BAHL
BHARTIA

Microwave Solid State Circuit Design

Second Edition

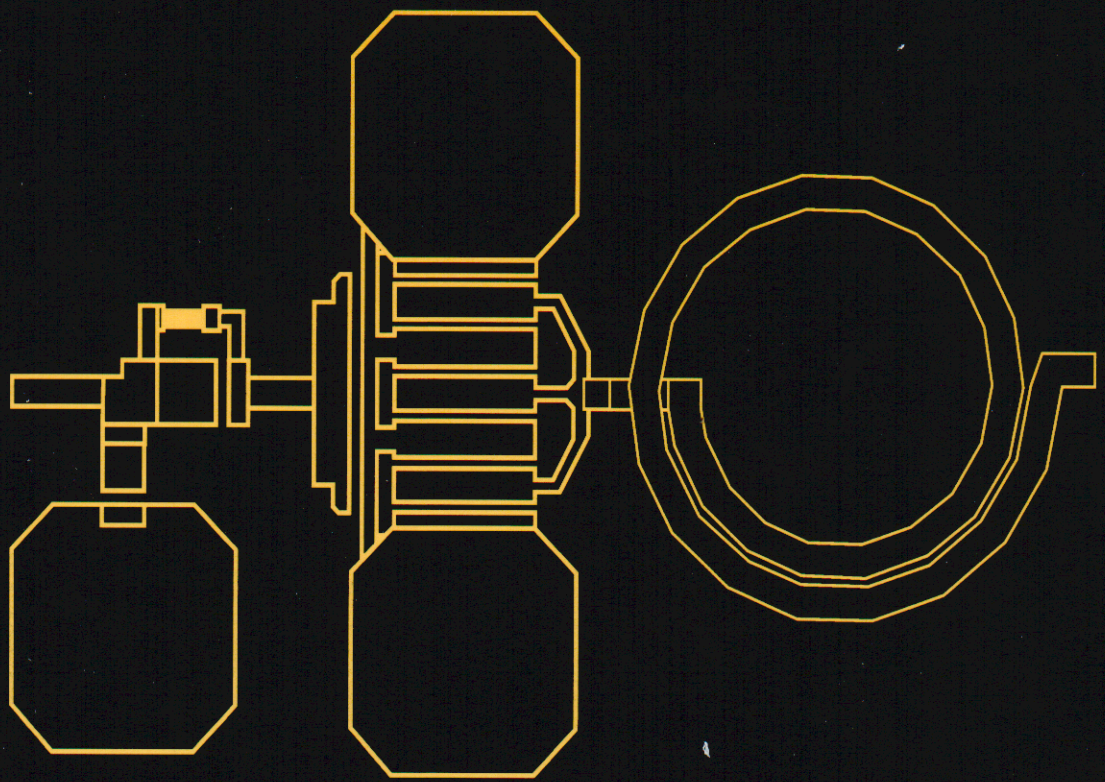
WILEY-INTERSCIENCE

WILEY

Microwave Solid State Circuit Design

Second Edition

Inder Bahl and Prakash Bhartia



MICROWAVE SOLID STATE CIRCUIT DESIGN

SECOND EDITION

Inder Bahl
M/A-COM
Roanoke, Virginia

Prakash Bhartia
Defense Research Establishment
National Defense Headquarters
Ottawa, Ontario, Canada

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

This book is printed on acid-free paper. ♾

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Microwave Solid State Circuit Design, Second Edition
Edited by Inder Bahl and Prakash Bhartia

ISBN 0-471-20755-1

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTRIBUTORS

Inder Bahl, Distinguished Fellow of Technology, M/A-COM, Inc., 5310 Valley Park Drive, Roanoke, VA 24019

Prakash Bhartia, Director General, DREO, 3701 Carling Avenue, Nepean, Ontario K1A 0Z4, Canada

Edward Griffin, Product Line Manager, M/A-COM, Inc., 5310 Valley Park Drive, Roanoke, VA 24019

K. C. Gupta, Professor, Department of Electrical Engineering, University of Colorado, Boulder, CO 80309-0425

Robert G. Harrison, Adjunct Research Professor, Department of Electronics, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

A. P. S. Khanna, R&D Section Manager, Agilent Technologies, 3175 Bowers Avenue, Mailstop 88A, Santa Clara, CA 95054

Protap Pramanick, Senior Staff Scientist, KL Microwave, 408 Coles Circle, Salisbury, MD 21804

Ramesh Ramadoss, Department of Electrical Engineering, University of Colorado, Boulder, CO 80309-0425

Arvind Sharma, Senior Staff Engineer, TRW Electronics and Technology Division, One Space Park, Redondo Beach, CA 90278

Prasad N. Shastry, Professor, Department of Electrical and Computer Engineering, Bradley University, 1501 W. Bradley Avenue, Peoria, IL 61625

Robert J. Trew, Willis G. Worcester Professor of Engineering, and Head ECE Department, Virginia Tech University, Blacksburg, VA 24061

CONTENTS

PREFACE

The first edition of this book was published almost fifteen years ago. Over the years the book has proved to be exceptionally well adopted and used by researchers, educators and designers alike. Technology has also evolved considerably, with significant interest in areas such as micro electro mechanical systems (MEMS), elecrooptics and nanotechnology. These topics are still evolving and some MEMS based devices are now commercially available. Optics based technology is already used extensively in communication and other RF based systems applications, while nanotechnology is still in its infancy. Thus, in revising this book, emphasis was placed on retaining and expanding topics of current and future interest and cutting back or deleting material or whole chapters that were outdated or of little utility. In this decision we were aided significantly by feedback from many of our readers and the contents of this book reflect these inputs. Since the commercial use of RF and microwave technologies is growing tremendously in the wireless communications area, much of the material is slanted to this application.

To start with, it was determined that the basic structure and organization of the book was sound. Following suggestions received to make the book self-contained, microwave network theory was incorporated into Chapter 1. The rest of the chapters were revised as required. For example, in Chapter 2, coaxial line discontinuities and rectangular waveguide discontinuities were dropped and the rest of the chapter updated with the most recent equations currently used for lumped inductors. In Chapter 3 on resonators, the sections on less popular resonators were dropped and sections on more current topics, such as, high Q, tunable, and active resonators were added. Chapter 4 on impedance transformation techniques was completely revised and authored by a new contributor. This chapter is absolutely new. The next chapter on hybrids and couplers contains about thirty percent new material and reflects current practices. Similarly in Chapter 6, the section on multiplexers was dropped and new sections on compact filters, filter tuning, and EM simulation of filters were included. While much of the material in Chapters 7 and 8 remains current, it was deemed appropriate to revise the write up and add information on heterostructure and wide band gap devices. The next chapter on oscillators incorporates oscillators for wireless circuits while the chapter on amplifiers also includes design procedures and examples on wireless amplifier designs.

Chapter 11 is completely new, including such current topics as cryogenic mixers and self-oscillating mixers. While Chapters 12 and 13 have also been revised, it was decided in view of the availability of a large number of computer-aided design software packages and many books on this topic to drop this subject and substitute a new chapter on the upcoming and popular subject of MEMS. Many researchers and commercial houses are engaged in this area and MEMS based devices hold significant promise in lowering power requirements and cutting back on size and weight. Finally, a new chapter on circuit manufacturing technologies has been inserted as Chapter 15. This chapter incorporates some of the continuing important topics from the old Chapter 15, but incorporates the most modern practices in printed circuit board fabrication, hybrid integrated circuits, and microwave monolithic integrated circuits.

Thus, overall this revised version has three new chapters, two others with major changes and the final ten with about 25% changes each. Each chapter still contains design equations, tables, appropriate figures, and a set of problems for students. Although the material has been revised, there is still adequate material and choice of subjects to cover a one semester, one quarter, two semester or a two-quarter course. The organization of topics as proposed in the Preface to the original volume is still valid, but it is recommended that MEMS be incorporated into any break-up, as the topic is expected to play a major role in future microwave circuits and systems design.

We believe that this new revised version will be as well accepted as the first edition by professors, students, and designers and researchers alike. As with the first edition, we hope it will simplify the understanding of the subject for students and facilitate the design of circuits and devices by the designer.

As with the first edition, extensive cooperation and coordination of effort amongst the contributors was required, particularly as new authors were also added. We thank all the contributing authors for their diligence and effort in preparing and revising their manuscripts and on their behalf thank all the people behind the scenes who made this possible. Finally, we appreciate the support and patience of our families throughout the preparation of this book.

INDER BAHL
PRAKASH BHARTIA

1

INTRODUCTION

P. Pramanick and P. Bhartia

The history of the development of microwave circuits has in many ways followed that of the lower frequency electronics circuits. There have been constant pressures in both these disciplines to move from tubes to solid state devices and from large components to small and to the development of integrated circuits, devices, and systems. However, unlike the electronics field, where a large public impact was possible and strong consumer demand generated through ownership of radios, digital watches, calculators, television sets, video cassette recorders, and so on, the microwave oven seems to be the single piece of equipment that comes to the layman's mind when the word *microwaves* is mentioned. The greatest impact and perhaps use of microwave circuits and systems has been in areas such as communications, radar, electronic warfare, navigation, surveillance, and weapon guidance systems, which are largely military in nature and have been supported strongly by the defense community. Although there is also a lucrative market for defense-products-oriented industries, the profits are often limited by small-volume, limited customers due to export restrictions, and hence the drive for development in the area of microwave circuits seems traditionally to have lagged behind that of electronic circuits. For example, electronic integrated circuits (ICs) on chips were available commercially long before microwave integrated circuits (MICs) were.

The current trend in microwave technology is toward circuit miniaturization, high-level integration, improved reliability, low power consumption, cost reduction, and high-volume applications. Component size and performance are prime factors in the design of electronic systems for satellite communications, phased-array radar systems, electronic warfare, and other military applica-

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

tions, while small size and low cost drive the consumer electronics market. Monolithic microwave integrated circuits (MMICs) based on gallium arsenide (GaAs) technology are the key to meeting the above requirements. They play an increasing role in consumer electronics dealing with information transfer, communications, automotive applications, and entertainment. With MMIC technology a typical microwave subsystem can be produced on a single chip at costs of less than $100 while simpler single-function chips cost less than $10. Some very simple function chips are now produced at costs as low as $1. While most MMICs currently in production operate in the 0.5- to 30-GHz microwave range, there are increasing applications in the millimeter-wave (mmW) spectrum (30–300 GHz) as higher frequency transistors mature. Monolithic technology is particularly beneficial to mmW applications through the elimination of the parasitic effects of bond wires that connect discrete components in conventional hybrid structures.

1.1 CHARACTERISTICS OF MICROWAVES/MILLIMETER WAVES

The term *microwave/millimeter wave* generally refers to the frequency range where wavelengths are of the order of centimeters down to 1 mm. Some authors have suggested that microwave/millimeter-wave spectrum corresponds to the frequency range 1–300 GHz (wavelength 30 cm–1 mm), but the generally accepted convention is that the frequency range 300 MHz–300 GHz (wavelength 100 cm–1 mm) is more appropriate. Figure 1.1 depicts the position of the microwave and millimeter-wave part of the spectrum in relation to the rest of the electromagnetic spectrum.

In the microwave/millimeter-wave region, for convenience, a letter band designation has also been used to indicate the portion of the spectrum being referred to. Over the years, the radar community, for example, has created its own band designations, leading to general confusion. For example, *K*-band designates the frequency ranges 18–26.5 GHz, 19–27 GHz, 10.9–36 GHz, and 20–40 GHz in the radar users terminology, the U.K. frequency designation, the old U.S. military, and the new U.S. military designations, respectively. For the sake of clarity and uniformity, only the new U.S. military designation should be used. Figure 1.2 depicts the frequency band breakdown for the microwave/millimeter-wave region of interest for the purposes of this test.

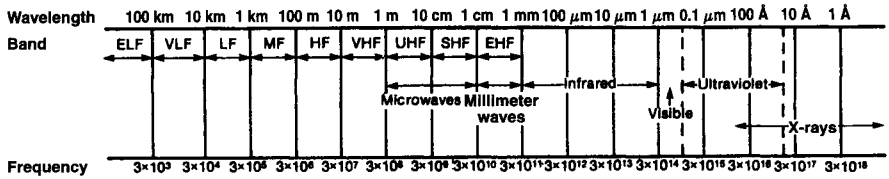


Figure 1.1 Electromagnetic spectrum.

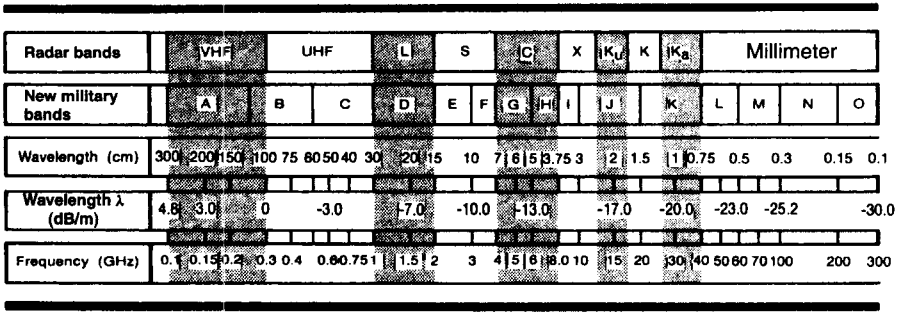


Figure 1.2 Band designation chart.

Some other features of microwave/millimeter waves are worth noting; for example, the decrease in wavelength as frequency increases corresponds to a reduction in component size, resulting in more compact systems and narrow beamwidths, which allow, for example, for greater resolution and precision in target tracking. Many advantages and disadvantages of microwave via-à-vis millimeter wave are discussed in Bhartia and Bahl [1].

Another interesting and important characteristic is the free-space propagation attenuation over the frequency range. As shown in Fig. 1.3, atmospheric absorption increases with frequency, but in addition, there exist high absorption windows in the millimeter-wave band due to atmospheric water vapor and oxygen. While this characteristic is not of significant interest for the purposes of

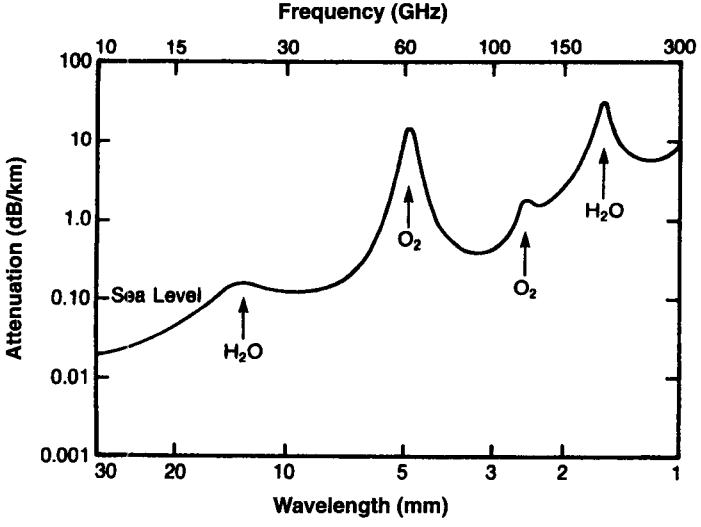


Figure 1.3 Atmospheric absorption as function of frequency.

this book in microwave circuits, it should be realized that a similar behavior occurs with transmission media, and one must choose the appropriate low-loss media for the frequency of operation of one's circuit.

1.2 HISTORY OF MICROWAVE PLANAR CIRCUITS

The first evolution of the conventional waveguide and coaxial line was a flat-strip coaxial transmission line used by Rumsey and Jamieson—as mentioned, for example, by Barrett [2] for producing an antenna system and power division network during World War II. This was gradually integrated with printed-circuit technology to result in the "microwave printed circuit" (MPC) reported by Barrett and Barnes [3] and soon developed into a printed-circuit waveguide handbook [4]. Shortly after Barrett and Barnes's [3] report on MPCs, the Federal Communications Research Laboratories announced the microstrip [5] and King [6] reported the "dielectric image line." The MPC concept was used in fabricating many components, such as directional couplers, filters, attenuators, and antennas. The microstrip is fundamental to IC design and is today being used extensively in its many forms in a large variety of circuits. Finally, the image line is an excellent transmission medium, particularly for millimeter-wave circuits.

Following the preceding developments, the MIC area grew rapidly in the 1960s. Between 1964 and 1968, the largest quantity production for a MIC was perhaps the 600 transmit/receive (T/R) modules produced by Texas Instruments for the molecular electronics for radar application (MERA) radar [7]. Many other significant developments also occurred over the 1960–1980 period, and it was evident that the field of MICs was fast maturing. Over these years, the idea of MMICs also evolved, where all microwave functions of analog circuits, as well as new digital applications, could be incorporated on a single chip [8]. In the earlier MICs and MMICs, high-resistivity *p*-type (boron) silicon was used as the microwave substrate and host for the devices. However, two factors have been primarily responsible for the emergence of GaAs in the development of MMICs. First, the semi-insulating substrate is almost an ideal dielectric medium for microstrip transmission; second is the GaAs field-effect transistor (FET), which is the workhorse of all analog ICs. The latter has benefited from the application of silicon processing technology. Thus, GaAs technology promises a new breed of components that will allow designers greater flexibility and lower cost approaches to achieve their design goals.

1.3 APPLICATIONS OF MICROWAVE PLANAR CIRCUITS

Microwave planar circuits can be applied to and substituted for the conventional form of microwave circuitry in virtually every application in the fields of communications, electronic warfare, radar, and weapon systems. In general,

the limitations are few, one fundamental one being the power-handling capability. However, realization of good matching circuitry has allowed high power levels to be achieved in some laboratory MMICs, but to achieve high powers while maintaining high yield still requires hybrid MIC as well as conventional waveguide and coaxial-line techniques. In many cases, requiring high power, such as space-based radar and the use of thousands of solid-state transmitter-receivers for an active aperture phased array, allows for power distribution and hence use of MMICs or MICs. Hopefully, the extensive number of T/R modules required will allow for a substantial cost reduction for MMICs over hybrid MICs or conventional circuitry.

Another area that should see high microwave planar circuit usage is electronic warfare. In particular, expendable systems such as expendable jammers and smart munitions using microwave/millimeter-wave guidance systems, both passive radiometer type and active radar, should benefit from the cost reductions and lower weight properties of these circuits. Naturally, satellite systems, where weight is an all-important issue due to high cost per kilogram of launched payload, also stand to benefit significantly from the use of this technology.

Similarly, receivers and transmitters for communications, electronic support measures, electronic countermeasures, and systems for electronic communication or signal intelligence (ELINT, COMINT, SIGINT) all make extensive use of microwave/millimeter-wave circuits and devices. As systems become more complex due to the nature of the electronic threat, hardware complexity increases, resulting in larger and heavier systems. Use of planar technology, and in particular MMICs, helps to achieve significant reduction in both these factors, thus making it possible to deploy these systems easily on aircraft, where space and power are driving constraints.

Most radars being built currently use hybrid circuitry that needs tweaking for optimum performance. In these cases, microwave planar circuits offer the advantages of smaller size, lighter weight, potentially lower cost, high reliability, broad-bandwidth capability, and function reproducibility. These advantages allow the development of active element phased radars with significant beam agility, multifunction capability, and reliability. In addition, in most applications the redundancy of the T/R modules that can be built in allows for the desirable feature of graceful degradation in case of failure. A typical active element configuration for one of these is shown in Fig. 1.4.

While the military and space applications stand to gain the most from use of planar microwave/millimeter-wave circuits, other unconventional applications, such as highway-traffic control using microwave systems and microwave sensor systems used in microwave heating and drying, will likewise see cost reductions and smaller sizes with the availability of multifunction monolithic circuits.

The advances in microwave/millimeter-wave planar circuits coupled with advances that are currently occurring in electro-optics, magneto-optics, microwave optics, and microwave acoustics point to exciting decades ahead for engineering and the sciences. In particular, these fields will have strong impact

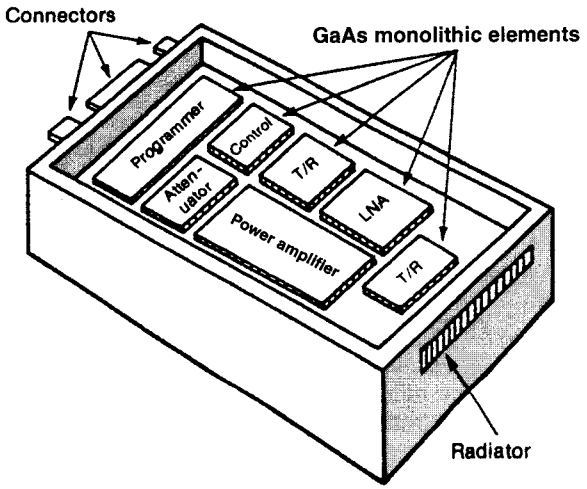


Figure 1.4 A T/R module using GaAs MMIC chips.

on consumer electronics and gadgets, while at the same time contributing in a large way to fields such as robotics, smart weapons, satellite technology and capabilities, smart-skins for aircraft, photonics, and smart built-in testing. Toward this objective, this text should serve as a reference for the fundamental microwave/millimeter-wave planar circuit active and passive components.

To make the book self-contained, in the next section, we discuss some fundamental concepts of microwave network theory and matrices of importance for circuit analysis and design.

1.4 MICROWAVE NETWORK THEORY

Microwave passive and active networks can be classified as multiport networks. Such networks are also known as N-port networks. Assuming the input and the output ports of an N-port microwave network are known, its frequency- and time-domain responses to a known excitation can be determined. For instance, if one of the ports of a transistor is terminated in a short circuit, the frequency response of the remaining two-port network can be obtained from the knowledge of the original three-port network. Also, in any microwave circuit, system or subsystem, there are many components connected in a certain fashion. The frequency response of this system can be obtained from knowledge of the individual components.

There are many equivalent ways in which the frequency response of a linear microwave network can be calculated [9]. This chapter deals with the representation of linear microwave networks as multiport black boxes.

1.4.1 Concepts of Equivalent Voltage and Current

The determination of network characteristics at microwave frequencies involves concepts that are substantially different from those used for low-frequency radio-frequency (RF) circuits for which voltages and currents, which determine impedance, can be uniquely defined. At microwave frequencies, use of neither high-frequency probes nor low-impedance current measurements are possible, because parasitic impedance and capacitance cannot be made small enough. In addition, the physical dimensions of a microwave circuit are no longer small compared to wavelength. Therefore, in most cases, the concepts of equivalent voltage and current are used. The equivalent voltage and current are so chosen that power transmitted along a transmission line is computed correctly using the equation

$$P = \tfrac{1}{2}\,\mathrm{Re}\int_S (E \times H)\,ds = \tfrac{1}{2}\,\mathrm{Re}(VI^*) \tag{1.1}$$

where '*' denotes the complex conjugate, and the power P is assumed to be flowing in the z direction; V and I are the equivalent voltage and current, respectively; E and H are the electric and magnetic fields, respectively, in the xy plane; and S is the cross-sectional area of the transmission line.

We can use the familiar definitions of voltage and current only in electrostatics. As soon as the electric field becomes time dependent, it generates a time-varying magnetic field, which in turn gives rise to a dynamic electric field. The electric voltage corresponding to this new dynamic electric field depends on the path of the line integral chosen to calculate the voltage.

Let us consider the transverse part of the electromagnetic (EM) field in a transmission line, for example, a coaxial line or a waveguide,

$$E_T^+ = \xi e_T(x, y)e^{-j\beta z} \tag{1.2a}$$

$$H_T^+ = \xi h_T(x, y)e^{-j\beta z} \tag{1.2b}$$

where ξ is a constant of proportionality, $e_T(x, y)$ and $h_T(x, y)$ are the normalized modal functions such that

$$\int_S [e_T(x, y) \times h_T(x, y)]\,ds = 1 \tag{1.2c}$$

and β is the propagation constant in the z direction.

According to Collin [9], the above fields are proportional to the equivalent voltage and current. Therefore,

$$E_T^+ = K_V V^+ e_T(x, y)e^{-j\beta z} \tag{1.3a}$$

$$H_T^+ = K_I I^+ h_T(x, y)e^{-j\beta z} \tag{1.3b}$$

where K_V is a constant of proportionality and V^+ and I^+ are the forward-going equivalent voltage and current, respectively. The transmitted power associated with the forward wave is given by

$$P^+ = \tfrac{1}{2}\,\mathrm{Re}\int_S (E_T^+ \times H_T^+)\,ds \tag{1.4}$$

Combining Eqs. (1.2c), (1.3a), (1.3b), and (1.4) gives

$$P^+ = \tfrac{1}{2}K_V K_I\,\mathrm{Re}(V^+ I^{+*}) \tag{1.5}$$

Comparing Eqs. (1.1) and (1.5) gives

$$K_V K_I = 1 \tag{1.6}$$

Using transmission line theory, we can write

$$\frac{V^+}{I^+} = Z_0 \tag{1.7}$$

where Z_0 is the characteristic impedance of the line. Comparing (1.2), (1.3), and (1.7), we get

$$\frac{K_I}{K_V} = Z_0 \tag{1.8}$$

Solving Eqs. (1.6) and (1.8) gives

$$K_V = \frac{1}{\sqrt{Z_0}} \tag{1.9a}$$

$$K_I = \sqrt{Z_0} \tag{1.9b}$$

The characteristic impedance Z_0 of a coaxial or two-conductor transmission line can be uniquely defined for the fundamental transverse-electromagnetic (TEM) mode. Therefore, the equivalent forward voltage and current can be uniquely expressed using Eqs. (1.2), (1.3), and (1.9) as

$$V^+ = \varsigma\sqrt{Z_0} \tag{1.10a}$$

$$I^+ = \frac{\varsigma}{\sqrt{Z_0}} \tag{1.10b}$$

Similarly, the backward voltage and current can be written as

$$V^- = \kappa\sqrt{Z_0} \tag{1.11a}$$

$$I^- = \frac{\kappa}{\sqrt{Z_0}} \tag{1.11b}$$

In Eqs. (1.10) and (1.11) ς and κ are proportionality constants.

If the transmission line supports a non-TEM mode, that is, transverse-electric (TE) or transverse-magnetic (TM) mode in a waveguide, the definition of the characteristic impedance cannot be unique. For example, if one considers the ratio of the fundamental mode average power flowing through a rectangular waveguide and the voltage at the center of the broad wall, the characteristic impedance assumes the form

$$Z_0(f) = \frac{2b}{a}\frac{\eta_0}{\sqrt{1-(f_c/f)^2}} \tag{1.12a}$$

where a and b are the width and the height of the waveguide cross section, respectively; η_0 is the free-space impedance; f is the operating frequency; and f_c is the cutoff frequency of the fundamental mode.

If one considers the ratio of the same voltage and the total longitudinal current, then the characteristic impedance assumes the form

$$Z_0(f) = \left(\frac{\pi b}{2a}\right)\frac{\eta_0}{\sqrt{1-(f_c/f)^2}} \tag{1.12b}$$

Also, if one considers the ratio of the fundamental mode average longitudinal power and the total longitudinal current, one gets

$$Z_0(f) = \left(\frac{\pi^2 b}{8a}\right)\frac{\eta_0}{\sqrt{1-(f_c/f)^2}} \tag{1.12c}$$

Therefore, unique definitions for equivalent voltage and current are not possible.

1.4.2 Admittance and Impedance Matrices

Consider Fig. 1.5. The accessible ports are denoted by $1, 2, 3, \ldots, N$. In addition, there is a ground terminal at each port. Let us assume that the I_i $(i = 1, 2, 3, \ldots, N)$ denote the port currents and the V_i $(i = 1, 2, 3, \ldots, N)$ denote the port voltages, respectively, at ports 1–N. The admittance matrix of the network is defined as

$$[I] = [Y][V] \tag{1.13}$$

where

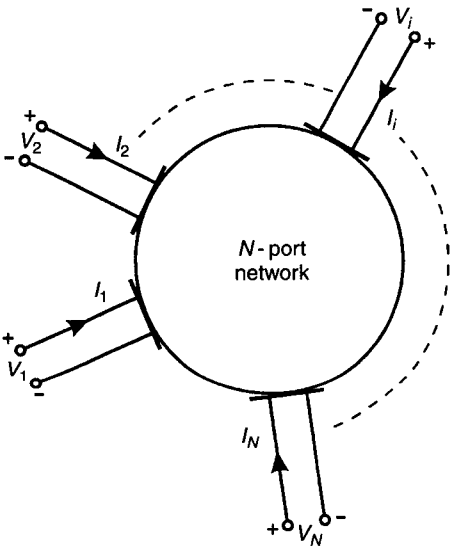


Figure 1.5 Schematic of N-port network.

$$[I] = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix} \qquad [V] = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix} \qquad [Y] = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1N} \\ Y_{21} & Y_{22} & \cdots & Y_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & Y_{N2} & \cdots & Y_{NN} \end{bmatrix} \tag{1.14}$$

The admittance matrix $[Y]$ is obtained by nodal analysis of the network and then solving for the port currents or voltages. The corresponding impedance matrix is defined as

$$[Z] = [Y]^{-1} \tag{1.15}$$

The above admittance and impedance matrices are also known as unnormalized admittance and impedance matrices, respectively.

A microwave component or a subsystem is connected to a larger system. As a result, each port of a component is terminated by an impedance offered by the larger system to which it is connected. Let us assume that the impedance vector gives the set of terminating impedances offered by the embedding system

$$[Z_I] = [Z_{I1} \quad Z_{I2} \quad \cdots \quad Z_{IN}] \tag{1.16}$$

We define a new set of port voltages and currents, known as the normalized port voltages and currents, as

$$v_i = \frac{V_i}{\sqrt{\text{Re } Z_{Ii}}} \qquad i_i = I_i \sqrt{\text{Re } Z_{Ii}} \tag{1.17}$$

Note that normalized voltages and currents have the same dimension, which is $\sqrt{\text{watt}}$. Hence the relationship between the normalized and the unnormalized parameters can be expressed as

$$[v] = [z][i] \tag{1.18}$$

where

$$[v] = [z_c]^{-1/2}[V] \qquad [i] = [z_c]^{1/2}[I] \tag{1.19}$$

and

$$[z_c] = \text{Re}\begin{bmatrix} Z_{I1} & 0 & 0 & \cdots & 0 \\ 0 & Z_{I2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & Z_{IN} \end{bmatrix} \tag{1.20}$$

$[z]$ is known as the normalized port impedance matrix, and

$$[y] = [z]^{-1} \tag{1.21}$$

is the corresponding normalized port admittance matrix.

Below we will show how normalized voltage and current matrices account for the interaction of the network with the system that embeds it.

1.4.3 Scattering Matrix

The scattering matrix concept with respect to positive and real terminating impedances was introduced by Penfield [10]. Kurokawa introduced the concept of power wave variables in 1965 [11] and generalized the concept of Penfield.

Following the method due to Kurokawa [11], let us consider a one-port network as shown in Fig. 1.6.

We define the following variables:

$$a_1 = \frac{V_1 + Z_{I1} I_1}{2\sqrt{\text{Re}\, Z_{I1}}} \tag{1.22a}$$

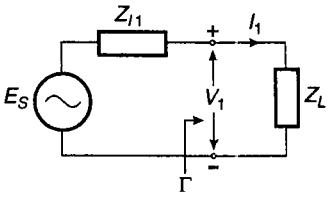


Figure 1.6 One-port network.

$$b_1 = \frac{V_1 - Z_{I1}^* I_1}{2\sqrt{\text{Re } Z_{I1}}} \tag{1.22b}$$

where the asterisk denotes the complex conjugate.

Solving Eqs. (1.22a) and (1.22b), we obtain

$$V_1 = \frac{Z_{I1}^* a_1 + Z_{I1} b_1}{\sqrt{\text{Re } Z_{I1}}} \tag{1.23a}$$

$$I_1 = \frac{a_1 - b_1}{\sqrt{\text{Re } Z_{I1}}} \tag{1.23b}$$

Now, what are the parameters a_1 and b_1?

From Fig. 1.6

$$V_1 = E_s - Z_{I1} I_1 \tag{1.24}$$

Combining Eqs. (1.22a) and (1.24) and subsequently multiplying a_1 by its complex conjugate, we obtain

$$|a_1|^2 = \frac{|E_S|^2}{4 \text{ Re } Z_{I1}} = P_{S,\max} \tag{1.25}$$

The right-hand side of Eq. (1.25) is easily recognized as the maximum power available from the source. Therefore, we can say that $|a_1|^2$ is the incident power from the source into the load Z_L. Using Eqs. (1.22a) and (1.22b), it can be shown that

$$|a_1|^2 - |b_1|^2 = \text{Re}\{V_1 I_1^*\} \tag{1.26}$$

The right-hand side of Eq. (1.26) is the total real power absorbed by the load. Therefore, we can come to the conclusion that $|b_1|^2$ is the power reflected from the load to the source. At this point we define the reflection coefficient

$$\Gamma = \frac{b_1}{a_1} = \frac{V_1 - Z_{I1}^* I_1}{V_1 + Z_{I1} I_1} = \frac{Z_L - Z_{I1}^*}{Z_L + Z_{I1}^*} \tag{1.27}$$

Using Eq. (1.27), the difference between incident and reflected power, or in other words the power absorbed by the load, can be written as

$$P_L = P_{S,\max} - P_{\text{refl}} = |a_1|^2 (1 - \Gamma^2) \tag{1.28}$$

From (1.27), under the matching condition, when $Z_L = Z_{I1}^*$, $\Gamma = 0$.

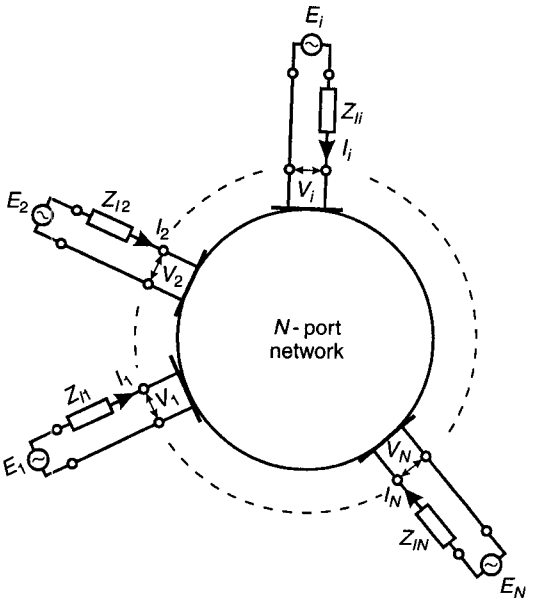


Figure 1.7 Multiport network.

Let us consider the multiport network shown in Fig. 1.7. The scattering matrix of a multiport network can be defined by the equation

$$[b] = [S][a] \tag{1.29}$$

where $[a]$ and $[b]$ are column matrices whose elements are the power wave amplitudes of the incident and reflected waves, respectively, at various ports. If the network has N ports, then $[S]$ is a square matrix of order N.

When the output impedances of the generators connected to the ports of the multiport network become purely real and equal to those of the transmission lines connected to the respective ports, a_i and b_i $(i = 1, 2, 3, \ldots, N)$ become the same as the complex incident and reflected voltages in transmission lines. Then we can write

$$V_i = V_i^+ + V_i^- \tag{1.30a}$$

$$I_i = I_i^+ + I_i^- \tag{1.30b}$$

$$\frac{V_i^+}{I_i^+} = \frac{V_i^-}{I_i^-} = Z_{Ii} \tag{1.31}$$

Combining Eqs. (1.30) and (1.22) gives

$$a_i = \frac{V_i^+}{\sqrt{Z_{Ii}}} = \sqrt{Z_{Ii}}\, I_i^+ \tag{1.32a}$$

$$b_i = \frac{V_i^-}{\sqrt{Z_{Ii}}} = \sqrt{Z_{Ii}}\, I_i^- \tag{1.32b}$$

and the reflection coefficient is given as

$$\Gamma = \frac{b_i}{a_i} = \frac{Z_L - Z_{Ii}}{Z_L + Z_{Ii}} \tag{1.33}$$

In Eqs. (1.30a)–(1.32), the plus and the minus superscripts denote the ingoing and outgoing parameters, respectively.

The total power absorbed by all the ports is given by the difference between the sum total of the incident and reflected powers of all the ports. Mathematically,

$$P_{\text{total}} = \sum_{k=1}^{N} P_k = \sum_{k=1}^{N} |a_k|^2 - \sum_{k=1}^{N} |b_k|^2 = [a^*]^T[a] - [b^*]^T[b] \tag{1.34}$$

Combining Eqs. (1.29) and (1.34) and the total power absorbed by the network for a lossless condition gives

$$P_{\text{total}} = [a^*]^T[[U] - [S^*]^T[S]][a] = 0 \tag{1.35}$$

where $[U]$ is the unity matrix of order N. From Eq. (1.35) we get

$$[S^*]^T[S] = [U] \tag{1.36}$$

For a reciprocal and symmetrical multiport network, Eq. (1.36) reduces to

$$[S^*][S] = [U] \tag{1.37}$$

which indicates that the scattering matrix of a symmetrical, lossless and reciprocal network is unitary.

Let us write the expanded form of Eq. (1.29):

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & S_{22} & \cdots & S_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1} & S_{N2} & \cdots & S_{NN} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_3 \end{bmatrix} \tag{1.38}$$

From Eq. (1.38) it can be seen that, in general,

$$S_{ij} = \frac{b_i}{a_j}\bigg|_{a_k = 0} \qquad k = 1, 2, 3, \ldots, N \qquad k \neq j \tag{1.39}$$

The condition

$$a_k = 0 \tag{1.40}$$

for all k's except $k = j$ is created by perfectly matching all but the jth port. The transmitted signal at the ith port and the incident signal at the jth port are appropriately monitored and S_{ij} is computed using Eq. (1.39). For a detailed description of the procedure the reader is referred to Schiek and Gronefeld [12].

For a given network terminated by a set of real impedances given by $[Z_I]$ in Eq. (1.16) the scattering matrix can be computed by using the following procedure.

- Obtain the port admittance matrix $[Y]$ using nodal analysis and then solving for the node currents.
- Invert the admittance matrix $[Y]$ to obtain the corresponding impedance matrix $[Z]$.
- Normalize $[Z]$ using

$$[z] = [z_c]^{-1/2}[Z][z_c]^{-1/2} \tag{1.41}$$

  where the diagonal matrix $[z_c]$ is obtained from Eq. (1.20).
- Obtain the matrix $[S]$ from

$$[S] = [[z] - [U]][[z] + [U]]^{-1} \tag{1.42}$$

Solving (1.42) gives

$$[z] = [[U] - [S]]^{-1}[[U] + [S]] \tag{1.43}$$

For a reciprocal network, all the matrices associated with the network are symmetrical matrices, which means

$$[Z] = [Z]^T \qquad [S] = [S]^T \qquad \text{etc.} \tag{1.44}$$

1.4.4 Transformation of Scattering Matrix Due to Shift in Reference Planes

Consider Fig. 1.8. The unprimed reference planes $t_j : j = 1, 2, \ldots, N$ are the original reference planes with respect to which the scattering matrix $[S]$ of the N-port network is defined. Now, let us assume that the reference planes are moved away from the network to their new positions marked by the primed letters $t'_j : j = 1, 2, \ldots, N$. The new scattering matrix of the network is given by

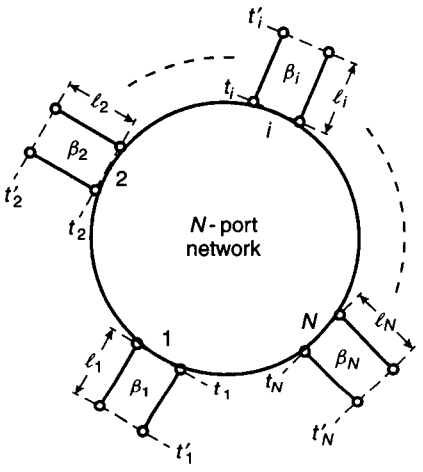


Figure 1.8 Multiport network

$$[S'] = [L][S][L] \tag{1.45}$$

where the matrix

$$[L] = \begin{bmatrix} e^{-j\beta_1 l_1} & \cdots & \cdots & 0 \\ & e^{-j\beta_2 l_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{-j\beta_N l_N} \end{bmatrix} \tag{1.46}$$

where β_k is the propagation constant of the wave at the kth port.

