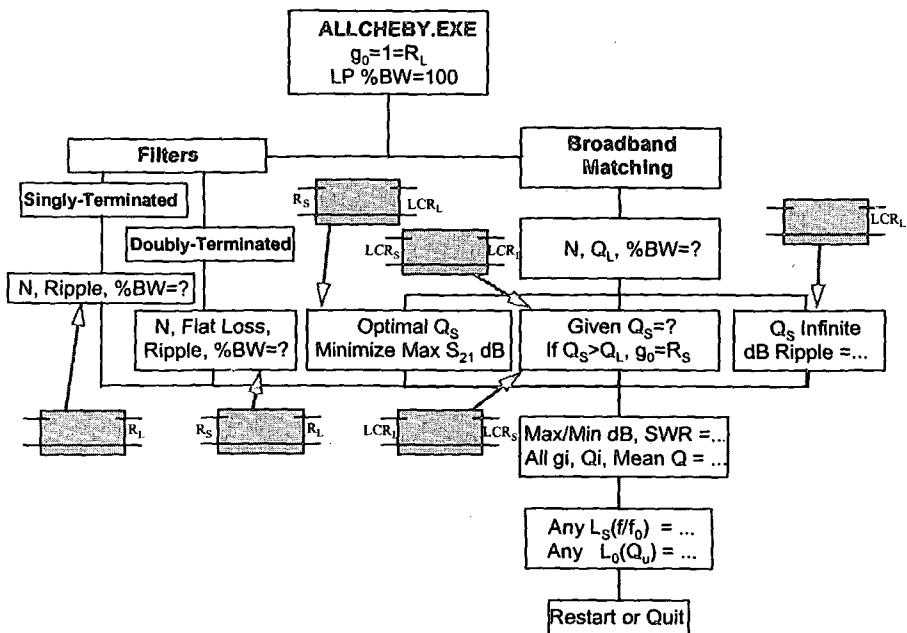


Figure 2.5.1. Flow chart of filter and matching program ALLCHEBY.EXE.

The exact Chebyshev insertion loss for lossless networks can be calculated for any given frequency. The normalized bandpass frequencies for 20-dB insertion loss are reported automatically to quantify the passband-to-stopband transition. For any solution, ALLCHEBY can estimate the midband dissipative loss given a uniform unloaded Q , Q_u , according to (2.3.8).

The theoretical origin and details of program ALLCHEBY are provided in Section 5.2. Two anomalies that arise in double-matching situations are discussed in Section 5.2.5. The emphasis here is on the ease of obtaining equal-ripple networks without resorting to tables and nomograms that may not be available or even exist.

Example 2.5.2. The requirement is for a four-resonator 0.1 dB equal-ripple filter having 50% bandwidth and the classical topology shown in Figure 2.4.4. *Problem:* Find all element values for the normalized filter. *Solution:* Program ALLCHEBY.EXE produced the data shown in Table 2.5.1. The lowpass prototype corresponds to the lower half of Figure 2.4.1, because the resonator next to the load is in series with the 1-ohm resistor. In Table 2.5.1, the corresponding left-hand sequence L_{ser} , C_{par} , etc. ends in Ohms for $G(5)$, i.e. $g_5=1.3554$ ohms. There is a ripple loss peak at the midband ω_0 in the bandpass topology corresponding to dc in the lowpass topology. Because the respective network input

impedances at those frequencies are 1 ohm, it is easy to confirm the 1-dB ripple using (2.1.10) and (2.1.11) with $R_1=g_5$.

Table 2.5.1. ALLCHEBY.EXE Output for Example 2.5.2.

Is this a Matching or a Filter (QL=0=QS) network (,_F)? F
Is this filter Singly- or Doubly-terminated (S,_)? D
N, dB FLAT LOSS, dB RIPPLE, %BW =? 4,0,.150
THIS INSERTION LOSS FROM 0.0000 TO 0.1000 dB
THIS dB FLAT LOSS = 0.0000
THIS dB RIPPLE = 0.1000
SWR FROM 1.0000 TO 1.3554
RETURN LOSS FROM 120.0000 TO 16.4276 dB
Lser or Cpar G(1) = 1.1088 Q(1) = 2.2176
Cpar or Lser G(2) = 1.3062 Q(2) = 2.6124
Lser or Cpar G(3) = 1.7704 Q(3) = 3.5407
Cpar or Lser G(4) = 0.8181 Q(4) = 1.6362
Ohms or Mhos G(5) = 1.3554
GEOMETRIC MEAN LOADED Q = 2.4069

2.5.3 Elliptic Function Lowpass Prototype

The lowpass prototype network topology for Cauer and Inverse Chebyshev filters is shown in Figure 2.4.2; these are filters that have equal-ripple or maximally-flat passbands, respectively, and poles of attenuation in the stop band. Lowpass Cauer filters can be designed using tables [Zverev] and programs [Cuthbert,1983:358-362], and lowpass inverse Chebyshev filters can be designed using some less common tables [Christian,1975:144], [Taylor].

Conversion of lowpass elliptic filters to classical bandpass forms is accomplished by replacing C's with parallel LC's and replacing L's with series LC's as shown in Figures 2.4.3, 2.4.5, and 2.4.6. An added difficulty with the bandpass case results from the branches that produce the poles of attenuation in the stopband. These pairs of frequencies, say ω_- and ω_+ , are geometrically related to the band center frequency ω_0 :

$$\omega_- \times \omega_+ = \omega_0^2. \quad (2.5.1)$$

Each pair of bandpass "trap" frequencies is found from the related lowpass "trap" frequency, Ω_∞ , found in tables:

$$\omega_- = \sqrt{1 + \left(\frac{\Omega_\infty}{2Q_{BW}} \right)^2} - \left(\frac{\Omega_\infty}{2Q_{BW}} \right), \quad (2.5.2)$$

when $\omega_0=1$ rad/s so that $\omega_+=1/\omega_-$.

The branch topologies in Figures 2.4.5 and 2.4.6 (called dipoles) can be replaced with any of several other topologies that do not change the impedance characteristic; these were given by [Zverev:524-526] and reproduced [Cuthbert,1983:Appen.H]. Those replacements are often

required to avoid high voltage points or to try for a better set of branch element values. Because the direct-coupled networks in Section 3.4.3.3 avoid these difficulties entirely, no further details of special transformations in elliptic function filters are described here.

2.5.4 Scaling

Impedance and frequency scaling of networks is convenient and enhances optimization. It is easy to rescale the normalized solution:

- Given an increased actual terminal resistance level, increase all R's and L's and decrease all C's by that factor, and
- Given an increased band center frequency in rad/s, decrease all L's and C's by that factor.

Impedance and frequency scaling do not affect the loaded Q's. Program DENORM.EXE simplifies scaling and rescaling and avoids many errors.

Example 2.5.3. Example 2.4.1 showed that the N=4 equal-element bandpass filter in Figure 2.4.4 had element values of $Q_i=8.25$ and $1/Q_i=0.1212 \forall i$ for a 3-dB bandwidth of 20%. *Problem:* Denormalize those element values from 1 ohm and 1 rad/s to 50 ohms and 100 MHz center frequency. *Solution:* Table 2.5.2. shows the user's interaction (<CAPS LOCK> is required) with program DENORM (with some annotation). Command 1 entered impedance factor 50, command 2 entered units 1E6, 1E-9, and 1E-12 for MHz, nanohenrys and picofarads, respectively, and command 3 entered frequency 100. Command 4 used the normalized reactance ω_0L (= $g_i \times Q_{BW}$), which is just the normalized inductance at 1 rad/s. Normally, the frequency, units and impedance factor are displayed before each answer. Commands 5-7 were not required in this case. Therefore, inductances are 656.5 nH in series and 9.6 nH in parallel; capacitances are 3.86 pF in series and 263 pF in parallel. The load resistance is 50 ohms.

Table 2.5.2. DENORM.EXE Output for Example 2.5.3.

```
***** MENU for DENORMALizing (Do 1, 2, & 3 First) *****
1. ENTER IMPEDANCE SCALING FACTOR (Default is Unity).
2. ENTER FREQUENCY AND L,C UNITS
3. ENTER FREQUENCY.
4. COMPUTE L & C GIVEN REACTANCE (or Normalized Inductance).
5. COMPUTE L & C GIVEN SUSCEPTANCE (or Normalized
Capacitance).
6. ENTER L AND COMPUTE IMMITTANCES AND RESONATING C.
7. ENTER C AND COMPUTE IMMITTANCES AND RESONATING L.
8. EXIT DENORM
*****
```

Table 2.5.2. (Continued) (After commands 1,2, and 3)

INPUT COMMAND NUMBER: ? 4

FREQUENCY = 100.0000 WITH UNITS = 1.E+06

INDUCTANCE UNITS = 1.E-09

CAPACITANCE UNITS = 1.E-12

IMPEDANCE SCALE FACTOR = 50.0000

REACTANCE (OHMS) =? 8.25 (User input)

INDUCTANCE = 656.5141

CAPACITANCE = 3.8583

PRESS <RETURN> KEY TO CONTINUE -- READY?

INPUT COMMAND NUMBER: ? 4

REACTANCE (OHMS) =? 0.1212 (User input)

INDUCTANCE = 9.6448

CAPACITANCE = 262.6319

2.6 Network Analysis

As a prelude to specific developments in subsequent chapters, the characterization of linear two-port networks by chain (ABCD) and scattering matrices is reviewed. The Hilbert transform is briefly described, because certain paired response data are interrelated, i.e., dependent. That is an essential component in the current academic approach to broadband impedance mapping.

2.6.1 ABCD Two-Port Parameters

Figure 2.6.1 shows a linear two-port; the output current is assumed to be leaving the output (b) port. A network is said to be *linear* if:

1. All voltages and current scale, i.e., if one is doubled, so are all the others, and
2. Superposition applies, i.e., the presence of sources with multiple frequencies does not generate additional frequencies.

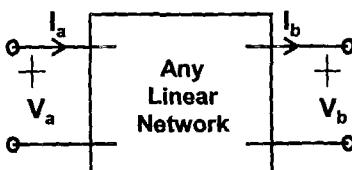


Figure 2.6.1. A linear two-port network with port voltages and currents.

The ABCD parameters are the coefficients in one of many possible sets of linear equations relating port voltages and currents:

$$\begin{aligned} V_a &= AV_b + BI_b \\ I_a &= CV_b + DI_b . \end{aligned} \quad (2.6.1)$$

Matrix notation in linear algebra allows a more compact equivalent expression:

$$\begin{bmatrix} V_a \\ I_a \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_b \\ I_b \end{bmatrix}. \quad (2.6.2)$$

The two vectors are matrices with only one column. It is convenient to denote the ABCD matrix as T :

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \quad (2.6.3)$$

The convention for deciphering (2.6.2) is that V_a , being in the first row and first column of its vector is related by superposing the first row of T , (A B), on the first column of the output vector and then adding the respective products, i.e. $V_a = A \times V_b + B \times I_b$. Similarly, I_a , being in the second row and first column of its vector is related by superposing the second row of T , (C D), on the first column of the output vector and adding the respective products, i.e. $I_a = C \times V_b + D \times I_b$. The utility in matrix notation is that it is just as compact when there are more than two equations in two unknowns. Generally, all the quantities in (2.6.1) and (2.6.2) are complex numbers.

2.6.2 Cascading Two-Port Subnetworks

The value of the ABCD formulation is in cascading two-port subnetworks as shown in Figure 2.6.2. Observing that adjoining input

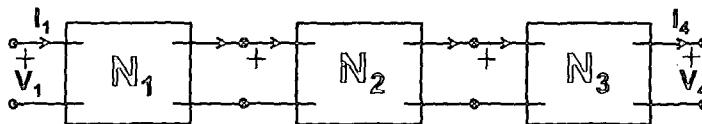


Figure 2.6.2. Cascaded subnetworks with common port voltages and currents.

and output port voltages and currents are shared, (2.6.2) shows that the overall chain matrix for the three subsections is

$$T_{14} = T_1 \times T_2 \times T_3 . \quad (2.6.4)$$

One application is finding the ABCD parameters for a ladder network that consists of a cascade of series impedance branches and parallel admittance branches. Consider the two branches in Figure 2.6.3.

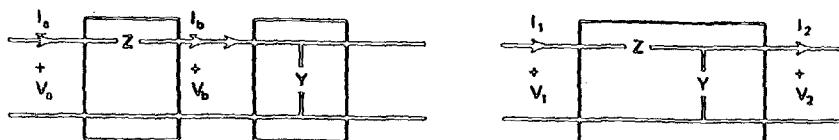


Figure 2.6.3. Cascaded series and parallel branches in a ladder network.

The ABCD matrices for those branches and their composite are

$$T = \begin{pmatrix} 1 & Z \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ Y & 1 \end{pmatrix} = \begin{pmatrix} 1+ZY & Z \\ Y & 1 \end{pmatrix}. \quad (2.6.5)$$

This example not only further illustrates the rules for matrix multiplication; it shows that there are several multiplications by the factors 1 and 0. It is shown in Section 6.2.3 that there are more efficient ways to obtain the ABCD parameters numerically.

If a network is lossless, then A and D are purely real, C and D are purely imaginary, and determinant $AD-BC=1$.

Example 2.6.1. Let Figure 2.6.3 represent a lossless lowpass network having $Z=j\omega L$ and $Y=j\omega C$. Then $A=1-\omega^2LC$, $D=1$, both real. Also, $B=j\omega L$ and $C=j\omega C$, both imaginary. Notice that $AD-BC=1$.

2.6.3 Scattering Parameters

Figure 2.6.4 again shows a linear network where the port parameters are power waves [Kurokawa]. Waves a_1 and a_2 are incident

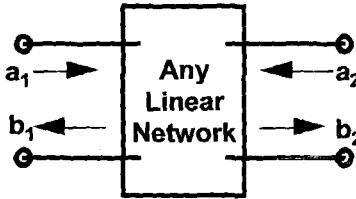


Figure 2.6.4. A linear two-port with incident and emerging power waves.

and waves b_1 and b_2 are emerging at their respective ports. The power incident on the input (left-hand) port is the maximum available power from a source as shown in (2.1.1):

$$|a_1|^2 = P_{as}, \quad (2.6.6)$$

and the net power delivered to a load connected to the right-hand port is

$$P_L = |b_2|^2 - |a_2|^2. \quad (2.6.7)$$

The complex a and b port variables are linearly related by complex (scattering) coefficients:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \quad (2.6.8)$$

The magnitudes of port variables a and b can be considered to have dimensions of the square root of power. Port variables can be normalized to their respective resistances or impedances [Cuthbert, 1983:93].

With any set of linear equations, it is important to interpret the meaning of each coefficient. Consider the meaning of S_{21} in the scattering parameter case; the second equation in the shorthand of (2.6.8)

is $b_2 = S_{21}a_1 + S_{22}a_2$. Suppose $a_2 = 0$, which implies a conjugate match at port 2; under that constraint, $S_{21} = b_2/a_1$. Therefore, $|S_{21}|^2$ is the power delivered to the load relative to the power available from a source, according to (2.6.7). Note that $a_2 = 0$ in Figure 2.6.4 means that there is no power reflected from the load. In the ordinary case, that means the load is 50 ohms (or 1 ohm if normalized), i.e., a "flat" load.

2.6.4 Special Relations for ABCD Parameters

The case just mentioned is shown in Figure 2.6.5. Of interest is

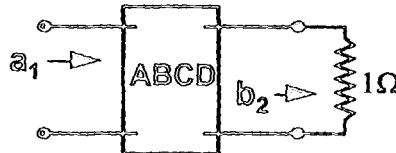


Figure 2.6.5. A flat-terminated two-port with incident and emerging waves.

the expression for S_{21} in terms of that network's ABCD parameters:

$$S_{21} = \frac{2}{A + B + C + D}, \quad a_2 = 0. \quad (2.6.9)$$

A similar but more general expression is given in Section 5.4.2 for the case where there is both a complex source and a complex load.

The input impedance of a two-port network as a function of the ABCD parameters and a complex load impedance, as shown in Figure 2.6.6, is easily derived using (2.6.1):

$$Z_{in} = R_{in} + j X_{in} = \frac{AZ_L + B}{CZ_L + D}. \quad (2.6.10)$$

This is just one of many important situations where the *bilinear form* in (2.6.10) occurs as a mapping of one network terminating impedance (Z_L) into another port parameter (Z_{in}). A bilinear function is linear in either variable, in this case in either Z_L or Z_{in} .

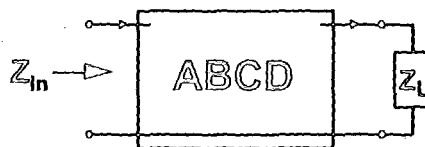


Figure 2.6.6. Input impedance of a two-port with arbitrary load impedance.

2.6.5 Hilbert Transform

One form of the *Hilbert transform* relates the input resistance and reactance of any linear ladder network. For example, if the input resistance, R_{in} , is known over all frequencies, then the reactance, X_{in} , at any particular frequency, say, ω , is

$$X_{in}(\omega) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{R_{in}(y)}{y - \omega} dy, \quad (2.6.11)$$

where y is a dummy variable of integration. Because resistance is an even function of frequency and is often band limited, evaluation of (2.6.11) with finite limits of integration is not a problem. In practice, even more simple integrations are possible; see [Cuthbert,1983:219-222] and [Carlin,1998]. The Hilbert transform plays a major role in the real-frequency broadband matching technique in Section 5.3.2.

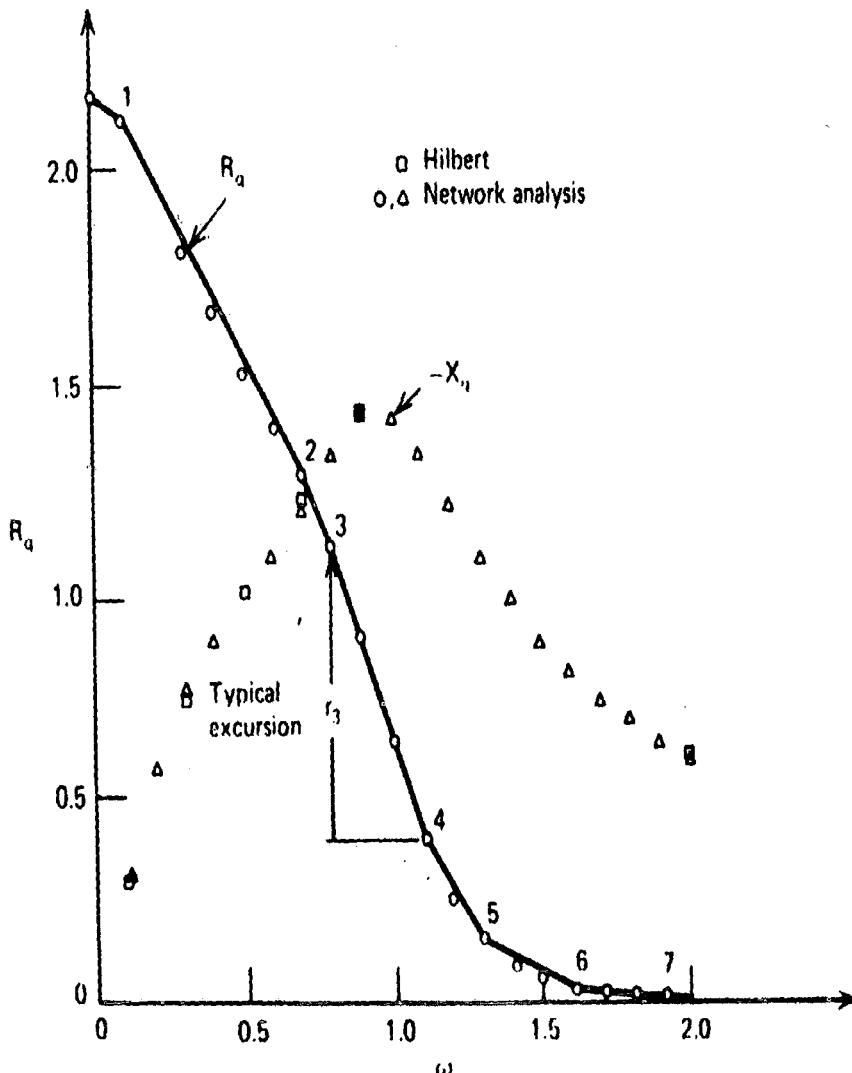


Figure 2.6.7. Input resistance and negative reactance for Example 2.6.2.

Example 2.6.2. Consider the $N=3$ lowpass network terminated by a 2.2-ohm resistance in Figure 2.6.8. An analysis of its input resistance, R_q , at

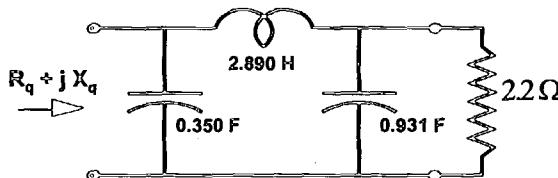


Figure 2.6.8. An $N=3$ lowpass network for analysis in Example 2.6.2.

a number of frequencies could be accomplished using the methods indicated in (2.6.5) and (2.6.10). The result is plotted in Figure 2.6.7. The negative of the reactance, $-X_q$, is also plotted using the triangle symbols. The square symbols represent $-X_q$ computed by (2.6.11); clearly the values by analysis and theory agree completely. Also, the network must have the “minimum reactance” property, i.e., in Figure 2.6.8 any series reactance added at the input terminals would not affect the input resistance function employed in (2.6.11) and thus could not affect the reactance computed by the Hilbert transform. See Section 2.1.5.

2.7 Summary of Fundamentals

For the purposes of this book, the fundamental concepts concern power transfer versus frequency, ladder network topologies and element values, and certain two-port network parameters. The loaded Q parameter is involved in several important ways, and the Hilbert transform is mentioned for the role it plays in current academic broadband matching methods.

Power transfer from a complex source through a lossless network to a complex load is the doubly-terminated situation, where the power can be no greater than some maximum available power that depends on the source resistance and source voltage or current. A generalized reflection coefficient is the comprehensive tool for determining the power and understanding the related impedance mappings (Appendix A). Power transfer from a lossless source through a lossless network to a complex load is the singly-terminated situation, where the power depends on the network input resistance or conductance and the source current or voltage, respectively.

The power versus frequency selectivity curve for a lowpass network is conventionally reproduced in a related bandpass network with appropriate scaling of the frequency axis. For purposes of this book, there are only a few passband shapes with certain properties of interest, such as ripple and flat loss. The stopband selectivity shape either is monotonic for all-pole networks or contains poles of attenuation (zeros of

transmission) for the elliptic-function and comprehensive cases. In every doubly-terminated case, the related frequency behavior of the network reflection coefficient is of interest. Presence of dissipation in the two-port network causes effects that have predictable properties that can be anticipated if not simply compensated.

3. Direct-Coupled Filters

Direct-coupled filters are bandpass ladder networks in dual forms of either all parallel resonators or all series resonators coupled in cascade by one or two reactive elements. A resonator consists of an LC pair that is tuned to the geometric band center frequency as described in Section 2.2.1. Direct-coupled filters have advantages in impedance, voltage, current, stray capacitance, and element value control that are not available in the classical bandpass network topology, which consists of alternating parallel and series resonators. Direct-coupled filter topologies carry to the limit certain alterations to the classical bandpass topology often made piecemeal, using Norton transformations.

Direct-coupled filters can be realized in a wide variety of physical forms, particularly at microwave frequencies. In any frequency range, the concept of inverters (ideal 90-degree transmission lines) that connect the resonators is a vital and simplifying tool; both resonators and inverters exist in many physical forms. Ideal inverters provide control of the resistance level at each resonator without affecting the filter's selectivity. In turn, those resistance levels enable control of element values and voltages and currents in simple and highly visible ways. Direct-coupled filters are easy to tune because of the inverters. Inverters also can incorporate stopband anti-resonances that provide the elliptic-function (Cauer) or comprehensive filter responses.

Finally, direct-coupled filters can be designed using the unifying loaded Q parameter as a meaningful guide for element values, selectivity, dissipation, sensitivity, and tuning. The effect of resonator loaded Q's and inverters on stopband selectivity is easily recognized in a simple Bode (semilog) graph. Resonator loaded Q's are also involved in straightforward design alterations to eliminate the passband distortion caused by non-ideal inverters as well as to avoid any negative element values. Elimination of passband distortion allows direct-coupled filters to escape their traditional limitation of being useful only for very narrow passbands; now they can be free of passband distortion over any band width.

3.1 Prior Technology

Bandpass filters have been an important part of radio engineering for at least 70 years, so it is important to identify those contributions that were stepping stones to direct-coupled filters as described in this chapter. An early example of a two-resonator filter was the intermediate frequency (I.F.) transformer that was crucial to the superheterodyne receiver introduced early in this century. It consisted of an RF transformer resonated on both primary and secondary sides by

capacitors. It is shown that a Pi of inductors that is equivalent to the transformer (Figure 2.4.10) contains an inverter flanked by inductors that, when resonated by capacitors, complete the primary and secondary resonators and thus the direct-coupled filter. More explicit recognition of coupled resonators is described in this section.

3.1.1 Classical Filters

Coupling coefficients were defined for adjacent parallel and series resonators in the classical bandpass filter topology (Figure 2.4.4) [Dishal, 1959]. For example, if a parallel resonator having capacitor C_i was followed by a series resonator having capacitor C_{i+1} , then Dishal defined a coupling coefficient

$$K_{i,i+1}^2 \equiv \frac{C_{i+1}}{C_i} = \frac{1/Q_{i+1}}{Q_i} = \frac{1}{Q_i Q_{i+1}} = \frac{1}{Q_{BW} g_i g_{i+1}}. \quad (3.1.1)$$

The three equalities on the right side of (3.1.1) follow from Figure 2.4.4 and (2.4.4). Because the bandpass network in Figure 2.4.4 depends on the lowpass prototype network in Figure 2.4.1, a related lowpass coupling coefficient was defined as well [Green, 1954]:

$$k_{i,i+1}^2 \equiv \frac{1}{g_i g_{i+1}}, \quad (3.1.2)$$

so that $k_{i,i+1}$ is a normalized coupling coefficient:

$$K_{i,i+1} \equiv \frac{k_{i,i+1}}{Q_{BW}}. \quad (3.1.3)$$

Green was mainly concerned with lowpass networks, where $Q_{BW}=1$.

3.1.2 Coupled Resonator Concept

Figure 2.4.7 shows adjacent parallel resonators top coupled by L , C , or parallel LC reactances as indicated by the “^” symbol. By writing the nodal equations for the transfer function, it was noted that the assumption of frequency independence of the top coupling elements produced the same transfer function obtained from classical filters having both kinds of resonators [Dishal, 1949]. These *constant reactance couplings* have been used extensively in narrow bandpass filter design methods [Humphery].

Tables of normalized terminal Q 's and coupling coefficients, q and k respectively, have been published in several resources [Zverev, 341], [ITT, 1975]. The two normalized terminal q 's are:

$$q_i = Q_i / Q_{BW}, \quad i = 1 \& n. \quad (3.1.4)$$

The constant reactance coupling elements between nodes i and j are defined by

$$K_{ij} = \frac{C_{ij}}{\sqrt{C_i C_j}}, \quad (3.1.5)$$

and there is a similar definition in terms of inductances for coupling L_{ij} . The constant-reactance design procedure begins with choice of a terminal

resistance at one end of the filter. Then that unnormalized Q , e.g., Q_1 , determines the adjacent resonator capacitor, C_1 . The next step uses (3.1.3) and (3.1.1) to find the capacitance in the next resonator and (3.1.5) to find the coupling capacitance, C_{12} . For the remainder of the filter, only the coupling coefficients are required until reaching the opposing terminating resistance, which is found by (3.1.4).

Dissipative elements have been anticipated by means of uniform predistortion of the response. Assuming that all L's and C's in the filter have the same unloaded Q factor, Q_u , then the real parts of the response function roots can be predistorted in the Laplace frequency s plane by the factor:

$$q_0 = Q_u / Q_{BW} . \quad (3.1.6)$$

The predistorted network element values are synthesized in the frequency variable ($s - 1/q_0$). Tables are arranged according to decreasing values of q_0 , from infinity (lossless) to some low value (very dissipative elements) [Zverev:341].

The next few sections show what the older coupled resonator design techniques did not explicitly identify: resonator loaded Q's, inverters, and the vital set of parallel resistances.

3.2 Properties

This section describes the prototype direct-coupled filter composed of resonators, inverters, and optional end couplings. The properties of these three main ingredients are described.

3.2.1 Topologies

The prototype direct-coupled filter network is shown in Figure 3.2.1. It is composed of resonators at each labeled node (I, II, ...) that are

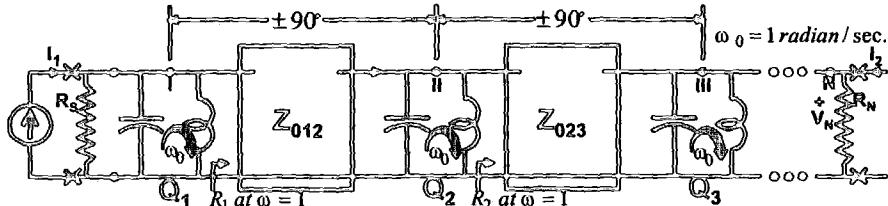


Figure 3.2.1. Prototype direct-coupled network with parallel resonators.

each tuned to the passband center frequency ω_0 , usually 1 rad/s. The resonators are connected by ideal inverters that act as frequency-independent 90-degree transmission lines; the node voltages are shown to differ by that phase angle. A network that is the dual of Figure 3.2.1 would contain series LC resonators connected by 90-degree ideal inverters; see Figure 2.4.12.

At ω_0 , the resonators appear to be open circuits. The parallel resistance following the last resonator in Figure 3.2.1 is transformed to another parallel resistance at the next-to-last resonator by an inverter; similarly, the other inverters determine all internal parallel resistances, until resistance R_1 appears at the input resonator. The input resistance R_1 may not match the source resistance, R_s , if there is to be a flat loss or doubly-terminated ripple response, and R_s is not present at all for singly-terminated applications. These easily observed internal parallel resistances are not considered in the older design procedure in Section 3.1.2, involving terminal Q's and coupling coefficients, K_{ij} .

The end resistances R_1 and R_N shown in Figure 3.2.1 may be obtained by end couplings, so that the fixed physical source and load resistances do not limit design choices. Figure 3.2.2 shows one possible end coupling arrangement based on $1+Q^2$ conversion of a parallel RC subnetwork into an equivalent series RC subnetwork at ω_0 . Other end

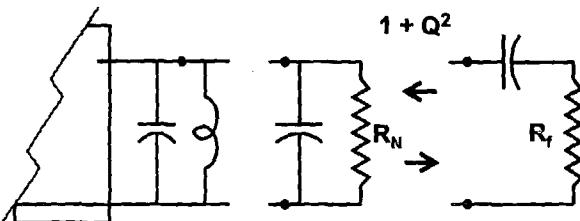


Figure 3.2.2. End coupling by $1+Q^2$ to control terminal resistance level.

coupling arrangements include RF transformers, Figure 2.4.10, or tapped reactances. These narrow-band end-coupling possibilities have been summarized in a convenient table [Zverev:567], and a similar design step for tapping into terminal coaxial resonators has been described [Dishal:1965].

As an aside, it is useful to note that the direct-coupled filter prototype topology in Figure 3.2.1 could be enhanced by bridging one or more non-adjacent nodes, e.g., an L or C coupling element from node I to node N. These can produce transmission zeros above and/or below the passband in real frequency or on the real axis in the Laplace s plane. The latter can improve the group delay response. A qualitative summary is available [Johnson], but the new "CQ" technique efficiently implements both lumped-element and microwave filters [Levy,1995]. Non-adjacent node bridging techniques are beyond the scope of this book.

3.2.2 Resonators

The parallel LC resonators in Figure 3.2.1 have loaded Q values that are established by the desired response shape and bandwidth. Each given loaded Q can be realized in terms of the parallel resistance and resonator reactance values at ω_0 according to (2.3.2). The inverters described in the next section enable wide ranging choices of parallel

resistance, so that both resonator and inverter element values can be controlled as desired without affecting the filter response.

The parallel LC resonators in Figure 3.2.1 could be replaced by many other resonant structures for use in narrow-band filters, e.g., shorted transmission-line stubs that are 90 degrees at ω_0 , or less than 90 degrees and capacitively resonated, or microwave cavities, etc. The main assumption is that the physical resonator's zero susceptance and slope match the equivalent LC resonator's at ω_0 . It is easy to show that the slope of parallel LC resonator susceptance, $B(\omega)$, at ω_0 in terms of resonator capacitance, C , is [Cuthbert,1983:297]:

$$\frac{dB}{d\omega} = \frac{1}{2C} \Big|_{\omega = \omega_0} \quad (3.2.1)$$

This is the usual starting relationship for microwave filter design, where resonators take many different forms but have zero susceptance and a known slope at the band center frequency. There is also a well-known relationship between resonator reactance slope and stored electric and magnetic energy [Drozd].

Unfortunately, the slope equivalence in (3.2.1) does not relate to loaded Q and broad bandwidths, because (3.2.1) is only the second coefficient in a Taylor series expansion of a broadband susceptance function. Therefore, resonator slope equivalence does not describe resonator dissipation, Section 2.3.3, or the stopband performance described in Section 3.3.1. It is the preoccupation with resonator slope equivalence in microwave filter literature [Matthaei,1964] that obscures the simple basis of direct-coupled filter design advocated in this book. See Section 5.2.6 for several ways to measure the loaded Q of a single resonator and Section 3.6.2 for a way to measure all loaded Q's in a tuned bandpass filter *in situ*.

3.2.3 Inverters

As noted, an ideal inverter acts like a transmission line with characteristic impedance Z_0 ohms and electrical length $\theta = 90$ degrees, that do not change with frequency. The pictorial representation is shown in Figure 3.2.3. Such a two-port network has a chain matrix:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 0 & jZ_0 \\ jY_0 & 0 \end{bmatrix}, \quad (3.2.2)$$

where $Y_0=1/Z_0$. Note that determinant $AD-BC=1$, as required for lossless

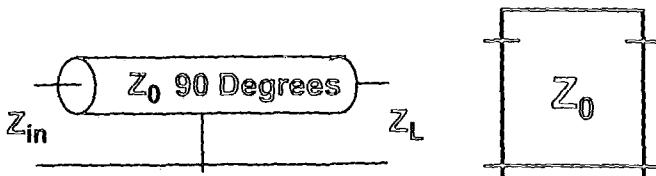


Figure 3.2.3. An ideal inverter and its conventional box symbol.

two-port subnetworks. The 90-degree port voltage/current relationship is indicated by (2.6.1) and (3.2.2). Also, the equation for the input impedance of a two-port network, Z_{in} , terminated by a load impedance, Z_L , as in (2.6.10) shows that, for an inverter,

$$Z_{in} = \frac{Z_0^2}{Z_L}. \quad (3.2.3)$$

The name "inverter" comes from (3.2.3) when $Z_0=1$ ohm. Because the direct-coupled filter is designed at midband frequency, ω_0 , the resonators in Figure 3.2.1 are open circuits. Then choices for the inverter Z_0 's transform the load resistance to R_3 , R_2 , and R_1 .

It is especially important to know that a series LC resonator can be replaced by a parallel resonator flanked by ideal inverters as shown in Figure 3.2.4. This can be verified by using (3.2.2) in the ABCD cascading

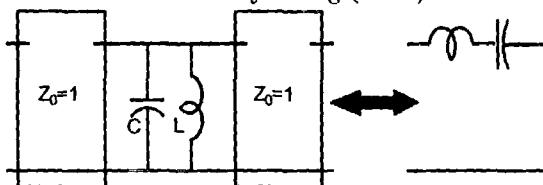


Figure 3.2.4. Ideal equivalence between series and parallel resonators.

method of Section 2.6.2. The indicated $Z_0=1$ in Figure 3.2.4 is not limiting, because other values simply scale the impedance level according to (3.2.3). Therefore, the direct-coupled filter prototype network in Figure 3.2.1 is derived by replacing the series resonators in the classical bandpass prototype network, Figure 2.4.4.

Three inverters of immediate interest are shown in Figure 3.2.5.

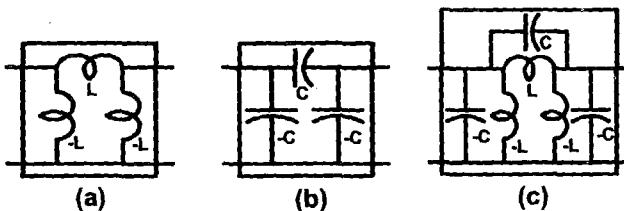


Figure 3.2.5. Narrow-band inductive, capacitive, and trap Pi inverters.

These inverters are Pi networks because they are associated with adjacent parallel resonators which can absorb the negative elements; the dual case, Figure 2.4.12, employs similar T inverter networks. The inductive inverter in Figure 3.2.5(a) is 90 degrees at all frequencies, a surprising fact that can be verified by the ABCD cascading method of Section 2.6.2. That also shows that the Z_0 of the inductive inverter is

$$Z_{0L} = \left(\frac{\omega}{\omega_0} \right) (\omega_0 L), \quad (3.2.4)$$

which is directly proportional to frequency, ω . It is assumed that $\omega \approx \omega_0$ in narrow pass bands. The capacitive inverter in Figure 3.2.5(b) has similar properties:

$$Z_{0C} = \left(\frac{\omega_0}{\omega} \right) \left(\frac{1}{\omega_0 C} \right), \quad (3.2.5)$$

which has Z_0 inversely proportional to frequency, ω . In both the L and C cases, the inverter Z_0 is equal to the top coupling reactance at ω_0 .

The trap inverter in Figure 3.2.5(c) furnishes one zero of transmission in the stopband at a null frequency, ω_n :

$$\omega_n = \frac{1}{\sqrt{LC}}. \quad (3.2.6)$$

Like the inductive or capacitive inverter, the characteristic impedance of the trap inverter is just the branch reactance at band center frequency, ω_0 :

$$Z_0 = 1 / \left[(\omega_n C) \left| \frac{\omega_0}{\omega_n} - \frac{\omega_n}{\omega_0} \right| \right]. \quad (3.2.7)$$

When $\omega_n > \omega_0$ the coupling branch reactance is inductive, and when $\omega_n < \omega_0$ it is capacitive. In any case, the sign of the reactance is ignored, and the Z_0 of the inverter is equal to that many ohms. For narrow-band design, one or more trap inverters can be used, with their null frequencies placed for selectivity by a well-known pole-placing technique [Daniels].

The great utility of inverters is confirmed by observing that any Pi or T reactive network having a center element with reactance of opposite sign to the other two branches is an inverter at that frequency. That supports the fact that every lossless two-port network contains an inverter; e.g., a circuit model of the short-circuit admittance equations shows a Pi inverter with y_{11} and y_{22} susceptances on either side of an inverter [Cuthbert, 1983, 291]. Particularly, the RF transformer in Figure 2.4.10 contains an inductive inverter.

3.2.4 Narrowband Choices in Parallel Resistance Space

Parallel resistance value choices at midband frequency ω_0 in Figure 3.2.1 affect impedance levels but do not affect selectivity. Parallel resistance values directly determine coupling element reactance values according to (3.2.3)-(3.2.7). Because resonator loaded Q values are fixed by selectivity requirements, another important effect of parallel resistance values is to control resonator reactance values. Furthermore, parallel resistance values also determine the resonator voltages and currents for a given amount of power passing through the lossless network.

There is always a region in parallel resistance space where every negative inverter element branch can be absorbed to leave a positive final element of that kind. This consideration is increasingly important with larger band widths (lower loaded Q's) and is developed and displayed in

Section 3.4.5. At this point in the development, inverter reactances have been assumed frequency independent in the pass band, a narrow-band assumption which is removed later. Still, it is useful to show how parallel resistances are involved in positive element values in the present situation.

For internal nodes in Figure 3.2.1, e.g., node II, and considering only L or C inverters, Figure 3.2.5, there are only three cases to consider for positive shunt element reactances at ω_0 :

Case 1. Both adjacent inverters are inductive/magnetic. The resonator reciprocal inductance, $L^{-1}=1/L$, is $L^{-1}=Q_2/R_2$. The coupling inductors and their shunt negative branches have reactance values that are the geometric means of their corresponding parallel resistances, i.e., $Z_{012}=X_{12}=\sqrt{R_1 R_2}$ and $Z_{023}=X_{23}=\sqrt{R_2 R_3}$. The final shunt reciprocal inductance at node II in Figure 3.2.1, after absorbing the negative inductances, is

$$L_2^{-1} = \left[Q_2 - \sqrt{R_2} \left(\frac{1}{\sqrt{R_1}} + \frac{1}{\sqrt{R_3}} \right) \right] / R_2. \quad (3.2.8)$$

Consider when L_2 is made to vanish, i.e., when $L_2^{-1}=0$:

$$R_2 \leq \frac{Q_2^2}{\left(\frac{1}{\sqrt{R_1}} + \frac{1}{\sqrt{R_3}} \right)^2} \text{ for } L_2 \geq 0. \quad (3.2.9)$$

Case 2. Both adjacent inverters are capacitive. In a similar way, it can be shown that

$$R_2 \leq \frac{Q_2^2}{\left(\frac{1}{\sqrt{R_1}} + \frac{1}{\sqrt{R_3}} \right)^2} \text{ for } C_2 \geq 0. \quad (3.2.10)$$

Case 3. One adjacent inverter is inductive and the other is capacitive. Let $R_{\min}=\min(R_1, R_3)$. Then

$$R_2 \leq R_{\min} Q_2^2 \text{ for } C_2 \geq 0 \text{ and } L_2 \geq 0. \quad (3.2.11)$$

Without loss of generality, setting $R_3=1$ implies that R_1 is within an order of magnitude of unity. Pass bandwidths of less than 20% ($Q_{BW}>5$) implies that $Q_2>5$ can be expected. It is concluded that $R_2 \leq Q_2^2$ is a constraint that almost always satisfied. In any event, (3.2.9)-(3.2.10) show that for any given R_3 value, there is always a feasible region in the R_1 - R_2 parallel resistance space for $L_2>0$ and $C_2>0$, because $Q_2>0$.

Case 4. Narrow-band capacitive end coupling. The capacitive end coupling illustrated in Figure 3.2.2 can be designed by the $1+Q^2$

method; see Example 3.3.2 in Section 3.3.2. It is easy to derive the constraint:

$$R_N \leq R_f (1 + Q_N^2) \text{ for } C_n \geq 0. \quad (3.2.12)$$

3.3 Response Characteristics

Inverters and end couplings have an important effect on both passband and stopband selectivity of direct-coupled networks. This section shows how to account for those effects using simple relationships using inverter properties and the resonator loaded Q's.

3.3.1 Stopband Selectivity

It has long been known that attention to terms in selectivity functions involving frequencies far from the passband leads to useful asymptotic relationships. For direct-coupled LC filters, that approach accurately estimates the stopband loss, L, when it is greater than about 20 dB:

$$L \approx -6 + 20 \log_{10} \prod_1^N Q_L + N 20 \log_{10} |F| + (NMI - NCI) 20 \log_{10} \left(\frac{\omega}{\omega_0} \right) \text{ dB}, \quad (3.3.1)$$

where

$$F \equiv \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right). \quad (3.3.2)$$

The frequencies in the ratios could be in any units, e.g. MHz, because the units cancel. The (-6) term in (3.3.1) is included for doubly-terminated filters and omitted for singly-terminated filters. NCI is the number of capacitive inverters, and NMI is the number of magnetic or inductive inverters.

Figure 3.3.1 shows the Bode semilog graph of the stopband selectivity relationship. The break point is shown for a singly-terminated filter, and it clearly shows the selectivity effect of large values of loaded Q product. Classical bandpass filters (no inverters, as in Figure 2.4.4) have stopband asymptotes that slope at 6N dB per octave, where N is the number of resonators. Direct-coupled filter asymptotes have a slope bias of 6 dB per octave for each excess inverter type (L or C); e.g., in Figure 3.3.1, (NMI-NCI)=2, so the upper stopband asymptote gains 12 dB per octave and the lower stopband asymptote loses a like amount. Figure 3.3.1 also indicates that passband distortion exists, even when NMI=NCI in ordinary direct-coupled filters. Section 3.4 shows how to eliminate that distortion.

Stopband selectivity estimate (3.3.1) also can provide useful estimates for direct-coupled filters with one or more traps. Figure 3.3.2 shows the selectivity effect of one trap inverter tuned to a null frequency in the upper stopband with one other inductive inverter separating N=3 resonators. Well above the null frequency, the trap acts like a

capacitance, so the asymptote there appears to have a slope that is the nominal $6N - 0 = 18$ dB per octave, i.e., $(NMI - NCI) = 0$.

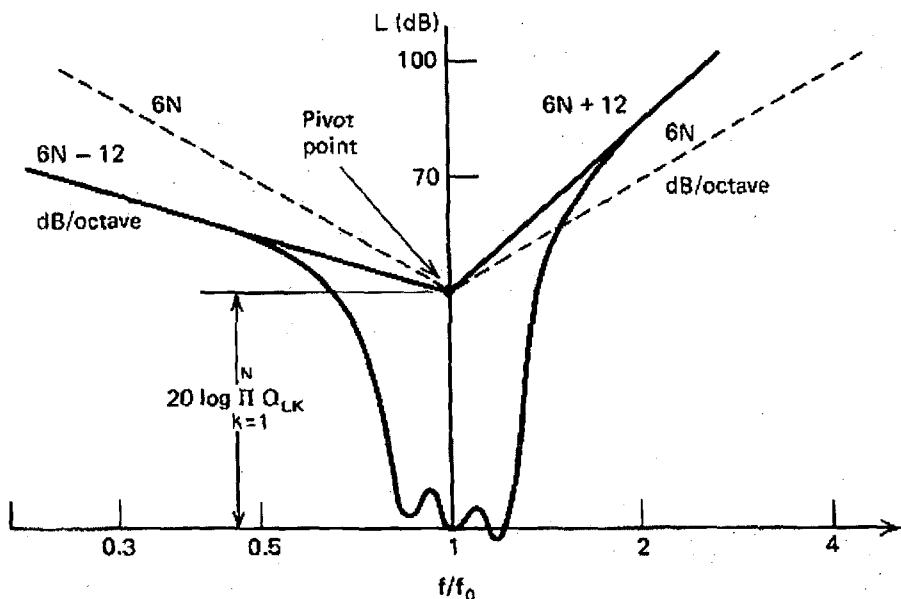


Figure 3.3.1. Stopband selectivity from direct-coupled filters.

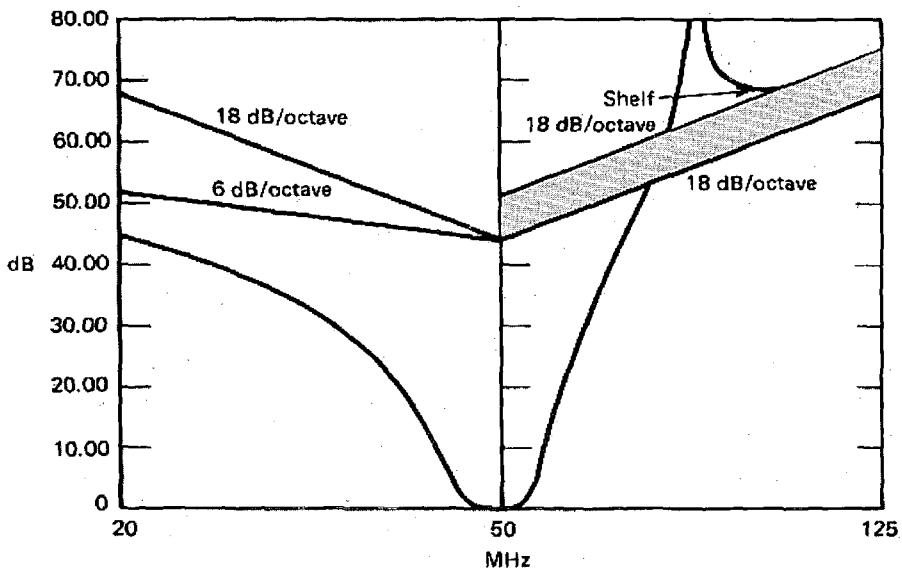


Figure 3.3.2. Stopband selectivity of a direct-coupled filter with a single trap.

Similarly, in the lower stopband, the trap acts like an inductor, and the apparent ($NMI-NCI=2$) causes the asymptote to have a slope of only $6N-12=6$ dB per octave. The main fact is that the pivot point on the ordinate has moved up for the upper stopband, providing increased selectivity without an increase in loaded Q product. That is the advantage of trap inverters, even though a price is paid elsewhere in the stopbands. The amount of apparent increase in loaded Q product has been quantified [Cuthbert,1983:293]. See Section 4.2.4 for a more accurate estimate of arbitrary stopband selectivity due to transmission zeros.

End coupling affects stopband selectivity in much the same way an inverter does, i.e. 6 dB per octave for end coupling that changes the resistance ratio by as much as 10:1 or $Q=3$ in (2.3.4) and less than 6 dB per octave for less resistance transformation. See [Cuthbert,1983:463] for details.

3.3.2 Passband Width

Passband width can be fixed and the stopband performance is dependent, or vice versa. Given lowpass prototype element values, g_i , the loaded Q's are obtained directly by specifying the bandwidth, viz., Q_{BW} , according to (2.4.4). When stopband performance is specified, the loaded Q product, ΠQ_L , in (3.3.1) and its major role illustrated in Figure 3.3.1 also determine the dependent passband width. It is easily shown that

$$Q_{BW} = \left[\frac{\Pi Q_L}{\Pi g_i} \right]^{\frac{1}{N}}, \quad (3.3.3)$$

where N is the number of resonators.

Example 3.3.1. From [Cuthbert,1983:286-9]. A three-resonator Butterworth filter is to be driven by a 50-ohm source and terminated in a 100-ohm load. The midband frequency is 50 MHz, and 60-dB attenuation is required at 90 MHz. Use inductive inverters to minimize the loaded Q values. Inductance values are to be in the range 20-300 nH. *Problem:* Find the network element values in ohms at 50 MHz and in pF and nH. Analyze the result with and without dissipation in the inductors (Q_u 's of 100 and ∞ , respectively) to show passband and stopband effects.

Solution: First, use program DENORM to determine that 20-300 nH corresponds to 6.28-94.25 ohms reactance at 50 MHz. Stopband attenuation estimate (3.3.1) provides the loaded Q product (ΠQ_L) because the other terms are known: $N=3$, $NMI=2$, $NCI=0$, and $\omega/\omega_0=90/50=1.80$. By (3.3.2) $|F|=1.2444$. Thus, (3.3.1) yields $\Pi Q_L=319.5424$ for 60 dB at 90 MHz. The g_i for a maximally-flat 3-dB passband are 1, 2, 1 [Matthaei:98]. That g_i product and the loaded Q product, according to (3.3.3), result in $Q_{BW}=5.4264$ and an ideal 3-dB bandwidth of 18.43%.

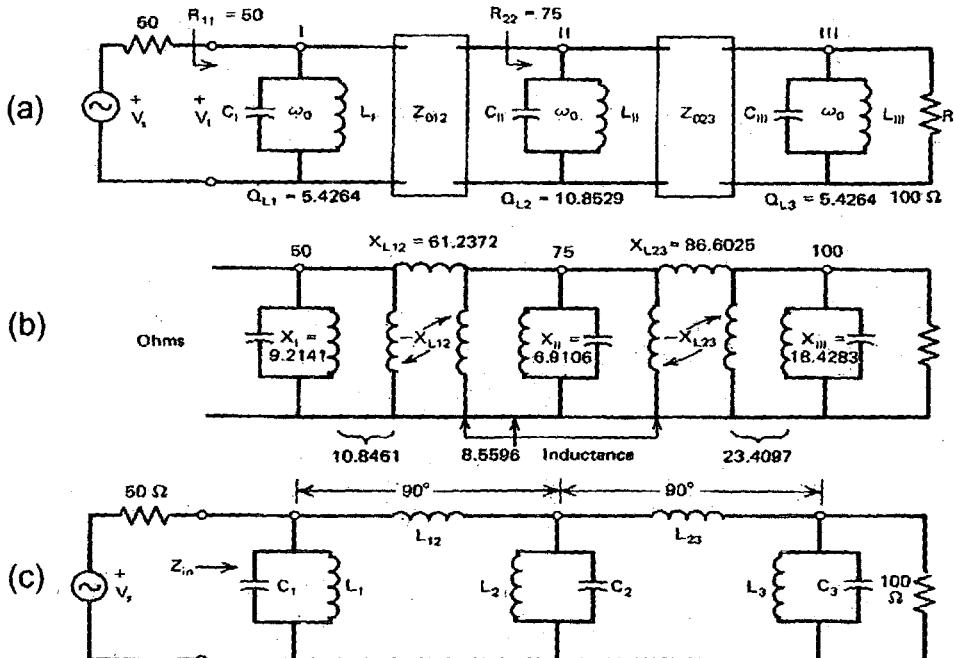


Figure 3.3.3. A three-resonator direct-coupled filter for Example 3.3.1.

Figure 3.3.3 (a) shows the prototype direct-coupled filter with the inverters and resonators. Each resonator loaded Q is $g \times Q_{BW}$ from (2.4.4) with the results shown in Figure 3.3.3(a). Because R_1 and R_3 are given, R_2 may be chosen to obtain inductive reactance values in the given range and/or to control the voltage at node II given the power delivered from the source. In this case, try $R_2=75$ ohms; then (3.2.3) requires $X_{L12}=61.2372$ and $X_{L23}=86.6025$ ohms. The associated negative reactances in the inverters are shown in Figure 3.3.3(b) along with the resonator reactances obtained from (2.3.2): $X_P=R_P/Q_P$. Combining adjacent inductive reactances having both signs produces the reactances at 50 MHz associated with the final elements in Figure 3.3.3(c): $X_{C1}=9.2141$, $X_{L1}=10.8461$, $X_{L12}=61.2372$, $X_{L2}=8.5596$, $X_{C2}=6.9106$, $X_{L23}=86.6025$, $X_{L3}=23.4097$, and $X_{C3}=18.4283$ ohms. These reactances can be converted to the capacitance and inductance values shown in Table 3.3.1.

Table 3.3.1. Elements in the Lossless Filter Shown in Figure 3.3.3.
Units: pF and nH.

C ₁	L ₁	L ₁₂	C ₂	L ₂	L ₂₃	C ₃	L ₃
345.46	34.52	194.92	460.61	27.25	275.66	172.73	74.52

A semi-log (Bode) graph of the selectivity of this lossless filter is shown in Figure 3.3.4. The break point predicted by (3.3.1) is shown at $-6+20\log(319) = 44$ dB. The actual attenuation at 90 MHz, all due to mismatch loss, is 60.15 dB. Figure 3.3.5 shows selectivity of the same filter with all inductors having unloaded $Q_u=100$. The midband insertion loss is nearly 1 dB and the passband center is slightly lower than 50 MHz. The 3-dB bandwidth relative to the midband loss is about 18% as predicted. Dissipation effects are predicted in Section 3.3.4 and passband distortion due to imperfect inverters is eliminated in Section 3.4.

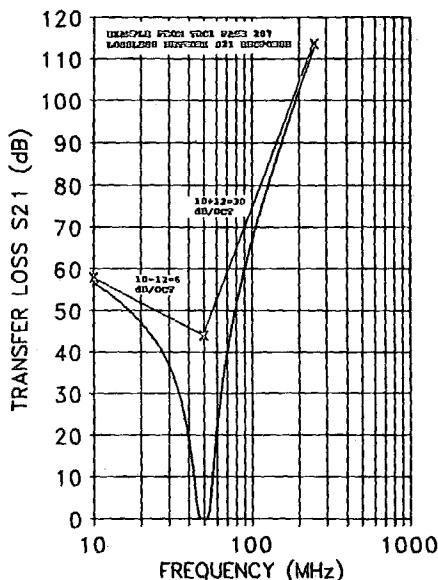


Figure 3.3.4. Selectivity of a lossless filter with inductive couplings.

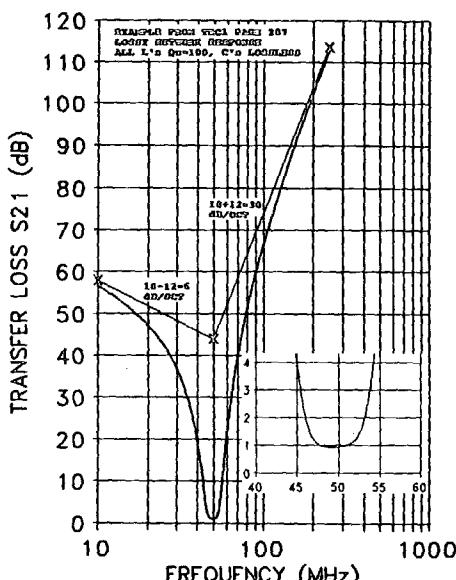


Figure 3.3.5. Selectivity of a lossy filter with inductive couplings.

Example 3.3.2. Change the output termination of the filter in Example 3.3.1 from 100 to 50 ohms. *Problem:* Design a top-C coupling from the output resonator. *Solution:* The situation is shown in Figure 3.2.2. The $1+Q^2$ equations in Section 2.3.1 apply to this case. Because $R_P=100$ and $R_S=50$, $Q=1$ according to (2.3.5). Then $X_P=R_P/Q_P=100$ and $X_S=Q_S R_S=50$ ohms by (2.3.2). At the band center frequency of 50 MHz, $C_P=31.83$ pF and $C_S=63.66$ pF. The modification to Example 3.3.1 consists of reducing C_3 in Table 3.3.1 to $(172.73-31.83)=140.90$ pF and adding a 63.66 pF capacitor in series with a 50-ohm load resistor as shown in Figure 3.2.2. Analysis shows that the top-C coupling to the load does not change the passband width but does reduce the attenuation at 90 MHz to about 57 dB, i.e., a 3-dB reduction due to capacitive cancellation of part of the two

inductive inverters. (The coupling $Q=1$ is less than $Q>3$ for full inverter asymptote behavior of 6-dB/octave as illustrated in Figure 3.3.4.)

3.3.3 Mismatched Terminations

Mismatch loss at midband frequency, ω_0 , in doubly-terminated filters is due to unequal terminating resistances as discussed in Section 2.2.4 and illustrated in Figure 2.2.2. It is designed to occur in even-order Chebyshev responses and/or in broadband matching applications, and the proper source resistance is available from lowpass prototype values as described in Section 2.5.2. Figure 3.2.1 shows the direct-coupled bandpass prototype network that is designed at the midband frequency, ω_0 , where R_s may not be equal to R_1 , the input resistance ω_0 .

The desired input resistance at midband, R_1 in Figure 3.2.1, is obtained by choosing inverter Z_0 's using (3.2.3). When normalized to a 1-ohm load, the guiding requirement is that

$$\frac{R_s}{R_1} = g_{N+1}, \quad (3.3.4)$$

where g_{N+1} is the normalized source resistance shown in Figures 2.4.1 and 2.4.4. Therefore, the prototype network is designed so that $R_1=R_s/g_{N+1}$. Then, after all element values are determined and the network is connected to a source with R_s , the proper mismatch will exist at the input (and everywhere else in the network).

As described in Section 2.1.5, the input resistance in singly-terminated (lossless-source) filters determines the input voltage relative to the voltage across the load resistance at midband frequency, ω_0 ; see (2.1.13). Adjustment of any of the inverter Z_0 's will change the input resistance to any value desired.

Example 3.3.3. Program ALLCHEBY, described in Section 2.5.2, shows that a Chebyshev $N=4$ doubly-terminated network response having 0.1 dB ripple requires a source resistance (R_s) of 1.3554 ohms for a 1-ohm load resistance. See Table 2.5.1. Because this is an even-order Chebyshev response, there is 0.1 dB loss at the midband frequency, ω_0 . Suppose that equal terminating resistances are required, i.e., $R_s=1$; then the prototype network must contain inverters that produce $R_1=1/1.3554=0.7378$ ohms. There would be three inverters for $N=4$; see Figure 3.2.1. If parallel resistances at the second and third resonators were arbitrarily chosen to be 2 ohms, then $Z_{034}=\sqrt{(1\times 2)}=1.4142$, $Z_{023}=\sqrt{(2\times 2)}=2.0000$, and $Z_{012}=\sqrt{(2\times 0.7378)}=1.2147$ ohms.