All the elements of the scattering matrix of a passive, lossless and reciprocal network cannot be chosen independently. Let us assume that Fig. 1.9 represents one such two-port network whose scattering matrix is given by

$$[S] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \tag{1.47}$$

From reciprocity, $[S]$ must be symmetrical. Therefore,

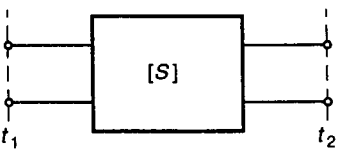


Figure 1.9 Two-port network.

$$[S] = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{11} \end{bmatrix} = \begin{bmatrix} |S_{11}|e^{j\theta_{11}} & |S_{12}|e^{j\theta_{12}} \\ |S_{12}|e^{j\theta_{12}} & |S_{11}|e^{j\theta_{11}} \end{bmatrix} \tag{1.48}$$

Combining Eqs. (1.37) and (1.48) gives

$$|S_{11}|^2 + |S_{12}|^2 = 1 \tag{1.49}$$

$$\theta_{11} = \theta_{12} \pm \tfrac{1}{2}\pi \tag{1.50}$$

This means that if S_{11} is known in complex form, S_{12} can be obtained from Eqs. (1.49) and (1.50). Equation (1.49) represents conservation of power in a lossless two-port network.

For a lossless nonreciprocal network, the above relations become

$$|S_{11}|^2 + |S_{21}|^2 = |S_{22}|^2 + |S_{12}|^2 = 1 \tag{1.51}$$

$$S_{11}^* S_{12} + S_{21}^* S_{22} = 0 \tag{1.52}$$

$$\theta_{11} + \theta_{22} = \theta_{12} + \theta_{21} \mp \tfrac{1}{2}\pi \tag{1.53}$$

where θ_{21} and θ_{22} are the phase angles associated with the elements S_{21} and S_{22}, respectively.

Scattering Matrix of a Lossless Three-Port Network. An application of unitary condition, given by Eq. (1.36), shows that it is impossible to simultaneously match all the ports of a three-port lossless reciprocal network. However, it is possible to match all three ports if the circuit is lossy or nonreciprocal. The examples are a Wilkinson power divider or a three-port circulator.

At this point it is worthwhile to discuss the usefulness of the scattering matrix. The concept of a scattering matrix is more general than admittance and impedance matrices. Many circuits may not possess an admittance or impedance matrix. Typical examples are ideal transformers having nonfinite elements. On the contrary, such transformers have scattering matrices. According to Carlin [13] and Kajfez [14] all passive networks possess scattering matrices.

In microwave engineering power flow is of primary consideration. Therefore, the scattering matrix is extremely useful. Let us consider the network in Fig. 1.10. Let P_a represent the available power from the generator and P_L the

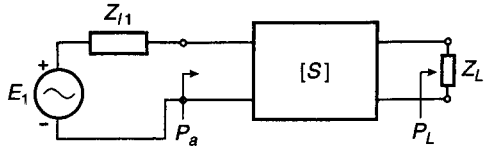


Figure 1.10 Circuit representation of two-port network.

power dissipated in the load R_L. Then it can be shown that the magnitude of the forward transmission coefficient is given by

$$|S_{12}|^2 = \frac{P_L}{P_a} \tag{1.54}$$

Scattering Matrix and the Concept of Insertion Loss. Let us consider the two-port network shown in Fig. 1.11*a*. The transfer coefficient of the network is given by

$$S_{21} = \frac{2V_L}{V_g}\left[\frac{R_g}{R_L}\right]^{1/2} \tag{1.55}$$

Now,

$$|S_{21}|^2 = S_{12}S_{12}^* \tag{1.56}$$

Therefore, from Eqs. (1.55) and (1.56) we obtain

$$|S_{21}|^2 = \frac{|V_L|^2/(2R_L)}{V_g/(8R_g)} = \frac{P_L}{P_a} \tag{1.57}$$

where P_a is the available power from the generator and P_L is the power dissipated in the load R_L. Now, let us consider the network in Fig. 1.11*a*. Suppose we have interposed a two-port network between the reference planes t_1 and t_2, as shown in Fig. 1.11*b*. Let the voltages across the load resistance R_L before and after the interposition of the two-port network be V_L and V_L^p, respectively. Also, let the powers dissipated in R_L before and after the interposition of the two-port network be P_L and P_L^p, respectively. The insertion power ratio of the two-port network is defined as

$$\text{IPL} = \frac{P_L^p}{P_L} \tag{1.58}$$

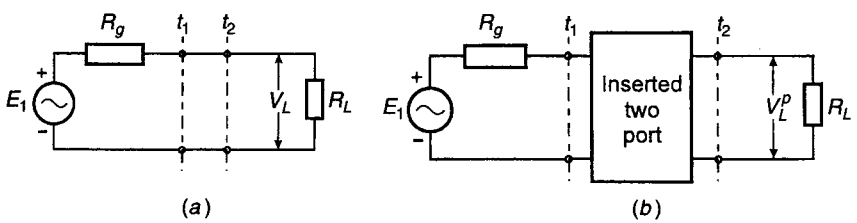


Figure 1.11 (*a*) Two-port network; (*b*) with two-port network inserted.

Analyzing the circuit in Fig. 1.11*b*, we obtain

$$\frac{P_L^p}{P_L} = \frac{|V_g|^2}{|V_L|^2} \frac{R_L^2}{(R_g + R_L)^2} \tag{1.59}$$

Combining Eqs. (1.57) and (1.59) gives

$$\frac{P_L^p}{P_L} = \frac{4R_L R_g}{(R_g + R_L)^2} \frac{1}{|S_{21}|^2} \tag{1.60}$$

Under a perfectly matched condition

$$\frac{P_L^p}{P_L} = \frac{1}{|S_{12}|^2} \tag{1.61}$$

Equation (1.61) is a very useful relationship in network synthesis.

1.4.5 Chain Matrix (*ABCD*) Representation

The chain, or *ABCD*, matrix is particularly useful in cascading or chain connecting microwave networks. The networks may be purely two-port ones or may have multiports. For a single isolated network, the *ABCD* matrix relates the output voltages and currents to the input voltages and currents. Let us consider the multiport network shown in Fig. 1.12. The input voltages and currents are $V_1, \ldots, V_N$ and $I_1, \ldots, I_N$. The output voltages and currents are $V_{N+1}, \ldots, V_{2N}$ and $I_{N+1}, \ldots, I_{2N}$, respectively. The input and the output parameters are related using the following matrix equation:

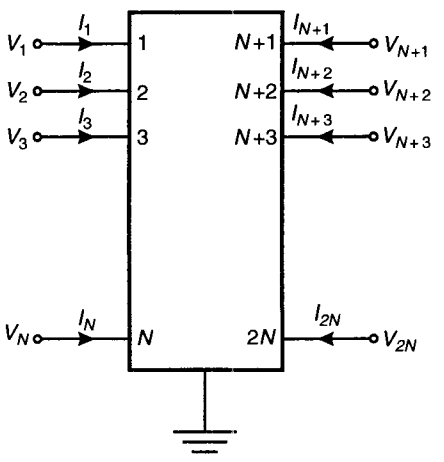


Figure 1.12 Multiport network.

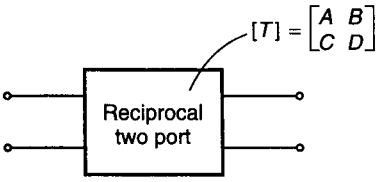


Figure 1.13 Reciprocal two-port network.

$$\begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \\ I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix} = \begin{bmatrix} [A] & [B] \\ [C] & [D] \end{bmatrix} \begin{bmatrix} V_{N+1} \\ V_{N+2} \\ \vdots \\ V_{2N} \\ I_{N+1} \\ I_{N+2} \\ \vdots \\ I_{2N} \end{bmatrix} \tag{1.62}$$

where $[A]$, $[B]$, $[C]$, and $[D]$ are $N \times N$ square matrices. For a two-port network, Eq. (1.62) reduces to

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \tag{1.63}$$

For a reciprocal two-port network, shown in Fig. 1.13, it can be shown that

$$AD - BC = 1 \tag{1.64}$$

or in general,

$$[A][D] - [B][C] = [U] \tag{1.65}$$

where $[U]$ is the identity matrix of order N.

As mentioned above, the $ABCD$ matrix is very useful in obtaining the overall response of a chain connection of a number of two-port networks. Let us consider the chain connection of two two-port networks, as shown in Fig. 1.14.

Let the $ABCD$ matrices of the individual networks be $[T_1]$ and $[T_2]$. Then it can be very easily shown that the $ABCD$ matrix of the cascaded network is given by

$$[T] = [T_1][T_2] \tag{1.66}$$

Equation (1.66) can be generalized for the cascade of more than two networks using the same principle.

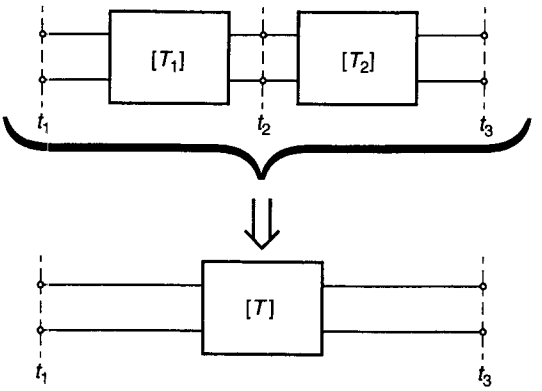


Figure 1.14 Two-port networks in tandem.

Use of ABCD Matrix in Computing Network Properties. In the previous section we have shown that a chain of cascaded two-port networks can be reduced to an equivalent two-port network by finding the overall *ABCD* matrix. Let such an equivalent two-port network be fed at the input port by a voltage source of output impedance Z_g and terminated at the output port by a load Z_L, as shown in Fig. 1.15.

Using Eq. (1.63) and the relationship

$$V_2 = Z_L I_2 \tag{1.67}$$

it can be shown that the input impedance of the network is given by

$$Z_{\text{in}} = \frac{Z_L A + B}{Z_L C + D} \tag{1.68}$$

and the output impedance is

$$Z_{\text{out}} = \frac{Z_g D + C}{Z_g B + A} \tag{1.69}$$

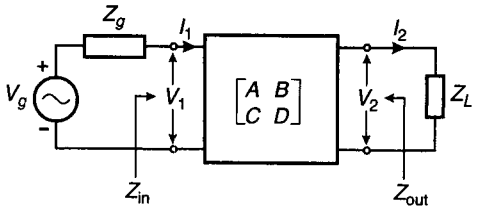


Figure 1.15 Two-port network *ABCD* matrix analysis.

The reflection coefficients looking into the input and the output are given by

$$\Gamma_{\text{in}} = \frac{Z_{\text{in}} - Z_g}{Z_{\text{in}} + Z_g} \tag{1.70}$$

and

$$\Gamma_{\text{out}} = \frac{Z_{\text{out}} - Z_L}{Z_{\text{out}} + Z_L} \tag{1.71}$$

respectively. The voltage gain is given by

$$A_v = \frac{V_L}{V_g} = \frac{V_2}{V_g} = \frac{Z_L}{Z_L A + B + Z_L Z_g C + Z_g D} \tag{1.72}$$

The power is given by

$$G = \frac{P_L}{P_1} = \frac{\text{Re } Z_{\text{in}}}{\text{Re } Z_L} \left| \frac{Z_L}{Z_L A + B} \right|^2 \tag{1.73}$$

where

$$P_1 = V_1 I_1 \tag{1.74}$$

is power delivered to the input port of the network and

$$P_L = V_L I_2 \tag{1.75}$$

is the power delivered to the load. The transducer power gain is given by

$$G_T = \frac{P_L}{P_{AG}} = 4 \frac{\text{Re } Z_g}{\text{Re } Z_L} \left| \frac{Z_L}{Z_L A + B + Z_L Z_g C + Z_g D} \right|^2 \tag{1.76}$$

where P_{AG} is available power from the source.

Normalized ABCD Matrix. Using the normalized voltage and current definitions in Eq. (1.17), we can define the normalized $ABCD$ matrix as follows [14]. Suppose the multiport network in Fig. 1.12 is terminated by a set of impedances Z_{Ii} $(i = 1, 2, \ldots, 2N)$. Then we can define a set of normalized voltages and currents v_i and i_i respectively $(i = 1, 2, \ldots, 2N)$ according to Eq. (1.17). These normalized voltages and currents can be related using the normalized $ABCD$ matrix as follows:

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \\ i_1 \\ i_2 \\ \vdots \\ i_N \end{bmatrix} = \begin{bmatrix} [A_n] & [B_n] \\ [C_n] & [D_n] \end{bmatrix} \begin{bmatrix} v_{N+1} \\ v_{N+2} \\ \vdots \\ v_{2N} \\ i_{N+1} \\ i_{N+2} \\ \vdots \\ i_{2N} \end{bmatrix} \tag{1.77}$$

where $[A_n]$, $[B_n]$, $[C_n]$, and $[D_n]$ are $N \times N$ square matrices. For a two-port network, Eq. (1.77) reduces to

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix} \begin{bmatrix} v_2 \\ i_2 \end{bmatrix} \tag{1.78}$$

The unnormalized and the normalized *ABCD* matrices are related through

$$\begin{bmatrix} [A] & [B] \\ [C] & [D] \end{bmatrix} = \begin{bmatrix} [Z_{IN}]^{-1/2} & [0] \\ [0] & [Z_{IN}]^{1/2} \end{bmatrix}^{-1} \begin{bmatrix} [A_n] & [B_n] \\ [C_n] & [D_n] \end{bmatrix} \begin{bmatrix} [Z_{I2N}]^{-1/2} & [0] \\ [0] & [Z_{I2N}] \end{bmatrix} \tag{1.79}$$

where $[Z_{IN}]$ is a diagonal matrix of real elements $Z_{I1}, Z_{I2}, \ldots, Z_{IN}$ and $[Z_{I2N}]$ is a diagonal matrix of real elements $Z_{I(N+1)}, Z_{I(N+2)}, \ldots, Z_{I(2N)}$. In case of a simple two-port network, Eq. (1.79) reduces to

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sqrt{Z_{I1}}} & [0] \\ [0] & \sqrt{Z_{I1}} \end{bmatrix}^{-1} \begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sqrt{Z_{I2}}} & [0] \\ [0] & \sqrt{Z_{I2}} \end{bmatrix} \tag{1.80}$$

Obviously the normalized *ABCD* matrix takes into consideration the effects of the impedances terminating the ports of a network. It can be shown that the scattering matrix of a two-port network is related to its normalized *ABCD* matrix via

$$[S] = \begin{bmatrix} A_n + B_n - C_n - D_n & 2(A_nD_n - C_nB_n) \\ 2 & D_n + B_n - A_n - C_n \end{bmatrix} \tag{1.81}$$

REFERENCES

1. Bhartia, P., and I. J. Bahl, *Millimeter Wave Engineering and Applications*, Wiley, New York, 1984.

2. Barrett, R. M., "Microwave Printed Circuits—the Early Years," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-32, Sept. 1984, pp. 983–990.
3. Barrett, R. M., and M. H. Barnes, "Microwave Printed Circuits," paper presented at the IRE National Conference on Airborne Electronics, Dayton, OH, May 23–25, 1951.
4. Sanders Associates, *Handbook of Tri-Plate Microwave Components*, Sanders Associates, Nashua, NH, 1956.
5. Greig, D. D., and H. F. Engelmann, "Microstrip—a New Transmission Technique for the Kilomegacycle L Range," *IRE Proc.*, Vol 40, Dec. 1952, pp. 1644–1650.
6. King, D. D., "Properties of Dielectric Image Lines," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-3, Mar. 1955, pp. 75–78.
7. Howe, H., "Microwave Integrated Circuits—an Historical Perspective," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-32, Sept. 1984, pp. 991–996.
8. Mcquiddy, Jr. D. N., J. W. Wassel, J. B. Lagrange, and W. R. Wisseman, "Monolithic Microwave Integrated Circuits: An Historical Perspective," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-32, Sept. 1984, pp. 997–1008.
9. Collin, R. E., *Foundations for Microwave Engineering*, McGraw-Hill, New York, 1986.
10. Penfield, P., "Noise in Negative Resistance Amplifiers," *IRE Trans. Circuit Theory*, Vol. CT-7, 1960, pp. 160–170.
11. Kurokawa, K., "Power Waves and the Scattering Matrix," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-13, No. 2, Mar. 1965, pp. 194–202.
12. Schiek, B., and A. Gronefeld, "Standing Wave Meters and Network Analyzers," in *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 20, Wiley, New York, 1999, pp. 403–423.
13. Carlin, H. J., and A. B. Giordano, *Network Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1964.
14. Kajfez, D., *Notes on Microwave Circuits*, Vol. 1, Kajfez Consulting, Oxford, MS, 1986.

2

TRANSMISSION LINES AND LUMPED ELEMENTS

Inder Bahl

Transmission lines in microwave circuits are normally used to carry information or energy from one point to the other and as circuit elements for passive circuits such as filters, impedance transformers, couplers, delay lines, and baluns. Passive elements in conventional microwave circuits are mostly distributed and employ sections of transmission lines and waveguides. This is because the sizes of discrete lumped elements (resistors, inductors, and capacitors) used in electronic circuits at lower frequencies become comparable to the wavelength at microwave frequencies. However, when the sizes of lumped elements are reduced to dimensions much smaller than the wavelength, they are also used at microwave frequencies.

This chapter is intended to provide accurate and simple closed-form expressions for characteristics of various types of lines and coupled structures. Brief descriptions of conventional waveguide and coaxial lines are included to help the reader in designing packages and good transitions between these media and planar transmission lines. Design information for lumped elements (inductors, capacitors, and resistors) is also included in the last section.

2.1 TRANSMISSION LINES

Multiconductor structures that support TEM or non-TEM modes of propagation are commonly referred to as "transmission lines." Waveguide (single con-

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

ductor) or dielectric rod (nonconductor) or their derivatives support the non-TEM mode of propagation. The TEM transmission lines are characterized by four basic parameters, the characteristic impedance Z_0, the phase velocity v_p, the attenuation constant α, and peak power-handling capability $P_{\max}$, in terms of physical parameters (such as the geometric cross section) and properties of the dielectric and the conductor materials used.

In a TEM line (perfectly terminated), the ratio of the voltage to the current at any point along the line is constant, having the units of resistance for a lossless medium. This ratio is defined as the characteristic impedance. The propagation constant for a lossy transmission structure is a complex quantity, comprising a real part known as the attenuation constant (which contains information about dissipation due to conductor and dielectric losses) and an imaginary part known as the phase constant (which contains information about phase velocity). The attenuation constant is defined as

$$\alpha = \frac{\text{average power lost per unit length}}{2 \times \text{power transmitted}} \tag{2.1}$$

Consider a uniform transmission line with series resistance ($\bar{R}$), series inductance ($\bar{L}$), shunt conductance ($\bar{G}$), and shunt capacitance ($\bar{C}$), all defined per unit length of the line, as shown in Fig. 2.1. Important transmission-line expressions are summarized in Table 2.1.

An extensive variety of transmission and waveguide structures are used at microwave frequencies. Figure 2.2 shows cross-sectional views of commonly used structures. Half-wavelength, quarter-wavelength, or smaller sections of these lines form the basic building blocks in most microwave circuits.

The power-handling capability of a transmission line is limited by dielectric breakdown and by heating due to attenuation. The electrical breakdown limits the peak power, while the increase in temperature due to conductor and dielectric losses limits the average power.

At normal temperature and pressure, the breakdown electric field of dry air is 2.9×10^6 V/m. Using this and by calculating the maximum field strength, the peak power-handling capability of a transmission line is readily determined.

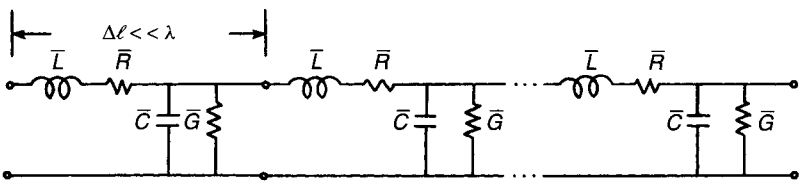


Figure 2.1 Lumped circuit representation of a transmission line.

Table 2.1 Summary of General Formulas for Transmission Lines

Quantity	General Line Expressions	Ideal Line Expressions	Approximate Results for Low-Loss Lines
Propagation constant $\gamma = \alpha + j\beta$	$\sqrt{(\bar{R}+j\omega\bar{L})(\bar{G}+j\omega\bar{C})}$	$j\omega\sqrt{\bar{L}\bar{C}}$	See α and β below
Phase constant β	$\mathrm{Im}(\gamma)$	$\omega\sqrt{\bar{L}\bar{C}} = \dfrac{\omega}{v_p} = \dfrac{2\pi}{\lambda}$	$\omega\sqrt{\bar{L}\bar{C}}\left(1 - \dfrac{\bar{R}\bar{G}}{4\omega^2\bar{L}\bar{C}} + \dfrac{\bar{G}^2}{8\omega^2\bar{C}^2} + \dfrac{\bar{R}^2}{8\omega^2\bar{L}^2}\right)$
Attenuation constant α	$\mathrm{Re}(\gamma)$	0	$\dfrac{\bar{R}}{2Z_0} + \dfrac{\bar{G}Z_0}{2}$
Characteristic impedance Z_0	$\sqrt{\dfrac{\bar{R}+j\omega\bar{L}}{\bar{G}+j\omega\bar{C}}}$	$\sqrt{\dfrac{\bar{L}}{\bar{C}}}$	$\sqrt{\dfrac{\bar{L}}{\bar{C}}}\left[1 + j\left(\dfrac{\bar{G}}{2\omega\bar{C}} - \dfrac{\bar{R}}{2\omega\bar{L}}\right)\right]$
Input impedance Z_i	$Z_0\left(\dfrac{Z_L \cosh \gamma l + Z_0 \sinh \gamma l}{Z_0 \cosh \gamma l + Z_L \sinh \gamma l}\right)$	$Z_0\left(\dfrac{Z_L \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_L \sin \beta l}\right)$	
Impedance of shorted line ($Z_L = 0$)	$Z_0 \tanh \gamma l$	$jZ_0 \tan \beta l$	$Z_0\left(\dfrac{\alpha l \cos \beta l + j \sin \beta l}{\cos \beta l + j\alpha l \sin \beta l}\right)$
Impedance of open line ($Z_L = \infty$)	$Z_0 \coth \gamma l$	$-jZ_0 \cot \beta l$	$Z_0\left(\dfrac{\cos \beta l + j\alpha l \sin \beta l}{\alpha l \cos \beta l + j \sin \beta l}\right)$
Impedance of quarter-wave line	$Z_0\left(\dfrac{Z_L \sinh \alpha l + Z_0 \cosh \alpha l}{Z_0 \sinh \alpha l + Z_L \cosh \alpha l}\right)$	$\dfrac{Z_0^2}{Z_L}$	$Z_0\left(\dfrac{Z_0 + Z_L \alpha l}{Z_L + Z_0 \alpha l}\right)$
Impedance of half-wave line	$Z_0\left(\dfrac{Z_L \cosh \alpha l + Z_0 \sinh \alpha l}{Z_0 \cosh \alpha l + Z_L \sinh \alpha l}\right)$	Z_L	$Z_0\left(\dfrac{Z_L + Z_0 \alpha l}{Z_0 + Z_L \alpha l}\right)$
Reflection coefficient ρ (at the load location)	$\dfrac{Z_L - Z_0}{Z_L + Z_0}$	$\dfrac{Z_L - Z_0}{Z_L + Z_0}$	
Standing-wave ratio	$\dfrac{1 + \lvert\rho\rvert}{1 - \lvert\rho\rvert}$	$\dfrac{1 + \lvert\rho\rvert}{1 - \lvert\rho\rvert}$	

Note: λ = wavelength measured along line; ω = angular frequency.

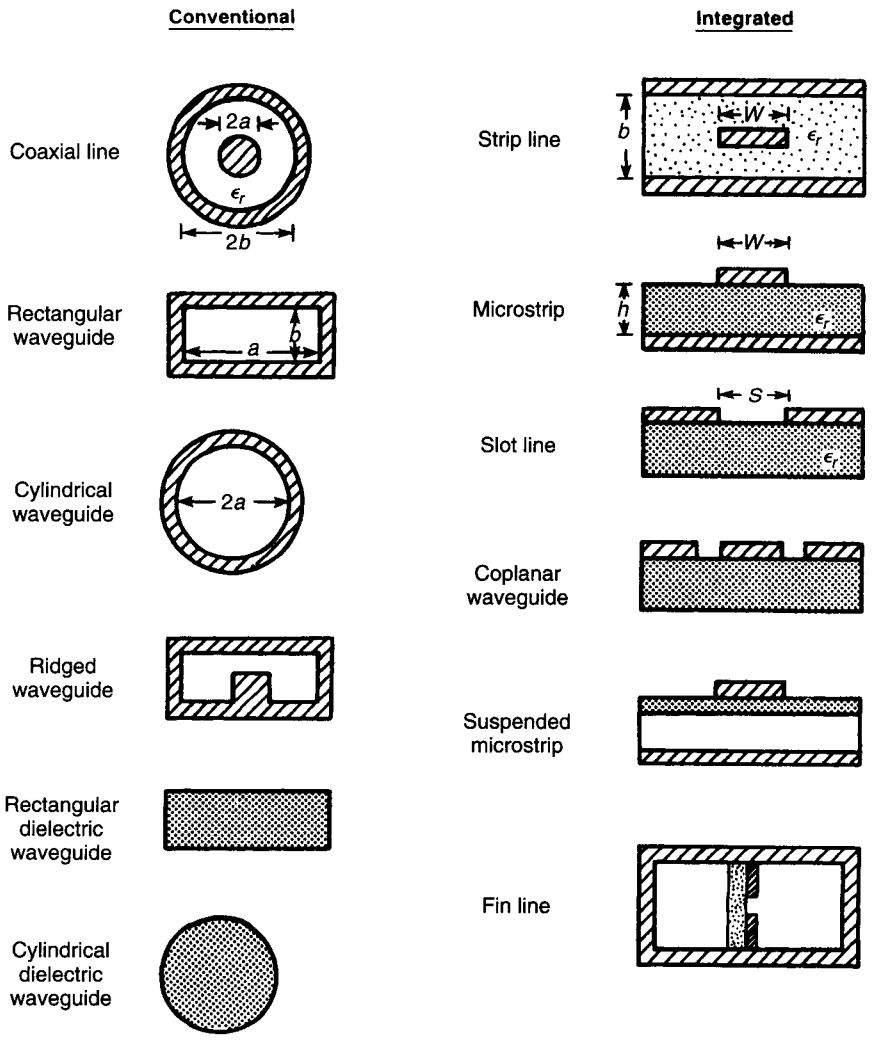


Figure 2.2 Transmission structures for microwave circuits.

The average power-handling capability of a transmission is determined by the temperature rise of the line in an air environment. The parameters that play major roles in the calculation of average power capacity are (1) attenuation constant, (2) surface area of the line, (3) maximum tolerable temperature rise, and (4) ambient temperature (i.e., the temperature of the medium surrounding the transmission structure. Thus the average power rating can be raised, if desired, by choosing a higher temperature limit and using forced cooling or cooling fins.

2.1.1 Characteristics of Conventional Transmission Structures

Coaxial lines and waveguides are frequently used in microwave circuits. These are briefly discussed in this section.

Coaxial Line. The dominant mode of propagation in a coaxial line is the TEM, and the characteristics obtained from static field analysis [1, 2] are given in Table 2.2. Shown in Figs. 2.2*a* and *b* are the radii of the inner and outer conductors, respectively. The effect of conductor temperature and roughness on the attenuation is very well described by Bhartia and Bahl [3]. To compute attenuation for temperatures other than 20°C, multiply the attenuation expression α_c in Table 2.2 by $[1 + 0.0039(T - 20)]^{1/2}$, where the temperature T is expressed in degrees Celsius.

Waveguide. Waveguides of rectangular and circular cross sections (Fig. 2.2) find frequent applications in high-power and low-loss systems. The properties of waveguide structures can be determined by solving the wave equation with appropriate boundary conditions [1–3]. The structure supports TE and TM

Table 2.2 Coaxial-Line Characteristics

Parameter	Expression	Units
Capacitance	$C = \dfrac{55.556\epsilon_r}{\ln(b/a)}$	pF/m
Inductance	$L = 200 \ln \dfrac{b}{a}$	nH/m
Characteristic impedance	$Z_0 = \dfrac{60}{\sqrt{\epsilon_r}} \ln \dfrac{b}{a}$	Ω
Phase velocity	$v_p = \dfrac{3 \times 10^8}{\sqrt{\epsilon_r}}$	m/s
Delay	$\tau_d = 3.33\sqrt{\epsilon_r}$	ns/m
Dielectric attenuation constant	$\alpha_d = 27.3\sqrt{\epsilon_r}\dfrac{\tan\delta}{\lambda_0}$	dB/unit length
Conductor attenuation constant (for copper at 20°C)	$\alpha_c = \dfrac{9.5 \times 10^{-5}\sqrt{f}(a+b)\sqrt{\epsilon_r}}{ab \ln(b/a)}$	dB/unit length
Cutoff wavelength for higher order modes	$\lambda_c = \pi\sqrt{\epsilon_r}(a+b)$	Unit of a or b
Maximum peak power	$P_{\max} = 44\lvert E_{\max}\rvert^2 a^2 \sqrt{\epsilon_r} \ln \dfrac{b}{a}$	kW

Note: λ_0 = free-space wavelength; $\tan\delta$ = loss tangent of dielectric; f = operating frequency in GHz; $E_{\max}$ = breakdown electric field.

Table 2.3 Properties of Waves in Empty Rectangular Waveguides

Property	TE_{mn} Modes	TM_{mn} Modes
Cutoff wavenumber k_c	$\sqrt{\left(\frac{m\pi}{a}\right)^2+\left(\frac{n\pi}{b}\right)^2}$	$\sqrt{\left(\frac{m\pi}{a}\right)^2+\left(\frac{n\pi}{b}\right)^2}$
Propagation constant γ_{mn}	$\sqrt{k_c^2-k_0^2}$	$\sqrt{k_c^2-k_0^2}$
Guide wavelength λ_g	$\frac{\lambda_0}{\sqrt{1-(k_c/k_0)^2}}$	$\frac{\lambda_0}{\sqrt{1-(k_c/k_0)^2}}$
Group velocity v_g	$c\frac{\lambda_0}{\lambda_g}$	$c\frac{\lambda_0}{\lambda_g}$
Phase velocity v_p	$c\frac{\lambda_g}{\lambda_0}$	$c\frac{\lambda_g}{\lambda_0}$
Wave impedance Z	$\frac{jk_0\eta_0}{\gamma_{mn}}$	$\frac{-j\gamma_{mn}\eta_0}{k_0}$
Longitudinal magnetic field H_z	$k_c^2\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)$	0
Longitudinal electric field E_z	0	$k_c^2\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)$

Transverse magnetic field		
H_x	$\frac{\gamma_{mn} m\pi}{a} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right)$	$\frac{k_0 n\pi}{b\eta_0} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right)$
H_y	$\frac{\gamma_{mn} n\pi}{b} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right)$	$-\frac{k_0 m\pi}{a\eta_0} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right)$
Transverse electric field		
E_x	$\frac{jk_0\eta_0 n\pi}{b} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right)$	$-\frac{\gamma_{mn} m\pi}{a} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right)$
E_y	$-\frac{jk_0\eta_0 m\pi}{a} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right)$	$-\frac{\gamma_{mn} n\pi}{b} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right)$

waves, and the properties of these modes in a rectangular waveguide are summarized in Table 2.3. The quantities η_0, k_0, and c are defined as

$$\eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \qquad k_0 = \omega\sqrt{\mu_0\epsilon_0} \qquad c = \frac{1}{\sqrt{\mu_0\epsilon_0}} \tag{2.2}$$

where ϵ_0 and μ_0 are the free-space permittivity and permeability, respectively. The lowest order propagating mode is the TE_{10} ($m = 1, n = 0$) and is also called the dominant mode of the rectangular waveguide. Since the modes of propagation in waveguides are non-TEM, the characteristic impedance cannot be defined uniquely [4].

For these structures, Z_0 can be defined either in terms of the voltage–current ratio or in terms of the power transmitted (W) for a given voltage or a given current; that is,

$$Z_0(v,i) = \frac{v}{i} \quad \text{or} \quad Z_0(W,i) = \frac{2W}{ii^*} \quad \text{or} \quad Z_0(W,v) = \frac{vv^*}{2W} \tag{2.3}$$

For TEM-mode transmission structures, these definitions are identical, but for waveguides they lead to three different values of Z_0. All of these three dif-

Table 2.4 Rectangular Waveguide Characteristics

Parameter	Expression	Units
Cutoff wavelength	$\lambda_c = 2a$	m
Guide wavelength	$\lambda_g = \dfrac{\lambda_0}{\sqrt{1-(\lambda_0/\lambda_c)^2}}$	m
Phase velocity	$v_p = \dfrac{c\lambda_g}{\lambda_0}$	m/s
Wave impedance	$Z = \dfrac{\eta_0}{\sqrt{1-(\lambda_0/\lambda_c)^2}}$	Ω
Characteristic impedance (v,i)	$Z_0 = \dfrac{\pi b}{2a}Z$	Ω
Attenuation constant (copper)	$\alpha = \dfrac{7.14\times10^{-2}\sqrt{f}}{b\eta_0\sqrt{1-(\lambda_0/\lambda_c)^2}}\left[1+\dfrac{2b}{a}\left(\dfrac{\lambda_0}{\lambda_c}\right)^2\right]$	dB/unit length
Maximum peak power	$P_{\max} = 6.63\times10^{-7}ab\lvert E_{\max}\rvert^2\sqrt{1-\left(\dfrac{\lambda_0}{\lambda_c}\right)^2}$	kW

Note: $c = 3\times10^8$ m/s, f in GHz.

ferent characteristic impedances may be expressed in terms of the wave impedance Z, which is defined as the ratio of transverse components of electric and magnetic fields. For the dominant TE_{10} mode the characteristics of a rectangular waveguide are summarized in Table 2.4.

2.1.2 Characteristics of Planar Transmission Lines

For a transmission structure to be suitable as a circuit element in MICs, one of the principal requirements is that the structure should be "planar" in configuration. A planar geometry implies that the characteristics of the element can be determined from the dimensions in a single plane. As shown in Fig. 2.2, various forms of planar transmission lines have been developed for use in MICs. The strip line [5–11], microstrip line [12–14], inverted microstrip line [15], slot line [14, 16], coplanar waveguide [14, 17], and coplanar strip line [14, 17] are representative planar transmission lines. The circuits realized using any one of the aforementioned transmission lines or combinations of them have distinct advantages, such as light weight, small size, improved performance, better reliability and reproducibility, and low cost, as compared to conventional microwave circuits. They are also compatible with solid state chip devices. Integrated circuits employing these structures at microwave frequencies have been widely discussed in the literature [9, 14, 18, 19].

Strip Line. The strip line is one of the most commonly used transmission lines for passive MICs. The dominant mode in a strip line is TEM, and the design data can be obtained completely by electrostatic analysis such as conformal mapping. Expressions for Z_0, $\alpha_T = \alpha_c + \alpha_d$, and f_c are given in Table 2.5, where K represents a complete elliptic function of the first kind with K' its complementary function. An approximate expression for K/K' (accurate to 8 ppm) is given by

$$\frac{K(k)}{K'(k)} = \begin{cases} \left[\dfrac{1}{\pi}\ln\left(2\dfrac{1+\sqrt{k'}}{1-\sqrt{k'}}\right)\right]^{-1} & \text{for } 0 \le k \le 0.7 \\ \dfrac{1}{\pi}\ln\left(2\dfrac{1+\sqrt{k}}{1-\sqrt{k}}\right) & \text{for } 0.7 \le k \le 1 \end{cases} \tag{2.4}$$

where $K'(k) = K(k')$, $k' = \sqrt{1-k^2}$. The characteristic impedance versus W/b for various values of t/b is shown in Fig. 2.3.

The total loss can be used to determine the quality factor Q of a $\frac{1}{2}\lambda$ resonator as

$$Q = \frac{8.68\pi\sqrt{\epsilon_r}}{\lambda_0 \alpha_T} \tag{2.5}$$

Table 2.5 Strip-Line Characteristics

Parameter	Expressions	Remarks
Characteristic impedance (Ω) for $t = 0$	$Z_0 = \frac{30\pi}{\sqrt{\epsilon_r}} \frac{K'(k)}{K(k)}, \qquad k = \tanh \frac{\pi W}{2b}$	Virtually exact [5]
Characteristic impedance (Ω) for $t \neq 0$	$Z_0 = \frac{30 \ln}{\sqrt{\epsilon_r}} \left\{ 1 + \frac{4}{\pi} \frac{b-t}{W'} \left[\frac{8}{\pi} \frac{b-t}{W'} + \sqrt{\left(\frac{8}{\pi} \frac{b-t}{W'} \right)^2 + 6.27} \right] \right\}$	For $W'/(b-t) < 10$ accuracy within 0.5% [11]
	$\frac{W'}{b-t} = \frac{W}{b-t} + \frac{\Delta W}{b-t}$	
	$\frac{\Delta W}{b-t} = \frac{x}{\pi(1-x)} \left\{ 1 - \frac{1}{2} \ln \left[\left(\frac{x}{2-x} \right)^2 + \left(\frac{0.0796x}{W/b + 1.1x} \right)^m \right] \right\}$	
	$m = 2\left[1 + \frac{2}{3} \frac{x}{1-x} \right]^{-1}, \qquad x = \frac{t}{b}$	
Attenuation constant (dB/unit length)	$\alpha_c = \frac{0.0231 R_s \sqrt{\epsilon_r}}{Z_0} \frac{\partial Z_0}{\partial W'} \left\{ 1 + \frac{2W'}{b-t} - \frac{1}{\pi} \left[\frac{3x}{2-x} + \ln \frac{x}{2-x} \right] \right\}$	[4]
	$\frac{\partial Z_0}{\partial W'} = \frac{30 e^{-A}}{W' \sqrt{\epsilon_r}} \left[\frac{3.135}{Q} - \left(\frac{8}{\pi} \frac{b-t}{W'} \right)^2 (1+Q) \right]$	
	$A = \frac{Z_0 \sqrt{\epsilon_r}}{30\pi}, \qquad Q = \sqrt{1 + 6.27 \left(\frac{\pi}{8} \frac{W'}{b-t} \right)^2}$	
	$\alpha_d = 27.3 \sqrt{\epsilon_r} \frac{\tan \delta}{\lambda_0}$	
Cutoff for higher order mode (GHz)	$f_c = \frac{15}{b\sqrt{\epsilon_r}} \frac{1}{(W/b + \pi/4)}, \qquad W$ and b in cm	TE is the lowest order mode [4]

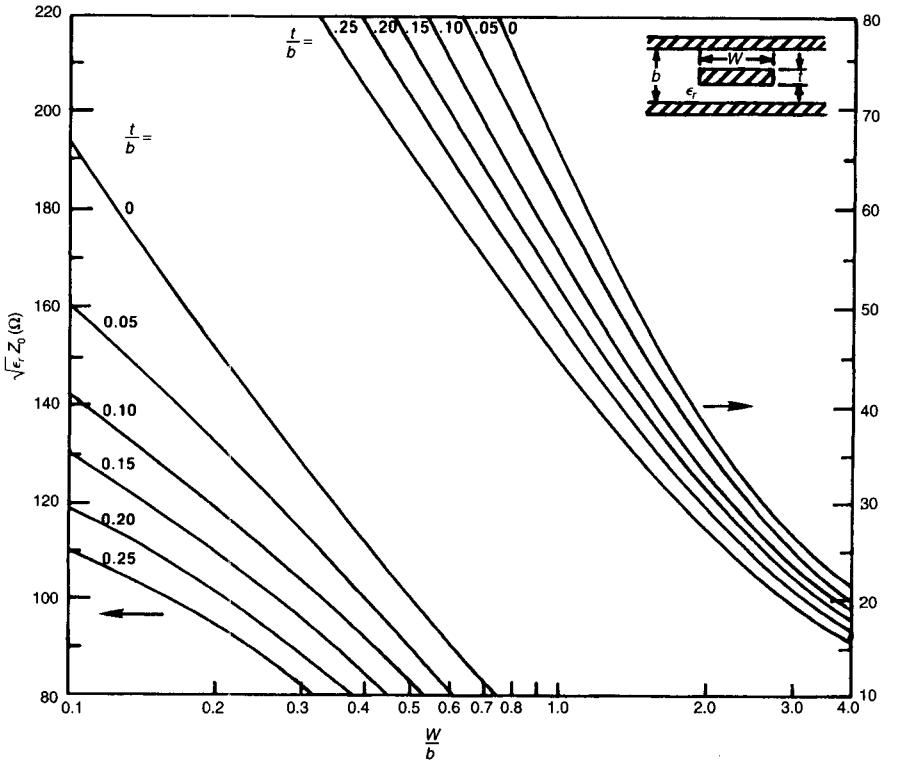


Figure 2.3 Strip-line characteristic impedance versus W/b for various values of t/b.