A different problem would be having a source $R_s=5.0$. Then, (3.3.4) would require $R_1=5.0/1.3554=3.6889$ ohms, and that might be obtained simply by changing $Z_{012}=\sqrt{(2\times 3.6889)}=2.7162$ ohms.

Example 3.3.4. Suppose that that a Chebyshev N=4 singly-terminated network response having 0.5 dB ripple is desired with a midband ($\omega=\omega_0$) voltage 3 times the load voltage. By (2.1.13), the input resistance must be $3^2=9$ times the load resistance at ω_0 . Then, if the loss at the midband frequency, ω_0 , is defined to be zero dB, there will be 4 peaks of 0.5 dB GAIN in the passband.

3.3.4 Effects of Dissipation

Dissipative elements in a network affect both the pass band and the stop band; remarkably, there are offsetting effects in the stopband so that the overall attenuation is essentially unchanged. The passband effects are much more severe, as described in Section 2.2.6.

The approximate loss due to dissipation in a resonator is given by (2.3.8): $L_0 \approx 4.34 Q_L / Q_u$ dB. Applying this equation for the three resonators in Example 3.3.1 and Figure 3.3.3(a), the unloaded $Q_u=100$ employed in Figure 3.3.5 indicates losses per resonator of 0.24, 0.47, and 0.24 dB, respectively. The sum is 0.95 dB, whereas the midband loss (all dissipative) by analysis is found to be 0.94 dB. This illustrates the fact that almost all the dissipative loss is in the resonators, not in the inverters [Cuthbert,1983:297].

The design of singly-terminated filters is shown to be especially dependent on the input resistance or conductance in Section 2.1.5 (see Figure 2.1.8). The direct-coupled bandpass prototype network in Figure 3.2.1 shows that at midband frequency ω_0 each resonator admittance is reduced by resonance to only dissipative conductance, if any. Then the input resistance is determined only by the load resistance, the inverter Z_0 's, and any resonator dissipation conductances. The designer must adjust at least one inverter Z_0 , e.g., the inverter between the first and second resonators, to obtain the desired input resistance when finite resonator Q_u exists. That is easily accomplished by either analyzing the lossless design with lossy elements, or by the continued-fraction expansion of the input resistance in terms of the ratios of Q_L/Q_u [Cuthbert,1983:296,466].

In the stopband, i.e., more than 20 dB attenuation, dissipative elements leave the total attenuation (insertion loss) constant by approximately dividing it between mismatch loss (2.1.11) and efficiency loss (2.1.12). In Example 3.3.1, the mismatch loss at 90 MHz for a lossless network was 60.15 dB. By analyzing the network in Figure 3.3.3 with inductors having $Q_u=100$, the attenuation shown in Figure 3.3.5 for 90 MHz is 60.13 dB, composed of 25.77 dB mismatch loss and 34.36 dB efficiency/dissipative loss. This universal phenomenon has never been contradicted or precisely explained (perhaps it is due to Boucherot's Theorem [Carlin,1998]). It can be relied upon.

3.3.5 Passband Distortion

There are no physical ideal inverters, i.e., two-port subnetworks that have constant characteristic impedance and constant 90-degree electrical length, both independent of frequency. Usually, one or both of these ideal characteristics vary with frequency. For example, the three LC inverters in Figure 3.2.5 are all 90 degrees long at all frequencies, but their Z_0 's vary with frequency. A coaxial transmission line has a constant Z_0 but its electrical length is proportional to frequency. Both Z_0 and electrical length of a rectangular waveguide vary with frequency. These non-constant properties affect the ideal selectivity of direct coupled filters. It is shown in Section 3.3.1 that the effect of LC inverters on the stopband selectivity is easily predicted.

There are no known ways to predict quantitatively the passband distortion due to inverters. It is possible to construct approximate frequency-mapping functions that predict passband edge frequencies for special cases [Cohn,1957]. The distortion in the maximally-flat filter, Example 3.3.1 and Figure 3.3.5, is not so obvious. The qualitative nature of distortion of equal-ripple passband response shapes can be seen in the following two examples, where narrow passbands are poor and wide passbands are unusable, respectively. Section 3.4 shows how to eliminate passband distortion in every situation.

Example 3.3.5. A two-resonator doubly-terminated filter must have a 10% pass band with a 0.9697 dB ripple. *Problem:* Design and analyze a normalized network to show the passband distortion. *Solution:* The ALLCHEBY.EXE program provides the design information in Table 3.3.2:

Table 3.3.2. A Two-Resonator Doubly-Terminated Filter from Program ALLCHEBY

Is this a Matching or a Filter ($QL=0=QS$) network (L,F)? F

Is this filter Singly- or Doubly-terminated (S,D)? D

N, dB FLAT LOSS, dB RIPPLE, %BW =? 2,0,0.9697,10

THIS INSERTION LOSS FROM 0.0000 TO 0.9697 dB

THIS dB RIPPLE = 0.9697

Lser or Cpar G(1) = 1.7994 Q(1) = 17.9936

Cpar or Lser G(2) = 0.6871 Q(2) = 6.8708

Ohms or Mhos G(3) = 2.6188

PASSBAND CENTERED AT 1 R/S HAS EDGES 0.9512 1.0512

20-dB STOPBAND HAS EDGES 0.8514 1.1746.

The normalized direct-coupled filter is shown in Figure 3.3.6. The equivalence in Figure 3.2.4 is used to convert a series resonator next to the load into a parallel resonator. (The inverter next to the load is assumed to be ideal and can be removed.) Therefore, as described in Section 2.4.2, the left-hand column is selected from Table 3.3.2, i.e., Lser,

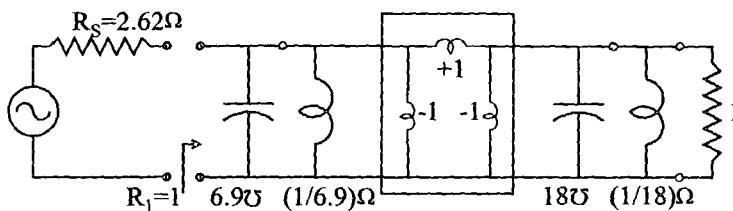


Figure 3.3.6. A normalized $N=2$ Chebyshev filter with an inductive inverter.

Cpar, and ohms. From that, it is seen that $g_{N+1} = 2.62$ ohms, and R_s and R_i must be related by (3.3.4). An arbitrary choice is to make the filter midband input resistance $R_i=1.0$ ohms, so that the proper mismatch exists at band center frequency of 1 rad/s when $R_s=2.62$ ohms. That explains the choice of inverter $Z_0=1$, according to (3.2.3). The loaded Q values in Table 3.3.2 are obtained in the resonators according to (2.3.2), i.e., $X_P=R_P/Q_P$, as shown in Figure 3.3.6.

The calculated passband response of the filter in Figure 3.3.6 is shown in Figure 3.3.7, where the passband distortion due to the inverter Z_0 's linear dependence on frequency is evident, even in a narrow (10%)

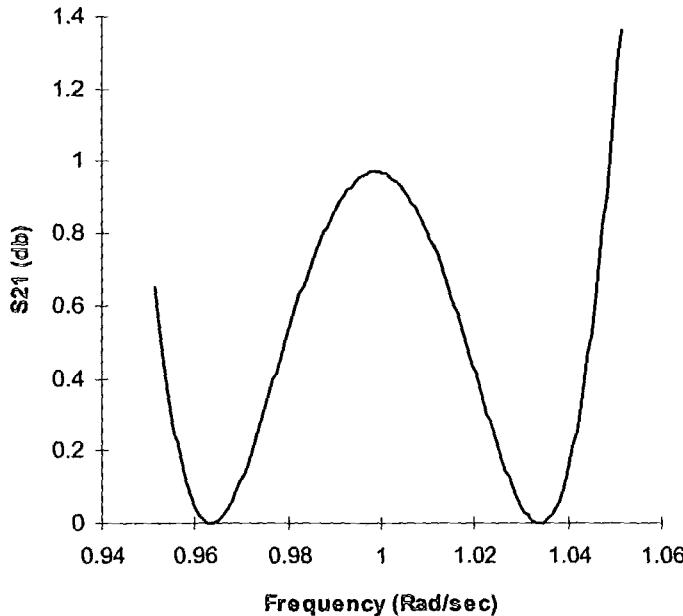


Figure 3.3.7. The distorted passband of an $N=2$ inductively-coupled filter.

band filter. The ideal passband extends from 0.95125 to 1.05125 rad/s according to (2.2.3), and those are the ends of the curve in Figure 3.3.7. At the two frequencies in Table 3.3.2 where the attenuation should be 20 dB, the inverter imperfection results in 18.40 and 21.61 dB for the lower and upper frequencies, respectively.

Example 3.3.6. Consider a three-resonator filter having 0.5 dB ripple in a 70% passband. Use two inductive inverters, as in Figure 3.3.3(b), and additional capacitive end couplings, as described in Example 3.3.2 and Figure 3.2.2. **Problem:** Design a normalized filter, and eliminate shunt capacitors at the ends and the shunt inductor in the middle. **Solution:** See Figure 3.3.8. Program ALLCHEBY shows that the loaded Q's are

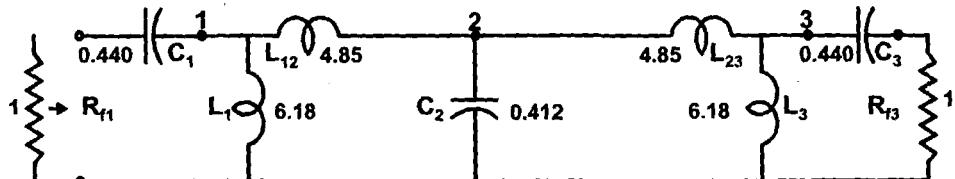


Figure 3.3.8. A filter with inductive inverters and capacitive end couplings.

$Q_1=Q_3=2.28$, and $Q_2=1.57$. The parallel resistances determine all element values. To eliminate the shunt capacitors in resonators 1 and 3, (3.2.11) shows that $R_1=R_3=1+2.28^2=6.20$ ohms. To eliminate the inductor in the middle resonator, (3.2.9) shows that $R_2=3.82$ ohms. Thus, (2.3.2) shows that $1/X_{II}=Q_2/R_2=0.412=C_2$ farads. Both inverter inductances are $\sqrt{(6.20 \times 3.82)}=4.87$ ohms. Also, (2.3.2) assigns the resonator reactances $X_I=X_{III}=6.20/2.28=2.72$ ohms. Combining each of those with the negative of the inverter reactances yields the shunt inductances in Figure 3.3.8: 6.18 henrys.

Figure 3.3.9 shows the terribly distorted response of this filter, which was designed using the narrow-band assumption that the inverters and end couplings are frequency independent (x markers). The

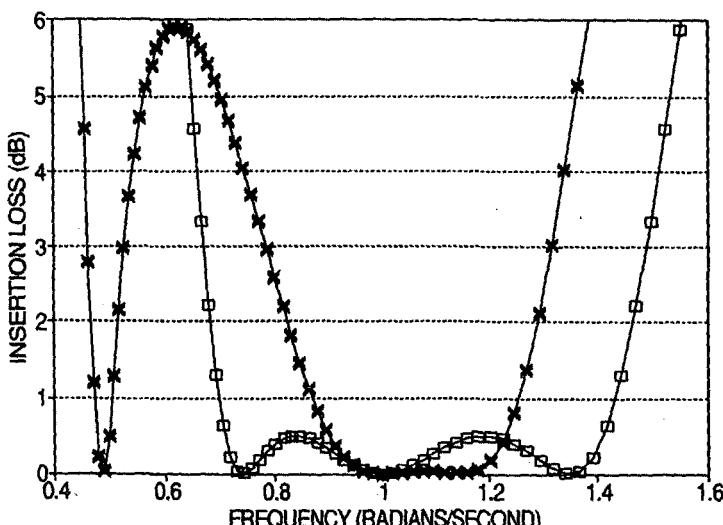


Figure 3.3.9. Passband of a 70% 0.5 dB filter with and without compensation.

undistorted response is shown by the \square markers. The next section shows what small changes in design procedure and element values produce the undistorted response.

3.4 Eliminating Passband Distortion

The way the inverter distortion can be corrected is shown in the next two sections, which deal with two- and three-resonator filters [Cuthbert,1996]. The solution is simply to stagger tune the resonators and end coupling by an amount that is trivial to calculate. Section 3.4.3 describes the procedure for replacing any series resonator by its exact shunt equivalent, similar to the narrow-band equivalent shown in Figure 3.2.4. That procedure is less confusing than designing by stagger tuning more than two or three resonators. Details of the Norton transformations are omitted, because the new results eliminate that complexity entirely.

3.4.1 Stagger-Tuning Two Resonators

Consider only resonators 1 and 2 in the bandpass prototype network shown in Figure 2.4.4. It is seen that there can be two adjacent capacitors (swap positions with the respective branch inductors) having values Q_2 and $1/Q_1$. Then insert an ideal transformer between capacitor $1/Q_1$ and inductor Q_1 and rescale the impedance of Q_1 and g_0 according to the turns ratio. The Norton transformation shown in Figure 2.4.11 enables replacement of the ideal transformer and the two capacitors by a Pi of capacitors that preserves the exact frequency response. The resulting topology is shown in Figure 3.4.1(c).

Figure 3.4.1(b) and (a) show the decomposition of the topology in (c) so that the capacitive Pi inverter (Figure 3.2.5) can be identified. Three surprising features are revealed:

1. Resonator inductor L_E in Figure 3.4.1(a) is repositioned in series as in (b), and $R_E=Q_E^2R_f$ defines the inverter impedance shown,
2. There is a trim capacitor, C_t , in Figure 3.4.1(b) that stagger tunes the input resonator below its resonant frequency, ω_0 , and
3. There is one inductive and one capacitive coupling with voltage transfer phases of less than 90° with opposing signs.

Repositioning L_E without changing its value has left the output resonator net capacitive and with a parallel resistance greater than R_E by exactly R_f according to $1+Q_E^2$. The inverter action according to (3.2.3) delivers a net inductive impedance to the input resonator. But that is exactly offset by C_t . The value of C_t is determined by considering that it has a loaded Q of $1/Q_E$, i.e. $C_t=1/(Q_ER_P)$.

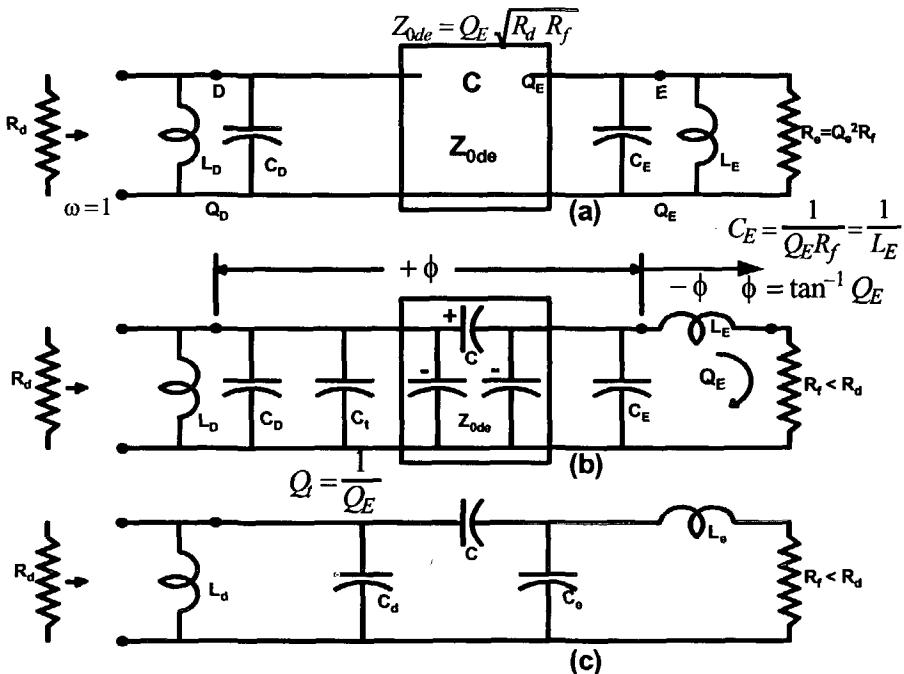


Figure 3.4.1. Undistorted prototype for $N=2$ with inductive end coupling.

Figure 3.4.2 shows an alternative replacement for resonators 1 and 2 in Figure 2.4.4. In this case, there is an inductive inverter and a capacitive top coupling analogous to the reverse situation in Figure 3.4.1. The relationships are also analogous. Repositioning C_E without changing its value has left the output resonator net inductive and with a parallel resistance greater than R_E by exactly R_f according to $1+Q_E^2$. The inverter action according to (3.2.3) delivers a net capacitive impedance to the input resonator. But that is exactly offset by L_t . The value of L_t is determined by considering that it has a loaded Q of $1/Q_E$, i.e. $L_t = R_p \times Q_E$.

The two topologies in Figures 3.4.1 and 3.4.2 modify series resonators at the ends of ladder networks by replacing one of the series elements (L or C) with a Pi subnetwork of like kind. They are remarkably similar in topology to the narrow-band end coupling designed by $1+Q^2$ as illustrated in Examples 3.3.2 and 3.3.6, but in the present stagger-tuned situation, there is no passband or stopband distortion. Note that as Q_E increases, the node-voltage phases approach $\pm 90^\circ$ and the trim elements tend to vanish, consistent with very narrow passband widths.

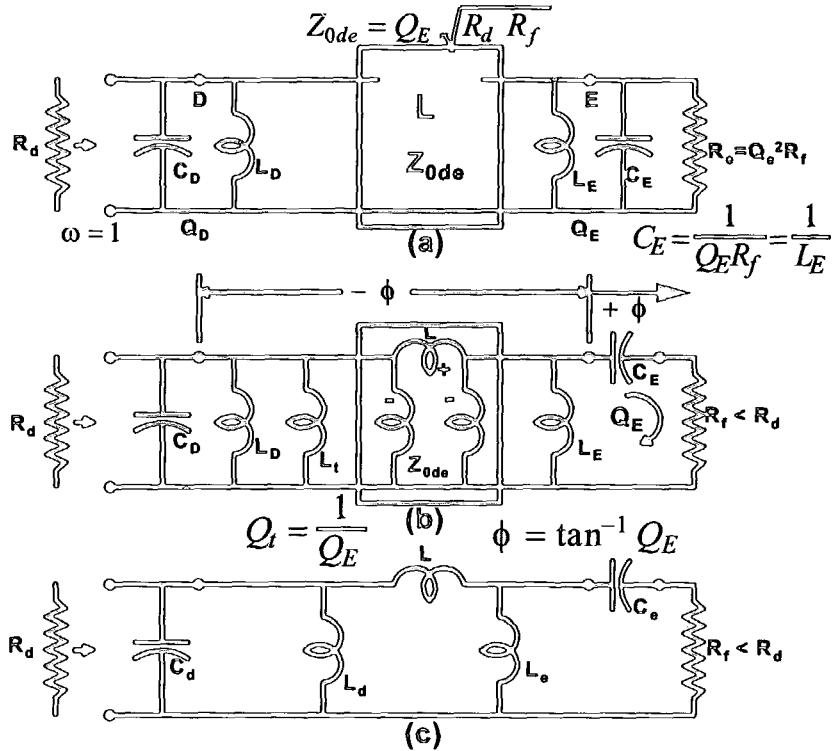


Figure 3.4.2. Undistorted prototype for $N=2$ with capacitive end coupling.

3.4.2 Stagger-Tuning Three Resonators

Consider only resonators 2, 3, and 4 in the bandpass prototype network in Figure 2.4.4. After insertion of ideal transformers, there are two adjacent capacitors and two adjacent inductors that can be replaced by their respective Pi networks according to the Norton transformation illustrated in Figure 2.4.11. The result in Figure 2.4.4 is a shunt LC pair from node 3 to ground, i.e., resonators 2, 3, and 4 have the appearance of three shunt resonators with one L and one C top coupling.

The direct-coupled $N=3$ topology obtained by two Norton transformations is partitioned into resonators, inverters, and two remaining elements as shown in Figure 3.4.3. There is one inductive inverter and one capacitive inverter and one trim element of like kind on their respective outboard sides, i.e. L_t and C_t . Qualitatively, resonator C is stagger tuned to a somewhat lower resonant frequency than ω_0 by C_t , i.e. more capacitive. Impedance Z shown in Figure 3.4.3 is somewhat inductive because of the Z_{0bc} inverter. Because resonator B is anti-resonant at ω_0 , inverter Z_{0ab} presents a somewhat capacitive impedance to resonator A, but trim inductor L_t staggers tunes that capacitance out.

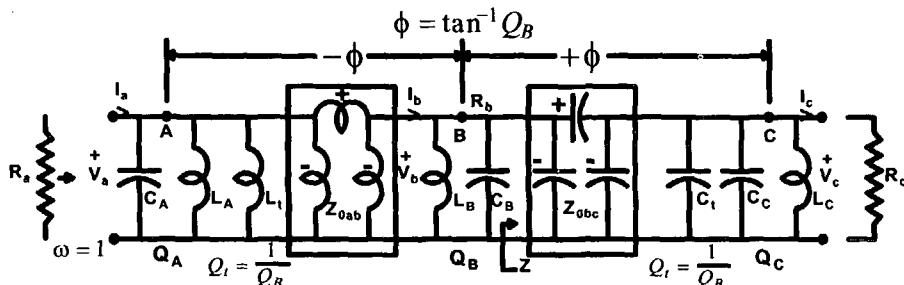


Figure 3.4.3. Undistorted prototype for $N=3$.

The effect of the stagger tuning is that there is no distortion of the ideal response at any frequency.

Figure 3.4.3 shows that the two trim elements are easily calculated: $C_t = 1/(Q_B R_c)$ and $L_t = Q_B R_a$. As in the two-resonator case, the node-voltage phases are less than 90° with opposing signs. Note that as Q_B increases, these phases approach $\pm 90^\circ$ and the trim elements tend to vanish, consistent with very narrow passband widths. Figure 3.4.3. could be turned end for end, of course.

3.4.3 Exact Replacement of Resonators

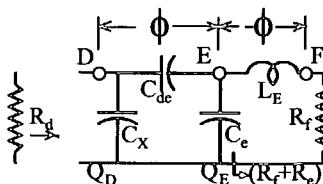
The stagger tuning concepts explained in the preceding two sections and [Cuthbert,1996] show how Norton transformations are applied to explicitly identify inverters and the adjustments necessary to eliminate passband distortion. Applying those stagger tuning results to more than three resonators would be confusing, so that the original concept of replacements for series (or parallel) resonators similar to Figure 3.2.4 is adopted. For example, compare resonators 2, 3, and 4 in Figure 2.4.4 to Figure 3.4.3; clearly all those elements between and including L_t and C_t compose an exact replacement for internal series resonator 3. (*Internal resonators* are those not on either end of the bandpass prototype network.) These new and useful results are summarized for easy application to design any direct-coupled filter that is free of distortion for any band width. Specific examples are provided in Sections 3.4.5 and 3.5.

3.4.3.1 Replacing External Series Resonators

Resonators 1 and 2 in Figure 2.4.4 are replaced by the equivalent networks with corresponding resonators E and D in Figure 3.4.1, and now all the network except resonator D is condensed into the subnetwork in Figure 3.4.4. The design starts with an external lowpass series inductor, g_i , connected to an end resistance g_0 or g_{N+1} (Figure 2.4.1), and that is converted to a series LC resonator scaled to $g_i \times Q_{BW}$. Leaving the series inductor connected to the load resistance, the series C is replaced by a Pi subnetwork composed of C_x , C_{de} , and C_e in Figure 3.4.4, which

includes their equations in terms of Q_E and parallel resistances. C_X is combined with the shunt resonator which also shares node D.

Figure 3.4.5 shows the analogous external resonator case, where the external series resonator L is replaced by the Pi subnetwork composed of L_X , L_{de} , and L_e ; their equations in terms of Q_E and parallel resistances are included. L_X is combined with the shunt resonator which also shares node D. In both the external resonator equivalences, there is one L and one C top coupling. Equations for shunt inverse inductances are employed to facilitate their vanishing, i.e., approaching zero, as well as for combining them with adjacent parallel inductors.



$$C_X = (1/Q_E - \sqrt{R_d/R_e})/R_d, \quad C_{de} = 1/\sqrt{R_e R_d}, \quad L_X^{-1} = (1/Q_E - \sqrt{R_d/R_e})/R_d, \quad L_{de} = \sqrt{R_d R_e}, \\ C_e = (Q_E - \sqrt{R_e/R_d})/R_e, \quad L_E = Q_E R_f, \quad L_e^{-1} = (Q_E - \sqrt{R_e/R_d})/R_e, \quad C_E = 1/(Q_E R_f), \\ R_e \equiv Q_E^2 R_f.$$

Figure 3.4.4. An exact replacement for the C in an end series resonator.

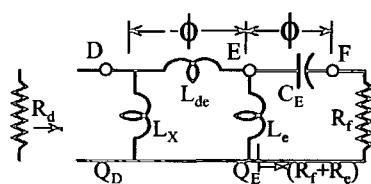
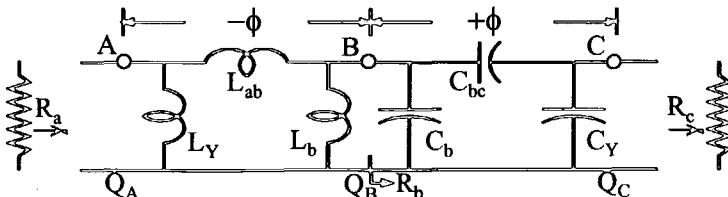


Figure 3.4.5. An exact replacement for the L in an end series resonator.

3.4.3.2 Replacing Internal Series Resonators

Resonators 2, 3, and 4 in Figure 2.4.4 are replaced by the equivalent networks with corresponding resonators A, B, and C in Figure 3.4.3, and now all the network between resonator A and resonator C is condensed into the subnetwork in Figure 3.4.6. This is an exact



$$L_Y^{-1} = (1/Q_B - \sqrt{R_a/R_b})/R_a, \quad L_{ab} = \sqrt{R_a R_b}, \\ L_b^{-1} = (Q_B - \sqrt{R_b/R_a})/R_b, \quad C_b = (Q_B - \sqrt{R_b/R_c})/R_b, \\ C_{bc} = 1/\sqrt{R_b R_c}, \quad C_Y = (1/Q_B - \sqrt{R_c/R_b})/R_c, \quad \phi = \tan^{-1} Q_B$$

Figure 3.4.6. An exact replacement for an internal series resonator.

replacement for a series resonator, as opposed to the approximate (narrow band) replacement shown in Figure 3.2.4, and one L and one C top coupling are mandatory.

The inverters are included in Figure 3.4.6 (see L_{ab} and C_{bc}), and the end elements L_Y and C_Y are combined with their respective shunt resonators, which share their respective nodes. Equations for shunt inverse inductances are employed to facilitate their vanishing, i.e., approaching zero, as well as for combining them with adjacent parallel inductors. The equations show that middle elements L_b and C_b in Figure 3.4.6 are not resonant at $\omega_0=1$ unless $R_a=R_c$.

3.4.3.3 Replacing Internal Series Resonators Having Traps

Figure 3.4.7 shows a lowpass to bandpass to series elliptic branch and its exact bandpass equivalent subnetwork. The lowpass prototype

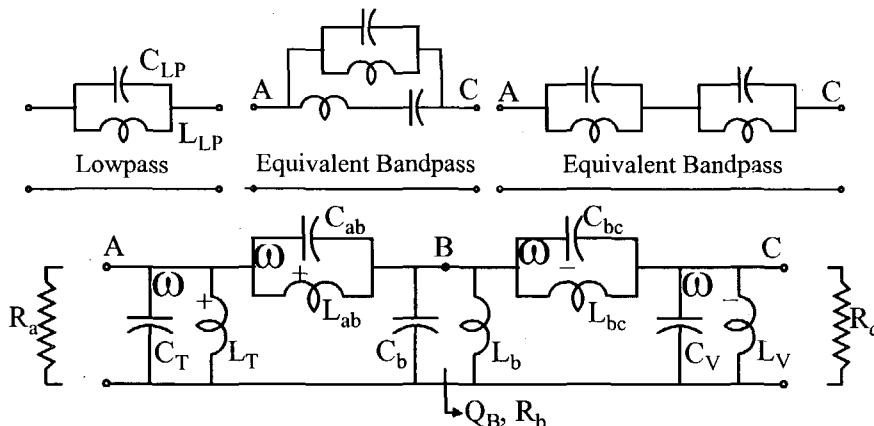


Figure 3.4.7. An exact replacement for a series elliptic branch.

inductor L_{LP} is elsewhere labeled g_i , and $C_{LP}=0$ is the case for all-pole filters having a monotonic stopband selectivity. When $0 < (L_{LP} \times C_{LP}) < 1$ in any lowpass prototype branch, a stopband zero of transmission occurs at their resonant (null) frequency. When that lowpass response is transformed to the conventional bandpass response by (2.4.3), the passband edge frequencies, ω_1 and ω_2 , are still computed by (2.2.3) and have the geometric symmetry of (2.2.1), i.e. $\omega_1 \times \omega_2 = 1$. The lowpass null frequency corresponds to two bandpass null frequencies that also have geometric symmetry about $\omega_0=1$. These null frequencies are produced by the two bandpass equivalent branches and the top-coupled equivalent subnetwork shown in Figure 3.4.7. Restating (2.5.2), the bandpass null frequencies are $\omega_- < \omega_1$ and $\omega_+ > \omega_2$:

$$\delta = 1 / \left(2Q_{BW} \sqrt{L_{LP} C_{LP}} \right), \quad (3.4.1)$$

$$\omega_- = \sqrt{\delta^2 + 1} - \delta, \quad \omega_+ = 1 / \omega_-.$$

Note that as $C_{LP} \rightarrow 0$, $\omega_- \rightarrow 0$.

The two inverter top-coupling branches, (C_{ab}, L_{ab}) and (C_{bc}, L_{bc}) , in Figure 3.4.7 produce the two zeros of transmission at the frequencies given by (3.4.1). At midband frequency $\omega_0=1$, branch $(C_{ab}, L_{ab}, \omega_+)$ appears to be inductive, and branch $(C_{bc}, L_{bc}, \omega_-)$ appears to be capacitive, with reactance values given by (3.2.7). Subnetwork elements C_T and L_T are combined with resonator A elements of like kind, and subnetwork elements C_V and L_V also are combined with resonator C elements of like kind. This subnetwork could be turned end-for-end, of course.

Table 3.4.1 summarizes the design equations for elements in Figure 3.4.7; do not overlook the two squared terms in the denominators. These equations were derived by applying a well-known transformation [Geffe, 1963]. For wide-band design, one or more pairs of trap inverters can be used, with their null frequencies placed for selectivity determined by a well-known pole-placing technique [Daniels]. Tables of element values and null frequencies for prototype elliptic filters are available [Zverev], and computer programs also have been provided [Amstutz], [Cuthbert, 1983:358-362].

Table 3.4.1. Element Equations Producing Nulls in Figure 3.4.7.

$$\begin{aligned} Q_B &= L_{LP} Q_{BW}, \quad L_T^{-1} = \left[\frac{1}{Q_B} \frac{(1+\omega^2)}{(1-\omega_-^2)^2} - \frac{\sqrt{R_a/R_b}}{(1-\omega_-^2)} \right] / R_a, \quad C_T = \omega_-^2 L_T^{-1}, \quad L_{ab} = (1-\omega_-^2) \sqrt{R_a R_b}, \\ C_{ab} &= \omega_-^2 / L_{ab}, \quad L_b^{-1} = \left[Q_B - \frac{\sqrt{R_b/R_a} + \omega_-^2 \sqrt{R_b/R_c}}{(1-\omega_-^2)} \right] / R_b, \quad C_b = \left[Q_B - \frac{\sqrt{R_b/R_c} + \omega_-^2 \sqrt{R_b/R_a}}{(1-\omega_-^2)} \right] / R_b, \\ C_{bc} &= 1 / [(1-\omega_-^2) \sqrt{R_b R_c}], \quad L_{bc} = 1 / (\omega_-^2 C_{bc}), \quad C_V = \left[\frac{1}{Q_B} \frac{(1+\omega^2)}{(1-\omega_-^2)^2} - \frac{\sqrt{R_c/R_b}}{(1-\omega_-^2)} \right] / R_c, \quad L_V^{-1} = \omega_-^2 C_V. \end{aligned}$$

The Geffe transformation has also been applied to derive a topology similar to the lower half of Figure 3.4.7 [Sabin]. However, that technique differs from this direct-coupled filter design method, where retention of a resonator between trap coupling is desired. The Sabin approach designs odd orders of elliptic bandpass filters with emphasis on absorbing the stray capacitance to ground between the traps.

3.4.4 All Possible All-Pole Topologies

Figure 3.4.8 duplicates Figure 2.4.7 to emphasize the possible combinations of top-coupling elements indicated by the “^” symbols. It is now clear from Figures 3.4.4 through 3.4.7 that the inductive and capacitive pairs of couplings can be employed in several combinations. These are shown in Table 3.4.2 (end-for-end not counted). For example, Figure 3.4.8 corresponds to the single N=4 line in Table 3.4.2; it is composed of resonator A followed by Figure 3.4.6, followed by resonator C

(the same as resonator D), followed by either Figure 3.4.4 or Figure 3.4.5.

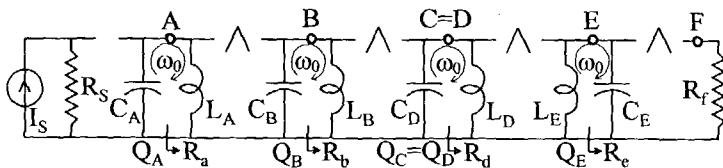


Figure 3.4.8. Direct-coupled network topology with only parallel resonators.

Table 3.4.2. Statistics for All Possible Direct-Coupled Filters.

N = Number of Resonators

NE = Number of Elements

NK = Number of Top Couplings (^)

KK = Number of Top Coupling Pair Combinations (LC, CL, LC, etc.)

N	TOPOLOGIES	NK	KK	Min NE	Max NE	Surplus NE
2	D^E^	2	2	4	5	1
3	A^B^C	2	2	6	8	2
	^E_1^D_1=D_2^E_2^	4	4	6	8	2
4	A^B^C=D^E^	4	4	8	11	3
5	A_1^B_1^C_1=A_2^B_2^C_2	4	4	10	14	4
	^E_1^D_1=A^B^C=D_2^E_2^	6	8	10	14	4
6	A_1^B_1^C_1=A_2^B_2^C_2=D^E^	6	8	12	17	5
7	A_1^B_1^C_1=A_2^B_2^C_2=A_3^B_3^C_3	6	8	14	20	6
	^E_1^D_1=A_1^B_1^C_1=A_2^B_2^C_2=D_2^E_2^	8	16	14	20	6

A particular N=4 case is shown in Figure 3.4.9. The minimum number of elements listed in Table 3.4.2 corresponds to a *canonic form*.

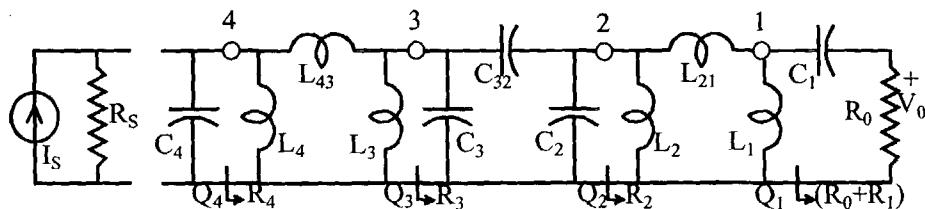


Figure 3.4.9. An N=4 direct-coupled network that can be distortion free.

3.4.5 Wideband Choices in Parallel Resistance Space

Wider passbands are possible when there is no passband distortion. That makes it more likely that negative elements may occur unless the parallel resistances have been chosen properly. This section describes the feasible regions for positive elements that always exist in parallel resistance space for the external and internal replacements for series resonators.

Table 3.4.2 shows that the two-resonator case occurs when resonator D is connected to the series-resonator equivalent subnetworks shown in either Figure 3.4.4 or Figure 3.4.5. That results in the networks shown in Figures 3.4.1 and 3.4.2, respectively, and the equations for their element values are shown in Table 3.4.3 for the normalization $R_f=1$ ohm. The choices in the boxes in Table 3.4.3 show that, for $R_f=1$, positive element values occur when R_d is within a range on a line, and some elements vanish when R_d is at the extremes of its range.

Table 3.4.3. Element Equations for the Undistorted N=2 Cases.

Figures 3.4.1 & 3.4.4

$$\begin{aligned} C_d &= [Q_D - (\sqrt{R_d} - 1)/Q_E]/R_d, \\ C_e &= (1 - 1/\sqrt{R_d})/Q_E, \\ D_{de} &= 1/(Q_E \sqrt{R_d}), \\ L_d &= R_d/Q_D, \quad \& \\ L_E &= Q_E. \quad \text{Thus,} \end{aligned}$$

$$(1 + Q_E Q_D)^2 \geq R_d \geq 1.$$

Figures 3.4.2 & 3.4.5

$$\begin{aligned} L_d &= R_d [Q_D - (\sqrt{R_d} - 1)/Q_E]^{-1}, \\ L_e &= Q_E (1 - 1/R_d)^{-1}, \\ L_{de} &= Q_E \sqrt{R_d}, \\ C_d &= Q_d/R_d, \quad \& \\ C_e &= 1/Q_E. \quad \text{Thus,} \end{aligned}$$

$$(1 + Q_D Q_E)^2 \geq R_d \geq 1.$$

Table 3.4.2 shows that a three-resonator case occurs when resonators A and C are connected at the ends of the series-resonator equivalent subnetwork shown in Figure 3.4.6. When all paralleled elements of like kind are combined, the resulting network is shown in Figure 3.4.10. The corresponding equations for the element values are shown in Table 3.4.4.

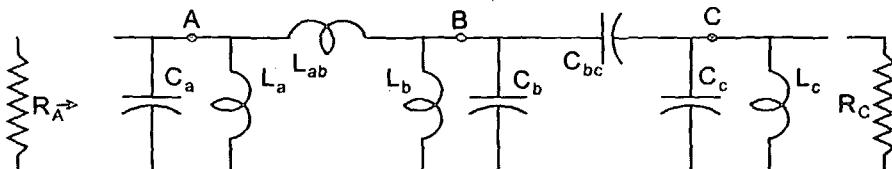


Figure 3.4.10. An N=3 network with an L and a C internal couplings.

Table 3.4.4. Element Equations for the First N=3 Case

$$\begin{aligned} L_a &= R_a (Q_A + 1/Q_B - \sqrt{R_a/R_b})^{-1}, \\ L_b &= R_b (Q_B - \sqrt{R_b/R_a})^{-1}, \\ C_b &= (Q_B - \sqrt{R_b/R_c})/R_b, \\ C_c &= [Q_C + 1/Q_B - \sqrt{R_c/R_b}]/R_c, \\ L_{ab} &= \sqrt{R_a R_b}, \\ C_{bc} &= 1/\sqrt{R_b R_c}, \\ C_a &= Q_A/R_a, \quad \& \\ L_c &= R_c/Q_C. \end{aligned}$$

Figure 3.4.11 shows the feasible region in parallel resistance space where all the elements in Figure 3.4.10 are positive. The four boundary

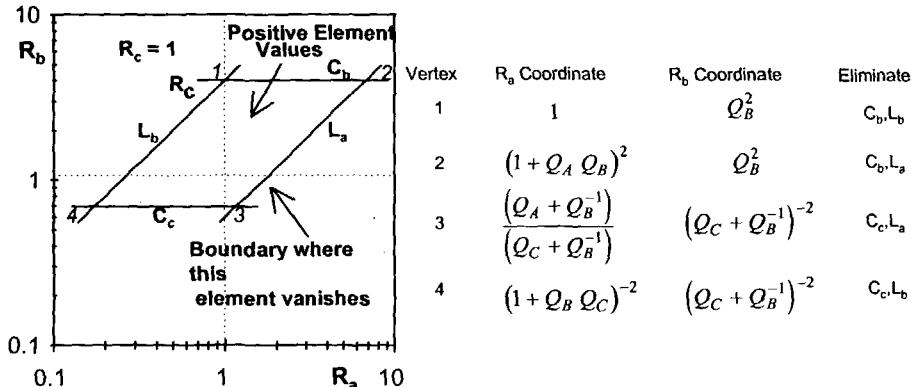


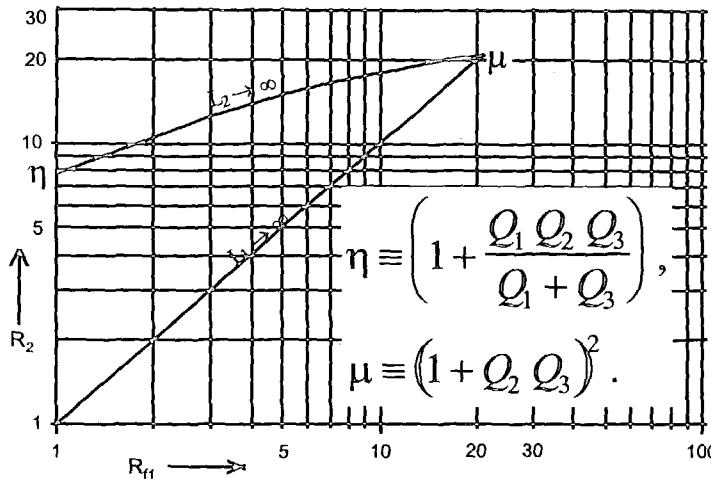
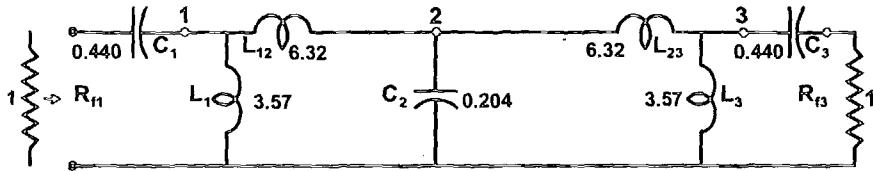
Figure 3.4.11. Feasible regions in parallel-resistance space for $N=3$.

lines represent loci where the indicated elements vanish. The table in Figure 3.4.11 shows the pairs of elements that vanish at each vertex and the vertex coordinates. Note that the relationships in Table 3.4.4 and in Figure 3.4.11 involve only parallel resistances that are related by the strictly-positive resonator loaded Q values. That guarantees that the feasible region cannot be empty. Logarithmic scaling of parallel resistances is advantageous for both graphing and computation.

Example 3.4.1. Design a three-resonator filter having the topology in Figure 3.3.8 but with exact 0.5 dB ripple over a 70% pass band width. (Example 3.3.6 designs the same filter based on the constant-reactance inverters and $1+Q^2$ end coupling.) *Problem:* Replace both external series resonators with an exact equivalent shunt resonator, and force the net shunt inductance in the middle of the filter to vanish. *Solution:* Start by considering $N=3$ in the lowpass prototype network of Figure 2.4.1(b). Program ALLCHEBY shows that the loaded Q 's are $Q_1=Q_3=2.28$, and $Q_2=1.57$ in the bandpass topology in Figure 2.4.4. Replace the two series resonators with the exact equivalent subnetworks in Figure 3.4.5. Resonator 2 has a shunt inductance equal to $L_{II}=R_2/Q_2$, where $Q_2=Q_D$ and $R_2=R_d$, and is flanked by two equal shunt inductors, L_x in Figure 3.4.5. Combining these three shunt inductances algebraically, the equation for the net inverse inductance is found to be

$$L_2^{-1} = \frac{Q_1 Q_2 Q_3 + (Q_1 + Q_3) - \sqrt{R_2 \left(\frac{Q_1}{R_{j3}} + \frac{Q_3}{R_{j1}} \right)}}{R_2 Q_1 Q_3}. \quad (3.4.2)$$

Figure 3.4.12 shows the feasible region for positive $L_1=L_3$ and L_2 as functions of R_2 versus $R_{j1}, R_{j3}=1$ assumed. The feasible region is a three-

Figure 3.4.12. Feasible region for positive $L_1=L_3$ in Example 3.4.1.Figure 3.4.13. Final $N=3$ filter in Example 3.4.1.

sided region bounded by a 45° line, $R_{ff} \geq 0$, and the monotone-increasing concave arc define by (3.4.2). See Figure 3.4.13; L_2 would appear in parallel with C_2 , and for L_2 to vanish, $L_2 \rightarrow 0$ in (3.4.2). That can occur for a wide range of input resistance R_{ff} values; in this case $R_1=1$ with the resulting value of parallel resistance R_2 at node 2 as shown in Figure 3.4.13. Therefore, $R_2=7.783$, which also determines the values of the other elements as shown in Figure 3.4.13. The response of this network is that shown in Figure 3.3.9, the curve with the \square markers. The ridiculous response shown by the x markers is that in Example 3.3.6, which corresponds to the element values in Figure 3.3.8. That has exactly the same topology, and element values that are similar. Clearly, both narrow and wide-band equal-ripple designs should utilize the exact equivalent subnetworks in Figures 3.4.4-3.4.7.

3.5 The Method of Choices

Spreadsheets are advocated for design of all the various topologies of both narrow- and wide-band direct-coupled networks. Like most programming situations, it is only necessary to check the correctness of one or two solutions manually; then many other solutions can be

examined for numerous choices of parameters, i.e., "what if". The defined parallel resistance space is an excellent basis for parameter variation in addition to varying passband width, flat loss, ripple, or selectivity parameters. Several PC spreadsheets include convenient "point-and-click" optimization features that greatly simplify minimization of element max/min ratios, forcing some elements to standard values, making groups of elements have the same value, etc. The wideband techniques described in Section 3.4 allow a designer to examine all possibilities while implicitly maintaining the exact response characteristics of the network. Many of these advantages are illustrated in the following sections.

3.5.1 A Spreadsheet for Four Resonators

Table 3.4.2 shows that for four resonators there are KK=4 top coupling combinations of LC pairs. Choose the LCLC combination from source to load as shown in Figure 3.4.9, and compare that to Figure 3.4.8. That shows that resonators 4, 3, and 2 in Figure 3.4.9 correspond to nodes A, B, and C in Figure 3.4.8, and nodes A and C are connected by the equivalent subnetwork in Figure 3.4.6. Similarly, resonators 2 and 1 correspond to resonators D and E, and those are connected by the equivalent subnetwork in Figure 3.4.5.

Each Kth resonator's reactance is related to its Q and parallel resistance by (2.3.2): $X_K = R_K/Q_K$. Those reactances can be combined with their adjacent parallel components of like kind in the subnetworks either numerically or by equation. For example, consider the situation at node 2 in Figure 3.4.9 and detailed in Figure 3.5.1. In addition to C_D and L_D

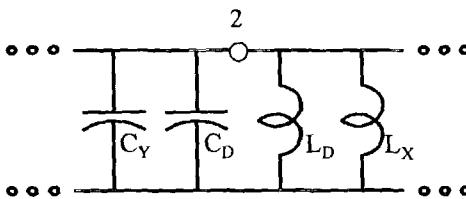


Figure 3.5.1. Design details at node 2 in Figure 3.4.9.

in resonator D, there is C_Y from Figure 3.4.6 and L_X from Figure 3.4.5. The optional equations for each element's net value are shown in Table 3.5.1.

Table 3.5.1. Optional Equations and Constraints in Figure 3.4.9.

$$\begin{aligned}
 C_1 &= 1/(Q_1 R_0), \quad L_1^{-1} = (1 - \sqrt{R_0/R_2})/(Q_1 R_0) \Rightarrow L_1 \geq 0 \text{ if } R_2 \geq R_0, \quad L_{21} = Q_1 \sqrt{R_2/R_0}, \\
 L_2^{-1} &= (Q_1 Q_2 + 1 - \sqrt{R_2/R_0})/(Q_1 R_2) \Rightarrow L_2 \geq 0 \text{ if } R_2 \leq R_0 (1 + Q_1 Q_2)^2, \quad C_{32} = 1/\sqrt{R_3 R_2}, \\
 C_2 &= (Q_2 + 1/Q_3 - \sqrt{R_2/R_3})/R_2 \Rightarrow C_2 \geq 0 \text{ if } R_3 \geq R_2 (Q_2 + 1/Q_3)^{-2}, \quad L_{43} = \sqrt{R_4 R_3}, \\
 C_3 &= (Q_3 - \sqrt{R_3/R_4})/R_3 \Rightarrow C_3 \geq 0 \text{ if } R_3 \leq R_2 Q_3^2, \quad C_4 = Q_4/R_4, \\
 L_3^{-1} &= (Q_3 - \sqrt{R_3/R_4})/R_3 \Rightarrow L_3 \geq 0 \text{ if } R_3 \leq R_4 Q_3^2, \\
 L_4^{-1} &= (Q_4 + 1/Q_3 - \sqrt{R_4/R_3})/R_4 \Rightarrow L_4 \geq 0 \text{ if } R_3 \geq R_4 (Q_4 + 1/Q_3)^{-2}.
 \end{aligned}$$

By observing Figure 3.4.9 and Table 3.5.1, it is noted that some elements cannot vanish (e.g., C_1 , L_{21} , etc.) and others can (e.g., L_1 , C_2 , etc.), depending on choices of parallel resistances. Constraints on the parallel resistances to keep elements positive are also shown in Table 3.5.1; they are derived from the equations for the respective elements. Again, it is emphasized that writing element and constraint equations such as those in Table 3.5.1 is optional, because the net values of the elements can be summed numerically from their several constituent parts. Similarly, the constraints can be enforced by equations and/or numerically by the optimization features included in spreadsheet programs.

It is instructive to utilize the optional element constraint equations in Table 3.5.1 to plot the feasible region shown in Figure 3.5.2. There can

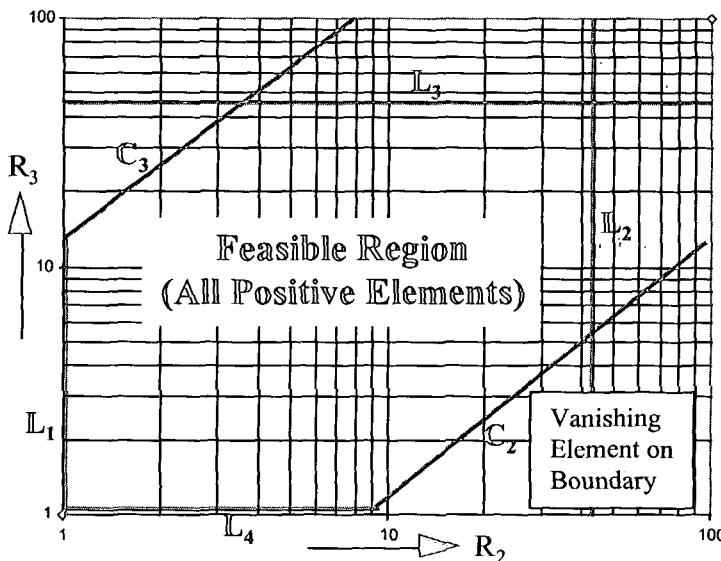


Figure 3.5.2. A feasible region for positive elements in Figure 3.4.9.

only be three parallel resistances involved in this case ($R_1=Q_1^2R_0$, according to Figure 3.4.5). By normalizing $R_0=1$, the other two parallel resistances can be visualized without loss of generality. It is usually best to utilize logarithmic scales for the two remaining parallel resistances, in this case R_3 versus R_2 in Figure 3.5.2, which is similar to Figure 3.4.11. Again, it is clear that the feasible region always exists, and that one or two elements may be made to vanish if desired.

Also note that the maximum possible value of R_4/R_0 is available as the product of the upper bounds on R_2 , R_3 , and R_4 given in Table 3.5.1 in terms of any set of loaded Q values. Therefore, one reason to go to the trouble of writing the constraint equations is to relate the bandwidth of

any response shape to the maximum possible transformation ratio while maintaining positive values for elements.

Figure 3.5.3 shows an EXCEL® version 5.0 spreadsheet that incorporates the equations in Table 3.5.1. Lines 3-10 contain filter

	A	B	C	D	E	F	G	H	I	J
1	File ---.XLS Example Using $A^2B^2C^2=D^2E^2$ with LCCL Couplings									
2	INPUT									
3	Load R0 =	1	Input R4 =	3.6889	Omega 0	Omega 1	Omega 2			
4	%BW =	50	QBW =	2.0000	1	0.7808	1.2808	2 P I/O		
5	10-MHz =	100.00	f1 MHz =	78.0776	f2 MHz =	128.0776			6.28E+08	
6	i =	1	2	3	4	5	NODES	V-V DEGS		
7	g1 =	1.1088	1.3061	1.7703	0.8180	1.3554	4-3	-74.23		
8	Qi =	2.2176	2.6122	3.5406	1.6360		3-2	74.23		
9	R1 (FIXED)		R2	R3	R4 (FIXED)		2-1	-65.73		
10	Ohms =	4.9177	4.4624	5.4099	3.6889		1-0	65.73		
11	VARIABLES	NORM UNNORM(R0=50)			FOR POSITIVE VALUES:					
12	None	C1 =	0.4509	14.35 pF						
13	R2, R3	C2 =	0.4451	14.17	R3 GE	0.5326				
14	R2, R3	C32 =	0.2035	6.48						
15	R2, R3	C3 =	0.4509	14.35	R3 LE	55.9400				
16	None	C4 =	0.4435	14.12					SOLVER:	
17	R2	L1 =	4.2111	335.11 nH	R2 GE	1.0000				
18	R2	L21 =	4.6845	372.78						
19	R2	L2 =	2.1143	168.25	R2 LE	46.1423				
20	R3	L3 =	2.3223	184.80	R3 LE	46.2435				
21	R3	L43 =	4.4673	355.50						
22	R3	L4 =	3.3760	268.65	R3 GE	1.0023				
23		MAX/MIN C's =	2.2156							
24		MAX/MIN L's =	2.2156							
25		WORST RATIO =	2.2156							

Optimized

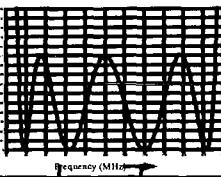


Figure 3.5.3. A spreadsheet for four-resonator all-pole filters.

specifications and calculated quantities that are not peculiar to the several LC coupling combinations. Lines 12-22 relate to the respective elements: column B lists the element names (Figure 3.4.9), column A shows which parallel resistances are independent variables for the respective elements, columns C and D show the normalized and unnormalized element values, and columns F and G show the constraints on parallel resistances for the elements to be positive. Columns I and J show the relative phases of the node voltages at the band center frequency. The values for R_2 and R_3 are arbitrary.

Example 3.5.1. Design an all-pole fourth-degree bandpass transformer to match a 50-ohm load to a 250-ohm source with a mismatch of no more than 0.1 dB over a 50% bandwidth centered at 100 MHz. *Problem:* This is essentially a filter, but the use of bandpass inverters simplifies the required impedance transformation. Minimize the spreads of L and of C values. *Solution:* The cross-hatched rectangles in Figure 3.5.3 show where data must be entered for particular problems that utilize this topology. Program ALLCHEBY provides the required lowpass prototype element values, and they are shown in Figure 3.5.3 with the corresponding resonator loaded Q values. The source resistance, $g_5=1.3554$, is not unity because of the mismatch at band center frequency;

see Section 3.3.3. The cells in rows 23-25 result from the standard spreadsheet operation to scan for extreme values in a range of cells and then to report both the ratios of extremes and the greater ratio.

A powerful alternative to manually experimenting with values of R_2 and R_3 to affect element values is to use the optimization capabilities included in the spreadsheet. For EXCEL®, the TOOLS menu offers the SOLVER optimizer, where cell D25 can be designated as the “target” cell and cells D10 and E10 as the “changing” cells. In OPTIONS, select automatic scaling, quadratic estimates, central differences, and Newton search. When started from cell values of $R_2=2$ and $R_3=3$, the optimizer will minimize the worst of both the L’s and C’s ratios to 2.2156 as shown in Figure 3.5.3. (Solutions are somewhat dependent on starting resistance values). Both sets of normalized and unnormalized element values are immediately available. The passband response, including the planned mismatch at band center, is included in Figure 3.5.3. It does not change as R_2 and R_3 are varied.

Example 3.5.2. Singly-terminated filters can be designed using a spreadsheet designed for a particular topology. *Problem:* Design a normalized fourth-degree singly-terminated all-pole filter having a maximally-flat (Butterworth) shape over a 50% bandwidth centered at 100 MHz. *Solution:* The table in [Matthaei,1964:107] provides the lowpass prototype element values: $g_1=0.3827$, $g_2=1.0824$, $g_3=1.5772$, and $g_4=1.5307$. The input resistance, R_4 , can have any value up to the limit defined by the inequalities in Table 3.5.1. As described in Section 2.1.5, let $R_4/R_0=9$ at the passband center frequency, so that (2.1.13) yields $|V_4/V_0|=3$. Therefore, for a 50-ohm load, an input resistance of $R_4=450/150=9$ ohms is entered on the spreadsheet in Figure 3.5.3 for the topology in Figure 3.4.9.

As a way to utilize the two degrees of freedom, R_2 and R_3 , the spreadsheet optimizer can enforce the constraints $L_1=L_2$, $L_2=L_3$, and $L_3=L_1$ so that $L_1=L_2=L_3$. The resulting element values and response shape are shown in Figure 3.5.4. This singly-terminated filter can be

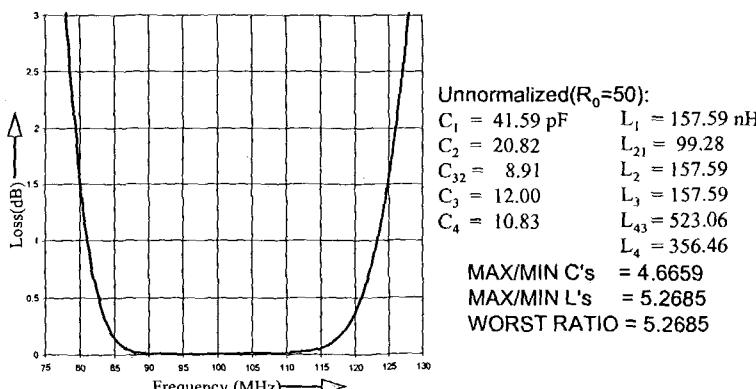


Figure 3.5.4. Passband and element values for the filter in Example 3.5.2.

employed in either of the two orientations shown in Figures 2.1.9 or 2.1.10.

3.5.2 A Spreadsheet for Five Resonators

The spreadsheet in Figure 3.5.5 was created for a topology including the coupling sequence CLCLLC from source to load, Figure 3.5.6. There are four data entry boxes enclosing the name and value cells: %BW; f_0 GHz; five g_i prototype values; and three independent parallel resistances, R_b , R_c , and R_d .

Example 3.5.3. Provide a fifth-degree doubly-terminated filter having 40% bandwidth at 1 GHz and make maximum use of magnetic coupling. **Problem:** Design the N=5 topology with capacitive top coupling at both ends in Table 3.4.2. **Solution:** The second N=5 entry in Table 3.4.2 shows that the external resonator equivalent subnetwork from Figure 3.4.5 is required at both ends, and the equivalent internal subnetwork from Figure 3.4.6 is required in the middle. This arrangement is programmed in the spreadsheet in Figure 3.5.5. The upper network in Figure 3.5.6 can be converted to utilize the three transformers shown along with six capacitors in the lower network.