Microstrip. Unlike the strip line, the microstrip line (shown in Fig. 2.4) is an inhomogeneous transmission line, since the field lines between the strip and the ground plane are not contained entirely in the substrate. Therefore, the mode propagating along the microstrip is not purely TEM but quasi-TEM.

Extensive literature dealing with the analytical and numerical solutions of this medium exists [14]. Of these solutions, the quasi-static approach is perhaps

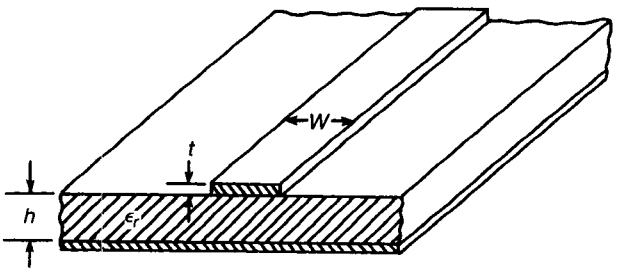


Figure 2.4 Microstrip-line configuration.

Table 2.6 Microstrip-Line Characteristics

Parameters	Expressions
Characteristic impedance (Ω)	$Z_0 = \begin{cases} \dfrac{\eta_0}{2\pi\sqrt{\epsilon_e}} \ln\left(\dfrac{8h}{W'} + 0.25\dfrac{W'}{h}\right), & \dfrac{W}{h} \le 1 \\ \dfrac{\eta_0}{\sqrt{\epsilon_e}}\left[\dfrac{W'}{h} + 1.393 + 0.667 \ln\left(\dfrac{W'}{h} + 1.444\right)\right]^{-1}, & \dfrac{W}{h} \ge 1 \end{cases}$
	$\eta_0 = 120\pi\Omega$
	$\dfrac{W'}{h} = \dfrac{W}{h} + \dfrac{1.25}{\pi}\dfrac{t}{h}\left(1 + \ln\dfrac{4\pi W}{t}\right), \quad \dfrac{W}{h} \le \dfrac{1}{2\pi}$
	$\dfrac{W'}{h} = \dfrac{W}{h} + \dfrac{1.25}{\pi}\dfrac{t}{h}\left(1 + \ln\dfrac{2h}{t}\right), \quad \dfrac{W}{h} \ge \dfrac{1}{2\pi}$
Effective dielectric constant	$\epsilon_e = \dfrac{\epsilon_r + 1}{2} + \dfrac{\epsilon_r - 1}{2} F\left(\dfrac{W}{h}\right) - \dfrac{\epsilon_r - 1}{4.6}\dfrac{t/h}{\sqrt{W/h}}$
	$F\left(\dfrac{W}{h}\right) = \begin{cases} \left(1 + 12\dfrac{h}{W}\right)^{-1/2} + 0.04\left(1 - \dfrac{W}{h}\right)^2, & \dfrac{W}{h} \le 1 \\ \left(1 + 12\dfrac{h}{W}\right)^{-1/2} & \dfrac{W}{h} \ge 1 \end{cases}$

Attenuation constant (dB/unit length)

$$\alpha_c = \begin{cases} 1.38\dfrac{R_s}{hZ_0}\dfrac{32-(W'/h)^2}{32+(W'/h)^2}\Lambda, & \dfrac{W}{h} \le 1 \\ 6.1\times 10^{-5}\dfrac{R_s Z_0 \epsilon_e}{h}\left(\dfrac{W'}{h}+\dfrac{0.667W'/h}{W'/h+1.444}\right)\Lambda, & \dfrac{W}{h} \ge 1 \end{cases}$$

$$\Lambda = \begin{cases} 1+\dfrac{h}{W'}\left(1+\dfrac{1.25t}{\pi W}+\dfrac{1.25}{\pi}\ln\dfrac{4\pi W}{t}\right), & \dfrac{W}{h} \le \dfrac{1}{2\pi} \\ 1+\dfrac{h}{W'}\left(1-\dfrac{1.25t}{\pi h}+\dfrac{1.25}{\pi}\ln\dfrac{2h}{t}\right), & \dfrac{W}{h} \ge \dfrac{1}{2\pi} \end{cases}$$

$$\alpha_d = 27.3\frac{\epsilon_r}{\sqrt{\epsilon_e}}\left(\frac{\epsilon_e-1}{\epsilon_r-1}\right)\frac{\tan\delta}{\lambda_0}$$

Dispersion

$$\epsilon_e(f) = \left(\frac{\sqrt{\epsilon_r}-\sqrt{\epsilon_e}}{1+4F^{-1.5}}+\sqrt{\epsilon_e}\right)^2$$

$$Z_0(f) = Z_0\frac{\epsilon_e(f)-1}{\epsilon_e-1}\sqrt{\frac{\epsilon_e}{\epsilon_e(f)}}$$

$$F = \frac{4h\sqrt{\epsilon_r-1}}{\lambda_0}\left\{0.5+\left[1+2\log\left(1+\frac{W}{h}\right)\right]^2\right\}$$

Quality factor

$$\frac{1}{Q} = \frac{1}{Q_0}+\frac{1}{Q_r}, \qquad Q_0 = \frac{8.68\pi\sqrt{\epsilon_e(f)}}{\lambda_0(\alpha_c+\alpha_d)}, \qquad Q_r = \frac{Z_0(f)}{480\pi(h/\lambda_0)^2 R}$$

$$R = \frac{\epsilon_e(f)+1}{\epsilon_e(f)}-\frac{[\epsilon_e(f)-1]^2}{2[\epsilon_e(f)]^{3/2}}\ln\left[\frac{\sqrt{\epsilon_e(f)}+1}{\sqrt{\epsilon_e(f)}-1}\right]$$

the simplest but has a limited range of validity, while the full-wave approach is complete and rigorous. In the quasi-static method, the nature of the mode of propagation is considered to be pure TEM, and the transmission-line characteristics are calculated from the electrostatic capacitances of the structure. It is found that this analysis is adequate for designing circuits when the strip width and the substrate thickness are much smaller than the wavelength in the dielectric material. In the quasi-static approach, the transmission characteristics are calculated from the values of two capacitances: one is C_a, for a unit length of the microstrip configuration with the dielectric materials replaced by air, and the other is C, for a unit length of the microstrip with the dielectric present. The characteristic impedance Z_0 and the phase constant β can be written in terms of these capacitances as

$$Z_0 = \frac{1}{c\sqrt{CC_a}} \tag{2.6}$$

$$\beta = k_0\left(\frac{C}{C_a}\right)^{1/2} = k_0\sqrt{\epsilon_e} \tag{2.7}$$

where

$$\epsilon_e = \left(\frac{\lambda_0}{\lambda_g}\right)^2 = \frac{C}{C_a} \tag{2.8}$$

λ_0 being the free-space wavelength and λ_g the guide wavelength. The effective dielectric constant ϵ_e takes into account the fields in the air regions. Numerical methods for the characterization of microstrip lines involve extensive computations. Closed-form expressions are necessary for optimization and computer-aided design of microstrip circuits. Closed-form expressions for Z_0 and ϵ_e have been reported by Wheeler [13, 20], Schneider [21], and Hammerstad [22, 23]. Both Wheeler and Hammerstad have also given synthesis expressions for Z_0. Closed-form expressions for microstrip characteristics are summarized in Table 2.6. The characteristic impedance and effective dielectric constant versus W/h are shown in Fig. 2.5.

The maximum frequency of operation with a microstrip is limited due to several factors such as excitation of spurious modes, higher losses, tight fabrication tolerances, handling fragility, pronounced discontinuity effects, low Q due to radiation from discontinuities, and, of course, technological processes. The frequency at which significant coupling occurs between the quasi-TEM mode and the lowest order surface-wave spurious mode is given in Bahl and Trivedi [24]:

$$f_T = \frac{150}{\pi h}\sqrt{\frac{2}{\epsilon_r - 1}}\tan^{-1}(\epsilon_r) \tag{2.9}$$

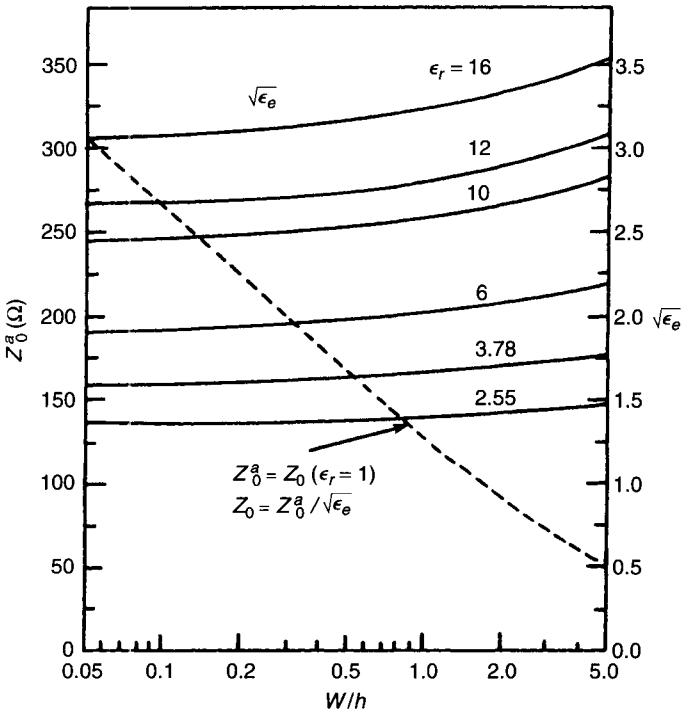


Figure 2.5 Characteristic impedance and effective dielectric constant of open microstrip lines.

where f_T is in gigahertz and h is in millimeters. Thus the maximum thickness of the GaAs substrate ($\epsilon_r \simeq 12.9$) for microstrip circuits designed at 100 GHz is less than 0.3 mm. The maximum thickness is also restricted by the high radiation losses incurred in microstrip discontinuities, such as open ends, gaps, slits, steps in widths, and bends. For a $\frac{1}{2}\lambda$ resonator, the radiation Q factor may be approximated as

$$Q_r = \frac{3\epsilon_r Z_0 \lambda_0^2}{32\eta_0 h^2} \tag{2.10}$$

Thus for thicker substrates, where $Q \simeq Q_r$, the variation of Q is proportional to $1/(fh)^2$. For example, a 50-Ω resonator on a GaAs substrate has a $Q_r \simeq 1.44 \times 10^4/(fh)^2$, where f is in gigahertz and h is in millimeters. At 100 GHz, the substrate thickness is <0.125 mm for $Q_r > 100$. A substrate thickness of this order results in attenuation of the order of 1 dB/cm. Fabrication tolerances and technological processes such as photoetching limit the minimum strip width and the spacing between two adjacent strips in the case of coupled lines. High-impedance lines of about 100 Ω require strip widths of the order of

0.01 mm on a 0.125-mm-thick GaAs substrate, thereby also setting a limit on the frequency of operation in microstrip lines because of low radiation Q_r.

Suspended and Inverted Microstrip Lines. Suspended and inverted microstrip lines (shown in Fig. 2.6) provide a higher Q (500–1500) than microstrip. The wide range of impedance values achievable makes these media particularly suitable for filters. Expressions for the characteristic impedance and effective dielectric constant for $t/h \ll 1$ are given as [25]

$$Z_0 = \frac{60}{\sqrt{\epsilon_e}} \ln \left[\frac{f(u)}{u} + \sqrt{1 + \left(\frac{2}{u}\right)^2} \right] \tag{2.11}$$

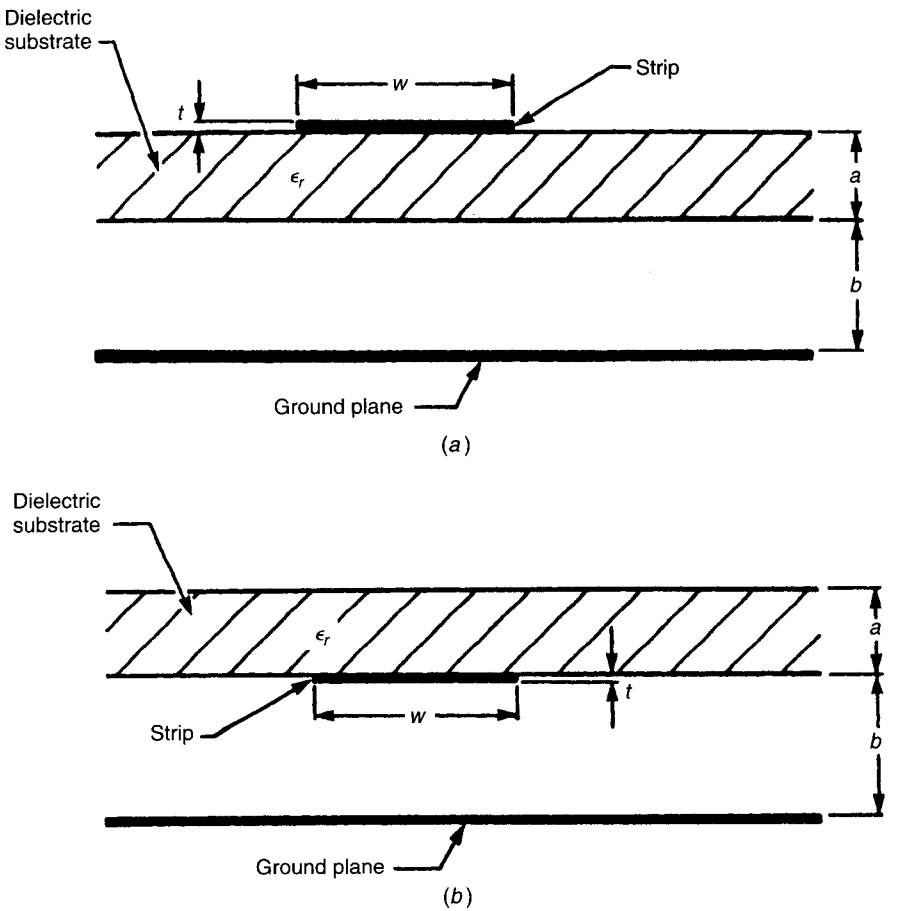


Figure 2.6 Suspended-substrate microstrip-line configurations: (*a*) suspended; (*b*) inverted.

where

$$f(u) = 6 + (2\pi - 6)\exp\left[-\left(\frac{30.666}{u}\right)^{0.7528}\right]$$

For suspended microstrip $u = W/(a+b)$, and for inverted microstrip $u = W/b$, where all the variables are defined in Fig. 2.6.

For suspended microstrip the effective dielectric constance ϵ_e is obtained from

$$\sqrt{\epsilon_e} = \left[1 + \frac{a}{b}\left(a_1 - b_1 \ln \frac{W}{b}\right)\left(\frac{1}{\sqrt{\epsilon_r}} - 1\right)\right]^{-1} \tag{2.12}$$

where

$$a_1 = \left(0.8621 - 0.1251 \ln \frac{a}{b}\right)^4$$

$$b_1 = \left(0.4986 - 0.1397 \ln \frac{a}{b}\right)^4$$

and for inverted microstrip the effective dielectric constant is obtained from

$$\sqrt{\epsilon_e} = 1 + \frac{a}{b}\left(\bar{a}_1 - \bar{b}_1 \ln \frac{w}{b}\right)(\sqrt{\epsilon_r} - 1) \tag{2.13}$$

where

$$\bar{a}_1 = \left(0.5173 - 0.1515 \ln \frac{a}{b}\right)^2$$

$$\bar{b}_1 = \left(0.3092 - 0.1047 \ln \frac{a}{b}\right)^2$$

The accuracy of (2.12) and (2.13) is within $\pm 1\%$ for $1 < W/b \leq 8$, $0.2 \leq a/b \leq 1$ and $\epsilon_r \leq 6$. For $\epsilon_r \simeq 10$, the error is less than $\pm 2\%$.

Slot Line. The slot-line configuration (shown in Fig. 2.7) is useful in circuits requiring high-impedance lines, series stubs, and short circuits and in hybrid combinations with microstrip circuits in MICs. The mode of propagation is non-TEM and almost TE in nature. Various methods of analysis discussed in the literature do not lead to closed-form expressions for slot-line wavelength and impedance. This becomes a serious handicap for circuit analysis and design, especially when computer-aided design techniques are used. However, closed-form expressions for characteristic impedance and slot-line wavelength have been obtained [26] by curve fitting the numerically computed results based on work by Cohn [27]. These expressions have an accuracy of about 2% for the

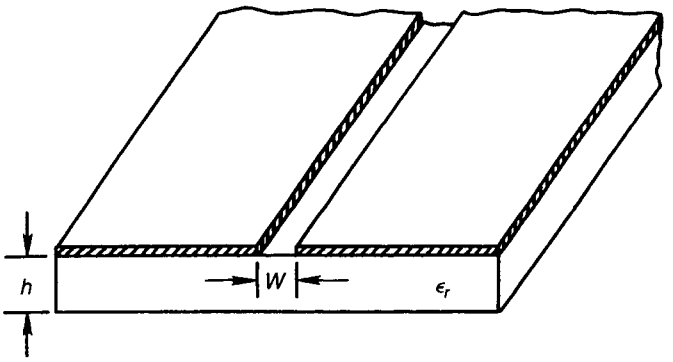


Figure 2.7 Slot-line configuration.

following sets of parameters:

$$9.7 \leq \epsilon_r \leq 20$$

$$0.02 \leq \frac{W}{h} \leq 1.0$$

$$0.01 \leq \frac{h}{\lambda_0} \leq \left(\frac{h}{\lambda_0}\right)_c$$

where $(h/\lambda_0)_c$ is the cutoff value for the TE_{10} surface-wave mode on the slot line and is given by

$$\left(\frac{h}{\lambda_0}\right)_c = \frac{0.25}{\sqrt{\epsilon_r - 1}} \tag{2.14}$$

1. For $0.02 \leq W/h \leq 0.2$

$$\frac{\lambda_g}{\lambda_0} = 0.923 - 0.195 \ln \epsilon_r + 0.2\frac{W}{h} - \left(0.126\frac{W}{h} + 0.02\right) \ln\left(\frac{h}{\lambda_0} \times 10^2\right) \tag{2.15}$$

$$Z_0 = 72.62 - 15.283 \ln \epsilon_r + 50\frac{(W/h - 0.02)(W/h - 0.1)}{W/h} + \ln\left(\frac{W}{h} \times 10^2\right)(19.23 - 3.693 \ln \epsilon_r) - \left[0.139 \ln \epsilon_r - 0.11 + \frac{W}{h}(0.465 \ln \epsilon_r + 1.44)\right] \times \left(11.4 - 2.636 \ln \epsilon_r - \frac{h}{\lambda_0} \times 10^2\right)^2 \tag{2.16}$$

2. For $0.2 \leq W/h \leq 1.0$

$$\frac{\lambda_g}{\lambda_0} = 0.987 - 0.21 \ln \epsilon_r + \frac{W}{h}(0.111 - 0.0022\epsilon_r) - \left(0.053 + 0.041\frac{W}{h} - 0.0014\epsilon_r\right) \ln\left(\frac{h}{\lambda_0} + 10^2\right) \tag{2.17}$$

$$Z_0 = 113.19 - 23.257 \ln \epsilon_r + 1.25\frac{W}{h}(114.59 - 22.531 \ln \epsilon_r) + 20\left(\frac{W}{h} - 0.2\right)\left(1 - \frac{W}{h}\right) - \left[0.15 + 0.1 \ln \epsilon_r + \frac{W}{h}(-0.79 + 0.899 \ln \epsilon_r)\right] \times \left[10.25 - 2.171 \ln \epsilon_r + \frac{W}{h}(2.1 - 0.617 \ln \epsilon_r) - \frac{h}{\lambda_0} \times 10^2\right]^2 \tag{2.18}$$

More accurate expressions for slot-line wavelength for $\epsilon_r = 9.7$ and 20 are also available [4, 14].

Coplanar Lines. Coplanar waveguides (CPWs) are finding extensive applications in MICs. Inclusion of CPWs in microwave circuits adds to the flexibility of circuit design and improves the performance for some circuit functions. The configuration of a CPW is shown in Fig. 2.8*a*. Another promising configuration that is complementary to a CPW is known as a coplanar strip (CPS) and is

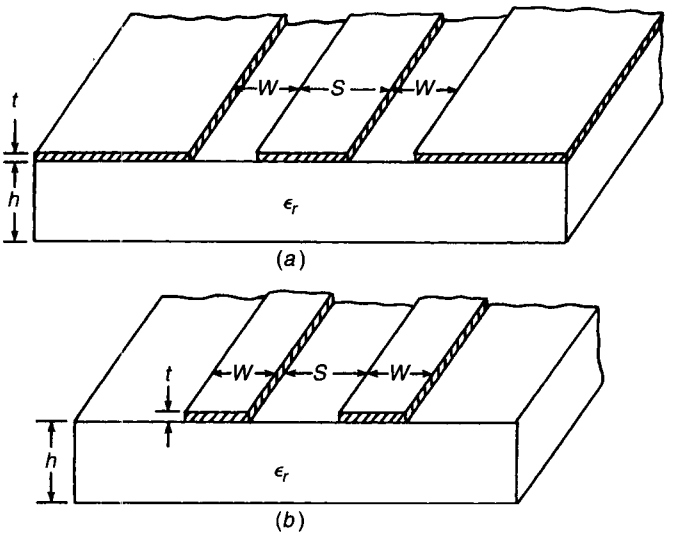


Figure 2.8 (*a*) Coplanar waveguide; (*b*) coplanar strips transmission line.

Table 2.7 Expressions for Coplanar-Line Characteristic Impedance and Effective Dielectric Constant

Structure	Characteristic Impedance (Ω)	Effective Dielectric Constant
Coplanar waveguide	$Z_0 = \dfrac{30\pi}{\sqrt{\epsilon_e}} \dfrac{K(k')}{K(k)}$	$\epsilon_e = 1 + \dfrac{\epsilon_r - 1}{2} \dfrac{K(k')K(k_1)}{K(k)K(k_1')}$
Coplanar strips	$Z_0 = \dfrac{120\pi}{\sqrt{\epsilon_e}} \dfrac{K(k)}{K(k')}$	$\epsilon_e = 1 + \dfrac{\epsilon_r - 1}{2} \dfrac{K(k')K(k_1)}{K(k)K(k_1')}$
	$k = \dfrac{a}{b}, \quad a = \dfrac{S}{2}, \quad b = \dfrac{S}{2} + W$	
	$k_1 = \dfrac{\sinh(\pi a/2h)}{\sinh(\pi b/2h)}$	

Source: Ghione, G., and C. Naldi, "Analytical Formulas for Coplanar Lines in Hybrid and Monolithic MICs," *Electron. Lett.*, Vol. 20, 1984, pp. 179–181.

shown in Fig. 2.8*b*. Both of these configurations belong to the category of "coplanar lines," wherein all the conductors are in the same plane (i.e., on the top surface of the dielectric substrate). A distinct advantage of these two lines lies in the fact that mounting lumped (active or passive) components in shunt or series configuration is much easier. Drilling holes or slots through the substrate is not needed.

Coplanar waveguides and coplanar strips support quasi-TEM modes and have been analyzed using quasi-static as well as full-wave methods [14]. Expressions for Z_0 and ϵ_e of CPWs and CPSs are given in Table 2.7. Expressions for $K(k')/K(k)$ are given in (2.4). Approximate expressions for attenuation for these lines are reported in Gupta et al. [14]. The variation of total loss $(\alpha_c + \alpha_d)$ for coplanar lines on alumina substrate of thickness 0.63 mm as a function of aspect ratio is plotted in Fig. 2.9. It is observed that loss decreases with decreasing impedance or increasing strip width.

2.1.3 Comparison of Various MIC Transmission Media

For hybrid MIC applications, microstrips, slot lines, CPWs, and CPSs have been used, whereas for monolithic MICs, microstrip has been used extensively, although there is an interest in CPW. Several other parameters of the four types of lines are compared qualitatively in Table 2.8. It can be generally seen that CPWs and CPSs combine some advantageous features of microstrip lines and slot lines. Their power-handling capabilities, radiation losses, Q factors, and dispersion behavior lie in between the corresponding values for microstrip and slot lines. Perhaps the best feature of the two coplanar lines is the ease of mounting components in series and shunt configurations, whereas microstrip lines are convenient only for series mounting and slot lines can accommodate only shunt-mounted components.

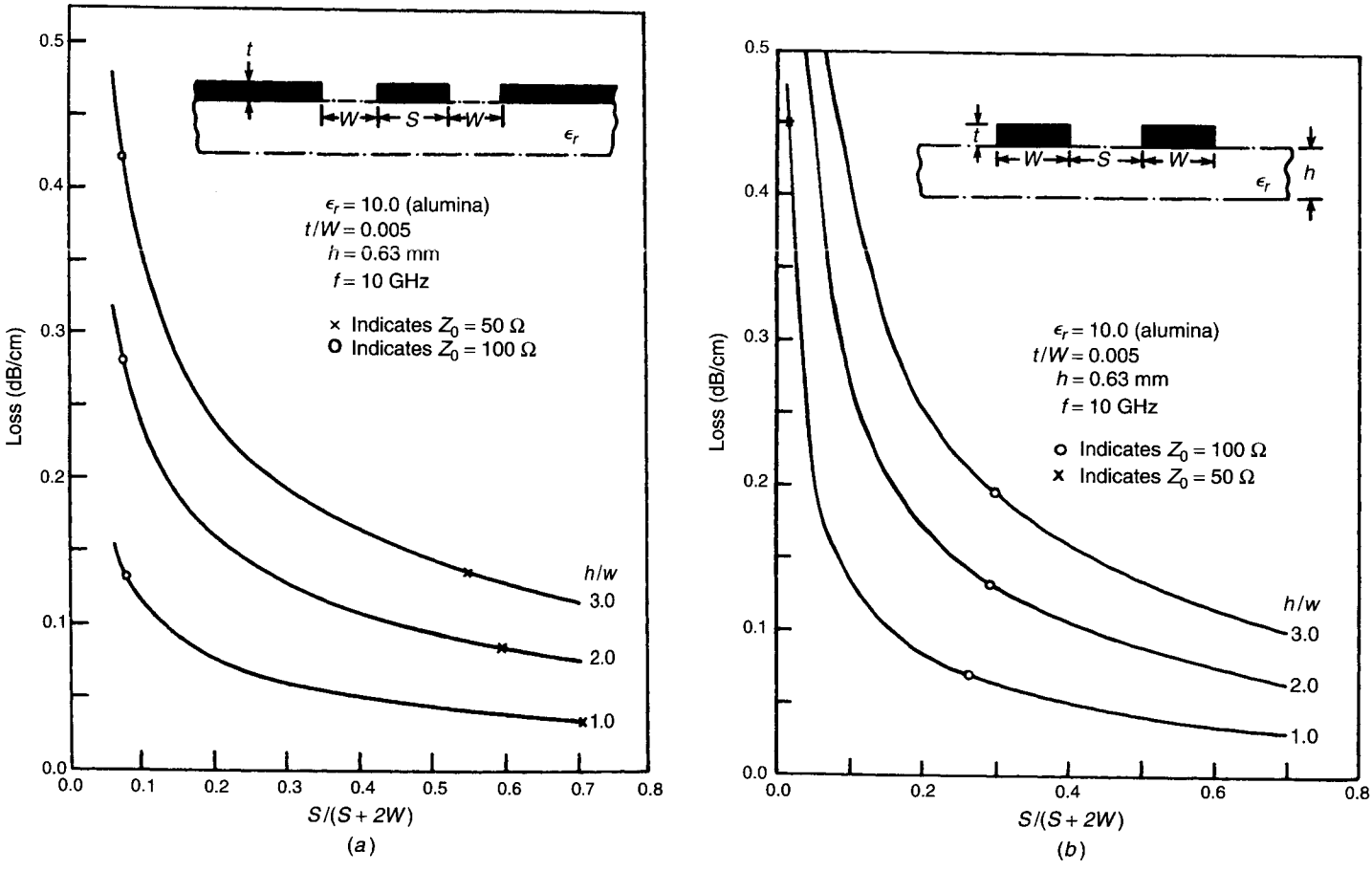


Figure 2.9 (*a*) Attenuation constant for CPW on alumina substrate; (*b*) attenuation constant for CPS on alumina substrate.

Table 2.8 Qualitative Comparison of Various MIC Lines

Characteristic	Microstrips	Slot Lines	Coplanar Waveguides	Coplanar Strips
Impedance range	20–110	55–300	25–155	45–280
Effective dielectric constant ($\epsilon_r = 10$ and $h = 0.025$ in.)	~6.5	~4.5	~5	~5
Power-handling capability	High	Low	Medium	Medium
Radiation loss	Low	High	Medium	Medium
Unloaded Q	High	Low	Medium	Low (lower impedance) High (higher impedance)
Dispersion	Small	Large	Medium	Medium
Mounting of components				
In shunt configuration	Difficult	Easy	Easy	Easy
In series configuration	Easy	Difficult	Easy	Easy
Technological difficulties	Ceramic holes Edge plating	Double side etching	—	—
Elliptically polarized magnetic field configuration	Not available	Available	Available	Available
Enclosure dimensions	Small	Large	Large	Large

2.2 COUPLED LINES

A "coupled-line" configuration consists of two transmission lines placed parallel to each other and in close proximity. In such a case, there is a continuous coupling between the electromagnetic fields of two lines. Coupled lines are utilized extensively as basic elements for directional couplers, filters, phase shifters, baluns, matching networks, and a variety of other useful circuits.

Data such as characteristic impedance, phase velocity, and insertion loss are needed for the design of circuit components. Because of the coupling of EM fields, a pair of coupled lines can support two different modes of propagation; the even and odd. These modes have different characteristic impedances. The velocity of propagation of these two modes is equal when the lines are imbedded in a homogeneous dielectric medium. This is a desirable property for the design of circuits such as high-directivity couplers. However, for transmission lines such as coupled microstrip lines the dielectric medium is not homogeneous. A part of the field extends into the air above the substrate. This fraction of the total field is different for the two modes of coupled lines. Consequently, the effective dielectric constants (and the phase velocities) are not equal for the two modes. This nonsynchronous feature deteriorates the perfor-

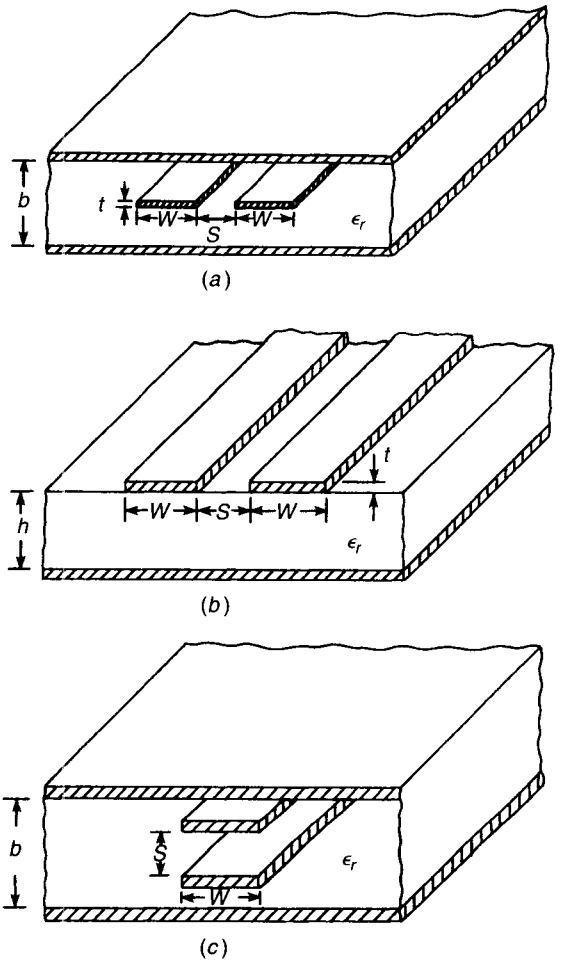


Figure 2.10 Coupled transmission lines: (*a*) strip; (*b*) microstrip; (*c*) broadside strip.

mance of circuits using these types of coupled lines. For the sake of simplicity only symmetrical configuration (both identical conductors) is considered here. Figure 2.10 shows coupled strip lines, microstrip lines, and broadside-coupled strip lines, and their characteristics are given in Table 2.9.

In coupled microstrip lines, the effect of thickness of conductors on capacitances can be evaluated by using the concept of effective width W_e [29]. The expressions given in Table 2.9 are valid for $S \geq 2t$, where $\Delta W = W_e - W$. Variation of even- and odd-mode total loss as a function of gap between the two conductors is shown in Fig. 2.11 [30].

Broadside-coupled strip lines have been widely used for realizing tight couplings (e.g., 3-dB hybrids) because for greater than −8-dB coupling, the spacing

Table 2.9 Characteristics for Coupled Structures

Characteristic Impedances (Ω)	Attenuation Constants (dB/unit length)	Remarks
	Strip Lines	
$Z_{0e} = \frac{30\pi}{\sqrt{\epsilon_r}} \frac{K(k_e')}{K(k_e)}$		$t = 0$, virtually exact [28]
$Z_{0o} = \frac{30\pi}{\sqrt{\epsilon_r}} \frac{K(k_o')}{K(k_o)}$		
$k_e = \tanh\left(\frac{\pi}{2}\frac{W}{b}\right) \tanh\left(\frac{\pi}{2}\frac{W+S}{b}\right)$		
$k_o = \tanh\left(\frac{\pi}{2}\frac{W}{b}\right) \coth\left(\frac{\pi}{2}\frac{W+S}{b}\right)$		
$Z_{0e} = \frac{30\pi(b-t)}{\sqrt{\epsilon_r}\left(W + \frac{bC_f}{2\pi}A_e\right)}$	$\alpha_c^e = \frac{0.0231 R_s \sqrt{\epsilon_r}}{30\pi(b-t)} \times \left\{60\pi + Z_{0e}\sqrt{\epsilon_r}\left[1 - \frac{A_e}{\pi}\left(\ln\frac{2b-t}{b-t} + \frac{1}{2}\ln\frac{t(2b-t)}{(b-t)^2}\right) + C_f \frac{(1+S/b)}{4\ln 2}\frac{\operatorname{sech}^2\theta}{1+\tanh\theta}\right]\right\}$	$\frac{t}{b} < 0.1$ $\frac{W}{b} \geq 0.35$ [28]
$Z_{0o} = \frac{30\pi(b-t)}{\sqrt{\epsilon_r}\left(W + \frac{bC_f}{2\pi}A_o\right)}$	$\alpha_c^o = \frac{0.0231 R_s \sqrt{\epsilon_r}}{30\pi(b-t)} \times \left\{60\pi + Z_{0o}\sqrt{\epsilon_r}\left[1 - \frac{A_o}{\pi}\left(\ln\frac{2b-t}{b-t} + \frac{1}{2}\frac{t(2b-t)}{(b-t)^2}\right) - C_f \frac{(1+S/b)}{4\ln 2}\frac{\operatorname{cosech}^2\theta}{1+\coth\theta}\right]\right\}$	
$A_e = 1 + \frac{\ln(1+\tanh\theta)}{\ln 2}$		
$A_o = 1 + \frac{\ln(1+\coth\theta)}{\ln 2}$		
$\theta = \frac{\pi S}{2b}$		
$C_f = 2\ln\left(\frac{2b-t}{b-t}\right) - \frac{t}{b}\ln\left[\frac{t(2b-t)}{(b-t)^2}\right]$	$\alpha_d^e = \alpha_d^o = 27.3\sqrt{\epsilon_r}\frac{\tan\delta}{\lambda_0}$	

Microstrip Lines

$$Z_{0i} = [c\sqrt{C_i C_i^a}]^{-1}, \qquad i = e \text{ or } o$$

$$C_e = C_p + C_f + C_f'$$

$$C_o = C_p + C_f + C_{ga} + C_{gd}$$

$$C_p = \epsilon_0 \epsilon_r \frac{W}{h}$$

$$2C_f = \sqrt{\epsilon_e}/(cZ_0) - C_p, \qquad c = 3 \times 10^8 \text{ m/s}$$

$$C_f' = \frac{C_f \sqrt{\epsilon_r/\epsilon_e}}{1 + \exp[-0.1 \exp(2.33 - 2.53W/h)](h/S)\tanh(10S/h)}$$

$$C_{ga} = \epsilon_0 \frac{K(k')}{K(k)}; \qquad k = \frac{S/h}{S/h + 2W/h}$$

$$C_{gd} = \frac{\epsilon_0 \epsilon_r}{\pi} \ln\left[\coth\left(\frac{\pi S}{4h}\right)\right] + 0.65C_f\left[\frac{0.02}{S/h}\sqrt{\epsilon_r} + 1 - \epsilon_r^{-2}\right]$$

$$\epsilon_e^i = \frac{C_i}{C_i^a}$$

$$\alpha_c^i = \frac{8.686R_s}{240\pi Z_{0i}} \frac{2}{h} \frac{1}{c(C_i^{at})^2} \times \left[\frac{\partial C_f^{at}}{\partial(W/h)}\left(1 + 2\frac{W}{2h}\right) - \frac{\partial C_i^{at}}{\partial(S/h)}\left(1 - 2\frac{S}{2h}\right) + \frac{\partial C_i^{at}}{\partial(t/h)}\left(1 + 2\frac{t}{2h}\right)\right]$$

$$\alpha_d^i = 27.3 \frac{\epsilon_r}{\sqrt{\epsilon_e^i}} \frac{\epsilon_e^i - 1}{\epsilon_r - 1} \frac{\tan\delta}{\lambda_0}$$

$$\frac{W_t^e}{h} = \frac{W}{h} + \frac{\Delta W}{h}\left(1 - 0.5 \exp\frac{-0.69\ \Delta W}{\Delta t}\right)$$

$$\frac{W_t^o}{h} = \frac{W_t^e}{h} + \frac{\Delta t}{h}, \qquad \frac{\Delta t}{h} = \frac{1}{\epsilon_r}\frac{t/h}{S/h}$$

Accuracy better than 3% for

$$0.2 \le \frac{W}{h} \le 2$$

$$0.05 \le \frac{S}{h} \le 2$$

Broadside-Coupled Strip Lines

$$Z_{0o} = \frac{60\pi/\sqrt{\epsilon_r}}{\frac{W}{b-S} + C_{f0} + 2\{(1 + t/S)\ln(1 + t/S) - (t/S)\ln(t/S)\}/\pi}$$

$$Z_{0e} = \frac{60\pi/\sqrt{\epsilon_r}}{\frac{W}{b-S} + 0.443 + \left(\ln\frac{b+2t}{b-S} + \frac{S+2t}{b-S}\ln[(b+2t)/(S+2t)]\right)/\pi}$$

$$C_{f0} = \frac{b}{S\pi}\left(\ln\frac{1}{1 - S/b} + \frac{S/b}{1 - S/b}\ln\frac{b}{S}\right)$$

$$\alpha_c^i = \frac{0.0231R_s\sqrt{\epsilon_r}}{Z_{0i}}\left(\frac{\partial Z_{0i}}{\partial b} + \frac{\partial Z_{0i}}{\partial S} - \frac{\partial Z_{0i}}{\partial W} - \frac{\partial Z_{0i}}{\partial t}\right)$$

$$\alpha_d = 27.3\sqrt{\epsilon_r}\frac{\tan\delta}{\lambda_0}$$

[31–33]

$$\frac{W/b'}{1 - S/b} \ge 0.35$$

$$\frac{W}{S} \ge 0.35$$

Note: Superscripts a and t denote air as dielectric and strip thickness effect, respectively.