In the spreadsheet in Figure 3.5.5, the 40% can be changed to any

Using ^E1^D1=A^B^C=D2^E2^							
$Rf1 = Rf2 = 1 \text{ ohm normalized}$							
INPUT			Omega 0	Omega 1	Omega 2		
%BW =	40	QBW =	2.5000	1	0.8198	1.2198	2PIf0
$f_0 \text{ GHz} =$	1.0000	$f_1 \text{ GHz} =$.8198	$f_2 \text{ GHz} =$	1.2198		6.28E+09
i =	1	2	3	4	5	NODES	V-V DEGS
$g_i =$	1.1468	1.3712	1.9750	1.3712	1.1468	A-B	73.74
$Q_i =$	2.8670	3.4280	4.9375	3.4280	2.8670	B-C	78.55
A (FIXED)	B	C	D	E (FIXED)		C-D	78.55
$R_i =$	8.2197	4.0000	8.2195	4.0000	8.2197	D-E	73.74
VARIABLES		NORM UNNORM(50-ohm)	FOR POSITIVE VALUES:				
None	$C_a =$	0.3488	1.1103	pF			
None	$C_e =$	0.3488	1.1103				
R_d	$C_d =$	0.8570	2.7279				
R_b, R_c	$C_{bc} =$	0.1744	0.5551				
R_b, R_c	$C_c =$	0.4263	1.3570	$R_b/R_c \text{ GE}$	0.0410	$R_b \text{ GE}$.3372
R_b, R_c	$C_b =$	0.7332	2.3340	$R_b/R_c \text{ LE}$	13.1808	$R_b \text{ LE}$	108.3391
R_b	$L_a =$	5.7340	45.6297	nH			
R_c, R_d	$L_{cd} =$	5.7339	45.6292				
R_d	$L_{de} =$	5.7340	45.6297				
R_b	$L_a =$	5.7340	45.6297	$R_b \text{ GE}$	1.0000		
R_d	$L_e =$	5.7340	45.6297	$R_d \text{ GE}$	1.0000		
R_b	$L_b =$	1.2990	10.3374	$R_b \text{ LE}$	117.2472		
R_c, R_d	$L_c =$	2.3457	18.6667	$R_d/R_c \text{ GE}$	0.0410	$R_d \text{ GE}$.3372
R_c, R_d	$L_d =$	1.5479	12.3179	$R_d \text{ LE}$	32.5394		
MAX/MIN C's =		4.9140					
MAX/MIN L's =		4.4140					
WORST RATIO =		4.9140					

Figure 3.5.5. A spreadsheet for the N=5 filter in Example 3.5.3.

other desired bandwidth. The values 1.1468, 1.3712, etc. can be the lowpass prototype element values for any 5th-order filter (e.g., the

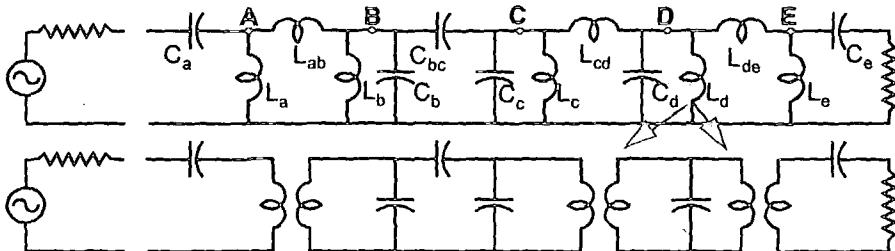


Figure 3.5.6. An $N=5$ CLCLLC network adapted for magnetic couplings.

Chebyshev 0.1 dB ripple doubly-terminated filter) normalized to 1 ohm and 1 radian per second.

The three parallel resistances (lowest hatched box) are arbitrary within the limits defined in lower right-hand columns once any value of R_c is entered. Then entry of values for R_b and R_d within the limits shown produces positive values for all filter elements, normalized and unnormalized to f_0 and equal 50-ohm terminations. Setting resistances R_b and/or R_d to their limits causes certain elements to vanish.

There are equations (not shown) that relate the element values to the three parallel resistances. That dependence is indicated in the spreadsheet column alongside the element labels; e.g., C_{bc} depends on R_b and R_c but not R_d . Choosing any set of parallel resistances does not affect the response of the filter, only the element values. The ratios of extreme values of C 's and L 's and the worst of those two ratios are shown at the bottom of the spreadsheet in Figure 3.5.5. The relative phases of node voltages A through E are also displayed. For example, see Figure 3.5.5: the phase of the voltage across inductor L_b lags that across inductor L_a by 73.74 degrees. That information is often useful for tuning and verification.

The greatest value of this design approach is the guaranteed exact filter response at any point in the R_b , R_c , R_d parallel resistance space. This allows adjustment of those three values within known limits to obtain the most acceptable set of positive element values. In this case, note that setting $R_b=R_d=4.000$ produces $L_{ab}=L_{de}=L_a=L_e=5.7340$ (normalized). (This is only obvious numerically, but it is consistent with the equations in Figure 3.4.5.) None of those L 's is a function of R_c but L_{cd} is, so by the equation for L_{cd} , (3.2.3), set $R_c=5.7340^2/4=8.2195$. The spreadsheet in Figure 3.5.5 produces all the element values shown, especially the five equal inductances. The conversion from Pi's of inductors to transformers indicated in Figure 3.5.6 can be accomplished with the equivalence shown in Figure 2.4.10; see [Zverev:529]. In particular, the transformer between nodes D and E in Figure 3.5.6 is obtained by splitting L_d into two parallel inductors. Then the primary and secondary transformer windings, Figure 2.4.10, also could be

programmed on the spreadsheet and re-optimized using the three degrees of freedom: R_b , R_c , and R_d .

3.5.3 A Spreadsheet for a Four Resonator Elliptic Filter

The lowpass prototype in Figure 3.5.7 defines the direct-coupled filter in Figure 3.5.8. The spreadsheet in Figure 3.5.9 is similar to that in Figure 3.5.3 except that trap couplings from Figure 3.4.7 are utilized.

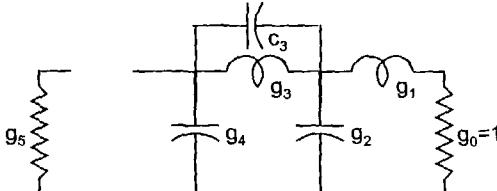


Figure 3.5.7. A degree 4 lowpass antimetric elliptic filter for Example 3.5.4.

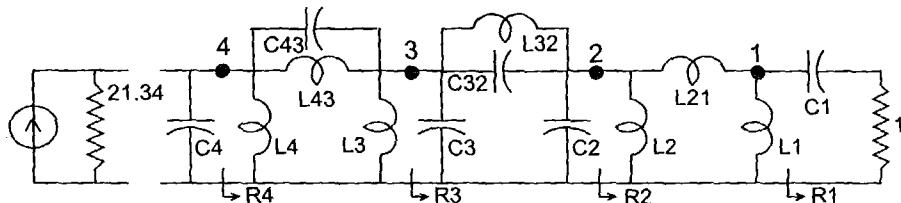


Figure 3.5.8. A four-resonator antimetric bandpass filter for Example 3.5.4.

ELLIPTIC FILTER TYPE C N=4 Example Using A^B^C=D^E^ with LCLC Couplings						
INPUT	Unnorm'ing R Ratios = 9					
Load R0 = 1	Input R4 = 21.3333	Omega 0	Omega 1	Omega 2		
%BW = 43.4123	QBW = 2.3035		1	0.8062	1.2403	2 P If0
f0 MHz = 161.25	f1 MHz = 130.0039	f2 MHz = 200.0061				1.01E+09
i = 1	2	3	4	5		
Trap C = .2359					NODES	V-V DEGS
gi = 0.8358	1.2531	1.0546	0.6373	1.0000	4-3	-67.63
Qi = 1.9252	2.8865	2.4293	1.4680		3-2	67.63
R1 (FIXED)	R2	R3	R4 (FIXED)		2-1	-62.55
Ohms = 3.7065	17.9178	10.7730	21.3333		1-0	62.55
VARIABLES	NORM	UNNORM(R0=9)		Nulls = 105.6796	246.0414	MHz
None	C1 = 0.5194	56.96 pF				
R2, R3	C2 = 0.1358	14.90	50			
R2, R3	C32 = 0.1262	13.84	45			
R2, R3	C3 = 0.0497	5.45	40			
R3	C43 = 0.0497	5.45	35			
R3	C4 = 0.0556	6.09	30			
R2	L1 = 2.5207	22.39 nH	25			
R2	L21 = 8.1493	72.39	20			
R2, R3	L2 = 17.6901	157.14	15			
R2, R3	L32 = 18.4530	163.92	10			
R2, R3	L3 = 17.9606	159.54	5			
R3	L43 = 8.6484	76.82	0			
R3	L4 = 26.3565	234.13	0			
MAX/MIN C's = 10.4586				Radians/Second		
MAX/MIN L's = 10.4560						
WORST RATIO = 10.4586						

Figure 3.5.9. A spreadsheet for the N=4 elliptic filter in Example 3.5.4.

Example 3.5.4. A maximally-selective 43% bandwidth filter centered at 161.25 MHz is desired to absorb both a load impedance consisting of a 56.96 pF capacitor in series with a 9-ohm resistor and a source impedance consisting of a 2.64 pF capacitor in parallel with a 192-ohm resistor. This is similar to an all-pole filter requirement posed by [Borlez]. *Problem:* Design a wideband fourth-degree elliptic filter with stopband selectivity and passband ripple that absorbs the given terminations. *Solution:* Although there are analytical all-pole broadband impedance matching programs, there are none for elliptic-function networks. However, a lowpass antimeric elliptic filter design program by [Amstutz] is available [Cuthbert,1983:356,352]. One lowpass transmission zero (trap) in a 4th-degree filter with an arbitrary stopband rejection of 30 dB is chosen; see Figure 3.5.7.

The corresponding bandpass network is shown in Figure 3.5.8. It is constructed by replacement of the g_3-C_3 lowpass branch by its bandpass equivalent in Figure 3.4.7 and by replacement of the g_1 lowpass branch by its bandpass equivalent in Figure 3.4.5. The series loaded Q of the given load at band center is $Q_1=1.925$ and the parallel loaded Q of the given source at band center is $Q_5=0.5142$. The one degree of freedom utilized in Amstutz' program is to vary the ripple in the passband (0 to 1 rad/s) until the given Q of the load is obtained. Normalized to a 1-ohm load and $Q_{BW}=2.3035$, that requires $g_1=Q_1/Q_{BW}=0.8358$, according to (2.3.2) and (2.4.4). A passband ripple of 0.0529 dB produces the required value for g_1 . Fortunately, $Q_4=g_4 \times Q_{BW}$ exceeds the Q_s required for the source.

The lowpass element values in Figure 3.5.7 appear on the spreadsheet in Figure 3.5.9. Note: $g_5=1$, which indicates that this is a modified elliptic-function filter of type b [Cuthbert,1983:351]. The spreadsheet shows that the unnormalized passband extends from 130 to 200 MHz, and the pair of zeros of transmission (trap null frequencies) occur at 105.7 and 246.0 MHz; both these frequency pairs are geometrically symmetric about the passband center frequency. The SOLVER optimization feature is used as described in Example 3.5.1 to minimize the worst of both the L's and C's ratios; that is shown to be 10.4586 in Figure 3.5.9. All corresponding element values are shown; in particular, $C_4=6.09$ pF includes the required 2.64 pF that belongs to the parallel-RC source. The frequency response is included in Figure 3.5.9, where the 30-dB stopband selectivity is clearly shown.

The spreadsheet utilized in Figure 3.5.9 also includes the two universal blocks of hidden cells that are shown in Table 3.5.2; these function as subroutines to calculate all N=2 and N=3 resonator subnetwork element values. The upper block solves the N=3 subnetwork given the six independent parameters %BW through R_c . Twelve dependent values calculated by equations (3.4.1) and Table 3.4.1 for the

%BW	L lowpass	C lowpass	Ra	Rb	Rc
43.4122	1.0546	.2359	21.3333	10.7730	17.9178
QB	LT inv	Lab	Lb inv	Cbc	CV
2.4293	-.0309	8.6484	.0557	.1262	-.0253
Omeg-sqd	CT	Cab	Cb	Lbc	LV inv
.4295	-.0133	.0497	.0497	18.4530	-.0108
%BW	L lowpass	Rd	Rf	QE	
43.4122	0.8358	17.9178	1	1.9252	
LX inv	Lde	Le inv	Re	CE	
-.0937	8.1493	.3967	3.7065	.5194	

Table 3.5.2. Subroutines for All N=2 and N=3 Element Values.

subnetworks in Figure 3.4.7 are returned in the fourth and sixth rows. The value labeled “Omeg-sqd” is in fact ω_-^2 from (3.4.1). It is noted in connection with (3.4.1) and the lowpass subnetwork in Figure 3.4.7 that as $C_{LP} \rightarrow 0$, $\omega_- \rightarrow 0$. This means that the subroutines in Table 3.5.2 that are in the CHOICES.XLS spreadsheet apply to both the elliptic (traps) and all-pass (Figure 3.4.6) N=3 subnetworks, with $\omega_- = 0$ for the latter. The lower four rows in Table 5.3.2 apply for both N=2 subnetworks according to the equations and Figures listed in Table 3.4.3.

3.6 Tuning

There are two basic tuning techniques that are applicable for direct-coupled filters of any bandwidth. Also, there are ways to measure the loaded Q values and the equivalent coupling coefficients of a filter that is properly tuned. The underlying phenomena are described below.

3.6.1 Alternating Open- and Short-Circuit Method

This method relies on the narrow-band prototype direct-coupled network shown in Figure 3.2.1, especially the action of the inverters between the resonators. At the band center frequency, ω_0 , each inverter is in resonance and, therefore, is an open circuit. When node II is short-circuited to ground, the inverter Z_{012} presents an open circuit at its input, i.e. $R_1 \rightarrow \infty$ in Figure 3.2.1, in accordance with (3.2.3). Similarly, if only node III is short-circuited to ground, then $R_2 \rightarrow \infty$ and $R_1 \rightarrow 0$.

As originally described [Dishal,1951], the source at frequency ω_0 is connected to the input of the filter and a high-impedance voltmeter is also connected to measure the voltage from node I to ground. The load

impedance is also connected. All the resonators are completely detuned. Then:

1. Tune resonator I for maximum voltage at node I,
2. Tune resonator II for minimum voltage at node I (input node),
3. Tune resonator III for maximum voltage at node I, and ...
4. Tune the last resonator for a maximum or minimum, depending on whether it is an odd or even numbered node, respectively.

If the resonators cannot be completely detuned (so they are nearly a short circuit), then the resonator nodes can be shorted to ground by suitable conducting straps.

Various considerations for tuning microwave filters using this method have been published [Matthaei]. Also, it has been observed that the end resonators are likely to be less accurately tuned by this method, and it is those end (terminal) resonators that have the greatest effect on passband SWR. Therefore, any unexpected SWR should be tuned by varying the end-resonator tuning after the open-short procedure has been completed. The final response should be that corresponding to the realized loaded-Q distribution, which is equivalent to the consecutive coupling coefficients defined by (3.1.1). Dissipative elements affect the response shape; see Section 2.2.6.

3.6.2 Reactive Input Reflection Function

It has been noted recently that the group delay of a reflection coefficient at the input of an open- or short-circuited network contains all the information required to tune direct-coupled filters [Ness]. For example, consider the three-resonator filter in Figure 3.3.3(a). When resonators II or III are short-circuited, the input impedance is $Z_{in}=0+jX_{in}$. So reflection coefficient (2.1.10) is defined wrt R_{11} , the desired midband input resistance of the filter (not wrt the source resistance, which may differ from R_{11}):

$$\Gamma_k \equiv \frac{jX_{in} - R_{11}}{jX_{in} + R_{11}}, \quad k = 1, 2, \dots, N, \quad (3.6.1)$$

where there are N resonators. Index k denotes that resonators 1 through k are not short circuited. Reflection coefficient Γ_k can be measured easily and accurately as S_{11} using a vector network analyzer, and the loci are on the perimeter of a Smith chart, because there is no real part ($|\Gamma|=1$).

Of interest is the group delay of Γ_k , which is similar to the ordinary transfer group delay and is as easily measured:

$$\Gamma_{dk} = -\frac{d\phi}{d\omega}, \quad (3.6.2)$$

where ϕ is the angle associated with Γ_k . To measure Γ_{d1} , for example, resonator 2 in Figure 3.3.3(a) is shorted, and input reactance X_{in} at ω_0 produces the reflection coefficient in (3.6.1) and its group delay in (3.6.2). The remarkable relationships to resonator loaded Q's are shown in Table 3.6.1.

Table 3.6.1. Input Reflection Coefficients versus Loaded Q Values.

$$\frac{\omega_0}{4}\Gamma_{d1} = Q_1, \quad \frac{\omega_0}{4}\Gamma_{d2} = Q_2, \quad \frac{\omega_0}{4}\Gamma_{d3} = Q_1 + Q_3, \quad \frac{\omega_0}{4}\Gamma_{d4} = Q_2 + Q_4,$$

$$\frac{\omega_0}{4}\Gamma_{d5} = Q_1 + Q_3 + Q_5, \quad \frac{\omega_0}{4}\Gamma_{d6} = Q_2 + Q_4 + Q_6, \quad \dots$$

The relations in Table 3.6.1. can be restated for loaded Q's as functions of reflection group delay:

$$Q_k = \frac{\omega_0}{4} (\Gamma_{dk} - \Gamma_{d(k-2)}), \quad k = 3, 4, \dots, N. \quad (3.6.3)$$

These reflection coefficients remain accurate in the presence of dissipative elements as long as the unloaded $Q_u > 10Q_k$ [Atia]. Also, see the measurement method by [Drozd] described in Section 5.2.6.

Example 3.6.1. Confirm the loaded Q values of the direct-coupled filter in Figure 3.3.3. *Problem:* Simulate measured S_{11} and confirm the loaded Q values. *Solution:* An analysis program was modified to compute the group delay in (3.6.2) by first-order finite difference in frequency (0.01% increment). Using the element values in Table 3.3.1 for a lossless network, it was found that $\Gamma_{d1}=0.06919$, $\Gamma_{d2}=0.1382$, and $\Gamma_{d3}=0.1382$ microseconds. Therefore, Table 3.6.1 and (3.6.3) show that $Q_1=5.4340$, $Q_2=10.8581$, and $Q_3=5.4192$. These values agree with the design values in Example 3.3.1. The unloaded $Q_u=100$ for the inductors was added to the simulation; the new reflection values were $\Gamma_{d1}=0.06939$, $\Gamma_{d2}=0.1374$, and $\Gamma_{d3}=0.1386$.

3.6.3 Narrow-Band Reflection Poles and Zeros

The conventional coupling coefficients between resonators are just the reciprocals of the mean loaded Q's (3.1.1). An accurate way to measure them for narrow-band filters has been described [Ness]. For bandwidths less than about 10%, it is well known that the bandpass reactance transformation having geometric symmetry is approximately linear:

$$\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) \approx \frac{2}{\omega_0} (\omega - \omega_0). \quad (3.6.4)$$

This fact is used in conjunction with the poles and zeros of reflection coefficient (3.6.1) to determine the coupling coefficients [Atia]. The

procedure is again to short and open circuit the resonators in sequence, but this time measuring the frequencies at which the S_{11} Smith chart loci intersect the real axis at the short- or open-circuit ends of the chart perimeter (because $|\Gamma|=1$). The reflection phase intercepts at $\arg(\Gamma)=0^\circ$ are the open-circuit or pole frequencies, and the intercepts at $\arg(\Gamma)=180^\circ$ are the short-circuit or zero frequencies. Those frequencies are used in simple equations to obtain the coupling coefficients, $K_{i,i+1}$; see [Ness].

3.6.4 Wideband Networks Having Exact Responses

Wideband direct-coupled filters have certain resonators detuned to eliminate the effects of inverter frequency dependence. Consider the topology in the spreadsheet for four resonators, Section 3.5.1 and Figure 3.4.9. Recall that the amount of resonator detuning is shown separately for the subsections in Figures 3.4.1, 3.4.2, and 3.4.3 in terms of trim capacitors C_t and inductors L_t .

One way to tune a wideband filter is first to ignore the stagger tuning and then use the open-short methods of Section 3.6.1 and 3.6.2, which are valid for any passband width. The appropriate resonators can be retuned to obtain the design response function versus frequency. In the topology shown in Figure 3.4.9, the load will be shorted and will provide accurate information for adjusting C_1 and/or L_1 , as well as providing information on the loaded Q and related coupling coefficient values. The success of the entire tuning operation may be confirmed by measuring the node-voltage phases, which are available from the design process (Figure 3.5.3).

3.7 Summary of Direct-Coupled Filters

Coupling coefficients began as the ratio of adjacent capacitors in the classical bandpass filter topology, consisting of alternating series and parallel LC resonators. That ratio was also applied to filters having resonators of one kind, either all parallel or all series, coupled by constant-reactance electric or magnetic elements. Thus, the coupling capacitors, inductors, or magnetic linkages were assumed independent of frequency, which limited passband widths to less than 20 per cent. The lowpass (normalized) coupling coefficients were scaled by Q_{BW} for the corresponding direct-coupled bandpass topologies. Published tables of terminal Q values (Q_1 and Q_n composed of the end elements and the respective resistive terminations), along with the sequence of coupling coefficients, provided all network element values. Uniformly dissipative elements were derived by predistortion of the response poles, and tables of filter design data were arranged in descending ratios of element unloaded Q_u to Q_{BW} .

A more comprehensive and flexible design concept adds inverters as the coupling mechanism, i.e. lossless frequency-independent 90°

transmission lines connecting the resonators. Ideal inverters flanking a series resonator make a two-port network exactly equivalent to a parallel resonator. Physical inverters include Pi networks of three inductors or of three capacitors, the shunt branches being negative elements that are absorbed into adjacent positive resonator elements of like kind. The Pi branch reactance magnitude is equal to the inverter Z_0 , and the inverter input impedance is simply Z_0^2/Z_L , Z_L being the impedance loading the inverter. This simple inverting action allows design of the filter at its midband frequency, where all the parallel resonators are anti-resonant and therefore transparent. Thus, the load resistance is inverted to the parallel resistance at the next resonator, and so on, to obtain the input resistance. Because resonator loaded Q's do not change with impedance scaling, the choices of parallel resistances provide control of resonator and inverter reactances and, therefore, all element values, without affecting the frequency response.

Stopband selectivity is highly visible in terms of the loaded Q product and the inverter frequency dependence, the latter being either directly or inversely proportional to frequency. Because one or more inverters can be anti-resonant traps tuned in the stopband, the selectivity can be enhanced asymmetrically with an offsetting reduction in required loaded Q product. The passband width is easily related to the resonator loaded Q product, so filter design can be based on either passband or stopband requirements. Passband flat loss is obtained by mismatch of the source and input resistances. The source resistance relative to one ohm is included in tables of lowpass prototype network element values. The ratios of individual resonator loaded Q's to their unloaded Q factors simply estimate the power dissipated in the resonators in dB, so that it is easy to predict the total midband dissipative loss. It turns out that dissipative loss in the stopband (>20 dB) adds to a reduced mismatch (reactive) loss, to sum to the original selectivity of the lossless filter, and so is of no consequence in the stopband.

The simplifying assumption of frequency-independent inverters (coupling elements) causes passband distortion, which is most severe for equal-ripple response shapes. Recently, it has been shown that passband distortion can be eliminated by stagger-tuning certain resonators by an easily calculated amount. A more organized approach is to utilize equivalent two-port networks that are exact replacements for external or internal series resonators, or for internal series resonators containing traps (pairs of zeros of transmission). Now it is possible to design wideband direct-coupled filters that have no distortion but do require equal numbers of capacitive and magnetic top couplings.

The parallel resistances (at midband frequency) constitute an ideal parameter space for control of element values without affecting the ideal response shape. A feasible region of parallel resistances always exists so that all elements can be positive. Spreadsheets incorporating point-and-

click optimization capability are ideal for enforcing various constraints and obtaining results otherwise unavailable, e.g., minimizing both the max/min ratios of L's and C's. The element values and constraints may be programmed on the spreadsheets by direct numeric summations or by evaluating optional equations that are obtainable by algebra. The equations also show the maximum possible load-source impedance transformation range in terms of the chosen loaded Q values. The method of choices does not require any polynomial algebra or synthesis techniques, and its use is more organized than the equivalent Norton transformations.

Finally, tuning direct-coupled filters is an organized process at midband frequency, based on inverter action and/or input reflection group delay or phase pole-zero intercepts. Information is available on the loaded Q and coupling values actually realized in the physical network as described in Section 3.6.2.

4. Comprehensive Equal-Ripple Filters —

4.1 Purpose

This chapter is a bridge between the highly structured approach to direct-coupled filter design in the preceding chapter and the pragmatic numerical procedures for broadband impedance matching in the next chapter. It illustrates what generality was sacrificed and exhibits phenomena that support efficient algorithms for broadband matching.

The zeros of transmission frequencies uniquely determine the entire frequency response of a Chebyshev equal-ripple filter, and that response can be computed easily. The bandpass filters of the last chapter have equal numbers of transmission zeros at dc and at infinite frequency, and optionally at one or more at finite frequencies. There are rigid constraints on the types of coupling between resonators and at the terminals. Those constraints can be avoided, but at a price of design complexity, namely, polynomial synthesis.

The basic steps of polynomial synthesis are described in terms of mathematical complexity and crucial decisions that may or may not avoid unacceptable results. The consequences of assuming particular transfer functions are noted, especially for the case of load constraints, which constitutes broadband matching.

Behavior of network reflectance, the magnitude of the input reflection coefficient, is observed as a function of each element's variation about its nominal value at sampled passband frequencies for typical filter and broadband match cases. The intrinsic characteristics of those functions show why an existing alternative to conventional synthesis is so much better conditioned and easily implemented. That method, iterated analysis, is described briefly; several of its features are employed for broadband matching in Chapter 5, and its optimization aspects are more thoroughly developed in Chapter 6.

4.2 Response Continuum

It has been shown that if the number and frequencies of transmission zeros are known, then the response of a filter that has an equal-ripple passband is known exactly in both the pass band and all of the stop band [Daniels]. A small computer program, RIPFREQS.EXE, will be used to demonstrate how easy it is to obtain and display these critical characteristics.

4.2.1 Transducer and Characteristic Functions

The *transducer function*, $H(\omega)$, relates source/load voltages of a doubly-terminated two-port network. See Figure 4.2.1. In terms of

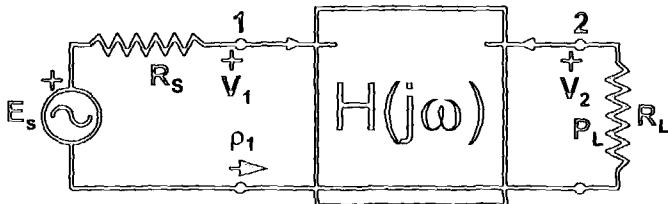


Figure 4.2.1. Doubly-terminated filters have resistances at both ports.

maximum power available from the source and power delivered to the load, the transducer function is

$$|H(\omega)|^2 = \frac{P_{as}}{P_L} = 1 + \xi^2 + |K(\omega)|^2. \quad (4.2.1)$$

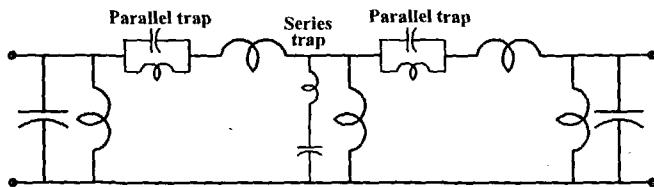
The transducer function is the inverse of the power ratio considered in Section 2.1.1, a convention chosen to ensure that $H(\omega) \geq 1$ for any ω . The *characteristic function* is $K(\omega)$ in (4.2.1), and it is generally a rational function, as is $H(\omega)$. When $K(\omega)=0$, $|H(\omega)|^2 = 1+\xi^2$ is the flat loss described in Section 2.2.4. For purposes of an equal-ripple passband, $K(\omega) \equiv \epsilon T(\omega)$, where $T(\omega)$ is a Chebyshev polynomial of the first kind, and ϵ is the ripple factor. These Chebyshev polynomials of various degrees all oscillate in an amplitude range between -1 to +1 within the normalized domain of $-1 \leq \omega \leq +1$. When $|T(\omega)|=1$, then the passband ripple is $10\log_{10}(1+\epsilon^2)$ dB. When $K(\omega)$ has a denominator polynomial, the denominator zeros, the poles of $K(\omega)$, produce transmission zeros (attenuation poles) at stopband frequencies.

To clarify what is meant by *Chebyshev* equal ripple, the *capture property* of Chebyshev polynomials is stated [Carlin,1998:307]. Consider the even positive polynomial of degree $2n$, $|K(\omega)|^2 = \epsilon^2 T_n^2(\omega)$, having maximum ripple factor ϵ^2 within a lowpass band $-1 \leq \omega \leq +1$. No other even positive polynomial of degree $2n$ which is greater than $|K(\omega)|^2$ outside the passband has a ripple factor less than ϵ^2 within the passband. See Figure 2.2.1: that is an $n=4$ Chebyshev passband, because it has exactly n valleys, i.e. where $|K(\omega)|^2 = 0$. Also, $|K(\omega)|^2 = \epsilon^2 T_n^2(\omega)$ has $n+1$ degrees of freedom, counting ripple factor ϵ .

Other "equal-ripple" responses can be obtained by certain frequency transformations [Daniels:103], but they do not satisfy the capture property. From here on, equal ripple means *Chebyshev* equal ripple. A test is to draw a horizontal line through the ripples: there must be N intercepts. Any number less than N is not Chebyshev equal ripple. For example, type c elliptic-function filters do not have Chebyshev equal-ripple passbands, but type b filters do. Both types have equal-ripple stopbands. See [Cuthbert,1983:351].

4.2.2 Transmission Zeros

Consider the ladder network in Figure 4.2.2. Transmission zeros occur when a short circuit exists in parallel or an open circuit exists in series in the network. The network has three traps, i.e. branches that cause transmission zeros at the respective trap null frequencies. There are two parallel resonant traps in series branches and one series resonant trap in a parallel branch. Count the number of trap zeros of transmission as $NT=3$.



Loss Zeros at DC ($NZ=1$), at Infinity ($NIN=3$), and at finite frequencies ($NT=3$). Filter Order $N=NZ+NIN+2NT=10$.

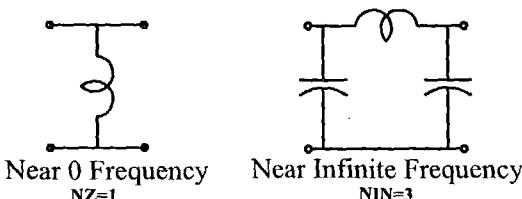


Figure 4.2.2. Ladder network with transmission zeros $NZ=1$, $NIN=3$ and $NT=3$.

Consider how the network in Figure 4.2.2 appears at frequencies very near dc: The parallel LC traps in series are essentially short circuits, and the series LC traps in parallel are open circuits; therefore, traps can be ignored. Working from left to right, the shunt C's and the series L's can be ignored, leaving just three shunt L's in parallel. But that is the same as the one L shown in Figure 4.2.2, so count the number of transmission zeros at zero frequency (dc) as $NZ=1$.

Consider how the network in Figure 4.2.2 appears as frequency approaches infinity: The traps can be ignored again. Working from left to right, the shunt L's can be ignored, leaving just the Pi network shown in Figure 4.2.2. Therefore, count the number of transmission zeros at infinite frequency as $NIN=3$. A more detailed discussion is available [Daniels:308].

As a consequence of the known transmission zeros for the network in Figure 4.2.2, it and its several equivalent networks must have the frequency response shown in Figure 4.2.3. The advantages of this remarkable situation are listed in Table 4.2.1 and are described in the following sections.

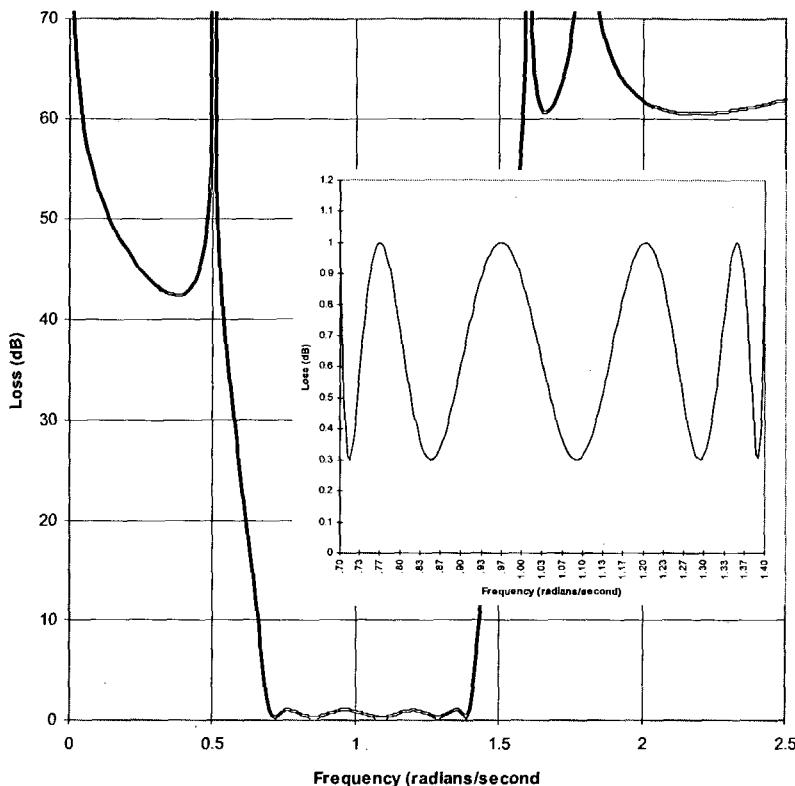


Figure 4.2.3. The unique frequency response of the network in Figure 4.2.2.

Table 4.2.1. Benefits of Knowing All Transmission Zeros of a Comprehensive Equal-Ripple Filter.

- Exact Lowpass or Bandpass Loss Function vs Frequency Is Determined,
- Bandpass Filters Do Not Require Geometric Frequency Symmetry,
- Bandpass Filters Need Not Be Based on a Lowpass Prototype Network,
- Loss Poles Can Be Asymmetric and Placed Automatically,
- The Passband Ripple Is Captured, and Has N Horizontal Intercepts for N+1 Adjustable Parameters (An Nth-Order Network),
- There are $N \setminus 2$ (integral part) Non-Zero Passband Valley Frequencies Which Can Be Calculated Easily and Accurately,
- Iterative Adjustment of Network Values at Frequencies of **Zero Loss** Yields Higher Accuracy Than Network Synthesis!

4.2.3 Passband Selectivity

It has been shown [Daniels:166-8] that there is a sinusoidal composite-function behavior in the pass bands of Chebyshev equal-ripple filters with an argument that is a nonlinear function of frequency. These relationships are shown in Table 4.2.2, where ω_a and ω_b are the lower and upper passband edge frequencies, respectively; $\omega_a=0$ for lowpass.

Table 4.2.2. Equal-Ripple Passband Equations in Terms of Transmission Zeros.

$$\text{Loss} = 10 \log_{10} \left[1 + \xi^2 + \varepsilon^2 T^2(\beta) \right] \text{dB.}$$

- Lowpass (NZ=0 & NIN Even) or Bandpass: (NZ+NIN Even): $T(\beta) = \cos(\beta)$,
- Lowpass (NZ=0 & NIN Odd): $T(\beta) = \sin(\beta)$,
- Flat Loss = $10 \log_{10} (1 + \xi^2) \text{dB}$.
- Ripple = $10 \log_{10} (1 + \varepsilon^2) \text{dB}$.

$$\begin{aligned} \beta &= NZ\beta_Z/2 + NIN\beta_{IN}/2 + \sum_1^{NT} \beta_i; \quad Z_i \equiv \sqrt{(\omega_i^2 - \omega_b^2)/(\omega_i^2 - \omega_a^2)}, \\ |Z| &\equiv \sqrt{(\omega_b^2 - \omega^2)/(\omega^2 - \omega_b^2)}, \quad \beta_i = \tan^{-1} \left[(-2|Z|Z_i)/(|Z|^2 - Z_i^2) \right], \quad \beta_{IN} = \tan^{-1} \left[(-2|Z|)/(|Z|^2 - 1) \right], \\ \beta_Z &= \tan^{-1} \left[\left(-2|Z|\omega_b / (\omega_a + 10^{-10}) \right) / \left(|Z|^2 - \left(\omega_b / (\omega_a + 10^{-10}) \right)^2 \right) \right], \quad \tan^{-1} \text{ is ATAN2}. \end{aligned}$$

The filter order or degree, N, is equal to

$$N \equiv NZ + NIN + 2NT, \quad (4.2.2)$$

where the numbers of transmission zeros are NZ at zero, NIN at infinity, and NT traps at finite frequencies. For example, the network in Figure 4.2.2 has N=10 as shown.

Consider the passband selectivity equations in Table 4.2.2 that are evaluated and plotted in Figure 4.2.3 for the network in Figure 4.2.2. The equations in Table 4.2.2 include the arbitrary flat-loss and ripple parameters, ξ and ε , respectively. Only β determines the frequencies of the passband peaks and valleys of the characteristic function, which is $|K(\omega)|^2 = \varepsilon^2 T^2(\beta)$, where β is the somewhat obscure function of frequency, ω , in Table 4.2.2. However, the function $\beta(\omega)$ is a one-to-one mapping, so its inverse is unique. Thus, the multiples $\beta=k\times(\pi/2)$, $k=1, 2, \dots, N+1$, can be found by *secant iterative search* on the frequency variable, ω [Cuthbert,1987:235]. That search algorithm, in program RIPFREQS.EXE, was found to converge within 5 to 10 iterations with an accuracy of 11 significant figures, as tested against an N=15 degree problem given by [Orchard].

Example 4.2.1. Consider the network in Figure 4.2.2. *Problem:* Find the frequencies of all the peaks and valleys in the passband for $0.7 \leq \omega \leq 1.4$ rad/s. Trap frequencies are at 0.5, 1.6, and 1.8 rad/s. *Solution:* Program RIPFREQS.EXE is run as shown in Table 4.2.3. Recall that the valleys are the zeros of the characteristic function.

Table 4.2.3. RIPFREQS.EXE Output for Example 4.2.1.

Passband lower, upper limits in rad/s =?	0.7, 1.4		
Number of zeros at dc (NZ: Lp & Cs branches) =?	1		
Number of zeros at infinity (NIN: Ls & Cp branches) =?	3		
Number of null (LC trap) frequencies =?	3		
Null frequency 1 (rad/s) =?	0.5		
Null frequency 2 (rad/s) =?	1.6		
Null frequency 3 (rad/s) =?	1.8		
Passband Loss Peaks (max) & Valleys (min):			
#	Radians/sec	Beta Degs	
1	0.70000000000	0.0000	PEAK
2	0.71605679461	90.0000	VALLEY
3	0.76600554266	180.0000	PEAK
4	0.85138964151	270.0000	VALLEY
5	0.96534181782	360.0000	PEAK
6	1.08983233414	450.0000	VALLEY
7	1.20382972986	540.0000	PEAK
8	1.29342515442	630.0000	VALLEY
9	1.35449345471	720.0000	PEAK
10	1.38896474129	810.0000	VALLEY
11	1.40000000000	900.0000	PEAK

4.2.4 Stopband Selectivity

It also has been shown [Daniels:165-6] that there is an uncomplicated function for loss in the stop bands of Chebyshev equal-ripple filters. These equations are displayed in Table 4.2.4 for lowpass or bandpass filters. These equations are also evaluated in program

Table 4.2.4. Stopband Loss Equations for Equal-Ripple Passband Filters.

$$\text{Loss} = 10 \log_{10} \left[1 + \xi^2 + \frac{\varepsilon^2}{4} \left(L + \sqrt{L} \right)^2 \right] \text{dB}, \quad Z = \sqrt{\frac{(\omega_a^2 - \omega_b^2)}{(\omega_a^2 - \omega_a^2)}},$$

$$L \equiv \left| \frac{\omega_a Z + \omega_b}{\omega_a Z - \omega_b} \right|^{NZ/2} \times \left| \frac{Z + 1}{Z - 1} \right|^{NIN/2} \times \prod_{i=1}^{NT} \left| \frac{Z + Z_i}{Z - Z_i} \right|$$

RIPFREQS.EXE as an option after the results in Table 4.2.3. The passband and stopband selectivity for as many as 251 frequencies in

either a linear or logarithmic progression between specified limits can be seen and stored in an ASCII file for plotting, such as in Figure 4.2.3.

Because it is easy to calculate the stopband selectivity for any distribution of transmission zeros, it is also straightforward to use software to determine the frequencies of stopband transmission zeros to meet arbitrary selectivity requirements. Such a loss “pole-placer” program has been described [Daniels:122-5]. Bandpass geometric symmetry is not required, so there can be just one or more such traps. Sections 4.3 and 4.5 deal with the last bullet in Table 4.2.1: How can the topologies and element values be obtained for networks specified by their passband flat-loss and ripple parameters and the transmission zero locations?

4.3 Challenges of Polynomial Synthesis

It has been observed that, although first-class synthesis programs are available commercially, they are beyond the reach of many potential users [Szentirmai], [Orchard]. The next chapter, on broadband impedance matching, describes two competing methods: a complicated synthesis procedure and a straightforward numerical procedure. Sections 4.3-4.5 are included to convince the reader that synthesis procedures are overly complicated and much more ill-conditioned than certain alternative numerical procedures for many practical but otherwise challenging network design tasks. A synthesis expert has noted that

“... the modern (insertion-loss) method of filter synthesis and design involves a very large amount of numerical computations, as well as, in most cases, the need to make choices that are anything but clear or simple. Furthermore, the numerical computations are nearly always very illconditioned, necessitating the use of either a large number of decimal places or esoteric procedures to overcome.” [Sentirmai].

4.3.1 Underlying Concepts

A recent book by an expert on wideband circuit design [Carlin,1998] has nine chapters, the first four of which are devoted to general properties, responses, energy relations, and various matrix representations of linear time-invariant circuits. In addition, there are more than 40 pages of appendices that outline the properties of analytic functions and the essentials of linear algebra. Apparently, that background is sufficient to synthesize wideband circuits.

4.3.2 Mathematical Operations and Sensitivities

The design of linear time-invariant circuits requires creating and evaluating real, rational polynomial functions. The computational tools

used in network synthesis span the realm of numerical analysis. Polynomial root finders are required, and roots are often allocated to left and right-half planes (*spectral factorization*) and then multiplied together into new polynomials. Many subsequent mathematical operations are trivial additions and subtractions of even and/or odd parts, where the challenge is programmed bookkeeping. Other procedures require solutions of sets of linear equations, algorithms for continued or partial fraction expansions, and synthetic division.

Non-trivial synthesis problems cannot be solved by straightforward manipulation of polynomials, as just suggested. The zeros of the filter polynomials are grouped close together near the edges of the passband, so that exceedingly small changes in the polynomial coefficients radically affect the locations of the polynomial zeros. This very large sensitivity of the roots with respect to (wrt) the coefficients is in stark contrast to the very small sensitivity of the element values to those same roots [Szentirmai]. Overall, the sensitivity of the element values wrt the coefficients of the polynomials in the matrix describing a network easily may be as high as 10^6 in very modest problems [Orchard]. It has been estimated that the necessary number of digits required for ordinary synthesis methods is roughly equal to the degree of the filter, N [Temes:113].

Historically, there have been three different approaches to avoiding the ill-conditioning. First, the *product method* represents the polynomials by their roots and a scaling multiplier at all times, so that the polynomials are never formed [Skwirzynski]. Second, a bilinear frequency mapping that spreads the locations of the zeros can be employed, thus accomplishing the synthesis in a less sensitive domain. Coincidentally, that is also the domain for observing the comprehensive filter properties described in Section 4.2 [Daniels:108]. Third, the filter element values can be adjusted to match the poles, zeros, and ripple factor of the characteristic function; a technique that trades computing time for high accuracy. This third method, *iterated analysis*, is described in Section 4.5. It also depends on the effect each element has on certain responses, as described in Section 4.4. In fact, it is the complexity of synthesis and the simple efficiency of iterated analysis that leads to the GRABIM technique described in Section 5.4.

4.3.3 The Approximation Problem

Consider equal-ripple passband filters with optional transmission zeros (traps) at finite frequencies in the stopbands. The equal-ripple response is one approximation to a constant, usually met with the Chebyshev polynomials of the first kind. Another common approximation is the maximally-flat criterion, which has the maximum number of error derivatives equal to zero at the reference frequency. Also, there are the least-squares method to minimize mean errors, and the interpolation method to make errors vanish at a number of discrete passband

frequencies. Section 5.3.1 explains that when filter loads are not both resistors, i.e., the broadband match situation, doubly-terminated filter transfer functions are not the optimal response functions.

4.3.4 Realization of Element Values

This is the most complex step, the synthesis of one of many possible equivalent network topologies. It is where the sensitivity problem plagues the result. When there are no traps, an LC ladder network is the result of applying a continued-fraction expansion of a rational polynomial, i.e. Cauer realization of a lossless one-port network. Similarly, applying a partial-fraction expansion of a rational polynomial results in a Foster realization, a less practical topology [Carlin,1998:217-223].

Lossless network realization employing continued- and partial-fraction expansions do not guarantee positive element values. However, there is a parametric representation of Brune functions that is similar to and slightly more complicated than the classical Foster functions [Fettweis], [Forster]. For example, varying the polynomial roots of filter functions to obtain some constrained selectivity objective might meet the objective but result in a non-physical network. The parametric Brune functions are guaranteed to result in a physical network, even before the network in its equivalent topologies is realized. That substantial advantage is seldom fully utilized.

Cascade synthesis is appropriate for realizing transmission zeros on the $j\omega$ axis, in the RHP, and those zeros (perhaps multiple) on the real (σ) axis of the Laplace s plane [Carlin,1998:252]. The latter are not academic, because they can have a beneficial effect on group delay in practical direct-coupled filters [Levy,1995].

4.3.5 Road Map for Topologies

Whatever network topology results from synthesis, there are usually many equivalent topologies to realize the same selectivity function. There are several small equivalency transformations available, as discussed in Section 2.4.4. The question of how many equivalent networks exist has been answered in terms of the locations of transmission zeros and terminal impedance behavior.

When no transmission zeros are at finite frequency, i.e. no traps, it is not difficult to tabulate the number of *canonic* equivalent networks – those having the minimum possible number of elements. This information appears in Table 4.3.1 [Kim]. Refer to (4.2.2); the data in Table 4.3.1 apply for $NT=0$, so that $N=NZ+NIN$. Lowpass networks, $NZ=0$ in Table 4.3.1, are unique (first column) as are highpass networks, $NIN=0$ (diagonal). Entries in each row are symmetric wrt the midpoint. Bandpass networks exist when $0 < NZ < N$, and they are most numerous at the midpoint of each row. The doubled squares contain the number of

Table 4.3.1. The Number of Equivalent Canonic Ladder Networks

NZ	0	1	2	3	4	5	6	7	8	9	10
1	1	1									
2	1	1	1								
3	1	2	2	1							
4	1	2	3	2	1						
5	1	3	5	5	3	1					
6	1	3	6	7	6	3	1				
7	1	4	9	13	13	9	4	1			
8	1	4	10	16	19	16	10	4	1		
9	1	5	14	26	35	35	26	14	5	1	
10	1	5	15	30	45	51	45	30	15	5	1

equivalent canonic ladder networks realized by the standard reactance transformation from lowpass to bandpass (2.4.3). Comparison of Tables 3.4.2 and 4.3.1 confirms that the consideration of only direct-coupled shunt resonator topology is indeed limiting.

There is a general classification of reactive ladder filters, including traps for real transmission zeros (on the real axis of the s plane) [Skwirzynski]. There are 24 classes of lowpass, highpass, and bandpass networks classified by the impedance behavior at the terminals at dc and infinite frequencies. Each class has dozens of permutations which are easily generated by programming a computer [Mellor].

Whatever topology is obtained and no matter what transformations are applied, an unacceptably wide range of element values, including negative values, is always possible. For example, the sequence of the set of null frequencies assigned to the three LC traps in Figure 4.2.2 has a drastic effect on element values. Negative element values may be avoided, whenever possible, by assigning the null frequencies nearest the passband edge to the trap branch situated in the middle part of the network [Saall].

4.4 Element Responses at Discrete Frequencies

The transducer and characteristic functions, H and K in (4.2.1), are two responses of a network. Because these functions can increase without limit, it is often more convenient to observe the reflection

coefficient in (2.1.2), which is bounded between zero and unity. The characteristic function is related to the reflection coefficient, ρ , by

$$|K|^2 = \frac{|\rho|^2}{1 - |\rho|^2}. \quad (4.4.1)$$

Therefore, $K=0$ when $|\rho|=0$ and $K \rightarrow \infty$ when $|\rho| \rightarrow 1$, with 1:1 correspondence in between. What is meant by "element response" is how reflectance $|\rho(\omega_i)|$ varies with any ONE branch value, e.g. versus any one of the L or C elements in Figure 4.2.2. Each element is varied from 0.1 to 10, a range that is reasonably inclusive in light of branch $1+Q^2$ impedance levels within the passband. This is rather like a wide-scale partial derivative in that all the other element values are fixed, i.e., a cross section. It clearly is useful to observe a set of reflectances corresponding to a set of passband frequency samples as the one element value is varied.

4.4.1 Filters

Doubly-terminated filters are LC two-port networks that have resistive terminations at both ports. Both lowpass and bandpass networks have similar element responses.

Example 4.4.1. Analyze the element responses for an $N=5$ lowpass Chebyshev prototype filter having 1 dB passband ripple. *Problem:* Design the filter, find the frequencies of the passband peaks and valleys, and analyze the input reflection magnitude versus each of the five branch element values. *Solution:* Use program ALLCHEBY.EXE to find the element values; this filter has a symmetrical topology. Use program RIPFREQS.EXE to find the passband peak and valley frequencies, and use an analysis program to calculate $|\rho|$ while varying each of the elements in turn.

The graph and notation in Figure 4.4.1 show the result for branch 1, capacitor C_1 that is in parallel with a 1-ohm load resistor (not shown). The abscissa is the value of C_1 over two decades centered on unity, and the ordinate is the reflection magnitude wrt 1 ohm, $|\rho|$, looking in at C_5 . The nominal (design) value is $C_1=2.1349$, and the six curves correspond to the numbered peak and valley frequencies listed. This filter has no flat loss, i.e. $R_1=R_2=1$, and $\delta=0$ in (4.2.1). Therefore, $|K|=0=|\rho|$ at dc (coincides with the abscissa). The other two valley frequencies shown have a clearly-defined reflection zero at $C_1=2.1349$. It is interesting to note that at $C_1=2.1349$ the three curves for the peak frequencies all intersect at $|\rho|=0.4535$, corresponding to 1.0 dB ripple.

The graph and notation in Figure 4.4.2 show the result for branch 2, the inductor L_2 . The nominal (design) value is $L_2=1.0911$. Again there are two distinct valley-frequency functions reaching $|K|=|\rho|=0$, and the

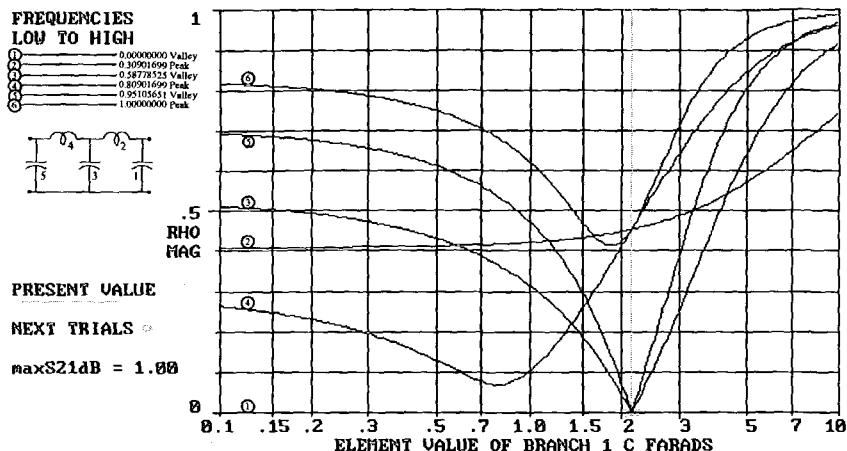


Figure 4.4.1. Reflectance versus branch 1 C for Example 4.4.1.

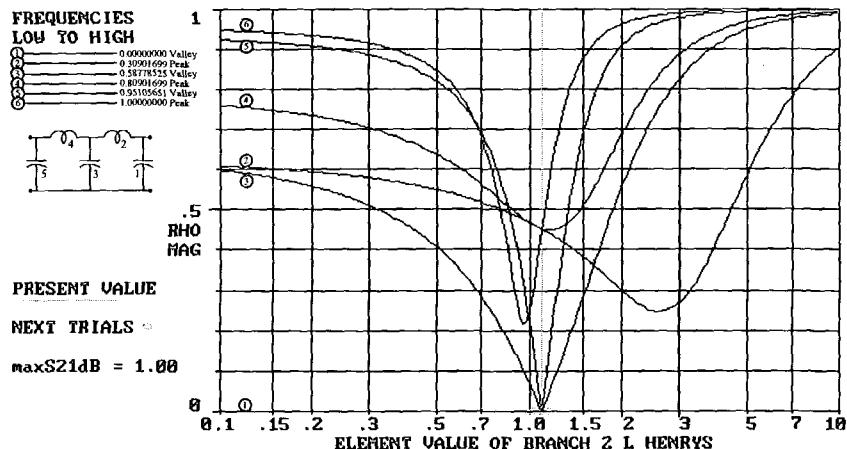


Figure 4.4.2. Reflectance versus branch 2 L for Example 4.4.1.

three peak-frequency functions intersect at $|\rho|=0.4535$. The *envelope function* is defined to be the arc segments of the frequency curves that constitute the worst-case $|\rho(\omega_i)|$ versus L_2 , $i = 1, \dots, 6$. Note that it is arc-wise continuous with discontinuous derivatives at the knots (arc joints) and has a minimum at $L_2=1.0911$. The same can be said for Figure 4.4.1, except that the minimum of the envelope function does not occur at the prototype value of $C_1=2.1349$. The zeros of characteristic function K (and ρ) are the basis for the iterated analysis method in Section 4.5, and the envelope function is the basis of the grid approach to broadband impedance matching (GRABIM) in Section 5.4.

It is instructive to consider the element sensitivities in light of Figures 4.4.1 and 4.4.2. Conventional *Bode sensitivities* are normalized

derivatives of a response wrt each element value [Cuthbert, 1983:101]. The derivative of the reflectance for each frequency can be seen as the slope of each curve in Figure 4.4.1. The reflectances for the two valley frequencies are equal to zero at $C_1=2.1349$ farads, where the slopes are large and depend on location for sign. The Bode sensitivities (slopes) at the peak frequencies are more benign, and they vary among the peak frequencies. However, it is clear that the Bode interpretation provides sensitivity information for each frequency only for small perturbations of the branch element value. Bode sensitivity is not applicable when the design goal is to minimize the worst-case reflectance over frequency, i.e. the envelope. The classical filter design in Example 4.4.1 is not optimal in the envelope sense; branch 1 is not optimal, while branch 2 appears to be optimal, in Figures 4.4.1 and 4.4.2, respectively. The design goal is met only when all branches are optimal.

4.4.2 Single-Match Broadband Networks

Figure 4.4.3 reproduces Figure 2.4.1 and adds the notation for both the load Q , Q_L , and the source Q , Q_s . It is noted in Section 4.2.1 that

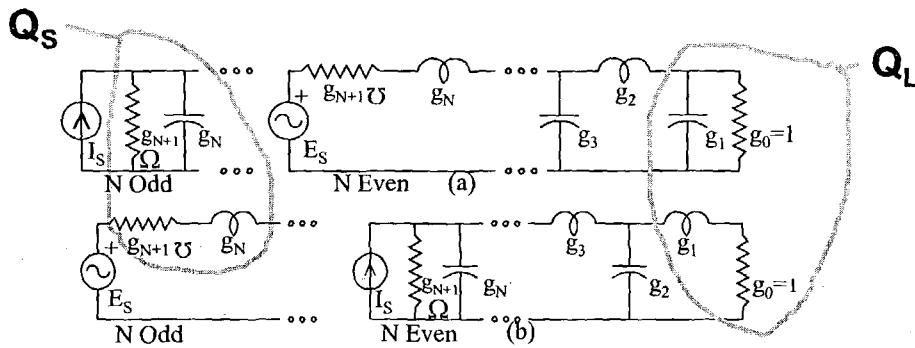


Figure 4.4.3. All-pole lowpass prototype networks with Q_L and Q_s .

equal-ripple filters have $N+1$ degrees of freedom: the N element values and the ripple factor, ϵ . A single-match network has one termination that includes reactance(s) along with the resistance. The classical broadband matching case described in Section 5.2 has the load composed of g_0 and g_1 , as indicated in Figure 4.4.3, so that $Q_L=g_1g_0$ is a constraint that removes one degree of freedom. (The first element of the matching network is g_2 , because g_1 is a part of the load.)

Over a given passband width, it is possible to minimize the maximum dB ripple IF the zeros of $K(\omega)$ are abandoned, i.e. if flat loss is accepted as shown in Figures 2.2.1 and 2.2.2. Typical single-match prototype design data are shown in Table 4.4.1. Note that there is 0.6772 dB flat loss and the maximum passband loss peaks at 0.8418 dB, the minimum possible peak dB given the $Q_L=2$ constraint and the Chebyshev

response. The insertion loss interval from 0.6772 to 0.8418 dB corresponds to $0.3800 \leq |\rho| \leq 0.4198$.

Table 4.4.1. ALLCHEBY Single-Match N=4, QL=2, 150% BW.

Is this a Matching or a Filter (QL=0=QS) network (_F)? M
 N,QLoad,%BW =? 4,2,150
 Is Source Loaded Q Optimal, Given, or Infinite (_G,I)? O
 MAXIMUM INSERTION LOSS FOR INFINITE N IS 0.5707 dB
 THIS INSERTION LOSS FROM 0.6772 TO 0.8418 dB
 THIS dB FLAT LOSS = 0.6772
 THIS dB RIPPLE = 0.1646
 SWR FROM 2.2256 TO 2.4469
 RETURN LOSS FROM 8.4051 TO 7.5399 dB
 Lser or Cpar G(1) = 3.0000 Q(1) = 2.0000
 Cpar or Lser G(2) = 0.6781 Q(2) = 0.4520
 Lser or Cpar G(3) = 3.3359 Q(3) = 2.2240
 Cpar or Lser G(4) = 0.3203 Q(4) = 0.2135
 Ohms or Mhos G(5) = 2.4469

Example 4.4.2. Analyze the element responses for an N=4 bandpass Chebyshev single-match broadband network based on the prototype filter data in Table 4.4.1. *Problem:* Design the filter with the wideband direct-coupled topology in Figure 3.4.9, find the frequencies of the passband peaks and valleys, and analyze the input reflection magnitude versus each of the five branch element values. *Solution:* Design the network according the spreadsheet in Figure 3.5.3, use program RIPFREQS.EXE to find the peak and valley frequencies, and then analyze the resulting network.

The graph in Figure 4.4.4 shows the result for the valley frequencies versus the parallel inductance, L_1 , in the second branch next to the specified R_0C_1 load. The nominal value is $L_1=4.2121$ henrys. The nature of broadband matching means that there are no zero values for either the characteristic function or the corresponding reflection function. The minimum reflection magnitude is 0.3800, corresponding to a transducer loss of 0.6772 dB.

The graph in Figure 4.4.5 shows the result for the peak frequencies, where the maximum reflection magnitude is 0.4198, corresponding to a transducer loss of 0.8418 dB. Program ALLCHEBY computes the element values so that the maximum reflection magnitude is the minimum possible (a *minimax*) for a specified $Q_L=2.0$. That produces the maximum-possible gain bandwidth when the characteristic function is Chebyshev.