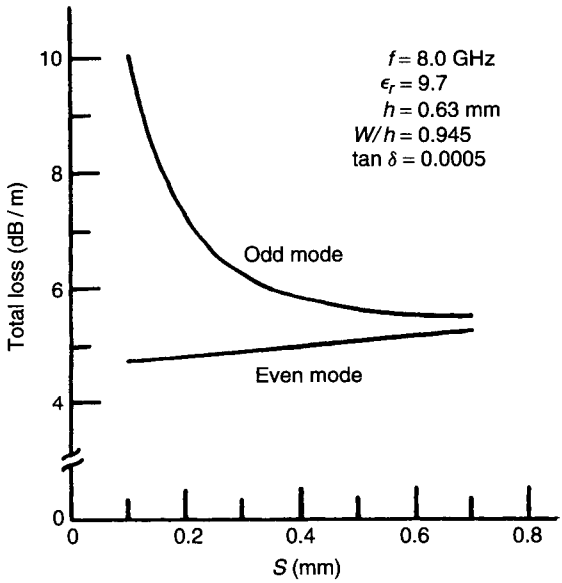


Figure 2.11 Even- and odd-mode losses in coupled microstrip lines.

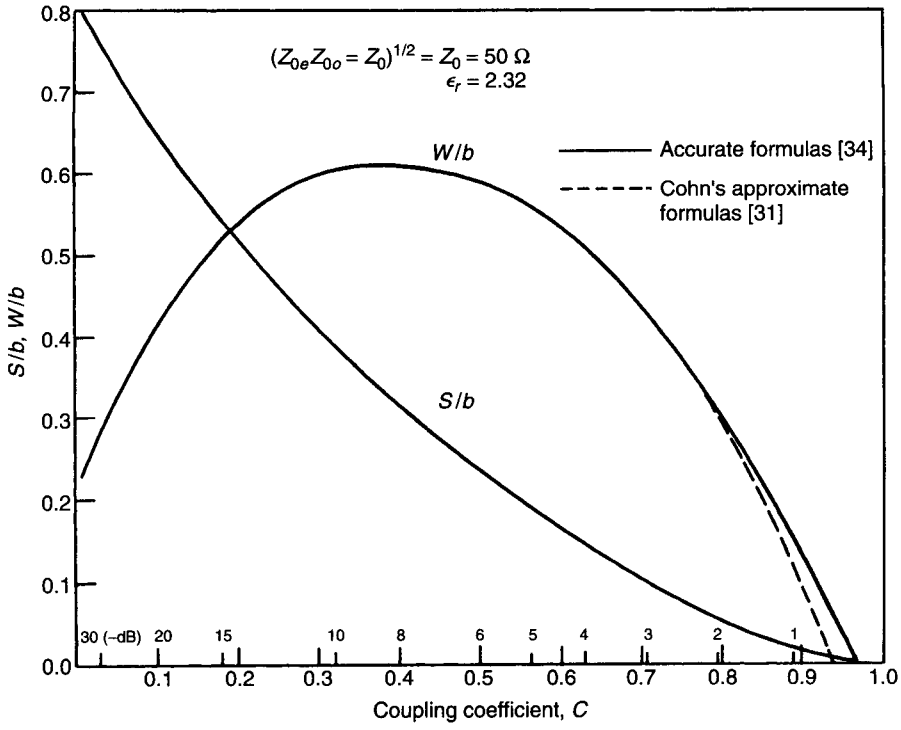


Figure 2.12 Variation of S/b and W/b with coupling coefficient for $\epsilon_r = 2.32$.

between the strips in the case of parallel coupled transmission lines (e.g., strip lines, microstrip lines) becomes prohibitively small. Expressions for characteristic impedances and attenuation constants are given in Table 2.9 [31–33]. Figure 2.12 shows the variation of W/b and S/b as a function of coupling coefficient for $\epsilon_r = 2.32$ and $t = 0$. More accurate results [34] are also included for comparison. The total loss $\alpha = \alpha_c + \alpha_d$ in broadside-coupled strip lines is plotted in Fig. 2.13 as a function of W/b for various values of S/b and $f = 3.0$ GHz. It may be observed that the odd-mode attenuation constant is always higher than the even-mode values.

Broadside-coupled strip lines in inhomogeneous media [35, 36] have also been used extensively for designing directional couplers and filters. However, to design high-directivity directional couplers, phase velocities for the even and

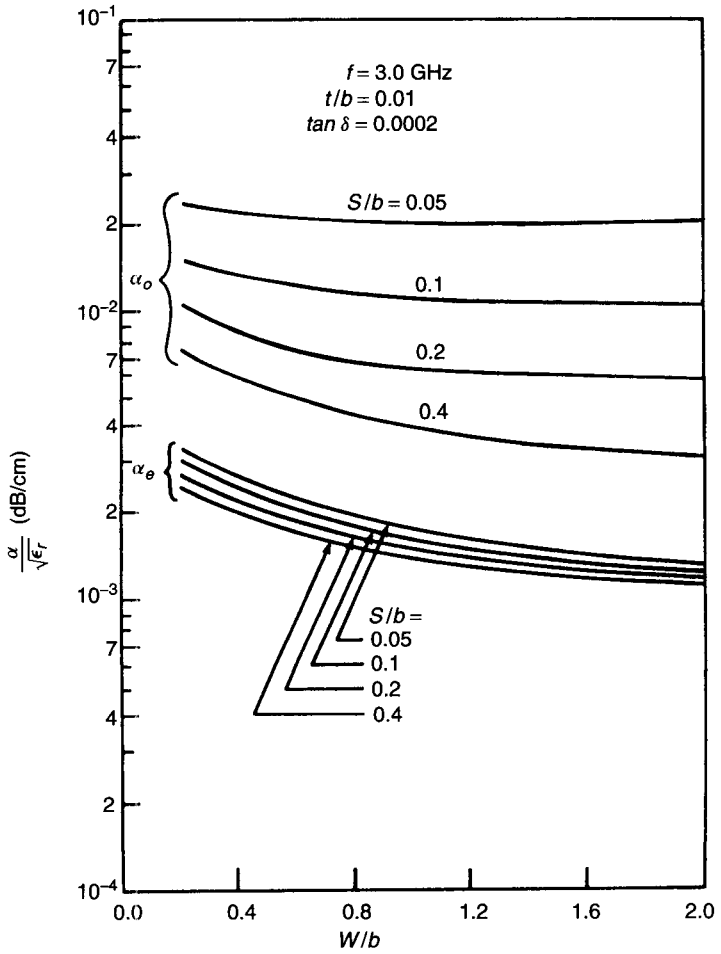


Figure 2.13 Attenuation constants for broadside-coupled strip lines.

odd modes must be the same. A detailed description of coupled lines and their components can be found in Mongia et al. [37].

2.3 DISCONTINUITIES

In microwave circuits, discontinuities between distributed elements, between lumped elements, and between distributed and lumped circuit elements always exist. Typical discontinuities as shown in Fig. 2.14 are (1) open circuits and

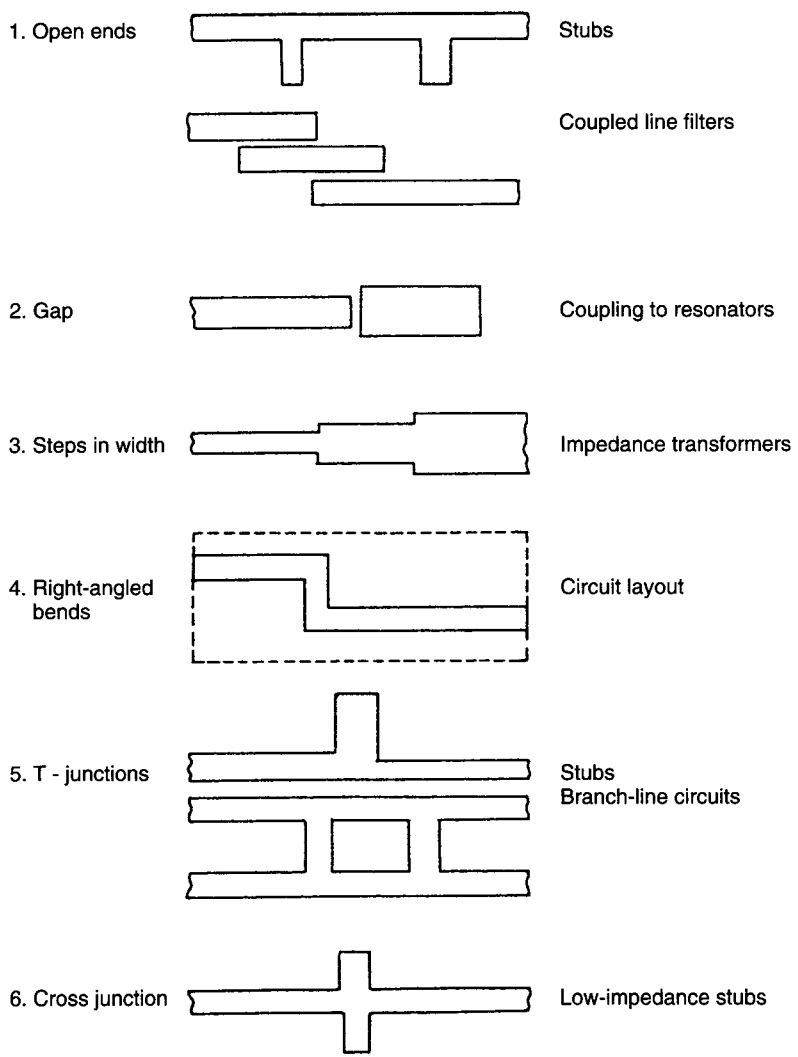


Figure 2.14 Typical planar strip transmission-line discontinuities.

short circuits, (2) gap between conductors, (3) step change in dimensions (introduced for a change in impedance level), (4) bends (right-angled and others), and (5) T and (6) cross-junctions. The reactances associated with these discontinuities may be called parasitic, as they are not introduced intentionally. Some of the effects of discontinuities on circuit performances are

(a) frequency shift in narrow-band circuits,
(b) degradation in input and output voltage standing-wave ratios (VSWRs),
(c) higher ripple in gain flatness of broadband ICs,
(d) interfacing problem in multifunction circuits,
(e) lower circuit yield due to degradation in circuit performance, and
(f) surface wave and radiation couplings that may cause oscillations in high-gain amplifiers.

The effect of discontinuities becomes more critical at higher frequencies. The discontinuities should be either taken into account or compensated for at the final stage of design. In most cases discontinuities are basically undesirable circuit reactances, and in a good circuit design, efforts are made to reduce or compensate for these reactances. A complete understanding of the design of microwave circuits requires characterization of the discontinuities present in these circuits. Since it is impossible to do tuning on GaAs MMICs, an accurate and comprehensive modeling of each device and circuit element is required to save expensive and time-consuming iteration of mask and wafer fabrication and evaluation. As the yield of MMICs depends on the size (the smaller the chip, the higher the yield) and acceptable circuits' electrical performance, discontinuities play an important part in the development of MMICs.

Since the discontinuity dimensions are usually much smaller than a wavelength, the discontinuities are represented by lumped-element equivalent circuits. In many cases, when the longitudinal dimension of a discontinuity is very short, the equivalent circuit consists of a single shunt or series-connected reactance located at the point of the discontinuity. However, when the discontinuity has a larger longitudinal extent, the equivalent circuit is usually a π or a T network.

Various general methods are used to determine discontinuity reactances. The most commonly used techniques are the variational method, mode matching, and spectral domain. A more complete characterization involves determination of the frequency-dependent scattering matrix coefficients associated with the discontinuity. Such analyses are available for several types of discontinuities [4, 10, 14, 38–41].

Compensation of discontinuity reactances in planar transmission lines can be obtained by removing appropriate portions from the discontinuity configurations, such as chamfering the bend. Compensations of step in width and T junctions have been treated in the literature [10, 42–45]. The step-in-width discontinuity can be easily compensated by gradually narrowing the wider strip,

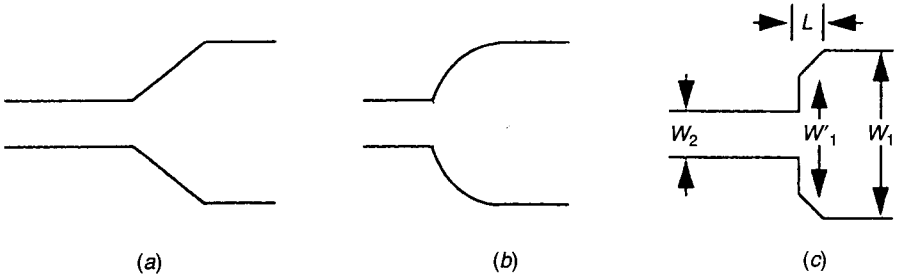


Figure 2.15 Three different kinds of compensated step-in-width discontinuity configurations with (*a*) linear taper, (*b*) curved taper, and (*c*) partial linear taper.

as shown in Fig. 2.15. For a step-width ratio ranging from 3 to 13, the step discontinuity reactance is negligible when $L = \frac{1}{8} W_1$ and $W_1' = 0.33 W_1$ and the GaAs substrate is 75–175 μm thick. Several types of right-angled bend chamfering, as shown in Fig. 2.16, have been studied, and it has been found that the configuration of Fig. 2.16*c* provides the best compensation on alumina and GaAs substrates. The T-junction discontinuity compensation is much more difficult than the step-in-width and right-angled bend discontinuity compensation techniques. Figure 2.17*a* shows T-junction compensation configurations using rectangular and triangular notches and their approximate dimensions for $h/\lambda \ll 1$. However, accurate dimensions of the compensated configuration depend upon the line widths, dielectric constant, and substrate thickness. Figure 2.17*b* illustrates T-junction discontinuity minimization configurations in which the line widths are tapered to minimize the junction effect. In this case the taper length is about twice the line width or GaAs substrate thickness, whichever is larger.

2.4 LUMPED ELEMENTS

The size of a lumped circuit element, by definition, is very much smaller than the operating wavelength and therefore exhibits negligible phase shift. Lumped elements for use at microwave frequencies are also designed on the basis of this consideration.

Lumped-element circuits that have lower Q than distributed circuits have the advantage of smaller size, lower cost, and wide-band characteristics. These are especially suitable for monolithic MICs and for broadband hybrid MICs where real estate requirements are of prime importance. Impedance transformations of the order of 20:1 can be easily accomplished using the lumped-element approach. Therefore, high-power devices that have very low impedance values can be easily tuned with large impedance transformers using lumped elements.

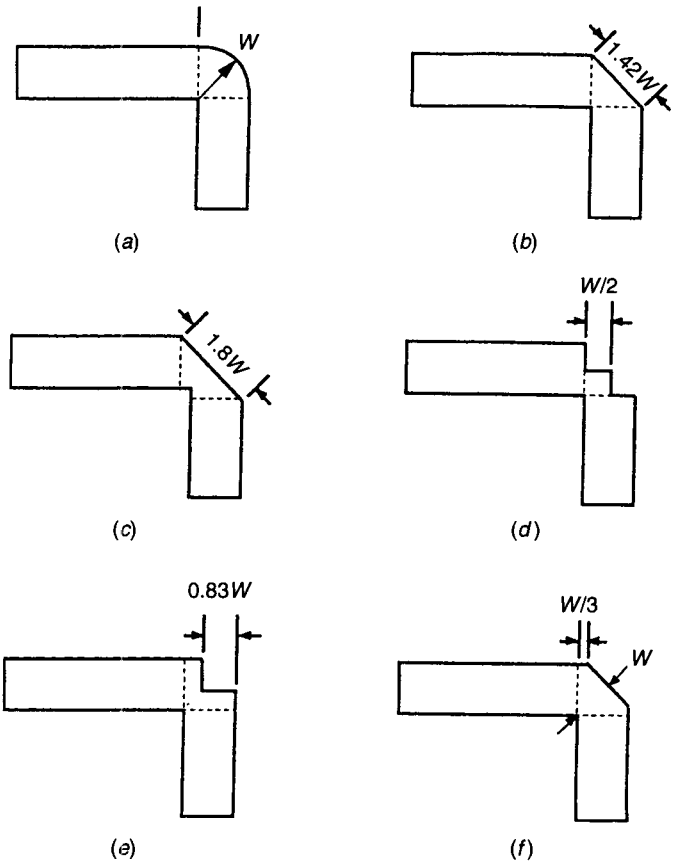


Figure 2.16 Six different configurations for compensated right-angled bends.

Consequently, lumped elements find applications in high-power oscillators, power amplifiers, broadband circuits, and low-cost wireless applications.

With the advent of new photolithographic techniques, the fabrication of lumped elements that was limited to X-band frequencies can now be extended to about 60 GHz. The three basic building blocks for circuit design—inductors, capacitors, and resistors—are available in lumped form. Computer-aided design of circuits using lumped elements requires a complete and accurate characterization of lumped elements at microwave frequencies. This necessitates the development of comprehensive mathematical models that take into account the presence of ground planes, proximity effect, fringing fields, parasitics, and so on. In this section we discuss the available design and characterization details of lumped elements [46–67].

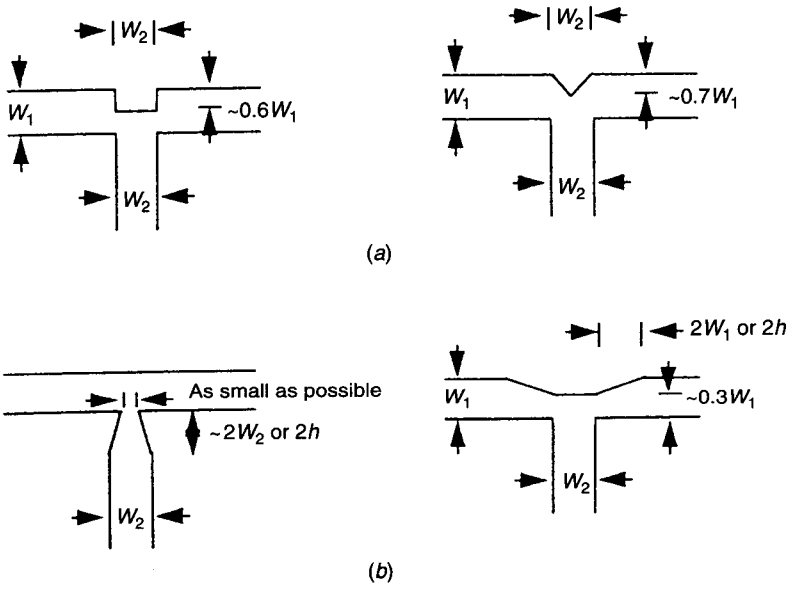


Figure 2.17 (*a*) T-junction discontinuity compensation configurations; (*b*) minimized T-junction discontinuity effect configuration.

2.4.1 Design of Lumped Elements

Design of lumped elements at RF, microwave, and the lower end of millimeter-wave frequencies may be arrived at by considering them as very short sections of (smaller than $\frac{1}{10}\lambda$) TEM lines. The driving point or input impedance of the line of length l given in Table 2.1 (when $\gamma l \ll 1$) may be written as

$$Z_{\text{in}} = Z_0 \frac{Z_L + Z_0 \gamma l}{Z_0 + Z_L \gamma l} \tag{2.19}$$

where Z_L is the terminating or load impedance. The characteristic impedance Z_0 and the propagation constant γ for the TEM line shown in Fig. 2.1 are given by

$$Z_0 = \sqrt{\frac{\bar{R} + j\omega\bar{L}}{\bar{G} + j\omega\bar{C}}} \tag{2.20a}$$

$$\gamma = \sqrt{(\bar{R} + j\omega\bar{L})(\bar{G} + j\omega\bar{C})} \tag{2.20b}$$

Equations (2.19) and (2.20) are used to derive equivalent circuits for L, C, and R elements.

2.4.2 Design of Inductors

The input impedance of a very small length of a short-circuited ($Z_L = 0$) line is expressed as

$$Z_{\text{in}} = Z_0 \gamma l = (\bar{R} + j\omega\bar{L})l = R + j\omega L \tag{2.21}$$

where R is the resistance and L is the inductance of the line of length l. When the thickness and width of conductors are much greater than the skin depth, for first-order approximation, conductors may be assumed lossless ($R = 0$).

However, in order to realize an inductor or a resistor, it is not necessary to have both conductors of a transmission line. A lumped inductor may be realized using a metallic strip or a wire either in the form of straight section or a circular or square spiral.

The important characteristics of an inductor are the inductance value and its parasitics, which determine its Q factor and resonant frequency.

Inductance. Straight sections of ribbons and wires are used for low inductance values typically up to 2–3 nH. Spiral inductors (circular or rectangular) have higher Q and can provide higher inductance values. These inductors are commonly used for high-density circuits. The presence of a ground plane also affects the inductance value, which decreases as the ground plane is brought nearer [56]. This decrease can be taken into account by means of a correction factor K_g. With this correction, the effective inductance L may be written as

$$L = K_g L_0 \tag{2.22}$$

where L_0 is the free-space inductance value. A closed-form expression for K_g of a ribbon is given by [4, 57]

$$K_g = 0.57 - 0.145 \ln \frac{W}{h} \qquad \frac{W}{h} > 0.05 \tag{2.23}$$

where W is the width and h is the substrate thickness. To first-order approximation, the above expression can also be used with other types of inductors. Typical inductance values for MMICs fall in the range from 0.2 to 10 nH.

Resistance and Unloaded Q. In order to find Q, it is necessary to determine R, the high-frequency resistance of the conductors. The expression for R is given by

$$R = \frac{KR_s l}{\text{perimeter}} \tag{2.24}$$

where R_s is the sheet resistance in ohms per square, l is the length of the con-

ductor, and K is a correction factor that takes into account the crowding of the current at the corners of the inductor conductors. For various structures, expressions for K are given as

$$K = \begin{cases} 1 & \text{(wire)} \quad (2.25a) \\ 1.4 + 0.217 \ln\left(\dfrac{W}{5t}\right) \qquad 5 < \dfrac{W}{t} < 100 & \text{(ribbon)} \quad (2.25b) \\ 1 + 0.333\left(1 + \dfrac{S}{W}\right)^{-1.7} & \text{(spiral)} \quad (2.25c) \end{cases}$$

where t is the thickness of the conductors. The unloaded Q of an inductor then may be calculated from

$$Q = \frac{\omega L}{R} \tag{2.26}$$

Table 2.10 gives expressions for inductances and resistances of various types of inductors. In the case of spirals n is the number of turns and S is the spacing between the conductors. More accurate expressions for R are given by Pettenpaul et al. [59].

It is recommended that for operating frequencies above one-third of the self-resonant frequency of the inductors, the parasitic capacitances associated with the inductors be included in the design. The values of shunt capacitors for strip and loop inductors can be easily obtained as described in the next section. However, there is no easy way to calculate shunt capacitances for a spiral. Measured inductance, parasitic capacitance, and Q values for a few typical spiral inductors are given in Table 2.11 [55]. The substrate was 0.63-mm-thick alumina ($\epsilon_r = 9.6$).

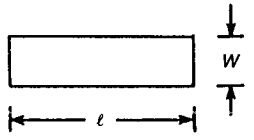
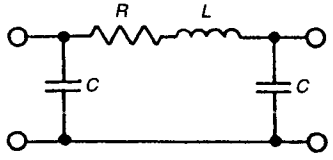
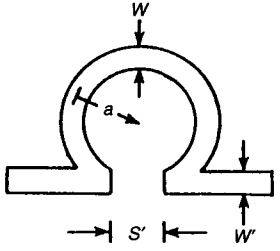
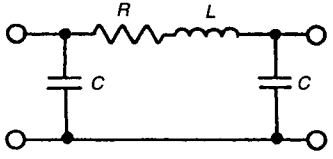
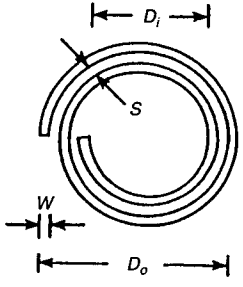
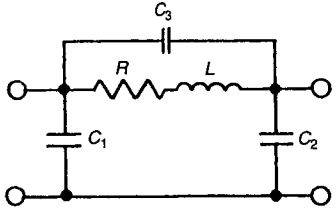
In practice, Q factors on the order of 50 are observed at the X band for MIC inductors, with higher values at higher frequencies. The maximum value is about 100. However, for MMIC inductors, the Q factor is lower than the Q factor for MIC inductors because of the highly unfavorable ratio of metal surface area to dielectric volume.

Accurate inductor models are developed using measured two-port scattering parameters. The inductors are represented by their inductance value, the effective quality factor Q_{eff}, and its resonant frequency f_{res}. The two-port lumped-element equivalent circuit model used to characterize inductors is shown in Fig. 2.18. The series resistance R used to model the dissipative loss is given by

$$R = R_{\text{dc}} + R_{\text{ac}}\sqrt{f} + R_d f \tag{2.27}$$

where R_{dc} represents direct-current (DC) resistance of the trace, and R_{ac} and R_d model resistances due to skin effect, eddy current excitation and dielectric

Table 2.10 Expressions for Lumped Inductors

Inductors Strip		Equivalent Circuit	Expressions
Strip			$L\ (\text{nH}) = 2 \times 10^{-4} l\left[\ln\left(\frac{l}{W+t}\right) + 1.193 + 0.2235\frac{W+t}{l}\right]K_g$ $R\ (\Omega) = \frac{KR_s l}{2(W+t)}$
Loop			$L\ (\text{nH}) = 1.257 \times 10^{-3} a\left[\ln\left(\frac{a}{W+t}\right) + 0.078\right]K_g$ $R\ (\Omega) = \frac{KR_s}{W+t}\pi a$
Spiral			$L\ (\text{nH}) = 0.03937\frac{a^2 n^2}{8a+11c}K_g$ $a = \frac{D_o + D_i}{4}, \quad c = \frac{D_o - D_i}{2}$ $R\ (\Omega) = \frac{k\pi a n R_s}{W}$ $C_3\ (\text{pF}) = 3.5 \times 10^{-5} D_o + 0.06$

Note: Dimensions in micrometers.

Table 2.11 Typical Spiral Inductor Parameters

W (mm)	S (mm)	n	D_o (mm)	L (theoretical) (nH)	L (measured) (nH)	C_3 (pF)	C_1 (pF)	C_2 (pF)	Q (at 4 GHz)
0.10	0.10	1.5	1.30	3.7	4.0	0.13	0.10	0.07	87
0.05	0.05	2.5	1.50	14.1	14.3	0.11	0.09	0.05	84
0.10	0.10	2.5	1.80	9.2	9.0	0.14	0.18	0.08	81
0.10	0.10	3.5	2.00	18.7	19.0	0.15	0.17	0.09	91

loss in the substrate. The frequency f is expressed in GHz. In the model $L_t(L + L_1 + L_2)$, R, and C's represent the total inductance, series resistance, and parasitic capacitances of the inductor, respectively. The Q_{eff} factor values are obtained by converting two-port S parameters data into one-port S parameters by placing a perfect short at the output port. The Q_{eff} factor was then calculated from $\text{Im}(Z_{\text{in}})/\text{Re}(Z_{\text{in}})$. The self-resonant frequency (f_{res}) of an inductor is calculated when $\text{Im}[Z_{\text{in}}] = 0$, that is, the inductive reactance and the parasitic capacitive reactance become equal. At this point $\text{Re}[Z_{\text{in}}]$ is maximum and the angle of Z_{in} changes sign. The inductor's first resonant frequency is of the parallel resonance type. Beyond the resonant frequency, the inductor becomes capacitive.

Table 2.12 summarizes typical model parameter values for various circular spiral inductors fabricated on GaAs substrate. The inductor pattern having a thickness of 4.5 μm was placed on a 3-μm polyimide layer backed by a 75-μm-thick GaAs substrate. Figure 2.19 shows the plots of L_t, Q_{eff}, and f_{res} versus the number of turns for inductors having $W = S = 8$ μm and inner diameter $D_i = 50$ μm. As expected, the inductance increases approximately as n^2, where n is the number of turns. For such inductors the value of the total inductance (in nanohenrys), Q_{eff} factor and resonance frequency (in gigahertz) may be approximately calculated using the following empirical equations:

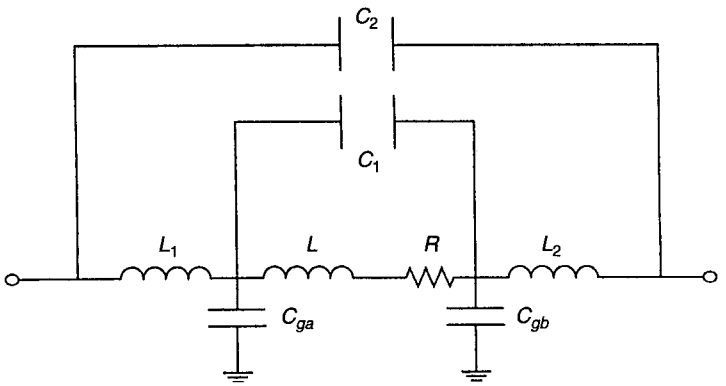


Figure 2.18 Lumped-element equivalent circuit model of the inductor.

Table 2.12 Model Parameter Values for Type A Inductors

Number of Turns	Dimensions (μm) W	S	D_i	R_{dc} (Ω)	R_{ac} (Ω)	R_d (Ω)	L_1 (nH)	L_2 (nH)	L (nH)	C_{ga} (pF)	C_{gb} (pF)	C_1 (pF)	C_2 (pF)	f_{res} (GHz)	Peak Q_{eff}
1.5	20	8	50	0.1	0.167	0.001	0.0001	0.0001	0.378	0.019	0.020	0.0001	0.00001	>40	49
1.5	20	8	108	0.18	0.15	0.009	0.0001	0.0001	0.563	0.028	0.027	0.0001	0.00001	37.6	52
1.5	20	8	158	0.23	0.16	0.018	0.0001	0.0001	0.760	0.038	0.036	0.0001	0.00001	27.2	48
1.5	20	8	210	0.24	0.24	0.02	0.092	0.077	0.85	0.048	0.046	0.0001	0.00001	22.3	40.5
1.5	12	8	50	0.10	0.19	0.0001	0.0001	0.0001	0.342	0.014	0.011	0.0001	0.00001	>40	41.5
1.5	8	8	50	0.20	0.155	0.01	0.0001	0.0001	0.32	0.010	0.008	0.00002	0.00001	>40	35
2.5	16	10	50	0.20	0.3	0.03	0.18	0.30	0.42	0.020	0.033	0.002	0.0096	28.7	37
2.5	16	10	108	0.30	0.25	0.06	0.16	0.10	1.10	0.039	0.037	0.0016	0.0043	21.00	39
2.5	16	10	158	0.35	0.27	0.1	0.21	0.23	1.45	0.048	0.054	0.0045	0.0045	16.2	38
2.5	16	10	210	0.38	0.3	0.15	0.24	0.26	1.91	0.059	0.065	0.0023	0.0052	13.15	35.5
2.5	12	8	50	0.25	0.33	0.021	0.028	0.029	0.751	0.022	0.021	0.00002	0.00001	34.2	35
2.5	8	8	50	0.25	0.40	0.02	0.033	0.032	0.696	0.019	0.015	0.00003	0.00005	38.1	30.0
3.5	12	14	50	0.38	0.55	0.06	0.32	0.32	1.18	0.036	0.039	0.0048	0.0075	18.7	30.0
3.5	12	14	108	0.45	0.52	0.13	0.26	0.29	2.08	0.047	0.053	0.0001	0.006	14.0	31.5
3.5	12	14	158	0.6	0.7	0.17	0.30	0.33	3.00	0.061	0.071	0.0001	0.0025	10.6	29.5
3.5	12	14	210	0.7	0.80	0.25	0.37	0.48	3.74	0.073	0.082	0.0003	0.0032	8.75	27.5
3.5	12	8	50	0.3	0.55	0.06	0.033	0.44	1.12	0.022	0.037	0.016	0.0001	20.5	30.0
3.5	8	8	50	0.3	0.60	0.07	0.036	0.034	1.407	0.023	0.023	0.001	0.0018	25.0	27.5
4.5	12	14	50	0.46	0.66	0.12	0.33	0.34	2.36	0.053	0.066	0.002	0.008	11.9	30.0
4.5	8	8	50	0.4	0.8	0.16	0.044	0.036	2.421	0.029	0.033	0.0023	0.002	17.5	25.5
5.5	8	8	50	0.5	1.0	0.21	0.047	0.038	3.64	0.044	0.054	0.0029	0.002	11.8	25.0

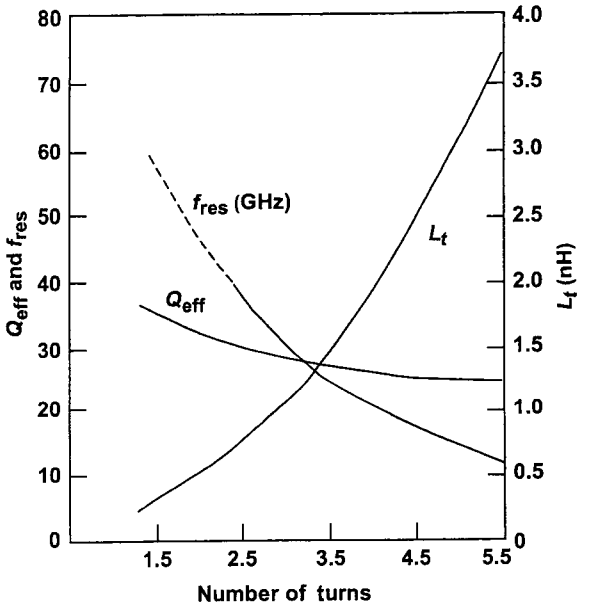


Figure 2.19 Variations of L_t, Q_{eff}, and f_{res} as a function of number of turns for compact inductors.

$$L_t = 0.04 + 0.12n^2 \tag{2.28a}$$

$$Q_{eff} = \frac{38}{n^{0.25}} \tag{2.28b}$$

$$f_{res} = \frac{36.52}{0.9432 + 0.01n^{3.15}} \quad \text{for } f_{res} < 30 \text{ GHz} \tag{2.28c}$$

Several techniques to enhance the Q factor of planar inductors have also been described in the literature, including the variable line width [68, 69], differential excitation technique [70], multilayer dielectric and metallization approach [67, 71, 72], and micromachining technique [73].

Current-Handling Capacity. The current-handling capability of a conductor is limited by the onset of electromigration. The conductor thickness and line width determine the current carrying capacity of the inductor. For a given conductor thickness, the inductor line width determines the maximum current-handling capacity. The maximum current density should not exceed 6×10^5 A/cm^2 for a gold conductor on a flat surface. Since in our case we have conductors on GaAs, we have derated this rule to 3×10^5 A/cm^2, which is the commonly accepted value in the literature. For example, for a 4.5-μm-thick conductor, the calculated maximum current-handling capacity per unit line width is 13.5 mA/μm.

Some salient considerations useful in the design of these inductors are summarized as follows:

1. In the spiral, $W \gg S$, that is, the separation S between the turns should be as small as possible. However, in monolithic circuits processing limits require $W > 5$ μm and $S > 5$ μm, and in hybrid MICs $W > 10$ μm and $S > 10$ μm.
2. There should be some space at the center of the spiral to allow the flux lines to pass through, which increases the stored energy. It has been found that $D_o/D_i = 5$ yields an optimum value of Q but not the maximum inductance value.
3. It is observed that for the same value of D_o, the Q of a circular spiral is higher than that of a square spiral by about 10%.
4. Multiturn inductors have a high Q due to higher inductance per unit area; also, they have lower self-resonance frequencies due to interturn capacitance.
5. The maximum diameter of the inductor should be less than $\frac{1}{30}\lambda$ in order to avoid distributed effects.
6. Spiral inductors require air bridges or dielectric crossovers.

Wire Inductor. In hybrid MICs, bond wire connections are used to connect active and passive circuit components, and in MMICs bond wire connections are used to connect the MMIC chip to the real world. The free-space inductance L (in nanohenrys) of a wire of diameter d and length l (in micrometers) is given by [4, 46, 61]

$$L = 2 \times 10^{-4} l \left[\ln \left\{ \frac{2l}{d} + \sqrt{1 + \left(\frac{2l}{d}\right)^2} \right\} + \frac{d}{2l} - \sqrt{1 + \left(\frac{d}{2l}\right)^2} + C \right] \tag{2.29}$$

where the frequency-dependent correction factor C is a function of bond wire diameter and its materials skin depth δ is expressed as

$$C = 0.25 \tanh \frac{4\delta}{d} \tag{2.30a}$$

$$\delta = \frac{1}{\sqrt{\pi \sigma f \mu_0}} \tag{2.30b}$$

where σ is the conductivity of the wire material. For gold wires, $\delta = 2.486 f^{-0.5}$, where frequency f is expressed in gigahertz. When δ/d is small, $C = \delta/d$. The wire resistance R (in ohms) is given by

$$R = \frac{R_s l}{\pi d} \tag{2.30c}$$

Ground Plane Effect. The effect of the ground plane on the inductance value of a wire has also been considered [4, 47]. If the wire is at a distance h above the ground plane, it sees its image at $2h$ from it. The wire and its image result in mutual inductance L_{mg}. Since the image wire carries a current opposite to the current flow in the bond wire, the effective inductance of the bond wire becomes

$$L_e = L - L_{mg} \tag{2.31a}$$

$$L_{mg} = 2 \times 10^{-4} l \left[\ln \left\{ \frac{l}{2h} + \sqrt{1 + \left(\frac{l}{2h}\right)^2} \right\} + \sqrt{1 + \left(\frac{l}{2h}\right)^2} + \frac{2h}{l} - \sqrt{1 + \left(\frac{2h}{l}\right)^2} + C \right] \tag{2.31b}$$

From Eqs. (2.29)–(2.31)

$$L_e = 2 \times 10^{-4} l \left[\ln \frac{4h}{d} + \ln \left(\frac{l + \sqrt{l^2 + d^2/4}}{l + \sqrt{l^2 + 4h^2}} \right) + \sqrt{1 + \frac{4h^2}{l^2}} - \sqrt{1 + \frac{d^2}{4l^2}} - 2\frac{h}{l} + \frac{d}{2l} \right] \tag{2.32}$$

Multiple Wires. In many applications, multiple wires are required to carry higher currents, reduce wire inductance, or improve wire reliability. For example, in the case of two wires placed parallel to each other at a distance S between their centers, above the ground plane, the total inductance of the pair becomes

$$L_{ep} = \tfrac{1}{2}(L_e + L_m) \tag{2.33}$$

Here both wires carry current in the same direction; therefore the mutual inductances L_m and self inductances are in parallel and for each wire the self and mutual inductances add up in series. In this case L_m is given by

$$L_m = 2 \times 10^{-4} l \left[\ln \left\{ \frac{l}{S} + \sqrt{1 + \left(\frac{l}{S}\right)^2} \right\} + \sqrt{1 + \left(\frac{l}{S}\right)^2} - \sqrt{1 + \left(\frac{S}{l}\right)^2} + \frac{S}{l} \right] \tag{2.34}$$

Equations (2.31b) and (2.34) are identical under $2h = S$ and $C = 0$. The same procedure can be carried out for multiple wires.

So far we have treated uniformly placed horizontal wires above the ground plane. However, in practice, the wires are curved, nonhorizontal, and not par-

Table 2.13 Wire Inductance of 1-mil-Diameter Gold Wires

Length (mils)	Number of Wires	Loop Height (mils)	Spacing between Wires (mils)	Inductance Value (nH)	Method
19	1	7	—	0.28	EM simulated [64]
34	1	—	—	0.49	Calculated [65]
45	1	—	—	0.67	Calculated [65]
34	2	—	2	0.39	Calculated [65]
34	2	—	6	0.31	Calculated [65]
34	2	—	15	0.27	Calculated [65]
45	2	—	2	0.54	Calculated [65]
45	2	—	6	0.44	Calculated [65]
45	2	—	15	0.37	Calculated [65]
45	3	—	2	0.46	Calculated [65]
45	3	—	6	0.34	Calculated [65]
45	3	—	15	0.26	Calculated [65]
57	2	20	—	0.93	Measured [66]
57	3	20	—	0.73	Measured [66]
57	9	20	—	0.43	Measured [66]
57	13	20	—	0.38	Measured [66]
93	2	20	—	1.22	Measured [66]
93	8	20	—	0.60	Measured [66]
93	12	20	—	0.40	Measured [66]
93	14	20	—	0.42	Measured [66]

allel to each other. In such situations average values of S and h can be used. Also wires have shunt capacitance that can be calculated [58]. Table 2.13 lists inductances of 1-mil-diameter gold wires.

When a large current is passed through a wire, there is a maximum current value the wire can stand due to its finite resistance. At this maximum value, known as the fusing current, the wire will melt or burn out due to metallurgical fatigue. The fusing current is given by

$$I_f = Kd^{1.5} \tag{2.35}$$

where K depends upon the wire material and the surrounding environment. When the wire diameter d is expressed in millimeters, the K values for gold, copper, and aluminum wires are 183, 80, and 59.2, respectively. For 1-mil-diameter wires, I_f values for gold, copper, and aluminum wires are 0.74, 0.32, and 0.24 A, respectively. A safer maximum value for the current in wires used in assemblies is about half of the fusing current value. For example, to apply 1 A current, one requires three 1-mil-diameter wires of gold. When a wire is placed on a thermally conductive material such as Si or GaAs, its fusing current value is higher than the value given above.

2.4.3 Design of Capacitors

A lumped capacitor may be visualized as a small length of an open-circuited line. For $\gamma l \ll 1$ and $\bar{G} \ll \omega\bar{C}$,

$$Z_{\text{in}} = \frac{Z_0}{\gamma l} = \frac{\bar{G}}{(\omega\bar{C})^2 l} + \frac{1}{j\omega\bar{C}l} + \frac{\bar{R}l}{3} \tag{2.36}$$

Its equivalent circuit will thus be a resistance $[\bar{R}l/3 + \bar{G}/(\omega\bar{C})^2 l]$ in series with the capacitor $\bar{C}l$.

There are three types of passive capacitors generally used in microwave circuits: microstrip patch, interdigital, and metal–insulator–metal (MIM). Microstrip patch capacitors are comprised of a conductor patch on a dielectric substrate having the ground plane on the other side. These capacitors can only be used for low capacitance values (<0.2 pF) due to practical considerations such as capacitance per unit area. In order to have a large value of capacitance per unit area, it becomes necessary to use the interdigital structure or to decrease the distance between the two conductors (top and bottom) of the transmission-line section. The choice between the interdigital and MIM capacitors depends on the capacitance value to be realized and the processing technology available. Usually for values less than 1 pF, interdigital capacitors can be used, while for values greater than 1 pF, MIM techniques are generally used to minimize the overall size. Four types of capacitors used in MMICs are compared in Table 2.14; Schottky capacitors are described in Chapter 8.

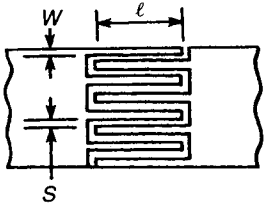
Capacitor configurations, equivalent circuits, and closed-form expressions for circuit elements are summarized in Table 2.15. Analysis of interdigital capacitors has been reported by Alley [49]. These capacitors can be fabricated employing an interdigital microstrip conductor by the technique used in the fabrication of MICs and do not require any additional processing steps.

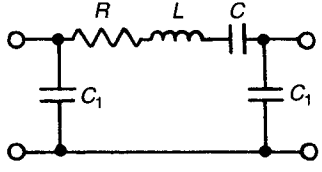
On the other hand, MIM capacitors can be used in chip form or can be fabricated by sandwiching a thin layer of dielectric between two parallel-plate conductors during the MIC process. In later stages the dielectric layer completely covers the bottom conductor pad with some overlap to prevent shorts. In MMICs, the dielectric layer normally used is silicon dioxide or silicon nitride and is 0.05–0.5 μm thick. The largest dimension is less than 0.1λ in dielectric

Table 2.14 Comparison of Various MMIC Capacitors

Capacitor Type	Capacitance Range (pF)	Tolerance
Microstrip	0–0.1 (shunt only)	2%
Interdigital	0.05–0.5	±10%
MIM	0.1–25	±10%
Schottky	0.5–100	Voltage dependent

Table 2.15 Expressions for Lumped Capacitors

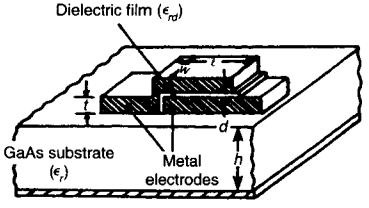


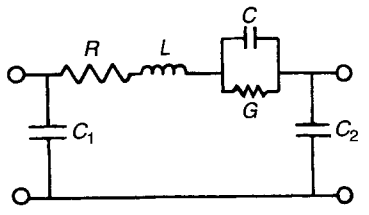


$$C\ (\text{pF}) = \frac{\epsilon_e 10^{-3}}{18\pi}\frac{K(k)}{K'(k)}(N-1)l$$

$$k = \tan^2\left(\frac{a\pi}{4b}\right), \qquad a = \frac{W}{2}, \qquad b = \frac{W+S}{2}$$

$$R\ (\Omega) = \frac{4}{3}\frac{R_s l}{WN}$$





$$C\ (\text{pF}) = \frac{10^{-3}\epsilon_{rd} Wl}{36\pi d}$$

$$R\ (\Omega) = \frac{KR_s l}{W+t}$$

$$G\ (\Omega^{-1}) = \omega C \tan\delta$$

$$C_1 = C_2\ (\text{pF}) = 10^{-2}\left[\frac{\sqrt{\epsilon_e}}{Z_0(W,h,\epsilon_r)} - \frac{\epsilon_r W}{360\pi h}\right]l$$

film. An accurate calculation for MIM capacitors should take into account the effect of fringing field [74].

For expressions in Table 2.15, N is the number of fingers, ϵ_e is the effective dielectric constant of microstrip line of width W, Z_0 is the characteristic impedance, h is the substrate thickness, ϵ_r is the dielectric constant of the substrate, ϵ_{rd} is the dielectric constant of the dielectric film, and $\tan\delta$ is its loss tangent. All dimensions are in micrometers. The elliptic functions $K(k)$ and $K'(k)$ are defined in (2.4).

When the capacitor is used in a circuit, the parasitic inductance L caused by the connection to the capacitor plates must be accounted for. The effective capacitance C_e is then given by

$$C_e = C\left(1 + \frac{\omega^2}{\omega_0^2}\right) \tag{2.37}$$

where $\omega_0^2 = 1/LC$ and ω is the operating angular frequency.

The Q of MIM capacitors is given by

$$Q = \frac{Q_d Q_c}{Q_d + Q_c} \tag{2.38}$$

where $Q_c = 1/\omega CR$ and $Q_d = 1/\tan\delta$. Quality factors of a square MIM capacitor as a function of capacitance for various values of the loss tangent are shown in Fig. 2.20.

2.4.4 Design of Resistors

Planar resistors can be realized either by depositing thin films of lossy metal on a dielectric base or by employing semiconductor films on a semi-insulating substrate. Nichrome and tantalum nitride are the most popular and useful film materials for thin-film resistors (thickness 0.05–0.2 μm). Resistors based on semiconductor (e.g., GaAs or Si) films can be fabricated by forming an isolated land of semiconductor conducting layer (thickness 0.05–0.5 μm).

These techniques are illustrated in Fig. 2.21 from the transmission-line analog. It can easily be shown that for a shorted line (neglecting L and G),

$$Z_{\text{in}} = \frac{R}{1 + j\omega CR/3} \tag{2.39}$$

where

$$R = R_s \frac{l}{W} \tag{2.40a}$$

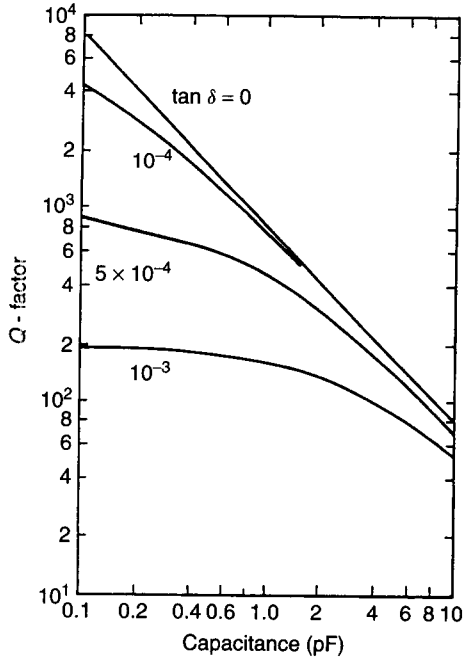


Figure 2.20 Quality factor of a square thin-film capacitor as a function of capacitance and dielectric loss tangent for $f = 10$ GHz.

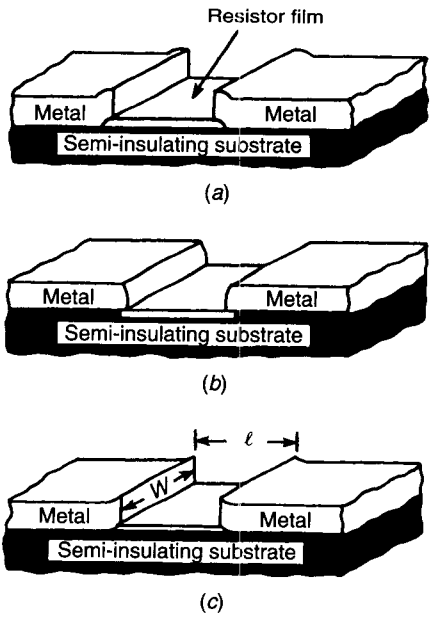


Figure 2.21 Planar resistors: (*a*) thin-film; (*b*) mesa; (*c*) implanted.

or

$$R = \rho_s \frac{l}{dW} \tag{2.40b}$$

Here R_s is the surface resistance (in ohms per square) and ρ_s is the specific resistivity (in ohm-meters) of the resistor film. The thickness d, width W, and length l of the film are measured in meters. The capacitance can be determined from the microstrip-line considerations. When film thickness $d \geq 1$ μm, the formula containing R_s should be used. However, for very thin films, $d \leq 1$ μm, the formula with ρ_s should be used. Desirable characteristics of film resistors are

- good stable resistance value, which should not change with time;
- low temperature coefficient of resistance (TCR);
- adequate dissipation capability;
- sheet resistivities in the range of 10–1000 Ω/square, so that parasitics can be minimized; and
- maximum resistor length less than 0.1λ if transmission-line effects are to be ignored.

A problem common to all planar resistors is the parasitic capacitance attributable to the underlying dielectric region and the distributed inductance, which makes such resistors exhibit a frequency dependence at high frequencies. If the substrate has a ground plane, one may determine the frequency dependence by treating the resistor as a very lossy microstrip line. Figure 2.22 shows how the VSWR increases dramatically at low values of ρ_s because the length of the resistor becomes too large. Also shown is the thermal resistance. Clearly, a trade-off is necessary between VSWR and thermal resistance.

Thin-film resistor materials are made of metals such as GeAu, Ta, Ti, Cr, and NiCr or of composite materials such as TiWN, TaN, and Ta_2N. Thin-film resistors are typically 0.1–0.4 μm thick and have limited current-carrying capacity. The maximum current density allowed by electromigration requirements in thin films is about 3×10^5 A/cm^2. Therefore, the current density per unit width for such resistors is of the order of 0.3–1.2 mA/μm.

High-power chip resistors and terminations are required to absorb unwanted power in RF and microwave couplers/hybrids and power dividers/combiners. Reliable design of such components requires substrate materials having high thermal conductivity and low dielectric constant values. Table 2.16 lists several potential candidates for chip resistors, and beryllium oxide (BeO) has been widely used for such applications as a substrate of choice. Both thin- and thick-film technologies are being used to manufacture high-power chip resistors. Nichrome (NiCr) and tantalum nitride (TaN) are commonly used as resistor

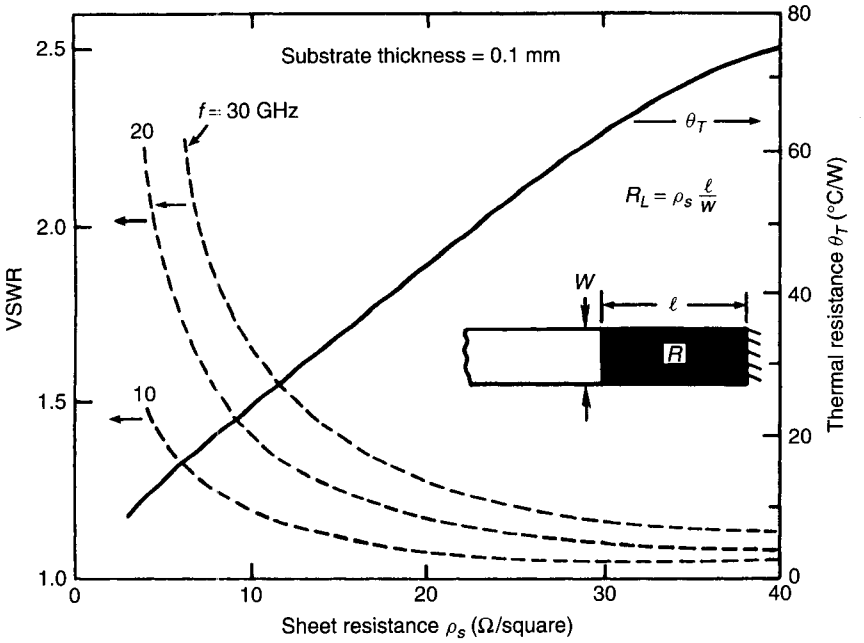


Figure 2.22 Thermal resistance and VSWR of planar resistor as function of sheet resistance and frequency.

thin films with TaN being preferred to NiCr, due to undesirable nickel in NiCr, which is believed to introduce unwanted intermodulation products in multi-carrier wireless systems. Thick-film resistors are manufactured using various compositions of ruthenium dioxide $(Ru)_2$ paste and a screen printing process.

Table 2.16 Comparison of Dielectric Substrates for High-Power Resistors

Property	Alumina	BN	BeO	AIN	SiC	Diamond
ϵ_r	9.9	4.2	6.7	8.5	45	5.7
K (W/m · °C)						
At 25°C	30	70	280	170	270	1400
At 100°C	25	—	200	150	190	—
At 200°C	—	—	150	125	150	—
Coefficient of thermal expansion (ppm/°C)	6.9	5.0	6.4	4.6	3.8	1.2
Shunt capacitance	Medium	Small	Small	Medium	Large	Small
Cost	Low	Low	Medium	Medium	Medium	High
Film adhesion	Excellent	Poor	Excellent	Good	Good	Poor
Machinability	Good	Good	Good	Good	Good	Poor

REFERENCES

1. Coaxial lines and waveguides have been discussed in several texts. For example, see: Ramo, S., J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, 2nd ed., Wiley, New York, 1984.
2. Montgomery, C. G., R. H. Dicke, and E. M. Purcell (Eds.), *Principles of Microwave Circuits*, McGraw-Hill, New York, 1948.
3. Bhartia, P., and I. J. Bahl, *Millimeter Wave Engineering and Applications*, Wiley, New York, 1984.
4. Gupta, K. C., R. Garg, and R. Chadha, *Computer-Aided Design of Microwave Circuits*, Artech House, Dedham, MA, 1981.
5. Cohn, S. B., "Characteristic Impedance of Shielded Strip Transmission Line," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-2, July 1954, pp. 52–55.
6. Cohn, S. B., "Problems in Strip Transmission Lines," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-3, Mar. 1955, pp. 119–126.
7. Vendelin, G. D., "Limitations on Stripline Q," *Microwave J.*, Vol. 13, May 1970, pp. 63–69.
8. Gunston, M. A. R., *Microwave Transmission-Line Impedance Data*, Van Nostrand-Reinhold, London, 1972, pp. 38–39.
9. Howe, H., Jr., *Stripline Circuit Design*, Artech House, Dedham, MA, 1974.
10. Bahl, I. J., and R. Garg, "Designer's Guide to Stripline Circuits," *Microwaves*, Vol. 17, Jan. 1978, pp. 90–96.
11. Wheeler, H. A., "Transmission Line Properties of a Stripline between Parallel Planes," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-26, Nov. 1978, pp. 866–876.
12. Barrett, R. M., "Microwave Printed-Circuit—Historical Survey," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-3, Mar. 1955, pp. 1–9.
13. Wheeler, H. A., "Transmission Line Properties of Parallel Strips Separated by a Dielectric Sheet," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-13, Mar. 1965, pp. 172–185.
14. Gupta, K. C., R. Garg, I. J. Bahl, and P. Bhartia, *Microstrip Lines and Slotlines*, 2nd ed., Artech House, Norwood, MA, 1996.
15. Schneider, M. V., B. Glance, and W. F. Bodtmann, "Microwave and Millimeter Wave Integrated Circuits for Radio Systems," *Bell Syst. Tech. J.*, Vol. 48, July–Aug. 1969, pp. 1703–1726.
16. Cohn, S. B., "Slot Line on a Dielectric Substrate," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-17, Oct. 1969, pp. 768–778.
17. Wen, C. P., "Coplanar Waveguide: A Surface Strip Transmission Line Suitable for Non-Reciprocal Gyromagnetic Device," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-17, Dec. 1969, pp. 1087–1090.
18. Gupta, K. C., and A. Singh (Eds.), *Microwave Integrated Circuits*, Wiley Eastern, New Delhi, India, 1974.
19. Young, L., and H. Sobol (Eds.), *Advances in Microwaves*, Vol. 8, Academic, New York, 1974.
20. Wheeler, H. A., "Transmission Line Properties of a Strip on a Dielectric Sheet on a

Plane," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-25, Aug. 1977, pp. 631–647.

21. Schneider, M. V., "Microstrip Lines for Microwave Integrated Circuits," *Bell Syst. Tech. J.*, Vol. 48, 1969, pp. 1422–1444.
22. Hammerstad, E. O., "Equations for Microstrip Circuit Design," in *Proceedings of the European Microwave Conference*, Microwave Exhibitors & Publishers, Kent, U.K., 1975, pp. 268–272.
23. Hammerstad, E. O., "Accurate Models for Microstrip Computer-Aided Design," *Dig. IEEE Int. Microwave Symp.*, 1980, pp. 407–409.
24. Bahl, I. J., and D. K. Trivedi, "A Designer's Guide to Microstrip Line," *Microwaves*, Vol. 16, May 1977, pp. 174–182.
25. Pramanick, P., and P. Bhartia, "CAD Models for Millimeter-Wave Finlines and Suspended-Substrate Microstrip Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-33, Dec. 1985, pp. 1429–1435.
26. Garg, R., and K. C. Gupta, "Expression for Wavelength and Impedance of Slotline," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-24, Aug. 1976, p. 532.
27. Cohn, S. B., "Slotline on a Dielectric Substrate," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-17, Oct. 1969, pp. 768–778.
28. Cohn, S. B., "Shielded Coupled-Strip Transmission Line," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-3, Oct. 1955, pp. 29–38.
29. Jansen, R. H., "High-Speed Computation of Single and Coupled Microstrip Parameters Including Dispersion, High Order Modes, Loss and Finite Strip Thickness," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-26, Feb. 1978, pp. 75–82.
30. Garg, R., and I. J. Bahl, "Characteristics of Coupled Microstriplines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, July 1979, pp. 700–705; also see correction in *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, Mar. 1980, p. 272.
31. Cohn, S. B., "Characteristic Impedances of Broadside-Coupled Strip Transmission Lines," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-8, Nov. 1960, pp. 633–637.
32. Howe, H., Jr., *Stripline Circuit Design*, Artech House, Dedham, MA, 1974, p. 125.
33. Shelton, J. P., Jr., "Impedances of Offset Parallel-Coupled Strip Transmission Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-14, Jan. 1966, pp. 7–15.
34. Bahl, I. J., and P. Bhartia, "The Design of Broadside-Coupled Stripline Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT 29, Feb. 1981, pp. 165–168.
35. Bahl, I. J., and P. Bhartia, "Characteristics of Inhomogeneous Broadside-Coupled Striplines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, June 1980, pp. 529–535.
36. Horno, M., and F. Medina, "Multilayer Planar Structures for High-Directivity Directional Coupler Design," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-34, Dec. 1986, pp. 1442–1449.
37. Mongia, R., I. Bahl, and P. Bhartia, *RF and Microwave Coupled-Line Circuits*, Artech House, Norwood, MA, 1999.
38. Wolff, I., and W. Menzel, "A Universal Method to Calculate the Dynamical Properties of Microstrip Discontinuities," in *Proceedings of the Fifth European Microwave Conference*, Microwave Exhibitors & Publishers, Kent, U.K., 1975, pp. 263–267.

39. Mehran, R., "The Frequency-Dependent Scattering Matrix of Microstrip Right-Angle Bends, T-Junctions and Crossings," *AEU: Archiv Für Elektronik und Übertragungstechnik*, Vol. 29, 1975, pp. 454–460.

40. Mehran, R., "Frequency-Dependent Equivalent Circuits for Microstrip Right-Angle Bends, T-Junctions and Crossings," *AEU*, Vol. 30, 1975, pp. 80–82.

41. Edwards, T. C., *Foundations for Microstrip Circuit Design*. Wiley, New York, 1981, Chap. 5.

42. Malherbe, J. A. G., and A. F. Steyn, "The Compensation of Step Discontinuities in TEM-Mode Transmission Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-26, Nov. 1978, pp. 883–885.

43. Larson, M. A., "Compensation of Step Discontinuities in Stripline," in *Proceedings of the Tenth European Microwave Conference*, Microwave Exhibitors & Publishers, Kent, U.K., 1980, pp. 367–371.

44. Chadha, R., and K. C. Gupta, "Compensation of Discontinuities in Planar Transmission Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Dec. 1982, pp. 2151–2155.

45. Hoefer, W. J. R., "A Contour Formula for Compensated Microstrip Steps and Open Ends," *IEEE Int. Microwave Symp. Dig.*, 1983, pp. 524–526.

46. Wheeler, H. A., "Simple Inductance Formulas for Radio Coils," *Proc. IRE*, Vol. 16, Oct. 1928, pp. 1398–1400.

47. Terman, F. E., *Radio Engineer Handbook*, McGraw-Hill, New York, 1943, p. 51.

48. Grover, F. W., *Inductance Calculations*, Van Nostrand, Princeton, NJ, 1946.

49. Alley, G. D., "Interdigital Capacitors and Their Applications to Lumped-Element Microwave Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-18, Dec. 1970, pp. 1028–1033.

50. Daly, D. A., et al., "Lumped Elements in Microwave Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-15, Dec. 1967, pp. 713–721.

51. Aitchison, C. S., "Lumped Components for Microwave Frequencies," *Philips Tech. Rev.*, Vol. 32, No. 9/10/11/12, 1971, pp. 305–314.

52. Aitchison, C. S., et al., "Lumped Microwave Circuits: Part III: Filters and Tunnel-Diode Amplifiers," *Design Electron.*, Vol. 9, Nov. 1971, pp. 42–51.

53. Aitchison, C. S., et al., "Lumped-Circuit Elements at Microwave Frequencies," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-19, Dec. 1971, pp. 928–937.

54. Caulton, M., "Lumped Elements in Microwave Integrated Circuits," in L. Young and H. Sobol (Eds.), *Adances in Microwaves*, Vol. 8, Academic, New York, 1974, pp. 143–203.

55. Pengelly, R. S., and D. C. Rickard, "Design, Measurement and Application of Lumped Elements up to *J*-Band," in *Proceedings of the Seventh European Microwave Conference*, Microwave Exhibitors & Publishers, Kent, U.K., 1977, pp. 460–464.

56. Gopinath, A., and P. Silvester, "Calculation of Inductance of Finite-Length Strips and Its Variations with Frequency," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-21, June 1973, pp. 380–386.

57. Chaddock, R. E., "The Application of Lumped Element Techniques to High Fre-

quency Hybrid Integrated Circuits," *Radio Electron. Eng.*, Vol. 44, 1974, pp. 414–420.

58. Caverly, R. H., "Characteristic Impedance of Integrated Circuit Bond Wire," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-34, Sept. 1986, pp. 982–984.
59. Pettenpaul, E., et al., "CAD Models of Lumped Elements on GaAs up to 18 GHz," *IEEE Trans. Microwave Theory Tech.*, Vol. 36, Feb. 1988, pp. 294–304.
60. Ladbrooke, P. H., *MMIC Design: GaAs FETs and HEMTs*, Artech House, Norwood, MA, 1989.
61. March, S. L., "Simple Equations Characterize Bond Wires," *Microwaves RF*, Vol. 30, Nov. 1991, pp. 105–110.
62. Wadell, B. C., *Transmission Line Design Handbook*, Artech House, Norwood, MA, 1991, Chap. 6.
63. Sadhir, V. K., I. J. Bahl, and D. A. Willems, "CAD Compatible Accurate Models of Microwave Passive Lumped Elements for MMIC Applications," *Int. J. Microwave Millimeter-Wave CAE*, Vol. 4, Apr. 1994, pp. 148–162.
64. Lee, H.-Y., "Wide-Band Characterization of a Typical Bonding Wire for Microwave and Millimeter-Wave Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. 43, Jan. 1995, pp. 63–68.
65. Patterson, H., "Analysis of Ground Wire Arrays for RFICs," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1997, pp. 765–768.
66. Liang, T., et al., "Equivalent-Circuit Modeling and Verification of Metal-Ceramic Packages for RF and Microwave Power Transistors," *IEEE Trans. Microwave Theory Tech.*, Vol. 47, June 1999, pp. 709–713.
67. Bahl, I. J., "High Performance Inductors," *IEEE Trans. Microwave Theory Tech.*, Vol. 49, Apr. 2001, pp. 654–664.
68. Lopez-Villegas, J. M., et al., "Improvement of the Quality Factor of RF Integrated Inductors by Layout Optimization," *IEEE RFIC Symp. Dig.*, 1998, pp. 169–172, and *IEEE Trans. Microwave Theory Tech.*, Vol. 48, Jan. 2000, pp. 76–83.
69. Bahl, I. J., "Improved Quality Factor Spiral Inductors on GaAs Substrates," *IEEE Microwave Guided Wave Lett.*, Vol. 9, Oct. 1999, pp. 398–400.
70. Danesh, M., et al., "A Q-Factor Enhancement Technique for MMIC Inductors," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1998, pp. 183–186.
71. Bahl, I. J., "High Current Capacity Multilayer Inductors for RF and Microwave Circuits," *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 10, Mar. 2000, pp. 139–146.
72. Bahl, I. J., et al., "Low Loss Multilayer Microstrip Line for Monolithic Microwave Integrated Circuits Applications," *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 8, Nov. 1998, pp. 441–454.
73. Ribas, R. P., et al., "Micromachined Microwave Planar Spiral Inductors and Transformers," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, Aug. 2000, pp. 1326–1335.
74. Wolff, I., and N. Knoppik, "Rectangular and Circular Microstrip Disk Capacitors and Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, Oct. 1974, pp. 857–864.

PROBLEMS

2.1 Calculate the magnitude of the characteristic impedance and the propagation constant for a lossy coaxial line at 2 GHz. Assume $b = 3a = 0.5$ cm and $\epsilon = (2.56 - j0.005)\epsilon_0$.

2.2 For a lossless transmission line terminated with a load, we have the characteristic impedance $Z_0 = 100\ \Omega$. The value of VSWR measured at the input is 4.0 and first voltage maximum is $\frac{1}{8}\lambda$ away from the load. What is the load impedance? (Use the Smith chart.)

2.3 A rectangular waveguide is to be designed for operation in the frequency range 7.5–10 GHz. Calculate the inside dimensions so that the following design criteria are satisfied: (a) There is only one mode of propagation, (b) the lowest usable frequency is 10% above the cutoff, and (c) highest usable frequency is 5% below the frequency where next higher mode can propagate.

2.4 A rectangular waveguide with inside dimensions 2.3×1.0 cm is used for operation at 10 GHz. Waveguide walls are made of copper (surface resistivity $R_s = 2.6 \times 10^{-7}\sqrt{f}$ ohms). Calculate the attenuation constant in nepers per meter and decibels per meter.

2.5 A rectangular waveguide with inside dimensions 3.1×1.55 mm is used for operation at 75 GHz. The waveguide walls are made of copper with a surface roughness of 1 μm root-mean-square (rms). Calculate the attenuation constant in decibels per centimeter for the waveguide.

2.6 A strip line of characteristic impedance 20 Ω is used at a frequency of 10 GHz with a load that is a microwave diode with conductance 0.05 S in parallel with a 1-pF capacitor. The line is one-eighth wavelengths long at the design frequency. Find the reflection coefficient at the load and the input admittance. If the dielectric constant of the substrate is 2.5, what is the physical length of the line?

2.7 Show that a pure TEM mode cannot be supported by a microstrip line. Calculate various parameters such as Z_0, β, and α for a microstrip with $\epsilon_r = 10$, $W = 1$ mm, and $h = 0.6$ mm at 10 GHz. Here gold conductors are 6 μm thick. Also calculate Q for a $\frac{1}{2}\lambda$ resonator.

2.8 A microstrip line of characteristic impedance Z_{01} is connected to another microstrip of higher characteristic impedance Z_{02}. If the step discontinuity reactance is represented by a series inductance (jX), determine the S matrix for the junction.

2.9 A microstrip line having a characteristic impedance of 50 Ω has a width $W = 1$ mm. The dielectric constant of the substrate is $\epsilon_r = 10$. What is the limiting frequency at which radiation starts? Is it necessary to correct for dispersion when operating at that frequency?

2.10 Since modes on microstrip lines are only quasi-TEM, derive from the basic theory of a lossless line that the inductance L and capacitance C per unit length of a microstrip line are given by

$$L = \frac{Z_0}{v} = \frac{Z_0\sqrt{\epsilon_e}}{c} \qquad C = \frac{1}{Z_0 v} = \frac{\sqrt{\epsilon_e}}{Z_0 c}$$

where Z_0 = characteristic impedance of the microstrip line
v = wave velocity in the microstrip line
$c = 3 \times 10^8$ m/s is the velocity of light in vacuum
ϵ_e = effective dielectric constant

Also determine L, C, and time delay per unit meter for the microstrip described in Problem 2.9.

2.11 A $\frac{1}{10}\lambda$ section of an open-circuited transmission line, with parameters given by $\bar{R} = 0.1$ Ω/cm, $\bar{L} = 0.1$ nH/cm, $\bar{G} = 0.2$ mΩ/cm, and $\bar{C} = 1$ pF/cm, is to be used as a lumped capacitor at 10 GHz. Find the value of capacitance, its equivalent circuit, and the Q factor.

2.12 Show that for a lumped element series mounted in the middle of a resonator, the value of reactance is given in terms of resonance frequency f as

$$X = 2Z_0 \cot \frac{\pi f}{2f_n}$$

where f_n is the resonant frequency for the line length l without reactance X being present and Z_0 is the characteristic impedance of the line. Stray reactances at open ends may be ignored.

3

RESONATORS

Arvind Sharma and A. P. S. Khanna

3.1 INTRODUCTION

Resonant structures are extensively used network elements in the realization of various microwave components [1]. At low frequencies, resonant structures are invariably composed of the lumped elements. As the frequency of operation increases, lumped elements in general cannot be used. Microwave resonant circuits can be realized by various forms of transmission lines. However, microwave resonant structures are almost invariably understood as the cavity resonator. Conventional resonators consist of a bounded EM field in a volume enclosed by metallic walls. The electric and magnetic energies are stored in the electric and magnetic fields, respectively, of the EM field inside the cavity, and the equivalent lumped inductance and capacitance of the structure can be determined from the respective stored energy. Some energy is dissipated due to finite conductivity of the metallic walls, and the equivalent resistance can therefore be determined from the currents flowing on the walls of the cavity resonator [2, 3].

In this chapter, a brief description of the resonant structures most commonly employed in various microwave components is presented. As far as possible, simple expressions have been provided for design applications. Basic parameters of microwave resonators are discussed in Section 3.2. The TE and TM cavity resonators are described in Section 3.3, and microstrip resonant structures are described in Section 3.4. Finally, dielectric resonators and yttrium–iron–garnet (YIG) resonators are treated in Sections 3.5 and 3.6, respectively. Resonator measurements are discussed in Section 3.7.

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

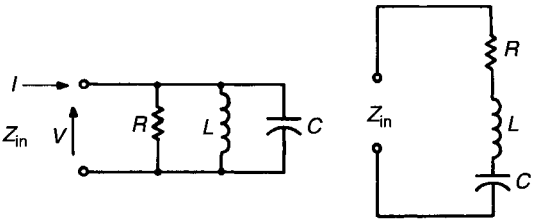


Figure 3.1 Lumped element (*a*) parallel and (*b*) series resonant circuits.

3.2 RESONATOR PARAMETERS

3.2.1 Resonant Frequency

The parameters of a resonator at microwave frequencies are essentially similar to those of a lumped-element resonator circuit at low frequencies. They can easily be described utilizing an *RLC* series or parallel network. Consider, for instance, an *RLC* parallel network as shown in Fig. 3.1. The input impedance of such a network as a function of frequency has both real and imaginary parts. At resonance, the input impedance is real and is equal to the resistance of the circuit. The electric and magnetic stored energies are also equal, leading to the expression for the resonant frequency as

$$\omega_0 = \frac{1}{\sqrt{LC}} \tag{3.1}$$

3.2.2 Quality Factor

The performance of a resonant circuit is described in terms of the quality factor Q, and various features such as frequency selectivity, bandwidth, and damping factors can be deduced from this. The quality factor is defined as

$$Q = \omega \frac{\text{time averaged stored energy}}{\text{energy lost per second}} \tag{3.2}$$

For the lumped resonant circuits

$$Q = \omega RC = \frac{R}{\omega L} \tag{3.3a}$$

for the parallel network (Fig. 3.1*a*) and

$$Q = \frac{1}{\omega CR} = \frac{\omega L}{R} \tag{3.3b}$$

for the series network (Fig. 3.1*b*).

3.2.3 Fractional Bandwidth

The input impedance of the parallel resonant circuit of Fig. 3.1 is given by

$$Z_{\text{in}} = \left(\frac{1}{R} + \frac{1}{j\omega L} + j\omega C\right)^{-1} \tag{3.4}$$

At a frequency $\omega_0 \pm \Delta\omega$ in the vicinity of the resonant frequency, (3.4) reduces to

$$Z_{\text{in}} = \frac{R}{1 \pm j2Q(\Delta\omega/\omega_0)} \tag{3.5}$$

From (3.5) it is clear that at $\omega = \omega_0$ the input impedance is only resistive. However, when

$$\Delta\omega = \frac{\omega_0}{2Q} \tag{3.6}$$

the magnitude of the input impedance decreases to $R\sqrt{2}$ of its maximum value R, and the phase angle is $\pi/4$ for $\omega < \omega_0$ and $-\pi/4$ for $\omega > \omega_0$. From (3.6) the fractional bandwidth BW is defined as

$$\text{BW} = \frac{2\Delta\omega}{\omega_0} = \frac{1}{Q} \tag{3.7}$$

3.2.4 Loaded Quality Factor

In practical situations, the resonant circuit is coupled to an external load R_L that also dissipates power, and the loaded quality factor Q_L is given by

$$\frac{1}{Q_L} = \frac{1}{Q} + \frac{1}{Q_e} \tag{3.8}$$

where Q_e is the external quality factor for a lossless resonator in the presence of the load.

3.2.5 Damping Factor

Another important parameter associated with a resonant circuit is the damping factor δ_d. It is a measure of the rate of decay of the oscillations in the absence of an exciting source. For high-Q resonant circuits, the rate at which the stored energy decays is proportional to the average energy stored. Consequently, the stored energy as a function of time is given by

$$W = W_0 e^{-2\delta_d t} = W_0 e^{-\omega_0 t/Q} \tag{3.9}$$

which implies that

$$\delta_d = \frac{\omega_0}{2Q} \tag{3.10}$$

Thus, we see that the damping factor is inversely proportional to the Q of the resonant circuit. In the presence of an external load, the Q should be replaced by Q_L.