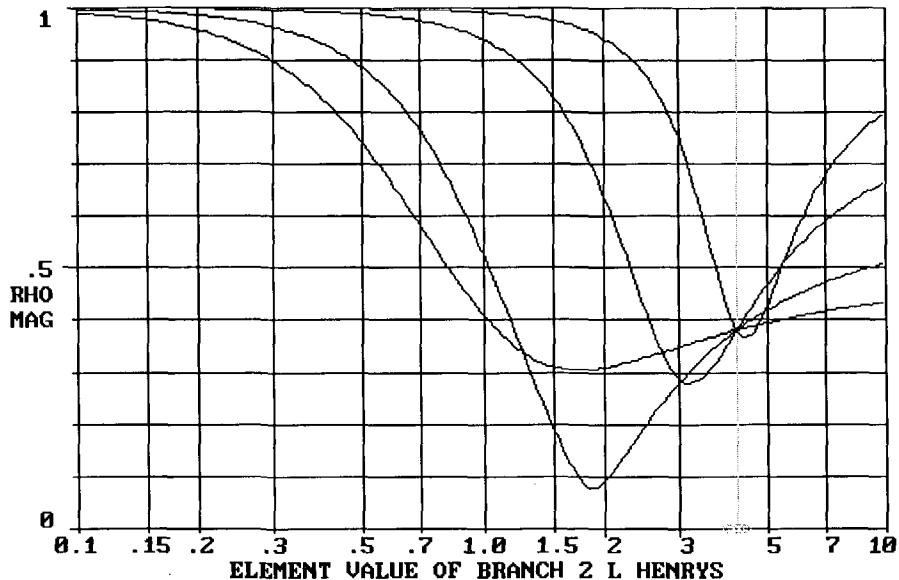


Figure 4.4.4. Valley-frequency reflectance versus L_1 for Example 4.4.2.

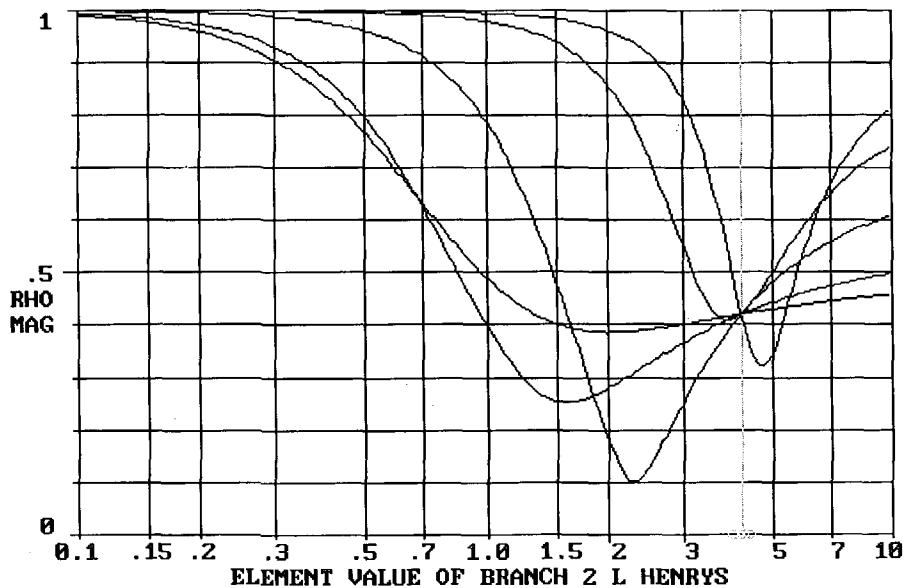


Figure 4.4.5. Peak-freauency reflectance versus L_1 for Example 4.4.2.

4.4.3 Double-Match Broadband Networks

As Figure 4.4.3 shows, the termination Q's could both be specified, so that in addition to $Q_L=g_1g_0$ there could be a specified $Q_S=g_Ng_{N+1}$. Now

there is no degree of freedom left to minimize the maximum ripple. In the single-match case, as in Table 4.4.1, increasingly larger values of degree N, the number of lowpass prototype elements, can force the dB ripple to zero and a flat transducer loss of 0.5707 as shown for $N \rightarrow \infty$. However, zero ripple and flat transducer loss over any finite passband cannot be obtained in double-matching problems no matter how great the degree N [Carlin,1998:413]. The reader is invited to try program ALLCHEBY with the entries as in Table 4.4.1 but specifying a given $Q_s=1$; for $N=15$ there is still a ripple of 0.7598 — it will not approach zero for $N \rightarrow \infty$.

4.4.4 Lessons Learned

Section 4.2 shows that knowledge of transmission zero locations for lowpass or bandpass filters having a Chebyshev equal-ripple passband completely determines the entire selectivity function of such filters. Especially, the passband frequencies of the valleys and peaks can be obtained easily and accurately.

Section 4.3 shows that getting the element values of such networks by synthesis techniques requires a sophisticated background in mathematics, numerical analysis, and filter theory. Even so, the resulting element values often are either negative or have unacceptably wide ranges. There are usually large numbers of equivalent canonical (minimum number of elements) topologies, and the application of the various transformations of topologies is tedious and confusing. "Bandpass filters are much more complicated than lowpass filters and the choice of best configuration often requires considerable skill and experience" [Orchard:1089]. It was concluded that either the designer could afford to have access to first-class synthesis programs or the designer needed to be a specialist in the synthesis field.

Section 4.4 shows that there are certain trends as to how individual elements affect filter response at critical frequencies. The next section reviews how the desired zeros of the filter characteristic function in (4.2.1) can be obtained simply and accurately by varying individual element values. The GRABIM technique in Section 5.4 shows how to obtain a starting point near the final values and how to work with arbitrarily-constrained matching problems, where the arc-wise continuous envelope function (Figures 4.4.1, 4.4.2, 4.4.4, and 4.4.5) must be minimized using all network elements, while discarding unneeded elements.

4.5 Synthesis by Iterated Analysis

Although synthesis of ladder filters in the Laplace s plane is badly conditioned numerically, analysis of such ladder networks at discrete frequency samples can be accomplished simply and very accurately. It is

also very little extra work to obtain the exact derivatives of the response at those frequencies wrt all the values of ladder elements. *Iterated analysis* [Orchard] is the automated adjustment of the element values (variables) to obtain characteristic function zeros at the desired frequencies. The number of degrees of freedom can be made equal to the number of constraints, so there is an exact solution, i.e. no least-squares or higher-order compromise is required as is often the case [Kintscher]. Iterated analysis produces element values that are as accurate as those obtained by explicit equations when available, e.g. those in Section 5.2.4.

4.5.1 Zeros and Poles of the Characteristic Function

Suppose that there is no flat loss in a doubly-terminated filter, i.e. $\delta=0$ in (4.2.1). Then the transducer loss is

$$\alpha = 10 \log_{10} \left\{ 1 + |K(\omega)|^2 \right\} dB. \quad (4.5.1)$$

The ripple factor, ϵ , is now contained as a scale factor within characteristic function K . Zero transmission loss corresponds to $K(\omega_i)=0$, and the ω_i are readily available from program RIPFREQS for Chebyshev equal-ripple passband filters, as illustrated in Table 4.2.3. There are $N\backslash 2$ (integral part) non-zero passband frequencies of zero transmission loss, where N is the degree of the network. The characteristic function $K(\omega)$ is rational, and the zeros of K are in its numerator. The poles of K are the transmission zeros ($\alpha \rightarrow \infty$), and they are zeros in the denominator of K at frequencies related to the trap branches in the ladder network.

By (4.2.2), the degree of the network is N , where $N=NZ+NIN+2NT$ for numbers of transmission zeros at zero, infinite, and trap frequencies, respectively. Chebyshev equal-ripple filters must be of even degree except in one lowpass case where $NZ=0$ and NIN is odd, so that N is odd; the latter are *symmetric filters*. Symmetric Chebyshev filters have equal resistive terminations, $R_1=R_2=1$, and can always be divided into two back-to-back identical networks.

4.5.2 Characteristic Zeros of Ladder Filters

The ABCD chain matrix of series and parallel branches is shown in (2.6.5), where it is noted that for lossless networks A and D are real and B and C are imaginary. From here on, all four ABCD parameters are considered real, so that jB and jC are used in place of the earlier B and C. For an ideal transformer with turns ratio t , $A=t$, $D=1/t$, and $B=0=C$. The overall ABCD chain matrix of an entire ladder network is just the product of all the component ABCD matrices, and it is easily computed at any particular frequency.

The overall ABCD chain matrix is related to the characteristic function by [Cuthbert, 1983:50]

$$2K(\omega) = (A - D) + j(B - C). \quad (4.5.2)$$

This means that at each non-zero passband frequency where $K=0$, there are two constraints: $(A-D)=0$ and $(B-C)=0$. So, for most filters where N is even, there are $2 \times (N/2) = N$ constraints. For symmetric lowpass filters where N is odd, there are $2 \times (N/2) = N-1$ constraints.

Notice that various sets of element values can make the constraint functions $(A-D)$ and $(B-C)$ positive or negative. These are constraint functions, and their zeros as functions of each element are portrayed in Figures 4.4.1 and 4.4.2, which plot the magnitude of the input reflection coefficient. However, the reflection magnitude is composed of the sum of the squares of the two constraints, and therefore has a distinct minimum. The slope (first derivative) of the curves in the figures is more representative of each of the two constraints. Any way they are viewed, the pair of constraint functions for $K=0$ in (4.5.2) are very well defined and conditioned for accurate optimization.

4.5.3 Balancing Variables and Constraints

Iterated analysis takes advantage of the fact that the number of constraints can be made equal to the number of variables. It has been stated that there are $N+1$ degrees of freedom in a Chebyshev equal-ripple ladder network; there are N of the L and C elements plus the ideal transformer's turns ratio to be varied. This statement assumes that the transmission pole (trap) frequencies are enforced; e.g., the parallel LC trap in a series branch in Figure 3.5.7 has g_3 as its variable parameter, with C_3 always adjusted so that branch is always anti-resonant at an assigned stopband null frequency.

For networks of even degree N , there are N constraints in (4.5.2). It is convenient to express the ripple factor in terms of the dB ripple, α_r :

$$\varepsilon = \sqrt{10^{\alpha_r/10} - 1}. \quad (4.5.3)$$

Then, for the lowpass, N -even ("antimetric") case, one of the $N+1$ degrees of freedom can be removed by assigning the transformer turns ratio the value to allow $R_1=R_2=1$:

$$t = \sqrt{1 + \varepsilon^2} + \varepsilon. \quad (4.5.4)$$

For all other cases of even degree N , which are bandpass networks, it is convenient to use the peak at the upper passband edge frequency as an added constraint [Orchard]. For symmetric bandpass networks

$$(B - C) - \Delta 2\varepsilon = 0, \quad (4.5.5)$$

and for antimetric bandpass networks

$$(A - D) - \Delta 2\varepsilon = 0, \quad (4.5.6)$$

where

$$\Delta \equiv \rho_1(\infty) \times (-1)^{(NU + NIN/2)}. \quad (4.5.7)$$

In (4.5.7), NU is the number of transmission zeros (trap nulls) in the upper stopband, and NIN is the number of transmission zeros at $\omega=\infty$.

Input reflection coefficient $\rho_1(\omega)$, Figure 4.2.1, at infinite frequency is equal to +1 (infinite input impedance) or -1 (zero input impedance), corresponding to two opposing points on the real axis of a Smith chart.

For the symmetric lowpass network case having odd degree N, there are $N-1$ constraints in (4.5.2) and the ideal transformer is not required, because $R_1=R_2=1$. The easiest way to balance the number of constraints and variables is to remove one variable by simply enforcing an arbitrary lower bound on one of the N elements and then observing the dB ripple at a peak frequency after optimization of the remaining elements (variables). The same method can be applied for the bandpass cases instead of (4.5.5)-(4.5.7). See Example 4.5.1 in Section 4.5.5. An alternate constraint for the symmetric lowpass network of odd degree N is to constrain the sum of the element values; see [Orchard:1092].

4.5.4 Efficient Network Analysis

The lossless ladder network elements in the iterated analysis method are inductors and capacitors and one ideal transformer. The ideal transformer is mentioned in Section 2.4.4, and its only parameter is its turns ratio, t , which causes an impedance transformation of t^2 that is independent of frequency. The overall ABCD chain matrix of an entire ladder network is just the product of all the component ABCD matrices; at any particular frequency, the component matrices can be processed one at a time using only four real operations: add, subtract, multiply and divide. The assimilation of the individual chain matrices to compute the overall ABCD matrix is computationally very fast, because 200 MHz personal computers (PC's) perform any of the four operations with high accuracy in about a microsecond. The details of assimilation of consecutive ABCD sections are presented in Section 6.2.3.

Theoretically, the exact derivatives of any standard transfer parameter of a lossless network with respect to (wrt) all of the elements can be obtained by just one analysis of the network per frequency. In practice, especially for derivatives of A, B, C, and D, it is computationally more efficient to make two analyses per frequency.

The equations and algorithm for computing the exact derivatives are provided in Section 6.2. It is sufficient here to indicate that initially all the unitized ($L=1$, etc.) branch reactances (or susceptances) for the L's, C's, and LC trap branches are computed and stored. The first analysis consists of starting with the 2×2 ABCD matrix of the ideal transformer at port 1 and assimilating the following LC branch elements in sequence, storing two of the four ABCD parameters accumulated up to each component. The second analysis at the same frequency starts at port 2 and works toward port one, evaluating the equations for the exact derivatives of A, B, C, and D wrt each component parameter.

4.5.5 Efficient Optimization

Although there are as many equations as unknowns, the equations are nonlinear, so that an iterative search must be made from a starting set of variables (L and C values and t) in the column vector \mathbf{x}^0 to find the set of variables, \mathbf{x}^* , that satisfies all the constraints. The constraints to be made equal to zero by varying \mathbf{x} are found in (4.5.2), (4.5.5), and (4.5.6). The constraints can be collected in a vector of constraint functions, $\mathbf{c}(\mathbf{x})$; e.g., $c_1(\mathbf{x})=A-D$ and $c_2(\mathbf{x})=B-C$ at $\omega_1, \omega_2, \dots$. Then the partial derivatives of each constraint function, c_i , wrt each variable x_j , are assembled in a *Jacobian* matrix, J :

$$J \equiv \left[\frac{\partial c_i}{\partial x_j} \right], \quad (4.5.8)$$

which designates the i th row and the j th column. Vector \mathbf{c} and matrix J can be formed two rows at a time corresponding to (4.5.2) and the two analyses at each frequency [Orchard:1093]. The Jacobian matrix is square in this situation.

The Newton iteration with the Jacobian matrix [Cuthbert, 1983:125] produces a sequence of corrections, $\Delta\mathbf{x}$, to the current vector of variables:

$$\Delta\mathbf{x} = -J^{-1}\mathbf{c}. \quad (4.5.9)$$

In practice, the Jacobian matrix is NOT inverted; it remains on the left-hand side of (4.5.9), and $\Delta\mathbf{x}$ is found by LU factorization [Cuthbert, 1987:98-9].

The starting vector, \mathbf{x}^0 , is not critical in lowpass filters. Because normalization with both frequency and impedance is employed, starting element values and a turns ratio somewhat larger than unity are appropriate; see the equal-element filter values in Figure 2.4.4. That start may not work reliably for bandpass filters, but methods to obtain good starting values for the Newton iterations and to avoid negative elements are described in detail in the next two chapters. Briefly, the starting vectors can be processed first with a reasonably efficient grid search to obtain an \mathbf{x}^0 suitable for the sequence of Newton steps defined by (4.5.9). The grid search and the Newton optimization can be accomplished in log space, which amounts to varying the element value in dB; in that way, no element value can become negative.

The element values obtained from this iterative process can be expected to be as accurate as those from explicit equations when available, e.g., those in Section 5.2.4.

Example 4.5.1. An equal-ripple filter passband extends from 0.7071 to 1.4142 rad/s with the topology shown in Figure 4.5.1. *Problem:* Design this filter using the iterated element method. *Solution:* Using NZ=5, NIN=1, and NT=0 in program RIPFREQS shows that the three valleys where K=0 are at 0.72905113, 0.92553926, and 1.31729198 rad/s. The

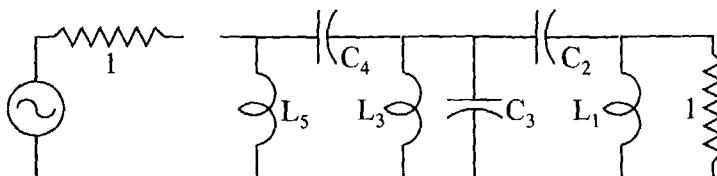


Figure 4.5.1. A bandpass network with NZ=5, NIN=1, and NT=0.

degree N=6, so there are six constraints according to (4.5.2). Arbitrarily adding the constraint that $C_3 = 2$ farads makes the number of equality constraints equal to the degrees of freedom, N+1=7.

The element values are obtained by a sequence of Newton steps, (4.5.9), until the magnitudes of all the constraints are less than 0.000005. The final element values are shown in Table 4.5.1. By analysis at the four peak frequencies from program RIPFREQS, the uniform ripple is found to be 0.268 dB. Decreasing the value assigned to C_3 would decrease the dB ripple.

Table 4.5.1. Element Values for Example 4.5.1 Obtained by Iterated Analysis.

In henries and farads:

Element	Value
L_1	0.91425
C_2	1.59397
C_3	2.00000
L_3	0.29803
C_4	1.59397
L_5	0.91425

4.6 Summary of Comprehensive Equal-Ripple Filters

This chapter is a bridge between the highly structured subject of direct-coupled filter design and the vast space of analytic functions that is encountered in broadband impedance matching. A useful concept is that of Chebyshev equal-ripple doubly-terminated filters, which depend only on the locations of transmission zeros on the $j\omega$ axis. Lossless LC network topologies have specific numbers of transmission zeros at zero, infinite, and finite frequencies. Once those numbers are known and the passband flat loss and ripple are given, then the selectivity of any such filter is uniquely determined. The passband frequencies where the peaks and valleys occur can be found with great accuracy by program RIPFREQS.

Naturally, it would be nice if there were an easy way to calculate the values of the elements in the selected network topology. Other than using first-class synthesis programs that are beyond the reach of many designers, practical application of network synthesis requires considerable skill and experience, especially for bandpass filters. A brief review of the scope of the underlying concepts, mathematical complexities, and numerical ill conditioning is provided to emphasize the difficulty of that process.

However, a very consistent picture emerges from observing the benign behavior of the reflection magnitude response (reflectance), as a function of one network element at a time, at critical passband frequencies, namely at the passband valley and peak frequencies. For filters without flat loss, that cross-sectional behavior shows that each valley has a single well-behaved zero of reflection versus each individual branch value. That characteristic leads to the method of iterated analysis. For filters with flat loss, necessarily those employed for broadband matching (i.e., maximum gain-bandwidth), the reflectance curves versus the branch variables at valley frequencies all intersect at the minimum reflection magnitude. In all cases, the curves versus each branch variables at peak frequencies all intersect at the maximum reflection magnitude. This consistent benign behavior is explained in Section 5.4.3.

The iterated analysis method of synthesis is outlined in order to provide a practical and less complicated filter synthesis technique, and to set the stage for the grid approach to broadband impedance matching (GRABIM) in Chapter 5. For no flat loss, the loss zeros and poles of the characteristic function are related to the response continuum phenomena and ladder network topologies. The numbers of variables and constraints are described in order that they be obtained in equal numbers, thus avoiding the overdetermined systems of equations that are typical of most optimization, e.g., nonlinear least squares. Pairs of constraint functions are simple combinations of the network's overall ABCD parameters obtained efficiently and numerically at each valley frequency. The exact partial derivatives of those constraints are also obtained at negligible cost in the algorithm described.

Finally, an iterated Newton search beginning at an arbitrary starting set of element values is described in terms of a square Jacobian matrix of partial derivatives. The methods suggested for finding reliable starting values and avoiding negative element values are mentioned because they are employed in Chapter Five and described in detail in Chapter Six. It is hoped that this chapter will assist the reader by providing insight and enhancing recognition in the next chapter's development of broadband impedance matching.

5. Matching Networks

This chapter shows how to design lossless two-port doubly-terminated networks that have specified reactances incorporated in one or both terminations – the single- or double-match cases, respectively. The objective is maximum power transfer over wide frequency bands. Generalized reflection coefficients play an essential role, because the solution is to design a network that will minimize the reflectance of the specified load over the passband. Because a constant reflectance magnitude describes a circular neighborhood of impedances centered on the normalizing impedance, broadband matching also concerns impedance matching over a pass band.

The first case considered is matching different terminating impedances at a single frequency. This is the case of matching an infinitesimal output neighborhood to its input image normalized to a source impedance, i.e., conjugate matching. The main tool is the loaded Q parameter, Q_L .

Analytic gain-bandwidth theory is described next. It maximizes power transfer over the passband from a resistive source to a simple load impedance, which may consist of an LR or a CR lowpass network termination, or the corresponding RLC bandpass network termination having a given Q_L . Although this concise theory has limited application to most practical problems, it provides insight into the fundamental limits of the gain-bandwidth tradeoff in terms of Q_{BW}/Q_L . At most, two lowpass load reactances, or the equivalent two bandpass resonator load terminations, can be processed using analytic theory as a design tool. The double-match problem also can be processed, but the result is even less likely to solve practical problems.

The escape from impractical design limitations came with introduction of the real-frequency technique (RFT) [Carlin, 1977]. Most terminating impedances are known by their tabulation as measured at a set of discrete passband frequencies. That approach is contrasted with trying to model a terminating impedance as a simple RLC one-port network. Reflectance at each tabulated frequency is considered at both ends of the lossless two-port matching network, which is synthesized in the last step. Unfortunately, intermediate steps require several kinds of numerical optimization as well as polynomial synthesis. The necessary skill and experience is often a serious obstacle for the designer. The real-frequency technique is reviewed because it introduces several important concepts, has displaced analytic theory, and is mentioned continually in the literature [Cuthbert, 1994a].

Finally, the grid approach to broadband impedance matching (GRABIM) is described because it requires minimal skill and produces practical results immediately. This technique is superior to the RFT in almost every case. It begins by designating a stored or a user-specified

LC and transmission-line ladder network topology of chosen complexity. A robust grid search over the logarithmic space of likely element values minimizes the maximum reflectance over all passband frequencies to find a small neighborhood of the likely global minimum. Then minimax-constrained optimization locates the precise minimum while maintaining positive elements and removing those that are unnecessary.

5.1 Single-Frequency Matching

This topic emphasizes the central role of the loaded-Q parameter in LC network design and the $1+Q^2$ method as the basis of a grid search technique for broadband impedance matching [Abrie, 1985, 1991]. The two-port cascade transmission line is also included, because it is an allowable element in the GRABIM method in Section 5.4.

The two-port matching networks are assumed lossless, and the conjugate impedance match (zero complex reflectance) required at the input terminals exists at every interface throughout the network at that frequency. Design of two-element "el" sections and their role in constructing three-element T and Pi networks are described, to introduce the role of transformation Q's for ladder networks of any length.

The generalized reflectance coefficients in (2.1.4) are bilinear functions of each network L or C branch and the load impedance, as defined in Appendix Section A.1. For the cascade transmission line (CASTL), its load impedance and the tangent of its electrical length are bilinearly related, but its characteristic impedance, Z_0 , is not bilinearly related to generalized reflection functions. However, the pertinent effect of Z_0 on reflectance is essentially the same as LC elements and can be explained easily. The same can be said for the ideal transformer as well.

5.1.1 Zero Reflectance

Recall the generalized reflection coefficients in (2.1.8):

$$|\alpha| = \left| \frac{Z_{in} - Z_s^*}{Z_{in} + Z_s} \right| = \left| \frac{Z_f - Z_T^*}{Z_f + Z_T} \right| = \left| \frac{Z_L - Z_b^*}{Z_L + Z_b} \right|. \quad (5.1.1)$$

Reflectance $|\alpha|$ is invariant at the input, interior, and load interfaces of lossless linear networks, including those containing CASTL's. In (5.1.1), Z_s is the source impedance, Z_{in} is the input impedance, Z_L is the load impedance; and both Z_T and Z_b are Thevenin equivalent impedances looking into the network toward the source at interior or load interfaces, respectively. See Figure 2.1.6. If the real parts of the impedances in (5.1.1) are strictly positive, then $|\alpha|=0$ only when there is a conjugate match at every interface, e.g. $Z_{in}=Z_s^*$. Therefore, zero reflectance implies a conjugate match at every interface in the network.

5.1.2 El Sections Matching Resistances

The objective is to design a two-element LC network that matches a load resistance, R_2 , to a source resistance, R_1 . The network topology is called an "el" section, and its design is a very simple but important application of the $1+Q^2$ series-parallel conversion method of Section 2.3.1.

Example 5.1.1. A $6+j0$ ohm load impedance must be matched to $25+j0$ ohms at a single frequency. *Problem:* Design the two el sections that produce this match. *Solution:* The steps and results are shown in Figure 5.1.1 (a) and (b), respectively. Equations (2.3.2) and (2.3.5) are repeated for convenience:

$$Q = \frac{X_s}{R_s} = \frac{R_p}{X_p}, \quad (5.1.2)$$

$$Q = \sqrt{\frac{R_p}{R_s} - 1}. \quad (5.1.3)$$

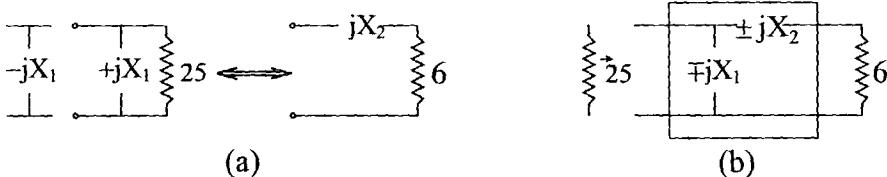


Figure 5.1.1. El section matching: resistive load and source in Example 5.1.1.

For a real Q value, (5.1.3) shows that $R_p > R_s$, so $R_p=25$ and $R_s=6$, yielding $Q=1.7795$. Series reactance X_2 in Figure 5.1.1(a) is found by using (5.1.2); $X_2=X_s=Q\times R_s$, so $X_2=10.6771$ ohms; the sign of the reactance is not important at this step. Again, (5.1.2) yields $X_1=X_p=R_p/Q$, so $X_1=14.0488$ ohms.

The series-parallel equivalence shown in Figure 5.1.1(a) is obtained with X_1 and X_2 having the same sign: they are either both L's or both C's. In order that the parallel-equivalent impedance be a real $25+j0$ ohms, it is necessary to add a $-X_1$ reactance in parallel to tune out (cancel) the $+X_1$. The result, in Figure 5.1.1(b), shows the two possible solutions: In the physical network, if X_2 is an L then X_1 is a C, and vice versa. It is important to observe that

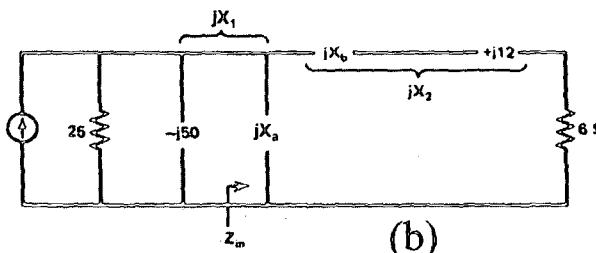
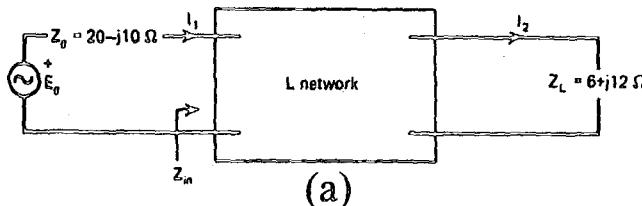
- $R_p > R_s$ mandates that el section topology be oriented as shown for matching a low resistance to a high resistance, and vice versa, and
- The branch reactances in an el section that match resistances must have opposite signs.

Also note that the reactances in ohms at a design frequency are readily converted into inductive and capacitive element values using program DENORM.

5.1.3 El Sections Matching Impedances

The more general application for el sections is to match impedances, one or both having a reactive component. An extension of Example 5.1.1 illustrates the method.

Example 5.1.2. A $Z_2 = 6+j12 \Omega$ load impedance must be matched to a $20-j10 \Omega$ source at a single frequency. *Problem:* Design all el sections that produce this match. *Solution:* Half the solutions to this problem are shown in Figure 5.1.2. Leaving the load in series form requires that the



X_2	± 10.68	$- 10.68$
X_1	$- 14.05$	± 14.05
X_b	$- 1.32$	$- 22.68$
X_a	$- 19.54$	± 10.97
X_{imp}	50	50
R_{imp}	25	25

Figure 5.1.2. El section matching complex load and source in Example 5.1.2.

source impedance be converted to parallel form; $1+Q^2$ can be used to obtain $R_P=25$ and $X_P=-50$ ohms as shown in Figure 5.1.2. The first step is to design an el network composed of X_1 and X_2 that matches 6 ohms to 25 ohms; that step is shown in Example 5.1.1. The physical matching network is X_a and X_b in Figure 5.1.2; if X_2 is inductive, as is the load, then $X_2=10.6771=X_b+12$, so $X_b=-1.3229$ ohms (capacitive). It remains to find what value of X_a in parallel with -50 ohms will produce reactance $X_1=-14.0488$ ohms, i.e.

$$\frac{1}{X_1} = \frac{1}{X_a} + \frac{1}{-50}. \quad (5.1.4)$$

The answer is $X_a=-19.5387$ ohms, so both X_a and X_b , which compose the matching network, turn out to be capacitive in this case.

A second solution is obtained when the signs of both X_1 and X_2 are reversed; the resulting answers are shown in Figure 5.1.2 also. Third and fourth solutions can be obtained by leaving the source in series form while converting the load to parallel form ($R_P=30$ and $X_P=+15$ ohms). These solutions have been previously published [Cuthbert, 1983:180].

5.1.4 Pi and T Sections

Pi and T sections composed of three reactances can be designed by creating two el sections in cascade. The easiest way to explain this is also by example. The concept of a hypothetical transformation resistance and its loaded Q is important.

Example 5.1.3. A $100+j0$ ohm load impedance must be matched to $50+j0$ ohms at a single frequency using a Pi section. *Problem:* Design all Pi sections that produce this match using el sections in cascade. *Solution:* The steps and results are shown in Figure 5.1.3 (a) through (h). The

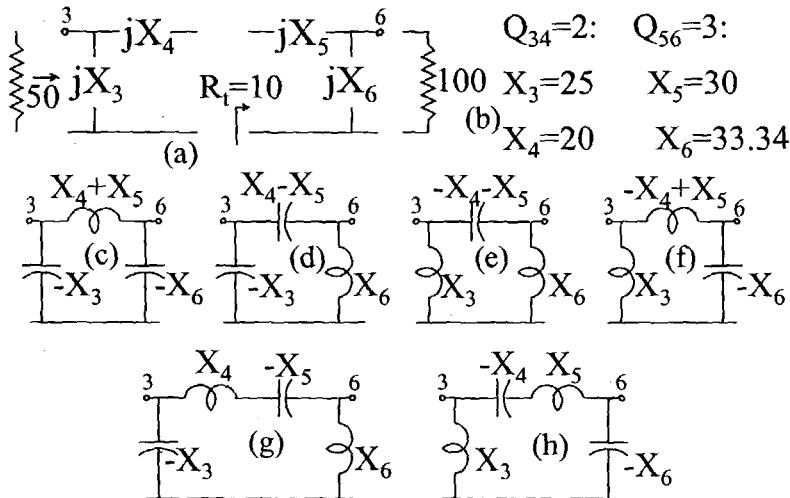


Figure 5.1.3. Networks matching 50 to 100 ohms in Example 5.1.3.

design begins at the 100-ohm load resistor by designing an el section that matches to a *transformation resistance*, R_T . Because the matching element next to the load is in parallel, the hypothetical resistance R_T is in series with reactance X_5 . Therefore, (5.1.3) shows that $R_T < 100$ is required. However, the el section composed of X_4 and X_3 is oriented so that $R_T < 50$ is also required. The latter constraint is dominant and determines the rms current in X_4 and X_5 . Given a power P in watts, the associated rms current is

$$I_T = \sqrt{\frac{P}{R_T}}. \quad (5.1.5)$$

Current I_T cannot be less than 0.1414 amperes/watt; a choice of $R_T=10$ causes 0.3162 amperes/watt. Parameter R_T also controls the transfer phase [Cuthbert, 1983:174]. Another consequence of the choice for R_T is the resulting element values as follows.

The resistance ratio 100/10 in (5.1.3) requires $Q_{56}=3$, so that $X_6=33\frac{1}{3}$ and $X_5=30$, ignoring signs. Similarly, $Q_{34}=2$, so that $X_4=20$ and $X_3=25$ ohms. As shown in the el section design in Example 5.1.1, each

pair of reactances must have opposite signs. If all the X_i values are defined positive, then the various combinations produce the Pi networks in Figure 5.1.3(c)-(h) that solve the problem. They are lowpass, highpass, and bandpass topologies of degree $N=2$ or $N=4$ with assorted NZ and NIN transmission zeros at dc and infinity, respectively. A spreadsheet and its built-in optimizer is an effective means for imposing user constraints.

Design of T networks is accomplished in dual fashion. In that case the transformation resistance, R_T , is a parallel resistance that controls the voltage at the middle node as well as the transfer phase and all element values. As a result of these examples, it should be observed that all the branch reactances in any ladder network can be quantified by a sequence of transformation resistances or their signed *transformation Q* values. Even though these methods relate only to a single frequency, transformation Q's are the basis of a wideband impedance matching technique [Abrie, 1985:79].

5.1.5 Cascade Transmission Lines

Figure 5.1.4 shows a CASTL terminated by load impedance $Z_2=R_2+jX_2$ and with input impedance $Z_1=R_1+jX_1$ at a single frequency. It

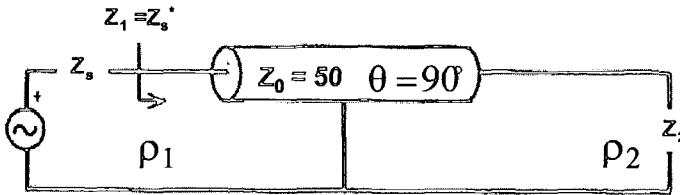


Figure 5.1.4. A terminated cascade transmission line.

interfaces with source impedance Z_S at port 1, and an impedance match implies the conjugate condition $Z_1=Z_S^*$. The input impedance is related to the load, real characteristic impedance, and electrical length by

$$Z_1 = Z_0 \frac{Z_2 + jyZ_0}{Z_0 + jyZ_2}, \quad y \equiv \tan \theta. \quad (5.1.6)$$

Relating the real and imaginary parts of (5.1.6) and some algebra enable Z_0 and θ to be expressed in terms of Z_1 and Z_2 :

$$Z_0 = \left(\frac{R_1 |Z_2|^2 - R_2 |Z_1|^2}{R_2 - R_1} \right)^{1/2} \quad (5.1.7)$$

and

$$\theta = \tan^{-1} \left(\frac{Z_0 (R_1 - R_2)}{R_1 X_2 + R_2 X_1} \right). \quad (5.1.8)$$

Clearly, a solution can exist only when $R_2 \neq R_1$ and the square root in (5.1.7) exists. Both of these equations are easily programmed

[Cuthbert, 1983:185], and many articles with various design aids based on these equations have been published.

For completeness, the reflection form for describing a CASTL should be noted:

$$\rho_1 = \rho_2 e^{-j2\theta}, \quad (5.1.9)$$

where ρ is the simple reflection coefficient wrt Z_0 ; see (2.1.10). One thing (5.1.9) shows is that the polar angle on a Smith chart is twice the electrical length of the CASTL.

The more important situation is when these CASTL's are embedded in an LC ladder network; then their ABCD parameters and equivalent Pi or T networks are of more interest, as described in Sections 5.4.3 and 5.4.4.

5.2 Analytic Gain-Bandwidth Theory

The first extensive analytical treatment of the gain-bandwidth problem [Fano] was followed by many other developments, especially its simplification [Levy, 1964], which is reviewed here. The situation is shown in Figure 5.2.1, which includes both a complex source and a

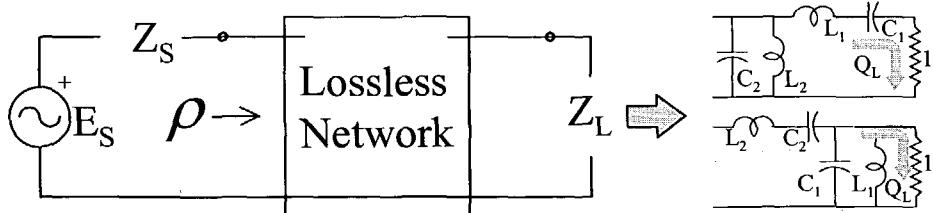


Figure 5.2.1. Fano broadband matching based on lowpass prototype data.

complex load. The practical bandpass cases only deal with a single dissipative load resonator and three kinds of source impedance: (1) a resistance, (2) a single dissipative resonator, or (3) an ideal voltage or current source having no resistance.

It is very important to have a clear understanding of broadband matching limitations in terms of simple Q parameters and network degree, N. Those limitations are described first, followed by the explicit equations for the element values which make comprehensive design programs such as ALLCHEBY possible. Examples are included, and references for modeling measured impedance versus frequency as a single resonator are discussed.

5.2.1 Gain-Bandwidth Limitation

The lossless matching network in Figure 5.2.1 is assumed to be derived from its lowpass prototype network, as described in Section 2.4.2. The load impedance, Z_L , usually is considered to be the first resonator ($L_1, C_1, 1$). Therefore, the matching network usually includes the second

resonator, composed of L_2-C_2 . Whether series or parallel, the load resonator has a loaded Q, Q_L , and the input reflection coefficient of the matching network is a function of Q_L . The reflection return loss is defined by (2.1.9) and (2.1.10), using ρ in place of Γ . These parameters are the ingredients in Fano's gain-bandwidth limitation:

$$\int_0^{\infty} \ln|\rho|^{-1} d\omega = \frac{\pi}{Q_L}. \quad (5.2.1)$$

The natural logarithm, \ln , differs from Log_{10} by just the constant 2.3026, so what (5.2.1) requires is that the area under the return-loss curve over all frequencies is a simple constant. (Fano originally stated the problem in terms of lowpass networks. In Figure 4.4.3, $Q_L=g_1 \times g_0$ when $\omega_0=1$, according to (2.3.2)).

By the theory of functions, continuous conjugate matching over a finite pass band is impossible [Carlin,1998:384]. Figure 5.2.2 shows the gain-bandwidth limitation: there can be low reflection and narrow

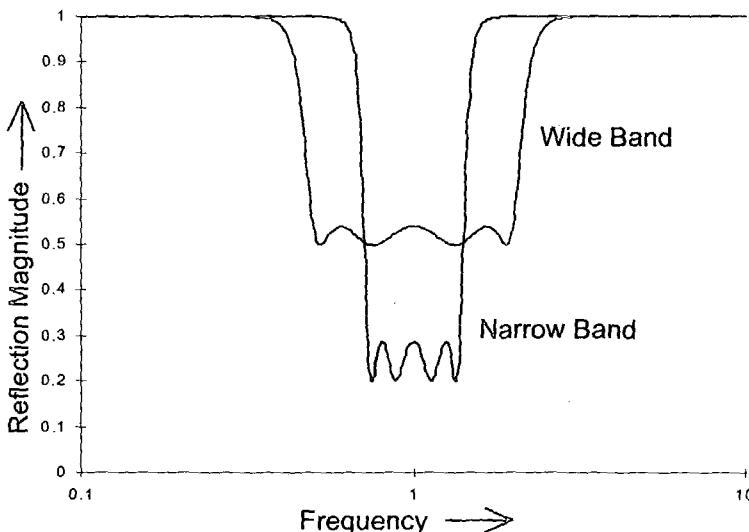


Figure 5.2.2. Fano's gain-bandwidth limitation for single matching.

bandwidth or wider bandwidth and greater reflection. The greater the loaded Q, the more drastic the compromise. When the ripples in Figure 5.2.2 result from the Chebyshev function, a network of higher degree, N, will have less ripple with lower reflection. It is useful to allow $N \rightarrow \infty$, so that there is constant reflection in the passband, e.g. $|\rho|_{\max}$ as illustrated in Figure 5.2.3. Then the integral in (5.2.1) becomes band limited so that the limits of integration can be from ω_1 to ω_2 . The result for the single-match case is

$$\text{Min} |\rho|_{\max} = e^{-\delta\pi}, \quad N \rightarrow \infty, \quad (5.2.2)$$

where *decrement*, δ , is defined as

$$\delta \equiv \frac{Q_{BW}}{Q_L}. \quad (5.2.3)$$

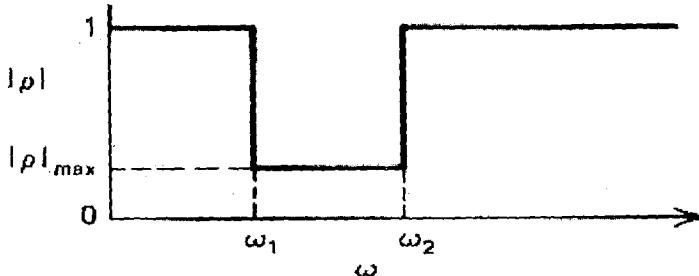


Figure 5.2.3. Limiting case of passband reflection as degree $N \rightarrow \infty$.

5.2.2 The Single-Match Minimization Problem

Of course N cannot be infinite, as in (5.2.2), so the question is how $|\rho|_{\max}$ varies with δ for a finite bandpass network of degree $2N$, where N is the number of resonators. Figure 5.2.4 shows the lowpass transducer function in (4.2.1), where the equal-ripple characteristic function

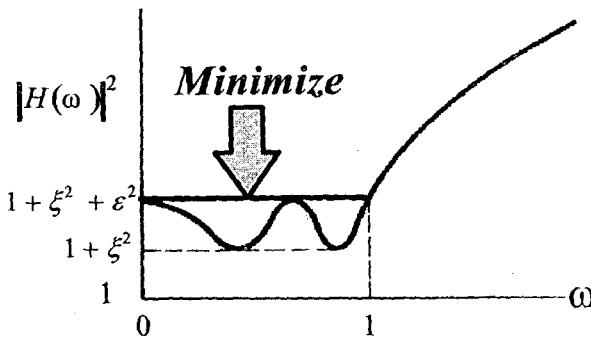


Figure 5.2.4. Lowpass equal-ripple (Chebyshev) transducer function.

$|K(\omega)|^2 = \varepsilon^2 T^2(\omega)$ and $T(\omega)$ is a Chebyshev polynomial of the first kind. The flat loss that is always associated with broadband matching is $1 + \xi^2$; see (4.2.1). The characteristic function is related to the input reflection coefficient by (4.4.1), so minimization of transducer function $|H|^2$ is equivalent to minimization of $|\rho|^2$ as well. See [Cuthbert, 1983:194] for details.

Fortunately, there are explicit expressions for the coordinates of the poles of $\rho(\omega)$ in the Laplace s plane for the Chebyshev case; they are on an ellipse having its major axis centered on the $j\omega$ axis. Any

particular ellipse has two defining variables, say a and b , that appear as arguments of hyperbolic sines ($\sinh z$). Location of the zeros of $\rho(\omega)$ is arbitrary, but they are usually placed in the LHP in order to maximize Fano's integral in (5.2.1) [Mellor:67], [Van der Walt]. The two degrees of freedom in a and b allow two constraints; for the single-match case they are (1) a given loaded Q_L , and (2) minimization of the maximum ripple as in Figure 5.2.4. The result for the single-match case (resistive source) is that constants a and b are determined in the process of minimizing $|\rho|_{\max}$. Flat-loss parameter ξ in (4.2.1) is also determined.

5.2.3 Single-Match Optimal Results

The bottom line on the single-match case is shown in Figure 5.2.5. The abscissa is the decrement $\delta = Q_{BW}/Q_L$. The ordinates are $|\rho|_{\max}$ on the right and the equivalent standing-wave ratio (SWR) on the left.

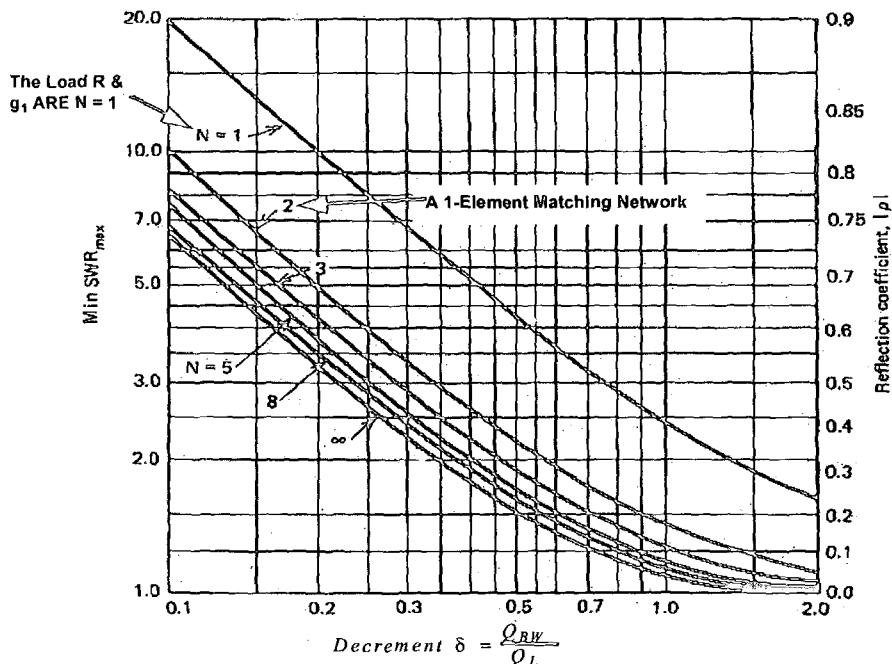


Figure 5.2.5. Fano's minimum possible reflection for the single match case.

Recall that $N=1$ is just the single RLC load resonator. So for a resistive source connected directly to that load without any matching network, $N=1$, the top line shows how $|\rho|_{\max}$ decreases as δ increases. The lowest line is the $N \rightarrow \infty$ case, which corresponds to (5.2.2). The significant trend is shown by the $N=2, 3, 4$, and 5 loci: for any given $\delta = Q_{BW}/Q_L$ there is a marked reduction in $|\rho|_{\max}$ for just one matching element ($N=2$), and a decreasing improvement as N increases. There is seldom any reason to

include more than four resonators ($N \leq 5$) in a single-match broadband matching network.

This important conclusion applies for bandpass situations, where N is the number of resonators (including the load). It also applies for the lowpass situations, where N is the number of elements including the load and $Q_{BW} \equiv 1$. In both cases, the source resistance is a dependent variable which is not unity because of flat loss. The value other than unity is easily accounted for in bandpass network design (Norton transformations in Section 2.4.4). It is useful to see how analytic theory has shaped the input reflection-coefficient frequency function; see Figure 2.2.7 for the $N=3$ lowpass response.

5.2.4 Chebyshev Network Element Values

In addition to knowing the pole locations for Chebyshev filters in terms of parameters a and b in Section 5.2.2, several investigators guessed or derived a recursive expression for the lowpass prototype element values [Green]. In the lowpass prototype, $Q_{BW} \equiv 1$ and load resistance $g_0 \equiv 1$, so the load decrement in (5.2.3) is simply $\delta_1 = 1/g_1$, where $\delta_1 = \delta$ by (5.2.3). Then,

$$g_1 = \frac{\sin\theta}{\sinh a - \sinh b}, \quad \theta \equiv \frac{\pi}{2N}, \quad (5.2.4)$$

$$g_{r+1} = \frac{4 \sin[(2r-1)\theta] \times \sin[(2r+1)\theta]/g_r}{\sinh^2 a + \sinh^2 b + \sin^2(2r\theta) - 2 \sinh a \times \sinh b \times \cos(2r\theta)}, \quad (5.2.5)$$

for $r=1, 2, \dots, N-1$, and

$$g_{N+1} = \frac{2 \sin\theta}{g_N (\sinh a + \sinh b)}. \quad (5.2.6)$$

These *recursion equations* define the elements for the prototype lowpass networks in Figure 4.4.3. In addition to degree N , constants a and b are always known, as in Sections 5.2.2 and 5.2.5, so that $g_1 = 1/\delta_1$ will satisfy (5.2.4). For $r=1$, g_1 appears in the numerator of (5.2.5) to determine g_2 , etc. The last value, g_{N+1} in (5.2.6), is the dependent source resistance or conductance. These equations are the basis for program ALLCHEBY.

For a single-match example, see Table 4.4.1 in Section 4.4.2. Element values for the classical single-match case have been plotted versus decrement δ_1 [Matthaei:126-9]. From that plot or program ALLCHEBY, it is readily observed that for $0.1 \leq \delta_1 \leq 2.0$ corresponding to Figure 5.2.5, the lowpass element values are contained within the range $0.9 \leq g_i \leq 12$ for $N \leq 5$. This fact is applied in GRABIM, Section 5.4.4.

5.2.5 Other Terminal Impedances

There are several other important cases of load and source impedances, and the network element values differ only by different

choices of constants a and b in (5.2.4)-(5.2.6). Doubly-terminated Chebyshev filters have resistive terminations, so a and b are determined by the specified ripple and flat loss. In singly-terminated Chebyshev filters (ideal source), $a=b$, and the one degree of freedom is determined by (1) a given Q_L , or (2) a given passband ripple.

For the double-match case, both Q_L and Q_S are specified (Figure 4.4.3), so there is no freedom to minimize $|\rho|_{\max}$ as in Section 5.2.2. Constants a and b are dependent, as are the ripple and flat loss, because the source decrement, δ_N , constrains (5.2.6):

$$\delta_N \equiv \frac{Q_{BW}}{Q_S} = \frac{1}{g_N g_{N+1}}. \quad (5.2.7)$$

Element equations (5.2.4)-(5.2.6) correspond to elements from load to source. However, it is required that $\delta_N > \delta_1$ [Cuthbert, 1983:201-2]. Therefore, when $Q_S > Q_L$, element numbering is reversed, with the source resistance equal to $g_0=1$ ohm and numbering increasing toward the load. The response is unaffected, because the reciprocity theorem also is valid for doubly terminated networks.

Strangely, there are a few double-match cases where the single-match relationships result in a lower maximum loss than when using the double-match relationships. In that event, the given Q_S must actually be increased, storing more energy and challenging common intuition. The required double-match source Q must be obtained with a "pad" on the source resonator as shown in Figure 5.2.6. Here, $Q_g \approx Q_S$.

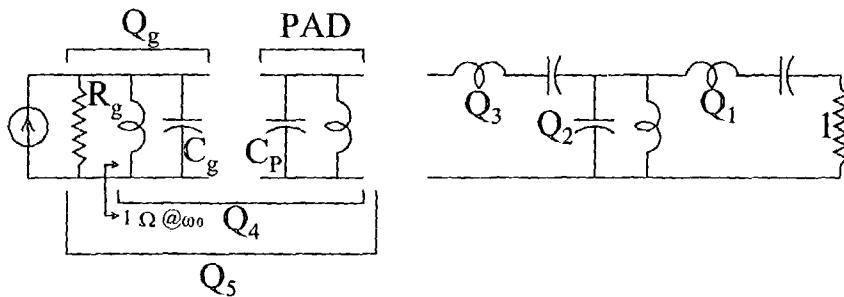


Figure 5.2.6. In some cases, a capacitive source pad minimizes reflection.

Example 5.2.1. A double-match problem specifies $Q_L=3$ and $Q_S=1.5$ to be matched over a 40% bandwidth using a three-resonator matching network. *Problem:* Design the network, including any source resonator pad reactances. *Solution:* The ALLCHEBY program provides the data in Table 5.2.1.

Table 5.2.1. ALLCHEBY Data for QL=3, QS=1.5 Over 40% BW.

Is this a Matching or a Filter (QL=0=QS) network (,F)? M
 N, QLoad, %BW =? 4,3,40
 Is Source Loaded Q Optimal, Given, or Infinite (,G,I)? G
 Q Source =? 1.5
 MAXIMUM INSERTION LOSS FOR INFINITE N IS 0.0232 dB
 THIS INSERTION LOSS FROM 0.0402 TO 0.0681 dB
 THIS dB FLAT LOSS = 0.0402
 THIS dB RIPPLE = 0.0279
 SWR FROM 1.2124 TO 1.2850
 RETURN LOSS FROM 20.3555 TO 18.0811 dB
 Lser or Cpar G(1) = 1.2000 Q(1) = 3.0000
 Cpar or Lser G(2) = 1.1322 Q(2) = 2.8306
 Lser or Cpar G(3) = 1.6173 Q(3) = 4.0431
 Cpar or Lser G(4) = 0.5180 Q(4) = 1.2949
 Ohms or Mhos G(5) = 1.2850 Q(5) = 1.6639
 ***** Q(4) includes pad to increase given terminal Q to Q(5) *****

When the last line in Table 5.2.1 occurs, three steps are required:

1. *Find* $C_g = \frac{1}{X_g} = \frac{Q_g}{R_g}$,
 2. *Find* $C_p = C_N - C_g$,
 3. *Note that* $Q_{N+1} \geq Q_s$:
- $$Q_{N+1} = R_g \times C_N . \quad (5.2.8)$$

Using the above data, $C_g=1.1673$, $C_4=1.2949$, and $C_p=0.1276$ farads.

In each of the cases in this section, $N \rightarrow \infty$ can be considered just as it is in (5.2.2) for the single-match case. As noted in Section 4.4.3, zero ripple and flat transducer loss cannot be obtained in the double-match case, no matter how large N. In the singly-terminated cases, the power ripple represents the ripple in input resistance or conductance; see Section 2.1.5. That ripple cannot approach zero as $N \rightarrow \infty$.

5.2.6 Measured Loaded Q

In those few cases where analytic gain-bandwidth theory can be applied to practical situations, it is necessary to measure the loaded Q of a terminating impedance. Accurate measurements at many frequency samples are readily obtained using an automatic network analyzer. Three references are discussed briefly to suggest effective ways to process those data, and they all are based on the circular image of resonator impedance versus frequency on a Smith chart.

The first method utilizes a printed image of the network analyzer's polar display of a resonator located at the end of a transmission line [Kajfez]. Dimensions and angles are measured, based on consideration of a single resonance with possibly an impedance in series. The frequencies of two different points on the circular image and its crossover points (zero and poles) are required. There is no data averaging to minimize error (other than that built into the analyzer).

The second method is also based on examining the circular image on a Smith chart, in this case at the $Q=1$ intercepts [Weiss]. The article is intended for measurement of high Q resonators, such as small antennas on aircraft. It does not account for incidental impedance and does not average data to minimize error.

The third method for measuring the loaded Q of a resonator applies for both high and low values of Q [Drozd]. It does average data and does not require reading information from a Smith chart, which is tedious and can lead to inaccurate results. The method is based on the relation of stored electric and magnetic energy to the slope of reactance (or susceptance) versus frequency. Because resonator reactance versus frequency is fairly linear in the vicinity of resonance, S parameter data closely-spaced in frequency can be used to estimate the reactance derivative (slope) by linear regression, which averages the data. The article includes a comparison with other methods.

The major limitation of any method of measuring loaded Q based on simple lowpass RC or bandpass RLC impedance behavior is that it does not fit the lumped-element model in many situations. For example, it is well known that the input impedance of many transistors can be modeled as a series RC circuit and the output impedance as a parallel RC circuit. One example where this is only half true is in the data for a 5 GHz FET [Ha:164]. The real-frequency methods in the following sections enable matching arbitrary impedance behavior.

5.3 Real-Frequency Technique

The real-frequency technique (RFT) was introduced in 1977 by H. J. Carlin and extended by his students and other researchers [Carlin,1977,1998]. The single-match problem is shown in Figure 5.3.1 when $Z_s=R_s+j0$. The load can be considered a one-port network with unknown characteristics other than impedance data tabulated at discrete passband frequencies, as shown in Figure 5.3.1.

The power transfer through the lossless matching network can be determined from either port, because reflectance $|\rho_{in}| = |\rho_q|$. The modulus of generalized reflection coefficient ρ_q , defined by (2.1.8), depends on back impedance $Z_q=R_q+jX_q$ as well as load impedance Z_L . The objective is to minimize the reflectance in the pass band.

ω	R_L	X_L
ω_1	R_{L1}	X_{L1}
ω_2	R_{L2}	X_{L2}
ω_3	R_{L3}	X_{L3}
⋮		
$\omega_{(k-1)}$	$R_{L(k-1)}$	$X_{L(k-1)}$
ω_k	R_{Lk}	X_{Lk}

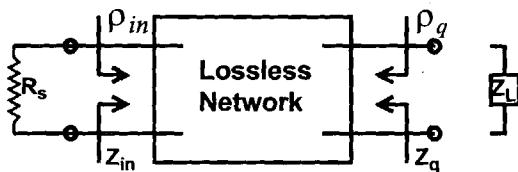


Figure 5.3.1. Discrete impedance data are used in the real-frequency method.

The RFT assumes a parameterized piece-wise linear function to represent $R_q(\omega)$ over all passband and stopband frequencies, calculates the $X_q(\omega)$ function (in similar form) dictated by the Hilbert transform, and optimizes the parameters of R_q to obtain a desired transducer loss over the passband. Next, a second optimization determines a rational resistance function, which is then converted into a rational function of Z_q that can be synthesized.

The need for numerical optimization as opposed to an analytical approach is discussed; an overview of Carlin's design steps is presented; and the difficulties, pitfalls, and advantages of the technique are mentioned.

A third approach has been proposed; treating the single-match broadband problem as a minimum-distance problem in the space of bounded functions. Network realizability is confirmed by testing a combination of Toeplitz and Hankel matrices for positive definiteness. No examples have been published so far [Helton], [Carlin,1992:497], [Carlin,1998:385,421].

5.3.1 Non-Analytic Real-Frequency Data

Single-match gain-bandwidth limitations (lowpass CR or LR load with NIN=1) pose only one constraint, a single nonlinear equation that can be solved to relate the a and b variables introduced in Section 5.2.2. An LCR lowpass load with two transmission zeros at infinity (NIN=2) introduces two nonlinear constraints with a more difficult solution [Chen]. No complete analytic solutions have been published for more complicated loads. In the real-frequency situation arising from a measured load, there is no way to know the number or nature of the

nonlinear constraints, and an appropriate transfer function cannot be prescribed.

When one or both terminations are not pure resistances, doubly-terminated filter transfer functions can no longer provide optimum properties; e.g., in the Chebyshev case, minimum passband loss and maximum stopband selectivity are not optimal. There is no known method for prescribing the best transfer function for an arbitrary terminating impedance function. However, numerical methods can directly optimize the passband minimax loss without wasting matching network elements on obtaining an inappropriate transfer function. At the same time, numerical methods can guarantee a desirable topology, namely a bandpass ladder network without the limitations of the lowpass-to-bandpass transformation. Therefore, numerical methods yield better results than the analytic solution [Carlin,1998].

5.3.2 Approximating the Network Resistance Function

Of several alternatives, Carlin introduced a piece-wise linear approximation for resistance function R_q , the real part of impedance Z_q in Figure 5.3.1. Figure 5.3.2 shows such a function, where the parameters for optimization are r_0 , r_1 , and r_2 . Frequencies for the joints are

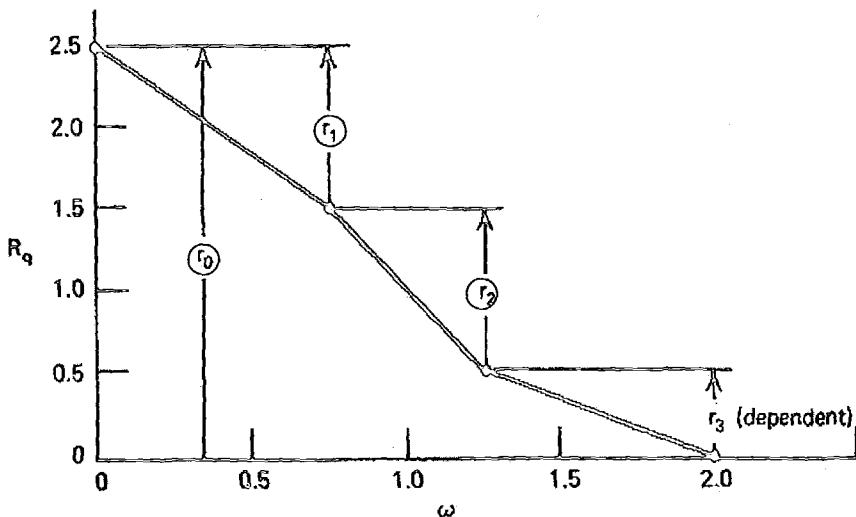


Figure 5.3.2. A piecewise linear approximation of transducer resistance.

arbitrarily chosen, and the function is band limited; i.e. $R_q=0$ for stopband frequencies, to provide finite integration limits in the Hilbert transform (2.6.11). For a given set of coefficients, r_i , resistance $R_q(\omega_i)$ can be obtained by the piece-wise approximation, and $X_q(\omega_i)$ is easily obtained by the Hilbert transform for any frequency, ω_i [Cuthbert,1983:220], [Carlin,1998:126-132]. See Section 2.6.5.