Alternately, the input impedance in the vicinity of resonance Z_{in} given by (3.5) can be rewritten to take into account the effect of losses in terms of the complex resonant frequency

$$\omega_c = \omega_0 + j\delta_d = \omega_0\left(1 + j\frac{1}{2Q}\right) \tag{3.11}$$

so that

$$Z_{\text{in}} = \frac{\omega_0 R/2Q}{j(\omega - \omega_c)} \tag{3.12}$$

In (3.12) the parameter R/Q is called the figure of merit and describes the effect of the cavity on the gain–bandwidth product. In terms of the lumped elements of the resonant circuit,

$$\frac{R}{Q} = \sqrt{\frac{L}{C}} \tag{3.13}$$

3.2.6 Coupling

Coupling structures provide a means of coupling energy into and/or out of the cavity. The excitation of the cavity can be accomplished by the following:

1. *Current Loops.* The plane of the loop is perpendicular to the direction of the magnetic field in the cavity (Fig. 3.2*a*).
2. *Electric Probes.* The direction of probe is parallel to the direction of the electric field in the cavity (Fig. 3.2*b*).
3. *Apertures.* The aperture is located between the cavity and input waveguide so that a field component in the cavity mode has the same direction to the one in the input waveguide (Fig. 3.2*c*).

3.3 CAVITY RESONATORS

At microwave frequencies, the dimensions of lumped resonator circuits become comparable to the wavelength, and this may cause energy loss by radiation.

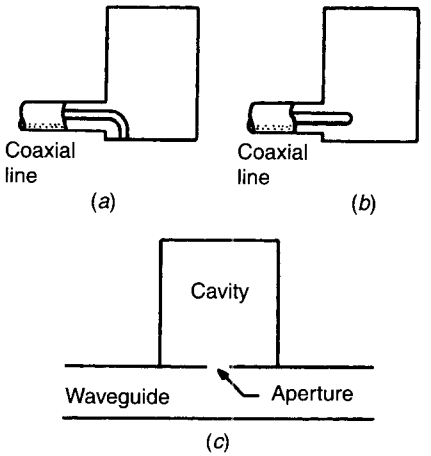


Figure 3.2 Cavity excitation using (*a*) loop coupling, (*b*) electric probe coupling, and (*c*) aperture coupling.

Therefore, resonant circuits at these frequencies are shielded to prevent radiation. Perfectly conducting enclosures, or cavities, provide a means of confining energy. Usually, cavities with the largest possible surface area for the current path are preferred for low-loss operation and the energy coupled to them by the methods previously described.

3.3.1 Coaxial Resonators

A coaxial cavity resonator (Fig. 3.3) supporting TEM waves can easily be formed by a shorted section of coaxial line. Resonances appear whenever the length $2d$ of the cavity is an integral number of wavelengths. The resonance modes occur at

$$f = \frac{nc}{2d} \qquad n = 1, 2, \ldots \tag{3.14}$$

where c is the speed of light. The lowest resonant frequency corresponds to

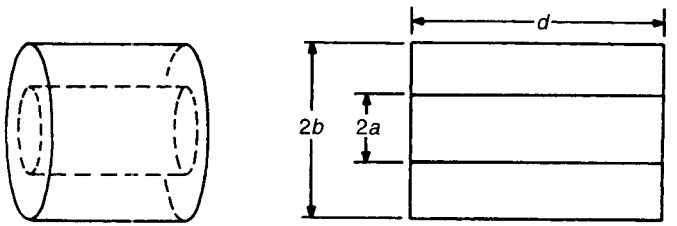


Figure 3.3 Coaxial cavity resonator and its cross section.

$n = 1$ and the Q of the cavity for this mode is given by [4]

$$Q\frac{\delta}{\lambda_0} = \frac{1}{4 + 2(d/b)(1 + b/a)/\ln(b/a)} \tag{3.15}$$

where δ is the skin depth and a and b are inner and outer radii, respectively.

It is also possible to have higher order resonance modes, depending on the structural parameters of the coaxial line. The first higher order mode appears when the average circumference is equal to the wavelength in the dielectric medium of the line. The cutoff frequency of this mode is

$$f_c = \frac{c}{\pi\sqrt{\epsilon_r}(a + b)} \tag{3.16}$$

where ϵ_r is the dielectric constant of the medium. Other higher order modes correspond to TE and TM waves that exist in a circular waveguide with the radius of the center conductor approaching zero. The resonance condition is

$$k_{nml} = \left[p_{nm}^2 + \left(\frac{l\pi}{2d}\right)^2\right]^{1/2} \tag{3.17}$$

where $k_{nml} = 2\pi f_{nml}/c$ and p_{nm} is the cutoff wavenumber that is obtained as the mth root of the transcendental equations

$$J_n'(ka)N_n'(kb) - J_n'(kb)N_n'(ka) = 0 \tag{3.18}$$

for TE modes and

$$J_n(ka)N_n(kb) - J_n(kb)N_n(ka) = 0 \tag{3.19}$$

for TM modes. Here J_n and N_n are the nth-order Bessel functions of the first and second kinds, respectively and the prime denotes their derivatives with respect to their arguments.

3.3.2 Reentrant Coaxial Resonators

Another coaxial cavity configuration consists of a short section of coaxial line with a gap in the center conductor. Figure 3.4*a* shows a capacitively loaded coaxial cavity. Radial cavity as shown in Fig. 3.4*b* is another possible variation. They are also referred to as reentrant coaxial cavities, since the metallic boundaries extend into the interior of the cavity. They are widely used in microwave tubes. The resonant frequency of such a structure can be evaluated from the solution of the transcendental equation [5]

$$\tan \beta l = \frac{dc}{\omega a^2 \ln(b/a)} \tag{3.20}$$

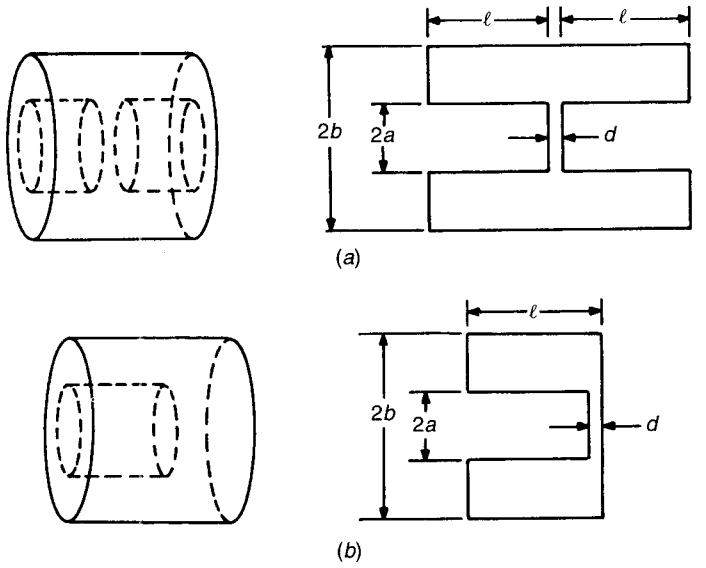


Figure 3.4 Reentrant coaxial cavity resonators: (*a*) capacitively loaded coaxial cavity resonator; (*b*) radial-cavity resonators.

where d is the gap in center conductor and $2l + d$ is the length of the cavity. From (3.20) it is obvious that the capacitively loaded coaxial cavity can have an infinite number of modes.

For the radial reentrant cavity of Fig. 3.4*b*, the resonant frequency can be evaluated by calculating the inductance and capacitance of the structure. The expression for the resonant frequency is

$$f = \frac{c}{2\pi\sqrt{\epsilon_r}} \left[al \left(\frac{a}{2d} - \frac{2}{l} \ln \frac{0.765}{\sqrt{l^2 + (b-a)^2}} \right) \ln \frac{b}{a} \right]^{-1/2} \tag{3.21}$$

An approximate expression for the Q of the cavity is

$$Q\frac{\delta}{\lambda} = \frac{2l}{\lambda} \frac{\ln(b/a)}{2\ln(b/a) + l[(1/a) + (1/b)]} \tag{3.22}$$

For a tunable reentrant cavity, d is large, and $l - d$ is also large compared with b. The resonances occur whenever the length of the center conductor is approximately a quarter wavelength.

3.3.3 Rectangular Waveguide Resonators

Rectangular resonant cavities are formed by a section of rectangular waveguide of length d. This cavity can also support an infinite number of modes. The field

configuration of the standing-wave pattern for the incident and reflected waves is not unique, that is, it depends on the assumed direction of propagation of the wave. In order to be consistent, we shall assume that wave propagation is in the positive z direction. The standing-wave pattern is then formed by the incident and reflected waves traveling in $+z$ and $-z$ directions, respectively.

The cutoff wavenumber k_{cmn} is given by

$$k_{cmn}^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 \qquad m = 0, 1, 2, \ldots \qquad n = 0, 1, 2, \ldots \tag{3.23}$$

where a and b are waveguide dimensions. The resonant wavenumber is then expressed as

$$k_{mnp} = \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2\right]^{1/2} \tag{3.24}$$

and the resonant frequency is defined as

$$f_{mnp} = \frac{k_{mnp}c}{2\pi} \tag{3.25}$$

From the preceding discussion, we see that the resonant frequency is the same for TE and TM modes. Therefore, they are referred to as degenerate modes. The field configuration of the dominant TE_{101} mode is shown in Fig. 3.5*b*. The quality factor Q of the dominant TE_{101} mode in the rectangular res-

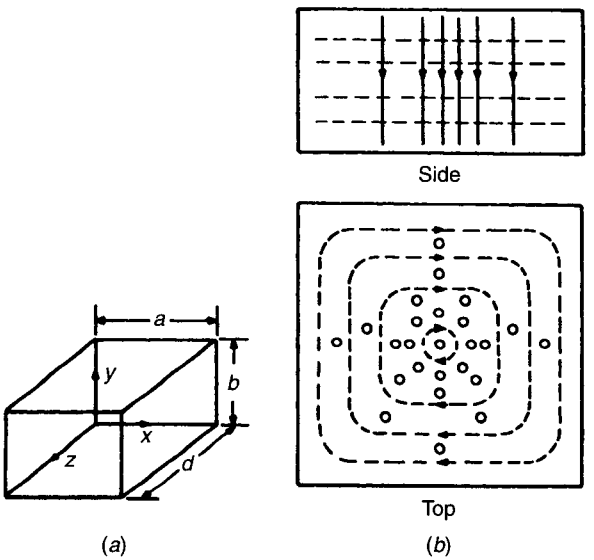


Figure 3.5 (*a*) Rectangular waveguide cavity resonator. (*b*) Field configuration of dominant TE_{101} mode.

onant cavity having surface resistance R_s can be evaluated using the expression

$$Q = \frac{120\pi^2}{4R_s}\left[\frac{2b(a^2+d^2)^{3/2}}{ad(a^2+d^2)+2b(a^3+b^3)}\right] \tag{3.26}$$

In rectangular cavities, the resonant frequency increases for higher order modes, as does the Q at a given frequency. Higher order mode cavity or "echo boxes" are useful in applications where a slow rate of decay of the energy stored in the cavity after it has been excited is required.

3.3.4 Circular Waveguide Resonators

Circular waveguide cavities are most useful in various microwave applications. Most commonly, they are used in wavemeters to measure frequency and have a high Q factor and provide greater resolution. It consists of a section of a circular waveguide of radius a and length d (Fig. 3.6*a*), with end plates to provide short circuit.

The resonance wavenumber of the circular waveguide cavity is given by

$$k_{nml} = \left[\left(\frac{x_{nm}}{a}\right)^2 + \left(\frac{l\pi}{d}\right)^2\right]^{1/2} \qquad l = 0, 1, 2, \ldots \tag{3.27}$$

where

$$x_{nm} = \begin{cases} p'_{nm} & \text{for TE modes} \\ p_{nm} & \text{for TM modes} \end{cases}$$

Values for p'_{nm} for various modes are given in Table 3.1. Field lines for TE_{111}, TM_{011}, and TE_{011} modes are shown in Fig. 3.6*b*. Simplifying (3.27) yields

$$(2af_{nml})^2 = \left(\frac{cx_{nm}}{\pi}\right)^2 + \left(\frac{cl}{2}\right)^2\left(\frac{2a}{d}\right)^2 \tag{3.28}$$

The Q of the circular cavity for TE_{nml} modes can be evaluated from

$$Q\frac{\delta}{\lambda_0} = \frac{[1-(n/p'_{nm})^2][(p'_{nm})^2+(l\pi a/d)^2]^{3/2}}{2\pi[(p'^2_{nm}+2a/d(l\pi a/d)^2+(1-2a/d)(nl\pi a/p'_{nm}d)^2]} \tag{3.29}$$

and for the dominant TE_{111} mode, Q can be obtained by substituting $n = m = l = 1$ in the preceding equation.

Using (3.28), plots of $(2af)^2$ versus $(2a/d)^2$ can be used to construct mode charts, as shown in Fig. 3.7. From this it can be seen that for the TE_{011} mode operation, the safe value of $(2a/d)^2$ is between 2 and 3.

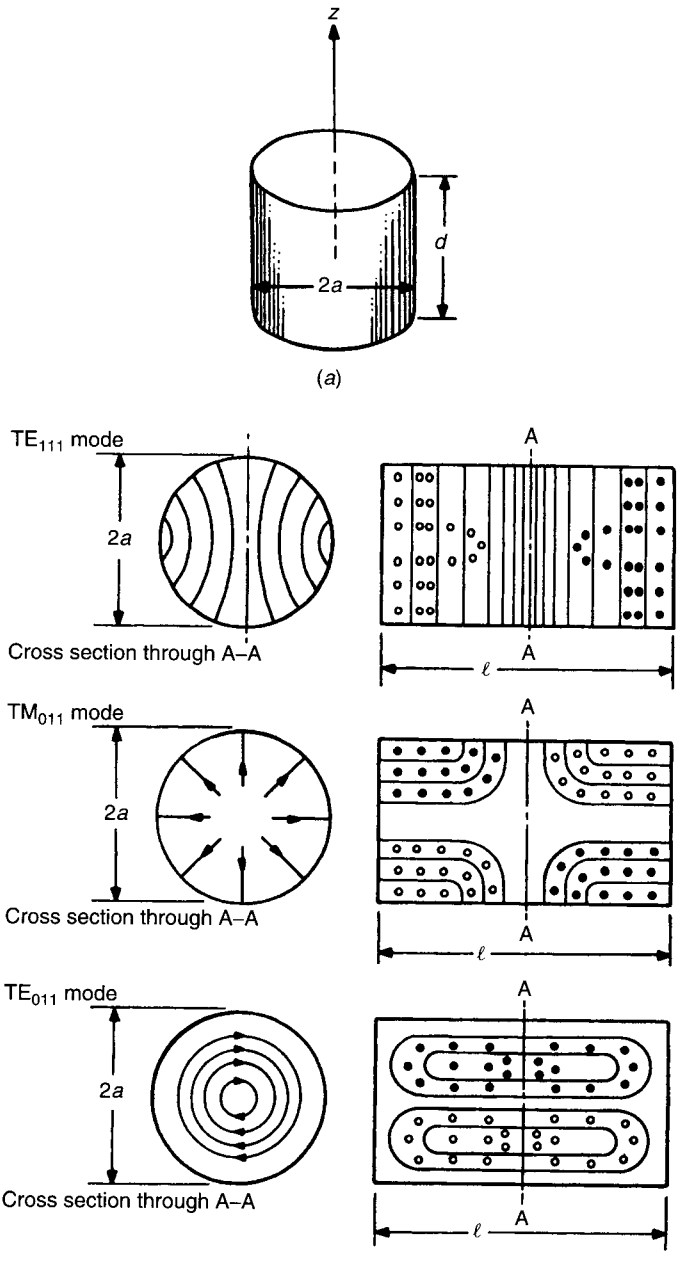


Figure 3.6 (*a*) Circular cylindrical waveguide cavity resonator. (*b*) Field configurations for TE_{111}, TM_{011}, and TE_{011} modes in cylindrical cavities.

Table 3.1 Roots of the Transcedental Equation $J_n'(ka) = 0$

Modes		
n	m	p'_{nm}
0	1	0
1	1	1.841
2	1	3.054
0	2	3.832
3	1	4.201
4	1	5.318

For TM modes, the Q is given by

$$Q\frac{\delta}{\lambda_0} = \frac{[p_{nm}^2 + (l\pi a/d)^2]^{1/2}}{2\pi(1 + 2a/d)} \quad \text{for } l > 0$$

$$\simeq \frac{p_{nm}}{2\pi(1 + a/d)} \quad \text{for } l = 0 \tag{3.30}$$

As with rectangular cavity resonators, the Q is higher for higher order modes.

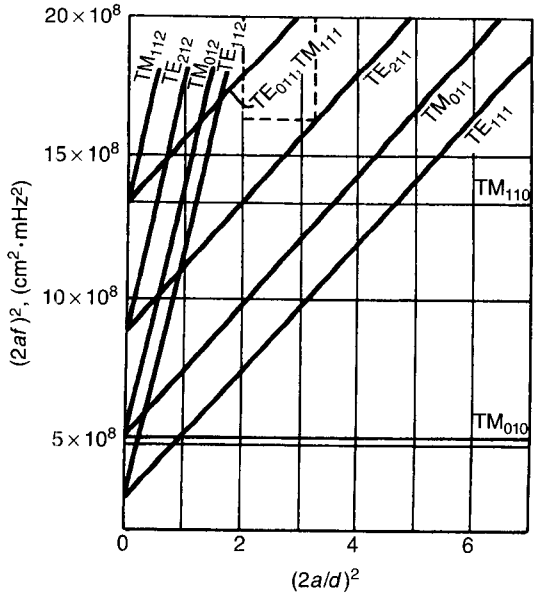


Figure 3.7 Mode chart of circular cylindrical cavity resonator.

3.3.5 Elliptic Waveguide Resonators

Elliptic resonant cavities formed using a section of an elliptic waveguide offer several advantages. There is no mode splitting due to slight deformations in the cavity surface, and the electric field configuration in the transverse plane is fixed with respect to its axes. Also, the longitudinal electric field of the TM_{111} mode in an elliptic cavity with semimajor axis a is always greater than the circular cavity with radius a. This feature may be useful in the dielectric material characterization utilizing perturbation techniques [6]. The resonance wavenumbers of an elliptic resonant cavity are given by Kretzschmar [7]. However, since elliptical waveguide cavities are not widely used, further details are not included and the reader is referred elsewhere [6–8].

3.4 PLANAR MICROSTRIP RESONANT STRUCTURES

Particularly in the last decade, there has been a spate of technical publications on various planar transmission lines, resonant structures, and circuits. Of the various planar transmission lines, the microstrip lines, slot line, and fin line have evolved as the most popular transmission lines in practice and have received the maximum attention from the theoretical point of view. This interest has stemmed from their applications in a wide variety of MIC components, such as directional couplers, filters, and oscillators. They are also used in the measurement of dielectric constant and of dispersion in a transmission line. The resonant structures have been analyzed using various analytical and numerical techniques. The main objective has been to evolve suitable design criteria so that complex microwave circuits can be designed to perform their prescribed functions as accurately as possible without resorting to the "cut-and-try" approach.

Microstrip lines are commonly employed in MICs. A microstrip resonator is a planar conductor patch on a dielectric substrate, generally of a regular shape. The EM energy is confined to the dielectric region between the top conductor patch and the bottom ground plane, surrounded by a perfect magnetic wall on its contour. The dielectric materials frequently used are alumina, sapphire, and glass or ceramic-reinforced Teflon. The properties of the dielectric material greatly influence the characteristics of the resonant structures. Both the dielectric and conductor contribute to the losses. In the open configuration, radiation from microstrip resonators depends mainly on the dielectric constant (ϵ_r) of the substrate material and its thickness (h). Microstrip resonators are normally coupled through microstrip lines that excite TM modes.

Of the various microstrip resonant structures, rectangular, circular disk, and ring geometries find extensive applications in oscillators, filters, and circulators. Microstrip resonant structures of complex geometric shapes can, in general, be fabricated to provide better performance and greater flexibility in the design. This has been demonstrated for the equilateral triangle, regular hexagon, elliptic disk, and ring resonators.

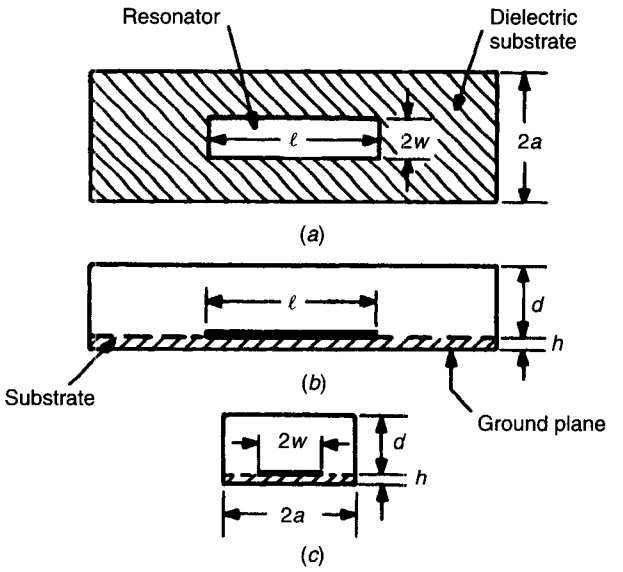


Figure 3.8 Rectangular microstrip resonator in shielding waveguide: (*a*) top view; (*b*) side view; (*c*) end view.

3.4.1 Rectangular Microstrip Resonators

Rectangular microstrip resonators are extensively used in filters, mixers, and other microwave circuit components. As shown in Fig. 3.8, the rectangular microstrip resonator consists of a rectangle of length l and width $2W$ on a dielectric substrate of thickness h and relative dielectric constant ϵ_r.

To a first approximation, it can be viewed as an open-ended transmission line of length l. The transmission line is resonant whenever the length of the line is an integer multiple of the half guide wavelength. The open end in a microstrip is not truly an open circuit due to the fringing fields [9].

The fringing capacitance can be modeled as an hypothetical extension Δl of the microstrip line at each end [10, 11], as given in Section 2.3. The resonant frequency of a rectangular microstrip resonator under quasi-static approximation is then given by [12]

$$f_r = \frac{c}{2\sqrt{\epsilon_e}(l + 2\Delta l)} \tag{3.31}$$

where ϵ_e is the effective dielectric constant of the microstrip, as given in Section 2.2.

An accurate description of the resonant frequency can be obtained from a solution of the three-dimensional electromagnetic boundary-value problem. Itoh [13] formulated the rectangular microstrip resonator problem utilizing

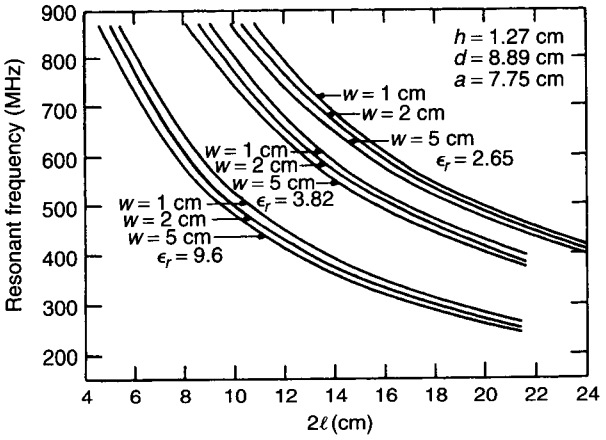


Figure 3.9 Resonant frequency of rectangular microstrip resonator as function of resonator length. (After Itoh [13]. Reprinted with permission of IEEE.)

hybrid modes in a rectangular waveguide shielding enclosure. The associated boundary-value problem was solved utilzing the spectral domain technique. The resonant frequencies computed using this technique are shown in Fig. 3.9.

3.4.2 Circular Microstrip Disk Resonators

Circular microstrip disk resonators have applications similar to the rectangular microstrip resonator. The resonant structure consists of a circular disk of radius a on a dielectric substrate, as shown in Fig. 3.10*a*. For a substrate thickness much less than the free-space wavelength ($h \ll \lambda_0$), the microstrip disk can be modeled as a cylindrical cavity with magnetic walls. The field inside the dielectric region corresponds to those of the TM_{nml} modes with $l = 0$. The resonant frequency of the cavity is given by [14]

$$f_{nm0} = \frac{p'_{nm}c}{2\pi a_e\sqrt{\epsilon_r}} \tag{3.32}$$

where p'_{nm} is mth root of the equation

$$J'_n(ka) = 0 \tag{3.33}$$

where J'_n is the derivative with respect to the argument of the Bessel function of the first kind and order n. For various values of n and m, the root of (3.33) can easily be found. Table 3.1 gives the first few roots. The effective radius a_e is given by [15–17]

$$a_e = a\left[1 + \frac{2h}{\pi a\epsilon_r}\left(\ln\frac{\pi a}{2h} + 1.7726\right)\right]^{1/2} \tag{3.34}$$

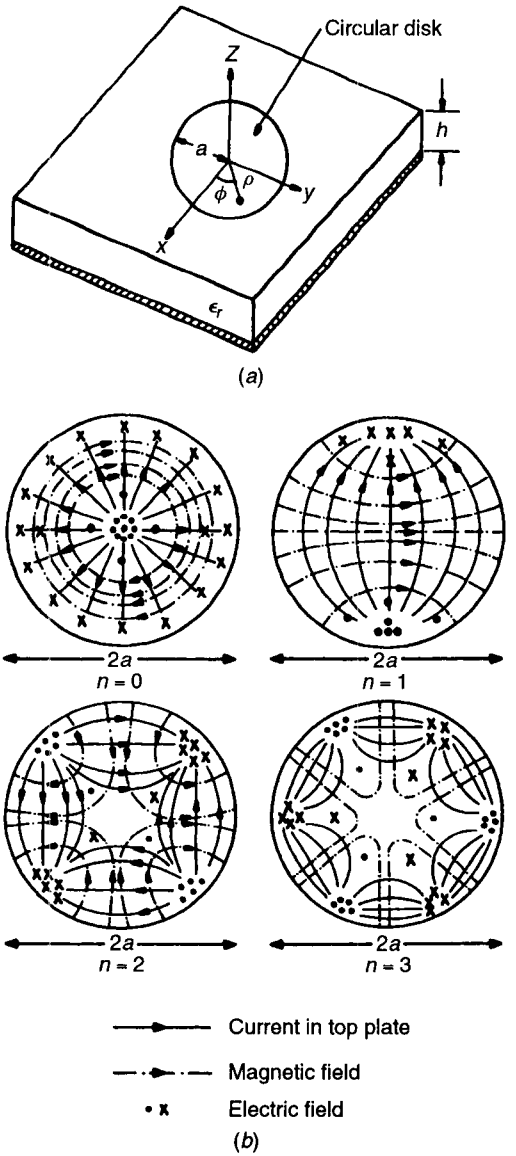


Figure 3.10 (*a*) Circular microstrip resonator. (*b*) Field configurations in circular microstrip disk resonator.

The field configurations of the first few resonance modes are shown in Fig. 3.10*b*.

The theoretical results calculated using a quasi-static formulation of the spectral domain technique [14, 15] yield results that are within 2.5% of the experimental values. To simplify the design of microstrip disk resonators [18],

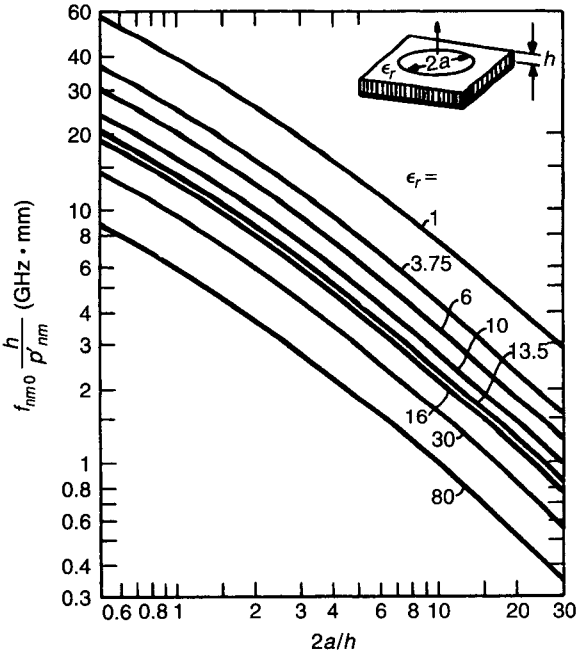


Figure 3.11 Design chart for microstrip disk resonator.

rearranging (3.32) produces

$$f_{nm0}\frac{h}{p'_{nm}} = F\left(\frac{h}{a};\epsilon_r\right) \tag{3.35}$$

The function on the left side is plotted in Fig. 3.11 as a function of $2a/h$ for various values of ϵ_r.

3.4.3 Circular Microstrip Ring Resonators

The microstrip ring resonator is another extensively used MIC component for the measurement of dispersion characteristics of microstrip lines [19–21]. The geometry of the microstrip ring resonator is shown in Fig. 3.12, with the inner and outer radii a and b, respectively. The resonance condition for the microstrip ring resonator can be obtained from EM fields existing in an annular cavity surrounded by magnetic walls. It is given by

$$J'_n(kb)\,Y'_n(ka) - J'_n(ka)\,Y'_n(kb) = 0 \tag{3.36}$$

where J_n and Y_n are Bessel functions of the first and second kind and order n and the prime denotes derivative with respect to the argument. The resonance

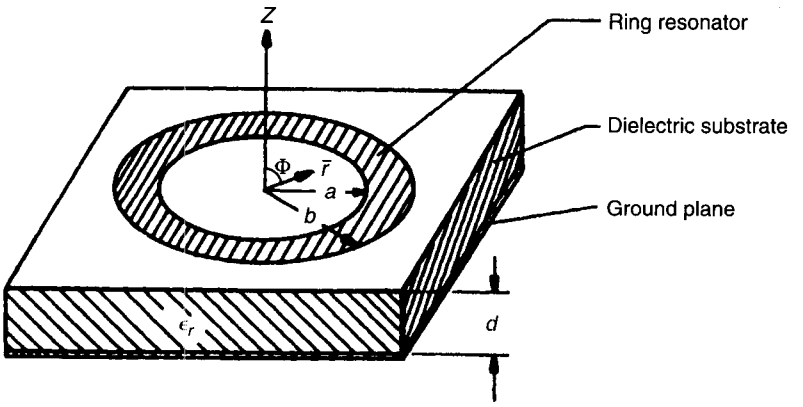


Figure 3.12 Circular microstrip ring resonator.

wavenumber k_{nm0} is obtained from the solution of the transcendental equation (3.36). The resonant frequency is then given by

$$f_{nm0} = \frac{ck_{mn0}}{2\pi\sqrt{\epsilon_e}} \tag{3.37}$$

The mode charts for a microwave ring resonator are shown in Fig. 3.13. The field patterns of some lower order resonance modes are shown in Fig. 3.14 [22].

3.4.4 Triangular Microstrip Resonators

Another important geometry for a wide variety of applications is the equilateral triangular microstrip resonator, as shown in Fig. 3.15. The 120° symmetry property of this element was utilized in an articulate design of circulators [23, 24]. Cuhaci and James [25] showed that, as a resonator, this element exhibits a slightly higher radiation Q factor (Q_r) than the corresponding circular microstrip disk resonator. This is a significant advantage in the design of low-loss MICs.

The resonance wavenumber for an equilateral triangular resonator is given as [23]

$$k_{nmp} = \frac{4\pi}{3a}\sqrt{m^2 + mn + n^2} \tag{3.38}$$

where a is the triangle side. The integers m, n, and p are such that

$$m + n + p = 0 \tag{3.39}$$

The dominant mode is achieved from one of the sets of the following integer

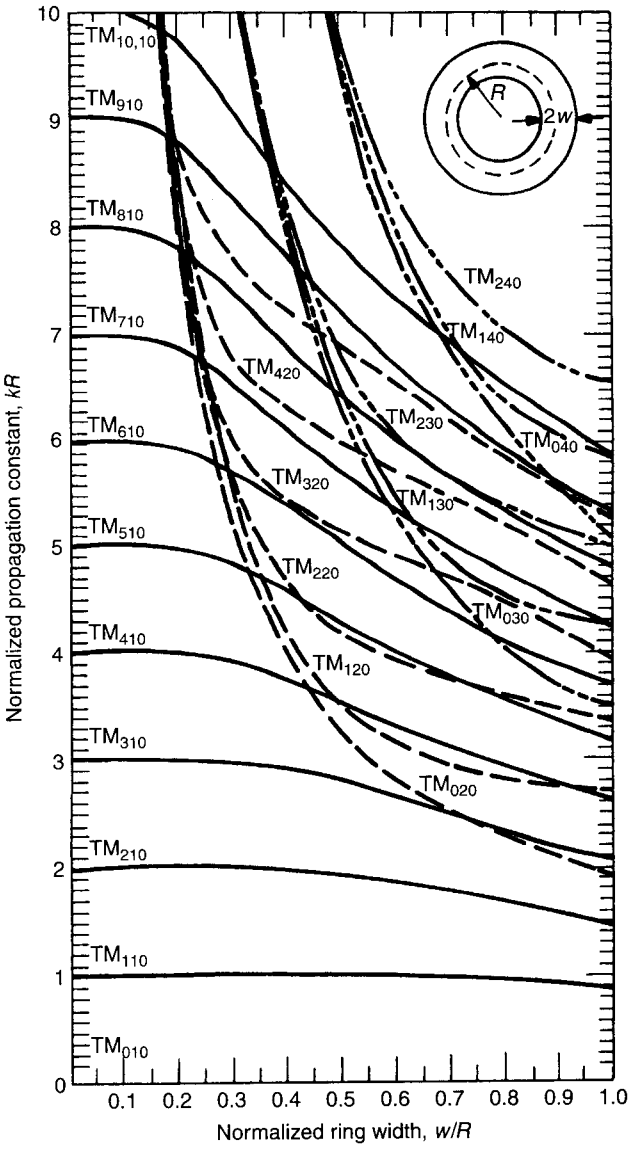


Figure 3.13 Mode chart for circular microstrip ring resonator. (After Wu and Rosenbaum [22]. Reprinted with permission of IEEE.)

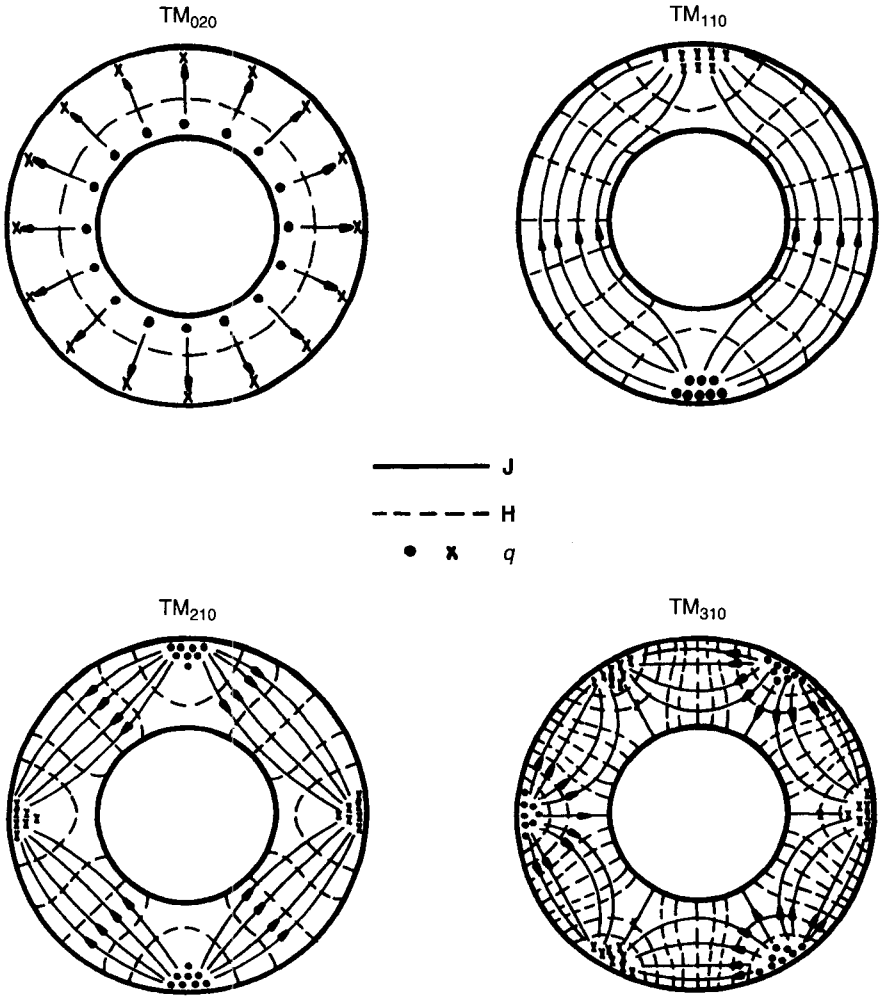


Figure 3.14 Field configurations for circular microstrip ring resonators.

values:

$$\begin{array}{llll} \text{(i)} & m=1 & n=0 & p=-1 \\ \text{(ii)} & m=0 & n=1 & p=-1 \\ \text{(iii)} & m=1 & n=-1 & p=0 \end{array} \tag{3.40}$$

Thus the dominant wavenumber is given by

$$k=\frac{4\pi}{3a} \tag{3.41}$$

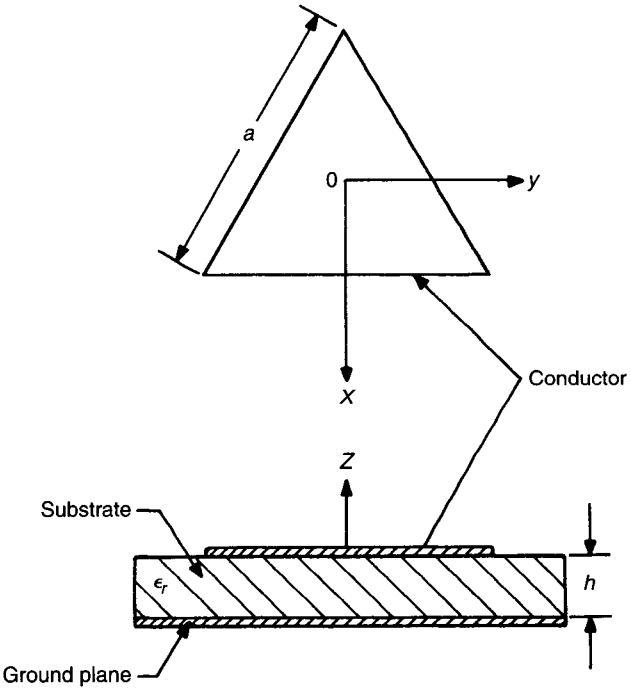


Figure 3.15 Equilateral triangular microstrip resonator.

The field patterns for the TM_{10} and TM_{11} modes in a triangular resonator are shown in Fig. 3.16. Another triangular shape that has not received much attention is the isosceles right-angled triangular microstrip resonator. The resonance condition, in this case, is given by [26]

$$k_{mn} = \left(\frac{\pi}{a}\right)\sqrt{m^2 + mn + 2n^2} \tag{3.42}$$

where a is the length of the isosceles triangle side. The integers m and n determine the mode of resonance, and the dominant mode is obtained when $m = 0$ and $n = 1$. The resonance wavenumber is then given by [27]

$$k_{01} = \sqrt{2}\frac{\pi}{a} \tag{3.43}$$

As such, no convenient exact method exists for solving the general triangular microstrip resonator except for equilateral and isosceles right-angled shapes [26]. A numerical method for the eigenvalues of the TE and TM modes propagating in a triangular waveguide of arbitrary dimension does exist [28]. Also, the general isosceles triangular microstrip resonator has been analyzed using

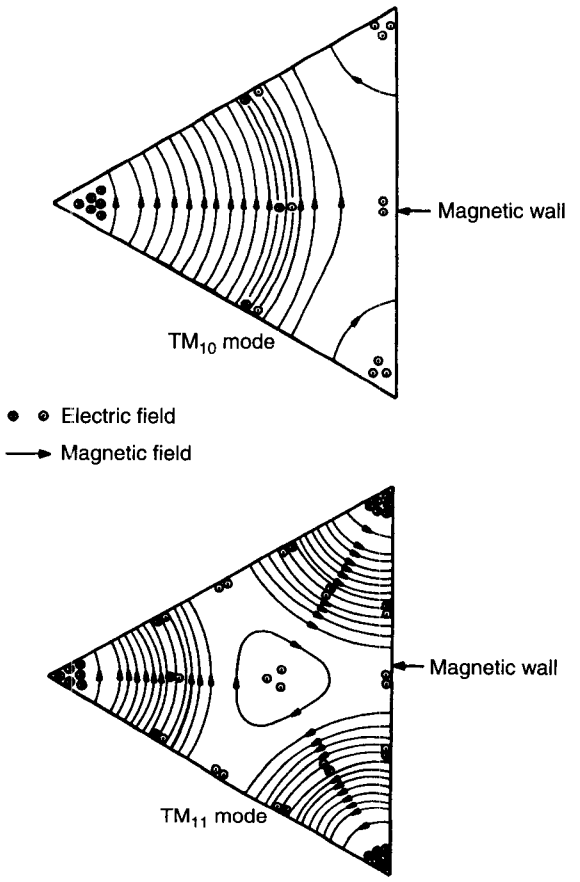


Figure 3.16 Field patterns for TM_{10} and TM_{11} modes in equilateral triangular microstrip resonator.

the spectral domain technique [29, 30]. The resonant frequency for an equilateral triangular microstrip resonator as a function of the triangle side is shown in Fig. 3.17. Those for an isosceles triangle as a function of apex angle (2α) for various values of triangle height are plotted in Fig. 3.18.

3.4.5 High-Q Resonators

Resonant structures with high Q are preferred in various applications involving voltage-controlled oscillators and filters. Consequently, planar transmission lines exhibiting lower dielectric, conductor, and radiation losses are preferred. Various transmission media have been proposed to realize lower loss. Whereas the valley microstrip line [31] and the micromachined suspended membrane transmission line [32–34] require special fabrication techniques, the high-

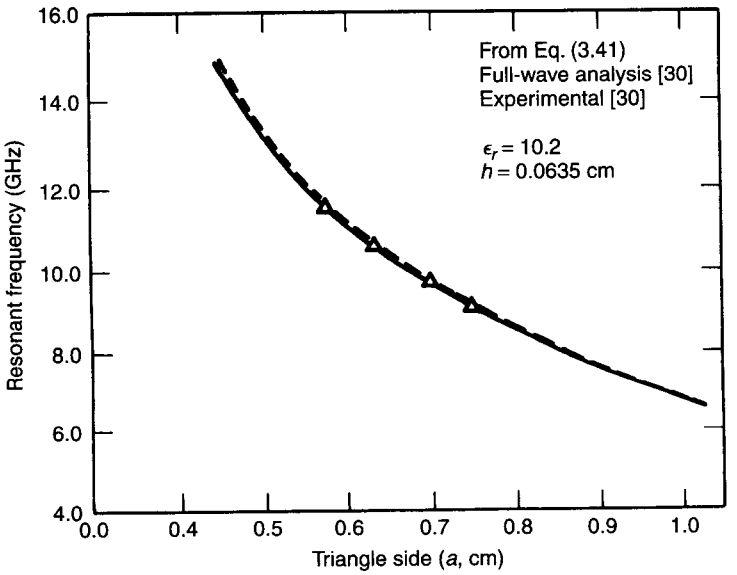


Figure 3.17 Resonant frequency as function of triangle side for equilateral triangular microstrip resonator.

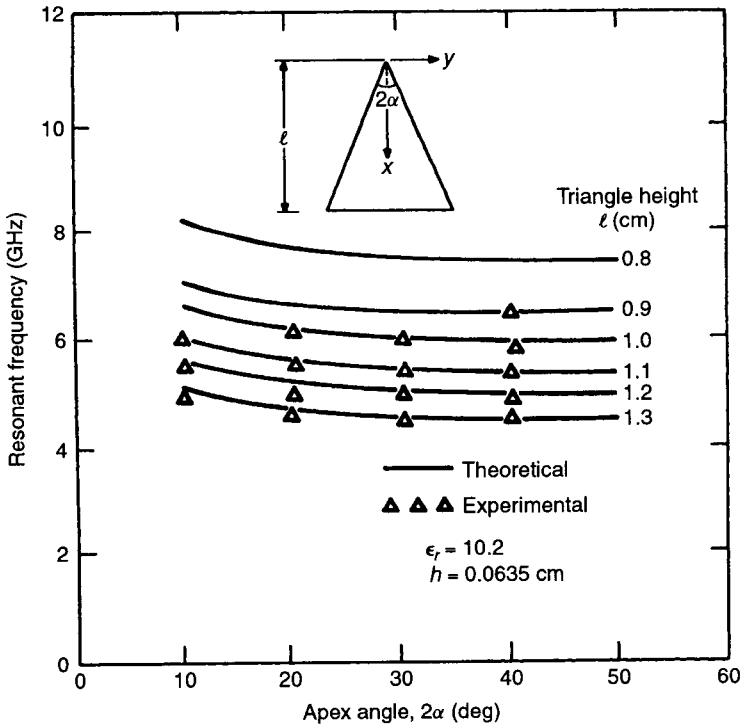


Figure 3.18 Resonant frequency as function of apex angle for isosceles triangular microstrip resonator.

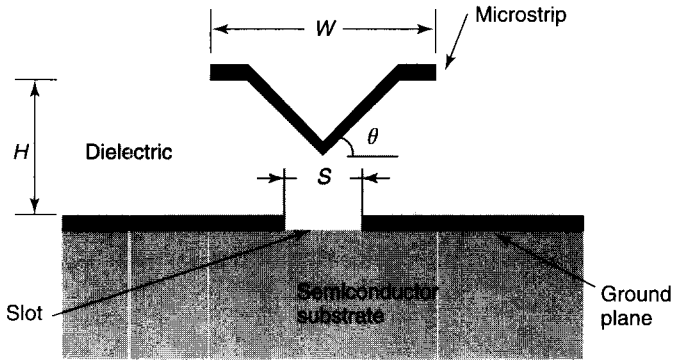


Figure 3.19 Valley microstrip.

impedance coplanar waveguide [35] and the multilayer substrate microstrip line [36] require simpler fabrication techniques that are compatible with monolithic fabrication processes.

The valley microstrip line [31], as the name suggests, consists of a microstrip conductor formed in the V-groove of the dielectric substrate and is shown in Fig. 3.19. This reduces the fringing field effects at the outer edges. Also, for a given characteristic impedance, the width of the microstrip conductor is larger than that of the conventional microstrip line, resulting in lower current density. Both of these factors help in lowering the conductor loss.

A coplanar waveguide realized by elevating the center conductor on a dielectric or forming an air-bridged center conductor results in higher characteristic impedance and lower effective dielectric constant. The conductor loss is about 0.1 dB/mm for a 97-Ω line for an air-bridged coplanar waveguide (Fig. 3.20), compared to the conventional coplanar line, which shows 0.5 dB/mm at 20 GHz [34].

The approach presented by Bahl et al. [36], shown in Fig. 3.21, provides a low-loss microstrip line by introducing low-dielectric-constant polyimide between the conductor and the GaAs substrate. It lowers the dissipation loss since the effective dielectric constant is lower.

Micromachined RF transmission lines and components fabricated as suspended membranes are especially suitable for low-resistivity substrates such as

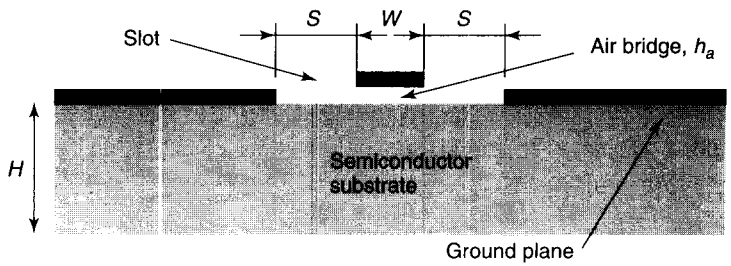


Figure 3.20 High-impedance coplanar waveguide.

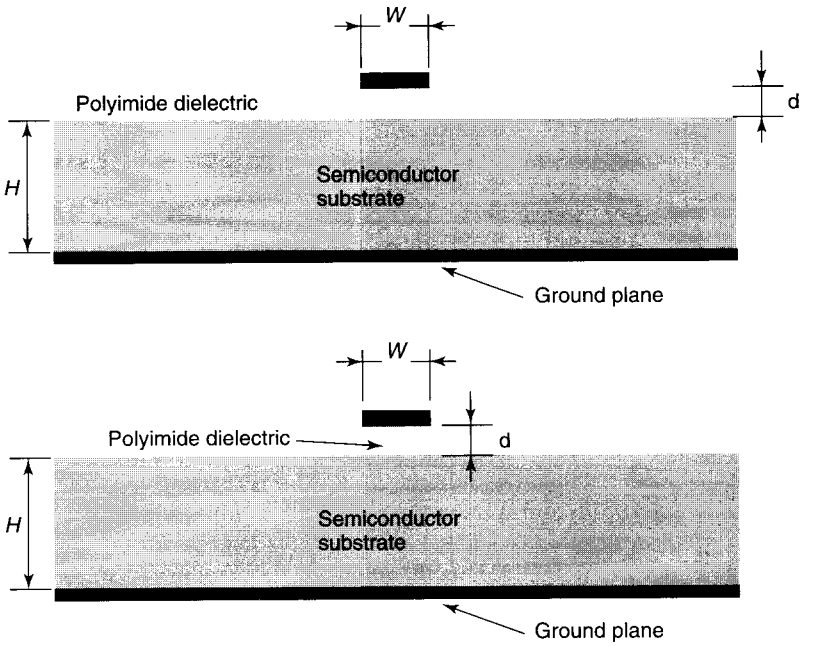
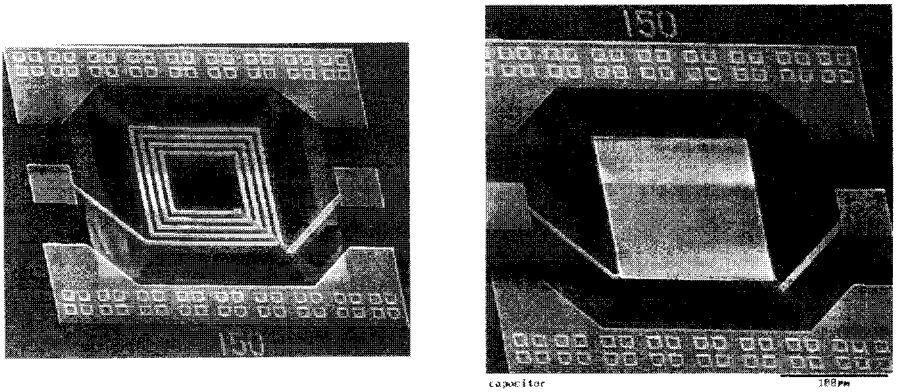


Figure 3.21 Multilayer microstrip line.

silicon at lower microwave frequencies where the excessive shunt capacitance lowers the resonant frequency and the Q of passive structures [32]. This approach can also be used at millimeter-wave frequencies to provide low-loss quasi-TEM operation [33]. A photograph of fabricated micromachined inductor and capacitor is shown in Fig. 3.22. Details of components realized using micromachined transmission lines are provided in Chapter 14.



(*a*) (*b*)

Figure 3.22 Micromachined inductor and capacitor elements. (After Sun et al. [32]. Reprinted with permission of John Wiley & Sons.)

In practice, any form of low dielectric and conductor loss transmission line can be used to form a resonant structure. Usually EM simulators are used to analyze these structures due to their nonconventional geometries.

3.4.6 Tunable Resonators

Resonant structures are basic building blocks in the realization of various components. The usefulness of these components can be extended tremendously if the resonant frequency of such structures can be varied using solid state devices such as PIN diodes, varactor or tunnel diodes, and field-effect or bipolar transistors. Such resonators can be used to realize switchable or tunable components such as filters, oscillators, amplifiers, and antennas.

Tuning Elements. Realizations of tunable resonators integrated with solid state devices are schematically shown in Fig. 3.23 for microstrip and CPW transmission lines commonly used in the realization of components [39]. Coplanar waveguides as well as coplanar strips are preferred since series and shunt configuration are easily realized. Typically, the variable capacitance can be achieved by changing the applied voltage across varactor diodes. The increase in the forward bias across the varactor results in narrowing of the depletion layer, which results in a higher capacitance value. There is a large forward leakage current, and increased forward resistance causes a lower Q value. Alternatively, increasing the reverse bias across the varactor causes the capacitance to decrease due to increase in the depletion layer. There is a small amount of leakage current, and the very low resistance value keeps the Q value high. Therefore, in practice, reverse-biased varactor diodes are preferred.

Tunable resonators can also be realized using variable inductors. They involve use of active devices such as field-effect or bipolar transitors.

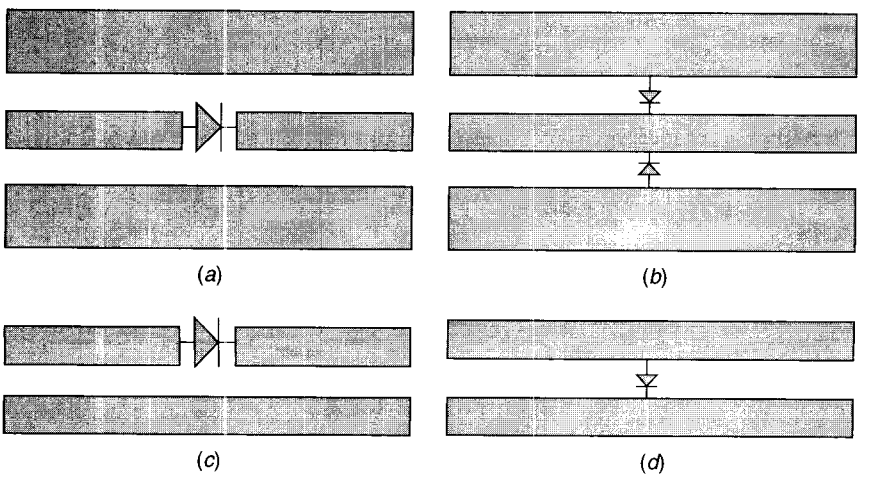


Figure 3.23 Transmission lines with solid state tuning elements.

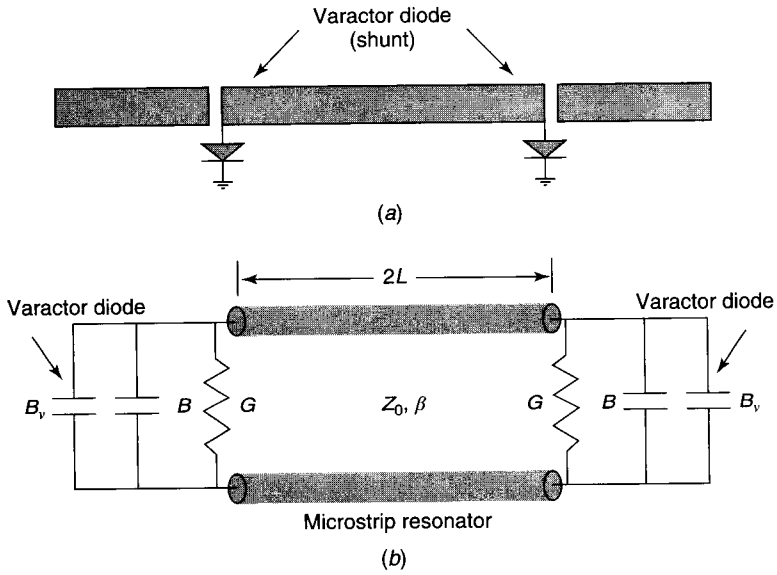


Figure 3.24 (*a*) Linear resonator with tuning elements. (*b*) Equivalent circuit of microstrip resonator with tuning varactor diodes.

Resonance Condition. In order to determine the resonant frequency of a resonator in the presence of tuning elements, the equivalent transmission-line structure must be analyzed. In case of a linear resonator, the capacitive loading provided by the varactor diodes increases its effective length and consequently reduces its resonant frequency. In the case of a microstrip ring with a series varactor diode, only the resonant modes that satisfy the necessary boundary conditions exist. The equivalent transmission-line simulation including the effect of the tuning elements is used to determine its resonant frequencies.

Linear Resonator. A linear resonator loaded with varactor diodes is schematically shown along with its equivalent circuit representation in Fig. 3.24. The diode admittances are connected with a transmission line of width W and length L with a characteristic impedance Z_0. The input admittance is given by [37]

$$Y_{\text{in}} = G + j(B + B_v) + Y_o \frac{G + j(B + B_v + Y_o \tan \beta L)}{Y_o + j[G + j(B + B_v)] \tan \beta L} \tag{3.44}$$

where $Y_o = 1/Z_0$; $\beta = 2\pi\epsilon_{\text{eff}}/\lambda_o$ is the phase constant, and ϵ_{eff} is the effective dielectric constant. In the above equation, the microstrip open end is described in terms of radiation conductance G and susceptance B. They are expressed as

$$G = \begin{cases} W^2/90\lambda_o^2 & \text{for } W \ll \lambda_o \quad (3.45a) \\ W^2/120\lambda_o^2 & \text{for } W \gg \lambda_o \quad (3.45b) \end{cases}$$

and $B = \beta \Delta l / Z_0$, where Δl is the equivalent line length due to fringing fields.

Since the imaginary part of the input impedance Y_{in} is zero at resonance, from (3.44)

$$\tan \beta L = \frac{2Y_o(B + B_v)}{(B + B_v)^2 + G^2 - Y_o^2} \tag{3.46}$$

or, alternatively, the resonant frequency f_r is given by

$$f_r = \frac{c}{2\pi L\sqrt{\epsilon_{\text{eff}}}} \tan^{-1}\left\{\frac{2Y_o(B + B_v)}{(B + B_v)^2 + G^2 - Y_o^2}\right\} \tag{3.47}$$

As can be seen from the above equation, when $B + B_v = 0$, the resonant frequency corresponds to that of a $\lambda/2$ resonator. For a given resonator length, the resonant frequency is reduced due to the presence of additional capacitive susceptance of the varactor diodes at its ends. By varying the capacitance, the resonant frequency is changed.

Ring Resonators. The microstrip ring resonator with a series tunable varactor diode is shown in Fig. 3.25. By changing the voltage across the varactor, a variable-capacitance value is realized. Increasing the capacitance increases the circumference of the ring resonator. In effect, the variable capacitance in a ring resonator results in a variable-circumference ring resonator. In the configura-

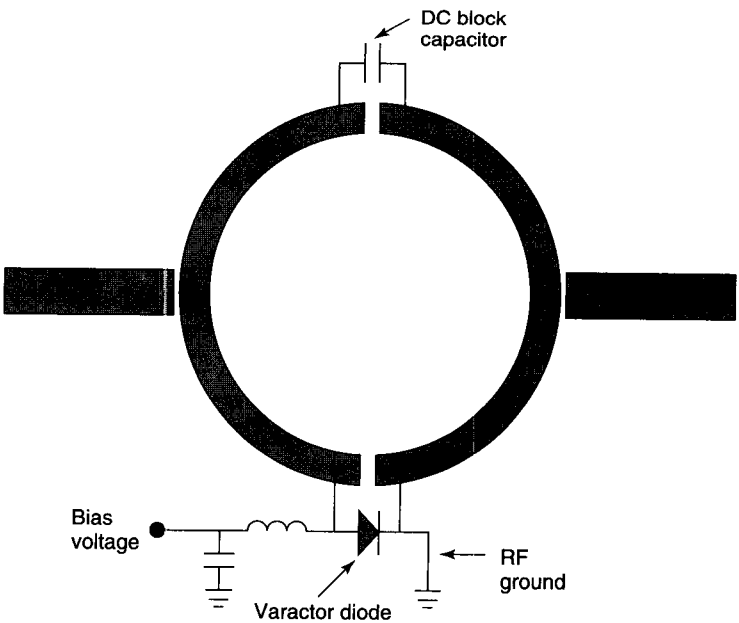


Figure 3.25 Ring resonator with tunable elements.

tion shown in Fig. 3.25, the series capacitor provides DC blocking and has a very low impedance in the frequency range of interest.

In general, it is difficult to analytically determine the resonant frequency of the structure due to the presence of the varactor diode in a ring resonator. Therefore, the equivalent transmission-line method is recommended [38]. The equivalent circuit of the tuning elements in various bias states can be used to determine the overall S-parameters of the ring resonator. Alternatively, one can use a commercially available circuit simulator for its computation. An accurate determination of the resonant frequency will require simulating the circuit with a very small frequency step. The resonant frequencies are then observable as peaks in S_{21} or S_{12}. Typical resonant frequencies are shown in Fig. 3.26*a* for a ring resonator and in Fig. 3.26*b* for a varactor-tuned ring resonator. A comparison of the modes shows that the even modes are not affected while the odd

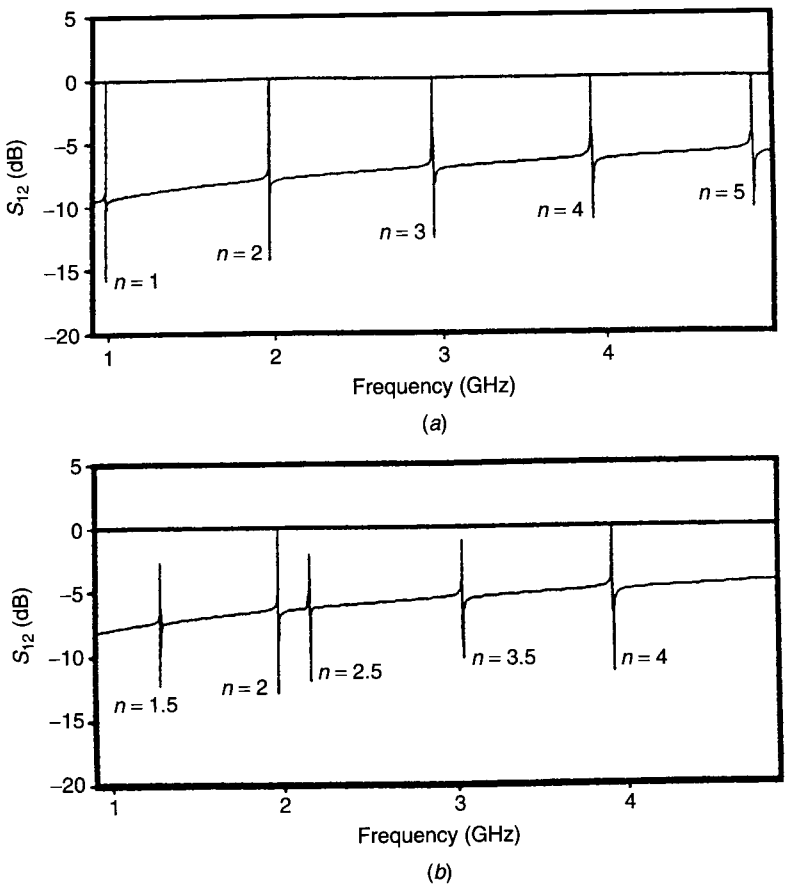


Figure 3.26 Typical frequency response of (*a*) ring and (*b*) varactor tuned ring resonator. (After Chung [38]. Reprinted with permission of John Wiley & Sons.)

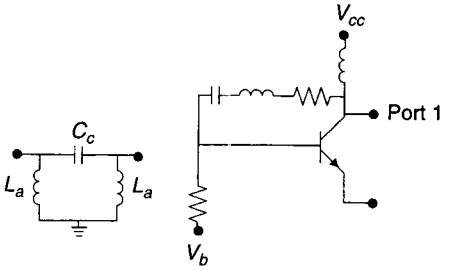


Figure 3.27 (*a*) Active resonator building block using active inductor. (*b*) Active inductor realized using bipolar transistor circuit.

modes vanish. Furthermore, the ring resonator now supports several "half-modes." By varying the voltage, the resonant frequencies of these half-modes can be changed [39]. Other resonator parameters are then easily determined by following their definitions presented earlier in the chapter.

Active Resonators. Recently, there has been considerable interest in MMICs fabricated using bipolar complementary metal–oxide–semiconductor (CMOS) processes. The easy availablity of this fabrication process has potential to provide small-size circuits at lower microwave frequencies. Unfortunately, the performance of passive elements is severely degraded on a lossy silicon substrate. These components typically have very low Q ($\simeq 5$) due to series resistances and parasitic inductances and capacitances.

In order to achieve high-Q performance of passive components, active devices such as bipolar junction or heterojunction bipolar transistors (BJTs, HBTs) are utilized. An active resonator, as shown in Fig. 3.27*a*, consists of active inductors realized using a feedback network around the active device. As an example, the feedback network as shown in Fig. 3.27*b* uses series R, L, and C. The input impedance looking into port 1 is inductive. The bias level is adjusted to produce low resistance and high inductance values over the frequency range of interest.

An active resonator is formed using the grounded active inductors described above with the MIM capacitors realized on a suspended membrane. The excess substrate capacitances may reduce the resonant frequency of the resonator. However, the bias level of the transistor can be adjusted to provide an optimum inductance value. The Q's of such active resonators are typically in the range of 75–100 [32].

Active resonators can be used in filters and oscillators. The bias tuning can be used to produce tunable filters and voltage-controlled oscillators.

3.5 DIELECTRIC RESONATORS

The dielectric resonator is made of a low-loss, temperature-stable, high-permittivity, and high-Q ceramic material in a regular geometric form. It reso-

nates in various modes at frequencies determined by its dimensions and shielding conditions. Because of its small size, low price, and excellent integrability in MICs, its application in active and passive microwave circuits has been rapidly increasing [40, 41].

The dimensions of a dielectric resonator are considerably smaller than those of an empty metallic cavity resonant at the same frequency, by a factor of approximately $1/\sqrt{\epsilon_r}$. If ϵ_r is high, the electric and magnetic fields are confined in and near the resonator, thus having small radiation losses. The unloaded quality factor Q_u is thus limited by the losses in the dielectric resonator. To a first approximation, a dielectric resonator is the dual of a metallic cavity. The radiation losses of the dielectric resonators with the commonly used permittivities, however, are generally much greater than the energy losses in the metallic cavities, which makes proper shielding of the dielectric resonator a necessity.

The shape of a dielectric resonator is usually a solid cylinder, but one can also find tubular, spherical, and parallelopiped shapes. A commonly used resonant mode in cylindrical resonators is denoted by $TE_{01\delta}$. The magnetic field lines are contained in the meridian plane, while the electric field lines are concentric circles around the z axis, as shown in Fig. 3.28. For a distant observer, this mode appears as a magnetic dipole, and for this reason sometimes this mode is referred to as a "magnetic dipole mode." When the relative dielectric constant is around 40, more than 95% of the stored electric energy of the $TE_{01\delta}$

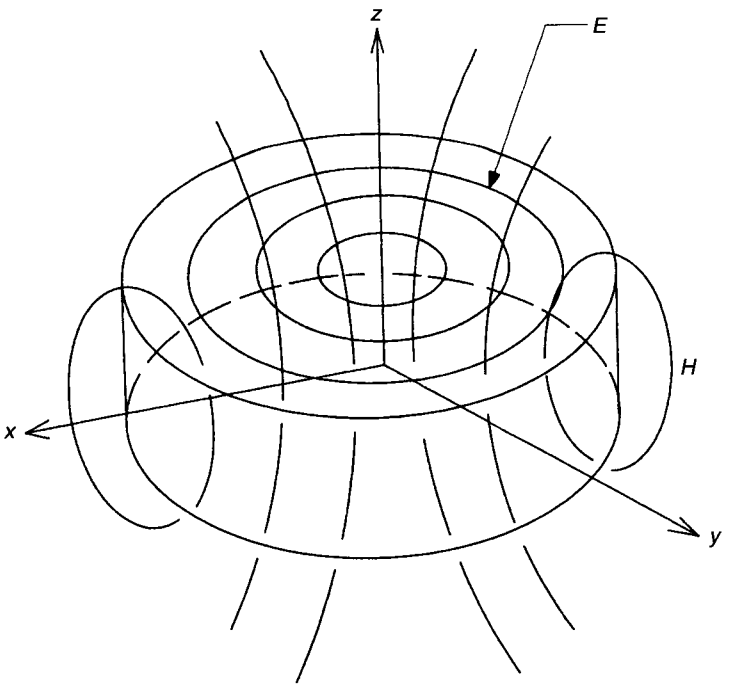


Figure 3.28 Field lines of resonant mode $TE_{01\delta}$ in isolated dielectric resonator.

mode as well as more than 60% of the stored magnetic energy is located within the cylinder. The remaining energy is distributed in the air around the resonator, decaying rapidly with distance away from the resonator surface.

3.5.1 Material

The important properties of the ceramic material to be used for dielectric resonators are

- the Q factor, which is equal to the inverse of the loss tangent;
- the temperature coefficient of the resonant frequency T_f, which includes the combined effects of the temperature coefficient of the dielectric constant and the thermal expansion of the dielectric; and
- the dielectric constant ϵ_r.

The Q, T_f, and ϵ_r values required for various applications differ, and in general proper combination can be achieved by choosing an appropriate material and composition. Until several years ago, the lack of suitable materials possessing Q, T_f, and ϵ_r all of acceptable values severely limited dielectric resonator applications. Materials such as TiO_2 (rutile phase), which has $Q \simeq 10{,}000$ at 4 GHz and $\epsilon_r \simeq 100$, were most often used for exploratory work. But TiO_2 has a T_f value of 400 ppm/°C, which makes it impractical for most of the applications.

A number of material compositions have been explored in attempts to develop suitable dielectric materials. These include ceramic mixtures containing TiO_2 [42], various titanates and zirconates [43, 44], glass ceramic [45], and alumina-based ceramics. Development of temperature-stable dielectric resonators dates back to the late 1970s. At present, several ceramic compositions have been developed offering excellent dielectric properties. Table 3.2 compares the important properties of different materials developed commercially.

It is not yet established if any of the dielectric compositions shown in Table 3.2 has overall superiority over the others, since many factors, such as ease of ceramic processing and ability to hold tolerances on the dielectric properties, must be considered. The performance limitations, if any, of the lower dielectric constant material remain to be determined, since most component work reported thus far has used dielectric resonators possessing ϵ_r in the range of 37–100. The lower dielectric constant material performance is likely to be more sensitive to shielding due to increase in fields outside the isolated resonator.

The temperature coefficient T_f of the resonator can be controlled in some materials, by modifying the composition, to be anywhere within +9 to −9 ppm/°C. Circuit effects also shift T_f by a few parts per million, depending upon the circuit configuration. An initial resonator T_f of +1 to +4 ppm/°C often

Table 3.2 Properties of Dielectric Resonators

Material Composition	Dielectric Constant	Loss Tangent at 4 GHz	Temperature Coefficient, ppm/°C	Manufacturer
$Ba_2Ti_9O_{20}$	40	1×10^{-4}	+2	Bell Labs
$BaTi_4O_9$	39	1×10^{-4}	+4	Raytheon, Transtech
$(Zr–Sn)TiO_4$	38	1×10^{-4}	−4–10 adjustable	Transtech, Murata, Siemens, Tekelec, NTK
$Ba(Zn_{1/3}Nb_{2/3})O_2$ $Ba(Zn_{1/3}Ta_{2/3})O_2$	30	4×10^{-5}	0–10 adjustable	Murata, Panasonic
$BaO–PbO–Nd_2O_3–TiO_2$	90	2×10^{-4} at 1 GHz	+10 to −10 adjustable	Murata, Transtech
$MgTiO_3–CaTiO_3$	21	1×10^{-4}	+10 to −10 adjustable	Murata

results in effective temperature compensation in transistor oscillators. Thus, the limit to achievable temperature compensation results from ceramic tolerances and the usual need for frequency tuning affecting the temperature effects.

The quality factor Q of the dielectric resonator decreases with the increase in frequency. Typically the product f_0 (in gigahertz) $\times$ Q_0 is constant. Some degradation of Q is usually incurred in component applications. Losses due to housing walls, dielectrics, and adhesives used to support the resonators, and other effects typically reduce Q by 10–20% as described later. Variations in the Q's of different materials being significant (Table 3.2), it is possible to select the right material for different applications. A higher Q material is preferred for lower noise oscillations as well as for sharp tuned filters, and lower Q for frequency-tunable wider band components.

3.5.2 Resonant Frequency

The resonant frequency of a resonator is determined by its dimensions and surroundings. Although the geometric form of a dielectric resonator is extremely simple, an exact solution of the Maxwell equations is considerably more difficult than for the hollow metallic cavity. For this reason, the exact resonant frequency of a certain resonant mode, such as $TE_{01\delta}$, can only be computed by rigorous numerical procedures. A number of theories on the subject are available in the literature that have an accuracy of $\pm 1\%$ for the given configuration. Unfortunately these methods call for the use of high-level computers. Kajfez and Guillon [46] have presented an approximate solution of the equations involved, both for the isolated case and for the more commonly used MIC configuration. These equations, given below, can be easily programmed on a desktop computer and give accuracies of better than $\pm 2\%$.

1. *Isolated Dielectric Resonator.* The resonant frequency is given in gigahertz by

$$f_r = \frac{34}{a\sqrt{\epsilon_r}}\left(\frac{a}{H} + 3.45\right) \tag{3.48}$$

where a represents the radius of the resonator in millimeters and H the height. This relation is accurate to about 2% in the range

$$0.5 < \frac{a}{H} < 2 \quad \text{and} \quad 30 < \epsilon_r < 50 \tag{3.49}$$

2. *Dielectric Resonator in MIC Configuration.* Figure 3.29 (inset) represents a dielectric resonator (DR) in a MIC configuration. The steps to determine the size of the dielectric resonator for a given frequency are as follows:

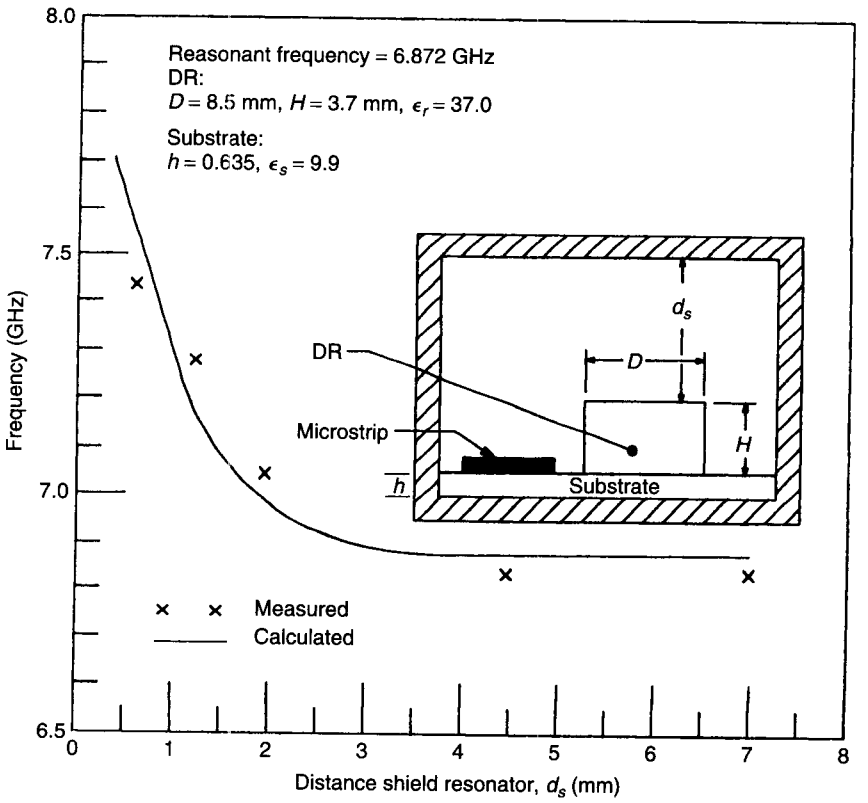


Figure 3.29 Resonant frequency as function of cover height.

(a) The diameter ($D = 2a$) of the resonator is selected to be

$$\frac{5.4}{k_0\sqrt{\epsilon_s}} > 2a > \frac{5.4}{k_0\sqrt{\epsilon_r}} \tag{3.50}$$

where ϵ_s and ϵ_r are the relative permittivities of the substrate and the resonator, respectively, and k_0 is the free-space propagation constant.

(b) Calculate k' from

$$k' = \frac{2.405}{a} + \frac{Y_0}{2.405a[1 + (2.43/Y_0) + 0.291\,Y_0]} \tag{3.51}$$

where

$$Y_0 = \sqrt{(k_0 a)^2(\epsilon_r - 1) - 2.405^2} \tag{3.52}$$

(c) Calculate propagation constant β for the $TE_{01\delta}$ mode as

$$\beta = \sqrt{k_0^2\epsilon_r - k'^2} \tag{3.53}$$

(d) Evaluate the attenuation constants α_1 and α_2:

$$\alpha_1 = \sqrt{k'^2 - k_0^2\epsilon_s} \qquad \alpha_2 = \sqrt{k'^2 - k_0^2} \tag{3.54}$$

(e) Find the resonator height H from

$$H = \frac{1}{\beta}\left[\tan^{-1}\left(\frac{\alpha_1}{\beta}\coth(\alpha_1 h)\right) + \tan^{-1}\left(\frac{\alpha_2}{\beta}\coth(\alpha_2 d_s)\right)\right] \tag{3.55}$$

The preceding relations can also be used to determine the frequency of a dielectric resonator with known dimensions. Figure 3.29 presents the variation of resonant frequency with the cover height. The measured points show that the accuracy of this method is within $\pm 2\%$.

3.5.3 Coupling of a Dielectric Resonator in MIC Configuration

The dielectric resonator is used in a number of different configurations depending upon the application. In order to effectively use dielectric resonators in microwave circuits, it is necessary to have an accurate knowledge of the coupling between the resonator and different transmission lines. The $TE_{01\delta}$ mode of the cylindrical resonator can be easily coupled to microstrip-line, fin-line, magnetic-loop, metallic, and dielectric waveguides [46]. In this section, we discuss the most commonly used configuration of the dielectric resonator, that is, $TE_{01\delta}$ mode coupling with a microstrip line.

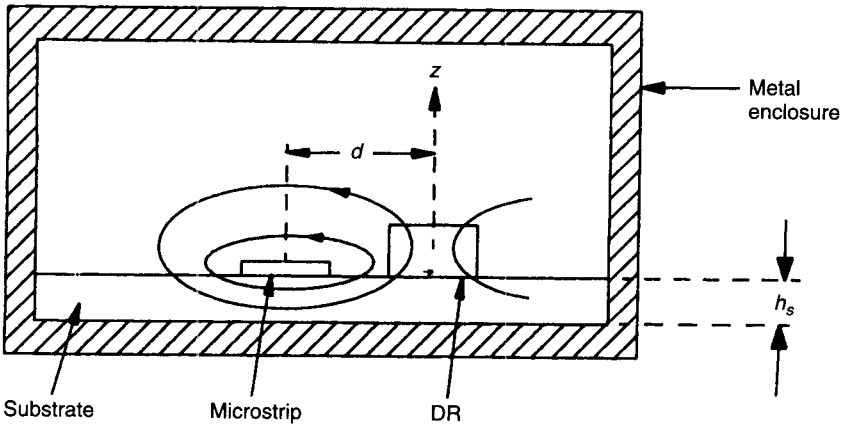


Figure 3.30 Coupling between microstrip line and dielectric resonator.