In general, $R_q(\omega)$ is expanded as a linear combination of even basis functions with coefficients, r_k , to be determined, and $X_q(\omega)$ is a linear combination of odd basis functions with the same coefficients [Carlin, 1992:498]. These linear combinations are

$$\begin{aligned} R_q(\omega) &= \sum_{k=0}^n r_k a_k(\omega), \\ X_q(\omega) &= \sum_{k=0}^n r_k b_k(\omega). \end{aligned} \quad (5.3.1)$$

The $a_k(\omega)$ basis functions that appear in Figure 5.3.1 are linear ramp segments, and the related $b_k(\omega)$ basis functions result from simplification of the Hilbert transform (2.6.11) by piece-wise linearizing the integrand numerator.

A resistance function of a resistively-terminated two-port ladder network, R_q , having all its transmission zeros at infinity and/or dc, must be a rational, even function of frequency [Carlin, 1983:21]:

$$R(\omega) = \frac{A_0 \omega^{2m}}{B_0 + B_1 \omega^2 + \dots + 1 \omega^{2n}}, \quad m \leq n. \quad (5.3.2)$$

The original RFT first obtained the piece-wise linear approximation by optimizing the transducer gain function, (2.1.2), using the r_i in Figure 5.3.1 as variables. That was followed by a second optimization to adjust the coefficients in (5.3.2) to fit the optimal piece-wise representation of R_q . Finding good initial values for optimization variables is usually a separate problem; in the piece-wise representation for R_q it is possible to assume a conjugate match at the load port in Figure 5.3.1, which provides initial values of the r_i variables that may lead to satisfactory convergence [Carlin, 1977].

Another way to approximate $R_q(\omega)$ so that parameters can be adjusted for optimal transducer gain is the Wiener-Lee transforms [Carlin, 1992]. This method approximates $Z_q(\omega)$ by a truncated Fourier series in the form of (5.3.1), where the resistance is the weighted sum of cosine basis functions, a_k , and the reactance similarly employs only sine functions, b_k . This technique avoids the singularities that occur in the line segment representation, which can impede gain optimization. A rectangular gain objective can be approximated by an integral in continuous frequency (as opposed to discrete samples) by using the fast Fourier transform algorithm. Starting values for gain optimization are obtained by using a simple application of the piece-wise linear approximation in Figure 5.3.2. Although there are some difficult details, a potential outcome of the Wiener-Lee transform method meets a fundamental need: finding a good approximation to the maximum-possible gain for $N \rightarrow \infty$, comparable to Section 5.2.5 for the analytic method.

5.3.3 Synthesis of a Resistance Function

The remaining RFT step is to synthesize a ladder network from a known resistance function (5.3.2). This is most easily accomplished by the *Gewertz procedure* [Carlin,1998:334], [Cuthbert,1983:58] which produces an RLC impedance

$$Z_q(s) = \frac{N(s)}{D(s)}. \quad (5.3.3)$$

Briefly, with the substitution $\omega^2 = -s^2$ in (5.3.2), the LHP roots of the denominator form the polynomial D(s). Coefficients of polynomial N(s) are found by the Gewertz procedure as the solution of a system of linear equations involving the other coefficients already obtained [Cuthbert,1983:58].

In order to realize the LC matching ladder network using continued-fraction expansion, the open-circuit impedance parameter, z_{11} , must be extracted from the known RLC impedance $Z_q(s)$ in (5.3.3). See [Carlin,1998:221,279], [Cuthbert,1983:60-61]. In this approach, there is no guarantee that all matching network elements will be positive.

5.3.4 Double-Matching Using the RFT

The double-matching problem is shown in Figure 5.2.1, where $Z_S = R_S + jX_S$ is obtained from tabulated data at the same frequencies as for the load impedance. It is shown in (2.1.8) that for lossless two-port networks the magnitude of the reflectance is constant at any interface, especially at the input and output ports. There is a relative phase-angle factor between the output and input generalized reflection coefficients, say $e^{j\phi}$, implicitly related to $Z_q(\omega)$, and angle ϕ must be known in order to compute the transducer power gain.

A complication arises because $Z_q(\omega)$ is defined as the back impedance at the output port when $X_S=0$. That is a necessary restriction in order that the network can be realized from $Z_q(s)$ according to the preceding section. See Figure 5.3.1; at any frequency and $Z_S=1+j0$,

$$\rho_{in} = \frac{Z_{in} - 1}{Z_{in} + 1} = e^{j\phi} \left| \frac{Z_L - Z_q^*}{Z_L + Z_q} \right|. \quad (5.3.4)$$

Then it can be shown [Carlin,1983] that the transducer power gain, T, for Figure 5.2.1 is

$$|S_{21}|^2 = \frac{P_L}{P_{as}} = \frac{\left(1 - |\rho_s|^2\right)\left(1 - |\rho_{in}|^2\right)}{\left|1 - \rho_s \rho_{in}\right|^2}, \quad (5.3.5)$$

where ρ_s is the reflection coefficient of the actual source impedance wrt 1 ohm:

$$\rho_s \equiv \frac{Z_s - 1}{Z_s + 1}. \quad (5.3.6)$$

The phase angle, ϕ in (5.3.4), is the transfer phase angle when there is a purely resistive match at the load port. It can be obtained from the resistance function in (5.3.2). It has been shown [Carlin,1983] that

$$\phi(\omega) = -m\pi - 2\arg D(\omega), \quad (5.3.7)$$

where m is from (5.3.2) and $D(\omega)$ is from (5.3.3).

A third nonlinear optimization is thus required, because the passband gain function involves ϕ in (5.3.4). Unfortunately, this last optimization includes the Gewertz procedure within the iterative process [Yarman,1990:214]. Otherwise, the double-match procedure is similar to the single-match design, but with another layer of complication and potential ill-conditioning.

5.3.5 Double-Matching Using Brune Functions

A notable improvement in real-frequency methods avoids explicit factorization of polynomials and repeated use of the Gewertz procedure with its solution of a system of linear equations. The major optimization step adjusts the poles of the impedance function $Z_q(s)$ to obtain a satisfactory transducer gain function. Not only is $Z_q(s)$ *positive definite* (guaranteeing all positive element values), but the broadband match performance is known before the network is realized [Yarman,1990], [Fettweis].

The improvement is obtained by expressing the back impedance, $Z_q(s)$ in Figure 5.3.1 ($X_S=0$), as a Brune function:

$$Z(s) = C_0 + \sum_{j=1}^n \frac{C_j}{s - s_j}, \quad s = \sigma + j\omega. \quad (5.3.8)$$

The coefficients, C_j , are computed in terms of the poles, s_j , of $Z_q(s)$, but their formula guarantees that $Z(s)$ will be positive definite, another consequence of the Hilbert transform [Forster]. Having a guess for values $s_j=\sigma_j+j\omega_j$, $j=1$ to n , immediately provides a way to compute $Z_q(j\omega_i)$, where each ω_i is a passband frequency sample for evaluating the transducer gain function. Therefore, the success of the optimization is known with the assurance that a real non-negative network can be found. Note that the degree of the ladder network, n , must be chosen arbitrarily.

The Brune function in (5.3.8) is similar to a partial fraction expansion, and a standard algorithm can collapse that summation into a rational form, as in (5.3.3), for network realization. The use of Brune functions reduces matching to a classical optimization problem having improved numerical stability. It also provides explicit expressions for the exact partial derivatives required by the optimization algorithm. Unfortunately, there is the usual problem of finding starting values of the s -plane poles, and Carlin's RFT for single matching is the recommended way to obtain them. Unfortunately, that process requires root finding [Yarman,1990:217].

5.3.6 Double-Matching Active Devices

Microwave amplifiers often consist of several transistors separated by matching networks as well as the terminal networks at both ends. The design of any internal network is a double-matching problem. Because the transistors are usually characterized by scattering parameters rather than impedance, a design scheme has been developed [Carlin, 1998:433] that differs from that in Sections 5.3.4 and 5.3.5.

Briefly, transducer power gain (5.3.5) is restated in terms of unit-normalized reflection coefficients similar to (5.3.6). Specifically, the input scattering parameter is

$$s_{11} = \frac{Z_{in} - 1}{Z_{in} + 1} = \frac{h(s)}{g(s)} = \frac{h_0 + h_1 s + \dots + h_n s^n}{g_0 + g_1 s + \dots + g_n s^n}. \quad (5.3.9)$$

Assuming a bandpass matching network, a guess for all the numerator coefficients, h_i , enables construction of the polynomial $g(s) \times g(-s)$. Spectral factorization (i.e., find all the LHP plane roots) provides a Hurwitz $g(s)$; therefore, $g(s)$ is obtained in a way that guarantees a matching network having all positive elements. Unfortunately, the optimization of the transducer power gain uses the numerator coefficients in (5.3.9) as variables, so the iteration may be ill-conditioned as well as having root-finding in each iteration. Also, the maximum possible gain factor must be guessed by cut-and-try. See [Yarman, 1982].

This so-called "simplified" RFT has been extended to incorporate transistor stability and input/out matching network design, all occurring concurrently in the optimization [Jung].

5.4 Introduction to GRABIM

The grid approach to broadband impedance matching (GRABIM) is much easier to understand and utilize than the RFT in Section 5.3. This section recalls the trends, observed in this and the preceding chapter, which suggest GRABIM. An overview of GRABIM concepts indicates how and why this new broadband matching method is so effective. This section emphasizes four innovations:

1. All reflection functions versus element parameters are known unimodal, monotonic, or similar shapes that compose the worst-case envelope over frequency,
2. The reflection functions are well-conditioned because of frequency and impedance scaling and transformations that control equivalent ladder elements for cascade transmission lines,
3. Grid searches over large, medium, and small hypercubes (lattices) that are repositioned in the solution space

approximately locate the likely global minimax solution while avoiding local anomalies in the envelope surface, and

4. The final minimax-constrained gradient optimization precisely locates the likely global solution and prunes the topology to produce a reduced-degree full-rank equalizer.

5.4.1 Thesis

The real-frequency technique (RFT) has evolved continually since its introduction more than 20 years ago. The ability to deal directly with terminal impedances known at discrete frequencies is far more realistic than hoping to find correspondences between a very few ideal load impedance models and given termination data. However, the ability to make good use of the RFT depends strongly on the designer's thorough background in academic circuit theory. That background is desirable and may be sufficient, but it is not necessary.

The RFT employs several uncertain optimization procedures and utilizes the mathematics of network synthesis throughout. Synthesis of filters by iterated analysis, Section 4.5, is far less complicated and depends on just one highly-structured numerical optimization – it is a big step forward [Sussman-Fort, 1991]. Iterated analysis also depends on knowing the frequencies where passband loss must be zero, and it is shown in Section 4.2 that any Chebyshev equal-ripple filter response is completely determined by the locations of its transmission zeros. For bandpass networks, there still is the uncertainty of finding reliable starting values for the optimization variables.

Section 4.4 shows that, for Chebyshev equal-ripple passband response, applying simple constraints to network terminations forces the passband to have flat loss at the valleys to maintain passband width. The ripple loss peaks still can be minimized when there is one load constraint, but the peak loss is strictly dependent when a second constraint is added, e.g. the double-match problem. However, Figures 4.4.4 and 4.4.5 show how the input reflectance at each frequency behaves as a function of a typical network branch value. That analytic solution brings all the valley frequency functions through the same valley flat loss (Figure 4.4.4) and all the peak frequency functions through the same peak loss (Figure 4.4.5) when plotted versus branch value. Filter transfer functions, like Chebyshev equal ripple, are not optimal for broadband matching.

It is possible to sample enough passband frequencies for a structured optimization to minimize the maximum loss over all the pass band, thus avoiding the limitation of a particular transfer function (as does the RFT). But GRABIM solves the problem of starting values for variables while avoiding the mathematics of network synthesis and uncertain optimization. It must be emphasized that GRABIM is not an all-out application of optimization *per se*, e.g. [Kintscher]. GRABIM is a

reasoned approach for getting the best possible broadband matching solution with the least possible human effort. The power of the personal computer is a key ingredient in this approach, as was argued in behalf of iterated analysis where its speed of operation allowed simplicity [Orchard].

The following sections define broadband matching as a highly structured optimization problem, based on frequency-sampled terminating impedances which may have a number of different values at each frequency. Ladder networks are considered as solutions, and their automatic generation is introduced. Then the grid search locates the neighborhood of the global minimum, and the minimax-constrained optimization removes excess network elements. Reasons why these techniques work so well and how they can be made efficient accompany their exposition. Surplus matching network branches are mentioned in the context of the surplus elements and degrees of freedom encountered in direct-coupled filters, Section 3.4.4.

5.4.2 Overview

Broadband matching requires design of a matching network (equalizer) placed between a complex source and a complex load, as shown in Figure 5.4.1. The matching network may contain any ladder

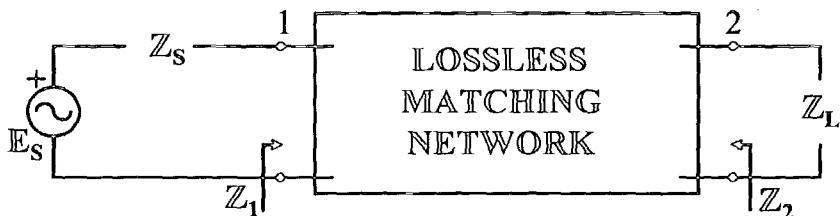


Figure 5.4.1. Broadband matching network terminations and impedances.

arrangement of inductors, capacitors, short- or open-circuit transmission-line stubs, cascaded transmission lines, and an ideal transformer.

The terminal impedance data are tabulated versus frequency as illustrated in Table 5.4.1. The data are normalized to one rad/s and 1 ohm, a necessity for GRABIM. (As explained below, it is unimportant which passband frequency corresponds to 1 rad/s, except when the matching network includes open- or short-circuit stubs.) The "Goal" or target data are the maximum desired network insertion loss at each frequency. These data are usually obtained by an S-parameter test set. Program S11TOZ.EXE is a convenient way to convert S-parameters in numeric/degree ASCII format to impedances normalized to one ohm.

Although Table 5.4.1 shows only a single impedance per frequency for a termination, there may be a cluster of discrete impedance data defining a neighborhood at each frequency. Impedance neighborhoods occur in beam-steered antenna arrays and mobile antennas, and they

arise in transistor impedance uncertainty because of either lot variation or operation into saturation.

Table 5.4.1. Measured Impedance Data for Source and Load.

ω	Goal	R_L	X_L	R_S	X_S
.3	0	0.59016	0.71680	0.71313	-0.45230
.4	0	0.71910	0.74944	0.56081	-0.49629
.5	0	0.80000	0.77500	0.41945	-0.49347
.6	0	0.85207	0.80503	0.29915	-0.45789
.7	0	0.88688	0.84174	0.20274	-0.40204
.8	0	0.91103	0.88470	0.12909	-0.33530
.9	0	0.92837	0.93288	0.07539	-0.26403
1.	0	0.94118	0.98529	0.03846	-0.19231

Separate computer data files

As presented in Section 2.1.4, the power transfer in Figure 5.4.1 is described by

$$T \equiv \frac{P_L}{P_{as}} = 1 - |\rho_1|^2 = 1 - |\rho_2|^2, \quad (5.4.1)$$

where the generalized reflection coefficients are

$$\rho_1 \equiv \frac{Z_1 - Z_s^*}{Z_1 + Z_s}, \quad \rho_2 \equiv \frac{Z_2 - Z_L^*}{Z_2 + Z_L}. \quad (5.4.2)$$

These coefficients are defined wrt fixed impedance data, i.e. Z_s and Z_L . Note that $|\rho_2|$ in (5.4.1) would be unchanged if ρ_2 were defined wrt Z_2 instead. Transducer power gain in terms of the ABCD parameters of the two-port network in Figure 5.4.1 is [Frickey]

$$T \equiv \frac{P_L}{P_{as}} = \frac{4R_s R_L}{(AR_L - Cq + DR_s)^2 + (B + Cp + DX_s + AX_L)^2}, \quad (5.4.3)$$

$$q = (X_s R_L + X_L R_s), \quad p = (R_s R_L - X_s X_L),$$

where A, B, C, and D are real numbers (the j operator is factored from B and C). The range of T is from 0 (worst) to 1.0 (best). It is efficient to evaluate T in (5.4.3), and then to use (5.4.1) to obtain $|\rho|$.

The objective of broadband matching is

$$\text{Minimize } |\rho|_{\max} \text{ over vector } v \quad (5.4.4)$$

subject to the inequality constraints

$$\rho_i(v) \equiv |\rho(v, \omega_i, D_i)| \leq |\rho|_{\max}, \quad i = 1 \text{ to } m. \quad (5.4.5)$$

There are m sampled frequency points in (5.4.4), e.g. $m=8$ as in Table 5.4.1. For each of the sampled frequencies, ω_i , there is at least one associated data subset, D_i , which contains the goal, R_L , X_L , R_S , and X_S

values for each frequency. The column vector v contains all the branch variables of the matching network², i.e., values of L , C , Z_0 , θ_0 , and t^2 :

$$v = (v_1, v_2, \dots, v_j, \dots, v_n)^T. \quad (5.4.6)$$

(A candidate network could have the PiCPiL topology in Figure 5.5.2, for example.) For a candidate network topology with assigned element values in v , there is a reflection magnitude function (reflectance), $|\rho(\omega)|$, that takes on values ρ_i at corresponding frequencies, ω_i . The constraints in (5.4.5) require each $\rho_i(v)$ to be less than $|\rho|_{\max}$ by varying the n network element values in vector v , while (5.4.4) requires $|\rho|_{\max}$ to be made as small as possible. This is a statement of a *minimax problem* (minimize the maximum). From the iterated analysis method in Section 4.5.3, it is necessary to have $m > (n+1)$ frequency samples spread over the given passband.

To further illustrate (5.4.5), consider the graphs of the eight $\rho_i(v_5)$ functions in Figure 5.4.2, and note that they are particularly well behaved (benign). The numbered frequencies correspond to those in Table 5.4.1 and are labeled by their respective curves. In this case, the minimum $|\rho|_{\max} \approx 0.5$ and $v_5 \approx 0.72$ is optimal. Notice that different segments of the ρ_i constraint functions constitute the worst-case envelope as only v_5 is varied. A closer view of the minimum $|\rho|_{\max}$ and the ρ_i

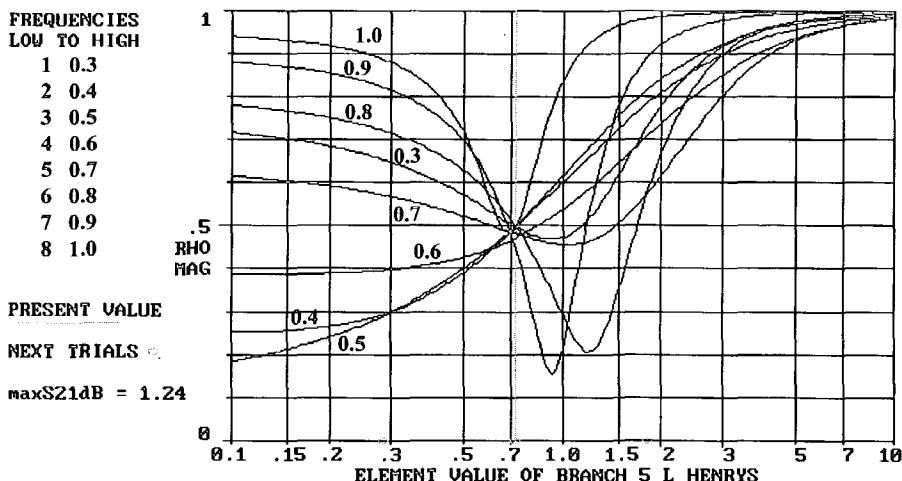


Figure 5.4.2. Typical reflectance functions versus a series branch inductance.

values at the eight frequencies is shown in Figure 5.4.3. Three of the eight ρ_i constraint functions in (5.4.5) are not *binding constraints* at the

² The reader is cautioned against confusing lower-case italicized v , v , with a Greek letter. The difference in this letter between the equations and the text is unfortunate but unavoidable.

optimum value of v_5 ; i.e., they are more than satisfied (not "active"), because their reflectances are less than the maximum at optimal v . (Figures 5.4.2 and 5.4.3 show data produced by Example 5.6.1.)

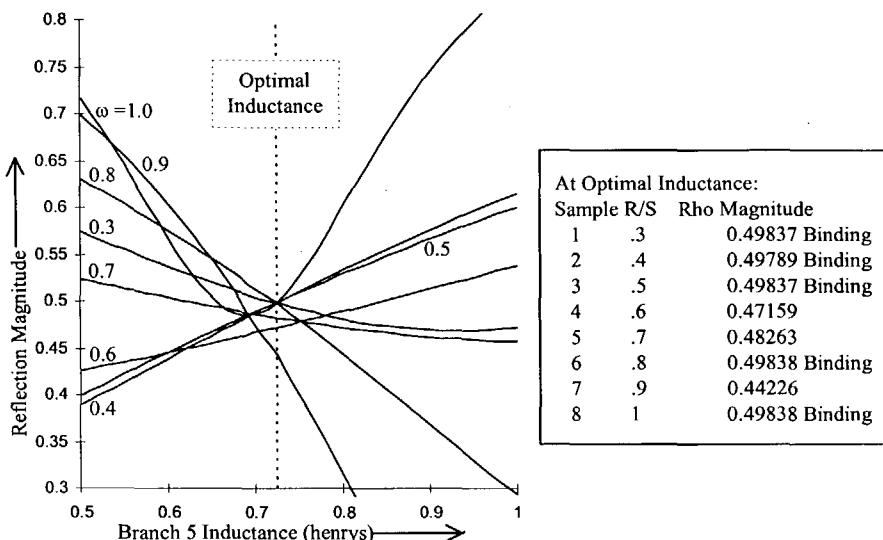


Figure 5.4.3. Closeup of binding and nonbinding reflectance functions.

A candidate ladder network having n variables is selected. Then, an approximate global minimax solution is obtained by testing a sequence of discrete combinations of all elements of v , followed by a strongly convergent minimax solution that obtains the global minimum and discards (prunes) branches not involved in the solution. It is important to justify the benign behavior of the $\rho_i(v_i)$ of each type of variable allowed in the ladder network is described. Then, cross sections of the resulting $\rho_i(v)$ surface are examined in the branch directions, as in Figure 5.4.2, and along the principal diagonals of hypercubes in n -space. In almost every case, the minimum so efficiently located is the global minimum for that minimal candidate topology, i.e., the likely least possible $|\rho|_{\max}$.

5.4.3 Branch Parameter and Reactance Effects

The variables in the candidate lossless networks are single L's and C's in series or parallel branches, the L in parallel LC in series branches or vice versa (traps), turns-squared ratio of an ideal transformer, and the characteristic impedances (Z_0) and electrical lengths (θ_0) of cascade transmission lines. Short- or open-circuited transmission-line stubs connected as series or parallel branches produce reactance values comparable to branch L's and C's except for stub periodicity, which is considered below.

The series branch reactance $X(\omega_i)$ in Figure 5.4.4 corresponds to a circular image in the input ρ Smith chart because of the inherent bilinear mapping in linear networks; see Appendix Section A.3.1. When $X \rightarrow \pm\infty$

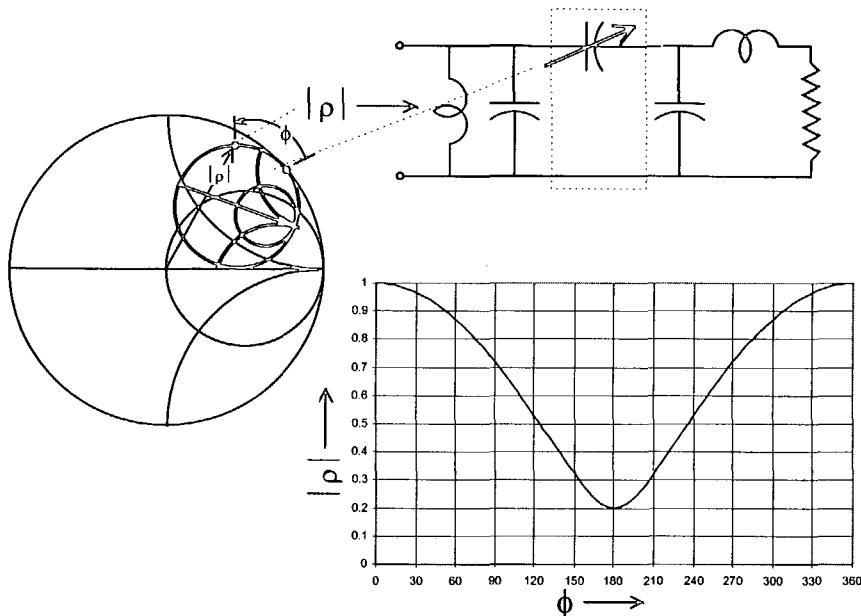


Figure 5.4.4. Origin of all ρ_i reflectance curves vs branch jX values.

the network between the series X branch and the input is purely reactive, so that ρ must be on the chart perimeter. The ρ image circle may encircle the chart origin, which represents the complex normalizing impedance described in Section 2.1.2. To cover the general case where the image circle may not touch the perimeter, a straightforward analysis of the geometry yields the relationship of $|\rho|$ in Figure 5.4.5 in terms of the ρ radius to the center of the image circle, R_{cen} , and the image circle's radius, R_{cir} :

$$|\rho| = \left[(R_{cir} + R_{cen})^2 - 2R_{cir}R_{cen}(1 - \cos\phi) \right]^{1/2}. \quad (5.4.7)$$

Angle ϕ is the angle around the smaller Smith chart image relative to the point of tangency with the larger Smith chart. The graph in Figure 5.4.4 plots (5.4.7) for $R_{cen}+R_{cir}=1$, i.e., when the image circle touches the perimeter. The relationship of ϕ to reactance X is a nonlinear but monotonic composite function, so that X would have to vary from $-\infty$ to $+\infty$ for ϕ to vary over its entire 360° range. The nature of the $|\rho_i(v_i)|$ reflectance curves in Figure 5.4.2 is now apparent: they are just skewed arcs of the curve plotted in Figure 5.4.4. The full 360° range is never traversed by the reactance of any single L or C, which can provide only all-positive or all-negative values, respectively. In Figure 5.4.2, the

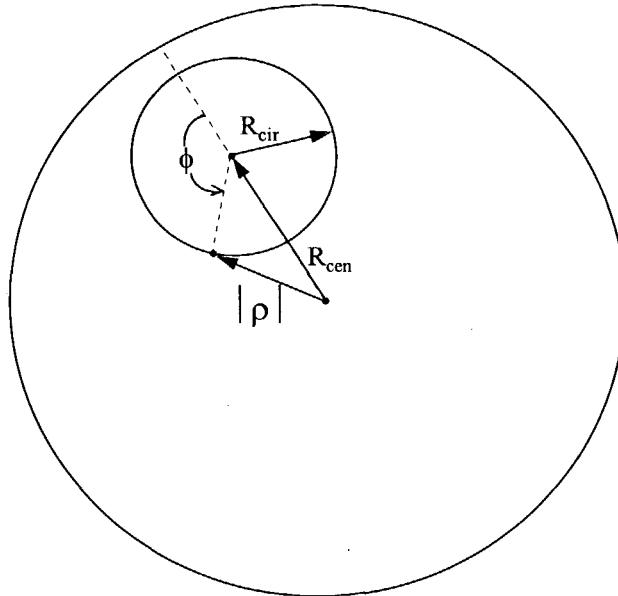


Figure 5.4.5. Geometry of a ρ vector terminating on a general circle image.

branch inductor must be in series, as seen by $|\rho| \rightarrow 1$ as $L \rightarrow +\infty$ to the right; to the left, $L \rightarrow 0$, and $|\rho|$ does not reach unity because of the effect of the load resistance.

A more organized analysis of any of the several types of variables is obtained by considering the internal interfaces a network component may have, as shown in Figure 5.4.6. A typical component two-port, such

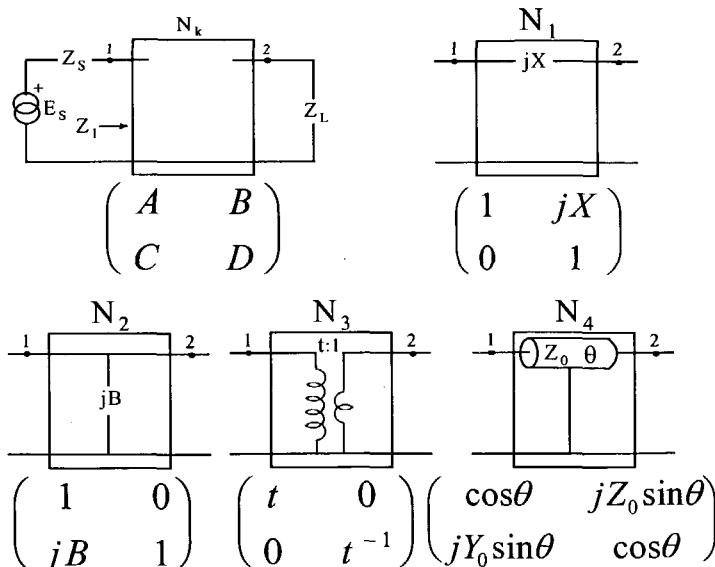


Figure 5.4.6. All lossless subnetworks for interfaces reflectance analyses.

as in Figure 5.4.6, is shown within dotted lines in Figure 5.4.4. Analysis of just the subnetwork is valid because of Thevenin equivalent impedances at any interface and the constant generalized reflection magnitudes (reflectances) throughout lossless networks, as in Figure 2.1.6. Therefore, application of (5.4.3), to find reflection magnitude $|\alpha(Z_1)|$ in Figure 5.4.6, also determines $|\rho|$, as in Figure 5.4.4.

The preceding paragraphs already predict the result of such analysis for the series branch, N_1 in Figure 5.4.6, for arbitrary values of Thevenin Z_S and Z_L . For example, Figure 5.4.7 shows reflectance $|\alpha|$ versus positive branch reactance for particular branch-interface impedances (reflectance at port 1 in Figure 5.4.6, or at the left side of the dotted rectangle in Figure 5.4.4). The right-hand side approaches unity

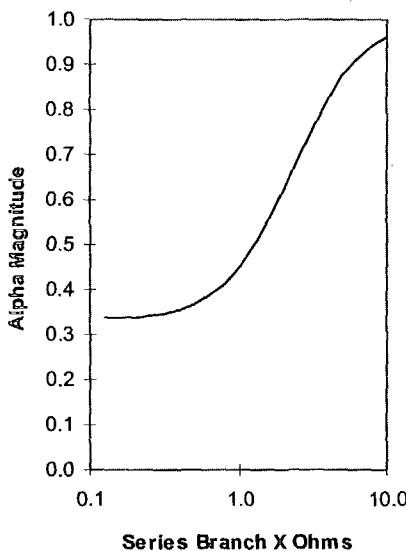


Figure 5.4.7. Reflectance versus a series branch reactance for network N_1 in Figure 5.4.6: $Z_S=2+j3$, $Z_L=1-j3$ ohms.

reflection as the series reactance approaches infinity. However, this is a case where the termination impedances restrict coverage of the composite function, i.e. $0 < \phi \leq 180^\circ$ in Figure 5.4.4. If this were the only branch involved, then it could be removed, because its reactance approaching zero yields the lowest reflectance, i.e. it is *monotonic*.

The shunt branch, N_2 in Figure 5.4.6, is the dual case and can be expected to provide results similar to series branches. For example, Figure 5.4.8 shows $|\alpha|$ versus positive branch susceptance for another particular pair of branch-interface impedances. The right-hand side approaches unity reflection as the shunt susceptance approaches infinity and has a single minimum, i.e. the reflectance is *unimodal*. Therefore,

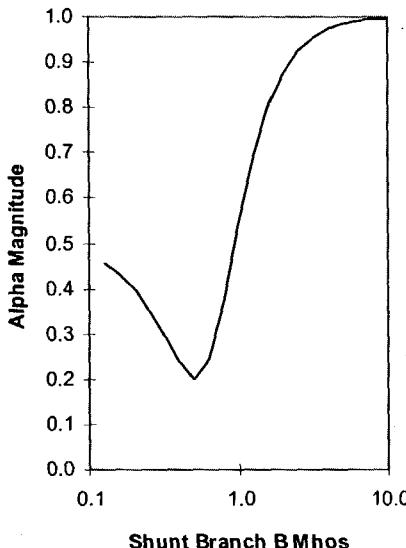


Figure 5.4.8. Reflectance versus a shunt branch susceptance for network N₂ in Figure 5.4.6: Z_S=3+j0.1, Z_L=1+j1 ohms.

series and shunt branches of either L or C cause reflectances that are either monotonic or unimodal.

Series branches may contain a parallel-LC pair to produce a zero of transmission (null) outside the passband; in those cases, the L is the variable and the C depends on both the L and the given null frequency. Similarly, shunt branches may contain a series-LC pair, and the C is the branch variable. Unlike single L's or C's in branches, the point of tangency between the input image circle and the reflection perimeter, Figure 5.4.4, does not correspond to either the zero or infinite value of the branch variable. This situation is comparable to an off-setting "phase" component of ϕ , which shifts and thus eliminates the monotonic or unimodal reflectance characteristics shown in Figure 5.4.7 and 5.4.8. A similar effect is encountered below for the transmission-line θ_0 variable; Section 5.5.2 shows that in both cases the grid search avoids these minor aberrations in those reflectance functions.

The ideal transformer, N₃ in Figure 5.4.6, scales an impedance, Z_L, seen to its right to the impedance Z₁=t²Z_L seen at its input port, where t is the turns ratio. Therefore, an ideal transformer preserves the loaded Q of the impedance, where Q_L=X_L/R_L. Algebra yields |ρ₁| from (5.4.2) as

$$|\rho_1| = \sqrt{\frac{(t^2 R_L - R_s)^2 + (t^2 R_L Q_L + X_s)^2}{(t^2 R_L + R_s)^2 + (t^2 R_L Q_L + X_s)^2}}. \quad (5.4.8)$$

Appendix Section A.3.2 verifies that the reflectance in (5.4.8) must approach unity for very large and very small t², and the reflectance is

unimodal. Figure 5.4.9 shows the reflectance for particular terminations. It is always unimodal over the complete range of turns ratio.

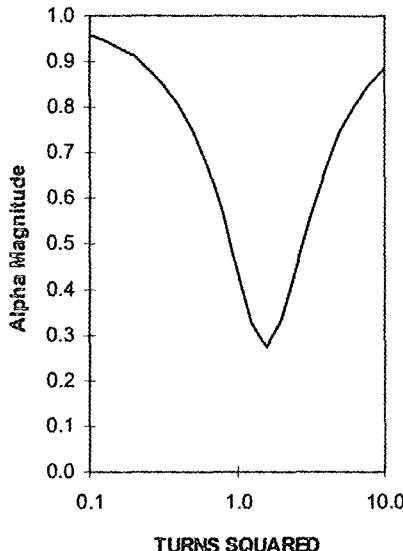


Figure 5.4.9. Reflectance versus transformer squared turns ratio for network N_3 in Figure 5.4.6: $Z_S = 1+j2$, $Z_L = 1-j1$ ohms.

The lossless transmission line, N_4 in Figure 5.4.6, has input impedance (5.1.6)

$$Z_1 = \frac{jyZ_0 + Z_L}{jyZ_L/Z_0 + 1}, \quad y \equiv \tan\theta. \quad (5.4.9)$$

First, consider the behavior of $|\rho_1|$ versus θ . According to Appendix A equation (A.1.1), Z_1 is a bilinear function of y , where y is a monotonic trigonometric function of θ over principal values. Also, (5.4.2) shows that ρ_1 is a bilinear function of Z_1 , so that ρ_1 is a bilinear function of y as well (composite bilinear functions are also bilinear). Appendix Figure A.1.1 shows that when $Z_f^* = Z_0$, a real number, the SWR circle is centered on the origin of the f plane and is eccentric inside the g -plane unit circle. To incorporate the source and load impedances in Figure 5.4.6, suppose that the Z -plane SWR circle passes through Z_L (Z_L causes that SWR wrt Z_0), and the g plane is normalized to Z_S^* . Then as θ varies from -90° to $+90^\circ$, y varies from $-\infty$ to $+\infty$, which traces over the entire SWR circles in the Z , f , and g planes of Figure A.1.1. In the g plane, there is a vector terminating on the image circle, $|g| = |\rho_1|$ as described by (5.4.7) and Figure 5.4.5. However, in this case of a circular image in the interior of the g plane, angle ϕ has an implicit relationship to transmission line length θ , so $|\rho_1|$ varies like a distorted sinusoid of arbitrary phase. A typical example is shown in Figure 5.4.10. Although this is not a

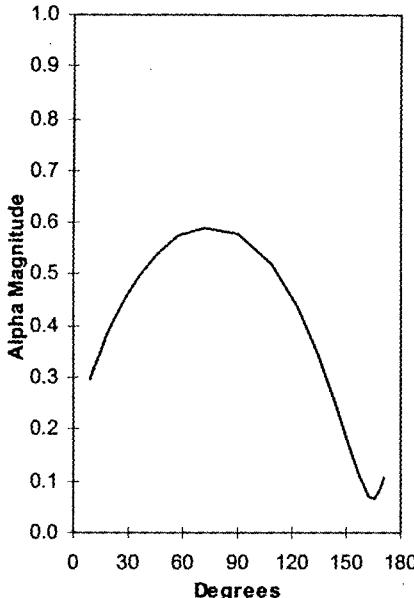


Figure 5.4.10. Reflectance versus cascaded transmission line length θ for network N_4 in Figure 5.4.6: $Z_S=1.0959+j0.211$, $Z_L=0.9615+j0.1923$ ohms.

unimodal response, it is well behaved. Section 5.5.2 shows that the grid search avoids this minor aberration in the reflectance function.

Finally, consider the behavior of $|\rho_1|$ versus Z_0 . Unfortunately, Z_1 in (5.4.9) is not a bilinear function of Z_0 . Fortunately, the following shows that $|\rho_1|$ is either a unimodal function of Z_0 or at worst has two minima (*bimodal*) for $0 \leq \theta \leq 180^\circ$. From N_4 in Figure 5.4.6, it is seen that parameters A and D are not functions of Z_0 , but B is linearly and C is inversely related to Z_0 . Therefore, (5.4.3) shows that $T \rightarrow 0$ when $Z_0 \rightarrow 0$ and also when $Z_0 \rightarrow \infty$. Consider the transducer function $P \equiv 1/T$; using the ABCD parameters given for N_4 in Figure 5.4.6 and the expressions in (5.4.3), it is straightforward to obtain the first derivative of P with respect to Z_0 . Setting that to zero as a necessary condition for points of inflection yields

$$\frac{dP}{dZ_0} = 0 \Rightarrow K_0 + K_1 Z_0 + AXZ_0^3 + (\sin\theta)Z_0^4 = 0, \text{ where}$$

$$K_0 \equiv (-\sin\theta)(q^2 + p^2), \quad K_1 \equiv A(Rq - Xp), \quad (5.4.10)$$

$$R \equiv R_L + R_S, \quad X \equiv X_L + X_S, \quad A = \cos\theta.$$

Constants q and p are defined in (5.4.3). When $Z_S = Z_L = 1+j0$ ohms, the four roots of P occur on the cardinal points of a unit circle in the complex Z_0 plane. For any terminating impedances and transmission line lengths in the range $0 < \theta < 180^\circ$, $\sin\theta > 0$ and $K_0 < 0$. Therefore, there can be only one or three positive-real roots of T in (5.4.10), corresponding to one or two

minima of $|\rho_1(Z_0)|$ for $0 < Z_0 < \infty$. Thus, reflectance function $|\rho_1|$ is either unimodal or bimodal on the real-valued range $0 < Z_0 < \infty$. Figure 5.4.11 shows the nominal unimodal behavior of reflectance versus Z_0 for particular Z_S and Z_L impedances in Figure 5.4.6. Note that the reflectance approaches unity for both extremes of positive-real Z_0 . Section 5.5.2 shows that a bimodal reflectance occurrence does not cause search problems, as confirmed in Figure 5.5.6.

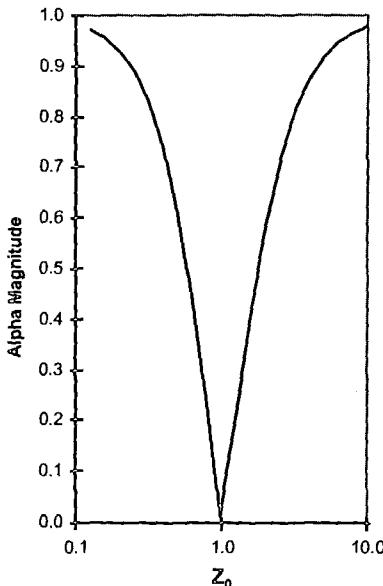


Figure 5.4.11. Reflectance versus cascaded transmission line Z_0 for network N4 in Figure 5.4.6: $Z_S = 1.0959 + j0.211$, $Z_L = 0.9615 + j0.1923$ ohms.

5.4.4 The Response Surface

The minimax response surface defined by (5.4.4) and (5.4.5) is examined by cross sections in one-dimensional subspaces, such as that for branch 5 inductance in Figure 5.4.2. Clearly, the response of interest is not just one reflectance frequency curve but the worst-case reflectance function composed of reflectance segments, i.e. a continuous, non-smooth *envelope function*. The envelope function in each L and C branch cross section is composed of segments of unimodal or monotonic reflectance curves, as described in Section 5.4.3. Thus, the envelope functions also are either unimodal, as in Figure 5.4.2, or monotonic, as in Figure 5.4.12. A typical minimax response function in two-space is shown in Figure 5.4.13. The negative of the surface is shown for an enhanced perspective.

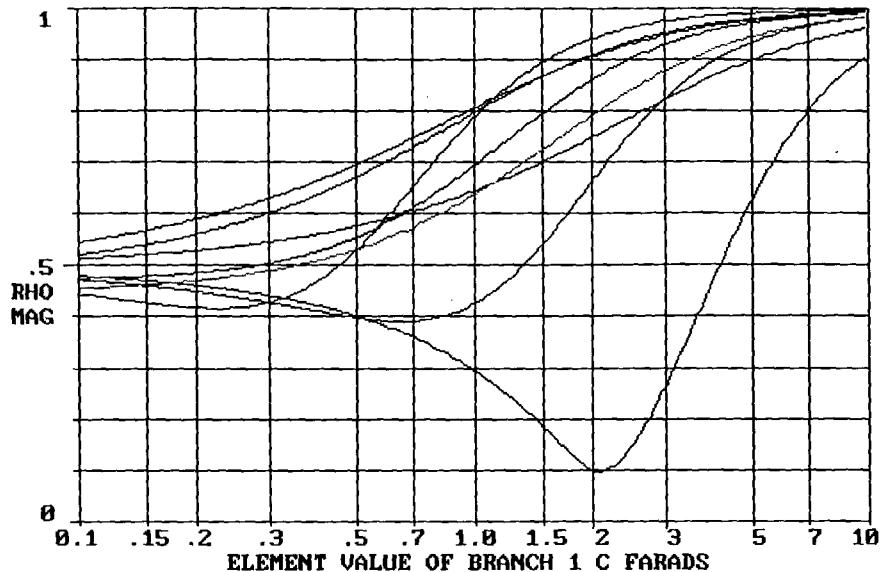


Figure 5.4.12. A monotonic envelope function for a branch leaving the solution.

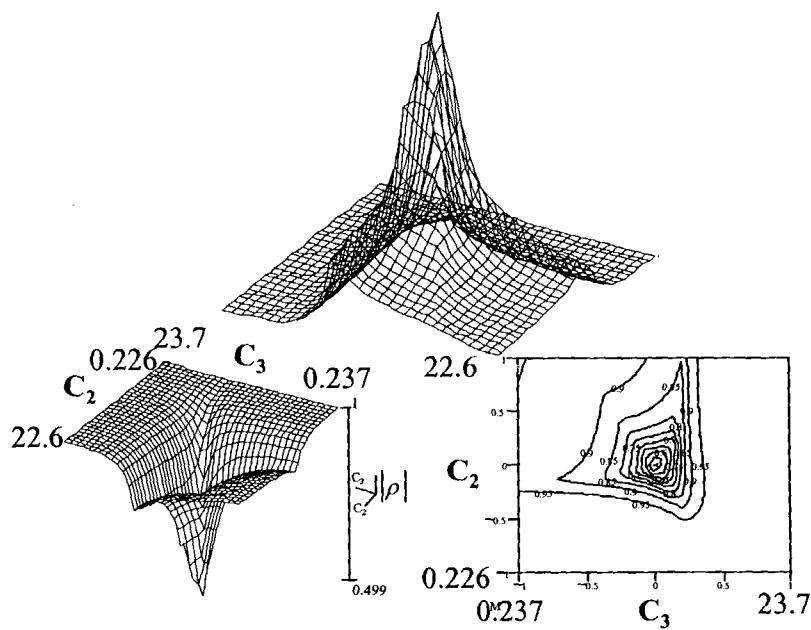


Figure 5.4.13. A typical response surface in two-space from Section 5.6.1.

To obtain well-scaled response surfaces, it is necessary to scale the effects of variables, i.e. values of L , C , squared turns ratio t^2 , Z_0 and θ_0 . Because normalization to 1 rad/s and 1 ohm is assumed, units are henrys, farads, numeric, ohms, and degrees at 1 rad/s, respectively. Practical ranges of variables are shown in Figure 5.4.14 for V , as suggested by transformation Q's in Section 5.1.4 and by broadband Chebyshev network

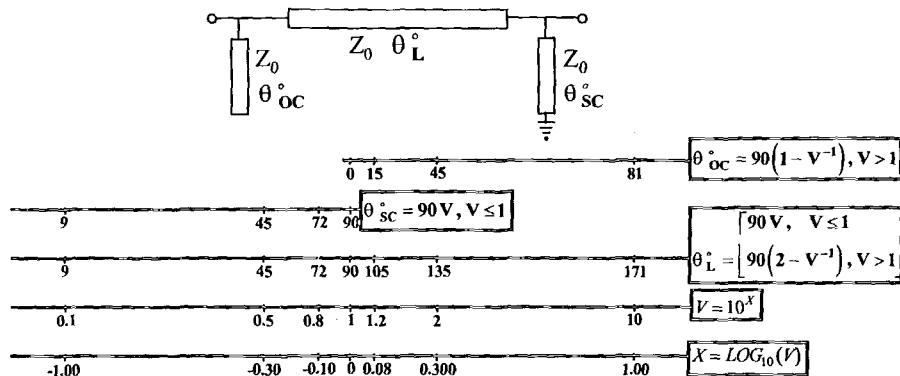


Figure 5.4.14. Mappings for variables to produce well-scaled responses.

element values in Section 5.2.4. The branch cross sections displayed, e.g. Figure 5.4.12, utilize L , C , t^2 , and Z_0 values in the logarithmic V space, which is simply related to the arithmetic X space as shown in Figure 5.4.14. Unbounded gradient optimization in the X space automatically ensures positive element values in the V space, because that amounts to quantifying variables in dB. Every value of x defines a corresponding value $v=10^x$. The value of v can be used directly as L , C , t^2 , or Z_0 . Otherwise, the value of v may define a corresponding value of θ_L , θ_{SC} , or θ_{OC} (at $\omega=1$ rad/s), according to the equations at the top three lines in Figure 5.4.14 or in (6.2.13)-(6.2.14).

Practical cascade transmission lines have reference electrical lengths between 0 and 180°. Transmission-line stubs have reference electrical lengths between 0 and 90°, with a split range to facilitate a grid search being able to select optionally either a short- or an open-circuited stub. The values of electrical length shown in Figure 5.4.14 are those at the defined reference frequency of $\omega_0=1$ rad/s. Therefore, the variable is θ_0 , with the understanding that at any other frequency there is an actual electrical length, θ , that is linearly related to θ_0 :

$$\theta = \theta_0 \frac{\omega}{\omega_0}. \quad (5.4.11)$$

Reference frequency $\omega_0 = 1$ may be at the ends or anywhere within the pass band. However, with stubs it is necessary to limit adjustment of the

variable θ_0 to those ranges that do not allow a zero of transmission at any pass band frequency.

The effects of network variables ranging in X and V in Figure 5.4.14 are well scaled. The L and C values range from approximately 0.1 to 10, as do their reactances at $\omega_0=1$ rad/s. The choices for transmission line lengths also result in ladder network branch reactances in the same range. Consider the T equivalent network for a lossless cascade transmission line in Figure 5.4.15. For $9 \leq \theta_0 \leq 171^\circ$, the series branch

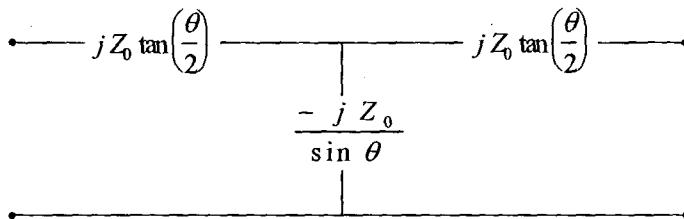


Figure 5.4.15. The T equivalent ladder network for a cascade transmission line.

reactances range from $Z_0/12.7$ to $Z_0 \times 12.7$, and the shunt branch reactance ranges from $Z_0/6.4$ to $Z_0 \times 6.4$ ohms at $\omega_0=1$ rad/s. Normalized Z_0 can range from 0.1 to 10, although a practical physical range is about 0.4 to 2.5. Transformer squared turns ratio, corresponding to impedance magnitude transformation, also can vary from 0.1 to 10.

Because the effects caused by multiple variables are reasonably scaled, it is useful to examine the minimax response surface defined by (5.4.4) and (5.4.5) on the diagonals of n-space hypercubes with edges $0.1 \leq v_i \leq 10$. To visualize how the minimax function behaves for more than the one and two variables already portrayed, consider a *hypercube* in the V space of N variables with each edge varying from 0.1 to 10 on a logarithmic scale. This is easily visualized in three dimensions as a cube [Wilde]. The interest is in the rays that connect vertices by passing through the center of the cube, i.e. the *principal diagonals*. There are exactly 2^{N-1} such diagonals when there are N variables. For example, when $N=8$, there are 128 principal diagonals through the hypercube. Figures 5.4.16, 5.4.17, and 5.4.18 show typical reflectance curves over hypercube diagonals for LC networks. The most important response is the envelope of worst-case reflectance versus *diagonal location factor*, i.e., the linear measure of travel along the diagonal.

The envelope functions in parameter cross sections are either unimodal or monotonic (*UM functions*) except where transmission lines and trap elements are involved. In the latter two cases, there may be simple aberrations in the UM envelope function that are avoided during a grid search. In one-dimensional diagonal subspaces, there are only minor aberrations in the envelope function, which again present no serious problems for the grid search described in Sections 5.5.2 and 6.3.

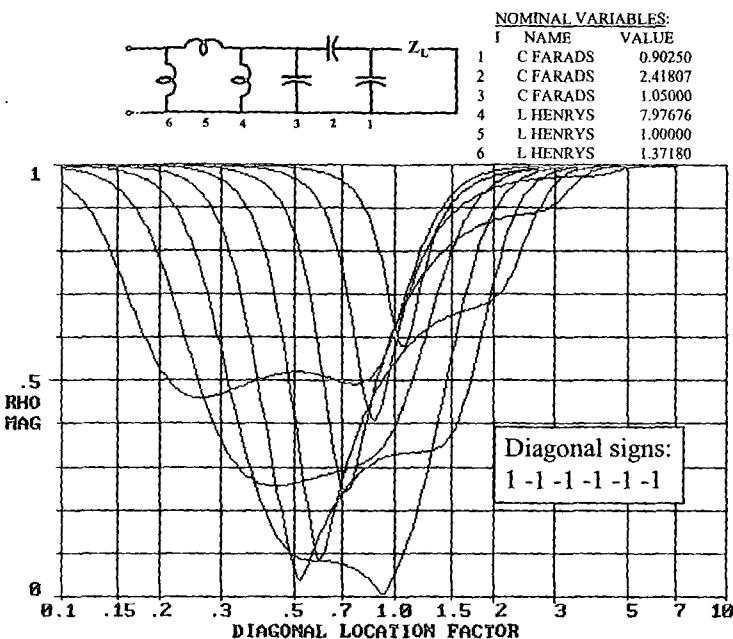


Figure 5.4.16. Reflectances over a diagonal for bandpass topology PiCPiL.

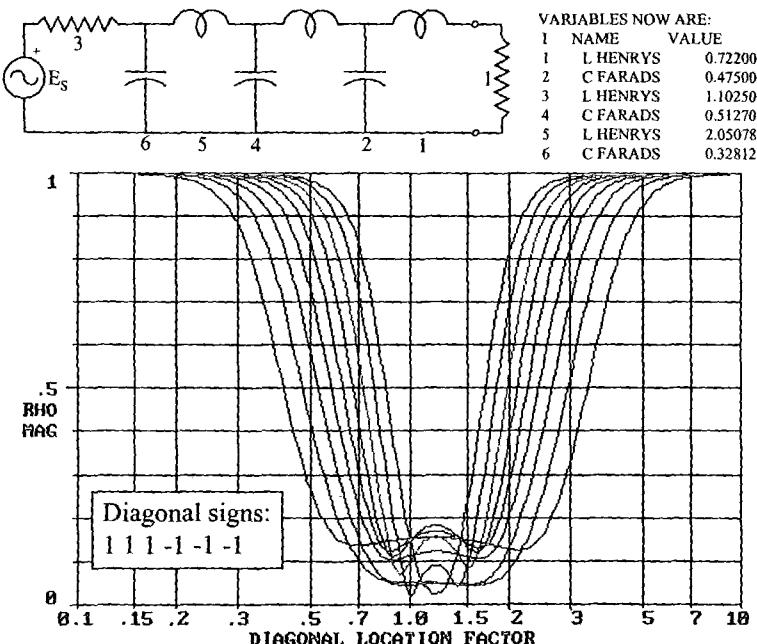


Figure 5.4.17. Reflectances over a diagonal for lowpass topology LPS6.

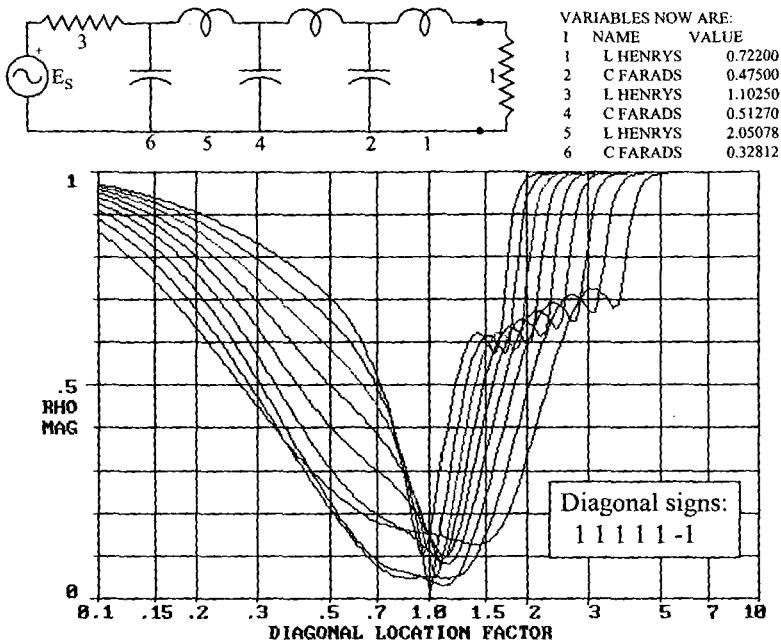


Figure 5.4.18. Reflectances over another diagonal for lowpass topology LPS6.

Examination of cross sections and diagonals for many realistic matching data sets and network topologies has shown it reasonable to expect the reflectance envelope minimum found to be global, i.e. the likely best-possible outcome with a candidate reduced-degree, full-rank topology.

5.5 Algorithms for GRABIM

There are three essential algorithms for GRABIM, the grid approach to broadband impedance matching: network analysis, grid search, and constrained optimization. Computational efficiency is mandatory, because there are several network topologies of varying complexity to be tested, and each is subjected to both direct and gradient search procedures. This section describes the general approach for each algorithm prior to providing specific examples. Details of the algorithms are provided in Chapter 6.

5.5.1 Efficient Network Selection and Analysis

The possible choices for elements in ladder networks are summarized in Table 5.5.1. The ideal transformer is included for lowpass networks, where the actual source resistance must be variable. In those cases, the source resistance can be fixed, and t^2 is varied. See Section 4.5.3.

Table 5.5.1. Twelve Lossless Network Elements And Parameters.

TYPE	CODE	NAME	VARIABLE(S)	CONSTANT
1	CS	Series Capacitor	C	
2	LP	Parallel Inductor	L	
3	LS	Series Inductor	L	
4	CP	Parallel Capacitor	C	
5	LCS	Parallel LC in Series	L	ω_∞
6	LCP	Series LC in Parallel	C	ω_∞
7	XFMR	Ideal Transformer	t^2	
8	CASTL	Cascade Transmission Line	Z_0, θ_0	ω_0
9	SCS	Short-Circuit Stub in Series	Z_0, θ_0	ω_0
10	SCP	Short-Circuit Stub in Parallel	Z_0, θ_0	ω_0
11	OCS	Open-Circuit Stub in Series	Z_0, θ_0	ω_0
12	OCP	Open-Circuit Stub in Parallel	Z_0, θ_0	ω_0

Although provisions for description of arbitrary connection of the elements in Table 5.5.1 is made available to a user (a topology code), it is often convenient to pre-program various standard topologies. For example, one such scheme [Cuthbert,1994b] presented the user with 12 different LC topologies, most providing a choice of from two to eight elements. Figure 5.5.1 shows such a choice, where BPP means "BandPass Parallel" because the network element next to the load impedance is in parallel. BPP3 would indicate that the user specified

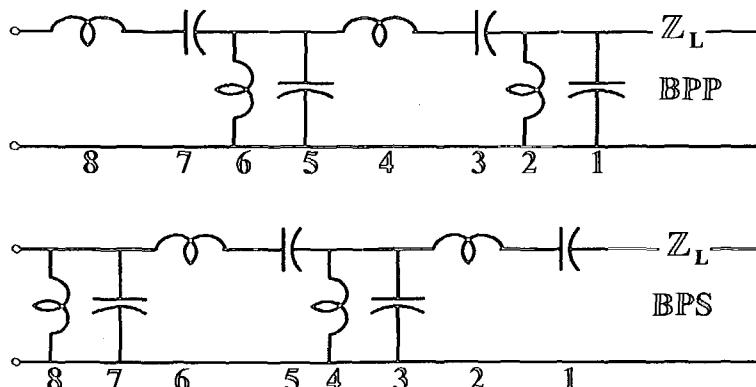


Figure 5.5.1. Two canonic bandpass topologies based on load connection.

just three elements in the matching network, starting from the load toward the source. A fixed number of elements in a useful bandpass topology is shown in Figure 5.5.2, where the impedance transformation capabilities of Norton transformers implicit in the Pi networks may be

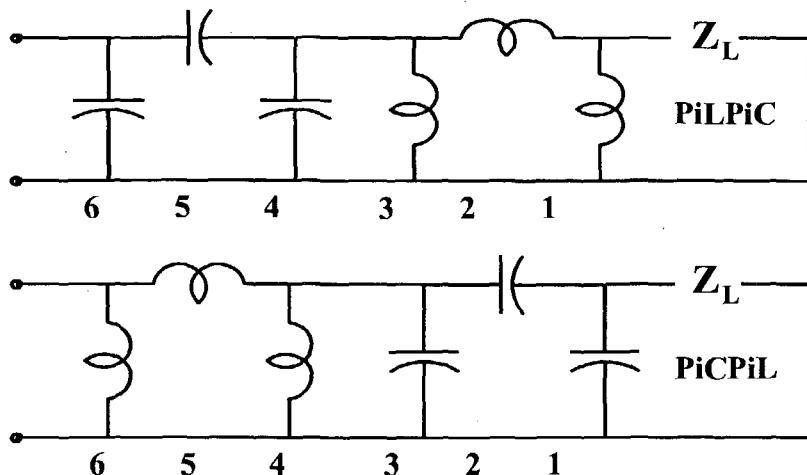


Figure 5.5.2. Two double Norton non-canonic bandpass topologies.

advantageous (Section 2.4.4 and Figure 2.4.11). These would be appropriate where the load is capacitive and the source is inductive, or vice versa. Also, note the Pi-T conversion in Figure 2.4.9.

A similar pre-programmed topological aid for distributed networks is provided. For example, the automatic arrangement of stubs and cascaded transmission lines shown in Figure 5.4.14 assumes that stubs are always separated by CASTL elements, so that the user can be presented with a sequence of interdependent choices:

- Is the element next to the load a stub?
- Lines only, or both lines and stubs?
- Total number of lines and stubs?
- Open- or short-circuit stubs or both considered?

In many cases a mixed distributed/lumped topology is desirable to create a band-limited response function. In those and other cases, the user specifies the desired sequence of element types, as listed in Table 5.5.1.

Once a candidate topology is defined and values are assigned to the elements, its overall ABCD parameters must be obtained at every sampled frequency to calculate the transducer gain T in (5.4.3) for the single or multiple sets of source and load terminations for the respective frequencies. The ABCD parameters for the four kinds of two-port subnetworks are shown in Figure 5.4.6. In addition to the reactances or susceptances of L's and C's, a stub in series or in parallel also has its input reactance or susceptance applied as X or B, respectively. From (2.6.10), it is observed that the input impedances of stubs are

$$Z_{inOC} = -jZ_0 \cot\theta, \quad Z_{inSC} = jZ_0 \tan\theta. \quad (5.5.1)$$

The reactance or susceptance for traps, applied as X or B for parallel LC in series or series LC in parallel, respectively, is

$$X = v_j \frac{\omega_i}{1 - \frac{\omega_i^2}{\omega_\infty^2}}, \quad LC = \frac{1}{\omega_\infty^2}, \quad (5.5.2)$$

where v_j is the variable L when X is used; otherwise, change X ohms to B mhos, and v_j is the variable C. The fixed stopband null frequency is ω_∞ , and ω_i is one of the m discrete frequencies in the passband sampling set.

It is noted in Section 2.6.2 that the overall network ABCD parameters are obtained by multiplying the individual element ABCD element matrices, starting from either the load or the source end. The numeric coprocessors in modern PC's perform any of the basic four operations (+, -, ×, ÷) in about 1 microsecond (μ S). With proper programming, lossless series or parallel element and transformer ABCD subsections can be assimilated with just four operations [Orchard], while the CASTL subsection requires eight operations. The final calculation for transducer gain T at one frequency by (5.4.3) requires 24 operations. Including about 50% overhead, the average analysis time per frequency is in the order of 100 μ S. The processing time not yet mentioned is for calculating the ABCD parameters at each frequency for the subsections. The equations can be as simple as $v_j \times \omega_i$, (5.5.2), or (5.4.11); some contain trigonometric terms, as in (5.5.1). Because the set of sampled frequencies is fixed, their contributions, as in (5.5.2), can be precomputed. It is shown in the next section that the grid search is combinatorial, meaning that sets of values for each variable are also known in advance. Therefore, all subnetwork ABCD parameters can be precomputed for the grid searches.

Incidentally, dissipative elements can be analyzed in twice the time required for lossless elements. However, allowing dissipation negates the basic advantage of lossless networks: The reflectance is invariant at every network interface (Figure 2.1.6). Also, the unimodal and monotonic (UM) envelope function behavior may be at risk. Reasonable dissipation and parasitic distortion are generally manageable by subsequent optimization of the lossless design, a final design step that is required in most situations where details of element geometry must be incorporated, e.g. bends in microstrip transmission lines and accurate models of surface-mounted components.

A speed consideration for personal computers is the fact that executable binary programs (.EXE) compiled from Microsoft Quick-BASIC® operate about four times faster in DOS "safe mode" than when the PC is configured for Windows®. Therefore, for fastest operation, those programs should be run in DOS only, using minimal AUTOEXEC.BAT and CONFIG.SYS files that do not invoke high-memory management and other interrupt features required to run Windows®.

5.5.2 Grid Searches

A grid search tests combinations of element values in a candidate network to obtain an approximate solution of the minimax matching objective stated in (5.4.4) and (5.4.5). It is basically a search in coordinate directions, Figure 5.4.14, using a fixed set of discrete values for each element. The best combination in the grid is found, then the grid is recentered there in each coordinate direction. Grid evaluation and repositioning are continued until no further improvement in minimax reflectance can be found in that grid. The spacing between grid points (granularity) is then reduced, and the evaluation and repositioning process is repeated. Usually, only 2-3 repositions are required for each grid, and only two reductions in granularity are required to arrive in the neighborhood of the likely global minimum. The benign nature of the envelope functions, recent research results in the pattern-search optimization field, and efficient implementation guarantee grid search effectiveness, as discussed here and explained in detail in Chapter 6.

Consider the grid defined for the two-element network in Figure

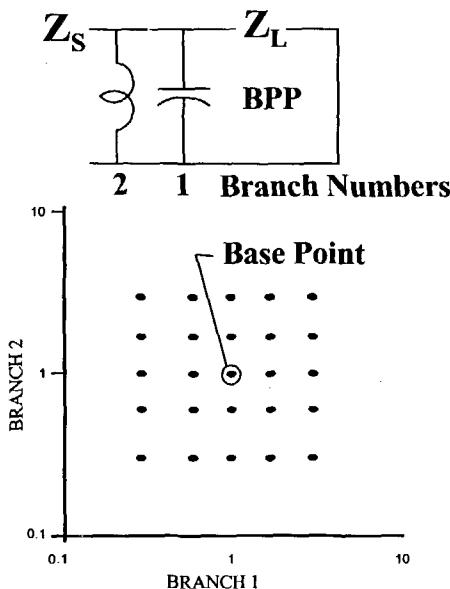


Figure 5.5.3. A 5×5 lattice (grid) for a two-branch matching network.

5.5.3. Five discrete values logarithmically centered about unity (the best guess available) are selected for each branch's L and C value. Each of the 25 combinations of the element value set {1, 0.6, 1.7, 0.3, 3} for each branch is tested at three passband frequency samples. Beginning with L=1 and C=1, compute $|\rho|$ using (5.4.3) and (5.4.1) for each frequency, and store the greatest (worst) value of $|\rho|$ at any frequency with the corresponding L and C values as the "best case". Then try L=1 and

$C=0.6$; if the greatest $|\rho|$ versus frequency is less than the stored case, then store that and the branch values as the best case. A frequency scan is abandoned whenever $|\rho|$ at any frequency exceeds the stored case, because that LC combination has thus failed. Next try $L=1$ and $C=1.7$, etc., until all five branch 1 values have been tried. Then set $L=0.6$ in branch 2, and go through the branch 1 set again. One of the 25 points in Figure 5.5.3 is the best case (minimax) for this iteration. Figure 5.5.4 illustrates a grid search using nested programming DO loops for as many as eight branches; the five-branch case is indicated. The “ ∞ dB” notation

The number of combinations for NT trial values in NB branches per ω :

NB	NT=3	5	7	9	11
2	9	25	49	81	121
3	27	125	343	729	1,331
4	81	625	2,401	6,561	14,641
5	243	3,125	16,807	59,049	161,051
6	729	15,625	117,649	531,441	1,771,561
7	2,187	78,125	823,543	4,782,969	19,487,171
8	6,561	390,625	5,674,801	43,046,721	214,358,881

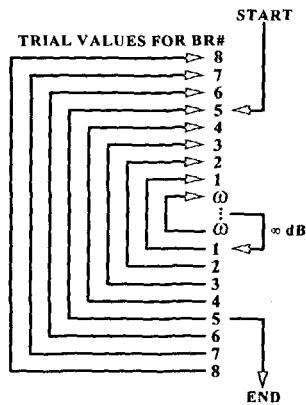


Figure 5.5.4. Programmed DO loops performing a simple grid search.

in Figure 5.5.4 indicates a premature failure that allows an early escape from the frequency loop. For each frequency there could be yet another loop, not shown in Figure 5.5.4, to process multiple impedances that define an impedance neighborhood.

In two-space as shown in Figure 5.4.13, the first grid pattern in Figure 5.5.3 locates the best result in the grid. Next, the grid pattern is repositioned with the center *grid base point*, Figure 5.5.3, repositioned to the stored best result. The branch value combinations (as multiplicative factors) are again repeated, without duplicating prior grid points, to find the best point in the repositioned grid. The grid is repositioned in sequence until the best grid point does not change; that usually requires about three base-point moves. Then the spacing between grid points (*granularity*) and perhaps the number of grid points in each variable are reduced, and a new set of iterations accomplished. Because the granularity is reduced twice by a factor of 4, the approximate location (neighborhood) of the global minimum is suitably obtained.

Grid iterations approximately locate the global minimum of an envelope reflectance function while avoiding minor $|\rho|$ surface aberrations. For example, Figure 5.5.5 shows a cross section for CASTL length θ_0 , where the lower abscissa is in logarithmic (V) space and the

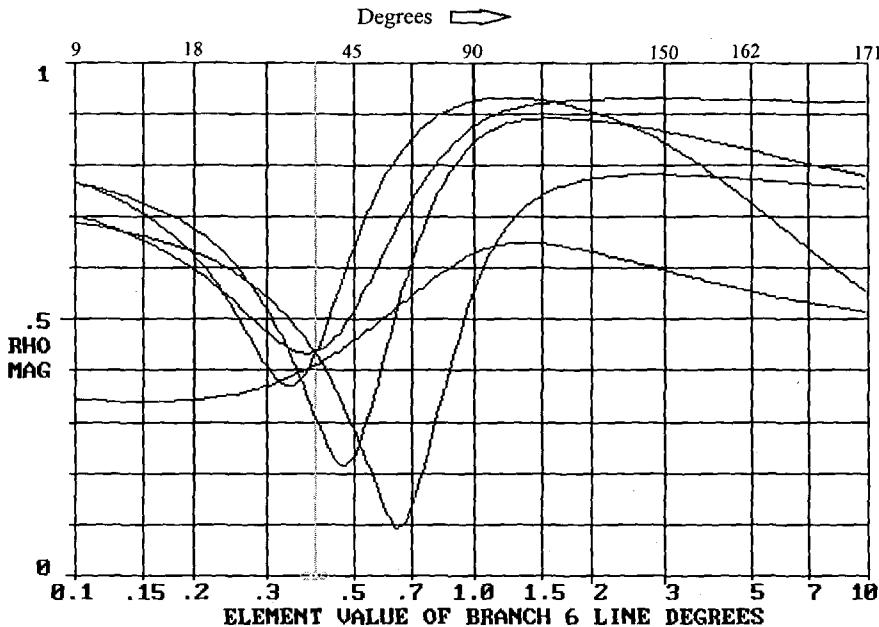


Figure 5.5.5. Reflectances versus CASTL electrical length (θ_0 at 1 rad/s).

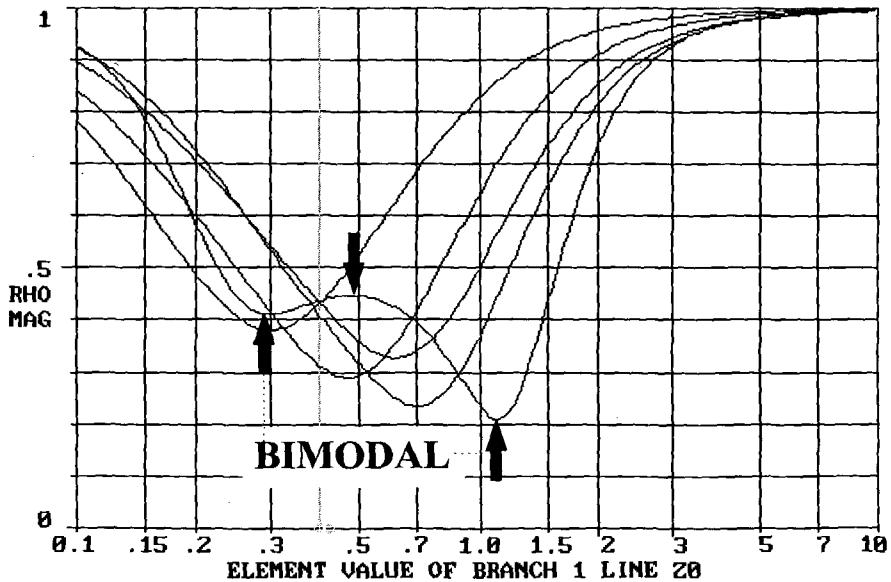


Figure 5.5.6. Reflectances versus CASTL impedances Z_0 .

upper abscissa is in the corresponding θ_L^o space (Figure 5.4.14 and (6.2.13)). There is an inferior minimum in Figure 5.5.5 at $v \approx 1.6$,

corresponding to $\theta_0 \approx 124 = 90(2 - 1.6^{-1})^\circ$, which the grid search avoided. Figure 5.5.6 shows the cross sections for a cascade transmission-line impedance, where the bimodal reflectance function at one frequency (two up arrows) did not affect the grid search result on the envelope. Figure 5.5.7 shows the cross sections for the electrical length of a short-circuit transmission-line stub in parallel (SCP), where the short circuits at two frequencies were avoided by the grid search; see 150° and 163° on the upper scale.

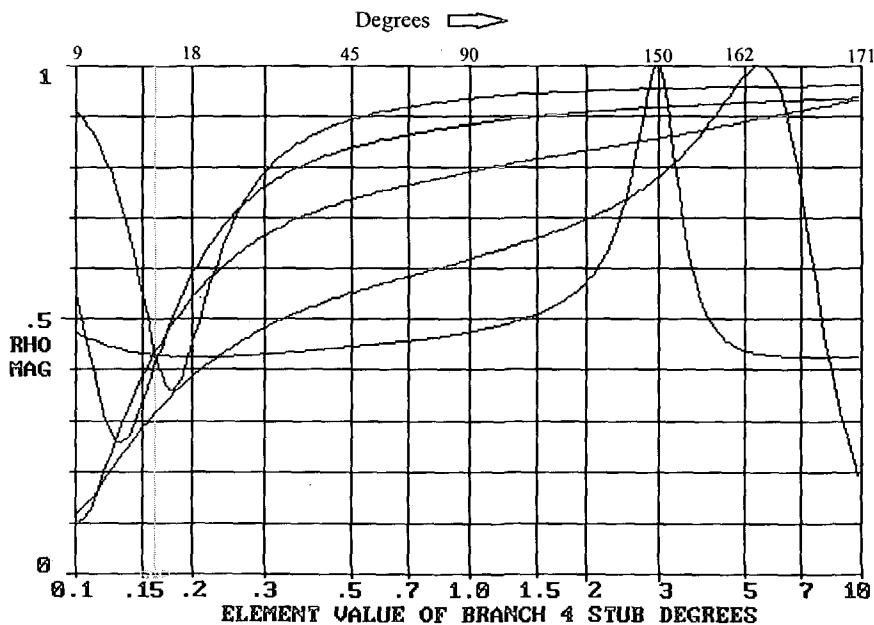


Figure 5.5.7. Reflectances versus SCP stub electrical length (θ_0 at 1 rad/s).

Early investigation explored as many as 11 trial grid points in each coordinate direction [Cuthbert, 1994b]. However, only three or, better yet, two trial points are required in each variable, as shown in Figure 5.5.8. The motivation is seen in the approximate time to complete

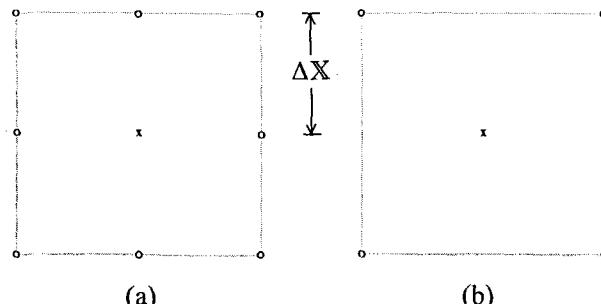


Figure 5.5.8. The two- and three-point grid patterns in two-space.

a full pattern for an LC ladder network:

$$T_s \approx 1.45 \times OT \times NS \times (24 + 7.7 \times NB) \times NT^{NB} \quad \text{Seconds}, \quad (5.5.3)$$

where OT is four-function ($+, -, \times, \div$) time, NS is the number of sampled passband frequencies, NT is the number of grid trial values, and NB is the number of branches. Factor 1.45 has been found to account for general programming overhead. There are 24 operations required to calculate transducer loss, T in (5.4.3), and the 7.7 factor includes the basic four operations to process a series or parallel branch and the related topological overhead.

Example 5.5.1. Suppose that a grid search is to be performed using a PC having four-function computing time of 1 microsecond. Consider a large LC network having ten branches that is sampled at 21 passband frequencies. *Problem:* Compare the computing times for a full two-point and three-point grid. *Solution:* In (5.5.3), $OT=1E-6$, $NS=21$, and $NB=10$, an upper limit on possible broadband matching network complexity. All but the number of combinations term (NT^{NB}) in (5.5.3) comes out to be about 3 milliseconds. The $NT=2$ hypercube in 10 space requires only 1024 combinations, while the $NT=3$ trial pattern requires 59,049 combinations. Therefore, the full two- and three-point trial patterns require no more than 3 seconds and 3 minutes, respectively. Times for the two patterns can be compared by calculating the ratio $3^{NB}/2^{NB}=1.5^{NB}$; see Table 5.5.2. Actual times are about half as long, because full frequency scans are seldom required.

Table 5.5.2. Comparing $NT=3$ to $NT=2$ Grid Combinations vs NB .

No. Branches	2	3	4	5	6	7	8	9	10
$3^{NB}/2^{NB}$	2.25	3.38	5.06	7.59	11.4	17.9	25.6	38.4	57.7

When the base point of the grid patterns in Figure 5.5.8 is repositioned to a stored best case, the repositioned grid will duplicate some of the points previously evaluated. Although those duplications can be avoided to reduce the computing time predicted by (5.5.3), it is clear from Table 5.5.2 that a practical number of trials per branch must be $NT=2$. No limitation is imposed, because $NT=2$ grids can visit any points that are available to an $NT=3$ grid. The grid granularity in Figure 5.5.8 refers to the X space in Figure 5.4.14 and is started with $\Delta X = 0.4$, so that a sequence of about three grid relocations will find no better combination in the approximate range of $-1 \leq x_j \leq +1$. The wide starting granularity avoids any of the $|\rho|$ surface aberrations illustrated above. Subsequent reductions to $\Delta X = 0.1$ and $\Delta X = 0.025$ will locate the

neighborhood of the global minimum (element values in vector v) to within about 3% ($V=10^{0.025}=1.059$).

As detailed in Chapter Six, the element search range in V space spans the range $0.04 \leq v_i \leq 25$. Because $x_i = \text{Log}_{10}(v_i)$, the corresponding range in X space is about $-1.40 \leq x_i \leq +1.40$. See Figure 5.4.14. The minimum granularity of $\Delta X = 0.025$ means that there are potentially 113 discrete values that each element might have. The algorithm by no means visits all the possible combinations. Instead, it uses function values to prune the number of grid points that are considered. The grid search is a scaled, translated integer lattice, and grid points are the lattice vertices. For scaling factors of powers of 2, all possible points of the lattice are listed in a matrix template using integer arithmetic, which is fast, uses little storage, and has no round-off error. Then, any duplicate testing of vertices is avoided by simply sorting an integer list [Dennis]. Element values are bounded by eliminating the disallowed values from the list, and holding elements constant reduces the combinatorial scope. The grid search is nonspeculative, unlike several simplex algorithms, for example Nelder-Mead and Hooke-Jeeves, which extend in promising directions by changing the basic search pattern.

In light of the preceding discussion, an overall algorithm for the GRABIM grid search is shown in Table 5.5.3. Although the description is given in terms of the X space, it should be noted that the actual network element parameters are in the V space or its subsets, Figure 5.4.14, i.e. $v=10^x$ for v_i and x_i .

Table 5.5.3. The GRABIM Grid Search Algorithm Without Details.

1. Set base point $x^b = (1 \ 1 \ \dots \ 1)^T$ in space E^N ,
2. Set $\Delta X = 0.4$ and $|\rho|_{\max} = 1.0$,
3. Evaluate $|\rho_i^k|$ at each previously untested k^{th} point in the grid ($\leq 2^N$ points) at all ω_i , $i=1, 2, \dots, m$, to find $|\rho_i^k|_{\max}$ that occurs at combination x^k ,
4. If $|\rho_i^k|_{\max} < |\rho|_{\max}$: then $|\rho|_{\max} = |\rho_i^k|_{\max}$ and $x^b = x^k$ and go to step 3, else
5. If $|\rho_i^k|_{\max} \geq |\rho|_{\max} \forall k$, then $\Delta X = \Delta X / 4$,
6. If $\Delta X < 0.025$, stop, else go to step 3.

The GRABIM grid search converges unfailingly to a point where the function is non-differentiable (at an envelope “knot”) or the gradient vector (all coordinate slopes) is zero [Torczon, 1991:143]. Although the grid search reliably locates the neighborhood of the minimum, its final rate of convergence is extremely slow.

5.5.3 Constrained Optimization for Element Removal

A *reduced-degree, full-rank network* is defined as one where all elements contribute to a minimax matching solution. A precise solution

of the minimax problem for a candidate network is required so that noncontributing elements are removed, i.e. elements that have monotonic envelope functions. Indeed, if one or more full-rank network topologies were known for a given data set, many common optimizers might find global matching solutions. The challenge is to process a network potentially more complicated than necessary and eliminate those elements that do not contribute to a minimax solution.

To obtain rapid convergence to a precise minimax solution, it is necessary to employ an optimization strategy based on the gradient of an objective function. The *gradient* is the vector (set) of first partial derivatives of an objective function with respect to (wrt) each variable. The objective envelope function of $|\rho|$ versus each variable is composed of arc segments (Figure 5.4.2). The envelope function is continuous but nonsmooth, and the partial derivatives are discontinuous at the knots (joints) of the segments. However, the minimax problem can be reformulated using only smooth functions, so that a special constrained optimization problem can be solved by numerically well-behaved algorithms. Those constrained optimization algorithms are only valid when started from the neighborhood of a minimum, which is always available by the grid search. Ultimately, the constrained optimization problem is solved by a sequence of unconstrained minimizations. A general explanation here is followed by more details in Section 6.4.

Figure 5.4.14 shows that the network variables, v_j , must remain strictly positive, but that the relation $x_j = \log_{10}(v_j)$ transforms the problem to a domain where there is no restriction on the x_j variables. For that and other reasons, the minimax problem originally defined by (5.4.4) and (5.4.5) is reformulated:

$$\text{Minimize } x_{N+1} \text{ such that } \rho_i(x) \equiv |\rho(x, \omega_i, D_i)| \leq x_{N+1}, \quad i = 1 \text{ to } m. \quad (5.5.4)$$

As before, there are m sampled frequency points in (5.5.4), e.g. $m=8$ in Table 5.4.1. For each of the sampled frequencies, ω_i , there is at least one associated data subset, D_i , namely the goal, R_L , X_L , R_S , and X_S values for each frequency. Reflectance $|\rho|_{\max}$ is formally identified as an *added variable*, namely x_{N+1} . The column vector x is related to all N variables of the matching network, e.g., values of L , C , Z_0 , θ_0 , and t^2 , and to $|\rho|_{\max}$ by

$$x = (x_1, x_2, \dots, x_N, x_{N+1})^T, \quad \text{where } x_j \equiv \log_{10}(v_j), \quad j = 1 \text{ to } N. \quad (5.5.5)$$

Section 6.4.4 describes one of several gradient-based optimizers that efficiently and accurately solve unconstrained problems, i.e., minimize some $F(x)$ using the first partial derivatives of F wrt each x_j . This second stage of GRABIM repeatedly solves a particular optimization problem:

$$\text{Minimize } F(x) \equiv x_{N+1} + \sum_{i=1}^m s_i \max \left\{ \left[\rho_i - (g_i + x_{N+1} - u_i) \right], 0 \right\}^2. \quad (5.5.6)$$

Minimizing objective function $F(x)$ minimizes $x_{N+1} = |\rho|_{\max}$ and a summation of positive terms. The "max" prefix denotes a choice of the strictly positive quantity within the {} brackets, either zero or $[\rho_i - (g_i + x_{N+1} - u_i)]$ if the latter is positive. The desired value of reflectance ρ_i at frequency ω_i is *goal*³ g_i , often zero but generally any value in the range $0 \leq g_i < 1$. The *float* to all goals is x_{N+1} , and the *offset* to each goal g_i at ω_i is u_i . The *weight* is s_i , a multiplier to add emphasis to $F(x)$ at frequency sample ω_i . There are two nested optimization loops; an outer loop adjusts s_i and u_i , and then an inner loop adjusts vector x to minimize (5.5.6). The following describes in general terms how and why those repeated optimizations of (5.5.6) determine vector values of $s=[s_i]$ and $u=[u_i]$ so that $F(x^*)=x_{N+1}$, where $x=x^*$ solves minimax problem (5.5.4).

This five-part explanation builds from a simple concept. First, suppose that $x_{N+1}=0$, $s_i=1$, and $u_i=0$ for all i in (5.5.6). Then (5.5.6) is reduced to the situation in Figure 5.5.9, corresponding to some choice of network element values in vector x . It is desired to vary x to satisfy the *inequality constraints*:

$$\rho_i \leq g_i, \quad i = 1, 2, \dots, m. \quad (5.5.7)$$

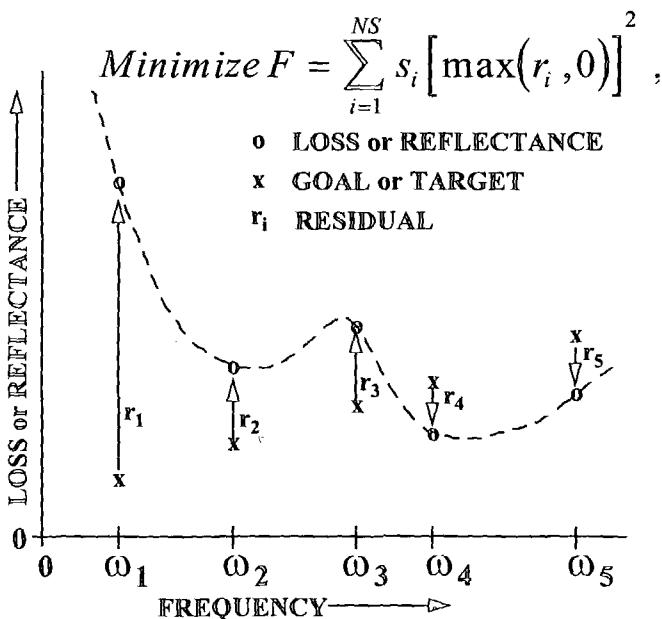


Figure 5.5.9. Minimizing binding residuals by an external (quadratic) penalty.

The *residual* error, $\rho_i - g_i$, at any frequency is just the difference between the reflectance and its goal, and it has a direction, depending on whether

³ All usage of g_i below represents a goal value and not a lowpass prototype value as above.

the residual is unsatisfied (binding) or more than satisfied. The residuals at frequencies ω_1 , ω_2 , and ω_3 are binding (positive), and each contributes a squared-error term in the summation. The residuals at frequencies ω_4 and ω_5 are satisfied (negative) and do not contribute to the sum of squared errors. A minimizer of $F(x)$ tends to work hardest on the squared residual error at ω_1 , because it is the largest term in the summation. As the set of network variables in vector x is adjusted, one residual error may be less prone to be reduced than the others, in which case its respective weight, s_i , can be increased to force more emphasis where it is needed. Unfortunately, that weight may have to be increased to infinity to force that error to the value of one or more of the other residual errors. Doing that causes ill-conditioning and is a poor strategy. Similarly, unlimited increases of the exponent on the residuals in Figure 5.5.9 also results in ill-conditioning; see Section 6.4.5. In most cases, (5.5.7) cannot be obtained, and only the largest positive residuals can be equalized.

Second, a minimax solution demands more than equal maximum errors – it requires that those maximum errors be minimized. See Figure 5.5.10. An added variable, x_{N+1} , “floats” all the goals, and x_{N+1} can be

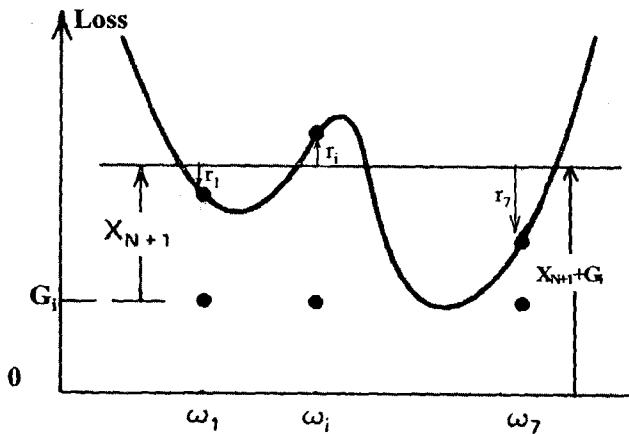


Figure 5.5.10. Goals are floated by an added variable, which is also minimized.

minimized to produce a minimax solution. The process of reducing all the goals and thus making equal maximum errors more difficult to obtain aggravates the problem of one or more weights, s_i , tending to infinity. However, that difficulty can be avoided by offsetting each goal by an amount u_i ; see (5.5.6). Suppose that the i th residual is less prone to be reduced as vector x is varied. Instead of increasing s_i without bound, u_i can be increased, which increases the size of its squared residual and its contribution to the error summation. Therefore, offsets u_i have the effect of requiring only moderate increases in weights s_i . This strategy is called

the *augmented Lagrangian* technique, which takes its name from Lagrange multipliers relating the weights and offsets.

Third, the key concept for solving minimax problem (5.5.4) is a theorem due to M. J. D. Powell [Powell, 1967:284]:

If the value of the variable vector x which minimizes $F(x)$ in (5.5.6) is $x^*(s, u^*)$, then x^* is a solution to the constrained problem

$$\text{Minimize } x_{N+1} \text{ such that } \rho_i \leq (g_i + x_{N+1}). \quad (5.5.8)$$

The weights in vector s must be suitably large but not infinite. The solution, x^* , is a function of only u , the vector of individual offsets, which must be adjusted so that $u=u^*$ makes the summation in (5.5.6) zero at x^* while x_{N+1} is minimized.

Fourth, it can be shown that, at the minimum $F(x^*)$ in (5.5.6), the weights and offsets are related by a set of constants called *Lagrange multipliers*, λ_i :

$$\lambda_i = s_i u_i, \quad i = 1, 2, \dots, m. \quad (5.5.9)$$

Originally, Lagrange multipliers were applied to equality constraints, i.e., where only an equality in (5.5.7) was considered. However, for binding inequality constraints, (5.5.9) applies; otherwise, $\lambda_i=0$ for satisfied constraints. Lagrange multipliers are sensitivities of an objective function wrt relaxation of the constraint constant. Suppose that goal g_i in (5.5.7) is increased by a small amount, say ϵ_i . Then, the Lagrange multiplier can be interpreted as

$$\lambda_i = \frac{\partial F(x^*)}{\partial \epsilon_i}, \quad (5.5.10)$$

for the binding i th constraint at the minimum point x^* . The Lagrange multiplier has been called the *shadow price* in economics, where λ_i shows how the objective function F will change for small changes in the constraint constant. In other words, tough, unyielding constraints have large Lagrange multipliers, whereas easily-satisfied constraints have much lower values.

Fifth, the fact that constraint weights and offsets, s_i and u_i , are related by a constant, as in (5.5.9), provides a way to adjust their values during a sequence of unconstrained minimizations of $F(x)$ in (5.5.6). An additional advantage is knowing when they should have a larger product, which is when their related residual errors are slow to reduce. A simplified algorithm for finding the solution of (5.5.4) is shown in Table 5.5.4. Powell's augmented Lagrangian technique converts the constrained problem in (5.5.4) to a sequence of unconstrained minimizations in (5.5.6). The details of step 5 in Table 5.5.4 are in Section 6.4.3; that simple algorithm for adjusting all s_i and u_i converges linearly. Step 3, for

the inner unconstrained minimization of $F(x)$ in (5.5.6), converges at a much faster (quadratic) rate, as explained in Section 6.4.4.

Table 5.5.4. The Augmented Lagrangian Algorithm Without Details.

See Equation (5.5.6).

1. For the initial x vector, set $x_{N+1} = \max\{p_i - g_i, i=1, 2, \dots, m\}$,
2. Set all u_i offsets to zero and all s_i weights equal to a value that makes the summation equal to $|x_{N+1}|$,
3. Minimize $F(x)$,
4. If the change of $|x_{N+1}|$ is less than some small criterion then stop, else
5. Adjust s_i and u_i according to the augmented Lagrangian strategy and go to step 3.

5.6 Examples Using GRABIM

Given frequency-sampled load and/or source impedance data normalized to one ohm and one rad/sec, each of the following examples is solved for one or more candidate matching networks. This multiplicity of candidates is a characteristic of other numerical matching methods [Dedieu:570] and is not a severe disadvantage, because of ease of use, rapid and reliable convergence, and the low order of broadband matching networks. Therefore, it is possible to test all feasible topologies systematically in a relatively short time. The special advantage of GRABIM is that it obtains the likely global solution for candidate network topologies that are pruned to full rank.

5.6.1 Example of a Non-Analytic Bandpass Problem

The data for this double-matching problem are obtained from source and load RLC models that could not conceivably have analytic solutions [Yarman,1990:220]. The source impedance is derived from a series resistance paralleled by a series LC branch that is resonant at 1.291 radians per second. Because the passband extends from 0.3 to 1.0 rad/s, the source impedance varies drastically, especially near the upper passband edge. The load impedance is derived from a parallel RL in series with another L, a situation where analytic theory is also unlikely to apply.

The normalized data are presented in Table 5.4.1, with eight passband frequencies at 0.1 rad/s intervals from 0.3 to 1.0 rad/s. The source resistance varies from 0.71 down to 0.04 ohms, and the source reactance is strictly capacitive. The load resistance varies from 0.6 to 0.95 ohms, and the load reactance is strictly inductive. The Norton transformation illustrated in Figure 2.4.11 suggests a way to deal with

an inductive load having resistance variation, and the dual Norton transformation similarly might equalize the capacitive source having drastic resistance variation. Therefore, the PiCPiL topology in the lower half of Figure 5.5.2 is tested by GRABIM.

Figure 5.6.1 shows the insertion loss after the grid search for the original six-element bandpass network. The result obtained by the

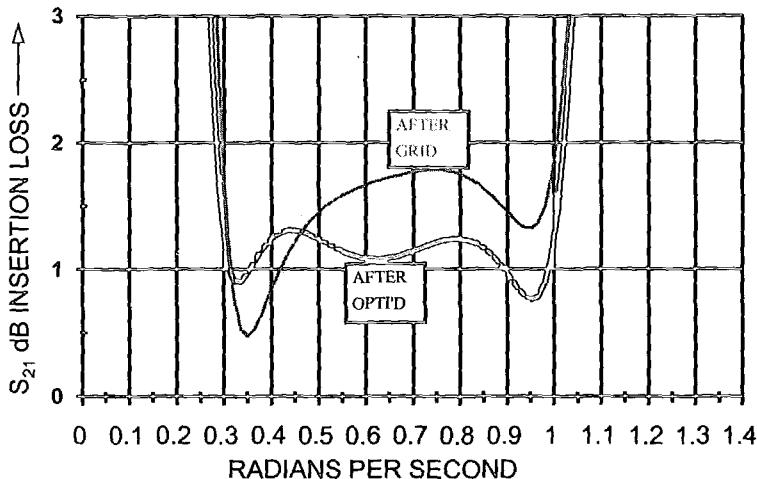
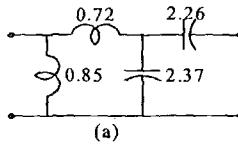


Figure 5.6.1. GRABIM results after each stage for Yarman example 3.

minimax-constrained optimization improves the insertion loss and also eliminates one capacitor and one inductor. The final topology and element values are shown in Figure 5.6.2(a). Other techniques also have

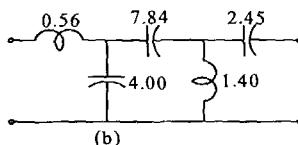
(a) Best case found:

$$0.97 < S_{21} \text{ dB} < 1.24$$



(b) Five branch bandpass

$$1.21 < S_{21} \text{ dB} < 1.63$$



(c) Five branch highpass

$$0.88 < S_{21} \text{ dB} < 1.75$$

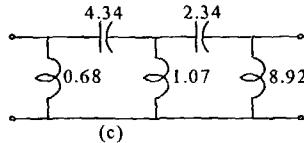


Figure 5.6.2. Three reduced-degree full-rank solutions for Section 5.6.1.

obtained the same solution [Sussman-Fort, 1991:Example 2]. Starting GRABIM with topology BPS6 (Figure 5.5.1) yields a similar solution,

except that a fifth element (a series inductor between the two C's) is slightly involved in the solution. Yarman's solution by the RFT required six elements (3 L's, 2 C's, and an ideal transformer) [Yarman,1990]. Figures 5.6.2 (b) and (c) show two other solutions obtained by GRABIM. Choosing a particular solution depends on physical and electrical requirements of the intended application. The main point is that several solutions can be obtained easily.

Several figures show aspects of this example after the complete optimization that started with the topology in the lower half of Figure 5.5.2. Figure 5.4.2 shows the cross sections for the branch 5 series inductor (unimodal envelope), and Figure 5.4.3 is a close-up of the five binding and three non-binding reflectances. Figure 5.4.12 shows the cross sections for the branch 1 parallel C, which is forced out of the network (monotonic envelope). Figure 5.4.13 shows the 3-D plot of the envelope surface over the subspace of C_2 versus C_3 . Finally, Figure 5.4.16 shows the reflectances along one particular hypercube principal diagonal for non-optimal values of the six network elements in the lower half of Figure 5.5.2.

5.6.2 Example of a Distributed Interstage Network

Field-effect transistor (FET) scattering parameters have been tabulated at 4.0, 4.5, 5.0, 5.5, and 6.0 GHz, so that an interstage matching network can be designed to transfer power between two of these active devices [Ha]. Program S11TOZ.EXE is used to convert the S_{22} data to normalized source impedances and the S_{11} data to normalized load impedances, as shown in Table 5.6.1. The same program is used to invert the source impedance to admittance.

Table 5.6.1. Tabulated Double-Match Data For Example 5.6.2.

Rad/s	g	R_L	X_L	R_S	X_S	G_S	B_S
.8	0	.2317099,	-.9217876	.8684981,	-.1220389	0.387	0.544
.9	0	.2313423,	-.7275047	.7550188,	-1.029207	0.463	0.632
1	0	.2284464,	-.5775817	.6481531,	-.914911	0.516	0.728
1.1	0	.2234532,	-.457425	.5405782,	-.8240196	0.557	0.848
1.2	0	.213391,	-.3397249	.4551464,	-.7313804	0.613	0.986

A simplified model of FET input and output immittances has series RC and parallel RC networks, respectively. The FET input R_L and X_L data in Table 5.6.1 do approximate a series RC network, as expected. However, the conductance for the FET output, G_S , is not at all constant. That anomaly does not affect the GRABIM technique.

The interstage network topology used by Ha is shown in Figure 5.6.3. There are six variables: Three Z_0 's and three line lengths at 1 rad/s. Each Z_0 is started at unity, as are the line-length arguments, which correspond to 90° , as shown in Figure 5.4.14. The grid search produces a trivial answer unless constraints are applied. Because

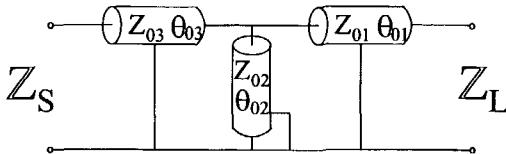


Figure 5.6.3. Transmission-line interstage matching network for Section 5.6.2.

physical Z_0 's occur between 20 and 125 ohms, the normalized Z_0 's are constrained: $0.4 \leq Z_0 \leq 2.5$. The grid search obtains a maximum $S_{21}=0.92$ dB across the passband, and minimax-constrained optimization reduces that to $S_{21}=0.53$ dB. The final element values in Figure 5.6.3 are $Z_{01}=2.5$ ohms (normalized), $\theta_{01}=6^\circ$, $Z_{02}=1.89$ ohms, $\theta_{02}=16.3^\circ$, $Z_{03}=1.75$ ohms and $\theta_{03}=21.6^\circ$.

5.6.3 Example of Neighborhood Matching

The source and load data in Table 5.6.1 for a FET are used to design an interstage network composed of lumped inductors arranged in a T topology. Furthermore, the FET S_{22} data often varies because of the device location on semiconductor wafers, as well as with high signal levels during operation in amplifiers. That uncertainty is illustrated in Figure 5.6.4 in the Smith chart (a circle of unit radius), where the circles of S_{22} data have a diameter of 0.07. The GRABIM technique is well

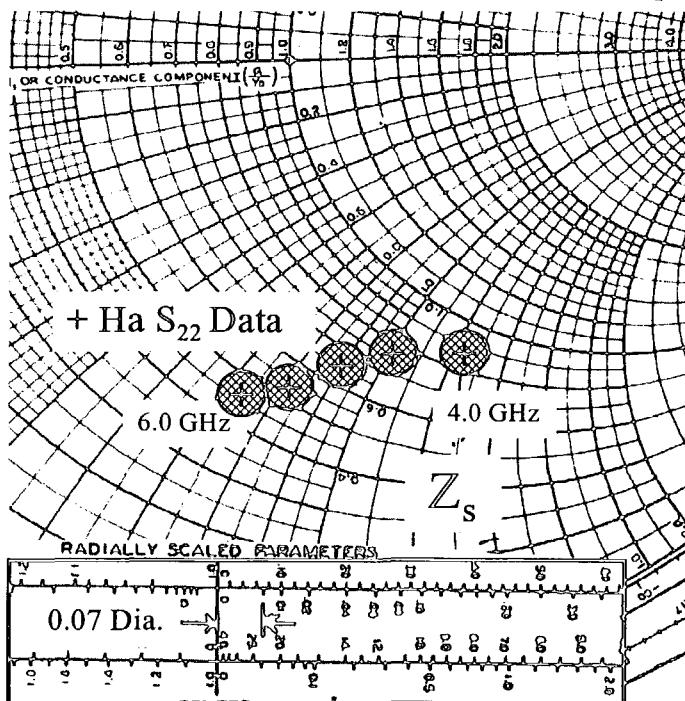


Figure 5.6.4. Scattering parameter uncertainty on a Smith chart.

suites to deal with multiple terminal data at each frequency, because the network analysis algorithm described in Sections 5.5.1 and 6.2.3 obtains the overall network ABCD matrix at each frequency and then applies (5.4.3) to find transducer gain T for each distinct terminal impedance. Only 24 basic operations are required for (5.4.3). The neighborhoods shown in Figure 5.6.4 are adequately modeled by five values at the vertices of pentagons, so only 24×5 basic operations are added to each frequency. This example assumes the impedance uncertainty and illustrates the effectiveness of compensating for that uncertainty.

Figure 5.6.5 shows a T subnetwork of lumped inductors followed by an augmenting el section, C_2 and L_1 . Table 5.6.2 summarizes performance of the original three-inductor design and the effect of uncertainty in S_{22} .

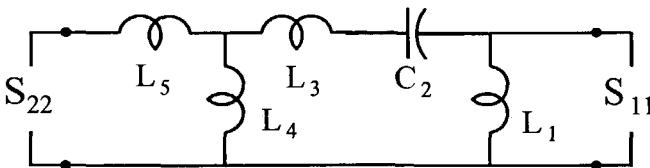


Figure 5.6.5. The T section is augmented by an el section for neighborhoods.

Table 5.6.2. Performance Before, After, And Compensated.

Performance of Network in Figure 5.6.5 Without Neighborhood S_{22} and C_2 & L_1 :

Design from	L_3 nH	L_4 nH	L_5 nH	4.0 GHz	4.5 GHz	5.0 GHz	5.5 GHz	6.0 GHz
Ha Figure. 5.25	0.42	0.94	1.05	0.57 dB	0.23 dB	0.44 dB	0.41 dB	0.75 dB
GRABIM	0.35	1.04	1.10	0.51 dB	0.43 dB	0.51 dB	0.34 dB	0.51 dB

Performance With Neighborhood S_{22} (Without C_2 & L_1) After Re-Optimization:

L_3 nH	L_4 nH	L_5 nH	4.0 GHz	4.5 GHz	5.0 GHz	5.5 GHz	6.0 GHz
0.33	0.78	0.60	1.17	1.19	1.18	0.96	1.19
		Max S_{21}					
		Min S_{21}	0.39	0.80	0.85	0.62	0.75

Performance of Full Network in Figure 5.6.5 With Neighborhood S_{22} and C_2 & L_1 :

L_1 nH	C_2 pF	L_3 nH	L_4 nH	L_5 nH
1.15	5.98	5.14	1.73	0.43
4.0 GHz	4.5 GHz	5.0 GHz	5.5 GHz	6.0 GHz
Max S_{21}	0.50	0.50	0.50	0.47
Min S_{21}	0.26	0.25	0.27	0.23

Using the original data without neighborhoods, the GRABIM result is comparable to the original design [Ha:Figure 5.25]; the maximum S_{21} is about 0.5 dB. When the neighborhoods are included, the original T network design causes 3.15 dB loss at one frequency. After re-optimizing the T network, that is reduced to no more than 1.19 dB, as shown in the

middle of Table 5.6.2. Augmenting the original T network with two more elements (an el section), as in Figure 5.6.5, provides a GRABIM solution having no more than 0.5 dB loss again as shown at the bottom of Table 5.6.2. That is a small price to pay for the severe loss penalty in not accounting for scattering parameter uncertainty. Note the fundamental relationship for sizes of matched neighborhoods in Appendix A, Section A.2.3 and Figure A.2.2.

5.6.4 Example of Topological Simplification and Sampling

Alternative solutions for an LC interstage network [Yarman,1982] have been published [Abrie,1991]. Data for the six linearly-spaced frequencies were tabulated from 8.0 to 13.0 GHz and are shown in Table 5.6.3, normalized to 50 ohms and 10.198 GHz. The normalizing frequency is the geometric mean of band edges 8 and 13 GHz, consistent with the

Table 5.6.3. Normalized Data for Yarman's Interstage Example.

Rad/s	g	R _L	X _L	R _S	X _S
0.7845	0.00	0.195	-0.889	2.170	-4.225
0.8825	0.00	0.193	-0.776	1.871	-3.944
0.9806	0.00	0.194	-0.697	1.740	-3.699
1.0786	0.00	0.193	-0.631	1.488	-3.487
1.1767	0.00	0.194	-0.577	1.270	-3.292
1.2748	0.00	0.194	-0.530	1.130	-3.125

candidate network used with GRABIM: BPP6, shown in the upper half of Figure 5.5.1. Four of the six elements vanish, leaving only L₂=0.863 and L₄=3.35 henrys and obtaining a maximum passband loss of 0.558 dB. This is a case starting with a marginal number of frequency samples for six variables (there should be at least N+1 and preferably 2N samples). The surviving two elements are obviously a minimal topology (rank 2). Candidate network BPP4 (Figure 5.5.1) has no sensitivity to the six samples and obtains the same result.

Yarman's solution required six elements and an ideal transformer, i.e. seven elements, and obtained a passband loss of 0.600 dB. Two solutions were obtained using Abrie's transformation-Q technique [Abrie,1991:149]: (1) three elements and 0.565 dB, and (2) five elements and 0.386 dB.

5.7 Summary of Matching Networks

Single-frequency impedance matching using el, T, and Pi sections to obtain conjugate match (zero reflectance) between real and complex terminating impedances is described. The roles of loaded Q in the 1+Q² method and transformation resistances are emphasized. The input impedance of a lossless (cascade) transmission line's dependence on its load impedance is the means for finding Z₀'s and θ's that provide zero

reflectance for given terminations. These parameters and concepts also play a role in broadband matching.

Fano's analytic gain-bandwidth limitation is described as the area under the return-loss curve, fixed strictly by the loaded Q of single RLC load resonators relative to the bandwidth measured by Q_{BW} . The main design parameter is the decrement: $\delta = Q_{BW}/Q_L$. Although not optimal, the Chebyshev equal-ripple function from filter theory is also applied to the broadband matching situation. There are two degrees of freedom available in this gain-bandwidth problem, and the one constraint of a single-match load allows minimization of the maximum passband reflectance. Adding a given source resonator consumes the second degree of freedom, so that the maximum passband reflectance is predetermined. The same is true for the singly-terminated case, for the same reason. In each case, it makes sense to consider the improvement available as the number of matching elements increases without bound; only in the single-match case can the reflectance approach zero. The recursive equation for matching network element values is provided, which applies to all cases including filters. References are provided for measuring the loaded Q of single resonators for those few situations where analytic gain-bandwidth theory applies.

The real-frequency technique (RFT) originated by Carlin is briefly described, especially its dependence on advanced mathematical considerations and a sequence of less-than-certain optimizations. The RFT was the first advanced technique to consider the practical situation where one or both matching network terminations are characterized only by a tabulation of impedances at a discrete set of frequencies. A piecewise-linear representation of the matching-network Thevenin resistance function over all real frequencies is a sum of weighted basis functions at the load port. Because the imaginary component can be determined using the Hilbert transform, that Thevenin impedance and the load impedance enable calculation of the reflectance at each of the sampled frequencies. The Thevenin resistance function has adjustable parameters to optimize the reflectance variation versus frequency. A rational, even resistance function of assumed degree is then fit to the Thevenin piecewise-linear resistance function by a second optimization. Finally, an LC ladder network can be synthesized from the latter function for a resistive source impedance, i.e. the single-match case. Solution of a set of linear equations is required for that synthesis.

The double-match case, where the source impedance is complex as well as the load impedance, is accommodated in the RFT by considering a transducer gain expression, T, involving the source reflectance wrt to 1 ohm and the transfer phase angle, ϕ , available from the rational resistance function. A third optimization that includes repeated solution of linear equations is required, because T involves ϕ .

An improved broadband matching technique employs Brune functions that represent the Thevenin impedance at the load port in a form similar to a partial fraction expansion. Nearly two-thirds of the original RFT procedure still is required to provide approximate locations of the impedance-function pole locations, which are then optimized over each $s=\sigma+j\omega$ pair of variables to obtain optimal T. The Brune impedance representation of chosen degree guarantees that all network elements subsequently synthesized will have positive values. Double-matching for active devices has also been formulated, again using optimization variables in the Laplace s plane, with enhancements that include device stability and terminal reflectance added concurrently in the optimization. None of the above techniques feature optimization with certain outcome.

Discrete sets of tabulated terminal impedance data also initiate the grid approach to broadband matching (GRABIM), a two-step optimization procedure that is robust and avoids most of the mathematics required for the RFT. Reflectances are reflection-coefficient magnitudes, $|\rho(\omega_i)|$, at each sampled frequency, ω_i . The benign behavior of reflectances versus any of the five types of network element parameters (L , C , Z_0 , θ_0 , and t^2) suggests that a very effective matching technique is to perform a minimax-constrained optimization on one or more candidate network topologies. The augmented Lagrange multiplier technique is a gradient-based algorithm that reliably and precisely solves inequality-constrained optimization problems when started from the neighborhood of a solution. It can also be modified to minimize the one or more maximum reflectances over frequency (minimax), and it removes network elements that are not contributing to the solution. This last feature is the primary advantage of GRABIM, because if a minimal network topology were known, then many ordinary optimizers might find the global solution.

The essential first step in GRABIM is to locate the neighborhood of the likely global minimax solution by a non-speculative grid search made feasible by efficient algorithms executed at the high speed of current personal computers. Worst-case reflectance envelopes versus parameters L , C , Z_0 , θ_0 , or t^2 are mostly unimodal or monotonic, arc-wise continuous, nonsmooth functions. The grid search evaluates the minimax function values at all vertices of a hypercube on the parameter coordinates, and the hypercube is then recentered on the least value found. That sequence terminates in two or three moves. The initial hypercube size (spacing between coordinate samples) spans about 40% of likely solution space, to reduce the function value while avoiding occasional minor local minima. That size is reduced twice by a factor of four after hypercube moves terminate, locating the likely global minimum in parameter space to within about 3%. The grid search definitely avoids the minor aberrations in the function surface that sometimes occur in the likely solution space.

Four examples of GRABIM are provided, with comparisons to RFT solutions. One example describes another advantage of GRABIM: It efficiently solves the problem of terminating impedance uncertainty that often occurs in steerable antenna arrays, mobile antennas, and transistors. In that example, it is shown that a small addition to network complexity can guarantee the same matching performance for 3.5% impedance tolerance compared to the original network with no load or source impedance variation. The fundamental relationship for the size of matched impedance neighborhoods is noted (Appendix A, Section A.2.3).

6. GRABIM in Detail

This chapter provides a concise description of the equations and algorithms that implement the grid approach to broadband impedance matching. Some of these algorithms also apply to similar tasks, e.g. to Orchard's iterated network analysis. In most cases, the several distinct procedures have been thoroughly documented in the literature. They are brought together and referenced here to implement the GRABIM technique, which finds the likely global solution for a reduced-degree full-rank network topology without the complications of polynomial synthesis. These procedures need to be programmed only once, not for each and every different problem. Alternative constrained optimization methods are also summarized.

6.1 Formulation

The broadband impedance matching problem using tabulated (real-frequency) impedance data belongs to the well-known category of *nonsmooth optimization*. "A function is smooth if it is differentiable and the derivatives are continuous" [Rockafellar,1994], [Polak]. A nonsmooth function lacks some of the properties usually relied upon in analysis, and that is the case with the reflectance envelope functions described in Section 5.4.4, where the first partial derivatives are not continuous. Both general and specific problem formulations are described.

6.1.1 The General Problem

One of the most common situations concerning envelope functions is that of minimizing a function $f(x)$ having the form

$$f(x) = \max_{i \in I} \rho_i(x) \quad \text{for } x \in R^n, \quad (6.1.1)$$

where \in is "belongs to", and I is some infinite or finite set. In the broadband matching problem, I is a set of frequencies, ρ_i is the reflectance (magnitude of a generalized reflection coefficient), and n is the number of matching-network element values related to the variables in vector x . The name *semi-infinite programming (SIP)* has been applied to (6.1.1), because x represents a finite number of variables, while I can be an infinite set. That subject has been surveyed [Hettich], and various specific solution techniques have been proposed [Charalambous], [Zhou], [Rustem], based on I being a finite, discretized set. Considerations for selecting the sample subset of I (frequency) are discussed in Section 6.4.6.

6.1.2 The General Solution

The usual approach to minimizing (6.1.1) is to define an added variable, x_{n+1} , and then to solve the problem

$$\min x_{n+1} \text{ s.t. } \max_{i \in I} (\rho_i(x) - x_{n+1}) \leq 0, \quad (6.1.2)$$

where "s.t." is "subject to." This shifts the effect of nondifferentiability of (6.1.1) into a set of inequality constraints. Where constraints are involved, Lagrange multipliers are not far behind [Rockafellar, 1993]. One Lagrange multiplier method for solving (6.1.2) is *Sequential Quadratic Programming (SQP)*, which linearizes the active (binding) constraints [Gill]. Also, there is often emphasis on heavily discretizing $i \in I$ [Zhou], which is not overly crucial to the broadband impedance matching problem. See Section 6.4.5.

Other methods for solving (6.1.2) include *barrier and penalty functions*, which add terms to the objective (x_{n+1} here) which vanish only at the constrained optimum. Current interior barrier methods add logarithmic terms, so that all values of x during minimization must remain feasible, i.e. all constraints always must remain satisfied [Polyak]. Conversely, exterior penalty functions add terms to the objective function that always have some constraints violated during minimization [Gill]. GRABIM has been implemented using the latter, quadratic penalty functions.

6.1.3 The Specific Problem

The complex transducer function for a lossless network terminated by complex load and source impedances, Z_L and Z_S , is [Frickey]

$$H = \frac{AZ_L + jB + jCZ_S Z_S + DZ_S}{2\sqrt{R_L R_S}}, \quad (6.1.3)$$

where each of the four ABCD parameters is real (B and C are real coefficients of imaginary numbers). Consistent with Sections 5.4.2 and 5.4.3, the *transducer power function* is

$$P \equiv |H|^2 = \frac{P_{as}}{P_L} \geq 1. \quad (6.1.4)$$

The transducer power function P_i at the i th frequency sample is monotonically related to its reflectance, $\rho_i = |\rho(\omega_i)|$; see (5.4.1). Although the reflectance is theoretically significant and scaled for easy display, the transducer function is used here as an optimization objective, because it varies from unity to infinity and is easier to manipulate. The related power loss function is

$$L = 10 \log_{10} P \text{ dB}. \quad (6.1.5)$$

Therefore, the specific GRABIM problem comparable to (6.1.2) is

$$\min x_{n+1} \text{ s.t. } \max_{i \in I} [(P_i - g_i) - x_{n+1}] \leq 0, \quad (6.1.6)$$

where goal or target loss, g_i , has been inserted for every sample:

$$g_i = 10^{\frac{L_i}{10}}, \quad L_i \text{ in dB at } \omega_i. \quad (6.1.7)$$

The numerical method for solving this problem is detailed in Section 6.4. It is noted in Section 5.5.3 that the neighborhood of the solution must be found in order to apply a Lagrange multiplier technique using partial

derivatives. The requisite grid search for that purpose does not require derivatives, but efficient computation of the transducer power function is essential. The next section provides those details.

6.2 Network Analysis

Computation of the transducer power function and its derivatives in terms of the matching network's overall ABCD parameters is described. There follows a concise listing of the ABCD parameters and their derivatives with respect to the variables in vector x for each constituent network element type. Finally, the algorithm for obtaining the overall network's transducer power function and its derivatives with respect to x at a frequency is detailed.

6.2.1 Transducer Function and Its Derivatives

The transducer power function is given in (5.4.3); for completeness, the inverted transducer power function is

$$P \equiv \frac{P_{as}}{P_L} = \frac{(AR_L - Cq + DR_s)^2 + (B + Cp + DX_s + AX_L)^2}{4R_s R_L}, \quad (6.2.1)$$

where frequency-dependent constants are

$$q \equiv (X_s R_L + X_L R_s), \quad p \equiv (R_s R_L - X_s X_L). \quad (6.2.2)$$

Partial derivatives are indicated by shorthand operator notation

$$\Lambda_y \equiv \frac{\partial}{\partial y}. \quad (6.2.3)$$

For example, $\Lambda_A P$ is the first partial derivative of P with respect to (wrt) chain parameter A . It is anticipated from (6.1.6) that such partial derivatives will be required for later conversion by the chain rule to be wrt the element parameters, namely L , C , t^2 , Z_0 , and θ_0 , and subsequently to the variables in vector x . The derivatives, ΛP , are available directly from (6.2.1), but it is easier to obtain them from (6.1.3) using the identity [Cuthbert,1983:102]

$$\Lambda_y |H|^2 \equiv 2\text{Re}\left[H^*(\Lambda_y H)\right], \quad (6.2.4).$$

where Re means "the real part of" and the $*$ superscript denotes complex conjugation. Conjugating (6.1.3) yields

$$H^* = (AZ_L^* - jB - jCZ_s^* Z_L^* + DZ_s^*) / 2\sqrt{R_s R_L}. \quad (6.2.5)$$

The partial derivatives of (6.1.3) wrt A , B , C , and D are

$$\begin{aligned} \Lambda_A H &= Z_L / (2\sqrt{R_s R_L}), \quad \Lambda_B H = j / (2\sqrt{R_s R_L}), \\ \Lambda_C H &= jZ_s Z_L / (2\sqrt{R_s R_L}), \quad \Lambda_D H = Z_s / (2\sqrt{R_s R_L}). \end{aligned} \quad (6.2.6)$$

Therefore, application of (6.2.4) yields the derivatives of transducer power function, P, wrt the overall matching network's A, B, C, and D:

$$\begin{aligned}\Lambda_A P &= \left(|Z_L|^2 A + X_L B - X_S |Z_L|^2 C + rD \right) / (2R_S R_L), \\ \Lambda_B P &= (X_L A + B + pC + X_S D) / (2R_S R_L), \\ \Lambda_C P &= \left(-X_S |Z_L|^2 A + pB + |Z_L|^2 |Z_L|^2 C - X_L |Z_S|^2 D \right) / (2R_S R_L), \\ \Lambda_D P &= (rA + X_S B - X_L |Z_S|^2 C + |Z_S|^2 D) / (2R_S R_L).\end{aligned}\quad (6.2.7)$$

Constant p in linear equations (6.2.7) is defined in (6.2.2), and a new constant also is employed:

$$r \equiv (R_S R_L + X_S X_L). \quad (6.2.8)$$

6.2.2 Derivatives with Respect to the Variable Space

Partial derivatives of P wrt variables in the x space are needed eventually. By the chain rule,

$$\frac{\partial P}{\partial x_k} = \frac{\partial P}{\partial E} \frac{\partial E}{\partial x_k}, \quad i.e., \quad \Lambda_{x_k} P = \Lambda_E P \times \Lambda_{x_k} E, \quad (6.2.9)$$

where x_k is the k th variable in vector x , and E stands for chain parameter A, B, C, or D. The middle term comes from (6.2.7). The following deals with the last term in (6.2.9).

Section 6.2.3 shows how to compute efficiently and accurately data items:

1. The overall matching network's ABCD parameters using each k th constituent element's $A_k B_k C_k D_k$ parameters, and
2. The derivatives wrt x of the overall matching network's ABCD parameters using the derivatives wrt x of each k th constituent element's $A_k B_k C_k D_k$ parameters.

Data item 2 above produces the needed last term in (6.2.9) if the derivatives wrt x of each k th constituent element's $A_k B_k C_k D_k$ parameters are made available as follows.

Recall from Section 5.4.4 and Figure 5.4.14 that the optimization variable space, X , was logarithmically related directly to element parameters L, C, t^2 , Z_0 , and, indirectly, to θ_0 in the V space. It is more convenient for both notation and computation to use the natural logarithm for that purpose:

$$x \equiv \ln(v), \quad v = e^x \quad (6.2.10)$$

Note that the partial derivative of v wrt x is

$$\Lambda_x v \equiv \frac{\partial v}{\partial x} = v. \quad (6.2.11)$$

Therefore, the chain rule can be applied to convert the derivatives wrt v_k of each k th constituent element's $A_k B_k C_k D_k$ parameters to derivatives wrt x :

$$\frac{\partial E_k}{\partial x_k} = \frac{\partial E_k}{\partial v_k} \frac{\partial v_k}{\partial x_k} = \frac{\partial E_k}{\partial v_k} v_k, \quad (6.2.12)$$

where v_k is the L , C , t^2 , or Z_0 of the k th network element, and E stands for chain parameter A , B , C , or D .

The transmission line length variable, θ_0 , is functionally related to the V space in order for cascade transmission-line (CASTL) lengths to be constrained to $0 \leq \theta_0 \leq 180^\circ \forall$ ("for all") v :

$$\theta_0 \equiv \begin{cases} \frac{\pi}{2} v, v \leq 1, \\ \frac{\pi}{2} (2 - v^{-1}), v > 1. \end{cases} \quad (6.2.13)$$

Transmission-line stub lengths are constrained to $0 \leq \theta_0 \leq 90^\circ \forall v$:

$$\theta_0 \equiv \begin{cases} \frac{\pi}{2} v, v \leq 1, SC\ Stub, \\ \frac{\pi}{2} (1 - v^{-1}), v > 1, OC\ Stub. \end{cases} \quad (6.2.14)$$

Because of (6.2.13), (6.2.14), and the linear dependence of electrical length on frequency, (5.4.11), the chain relationship equivalent to (6.2.12) for transmission-line length has two additional terms:

$$\frac{\partial E_k}{\partial x_k} = \frac{\partial E_k}{\partial \theta} \frac{\partial \theta}{\partial \theta_0} \frac{\partial \theta_0}{\partial v_k} v_k. \quad (6.2.15)$$

These considerations lead to the summary of element ABCD parameters and their derivatives in Table 6.2.1. Note that $j = \sqrt{-1}$, and the four column labels containing " j " result in real coefficients.

6.2.3 Lossless Ladder Network Analysis Equations

Section 2.6.2 describes how the overall chain matrix of a cascaded network is simply the product of each subnetwork's chain matrix:

$$T(x) = T_1 T_2 \cdots T_{k-1} T_k T_{k+1} \cdots T_K, \quad (6.2.16)$$

where T_k is the $A_k B_k C_k D_k$ matrix of the k th subnetwork, $k=1, 2, \dots, K$, and x is the vector of variables in (6.2.10). Components $A_k B_k C_k D_k$ of matrix T_k for the 12 kinds of element types are shown in four columns of Table 6.2.1. It is useful to identify the overall matrices to the left and to the right of the k th subnetwork's matrix T_k in (6.2.16):

$$L_k \equiv T_1 T_2 \cdots T_{k-1}, \quad R_k \equiv T_{k+1} \cdots T_K. \quad (6.2.17)$$

The chain matrix is formed by starting with the unit matrix and multiplying it consecutively as indicated, working from port 1 to port 2.

Table 6.2.1. Element Parameters And Derivatives In V Space.

TYPE	CODE	VAR	CON	A_k	B_k/j	C_k/j	D_k	$\Lambda_x A_k$	$\Lambda_x B_k/j$	$\Lambda_x C_k/j$	$\Lambda_x D_k$
1	CS	C		1	$-1/(\omega_0 C)$	0	1	0	$+1/(\omega_0 C)$	0	0
2	LP	L		1	0	$-1/(\omega_0 L)$	1	0	0	$+1/(\omega_0 L)$	0
3	LS	L		1	$\omega_0 L$	0	1	0	$\omega_0 L$	0	0
4	CP	C		1	0	$\omega_0 C$	1	0	0	$\omega_0 C$	0
5	LCS	L	ω_0	1	ΩL	0	1	0	ΩL	0	0
6	LCP	C	ω_0	1	0	ΩC	1	0	0	ΩC	0
7	XFMR	t^2		t	0	0	$+1/t$	t	0	0	$-1/t$
8	CASTL	Z ₀	ω_0	$\cos\theta_0$	$Z_0 \sin\theta_0$	$Y_0 \sin\theta_0$	$\cos\theta_0$	0	$Z_0 \sin\theta_0$	$-Y_0 \sin\theta_0$	0
		θ_0	ω_0					$-\Psi \sin\theta_0$	$Z_0 \Psi \cos\theta_0$	$Y_0 \Psi \cos\theta_0$	$-\Psi \sin\theta_0$
9	SCS	Z ₀	ω_0	1	$Z_0 \tan\theta_0$	0	1	0	$Z_0 \tan\theta_0$	0	0
		θ_0	ω_0					0	$Z_0 \Psi \cos^2\theta_0$	0	0
10	SCP	Z ₀	ω_0	1	0	$-Y_0 \cot\theta_0$	1	0	0	$+Y_0 \cot\theta_0$	0
		θ_0	ω_0					0	$-Y_0 \Psi / \sin^2\theta_0$	0	0
11	OCS	Z ₀	ω_0	1	$-Z_0 \cot\theta_0$	0	1	0	$-Z_0 \cot\theta_0$	0	0
		θ_0	ω_0					0	$-Z_0 \Psi / \sin^2\theta_0$	0	0
12	OCP	Z ₀	ω_0	1	0	$+Y_0 \tan\theta_0$	1	0	0	$-Y_0 \tan\theta_0$	0
		θ_0	ω_0					0	$Y_0 \Psi / \cos^2\theta_0$	0	0

$$\Omega = \omega / \left(1 - \frac{\omega^2}{\omega_0^2} \right), \quad \theta = \omega \theta_0, \quad \Psi = \frac{\omega}{\omega_0}, \quad \Psi = \frac{\partial \theta}{\partial x} = \begin{cases} \text{if } \theta_0, \theta_0 \leq \pi/2 \text{ CASTL \& SC stubs,} \\ \omega (\pi - \theta_0), \theta_0 > \pi/2 \text{ CASTL \& } \theta_0 \leq \pi/2 \text{ OC stubs.} \end{cases}$$

Only four operations are required to add each subnetwork in Table 6.2.1, except for type 8. Each other even-numbered type adds a series reactance, B_k , so that

$$\begin{bmatrix} A_L & jB_L \\ jC_L & D_L \end{bmatrix} \begin{bmatrix} 1 & jB_k \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} A_L & j(B_L + A_L B_k) \\ jC_L & D_L - C_L B_k \end{bmatrix}. \quad (6.2.18)$$

Each odd-numbered type (except type 7) adds a parallel susceptance, C_k , so that

$$\begin{bmatrix} A_L & jB_L \\ jC_L & D_L \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jC_k & 1 \end{bmatrix} = \begin{bmatrix} A_L - B_L C_k & jB_L \\ j(C_L + D_L C_k) & D_L \end{bmatrix}. \quad (6.2.19)$$

The ideal transformer, type 7 in Table 6.2.1, is updated by

$$\begin{bmatrix} A_L & jB_L \\ jC_L & D_L \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix} = \begin{bmatrix} t A_L & jB_L / t \\ j t C_L & D_L / t \end{bmatrix}. \quad (6.2.20)$$

The cascade transmission line, type 8 in Table 6.2.1, has a full T_k matrix, requiring eight operations for updating.

Derivatives wrt x_k of the overall matching network's ABCD parameters (T) are related to derivatives wrt x_k of each k th constituent element's $A_k B_k C_k D_k$ parameters [Iobst]:

$$\Lambda_{x_k} T \equiv \begin{bmatrix} \frac{\partial A}{\partial x_k} & \frac{\partial B}{\partial x_k} \\ \frac{\partial C}{\partial x_k} & \frac{\partial D}{\partial x_k} \end{bmatrix} = L_k \times \Lambda_{x_k} T_k \times R_k, \quad (6.2.21)$$

where L_k and R_k are defined in (6.2.17), and $\Lambda_{x_k} T_k$ is the matrix with elements from the right-hand four columns of Table 6.2.1. The four dependent partial derivatives found numerically by (6.2.21) are used in (6.2.9) where $\Lambda_{x_k} E$ is required, E standing for A, B, C, or D.

It is possible to obtain the exact partial derivatives of the ABCD parameters of a lossless network in just one forward analysis (port 1 to port 2) at a frequency, as in (6.2.16), using Tellegen's theorem [Cuthbert, 1983:108]. Surprisingly, it is computationally cheaper and conceptually simpler to find them from a second analysis made from port 2 to port 1 [Orchard]. In the first analysis, from port 1 to port 2, the L_k matrices are composed subnetwork by subnetwork as described, storing the four real values in each matrix, L_k . In the second analysis, from port 2 to port 1, the R_k matrices are likewise composed and the four real values are used immediately in (6.2.21).

6.2.4 Lossless Ladder Network Analysis Algorithm

The fruit of the labor in Section 6.2 is to fill vector e with the errors at each of the m frequencies; see (6.1.6):

$$e = [e_i] \equiv [P_i - g_i], \quad i = 1, 2, \dots, m. \quad (6.2.22)$$

When gradient optimization is required, it is also necessary to construct the non-square ($m > n$) Jacobian matrix, J

$$J \equiv \begin{bmatrix} \Lambda_{x_1} P_1 & \Lambda_{x_2} P_1 & \cdots & \Lambda_{x_n} P_1 \\ \Lambda_{x_1} P_2 & \Lambda_{x_2} P_2 & \cdots & \Lambda_{x_n} P_2 \\ \Lambda_{x_1} P_3 & \Lambda_{x_2} P_3 & \cdots & \Lambda_{x_n} P_3 \\ \Lambda_{x_1} P_4 & \Lambda_{x_2} P_4 & \cdots & \Lambda_{x_n} P_4 \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{x_1} P_m & \Lambda_{x_2} P_m & \cdots & \Lambda_{x_n} P_m \end{bmatrix}. \quad (6.2.23)$$

The analysis sequence is:

1. For each frequency ω_i , $i=1$ to m , and its related constants R_S , X_S , R_L , X_L , precompute q and p (6.2.2), r (6.2.8), and all the other coefficients in (6.2.7).
2. Assign values to the components of variable vector x , then convert them to v by (6.2.10).
3. Assign a frequency, $\omega=\omega_i$, then compute Ω and θ for transmission-line elements using (6.2.13), (6.2.14), and the equations at the end of Table 6.2.1.
4. Start with the unit matrix, and multiply it consecutively from port 1 to port 2 as indicated in (6.2.16)-(6.2.20), storing all L_k matrices.
5. Use the ABCD obtained to evaluate (6.2.1), store $(P_i - g_i)$ into e_i , and evaluate and store (6.2.7).

6. Start with the unit matrix, and multiply it consecutively from port 2 to port 1 to evaluate (6.2.21).
7. Use those values in (6.2.9), and store the results in row i of (6.2.23).
8. If $i \neq m$, go to 3, else stop.

6.3 Grid Search

The purpose of a grid search is to locate the likely global minimax solution to within about 3 percent in the element parameter space. That is a necessary starting point for the gradient-based method of multipliers (Section 6.4). The grid search of vertices of large, medium, and small hypercubes, each repositioned in the variable space, approximately locates the likely global minimax solution while avoiding local anomalies in the envelope surface. Fortunately, considerable recent research describes notation and convergence of such pattern-search algorithms [Torczon, 1991, 1997], [Dennis]. This section shows how the 2^n -factorial pattern search can be applied in GRABIM.

6.3.1 Minimax Objective Function

The smooth error function, e_i , at each i th frequency is defined in (6.2.22) as the difference between the transducer power function and its goal value. However, the non-differentiable objective according to (6.1.6) minimizes the maximum of such error functions. (Derivatives of the envelope exist, but are discontinuous.) Each error, e_i , is a function of the vector of element parameters, v , and that in turn is related logarithmically to the variable vector x , as stated by (6.2.10). The minimax problem ultimately iterates on the variables in vector x .

The grid search evaluates functions at the vertices of hypercubes centered at a sequence of iterates, $x_1, \dots, x_k, x_{k+1}, \dots$, where x_k now denotes the vector value of x at the k th iteration. The minimax objective function at x_k has a value, called *best*, which is set equal to the largest error over the m frequency samples. It is important to note that at the next iterate, x_{k+1} , the frequency scan can be abandoned immediately upon encountering any error greater than *best* from the preceding pattern about x_k . That reduces the required computing effort by about one half on average.

6.3.2 Pattern Searches

Direct search methods for numerical optimization depend only on simple decreases in a finite sequence of smooth or nonsmooth objective function values. Direct search methods are not only derivative-free, they do not model the objective function; e.g. they do not assume it is quadratic. They are valued not for fast local convergence, but for reliable

steady improvement. "The most important distinction, for theoretical purposes, is between those methods for which the set of search directions is modified at the end of each iteration and those methods for which the set of search directions remains fixed across all iterations" [Torczon,1991]. Probably the most popular direct search algorithm is Nelder-Mead, which utilizes a simplex geometric solid which adapts in size and shape after each iteration. Its deficiency is that it does not search in each of n linearly-independent directions at every iteration, so convergence cannot be guaranteed. It is known that Nelder-Mead sometimes fails when the number of variables becomes large enough, e.g. $n=16$.

A particular subset of direct search methods, pattern search methods, have fixed linearly-independent search directions across all iterations and have been generalized to include a global convergence theory [Torczon,1997]. They are distinguished by iterates that lie on a scaled, translated integer lattice and are backed by convergence theorems that are borne out in practice by numerical tests. They are robust, converging unfailingly to a point where the objective function is non-differentiable or the gradient is zero. Members of the class of direct search methods include

- Coordinate search with fixed step lengths,
- Evolutionary operation (EVOP) using factorial searches,
- The original pattern search by Hooke and Jeeves, and
- The new multi-directional search algorithm by Dennis and Torczon.

GRABIM utilizes factorial search, a 1950's method enhanced by 1990's research. The several reasons factorial searches are mandatory for broadband impedance matching become apparent in the following description. EVOP, a statistical analysis technique, today goes by the name Design of Experiments or DOE [Montgomery]. However, GRABIM is similar only to the factorial search aspect of EVOP or DOE and incorporates no aspect of chance. There has been at least one attempt to combine pattern (Hooke-Jeeves algorithm) and random searches [Bandler,1969].

6.3.3 Grid Geometry

Iteration is the process of selecting a value set sequence (iterates) for variables and testing a set of function values in the vicinity. Figure 6.3.1 illustrates the variable space for two dimensions; it is still representative when there are more than two dimensions. The space has been discretized into $101 \times 101 = 10201$ values where the grid lines intersect. Each grid point has integer coordinates in terms of $16x_1$ and $16x_2$; e.g. the upper left-hand dot has integer coordinates $(-29, +31)$, representing $x_1 = -29/16 = -1.8125$ and $x_2 = +31/16 = 1.9375$. The choice of

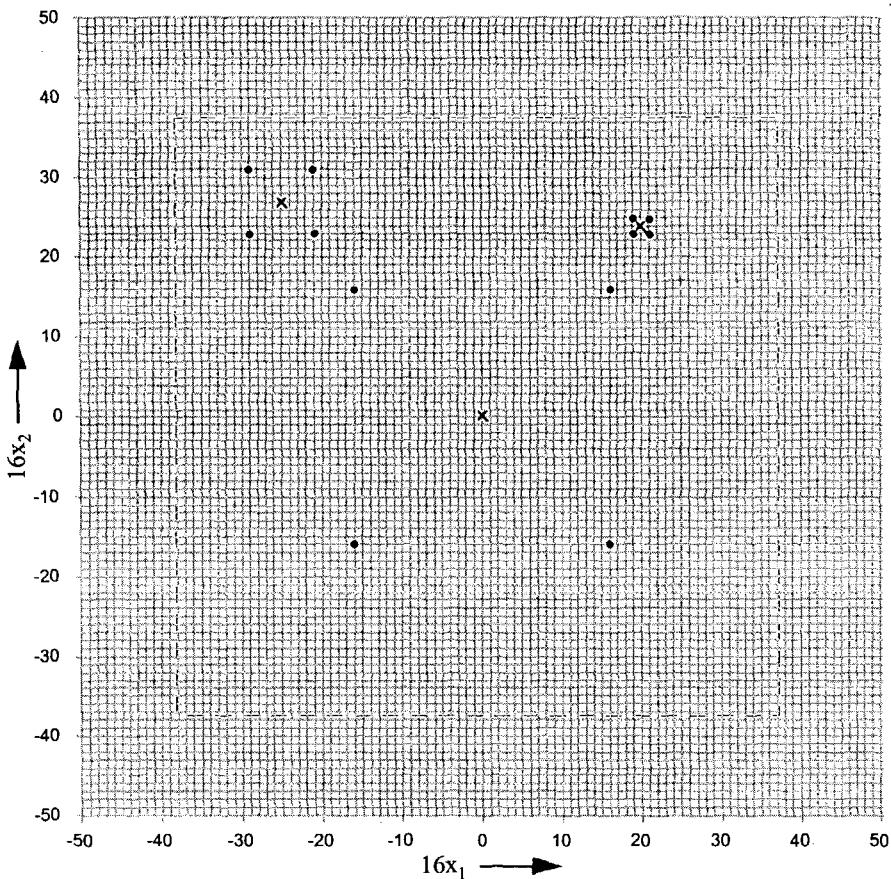


Figure 6.3.1. Granularity and hypercubes in a solution two-space.

factor 16 is made clear below. Some important integer and real values in the variable (x) and parameter (v) are summarized in Table 6.3.1, using the factor 16 and (6.2.10).

Table 6.3.1. Key Integer and Real Values in the X and V Spaces.

$16x$	-1	+1	-4	+4	-16	+16	-36.8	+36.8	-50	+50
x	-0.0625	0.0625	-0.25	0.25	-1	+1	-2.3	2.3	-3.12	3.12
v	0.9394	1.0645	0.78	1.28	0.37	2.7	0.1	10	.044	22.8

Consider the pattern represented by the central pattern of four dots in Figure 6.3.1 located at coordinates $(\pm 16, \pm 16)$. According to Table 6.3.1, parameters such as L , C , Z_0 , etc., have corresponding values of $v=0.37$ and $v=2.7$. It is now useful to revisit Figures 5.4.2, 5.4.12, 5.5.5, 5.5.6, and 5.5.7. These display cross sections of the reflection envelopes versus the parameter over a range from $0.1 \leq v \leq 10.0$; see Table 6.3.1 and

the dashed rectangle in Figure 6.3.1. It is important that the initial exploratory coordinates in the pattern are spaced widely apart to avoid anomalies such as in Figure 5.5.7. The two-variable ($n=2$) case in Figure 6.3.1 shows large, medium, and small hypercubes having sides of scaled lengths 32, 8, and 2, respectively. The 2^n -factorial algorithm "constructs a hypercube centered at the current iterate and then computes the function values at the vertices to find a function value that is strictly less than the function value at the current iterate. If a new best point is found, the hypercube is then centered on the new best iterate and the search is restarted. If not, the size of the hypercube is reduced" [Torczon,1991]. The reduction shown in Figure 6.3.1 is by a factor of four, performed only twice for clarity of the illustration.

In practice, it has been found that three reductions (four hypercube sizes or cycles) are useful, starting at $\Delta x=1$. Thus, the normalization similar to Figure 6.3.1 is to 64 instead of 16, and the final two factors in V space are 0.9845 and 1.0157. Therefore, the X space in the domain $-3.125 \leq x \leq +3.125$ is discretized into 401 points in each coordinate. The factorial search by no means visits all the 401^n grid points in multi-dimensional n space. Example 5.6.1 utilizes the termination impedance data in Table 5.4.1; the 2^6 -factorial search with four cycles of $\Delta x=1$, $1/4$, $1/16$, and $1/64$ converges in 0.01 minutes and visits 7026 points using a 200 MHz PC in safe mode (4 times faster than in Windows[®]). Indeed, knowledge of the response surface (Section 5.4) makes visiting all 401^n points unnecessary, thus avoiding the "curse of dimensionality" in the "vastness of hyperspace" [Wilde:279-281].

The grid search is not just about searching along individual coordinates, because all possible combinations of all coordinate trial values are evaluated for every pattern. Those patterns should be viewed as an archeological grid that is being repositioned and eventually refined to locate an irregular minimum; see Figure 5.4.13. In that context, the crude convergence of a grid search need only locate the neighborhood of the minimum in order that the minimax-constrained gradient optimization can carry on from there.

6.3.4 Grid Algorithm

A basis matrix, B , and a generating matrix, C , are required to define a pattern search [Torczon,1997]. The GRABIM technique is based on the benign properties of the function envelope in each coordinate direction, so a strategic choice is $B=I$, the unit matrix. A generating matrix contains in its columns all possible combinations of $\{-1, +1\}$ and a column of zeros; see Figure 6.3.1 for $n=2$. In that case,

$$C = \begin{bmatrix} 1 & 1 & -1 & -1 & 0 \\ 1 & -1 & -1 & 1 & 0 \end{bmatrix}. \quad (6.3.1)$$

If the current iterate, x_k , is the center of a hypercube, a square in Figure 6.3.1, then the q th exploratory step from there is

$$s_k^q = \Delta_k c_k^q, \quad (6.3.2)$$

where Δ_k is the step length and c_k^i denotes a column of matrix C_k :

$$C_k = [c_k^1, \dots, c_k^q, \dots, c_k^{2^n+1}]. \quad (6.3.3)$$

The 2^n -factorial pattern search makes a series of well-defined deterministic exploratory moves about the current k th iterate, x_k :

$$x_{k+1}^q = x_k^q + s_k^q, \quad q = 1, \dots, 2^n + 1, \quad (6.3.4)$$

which includes the current iterate (hypercube center) as well, for convenience.

As a further example, a generating matrix for $n=4$ is

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & 0 \end{bmatrix}. \quad (6.3.5)$$

Generating matrices can be put in a standard form [Torczon, 1997]:

$$C_k = [M_k \ M_k \ L_k], \quad (6.3.6)$$

where L_k is just a column of zeros in a 2^n -factorial search. However, for programming purposes, the combinations can be generated by nested loops, as in Figure 5.5.4, which is equivalent to a tree diagram.

It is important to note that (6.3.4) is a vector equation; (6.3.2) and (6.3.3) show that each row of matrix C corresponds to a variable parameter in the matching network. Looking at (6.3.5), each row contains n values that are $+1$ and n values that are -1 . That means each of the two possible element parameter values will be used n times at each of the m frequencies in the current exploratory pattern. Furthermore, when those values are precomputed before each pattern is explored (*static initialization*), they can be tested then for any element parameter's upper and/or lower bounds. If any bounds are violated in the current lattice, then the precomputed value should be coded (e.g., set to 999) so that when encountered in the ABCD ladder network analysis, as in (6.2.18), it will cause immediate termination (e.g., returning $P_i=1E10$). Finally, if the precomputation finds a "hold" placed on a parameter, then the ± 1 in the C matrix can be replaced by 0 to maintain the nominal value of that variable.

The algorithm for the 2^n -factorial pattern about the k th iterate (center) is shown in Table 6.3.2 [Torczon, 1997]. The function value $f(x_k)$ is defined by (6.1.1), where P_i in (6.1.4) can be substituted for ρ_i without effect.

Table 6.3.2. Exploratory Moves in the 2^n -Factorial Algorithm.

Given x_k , Δ_k , $f(x_k)$, and C as in (6.3.3), set $s_k = 0$ and $best = f(x_k)$;

For $q = 1$ to 2^n do

(a) $s_k^q = \Delta_k C^q$, $x_k^q = x_k + s_k^q$, and compute $f(x_k^q)$;

(b) If $f(x_k^q) < best$, then $best = f(x_k^q)$ and $s_k = s_k^q$;

Return.

The pattern searches are started in cycle 1 with $\Delta_1 = 1.0$ from the point $x=0$ with $best=f(0)$. That sequence of fixed large patterns (first cycle) ends when $best$ fails to change after a complete pattern. In practice, because of the slow convergence of factorial searches, it is helpful to apply a relative change tolerance of 0.001 so that the cycle ends when the change in $best$ is insignificant. Then the *cycles* are repeated with $\Delta_2 = 1/4$, $\Delta_3 = 1/16$, and $\Delta_4 = 1/64$.

Examination of Figure 6.3.1 shows that some vertices of a k th, $k+1$ and perhaps subsequent patterns are duplicated. (By using factors of four instead of two in reducing Δ , the duplication among cycles is less likely.) Note the integer coordinates illustrated in Figure 6.3.1, $16x_1$ and $16x_2$; a particular iterate is the center of one of four hypercubes defined by (6.3.2)-(6.3.4). Therefore, the integer coordinates of each new exploratory point can be listed efficiently during static initialization and sorted with the stored points already explored to eliminate duplicate function evaluations. Without this static test before each pattern is evaluated, there would be at most about 100,000 function evaluations consuming about 0.5 minute over the four cycles. Testing for prior exploratory points in just the last several patterns is probably sufficient. Aside from the memory required, complete testing is not likely to be advantageous, because the effort may approach that required to obtain function values.

It is noted above that both basis and generating matrices, B and C , are required to define a pattern search. Then, the search directions at each iterate are found from the matrix product $B \times C$. For the K^n -factorial searches in coordinate directions, $B=I$, the unit matrix, so further consideration of the basis matrix is not required; however, the next section mentions a rotated, orthogonal basis matrix.

6.3.5 Other Factorial Search Algorithms

Figure 6.3.1. illustrates the 2^n -factorial grid search. Section 5.5.2 compares the 3^n -factorial grid search to the 2^n -factorial grid search, where each pattern has 3^n and 2^n points, respectively. Before the underlying reflectance behavior in coordinate directions was recognized, K^n -factorial grid searches were studied, $K=3, 5, 7, 9$, and 11 [Cuthbert, 1994b]. For the fixed-pattern searches which do not use an adaptable simplex, the number of points in the lattice is exponential in n , the

number of variables. However, there are also advantages in more dense lattices.

Lattices locate a minimum more precisely. See Figure 5.5.8; exploration from the center is not just in coordinate directions. In Figure 5.5.8 in the 3^2 case, there are four search directions, while the 2^2 case explores in only two directions (arguably). Clearly, the 5^2 case in Figure 5.5.3 explores in eight directions. This property produces a noticeable effect, because each lattice is not allowed to rotate and, thus, better to locate an irregular minimum (Figure 5.4.13). Although a precise location of the minimum is not required of the grid search, it does shorten the subsequent gradient search.

Perhaps the best strategy is to use more coordinate trial points for lower numbers of network parameters (branches). Timing relationship (5.5.3) shows that the dominant factor is the exponential relationship involving NT, the number of grid trial values, and NB, the number of branches or parameters: NT^{NB} . Table 6.3.2 suggests a strategy.

Table 6.3.2. Number of Trial Combinations Versus NT and NB.

NB	NT	NT^{NB}
2	31	961
3	9	729
4	5	625
5	4	1024
6	3	729
7-10	2	≤ 1024

Such a strategy is easy to implement in the nested DO loops in Figure 5.5.4. The $NT=31$ points for $NB=2$ is probably an extravagant use of memory, so that $NT=9$ would be a better choice in that case.

An additional enhancement at the end of the grid search is to rotate, randomly or otherwise, the orthogonal axes of the final hypercube for a few additional trials. Previous pattern search methods continually utilized orthogonal search directions rotated randomly or speculatively [Bandler,1972a], [Wilde 312-3]. It is shown in Section 5.4.4 that non-adaptive K^n -factorial grid searches avoid minor aberrations, some of which might be local minima. In the neighborhood of the likely global minimum, it is safe to compensate for the limited number of search directions on the chance that a few coordinate rotations might refine the location with little additional effort. Efficient computation of a set of orthonormal directions corresponding to columns of a rotated basis matrix, B , has been described [Himmelblau:161].

The multi-directional line search algorithm, listed in Section 6.3.2 as a member of the distinguished class of pattern search algorithms, requires only a linear increase of exploratory points versus n ; and B can be any nonsingular matrix, including I [Torczon,1991]. It may be

possible to employ that search algorithm in the GRABIM technique and still avoid any envelope surface anomalies. However, the broadband matching problem fundamentally limits the number of variables, so that the factorial grid search phase should be allocated about the same time as the minimax-constrained gradient optimization phase. In other applications having large n , the multi-directional search algorithm appears to be very attractive, especially using parallel machines [Dennis].

6.4 Method of Multipliers

The broadband impedance matching minimax problem in (6.1.1) is solved approximately by a grid search. Now, an exact solution of the equivalent problem in (6.1.6), which shifts the effects of nondifferentiability into a set of inequality constraints, is detailed. There are many excellent references, particularly general surveys [Fletcher, 1974, 1981b], [Ryan], and detailed mathematical descriptions [Fletcher, 1981a], [Bertsekas]. The method of multipliers and alternative constrained optimization methods mentioned in Section 6.4.5 all depend on the modern interpretation that Lagrange multipliers arise from subdifferentiation of nonsmooth objective functions [Rockafeller, 1993]. The method of multipliers is developed concisely to facilitate subsequent computer programming.

6.4.1 The Problem

The general *nonlinear programming problem* is

$$\text{Minimize } f(x) \quad \text{s.t.} \quad h(x) = 0 \quad \text{and} \quad c(x) \geq 0. \quad (6.4.1)$$

The objective function f is a scalar function of the variable vector x , and equality constraints in vector h and inequality constraints in vector c are also functions of x . All functions may be nonlinear in x . A very readable and brief introduction to methods for solving the many practical problems having the form in (6.4.1) has been published [Fletcher, 1981b]. The following sections detail the method of multipliers for solving this problem and end with mention of several alternative methods.

Several authors express the inequality constraints as the opposing inequality $c \leq 0$; the difference is only a change of sign. Comparing GRABIM problem (6.1.6) to (6.4.1), it is seen that $f(x) \equiv x_{N+1}$, an added variable, and

$$c_i(x) \equiv (g_i + x_{N+1}) - P_i, \quad i = 1, \dots, m. \quad (6.4.2)$$

This says that all constraints should be positive, i.e., all responses, P_i , should be less than their respective goals as floated by a minimal bias, x_{N+1} . Having related the specific to the general problem, the latter is described in detail as it appears in optimization literature.

6.4.2 Quadratic Penalty Functions

For the equality-constrained problem in (6.4.1), a simple penalty function for equality constraints is [Courant]

$$F(x) = f(x) + s \sum_{r=1}^l [h_r(x)]^2 . \quad (6.4.3)$$

The penalty is the sum of squared errors weighted by a scalar parameter, s . If $x(s)$ represents the minimizer of (6.4.3) for a fixed s , then it can be shown that the solution⁴ of (6.4.1), x^* , is the limiting value of $x(s)$ as $s \rightarrow \infty$. Numerically, the solution is obtained by the sequence of unconstrained minimizations technique (SUMT), employing $s = \{1, 10, 100, \dots\}$. Severe numerical difficulties occur, leading to increasingly poor accuracy for larger value of s . A great deal of analysis was devoted to overcoming that problem [Fiacco]. The comparable approach for inequality constraints is

$$F(x) = f(x) + s \sum_{i=1}^m [\min(c_i(x), 0)]^2 , \quad (6.4.4)$$

where the $\min(\cdot)$ function selects the smaller of its two arguments. Another, less troublesome, defect was thus added: The first derivative of F is continuous, but the second derivative is discontinuous where $c_i(x)=0$.

The method of multipliers avoids the requirement that $s \rightarrow \infty$ by adding a term linear in $c(x)$:

$$F(x) = f(x) - \lambda^T c + \sum_{i=1}^m s_i [\min(c_i(x), 0)]^2 , \quad (6.4.5)$$

where λ is a vector of Lagrange multipliers, and there there is a weighting element, s_i , for each constraint, $c_i(x)$. This is the *augmented Lagrangian penalty function*. In fact, the first two terms in (6.4.5) constitute the classical *Lagrangian function*. The reason the method of multipliers MUST start from a neighborhood of the constrained minimum is that the first- and second-order conditions for the Lagrangian function must be satisfied. Those conditions at the constrained minimum are that its first partial derivatives wrt x are zero, and its Hessian matrix of second partial derivatives wrt x is positive definite. When these conditions are satisfied, there exist suitably large values of weights, s_i in (6.4.5), such that a solution of (6.4.1) exists [Fletcher, 1981a:132]. Because $f(x)$ is linear in this case, the second-order (curvature) requirement reduces to the curvature of all active constraints being non-negative in some neighborhood of the solution.

There is an expression that differs from (6.4.5) only by an amount that is not a function of x [Cuthbert, 1987:296-8]:

⁴ The * superscript in Chapter Six denotes optimal real value(s), not complex conjugation.

$$F(x, s, u) = f(x) + \sum_{i=1}^m s_i \left[\min(c_i(x) - u_i, 0) \right]^2, \quad (6.4.6)$$

where the Lagrange multipliers at the solution, x^* , can be expressed in terms of finite weights, s_i , and offsets, u_i , by

$$\lambda_i = s_i u_i, \quad i = 1, \dots, m. \quad (6.4.7)$$

"Each constraint has an associated Lagrange multiplier (at the solution) which is conventionally zero if the constraint is inactive. The Lagrange multiplier λ^*_i can be interpreted (to first order) as the rate of change in $F(x^*)$ that would result from a perturbation in the constraint function c_i " [Fletcher, 1981b].

Note that setting all $u_i = 0$ and $s_i = s$ reduces (6.4.6) to (6.4.4). More importantly, minimizing (6.4.6) obtains the same optimal variable value, x^* , as when minimizing (6.4.5). Unfortunately, the optimal Lagrange multipliers, or the corresponding weights and offsets, are not known in advance and must be found by a sequential process. These Lagrange multiplier penalty functions were first developed for equality constraints [Powell, 1969], [Hestenes] and later extended to inequality constraints [Rockafellar, 1973].

6.4.3 Adjusting the Multipliers

It is first explained why the Lagrange multipliers, or the weights and offsets according to (6.4.7), must be adjusted. Suppose the corresponding sets of s_i and u_i have been assigned values; then $F(x, s, u)$ in (6.4.6) can be minimized, thus finding some value $x = x'$. That process defines the function $x'(s, u)$ and leads to a theorem [Powell, 1969]:

If the value of the variable, x , which minimizes $F(x, s, u)$ is $x'(s, u)$, then $x'(s, u)$ is a solution of the constrained problem

$$\text{Minimize } f(x) \quad \text{s.t. } c(x) = c[x'(s, u)], \quad (6.4.8)$$

where all the constraints, $c(x)$, are active (binding, equality) constraints.

This is true because the summation term in (6.4.6) is constant for any value of x satisfying the constraints of (6.4.8). Therefore, considering only those values of $x = x'$, the value of x which minimizes (6.4.6) also minimizes $f(x)$ subject to the constraints in (6.4.8) being satisfied [Walsh:200]. This theorem is important because it states that only the vector values of s and u need be found so that the right-hand side of (6.4.8) is zero, i.e.

$$c[x'(s, u)] = 0. \quad (6.4.9)$$

Therefore, the problem has been reduced to adjusting vectors s and u to solve the set of nonlinear equations expressed by (6.4.9). Actually,

the Lagrange multipliers at the ultimate solution, x^* , are fixed values, so (6.4.7) implies that only one vector, say u , need be adjusted. If the weights, s_i , are sufficiently large, iteration $k+1$ on vector u is:

$$u^{(k+1)} = u^{(k)} - c[x'(u^{(k)})]. \quad (6.4.10)$$

The iteration in (6.4.10) can be made to have linear convergence at any rate desired. (If second derivatives of $F(x,s,u)$ were available, a quadratic convergence could be obtained.) Correction (6.4.10) is applied unless it happens that the maximum-constraint moduli ratio⁵, $\|c^{(k+1)}\|_\infty / \|c^{(k)}\|_\infty$, is not sufficiently small (0.25 or less was recommended). In that case, certain weights, s_i , should be increased (a factor of 10 was recommended). Once the weights, s_i , become sufficiently large and c sufficiently small, the maximum-constraint moduli ratio will always pass the above test. A detailed flow diagram for these adjustments has been published [Powell, 1969:286].

A reliable initialization before sequentially minimizing (6.4.6) is to set all weights to unity, $s_i=1$, all offsets to zero, $u_i=0$, and $f(x)=x_{N+1}$ to the infinity norm of the error vector, $e \equiv [e_i]$, where $e_i \equiv (P_i - g_i)$:

$$x_{N+1}^{(0)} = \|e\|_\infty \equiv \max_i |e_i|, \quad i = 1 \text{ to } m. \quad (6.4.11)$$

With this initial value for the added variable, the constraint functions from (6.4.2) are

$$c_i(x) = x_{N+1} - e_i, \quad i = 1, \dots, m. \quad (6.4.12)$$

This substitution makes only a few of the constraints active ($c_i \leq 0$) during the first minimization, so that the objective function (6.4.6) is not dominated completely by the summation.

Finally, the method of multipliers algorithm is summarized in Table 6.4.1.

Table 6.4.1. Multiplier Penalty Method Steps to Minimize (6.4.6).

1. Initialize offset $u=0$, set $s_i=1$, and set added variable x_{N+1} to the maximum error, $e_i = (P_i - g_i)$, $i=1$ to m .
2. Minimize $F(x,s,u)$ to find $x'(s,u)$ and $c[x'(s,u)]$.
3. Stop if $\|c(x')\|_\infty$ is suitable small, but if $\|c(x')\|_\infty$ increased, go to step 5.
4. If each $|c_i(x')|$ decreased by factor 4 or more, set $u=u-c(x')$ and go to step 2.
5. Corresponding to each $|c_i(x')|$ not decreasing by factor 4, adjust $s_i=10s_i$ and $u_i=u_i/10$, then go to step 2.

⁵ These *infinity norms* are defined as in (6.4.11).

The minimizations are well scaled and require progressively fewer iterations, with the weights seldom exceeding 1000.

6.4.4 Gauss-Newton Unconstrained Minimizer

This section describes an unconstrained minimization method that is particularly well suited for step 2 in Table 6.4.1. It has many advantages:

- Only first partial derivatives of $F(x)$ are required,
- A positive-definite matrix of second derivatives (the Hessian) is estimated,
- No memory of results from prior search directions is required,
- The line-search criterion is simple and back tracks intelligently,
- Variables can be bounded or held easily, and
- Final convergence is quadratic.

The unconstrained function to be minimized is restated from (6.4.6) and (6.4.2):

$$F(x) = x_{N+1} + r^T r / 2, \quad r \equiv [r_i] = [\sqrt{s_i} (x_{N+1} - P_i(x) + g_i - u_i)], \quad i \in A, \quad (6.4.13)$$

where set A contains the integer indices of the α active constraints, i.e. $c_i(x) < 0$; the *residuals*, r_a , are the augmented errors at each frequency; and divisor 2 is added for subsequent notational convenience. The first partial derivatives of $F(x)$ wrt x are:

$$\Lambda_j F = \sum_{a=1}^{\alpha} r_a (\Lambda_j r_a), \quad j = 1, \dots, N; \quad \Lambda_{N+1} F = 1 + \sum_{a=1}^{\alpha} \sqrt{s_a} r_a. \quad (6.4.14)$$

The first N partial derivatives of r_a wrt x_j , $(\Lambda_j r_a)$, are simply $-\Lambda_j P_a \times \sqrt{s_a}$ by differentiation of (6.4.13). The partial derivatives of P_a are available from the Jacobian matrix in (6.2.23). Therefore, it is straightforward to construct the gradient vector, g , of function $F(x)$, having constituents according to (6.4.14). Holding a variable constant is accomplished simply by setting its component in the gradient vector to zero. Also, a practical computer program would be dimensioned to accommodate all possible residuals, r_i in (6.4.13). Then, those satisfying $r_i > 0$ are set equal to zero so that they no longer affect the immediate search step, which is described next.

A Newton search has two distinct phases. First, a search direction is constructed, and second, a step in that direction, decreasing the function value, is determined. A *Newton step* in variable space from iterate point x_k is a vector that has both direction and length [Cuthbert, 1983:124]:

$$dx = -H^{-1} g, \quad (6.4.15)$$

where H is the Hessian matrix, composed of all second partial derivatives of objective function $F(x)$. The Gauss-Newton approximation of the Hessian, including the Levenberg-Marquardt term, vI , is

$$H \approx J^T J + v I, \quad (6.4.16)$$

where I is the unit matrix. This approximation is possible because of the sum of squared residuals, $r^T r$ in (6.4.13); note that the second partial derivative of x_{N+1} is zero. The *normal matrix* $J^T J$ is square, symmetric, and positive definite. The Jacobian matrix, J , is formed from the partial derivatives of the residuals in (6.4.13); its negative is simply obtained from (6.2.23), and a last column is added according to (6.4.14):

$$-J \equiv \begin{bmatrix} \sqrt{s_1} \Lambda_1 P_1 & \cdots & \sqrt{s_1} \Lambda_N P_1 & -\sqrt{s_1} \\ \sqrt{s_2} \Lambda_1 P_2 & \cdots & \sqrt{s_2} \Lambda_N P_2 & -\sqrt{s_2} \\ \sqrt{s_3} \Lambda_1 P_3 & \cdots & \sqrt{s_3} \Lambda_N P_3 & -\sqrt{s_3} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{s_m} \Lambda_1 P_m & \cdots & \sqrt{s_m} \Lambda_N P_m & -\sqrt{s_m} \end{bmatrix}. \quad (6.4.17)$$

When scalar $v=0$, dx in (6.4.15) approximates a Newton step, and when $v \rightarrow \infty$ the step is an infinitesimal steepest descent. A superior feature is to replace the unit matrix, I , by an adaptive diagonal matrix, D^2 , that functions as an implicit scaling of variables [Cuthbert,1987:208].

The Gauss-Newton unconstrained optimizer solves a system of linear equations to define each trial step, dx :

$$[J^T J + vD^2] dx = J^T r - (0 \ 0 \ \dots \ 0 \ 1)^T, \quad (6.4.18)$$

where the right-hand side is the negative gradient, $-g$. A computer program requires a number of strategic policies, e.g. how to increase parameter v to decrease the step length when the dx step in (6.4.15) produces an increase in objective function value. The matrix inverse indicated in (6.4.15) is only symbolic; the system of equations in (6.4.18) should be solved by matrix decomposition techniques. The reader is referred to [Cuthbert,1987:209-214,421-425] for a detailed flow chart, discussion, examples, and listing of the BASICA code for program LEASTP ($p=2$ in this application). Modifications to include the added variable are minimal.

6.4.5 Alternative Constrained Optimization Methods

The literature describes many varied approaches for solving the constrained optimization problem (6.4.1). The quadratic penalty functions described in Section 6.4.2 belong to one of several sequential unconstrained minimization techniques (SUMT) classes: they are *exterior-point methods* characterized by their use of only infeasible ($c(x) < 0$) points.

The first provably convergent direct search method for nonlinear programming has been applied recently to the augmented Lagrangian objective function in (6.4.5) [Torczon:1999]. The particular outer iteration in Section 6.4.3 that adjusts Lagrange multipliers does not require derivatives and converges at a linear rate. The inner iteration in step 2 of Table 6.1.1 is an unconstrained minimization; the Gauss-Newton minimizer described in Section 6.4.4 converges exactly and at a much faster quadratic rate. Torczon has proposed that a pattern search minimization algorithm also be employed in the inner iteration; it would converge linearly and is best suited for exterior-point constrained optimization applications. It is claimed that successive inexact pattern search minimizations can be tolerated, based on other recent research [Conn]. No data are available to support this claim. Indeed, the grid search described in Chapters Five and Six utilizes the objective criteria in Table 6.3.2 to achieve the same approximate result. The more highly convergent gradient-based multiplier technique still is required to prune network branches reliably and efficiently.

A more popular exterior-point technique is the sequential quadratic programming (*SQP*) method, which applies Newton's method (requiring second partial derivatives) to solve the nonlinear equations in (6.4.9) that arise from the method of Lagrange multipliers. Both x and λ are iterated, using a linearization of both the gradient and the constraint vectors to produce a sequence of quadratic programming (QP) subproblems. The minimax problem solved using SQP has been reported [Rustem], [Zhou]. The required second partial derivatives pose a serious disadvantage.

Another SUMT class is *interior-point methods*, characterized by their property of requiring constraint feasibility ($c(x) \geq 0$) at all times. They are often called *barrier methods*, and introduce two added difficulties: the barrier function is not defined outside the feasible region, and an initial strictly feasible starting point is required. Although sometimes utilized for RF network design [Waren], they remained unpopular until 1984 when N. A. Karmakar introduced a highly successful barrier solution for linear programming. The re-examination that his approach produced for both linear and nonlinear programming is ongoing [Powell,1995], [Polyak]. It is not clear that barrier methods are superior to quadratic penalty methods in any substantial way.

There are numerous other approaches for solving the nonlinear programming problem, including exact penalty functions and feasible direction methods [Fletcher,1981a], [Bertsekas]. The exact penalty function method attempts to employ a single parameter-free optimization to a solution, often involving nondifferentiable objective functions. Feasible direction methods date back to at least 1960, and attempt to maintain feasibility by searching from one feasible point to another along feasible arcs. One popular method related to this class is generalized

reduced gradients (GRG). In contrast, an appealing feature of the multiplier method is the minimal need for sophisticated linear algebra and multidimensional calculus.

A direct solution of the minimax problem in (6.1.1) uses a sequence of least- p th error minimizations so that $\|e\|_p \rightarrow \|e\|_\infty$ as $p \rightarrow \infty$ [Bandler, 1972b]; see (6.4.11). Starting with a gradient minimization using $p=2$, sequential minimizations are restarted with $p=4, 16, 256, \dots$. To avoid eventual ill-conditioning and over- and under-flows, the least- p th objective function is normalized to the largest sampled error, and extrapolation of p to infinity is employed [Bandler, 1975]. This approach is limited because the minimax solution is only achieved when $p \rightarrow \infty$, similar to the original SUMT difficulties with the earlier exterior and interior penalty functions [Fiacco].

6.4.6 Frequency Sampling Strategy

To return to the general problem of minimizing envelope functions cited in Section 6.1.1, recall (6.1.1):

$$\text{Minimize } f(x) = \max_{i \in I} \rho_i(x) \quad \text{for } x \in R^n, \quad (6.4.19)$$

where \in means “belongs to”, and I is some infinite or finite set. In the broadband matching problem, I is a set of frequencies, ρ_i is the reflectance (magnitude of a generalized reflection coefficient), and n is the number of matching-network element values related to the variables in vector x . The name *semi-infinite programming (SIP)* has been applied to (6.4.19). From Figure 5.6.1, which relates to an example having frequencies sampled at intervals of 0.1, one can imagine that some of the function’s peaks and valleys might escape detection for samples spaced too widely apart.

One solution is to sample by finite one-dimensional grids of successively finer mesh sizes to guarantee convergence to a solution of the semi-infinite problem [Hettick]. It has been suggested that, on the domain $[0, 1]$, normalized frequencies be selected by

$$\Omega = \left\{ 0, \frac{1}{q}, \frac{2}{q}, \dots, \frac{(q-1)}{q}, 1 \right\}, \quad (6.4.20)$$

where q , a positive integer, is progressively increased [Zhou]. Typically, if $q >> n$, only relatively few of the many smooth constraints are active (binding) at the solution. Only $n+1$ data points are required to fit exactly an n th-degree polynomial, and more sophisticated interpolation theory also can be applied as a sampling guide. The same could be said for quadrature (numerical integration) theory, especially Gaussian integration [Cuthbert, 1987:201].

Note that these constraint functions are “sequentially related” in that small frequency differences do not substantially alter the functions; see Figure 5.4.2, for example. That relationship has led to both a direct search scheme, called RIPPLE [Bandler, 1969] and a gradient method

applied only to the larger ripples [Bandler,1972a]. An iterative procedure, with the designer closing the loop, also has been suggested: Pick a set of samples, observe the characteristics of the solution, then pick subsequent other sets of samples for revised solutions [Waren].

A much more thoughtful approach examines the condition number of the normal matrix, $J^T J$, in (6.4.18) [Zhang]. Zhang noted that a least-squares solution for n variables employing at least $m=2n+1$ samples is always possible, but the normal matrix might be far from the unit matrix, I , by distance $d \equiv \|J^T J - I\|$. Starting with many more samples, m , than variables, n , and using the Frobenius norm, sets of row vectors from J can be tested to eliminate those rows (samples) causing a large distance, d . That is repeated until there are only n rows left so that J is nearly orthogonal.

Nevertheless, selecting a fairly uniform distribution of $n+1$ to $2n$ frequency samples and checking the solution obtained is a reasonable engineering technique.

6.5 Summary of GRABIM in Detail

The mathematical details for programming the grid approach to broadband impedance matching (GRABIM) are summarized. The general semi-infinite programming problem and the ordinary minimax problem with a finite number of frequency samples are outlined in terms usually employed in current technical literature. The transducer power function is selected for the specific GRABIM problem because it is well scaled and easy to manipulate mathematically, especially where partial derivatives are involved.

Orchard's ladder network analysis method is employed for both the transfer function and all its derivatives. The primary space selected for optimization is the natural logarithm of element variable space, $x = \ln(v)$, because of its excellent scaling and inherently positive element values. Equations are summarized for all partial derivatives of the transducer power transfer function wrt chain parameters and ultimately with respect to x , where x is related logarithmically to each of the five kinds of element variables: L , C , t^2 , Z_0 , and θ_0 . Those equations require the numerical values relating the derivatives of the network's ABCD parameters to the derivatives of the element parameters. Details of that algorithm are provided. Procedures for efficient computation and various strategies to minimize run time are described.

The 2^n -factorial grid search is related to current research of pattern search methods; the hypercube lattice is described as a member of that class. The basis and generating matrices are defined so that efficient precomputation can reduce the search effort. The discretization of X space within the range of all feasible solutions is shown to define 401 points on the line $-3.125 \leq x_j \leq +3.125$ in each x_j coordinate. Although very

few of those points are explored, it is emphasized that initial exploratory points in the pattern are spaced widely apart to avoid minor anomalies in the objective's envelope surface. Even 10 element variables could be subjected to a grid search in less than 30 seconds, using a 200 MHz PC.

The approximate minimax solution obtained by the grid search is a mandatory starting point for the precise solution of the equivalent constrained optimization problem available by the method of multipliers. That problem is framed as in current literature and again related to the specific GRABIM application. The major role played by quadratic (exterior) penalty functions is emphasized, to make it easier to appreciate the role of the Lagrange multipliers, which are central to the method. The augmented Lagrangian penalty function is defined with specific references to methods and the rationale for adjusting the two sets of algorithmic parameters, the weights and the offsets. The sequential unconstrained minimization technique (SUMT) algorithm is described in a table, and reference is made to a specific flow chart in the literature.

This SUMT algorithm has an outer loop for adjustment of weights and offsets by a simple iteration that converges linearly. The important inner loop is a Gauss-Newton unconstrained optimizer that converges quadratically. Its advantages over alternative methods are listed and the main mathematical features defined. It utilizes the exact first partial derivatives obtained from network analysis. The specific transfer of numerical results is indicated, which involves the Jacobian matrix and the related normal matrix to approximate the second partial derivatives in a very significant way. The Gauss-Newton technique and the automatic step control and scaling by the modified Levenberg-Marquardt technique are outlined.

Alternative methods for solving the constrained optimization problem are mentioned, especially the barrier (interior) penalty method. The main advantage of the method of multipliers is that no second derivatives are required. The other methods not only lack that advantage, but have more sophisticated and perhaps obscure roots in linear algebra and multi-variable calculus.

Finally, the strategy for determining the number and location of frequency samples is discussed. Several different schemes in the technical literature are mentioned, including simply involving the intuition and judgment of the designer after evaluation of a nominal design. It is concluded that selecting at least $n+1$ samples for n variables, and perhaps as many as $2n$ samples, usually suffices if the solution is tested to discover any shortcomings.

Appendix A — Circle to Circle Mapping

Bilinear transformations occur throughout network theory, mainly because circuit connections require bilinear operations, i.e. inversion, rotation, and scaling of complex quantities. Even conversion relationships among impedance, admittance, and scattering formalisms are bilinear. Bilinear mappings from a rectangular plane into and between unit circles are utilized in two ways in this book:

- Network input reflection behavior as a function of each element parameter (Section 5.4.3), and
- Impedance neighborhood size as a function of radial location in a unit circle.

The following development extends that previously presented [Cuthbert, 1983:370-2].

A.1 Bilinear Mapping

All bilinear functions that map the Z domain into the w range have the form

$$w = \frac{a_1 Z + a_2}{a_3 Z + 1} = T + U \frac{Z - Z_C}{Z + Z_C^*}, \quad (\text{A.1.1})$$

where all the quantities are complex and the asterisk (*) denotes conjugation⁶ [Cuthbert, 1983:242]. The function w might be a network's generalized input reflection coefficient, and Z might be the impedance of one branch at a particular frequency. Three pairs of (w, Z) complex data are sufficient to determine the three complex coefficients a_1 , a_2 , and a_3 . There are simple relationships for U , T , and Z_C as functions of the three a_i coefficients. Note that the rational bilinear expression following the complex constant U is also a generalized reflection coefficient.

Bilinear transformations map circles into circles, and lines are included as circles with infinite radius. Bilinear transformations also preserve angles of intersection.

A.1.1 Right-Half Plane to Circle Mappings

Figure A.1.1 shows a rectangular Z plane and two reflection planes, f and g , each containing a unit circle about its origin. Functions $F(Z)$ and $G(Z)$ map the right-half Z plane (RHP) into unit circles in the f and g planes, respectively. Those two transformations are

$$f = F(Z) = \frac{Z - Z_f^*}{Z + Z_f} \equiv Re^{j\theta} \quad (\text{A.1.2})$$

⁶ In this appendix, the * superscript conjugates complex values, in contrast to Chapter Six.

$$g = G(Z) = \frac{Z - Z_g^*}{Z + Z_g} \quad (\text{A.1.3})$$

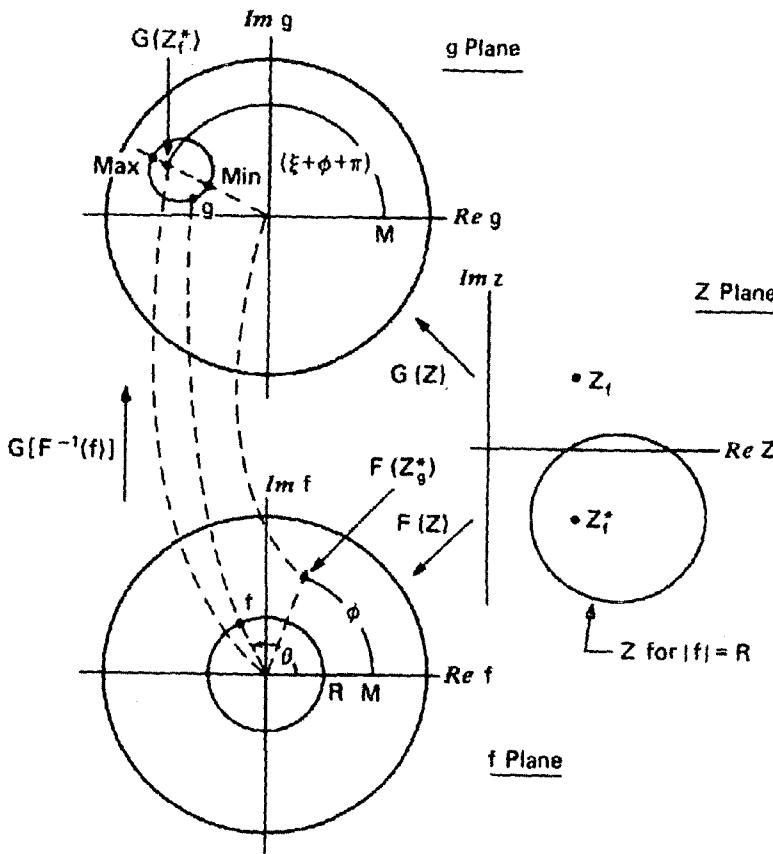


Figure A.1.1. Bilinear transformations between the Z plane and unit circles.

The real parts of the two normalizing impedances, Z_f and Z_g , must be strictly positive (nonzero). The Z plane in Figure A.1.1 shows the point Z_f and its conjugate, Z_f^* . These mappings define generalized Smith charts. Because these mappings are 1:1 and reciprocal, the left-half Z plane maps into the regions outside the unit circles in the f and g planes. The centers of the unit circles are the points where $f=0$ and $g=0$, implying that those points represent Z_f^* and Z_g^* , respectively. In particular, these transformations map the $\text{Re } Z = 0$ line (the $\text{Im } Z$ axis) onto unit circles.

Reciprocal functions are easily derived. For example, (A.1.2) can be solved for Z :

$$Z = F^{-1}(f) = \frac{Z_f^* + f Z_f}{1 - f} \quad (\text{A.1.4})$$

A.1.2 Circle to Circle Mapping

Bilinear functions that map unit circles into unit circles must have the form [Churchill]

$$g = E(f) = e^{j\gamma} \frac{f - f_0}{1 - f f_0}, \quad (\text{A.1.5})$$

where $|f_0| < 1$. Although (A.1.3) shows the g-plane origin, (A.1.5) also shows that it corresponds to $f=f_0$. Therefore,

$$f_0 = F(Z_g^*) = \frac{Z_g^* - Z_f^*}{Z_g^* + Z_f^*} \equiv M e^{j\phi}. \quad (\text{A.1.6})$$

Figure A.1.1 shows that $f=f_0=F(Z_g^*)$ in the f plane is the image of the g-plane origin. Also note that

$$M = |F(Z_g^*)| = |G(Z_f^*)|. \quad (\text{A.1.7})$$

A.2 Interior Circular Images

Most useful bilinear transformation features usually involve corresponding images of lines and/or circles that appear in the domain and range.

A.2.1 Concentric Circle in a Unit Circle

Consider the concentric circle about the f-plane origin in Figure A.1.1. According to (A.1.2), it is the locus of all points where $|f|=R<1$. The inverse transformation in (A.1.4) can be used to show that the concentric f-plane circle has an image in the Z plane that has geometric symmetry about the line through Z_f^* in the direction of $\text{Re } Z$. See Figure A.1.1: the Z-plane image is reflected about $Z=Z_f^*+\sigma$, where σ is real and $-\infty < \sigma < +\infty$.

The Z-plane circle in Figure A.1.1 corresponding to $|f|=R$ is commonly observed when Z_f is purely real; then the circle's center is on the real axis, the f plane is an ordinary Smith chart normalized to some resistance R_f , and the two corresponding images in the f and Z planes are said to be constant standing-wave ratio (SWR) circles. It is not surprising that in the general case the "SWR" circle is displaced vertically by exactly $-Im Z_f$, where "Im" means "the imaginary part of". For $Z=R+jX$ and $Z_f=R_f+jX_f$, the imaginary parts of both the numerator and denominator of (A.1.2) are just $X+X_f$. When Z_f is real, $X_f=0$.

A.2.2 Eccentric Circle in a Unit Circle

Another important image of the concentric circle about the f-plane origin appears in the g-plane in Figure A.1.1. The eccentric circle may or may not encircle the g-plane origin. The *central point* $G(Z_f^*)$ is the image of the f-plane origin, $f=F(Z_f^*)=0$; see Figure A.1.1. The angle of the radial through the g-plane image is shown in Figure A.1.1, where

$$\xi = 2 \arg \left(Z_g^* + Z_f \right). \quad (\text{A.2.1})$$

The g-plane image is symmetrical about this radial. The maximum g-plane image modulus is [Cuthbert, 1980]

$$|g|_{\max} = \frac{M + R}{1 + MR}, \quad (\text{A.2.2})$$

where R is fixed as in (A.1.2) and M is found from (A.1.7). The minimum g-plane image modulus is

$$|g|_{\min} = \frac{M - R}{1 - MR}. \quad (\text{A.2.3})$$

A.2.3 Neighborhood Parameters

The Thevenin central-point impedance $G(Z_f^*)$ in the G plane can be re-normalized to the origin of the f plane. Next, Z_f^* can be determined from the size and location of the reflection-chart circle image, which is an *impedance neighborhood*.

For example, suppose that the reflection circle in the g plane in Figure A.1.1 is normalized to $50+j0$ ohms. Then the g-plane eccentric circle corresponds to the concentric circle about Z_f^* in the Z plane as in Figure A.1.1, and the f-plane concentric circle has an SWR circle image (not shown) in the Z plane, centered on the $\text{Re } Z$ axis. The neighborhood of impedances in the g plane can be transformed to an SWR neighborhood about 50 ohms in the f and Z planes by a lossless network matching impedance Z_f^* to $50+j0$ ohms. This transformation exploits the conservation of power (invariant reflectance), described in Section 2.1.4, and is a statement of the invariant Poincaré metric used by mathematicians [Helton].

Incidentally, this concept can also be extended to finding the normalizing load impedance Z_f^* that allows two given load impedances, Z_{on} and Z_{off} , to produce a given reflection phase change while maintaining the same reflection magnitude. Then, designing the lossless network matching Z_f^* to $50+j0$ ohms is a part of designing hybrid phase shifters, because the phase change is preserved at both ports of the network [Atwater].

The first of three relationships is provided to simplify finding radius M when the location and size of the eccentric circle in the g-plane of Figure A.1.1 are given. The parameters are R_{cen} , the vector length to the arithmetic center of the image, and R_{cir} , the radius of the image circle; see Figure 5.4.5. Eliminating R from (A.2.2) and (A.2.3) yields

$$M = B - \sqrt{B^2 - 1} \quad \text{where } B \equiv \frac{1 + R_{\text{cen}}^2 - R_{\text{cir}}^2}{2R_{\text{cen}}}, \quad R_{\text{cen}} > 0. \quad (\text{A.2.4})$$

This function is plotted in Figure A.2.1. When $Z_g=50+j0$, the g-plane is an ordinary Smith chart. To match a neighborhood of impedances optimally at a frequency, M is found from (A.2.4) to locate the point

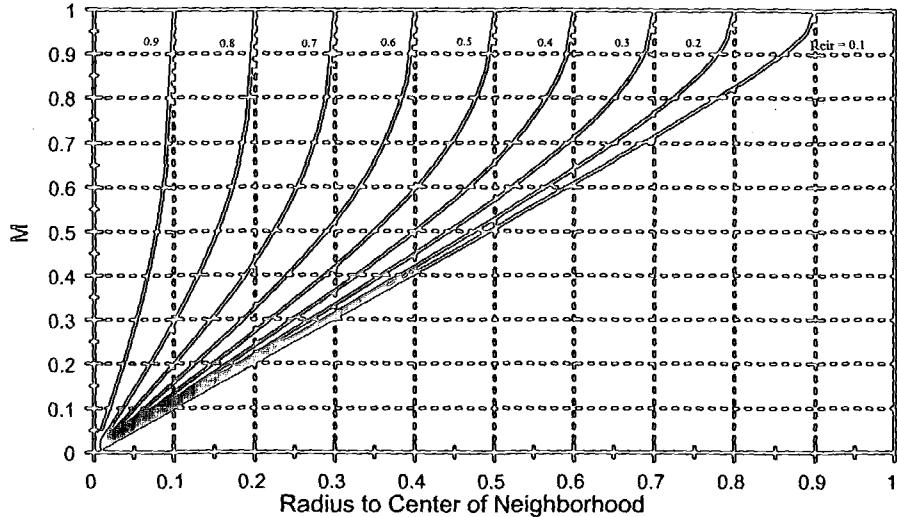


Figure A.2.1. Radius M in the g plane versus R_{cen} and R_{cir} .

$G(Z_f^*)$; see Figure A.1.1. Then Z_f^* is read directly from the R and X chart loci at that point. Any lossless network that matches Z_f to $50+j0$ at that frequency will transform the given neighborhood of impedances into an SWR circle having minimum radius R in the f-plane, Figure A.1.1.

The second relationship provides the radius of the matched SWR circle, R, in the f-plane Smith chart. Again, (A.2.2) and (A.2.3) are used to find R_{cir} in terms of R and M and then to find

$$R = \frac{1}{M} \left(\sqrt{K^2 + 1} - K \right), \quad \text{where } K \equiv \frac{1 - M^2}{2MR_{cir}}, \quad R_{cir} > 0. \quad (\text{A.2.5})$$

This function is plotted in Figure A.2.2 and shows R versus R_{cen} with families of R_{cir} . That shows that $R > R_{cir}$, which is increasingly true as R_{cen} increases. Nevertheless, reflection magnitude R is the least possible radius of a matched concentric neighborhood.

The third relationship is an alternative view of $R > R_{cir}$, obtained by solving

$$M = \frac{(R^2 - 1) + \sqrt{(R^2 - 1)^2 + 4R_{cen}^2 R^2}}{2R_{cen} R^2}, \quad \text{and} \quad (\text{A.2.6})$$

$$R_{cen} = \frac{R(1 - M)}{1 - M^2 R^2}.$$

Figure A.2.3 shows R_{cir} versus R_{cen} with families of constant R.

Reflection magnitude R can be converted to SWR using

$$\text{SWR} = \frac{1 + R}{1 - R}. \quad (\text{A.2.7})$$

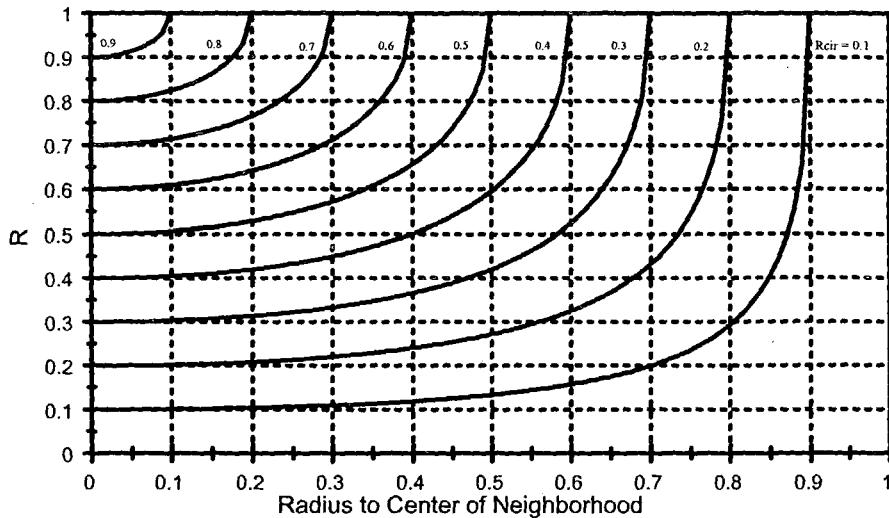


Figure A.2.2. Radius R in the f plane versus R_{cen} and R_{cir} .

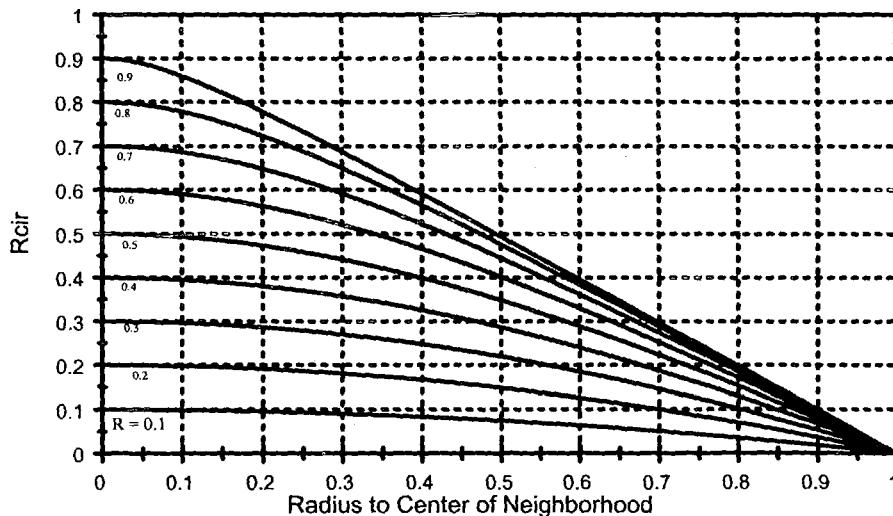


Figure A.2.3. Neighborhood radius R_{cir} versus R_{cen} and f -plane radius R .

The method in this section can also be applied to conjugate matching to a complex source as well. In that case, the SWR in (A.2.7) is a valid mapping but does not relate to standing waves on a transmission line.

Broadband impedance matching does not match exactly at any frequency in the band. Therefore, the matching to complex source impedances is inexact, resulting in a cluster of circles of different sizes at different frequencies, each somewhat off center from the complex Smith chart origin. The goal is to contain these circles within the smallest

possible circular corral centered on the chart origin. See the numerical example in Section 5.6.3.

Example A.2.1. A neighborhood of impedances having a radius of 0.306 is centered at a radius of 0.269 on the 118° radial on a $50+j0$ ohm Smith chart. *Problem:* Find impedance Z_f that must be matched to $50+j0$ ohms and the SWR about 50 ohms when matched. *Solution:* Refer to Figure A.2.1 with $R_{cen}=0.269$ and $R_{cir}=0.306$ to find $M=0.30$. Plot 0.30 at angle 118° on a 50-ohm Smith chart and read the $R+jX$: $Z_f^*=33.2+j19.9$ ohms. Therefore, a lossless network that matches $33.2+j19.9$ to $50+j0$ is required. Enter Figure A.2.2 with $R_{cen}=0.269$ and $R_{cir}=0.306$ to find $R=0.33$; according to (A.2.6), that corresponds to a SWR of 2.0:1. The conclusion is that the given neighborhood of impedances can be centered on $50+j0$ ohms within a 2:1 SWR circle at that frequency.

A.3 Planar Lines to Circles

So far only circles in the Z plane are considered. There are two important cases of lines in the Z plane, other than the jX axis, to consider.

A.3.1 Constant Resistance

See Figure A.3.1. In the Z plane, point Z_f^* is mapped to the center

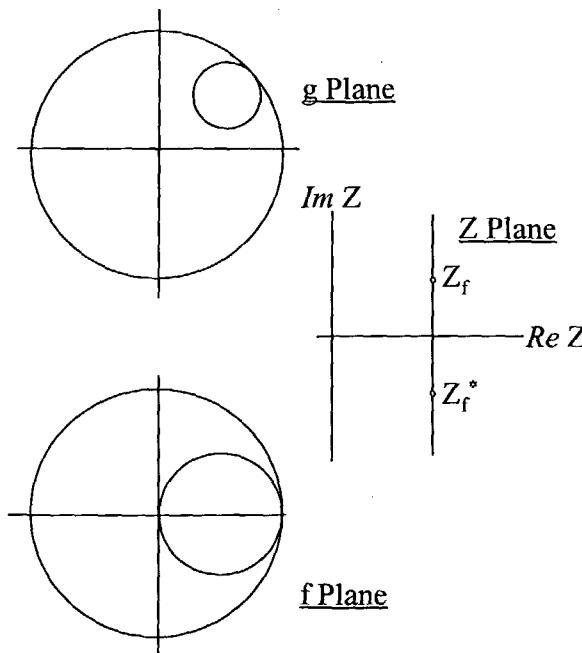


Figure A.3.1. Constant resistance loci about Z_f^* in three planes.

of the f plane. It is important to locate the locus in the f plane when a pure reactance is added to or subtracted from Z_f^* , i.e. the locus of $\operatorname{Re} Z = R_f$. Note that (A.1.2) shows that the point $Z = j\infty$ maps into $f=1+j0$, and that the f -plane locus must be a circle, as shown in Figure A.3.1. The f -plane circle must map to the g -plane so as to be tangent at some point on the unit circle, as in Figure A.3.1, a situation distinctly different from the interior circle image in Figure A.1.1. These two points in (A.2.3) show that $g_{\min}=M$. Furthermore, the area between $\operatorname{Re} Z = R_f$ and $\operatorname{Re} Z = 0$ maps inside the f -plane circle. Therefore, (A.1.6) shows that f_0 is inside that region when $R_g < R_f$, which is the condition for encirclement of the g -plane origin.

A.3.2 Constant Q

The last case considers an impedance of variable magnitude and constant angle, i.e. constant Q , as shown in Figure A.3.2. The line in the

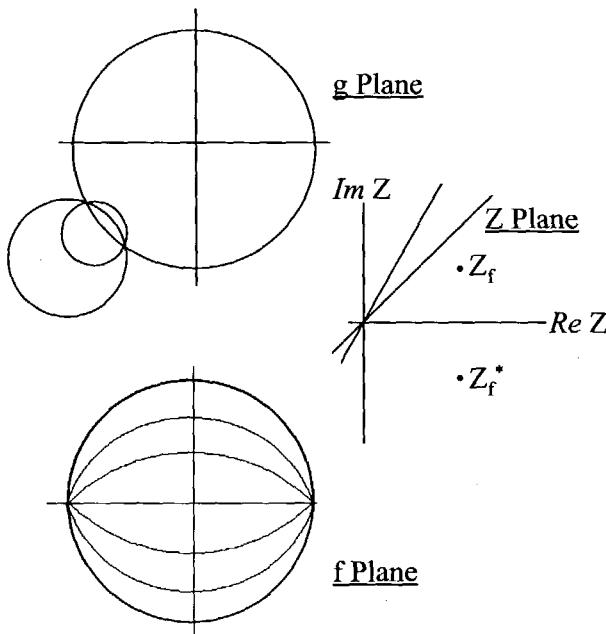


Figure A.3.2. Images of impedances having constant Q in three planes.

Z plane starts at $Z=0$ and extends to $Z \rightarrow \infty$, two points that appear in the f plane where the unit circle intersects the real axis. The circular arcs in the f plane are well known as lines of constant Q in the Carter charts, the less well-known polar variation of a Smith chart. The corresponding images in the g plane are circles that are partially within the unit circle and have intercept points in common.

A.3.3 Eccentric Vector Magnitude

See Figure 5.4.5. It is informative to quantify the behavior of $|\rho|$ as the vector that terminates on the g-plane circular images in Figures A.1.1, A.3.1, and A.3.2 versus variable f-plane angle, Z-plane reactance, and impedance magnitude, respectively. This would not be difficult if the three a_i constants in (A.1.1) were known. In most cases, two pairs of (Z, w) data are readily available, e.g. $Z=0$ and $Z=\infty$ in Figure A.3.2. Certainly a third data pair could be obtained, so that the exact bilinear transformation at that frequency would be known. Then $|\rho(Z)|$ would be determined for all values of Z , however constrained, i.e. on an SWR circle in Z , or on $\text{Im } Z$, or on $|Z|$. Otherwise, it may be sufficient to know that there are certain extrema of $|\rho|$ and that $|\rho|$ varies in a benign manner, especially that there is a monotone relationship between reactance, X , and angle ϕ in Figure 5.4.4. See Section 5.4.3.

Appendix B — Abbreviations and Symbols

wrt	With respect to
\in	Belongs to
\forall	For all
s.t.	Subject to
\wedge	Filter coupling reactance
Λ_y	$\frac{\partial}{\partial y}$, Partial derivative operator
Re	Real part of a complex number
Im	Imaginary part of a complex number
\setminus	Integral division operator
\rightarrow	Approaches
∞	Infinity
Π	Product operator
ω	Radian frequency, rad/sec
ω_0	Reference frequency; usually unity
ω_i	i^{th} frequency sample
θ	Angle
θ_0	Angle at reference frequency ω_0
α	Generalized reflection coefficient
ρ	Reflection coefficient
Γ	Reflection coefficient
λ	Lagrange multiplier
ϕ	Angle
$*$	Complex conjugate operator (superscript)
$F(x)$	Scalar function of a vector variable
g_i	i^{th} prototype network element (except Chapter Six)
g_i	Goal (target) at the i^{th} frequency (Chapter Six)
P_i	Inverted transducer power function at frequency ω_i
u_i	Lagrangian residual offset variable at frequency ω_i
v	Column vector of element values in the V space
v_k	k^{th} element in column vector v
w_i	Penalty function weight on i^{th} residual at frequency ω_i
x	Column vector of element values in the X space
x^k	Value of x at k^{th} iteration
x^0	Initial vector of starting values for optimization
x'	Variables vector after a minimization in SUMT
x^*	Final optimal value of variables vector
Z_S	Source impedance
Z_L	Load impedance
Z_0	Characteristic impedance of a uniform transmission line

References

- Abrie, P. L. D. (1985). *The Design Of Impedance-Matching Networks For Radio-Frequency And Microwave Amplifiers*. Boston: Artech House.
- Abrie, P. L. D. (1991). The transformation-Q approach to synthesizing lumped-element impedance-matching networks. *Suid-Afrikaanse Tydskrif vir Wetenskap*, V89. March/April: 147-149.
- Amstutz, P. (1978). Elliptic approximation and elliptic filter design on small computers. *IEEE Trans. Circuits and Systems*, pp. 1001-1011.
- Atia, A. E. and A. E. Williams (1975). Measurements of intercavity couplings. *IEEE Trans. Microwave Thy. Tech.*, V23, N6. June:519-522.
- Atwater, H. A. (1980). Reflection coefficient transformations for phase-shift circuits. *IEEE Trans. Microwave Thy. Tech.*, V28, N6. June:563-568.
- Bandler, J. W. and P. A. McDonald (1969). Optimization of microwave networks by razor search. *IEEE Trans. Microwave Thy. Tech.*, V17, N8. Aug.:552-563.
- Bandler, J. W., T. V. Srinivasan, and C. Charalambous (1972a). Minimax optimization of networks by grazor search. *IEEE Trans. Microwave Thy. Tech.*, V20, N9. Sept.:696-604.
- Bandler, J. W. and C. Charalambous (1972b). Practical least pth optimization of networks. *IEEE Trans. Microwave Thy. Tech.*, V20, N12. Dec.:834-840.
- Bandler, J. W. and W. Y. Chu (1975). *Nonlinear Programming Using Least Pth Optimization with Extrapolation*. Hamilton, Ontario, Canada: McMaster University, Report No. SOC-78.
- Bertsekas, D. P. (1982). *Constrained Optimization and Lagrange Multiplier Methods*. Belmont, MA: Athena Scientific.
- Blinchikoff, H. J., and A. I. Zverev (1976). *Filtering in the Time and Frequency Domains*. NY: John Wiley.
- Borlez, Y. (1996). A simple Chebychev bandpass impedance-matching program. *RF Design*. Oct.: 80-89.

- Carlin, H. J. (1977). A new approach to gain-bandwidth problems. *IEEE Trans. Circuits and Systems*, V24N4. April:170-175.
- Carlin, H. J. and B. S. Yarman (1983). The double matching problem: analytic and real frequency solutions. *IEEE Trans. Circuits and Systems*, V30N1. Jan.:15-28.
- Carlin, H. J. and P. P. Civalleri (1992). An algorithm for wideband matching using Wiener-Lee Transforms. *IEEE Trans. Circuits and Systems*, V39N7. July:497-505.
- Carlin, H. J. and P. P. Civalleri (1998). *Wideband Circuit Design*. NY: CRC Press.
- Chen, W. K. (1977). Synthesis of optimum Butterworth and Chebyshev broadband impedance matching networks. *IEEE Trans. Circuits and Systems*, V24: April:152-169.
- Christian, E. (1975). *Introduction to the Design of Lossless Transmission Networks*. Raleigh, NC: by author.
- Christian, E. and E. Eisenmann (1977). *Filter Design Tables and Graphs*. Raleigh, NC: by author.
- Churchill, R. V. (1960). *Complex Variables and Applications*. NY: McGraw-Hill.
- Cohn, S. B. (1957). Direct-coupled resonator filters. *Proc. IRE*, V45. Feb. 187-196.
- Cohn, S.B. (1959). Dissipation loss in multiple-coupled resonator filters. *Proc. IRE*, V47. Aug.: 1342-1348.
- Conn, A. R., N. I. M. Gould, and P. L. Toint (1991). A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM Jour. Numerical Analysis*, 28, pp. 545-572.
- Courant, R. (1943). Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math Soc.*, V49. Pages 1-23.
- Cuthbert, T. R. (1980). *Filter Load Effects, Absorption, and Loaded-Q Design*. Dallas: Southern Methodist University Ph.D.Dissertation

- Cuthbert, T. R. (1983). *Circuit Design Using Personal Computers*. NY: John Wiley. Also, Melbourne, FL: Krieger Publishing Co. (1994).
- Cuthbert, T. R. (1987). *Optimization Using Personal Computers with Applications to Electrical Networks*. NY: John Wiley.
- Cuthbert, T. R. (1994a). Broadband impedance matching methods. *RF Design*. Aug.:64-71.
- Cuthbert, T. R. (1994b). Broadband impedance matching – fast and simple. *RF Design*. Nov.:38-50.
- Cuthbert, T. R. (1996). Wideband direct-coupled filters having exact response shapes. *IEEE Trans. Microwave Thy. Tech.*, MTT-S Newsletter, Spring: 27-38.
- Cuthbert, T. R. (1997). Wide-Band Filters with Wide-Ranging Choices. *RF Design*. July:48-61.
- Daniels, R. W. (1974). *Approximation Methods for Electronic Filter Design*. NY: McGraw-Hill.
- Dedieu, H., C. Dehollain, J. Neirynck, and G. Rhodes (1994). A new method for solving broadband matching problems. *IEEE Trans. Circuits and Systems*, V41N9. Sept.:561-571.
- Dennis, J.E. and V. Torczon (1991). Direct search methods on parallel machines. *SIAM J. Optimization*, V71,N4. Nov.:448-474.
- Dishal, M. (1949). Design of dissipative band-pass filters producing desired exact amplitude-frequency characteristics. *Proc. IRE*, V37. Sept.:1050-1069.
- Dishal, M. (1951). Alignment and adjustment of synchronously tuned multiple-resonant-circuit filters. *Proc. IRE*, V39. Nov.:1448-1460.
- Dishal, M. (1965). A simple design procedure for small percentage bandwidth round-rod interdigital filters. *IEEE Trans. Microwave Thy. Tech.*, V13, N5. Sept.:696-698.
- Drozd, J. M. and W. T. Joines (1996). Determining Q using S parameters. *IEEE Trans. Microwave Thy. Tech.*, V44, N11. Nov.:2123-7.
- Fano, R. M. (1950). Theoretical limitations on the broadband matching of arbitrary impedances. *J. Franklin Inst.* Feb.:139-154.

- Fiacco, A. V. and G. P. McCormick (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. NY: John Wiley.
- Fettweis, A. (1979). Parametric representation of Brune functions. *Ckt. Thy. & Applis.*, V7:113-119.
- Fletcher, R. (1974). Methods related to Lagrangian functions, in *Numerical Methods for Constrained Optimization*, P.E. Gill and W. Murray, Eds. NY: Academic. Pages 219-227.
- Fletcher, R. (1980-1981a). *Practical Methods of Optimization: Unconstrained Optimization*, Vol.1, and *Constrained Optimization*, Vol.2. New York: John Wiley.
- Fletcher, R. (1981b). Methods for nonlinear constraints. In *Nonlinear Optimization 1981*. NY: Academic, pages 185-211.
- Forster, U. and R. Behauen (1980). Comments on parametric representations of Brune functions. *Int. J. Circuit Theory Appl.*, V8.
- Frontline Systems, Inc., Incline Village, NV. (Developers of the Microsoft EXCEL® SOLVER). Internet Address: <http://www.frontsys.com>
- Geffe, P. R. (1963). *Simplified Modern Filter Design*. NY: Rider.
- Geffe, P. R. (1997). LC filter design, in *CRC Handbook of Electrical Filters*, J. T. Taylor and Q. Huang, Eds. NY: CRC Press: 45-79.
- Gill, P. E., W. Murray, and M. H. Wright (1981) *Practical Optimization*. NY: Academic.
- Green, E. (1954). *Amplitude-Frequency Characteristics of Ladder Networks*. Chelmsford, U.K.: Marconi
- Ha, T. T. (1981). *Solid-State Microwave Amplifier Design*. NY: Wiley.
- Helton, J. W. (1981). Broadbanding: gain equalization directly from data. *IEEE Trans. Circuits and Systems*, V28N12. Dec.:1125-1137.
- Hestenes, M. R. (1969). Multiplier and gradient methods, *J. Optim. Theory Appl.*, V4. Pages 303-320
- Hettick, R. and K. O. Kortanek (1993). Semi-infinite programming: theory, methods, and applications. *SIAM Review*, V35, PP. 380-429.

- Himmelblau, D. M. (1972). *Applied Nonlinear Programming*. NY:McGraw-Hill.
- Humpherys, D. S. (1970). *The Analysis, Design, and Synthesis of Electrical Filters*. Englewood Cliffs, NJ: Prentice-Hall.
- Iobst, K. W. and K. A. Zaki (1982). An optimization technique for lumped-distributed two ports, *IEEE Trans. Microwave Thy. Tech.*, V30, N12. Dec.:2167-2171
- ITT (1975). *Reference Data for Radio Engineers*, Sixth Edition. NY: H. W. Sams.
- Johnson, R. A. (1983). *Mechanical Filters in Electronics*. NY: Wiley. pp184-6.
- Jung, W. and W. Chiu (1993). Stable broadband microwave amplifier design using the simplified real frequency technique. *IEEE Trans. Microwave Thy. Tech.*, V41, N2. Feb.:336-340.
- Kajfez, D and E. J. Hwan (1984). Q-factor measurement with network analyzer. *IEEE Trans. Microwave Thy. Tech.*, V32, N7. July:666-670.
- Kim, H. K. and E. Kim (1981). Generation of equivalent Cauer-type canonic ladder networks. *IEEE Trans. Circuits & Sys.*, V28, N10. Oct.:1004-1006.
- Kintscher, J. and R. Unbehauen (1997). Approximation methods, in *CRC Handbook of Electrical Filters*, J. T. Taylor and Q. Huang, Eds., pp.3-18. NY:CRC Press.
- Kurokawa, K. (1965). Power waves and the scattering matrix. *IEEE Trans. Microwave Thy. Tech.*, V13, N2. March:194-202.
- Levy, R. (1964). Explicit formulas for Chebyshev impedance-matching networks. *Proc. IEE*. June:1099-1106.
- Levy, R. (1995). Direct synthesis of cascaded quadruplet (CQ) filters. *IEEE Trans. Microwave Thy. Tech.*, V43, N12. Dec.:2940-45.
- Matthaei, G. L., L. Young, and E. M. T. Jones (1964). *Microwave Filters, Impedance Matching Networks, and Coupling Structures*. New York: McGraw-Hill. Also, Boston: Artech House (1980).

- Mellor, D. J. (1975). *Computer-aided synthesis of matching networks for microwave amplifiers*. Stanford Electronics Lab., CA. Dissertation: University Microfilm #75-21887.
- Microsoft EXCEL® User's Guide for Version 5.0. (1993). Chapters 27-30.
- Montgomery, D. C. (1984). *Design and Analysis of Experiments*, 2nd Ed. NY: John Wiley.
- Ness, J. B. (1998). A unified approach to the design, measurement, and tuning of coupled-resonator filters. *IEEE Trans. Microwave Thy. Tech.*, V46, N4. April:343-351.
- Orchard, H. J. (1985). Filter design by iterated analysis. *IEEE Trans. Circuits & Sys.*, V32, N11. Nov.:1089-1096.
- Pardalos, P. M. (1994). On the passage from local to global in optimization, in *Mathematical Programming State of the Art 1994*, J. R. Birge and K. G. Murty, Eds. University of Michigan. Pages 220-247.
- Polak, E. (1987). On the mathematical foundations of nondifferentiable optimization in engineering design. *SIAM Review*, V29, N1, pp. 1-20.
- Polyak, R. (1992). Modified Barrier Functions (theory and methods). *Math. Programming*, V54, pp. 177-222.
- Powell, M. J. D. (1969). A method for nonlinear constraints in minimization problems, in *Optimization*, R. Fletcher, Ed. NY: Academic Press, pp. 283-297.
- Powell, M. J. D. (1995). Some convergence properties of the modified log barrier method for linear programming. *SIAM J. Opti.*, V5N4. Nov.:695-739.
- Rockafellar, R. T. (1973). The multiplier method of Hestenes and Powell applied to convex programming. *J. Optim. Theory Appl.* V12. Pages 555-562.
- Rockafellar, R. Tyrrell (1993). Lagrange multipliers and optimality. *SIAM Review*, V35N2. June:183-238.
- Rockafellar, T. (1994). Nonsmooth Optimization, in *Mathematical Programming State of the Art 1994*, J. R. Birge and K. G. Murty, Eds. University of Michigan. Pages 248-258.

- Ryan, D. M. (1974). Penalty and barrier functions, in *Numerical Methods for Constrained Optimization*, P.E. Gill and W. Murray, Eds. NY: Academic. Pages 186-8.
- Saal, R. and E. Ulbrich (1958). On the design of filters by synthesis. *IEEE Trans. Circuit Thy.*, V5. Dec.:284-317.
- Sabin, W. E. (1992). The elliptic bandpass filter. *RF Design*. June:82-85.
- Skwirzynski, J. K. (1971). On synthesis of filters. *IEEE Trans. Circuit Thy.*, V18, N1. Jan.:152-163
- Sussman-Fort, S. (1987). Approximate direct-search minimax circuit optimization. *Intl J. Numerical Methods in Engineering*, V28. NY: John Wiley, pages 359-368.
- Sussman-Fort, S. (1991). The computer-aided design of microwave matching circuits. *Intl. J. Microwave and Millimeter-Wave Computer-Aided Engineering*, V1, N3, 288-305.
- Szentirmai, G. (1997). Computer-aided design methods in filter design: S/FILSYN and other packages, Chap. 3 in *CRC Handbook of Electrical Filters*, J. T. Taylor and Q. Huang, Eds. NY: CRC Press.
- Taub, J. J. (1963). Design of minimum loss band-pass filters. *Microwave Journal*, Nov.:67-76.
- Taub, J. J. And H. J. Hindin (1964). Minimum insertion loss microwave filters. *Microwave Journal*, Aug.:41-45.
- Temes, G. C. and S. K. Mitra (1973). *Modern Filter Theory and Design*. NY: Wiley.
- Torczon, V. (1991). On the convergence of the multidirectional search algorithm. *SIAM J. Optimization*, V1,N1. Feb.123-145.
- Torczon, V. (1997). On the convergence of pattern search algorithms. *SIAM J. Optimization*, V7,N1. Feb.:1-25.
- Torczon, V. and R. M. Lewis (1999). A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. Submitted to *SIAM Jour. on Optimization*.

- Van Der Walt, P. W. (1986). On optimum Chebyshev broad-band impedance-matching networks. *IEEE Trans. Circuits and Systems*, V33N24. Oct.:1010-1012.
- Walsh, G. R. (1979). *Methods of Optimization*. Chichester, U.K.: Wiley.
- Waren, A. D., L. S. Lasdon, and D. F. Suchman (1967). Optimization in engineering design. *Proc. I.E.E.E.*, V55N11. Nov.:1885-1897.
- Weiss, S. J. and W. K. Kahn (1996). Measurement of simple resonant equivalent circuits for microstrip antennas. *IEEE Trans. Microwave Thy. Tech.*, V44, N6. Aug.:1513-6.
- Wilde, D. J. and C. S. Beightler (1967). *Foundations of Optimization*. Englewood Cliffs, NJ: Prentice-Hall.
- Williams, A. B. and F. J. Taylor (1991). *Electronic Filter Design Handbook*, 2nd Ed. NY: McGraw-Hill.
- Yarman, B. S. (1982). *Broadband matching a complex generator to a complex load*. Ph.D. Thesis, Ithaca, NY: Cornell University.
- Yarman, S. Y. and A. Fettweiss (1990). Computer-aided double matching via parametric representation of Brune functions. *IEEE Trans. Circuits & Sys.*, V37, N2. Feb.:212-222.
- Zhang, X. D., X. N. Hong, and B. X. Gao (1990). An accurate Fourier transform method for nonlinear circuits analysis with multi-tone driven. *IEEE Trans. Circuits & Sys.*, V37, N5. May:668-9.
- Zhou, J. L. and A. L. Tits (1996). An SQP algorithm for finely discretized continuous minimax problems and other minimax problems with many objective functions. *SIAM J. Opti.*, V6N2. May: 461-487.
- Zverev, A. I. (1967). *Handbook of Filter Synthesis*. New York: John Wiley.

Author Index

—A—

- Abrie, P. L. D., 113, 117, 167
 Amstutz, P., 71, 83
 Atia, A. E., 86
 Atwater, H. A., 198

—B—

- Bandler, J. W., 179, 184, 192
 Bertsekas, D. P., 185, 191
 Borlez, Y., 34, 83

—C—

- Carlin, H. J., 44, 61, 91, 96, 98,
 105, 112, 119, 125, 126, 127,
 128, 129, 130, 131, 168
 Chen, W. K., 127
 Christian, E., 22, 38
 Churchill, R. V., 197
 Cohn, S. B., 62
 Conn, A. R., 191
 Courant, R., 186
 Cuthbert, T. R., 1, 19, 24, 30, 36,
 38, 42, 44, 51, 53, 57, 61, 65,
 68, 71, 83, 91, 94, 102, 106,
 109, 112, 116, 117, 118, 120,
 123, 128, 129, 149, 155, 173,
 177, 183, 186, 189, 190, 192,
 195, 198

—D—

- Daniels, R. W., 19, 53, 71, 90, 91,
 92, 94, 95, 96, 97
 Dedieu, H., 162
 Dennis, J. E., 157, 178, 179, 185
 Dishal, M., 48, 50, 84
 Drozd, J. M., 51, 86, 125

—E—

- Fano, R. M., 118, 119, 121, 168
 Fettweis, A., 98, 130
 Fiacco, A. V., 186, 192
 Fletcher, R., 185, 186, 187, 191
 Forster, U., 98, 130

—F—

- Geffe, P. R., 71
 Gill, P. E., 172
 Green, E., 48, 122

—G—

- Ha, T. T., 125, 164, 166
 Helton, J. W., 126, 198
 Hestenes, M. R., 187
 Hettick, R., 192
 Himmelblau, D. M., 184
 Humpherys, D. S., 48

—H—

- International Telephone &
 Telegraph Corp., 14, 48
 Iobst, K. W., 176

—I—

- Johnson, R. A., 50
 Jung, W., 131

—J—

- Kajfez, D., 125
 Kim, H. K., 98
 Kintscher, J., 106, 133
 Kurokawa, K., 11, 42

—L—

Levy, R., 50, 98, 118

—M—

Matthaei, G. L., 35, 51, 57, 79, 85, 122

Mellor, D. J., 99, 121

Microsoft Corp., 151

Montgomery, D. C., 179

—N—

Ness, J. B., 85, 86, 87

—O—

Orchard, H. J., 6, 94, 96, 97, 105, 106, 107, 108, 109, 133, 151, 171, 177, 193

—P—

Polak, E., 171

Polyak, R., 172, 191

Powell, M. J. D., 161, 187, 188, 191

—R—

Rockafellar, R. T., 171, 172, 187

Ryan, D. M., 185

—S—

Saal, R., 99

Sabin, W. E., 71

Skwirzynski, J. K., 97, 99

Sussman-Fort, S., 132, 163

Szentirmai, G., 96, 97

—T—

Taub, J. J., 24, 36

Temes, G. C., 97

Torczon, V., 1, 157, 178, 179, 181, 182, 184, 191

—V—

Van Der Walt, P. W., 121

—W—

Walsh, G. R., 187

Waren, A. D., 191, 193

Weiss, S. J., 125

Wilde, D. J., 146, 181, 184

Williams, A. B., 35

—Y—

Yarman, B. S., 130, 131, 162, 163, 164, 167

—Z—

Zhang, X. D., 193

Zhou, J. L., 171, 172, 191, 192

Zverev, A. I., 19, 22, 33, 34, 35, 38, 48, 49, 50, 71, 81

Subject Index

—A—

- Abbreviations, 204
- ABCD parameters, 40
 - Cascade transmission line, 139
 - Matrix, 41
 - Parallel branch, 139
 - Precomputed, 151
 - Parallel admittance branch, 42
 - Series branch, 41, 139
 - Subnetworks, 41
 - Transformer, ideal, 139
- Aberrations, in minimax
 - reflectance surfaces, 146, 153, 156
- Added variable, 158, 160, 171, 185, 188
- All-pole filter, 19
- Analytic gain-bandwidth theory
 - Overview, 118
- Antimetric filters, 107
- Approximation
 - Piecewise linear function, 127
 - Wiener-Lee series, 128
- Arithmetic symmetry, 86
- Augmented Lagrangian
 - Algorithm, 162
 - Avoiding infinite weights, 161

—B—

- Barrier methods, 191
- Base point, repositioning, 156
- Basis functions, 128
- Bilinear function, 43, 97
 - CASTL tangent of length, 141
 - Composites are bilinear, 141
 - Definition, 195
 - Determined by three complex constants, 203
 - Mapping circles into circles, 197
 - Mapping planes into circles, 195
- Bimodal, reflectance function, 142, 155
- Bode semilog graph
 - Lossless and lossy filter responses, 59
 - Selectivity estimate, 55
- Bounds on variables, 182
- Broadband matching
 - Overview, 133
 - Objective, 134
- Brune functions, 98
- Butterworth response, 19

—C—

- Canonic networks, 98
 - Two classical bandpass topologies, 149
 - Form, 72
- Capacitance, parasitic, 31, 32
- Cartesian plane, 9, 195

- Cascade transmission line
 - ABCD parameters, 139
 - Input impedance, loaded, 117
 - Matching using Z_0 and length given terminations, 117
 - Possible bimodal reflectance versus Z_0 , 142
 - Cascading subnetworks, 41
 - Cauer
 - Filter, 22, 27
 - Realization, 98
 - Topologies, 30
 - Central point of eccentric circle image, 197
 - Chain rule, 174, 175
 - Characteristic function, 91, 106, 120
 - Characteristic impedance, 10
 - Chebyshev
 - Capture property, 91
 - Formula for network element values, 122
 - Response, 18
 - Ripple at dc, 21
 - Classical bandpass filter, 29, 38
 - Clusters of neighborhoods, 200
 - Coefficients
 - Scattering, 42
 - Interpreting in linear equations, 42
 - Comprehensive equal-ripple filters, transmission zeros, 90, 93
 - Computing
 - Math operations count, 176
 - PC speed, 151
 - Conjugate match, 43, 113, 119
 - Conjugation, 9
 - Constant reactance couplings, 48
 - Constraints
 - Active (binding), 136, 172
 - Inequality, 159, 185
 - For positive elements, 77
 - In broadband matching, 134
 - Continued-fraction expansion, 129
 - Conversion
 - Lowpass to bandpass elements, 28, 38
 - Lowpass trap to bandpass, 30
 - S parameters to normalized impedance, 133
 - Series-parallel impedances, 24
 - Coordinate searches, 181
 - Coordinates, integer, 183
 - Coprocessors, speed, 151
 - Coupling, constant reactance, 48
 - Coupling coefficients
 - Bandpass, 48
 - Normalization, 48
 - Lowpass, 48
 - Measured on Smith charts, 87
 - Cross section, reflectance, 100, 136
 - Curvature, 186
 - Cycles of lattice granularity, 183
- D—**
- Decrement, normalized Q, 120
 - Degrees of freedom, 91, 102, 107
 - Denormalizing element value, 39
 - Derivative, partial
 - For all 12 element types in V space, 175
 - Identity, 173
 - Shorthand notation, 173
 - Design of Experiments (DOE), 179
 - Determinant of lossless network, 51
 - Diagonal location factor, 146
 - Diagonals, hypercubes, 146
 - Dipoles, 38
 - Direct search methods
 - Definition, 178
 - Pattern search subset, 179

- Direct-coupled filters
 - All coupling topologies, 71
 - Design Example 3.3.1, 57
 - Example 3.3.5, 62
 - Overview, 47
 - Prior technology, 47
 - Prototype, 49
 - Topologies, 49
 - Tuning, 84
 - Dissipation
 - Effect on analysis speed, 151
 - Effect on input resistance, 61
 - Effect on filters, 49
 - Midband loss, 61
 - Resonator, 26
 - DO loops, 153
 - Double match, 104
 - Anomalies, 37, 123
 - Between active devices, 131
 - Example 5.2.1 requiring source pad, 123
 - Real-frequency technique, 129
 - Source decrement, 123
 - Using Brune functions, 130
 - Duality, 34
- E—
- Efficiency, 15
 - EI section matching
 - Example 5.1.1 for resistances, 114
 - Example 5.1.2 for impedances, 115
 - Element numbering, 27
 - Element Response
 - Overview, 99
 - Single-match Example 4.4.2, 103
 - Eliminating PB Distortion, 65
 - Elliptic-function filter, 27, 71
 - Example 3.5.4, 82
 - Type b, 83
 - Type c, 91
- End coupling
 - Effect on stopband selectivity, 57
 - Example 3.3.2., 59
 - Narrow band, 50
 - Undistorted wide band, 65
 - Envelope function, 171
 - Aberrations or anomalies, 146
 - Definition, 101
 - Reflectance cross sections, 146
 - Unimodal or monotonic, 143
 - Equal-element response, 19
 - Equations, linear, 40, 42
 - Equations, nonlinear, 187
 - Evolutionary Operation (EVOP), 179
 - Exact derivatives, 108
 - Exact penalty function, 191
 - Exploratory step, 182
 - Exterior-point methods, 190
- F—
- Factorial search, 179, 183
 - Feasible direction searches, 191
 - Feasible region
 - Example 3.4.1 for three resonators, 74
 - In parallel resistance space, 74
 - Log graphs, 77
 - Narrow band cases, 54
 - On a line, 73
 - Vanishing elements, 77
 - Filters
 - All-pole, 19
 - Antimetric, 107
 - Classification, 99
 - Degree, 94
 - Symmetric, 106
 - Flat load, 43
 - Flat loss, 36, 91
 - Flowchart, ALLCHEBY, 36
 - Foster realization, 98

—G—

Gain bandwidth, 103, 118, 119
 Gauss-Newton step, 190
 Generalized reduced gradients (GRG), 192
 Geometric mean
 Passband center frequency, 18
 Geometric symmetry, 86
 Of Z-plane circle image, 197
 Trap, 70
 Geometry
 Eccentric circles, 137
 Gewertz procedure, 129, 130
 Global minimum, 153
 Goal (target) loss, 172
 GRABIM
 Algorithms, 148
 Four innovations, 131
 In detail, 171
 Introduction, 112, 131
 Prunes candidate network to full rank, 162
 Thesis, 132
 Granularity, search, 153, 156
 Grid base point, 153
 Grid Search
 Base point, 153
 Convergence, 157
 Detailed overview, 152
 History, 178
 Non-speculative, 157
 Premature failure in frequency scan, 153
 Group delay, 29, 85, 98

—H—

Hilbert transform, 43, 126, 128, 130
 Holds on variables, 182
 Hurwitz polynomial, 131
 Hypercube
 Diagonal, principal, 136, 146
 Sizes in cycles, 181
 Trial pattern Ex. 5.5.1, 156

—I—

Ill-conditioning
 Least-*p*th objective function, 192
 Weights, quadratic penalty function, 186
 Images
 Constant Q, 202
 Constant resistance, 201
 Preserved angles of intersection, 195
 Impedance
 Conjugation, 10
 Neighborhood examples, 134
 Open-circuit parameter, 129
 Positive definite, 130
 Initialization, static, 183
 Input impedance
 Cascade transmission line with load, 117
 Function of ABCD and load, 43
 Inverter with load impedance, 52
 Transmission-line stubs, 150
 Input resistance, computing for Mismatch, 60
 Singly-terminated filters, 60
 Insertion loss definition, 15
 Interior-point (barrier) methods, 191
 Interstage network
 Distributed Example 5.6.2, 164
 Lumped Example 5.6.3, 165
 Inverse Chebyshev response, 22
 Inverter
 Characteristic Z_0 , 52
 Every lossless subnetwork, 53
 Ideal, 49, 51
 Trap, 53
 Iterated analysis
 Accuracy, 109
 Example 4.5.1, 110
 Overview, 105
 Iteration, 179

—J—

Jacobian matrix, 109, 177

—K—

Knot (joint) in envelope, 101, 157, 158

—L—

Lagrange multiplier

Associated with constraints, 172

Related to weights and offsets, 161

Sensitivity (shadow price), 161

Lagrangian function, 186

Laplace s-plane, 49, 130

Lattice

Number of trial points, 155

Two-dimensional example, 5x5, 152

Typical patterns in two-space, 180

Levenberg-Marquardt in Gauss-Newton steps, 190

Likely global solution for three networks from Example 5.6.1, 163

Load, flat, 14

Load Q, 28

Loaded Q

Definition, 25

From input reflection delay, 86

Measuring techniques, 124

Loss at dc, 21

—M—

Mapping

Applications, bilinear, 195

Bilinear, 12

Bilinear inherent in networks, 137

Bilinear, CASTL length parameter, 141

CASTL length from v to θ_0 , 175

Circle to circle, 12

Frequency, 97

Frequency, LP to BP, 29

Half plane to circle, 12

Lowpass to bandpass, 19

Smith chart, 9

Transmission line lengths, 145

Variables in log (dB) space, 145

Variables in domains and ranges, 145

Matching

Conjugate, 9

Impedance neighborhoods, 165

Networks overview, 112

Single frequency, 113

Using cascade transmission lines, 118

Using el sections, 114, 115

Using Pi sections for resistances, 116

Matrix

ABCD of ideal inverter, 51

Basis, 181, 183

Generating, 181, 183

Hessian, 190

LU factorization, 109

Multiplication rules, 42

Normal, 190, 193

Notation, 41

Rotation for orthogonal search, 184

Max operator, 159

Maximally flat, 19

Method of choices, 75

Method of multipliers, 185

Adds a term in objective, 186

Augmented Lagrangian function, 186

Algorithm in detail, 188

Initialization, 188

Midband

Dissipation, 26, 37

Voltages and currents, 31

Loss, Example 2.3.2., 26

Mismatch loss ($R_1 \neq R_2$), 60

Minimax

- General problem and solution statements, 171
 - Objective function, 103, 135
 - Objective with added variable, 160
 - Problem solved by SQP, 191
 - Reformulated with added (float) variable, 158
 - Specific matching problem statement, 172
- Minimax reflectance surface in two-space**, 144
- Minimizer**, Gauss Newton, 189
- Minimum**
 - Reactance, 15, 45
 - Susceptance, 15
 - Inferior in reflectance envelope function, 154
- Minimum-loss response**, 19
- Mismatch loss**, 15
- Multi-directional line search**, 184

—N—

- Neighborhood**
 - Clusters, 200
 - Impedance, 134, 198
 - Global minimum, 153, 157
- Neighborhood, impedance**, 198
- Network**
 - ABCD subnetworks, 175
 - Analysis, 40, 108, 177
 - Augmented for neighborhood matching, 166
 - Automatic description, 148, 150
 - Candidate, 135, 148
 - Dissipative, 14, 19
 - Doubly terminated, 13
 - Double Norton candidates, 150
 - Dual, 27, 32, 34
 - Elements, 12 types, 149
 - Internal reflectance interfaces, 139
 - Linear, definition, 40

- Lossless ABCD parameters**, 42
- Lowpass prototype**, 27
- Pruning branches**, 136
- Reduced-degree, full rank**, 157
- Singly terminated**, 15
- Topologies**, 27
- Voltage/current extremes**, 14, 31
- Newton search**, 189
- Node-voltage phases**, 66, 81, 87
- Nomographs**, filter response, 22
- Nonlinear programming**, 185
- Norton**
 - Equivalent source, 17, 27
 - Transformation, 65
- Null frequencies**, 53, 71, 92
- Assignment order in synthesis**, 99
- Trap reactance in terms of**, 151

—O—

- Optimization, nonlinear**, 109
- Gradient for fast convergence**, 158
- In spreadsheets**, 76
- Newton iteration**, 109, 189
- Non-smooth**, 171
- Over-coupled response**, 19

—P—

- Parallel branch ABCD parameters**, 139
- Parallel resistance space**, 53, 72
- Parameters**
 - ABCD (Chain), 40
 - Scattering, 42
- Passband**
 - Center frequency, 18
 - Distortion due to inverters, 62, 64
 - Edge frequencies, 18, 94
 - Lowpass normalization, 20
 - Peak and valley frequencies, 95
 - Width, 57

- Pattern center (base point), 182
 Penalty function, 172
 Augmented Lagrangian, 186
 Exterior, 172
 Interior, 191
 Quadratic (exterior), 172, 186
 Permutation of elements, 99
 Pi-to-T transformation, 33
 Piece-wise linear function, 126
 Poincaré metric, 198
 Positive element values, 53
 Maintained by log
 transformation, 109
 Powell's theorem, 161
 Power
 Available from source, 43
 Current, 116
 Infinite from ideal source, 16
 Maximum available, 8, 10, 42
 Net entering port, 42
 Range given load SWR, 12
 Reactive, 25
 Real, 25
 Transfer, 8, 134
 Transfer in lossless network,
 14
 Waves, 11, 42
 Precomputation, 182
 Predistortion, 49
 Product method in synthesis, 97
 Program
 ALLCHEBY, 6, 36, 60, 62, 124
 CHOICES, 7
 CNETOPM, 6
 DENORM, 6, 39
 Elliptic filter design, 83
 Pole placer, 96
 RIPFREQS, 6, 94
 S11TOZ, 6, 133
 Prototype, 27
 Bandpass, 28
 Direct-coupled network, 31
 Element values, 19, 35
 Elliptic-function, LP, 28, 38
 Immittances, 29
 Normalization, 35
- Q—
- Q (quality factor)
 Bandwidth, 18
 Doubly-loaded, 30
 Load, 102
 Loaded, 24
 Singly-loaded, 29
 Source, 102
 Stress magnification, 26
 Terminal, normalized, 48
 Unloaded, 26
 Quadratic programming (QP),
 191
 Quadrature (numerical
 integration), 192
- R—
- Random searches, 179
 Ratio
 Element values, 31, 32
 Ideal transformer turns, 34
 Reactance
 Network branch effect on
 reflectance, 136
 Odd function of frequency, 45
 Parallel, 25
 Reactance transformations, 29
 Real-frequency technique
 Features, 125
 Introduction, 112
 Reciprocal functions, 196
 Reciprocity theorem, 16
 Recursion equations, 122
 Reference voltage, singly-
 terminated network, 16
 Reflectance
 Cross section, 136
 Definition, 14
 Element response, 100
 Filter Example 4.4.1, 100
 Invariant, 198
 Monotonic, 139

Reflectance (Continued)

- Related to transducer power function, 172
- Skewed arc of a curve, 137
- Subnetwork interface, 139
- Transformer, ideal, 141
- Trap branches, 140
- Unimodal, 140
- Unit, branch relationship, 138
- Zero for single-frequency matching, 113
- Reflection, transmission line, 10
- Reflection coefficient, 100
 - Generalized, 9, 113, 134, 195
 - Real and imaginary parts, 9
 - Smith chart, 20
- Reflection mapping, transfer phase change, 198
- Relative-change stopping tolerance, 183
- Residual error, 159, 189
- Resistance
 - Even function of frequency, 44
 - Parallel, 24, 32
 - Rational function, 126
- Resonator
 - Coupling concept, 48
 - Direct-coupled, parallel, 31
 - Efficiency, 26
 - Exact replacement, 68
 - In direct-coupled filters, 49
 - LC, 26, 50
 - Loaded Q, 25
 - Parallel (nodal), 35
 - Replacing external series resonator, 68
 - Replacing internal series resonators, 69
 - Replacing internal trap resonators, 70
 - Resonance frequency, 25
 - Series (mesh), 35
 - Slope equivalence, 51
 - Stagger tuning, 65, 67
 - Stored energy, 25

Response

- Bandpass shape, 18
- Butterworth, 19
- Chebyshev shape, 18
- Effects of dissipation, 23
- Elliptic function, 22
- Equal-element, 23, 35
- Group delay in BP filters, 29
- Inverse Chebyshev, 22
- Lowpass, 17
- Major shapes, 17, 19
- Reflectance surface, 143
- Open-circuit voltage, 16
- Short-circuit current, 16
- Voltage transfer ratio, 16
- Return loss
 - Definition, 14
 - Gain-bandwidth limitation, 119
- RF transformer, 32, 33, 80
- Ripple, 36, 91
- Rotated basis matrices, 184

—S—

Sampling, frequency, 134, 192

- Scaling
 - CASTL T equivalent branch reactances, 146
 - Frequency, 39
 - Impedance, 39
 - Rules, 39
 - Response surface, 145
- Scattering parameters, 42
 - In double matching, 131
 - In terms of ABCD parameters, 43
 - Normalization, 42
- Secant search, 94
- Selectivity, 3, 17, 32, 45, 47, 53, 55, 56, 57, 59, 62, 70, 71, 76, 83, 88, 94, 95, 96, 98, 105, 110, 127
- Asymptotes, 55
- Passband, 94
- Stop band, 55, 95

- Semi-infinite programming (SIP), 171, 192
 - Sensitivity
 - Bode, 101
 - Element value, 97
 - Lagrange multiplier, 187
 - Polynomial coefficient, 97
 - Root, 97
 - Sequential unconstrained minimization technique (SUMT), 186, 190
 - Sequential quadratic programming (SQP), 172, 191
 - Series branch ABCD parameters, 139
 - Shadow price, 161
 - Single-match
 - Dependent source resistance, 122
 - Overview, 120
 - Main result, 121
 - Singly-terminated filter
 - Example 3.5.2, 79
 - Smith chart
 - Center, 9
 - Description, 9
 - Generalized, 12
 - In f plane, 199
 - In measuring loaded Q, 125
 - Normalization, 9
 - Reflectance behavior for broadband matching, 137
 - Transmission-line application, 9
 - Source
 - Ideal, 15
 - Q, 28
 - Spectral factorization, 97, 131
 - Spreadsheet
 - Applications, 75
 - Example 3.5.1., 78
 - Example 3.5.2., 79
 - Example 3.5.3, 80
 - Optimization feature, 76, 83
 - Subroutines, 83
 - Stagger tuning, 65, 67, 87
 - Standing-wave ratio (SWR), 11
 - Circle in impedance and reflection planes, 197, 199
 - Relation to reflection coefficient, 11
 - Voltage/current extremes, 14
 - Without transmission line, 12
 - Starting vector in iterated analysis, 109
 - Stopband ripple, 22
 - Subnetworks
 - Cascading, 41
 - Summaries
 - Comprehensive equal-ripple filters, 110
 - Direct-coupled filters, 87
 - Fundamentals, 45
 - GRABIM in detail, 193
 - Matching networks, 167
 - Superposition principle, 40
 - Symbols, 204
 - top coupling, 32
 - Symmetric filters, 106
 - Synthesis
 - Approximation problem, 97
 - Cascade, 98
 - Continued fraction expansion, 98
 - Polynomial, 35
 - Underlying concepts and operations, 96
- T—
- T equivalent network, CASTL, 146
 - T-to-Pi transformation, 33
 - Tellegen's theorem, 177
 - Terminations, unequal, 21, 36
 - Thevenin
 - Equivalent source, 14, 17, 27
 - Impedance, 14
 - Time for lattice pattern search, 156
 - Topology, 27

- Distributed interstage network, 165
- El section, 114
- Element type codes, 149
- Elliptic branches, 38
- Permutations, element, 99
- Pi section, 33
- Reduced-degree, full rank, 148
- T section 33
- Transducer function, 90, 106, 120, 129, 172
 - In terms of ABCD, 134
 - Inverted, power, 173
 - Power gain, 9
- Transformation
 - Bandpass network, 32
 - Bilinear, 12
 - Frequency, LP to BP, 29
 - Geffe's, 71
 - Norton, 33
 - Pi-to-T, 33
 - Reactance, LP to BP, 19
 - Series-to-parallel resonator, 52
 - Star-delta, 33
 - T-to-Pi, 33
- Transformation Q, 117, 167
- Transformation resistance, 116
- Transformer, ideal, 34
 - ABCD parameters, 139
 - For lowpass network source resistance, 148
 - Reflectance, unimodal, 141
- Transient overshoot, 19
- Transmission line Z_0 , 10
- Transmission zeros, 53, 92
 - On the $j\omega$ s-plane axis, 98
 - On the s-plane real axis, 99
- Trap, 23, 30, 53
 - Asymmetric, 88
 - Reactance effects on reflectance, 140
- Reactance given null ω , 151
- Stopband selectivity effects, 55
- Symmetrical frequency pairs, 38
- Tree diagram, 182
- Trial points in grid or lattice, 155, 184
- Tuning, filter
 - alternating open/short method, 84
 - By input reflection delay, 85
- U—
- UM functions (unimodal or monotonic), 146
- Undercoupled response, 19
- V—
- Variable space, discretized, 179
- Vector, column
 - Definition, 41, 135
 - Gradient, 189
 - Voltage stress, Ex. 2.4.1, 31
- W—
- Waves
 - Incident and emerging, 42
 - Power, 11, 42
 - Traveling on transmission line, 10
- Weights, 187
- Wiener-Lee transforms, 128
- Z—
- Zeros of transmission, 53, 92