Figure 3.30 shows the magnetic coupling between a dielectric resonator and microstrip. The resonator is placed on the top of the microstrip substrate. The lateral distance between the resonator and the microstrip conductor primarily determines the amount of coupling between the resonator and the microstrip transmission line. Proper metallic shielding required to minimize the radiation losses (hence to increase Q) also affects the resonant frequency of the $TE_{01\delta}$ mode. The reason for the modification of the resonant frequency can be explained by the cavity perturbation theory. Namely, when a metal wall of a resonant cavity is moved inward, the resonant frequency will decrease if the stored energy of the displaced field is predominately electric. Otherwise, when the stored energy close to the metal wall is mostly magnetic, as in the case of the shielded $TE_{01\delta}$ dielectric resonator considered here, the resonant frequency will increase when the wall moves inward.

The $TE_{01\delta}$ mode in a dielectric resonator can be approximated by a magnetic dipole of moment M. The coupling between the line and the resonator is accomplished by orienting the magnetic moment of the resonator perpendicular to the microstrip plane so that the magnetic lines of the resonator link with those of the microstrip line, as shown in Fig. 3.30. The dielectric resonator placed adjacent to the microstrip line operates like a reaction cavity that reflects the RF energy at the resonant frequency [47]. The equivalent circuit of the resonator coupled to a microstrip line is shown in Fig. 3.31. In this figure L_r, C_r, and R_r are the equivalent parameters of the dielectric resonator; L_1, C_1, and R_1 are the equivalent parameters of the microstrip line; and L_m characterizes the magnetic coupling. The transformed resonator impedance Z in series with the transmission line is easily determined to be

$$Z = j\omega L_1 + \frac{\omega^2 L_m^2}{R_r + j\omega(L_r - 1/\omega^2 C_r)} \tag{3.56}$$

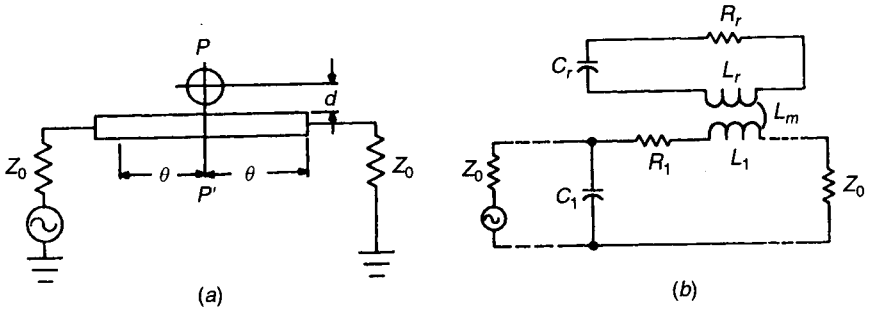


Figure 3.31 Equivalent circuit of dielectric resonator coupled with line.

Around the center frequency, ωL_1 can be neglected and Z becomes

$$Z = \omega Q_u \frac{L_m^2}{L_r} \frac{1}{1 + jX} \tag{3.57}$$

where $X = 2Q_u(\Delta\omega/\omega)$, and the unloaded Q and the resonant frequency of the resonator are given by

$$Q_u = \frac{\omega_0 L_r}{R_r} \tag{3.58a}$$

$$\omega_0 = \frac{1}{\sqrt{L_r C_r}} \tag{3.58b}$$

At the resonance frequency, $X = 0$ and

$$Z = R = \omega_0 Q_u \frac{L_m^2}{L_r} \tag{3.59}$$

Equation (3.59) indicates that the circuit shown in Fig. 3.31 can be represented by the simple parallel tuned circuit shown in Fig. 3.32, where L, R, C satisfy the following equations:

$$L = \frac{L_m^2}{L_r} \qquad C = \frac{L_r}{\omega_0^2 L_m^2} \qquad R = \omega_0 Q_u \frac{L_m^2}{L_r} \tag{3.60}$$

The coupling coefficient β at the resonant frequency ω_0 is defined by

$$\beta = \frac{R}{R_{\text{ext}}} = \frac{R}{2Z_0} = \frac{\omega_0 Q_u}{2Z_0} \frac{L_m^2}{L_r} \tag{3.61}$$

If S_{110} and S_{210} are defined as the reflection and transmission coefficients of the

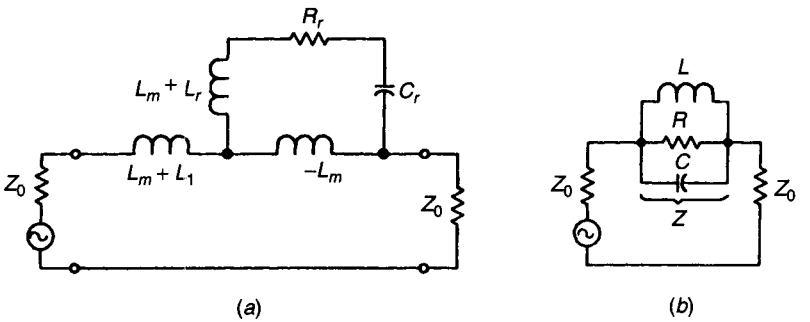


Figure 3.32 (*a*) Simplified equivalent circuit. (*b*) Final equivalent of dielectric resonator coupled with microstrip line.

resonance frequency of the resonator coupled to the microstrip, β can be shown to be given by [48]

$$\beta = \frac{S_{110}}{1 - S_{110}} = \frac{1 - S_{210}}{S_{210}} = \frac{S_{110}}{S_{210}} \tag{3.62}$$

This relation can be used to determine the coupling coefficient from the directly measurable reflection and transmission coefficients. The value of β can also be accurately calculated from a knowledge of the circuit configuration. The quantity L_m^2/L_r in (3.61) is a strong function of the distance between the resonator and the microstrip line for given shielding conditions and substrate thickness and permittivity. The analysis of β involves the use of known EM concepts and finite-element techniques.

The relation between different quality factors is well known and given by

$$Q_u = Q_L(1 + \beta) = Q_e\beta \tag{3.63}$$

The external quality factor Q_e $(=Q_u/\beta)$ is generally used to characterize the coupling. Figure 3.33 shows an example of the variation of Q_e with the distance between the resonator and the line.

The S parameters of the dielectric resonator coupled to a microstrip with the lengths of transmission lines on input and output, as shown in Fig. 3.31, can be determined from the relations previously presented and given by [46]

$$S = \begin{bmatrix} \dfrac{\beta}{\beta + 1 + jQ_u\,\Delta\omega/\omega_0} e^{-2j\theta} & \dfrac{1 + jQ_u\,\Delta\omega/\omega_0}{\beta + 1 + jQ_u\,\Delta\omega/\omega_0} e^{-2j\theta} \\ \dfrac{1 + jQ_u\,\Delta\omega/\omega_0}{\beta + 1 + jQ_u\,\Delta\omega/\omega_0} e^{-2j\theta} & \dfrac{\beta}{\beta + 1 + jQ_u\,\Delta\omega/\omega_0} e^{-2j\theta} \end{bmatrix} \tag{3.64}$$

where 2θ is the electrical line length between the input and output planes.

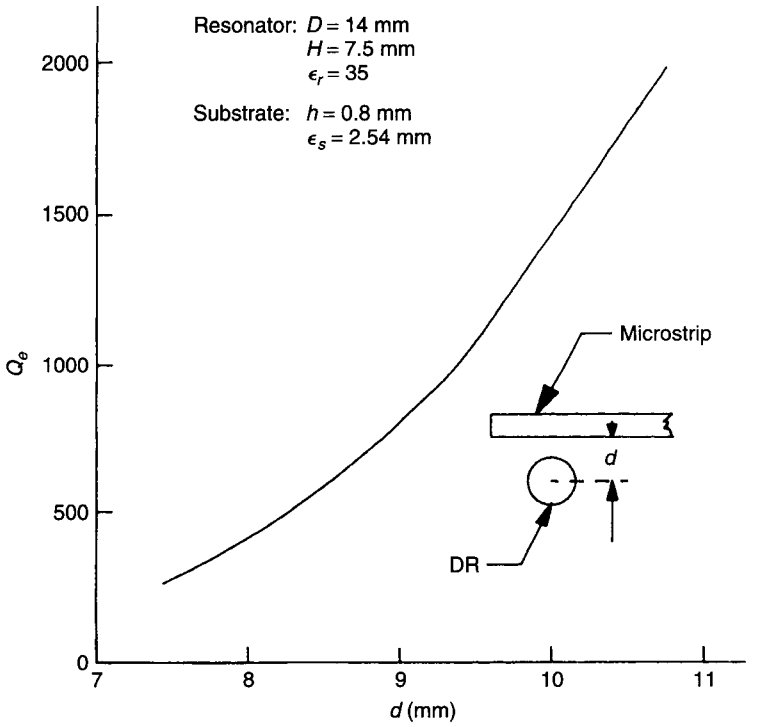


Figure 3.33 External Q factor as function of distance between line and resonator. (After Kajfez and Guillon [46]. Reprinted with permission of Artech House.)

3.5.4 Spurious Modes

As in the case of all resonant cavities, there are many possible resonant modes that can be excited in dielectric resonators. These modes can be divided into three families: transverse-electric (TE), transverse-magnetic (TM), and hybrid electromagnetic (HEM) modes. Each of the three families have a large number of individual modes, so that one encounters a dilemma in choosing which mode is best suited for a particular application. The $TE_{01\delta}$ is the principal or main mode traditionally used, but for certain applications, such as for a dual-mode filter, the $HEM_{11\delta}$ mode has definite advantages.

The resonant frequency of the principal and spurious modes is determined by the physical dimensions and the dielectric constant of the resonator for fixed shielding conditions. The resonant frequency of some spurious modes, like $TM_{01\delta}$ and $HEM_{11\delta}$, may be close to the resonant frequency of the principal $TE_{01\delta}$ mode. Figure 3.34 shows a mode chart of the dielectric resonator in the shown shielded configuration [49]. The aspect ratio D/H of the dielectric resonator can be used as a design parameter to place the resonance of spurious

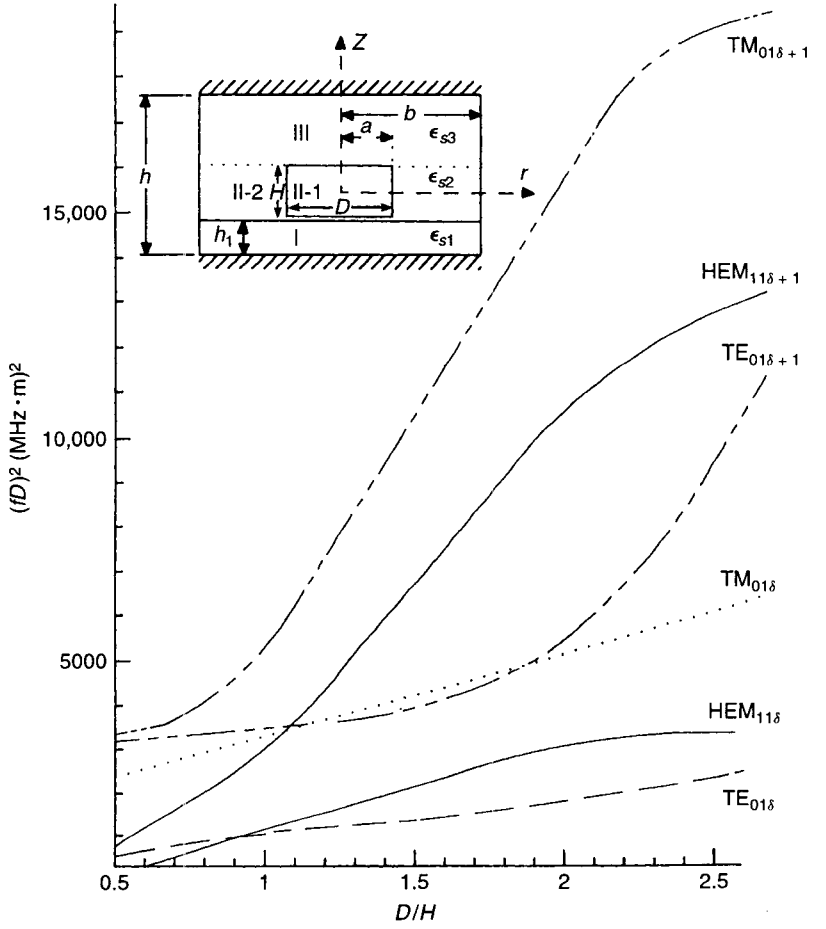


Figure 3.34 Mode chart of dielectric resonator in cavity: $\epsilon_r = 37.2$, $\epsilon_{s1} = \epsilon_{s2} = \epsilon_{s3} = 1.0$, $b = 16$ mm, $h = 21$ mm, $H = 8$ mm, and $h_1 = 1$ mm.

modes outside the operating frequency band of the principal $TE_{01\delta}$ mode, thus minimizing undesirable interference. As is clear from Fig. 3.34, an aspect ratio between 2 and 2.5 results in the best separation of the spurious modes.

The $TM_{01\delta}$ and $HEM_{mn\delta}$ modes are far more sensitive to frequency-tuning screws than the principal mode. For example, the same amount of tuning that will tune the principal mode of a 6-GHz resonator by 25 MHz can move the resonant frequency of the $TM_{01\delta}$ or $HEM_{01\delta}$ in the opposite direction across the entire 500-MHz radio band. Frequency tuning must be limited to avoid a significant reduction of mode separation, which can bring the spurious response into the frequency band of interest.

3.5.5 Frequency Tuning

Most applications of the dielectric resonator demand frequency tunability over a narrow band. The resonant frequency of the dielectric resonator can be tuned with the help of a tuning screw placed directly above the resonator, in the metallic shield, as shown in Fig. 3.35. The resonant frequency being sensitive to the distance between the resonator and the shield, as explained earlier, the resonant frequency increases with the tuning-screw depth. This effect can be calculated using the finite-element method [50]. Figure 3.35 also shows the possible frequency tuning for several tuning-screw diameters. The curves presented in this figure allow one to choose an optimum screw diameter to obtain the desired tuning margin.

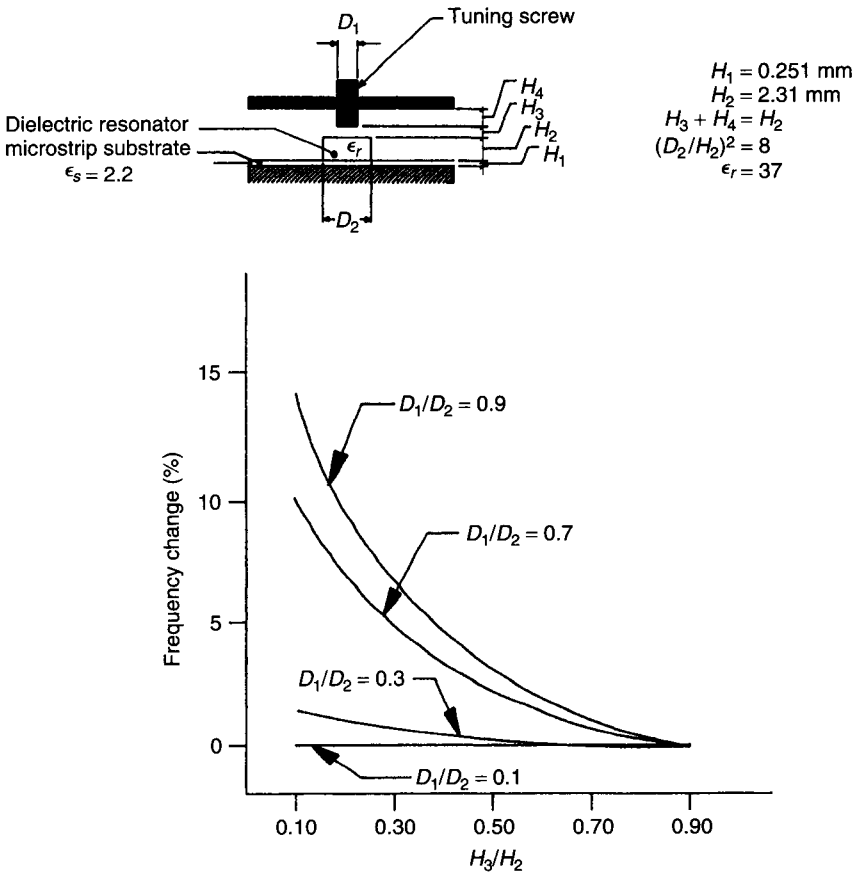


Figure 3.35 Frequency change of dielectric resonator for several tuning-screw diameters. (After Gil et al. [56]. Reprinted with permission of IEEE.)

3.6 YIG RESONATORS

The YIG resonator is a high-Q, ferrite resonator that can be tuned over a wide band by varying the biasing DC magnetic field. Its high performance and convenient size for applications in MICs make it an excellent choice in a large number of commercial and military applications, such as filters, oscillators, frequency multipliers, discriminators, and limiters. A YIG resonator makes use of the ferromagnetic resonance that occurs when a small magnetic microwave field is applied perpendicular to a static magnetic field. Depending on the material composition, size, and applied field, resonant frequencies between 500 MHz and 50 GHz can be achieved [51]. Single-crystal YIG ($Y_3Fe_5O_{11}$) and gallium-doped YIG are part of the family that resonates at microwave frequencies when immersed in a magnetic field. This resonance is directly proportional to the applied magnetic field, and linear tuning can be achieved by changing the magnetic field with an electric current. The resonator consists of a YIG sphere, an electromagnet, and a coupling loop. A typical YIG resonator in a MIC configuration is shown in Fig. 3.36. The YIG tuning magnets are electromagnets with a single air gap. The DC current through the two series-connected main coils provides the required DC magnetic field across the gap.

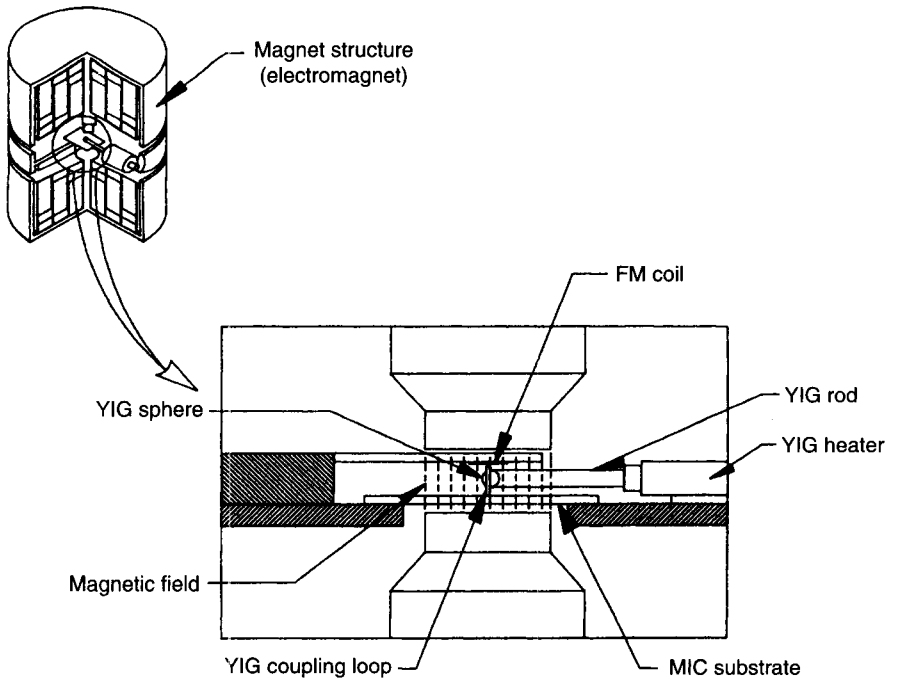


Figure 3.36 YIG oscillator elements. (After Osbrink [51]. Reprinted with permission of MSN & Communications Technology.)

The YIG sphere, frequency-modulated (FM) coil, and MIC substrate are placed between the poles of the electromagnet.

3.6.1 Resonant Frequency and Quality Factor

Microwave resonance occurs when the frequency of the RF magnetic field coincides with the natural electron dipole precessional frequency of the material. For spherical resonators, this resonant frequency is given by

$$f_0 = \Upsilon(H_0 \pm H_a) \tag{3.65}$$

where H_0 is the applied biasing field measured in oersteds, H_a is an internal field within the crystal known as the anisotropic field, and Υ is the charge–mass ratio of an electron and has a value of 2.8 MHz/Oe.

The upper frequency of YIG resonance is limited by the available magnetic field, and the lower frequency limit of operation is directly proportional to the value of its saturation magnetization ($4\pi M_s$), which is a measure for net density of precessing electron spins in the material. The externally applied tuning field must be sufficient to produce alignment of all magnetic dipoles within the crystal. A pure YIG material has a $4\pi M_s$ value of 1780 G at room temperature. A YIG of substantially lower saturation magnetization value (typically 250 G) can be grown by doping the crystal with gallium. However, doping increases YIG resonator losses and can be expressed by the linewidth parameter (ΔH).

The linewidth, expressed in oersteds, is a direct measure of the unloaded Q_u of the YIG resonator and can be compared to the 3-dB bandwidth of an unloaded cavity. Linewidth is related to Q_u and M_s by the following relation [52]:

$$Q_u = \frac{H_0 - 4\pi M_s/3}{\Delta H} \tag{3.66}$$

Where H_0 is the DC biasing field strength at resonance. Typical values of linewidths are less than 0.1 Oe for pure YIG to 1.5 Oe for doped 400 G YIG. Using (3.65)

$$Q_u = \frac{f_0 - \Upsilon 4\pi M_s/3}{\Delta H \Upsilon} \tag{3.67}$$

This expression demonstrates an interesting phenomenon in YIG, that is, the unloaded Q of the resonator decreases with the decrease in frequency and the cutoff frequency f_c, where $Q_u = 0$ is given by

$$f_c = \tfrac{1}{3}\Upsilon 4\pi M_s \tag{3.68}$$

For a pure YIG, $f_c = 1670$ MHz.

The resonant frequency of the YIG sphere as well as its frequency drift over temperature depends strongly upon its orientation. Orientation is commonly adjusted to compensate for the change in frequency caused by variation of the EM gap with temperature.

3.6.2 Coupling and Equivalent Circuit

The RF coupling between the YIG and the microwave circuit is realized by a loop around the sphere, as shown in Fig. 3.37*a*. Magnetostatic mode 110 of the resonator is thus magnetically coupled to the transmission line. The equivalent circuit of the loop-coupled YIG is shown in Fig. 3.37*b*. The parallel resonant circuit is induced in series with the loop impedance by coupling of the YIG sphere. Here L_L and R_L are the inductance and resistance, respectively, of the coupling loop.

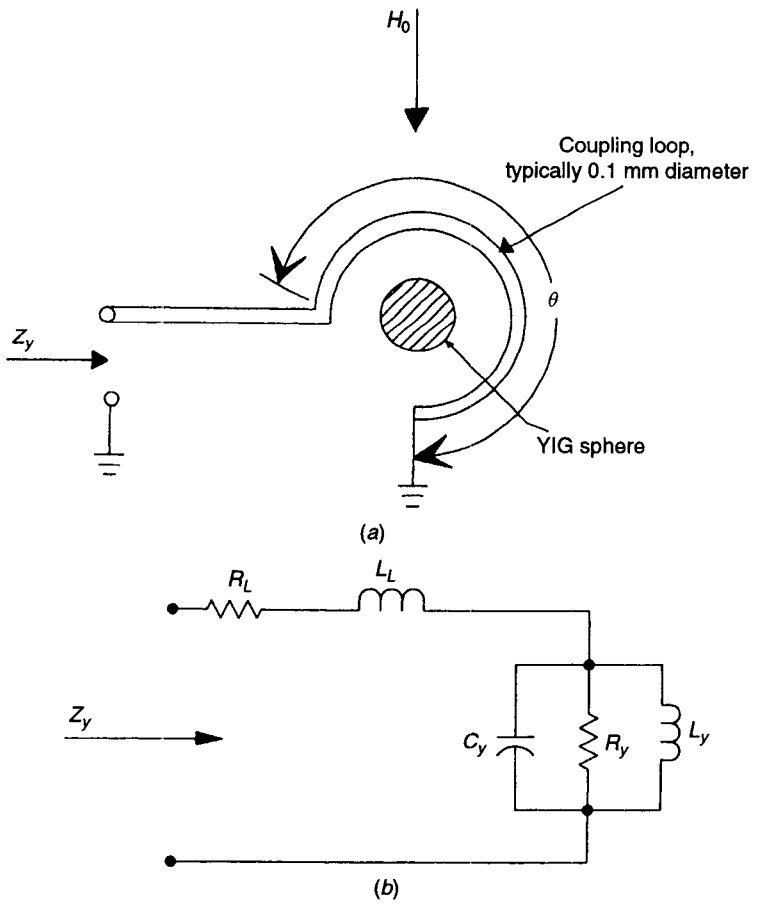


Figure 3.37 (*a*) Loop coupling circuit. (*b*) Equivalent circuit of loop-coupled YIG resonator.

The YIG sphere is located at the center of the loop of wire. The coupling coefficient is a direct function of the diameter of the sphere, saturation magnetization $4\pi M_s$ and loop angle θ and an inverse function of the loop diameter. In order to achieve wide-band tuning, the YIG sphere needs to be tightly coupled. The diameter of the YIG sphere is limited by high-frequency spurious modes. Typically a diameter between 12 and 35 mils is used for most applications. A high value of $4\pi M_s$ is desired. Pure YIG, which has $4\pi M_s$ of 1780 G, possesses a spurious mode in the S/C band. The frequency of this spurious mode is decreased to lower than the useful frequency by doping with gallium at the cost of reduced $4\pi M_s$. Reduced loop diameter can help increase coupling and hence bandwidth but can cause surface spurious modes to appear.

3.6.3 Magnetic Circuit

Design of the magnetic circuit forms an important aspect of any YIG device design due to the fact that it is the value of the magnetic field that determines the resonance frequency of the YIG resonator. Frequency and linearity accuracy of the YIG device depends directly on the magnetic circuit design. Important characteristics of the magnetic circuit include air gap, pole piece diameter, and number of coil turns. These elements are optimized for necessary performance requirements. Another important requirement of the magnetic circuit is minimum hysteresis, which represents a difference in resonance frequency, for the same current, between up and down current sweep. High-nickel magnetic steels are commonly used for their high saturation magnetic field and low hysteresis after proper thermal treatment.

The YIG magnetic tuning circuits dissipate large DC power. Typically it is about 5 W at 12 GHz and the tuning sensitivity lies between 15 and 25 MHz/mA. The latter implies that the current supply must have very low noise and ripple to minimize FM noise in oscillators. As an example, if the inherent spectral linewidth is expected to be better than 10 kHz, the magnet power ripple should not exceed 0.5 μA. Due to the large value of the inductance of the main tuning coil, the modulation bandwidth in oscillator applications is of the order of a few kilohertz, which is not sufficient for most commercial or military applications. A small coil called an FM coil is generally placed on the pole piece. This coil offers a small frequency deviation, but a modulation bandwidth of up to a few megahertz can be achieved.

In addition to the standard magnetic structure, permanent magnet structures are also used in some applications. Permanent magnets require smaller DC power to tune the resonator to the low end or middle of its tuning range, and electronic sweep adds or substracts a magnetic field sufficient to tune the entire range of the resonator. This configuration is commonly used for narrow-band applications offering significant reduction of size and power dissipation. A permanent magnet also provides considerably faster switching speed, and the low magnet main coil dissipation, especially for less than one octave coverage, makes it suitable for easier temperature stabilization.

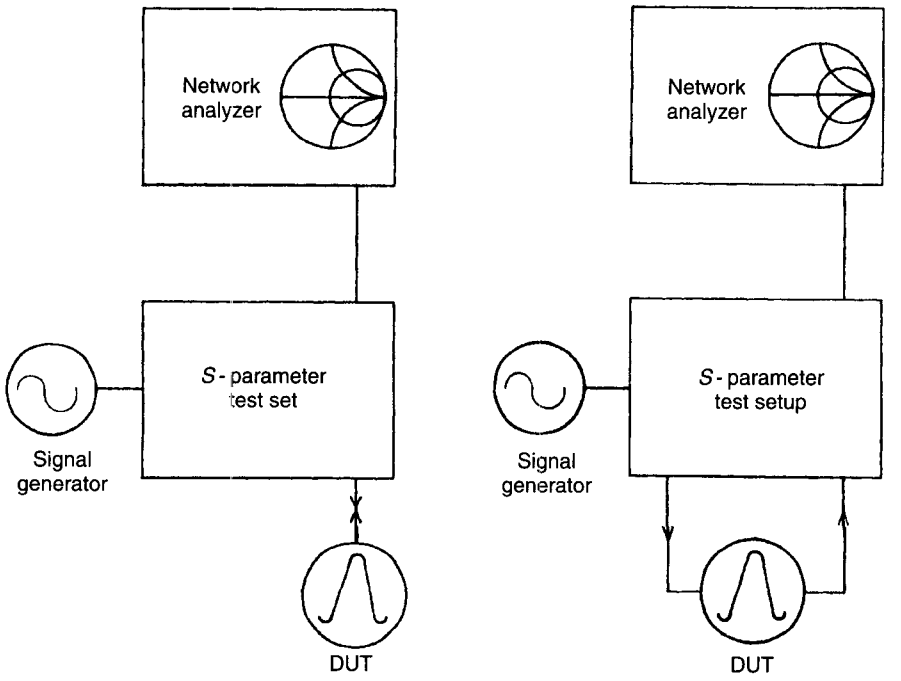


Figure 3.38 Measurement setups for (*a*) reflection and (*b*) transmission resonators.

3.7 RESONATOR MEASUREMENTS

Accurate characterization of microwave resonators is essential for their effective use. The important parameters that are required to fully describe a resonator for a given mode are the resonant frequency f_0, the coupling coefficient, and the quality factors Q_u (unloaded Q), Q_L (loaded Q), and Q_e (external Q).

Figure 3.38 shows an experimental setup for the measurement of resonator parameters using a commonly available network analyzer. The network analyzer displays the magnitude and phase of the reflection and transmission coefficients as required for the single- or two-port resonators. Many methods for the Q measurement are possible, but we will describe here only one simple technique using Q loci on the Smith chart.

3.7.1 Single-Port Resonator

The single-ended resonator is the most commonly used configuration among microwave resonant circuits. The equivalent circuits of the two possible configurations are shown in Fig. 3.1, where R, L, and C are the equivalent lumped resistance, inductance, and capacitance. The parallel tuned circuit of Fig. 3.1*a* is known as the detuned short configuration, and the series-tuned circuit of

Table 3.3 Important Parameters of a Resonant Circuit

Parameter	Series Tuned	Parallel Tuned
f_0	$\frac{1}{\sqrt{LC}}$	$\frac{1}{\sqrt{LC}}$
Q_u	$\frac{\omega L}{R}$	$\frac{R}{\omega L}$
β	$\frac{Z_0}{R}$	$\frac{R}{Z_0}$
Q_L	$\frac{Q_u}{1+\beta}$	$\frac{Q_u}{1+\beta}$

Fig. 3.1*b* is known as the detuned open configuration. As shown, either configuration can be converted to the other by displacing the reference plane by a quarter wavelength. The important parameters of these resonant circuits are defined in Table 3.3.

Further analysis of the resonant circuits being similar for the two configurations, we restrict our discussion to only the parallel tuned circuit. The input impedance of the circuit in Fig. 3.1*a* can be written as

$$\frac{1}{Z_{\text{in}}} = \frac{1}{R} + \frac{1}{j\omega L} + j\omega C \tag{3.69}$$

or

$$Z_{\text{in}} = \frac{R}{1 + 2jQ_u\delta} \tag{3.70}$$

where $\delta = (\omega - \omega_0)/\omega_0$ represents the frequency-detuning parameter.

The locus of the impedance, using (3.70), can be drawn by varying δ [53]. As impedance is a linear function of frequency, a circular locus will be produced when plotted on the Smith chart, as illustrated by circles *A*, *B*, and *C* in Fig. 3.39. Circle *A*, for which $R = Z_0$ passes through the origin, is called the condition of critical coupling ($\beta = 1$ from the preceding table), since it provides a perfect match to the transmission line at resonance. Circle *C*, with $R > Z_0$, is said to be overcoupled ($\beta > 1$ from the preceding table), and circle *B*, with $R < Z_0$, is undercoupled. The coupled coefficient for any given impedance locus can be easily determined by measuring the reflection coefficient S_{110} at resonance.

For the undercoupled case

$$\beta = \frac{1 - S_{110}}{1 + S_{110}} \tag{3.71}$$

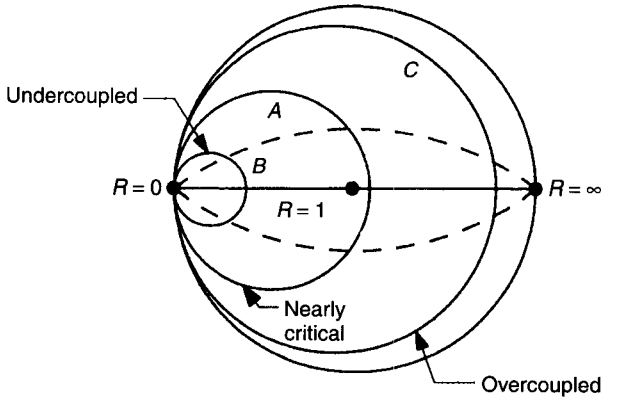


Figure 3.39 Input impedance of resonant cavity referred to detuned-short position plotted on Smith chart for three degrees of coupling.

and for the overcoupled case

$$\beta = \frac{1 + S_{110}}{1 - S_{110}} \tag{3.72}$$

The evaluation of β locates the intersection of the impedance circle with the real axis, as shown in Fig. 3.40.

In order to measure various quality factors, (3.70) can be written as

$$\bar{Z}_{\text{in}} = \frac{Z_{\text{in}}}{Z_0} = \frac{\beta}{1 + 2jQ_u\delta} = \frac{\beta}{1 + 2jQ_L(1+\beta)} = \frac{\beta}{1 + 2jQ_e\beta} \tag{3.73}$$

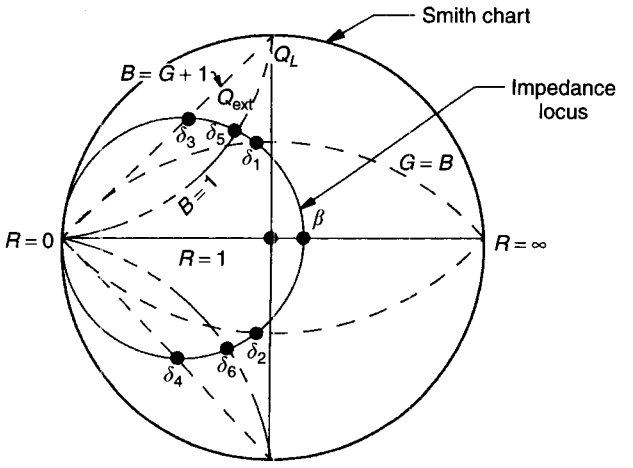


Figure 3.40 Identification of half-power points from Smith chart; Q_0 locus is given by $B = G$, Q_L by $B = G + 1$, and Q'_{ext} by $B = 1$.

where Q_u, Q_L, and Q_e are interrelated by the well-known relation

$$Q_u = Q_L(1+\beta) = Q_e\beta \tag{3.74}$$

The normalized frequency deviations corresponding to various quality factors are given by

$$\delta_u = \pm\frac{1}{2Q_u} \qquad \delta_L = \pm\frac{1}{2Q_L} \qquad \delta_e = \pm\frac{1}{2Q_e} \tag{3.75}$$

The impedance locus of Q_u, for example, can be determined by using (3.75) in (3.73) and is given by

$$(Z_{\text{in}})_u = \frac{\beta}{1 \pm j} \tag{3.76}$$

Equation (3.76) represents the points on the impedance locus where the real and imaginary parts of the impedance are the same. Figure 3.40 represents the locus of these points (corresponding to $R = X$) for all possible values of β. This locus is an arc whose center is at $Z = 0 \pm j$ and the radius is the distance to the point $0 \pm j0$. The intersection of this arc with the impedance locus determines the Q_u measurement points

$$Q_u = \frac{f_0}{f_1 - f_2} \tag{3.77}$$

The frequencies f_1 and f_2 are called half-power points because these points correspond to $R = X$ on the impedance locus.

The loaded and external Q values can be determined in a similar way. From (3.73) and (3.75), the impedances corresponding to Q_e and Q_L are given by

$$(Z_{\text{in}})_e = \frac{\beta}{1 \pm j\beta} \tag{3.78}$$

$$(Z_{\text{in}})_L = \frac{\beta}{1 \pm j(1+\beta)} \tag{3.79}$$

Using (3.78) and (3.79), the Q_e and Q_L loci can be easily determined. These loci are shown in Fig. 3.40.

3.7.2 Two-Port Resonator

Two-port resonant circuits are commonly used as transmission components in a number of applications. The equivalent circuit of a commonly used two-port resonator is shown in Fig. 3.41*a*. The input and output coupling coefficients are represented by β_1 and β_2 in

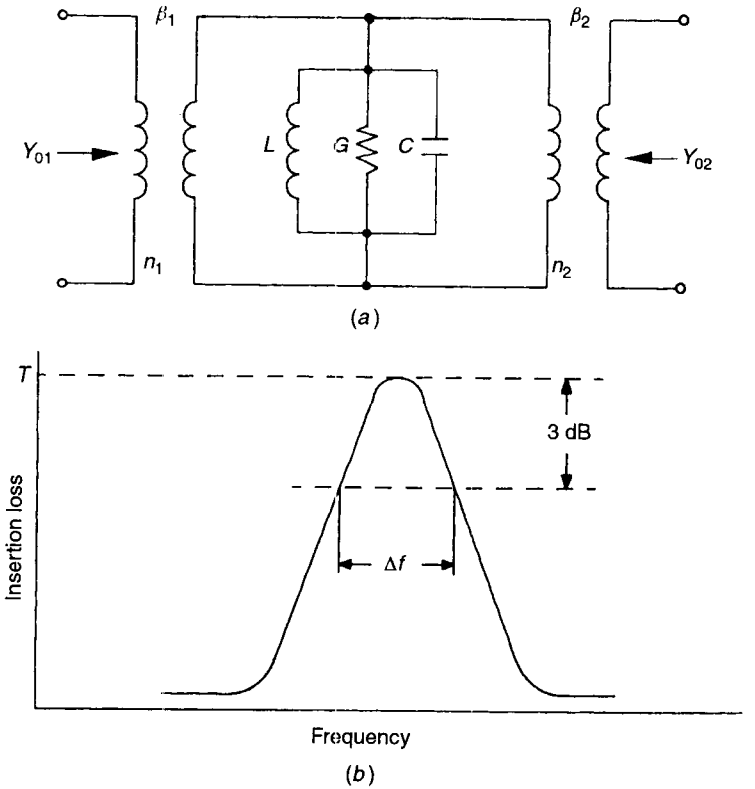


Figure 3.41 (*a*) Equivalent circuit of two-port resonator; (*b*) transmission response of two-port resonator.

$$\beta_1 = \frac{Y_{01}}{n_1^2 G} \qquad \beta_2 = \frac{Y_{02}}{n_2^2 G} \tag{3.80}$$

where Y_{01} and Y_{02} are the input and output transformed admittances. The coupling coefficients can be directly determined by measuring the VSWR at the input and output ports with the other port open circuited. The transmission response of such a resonant circuit measured using the setup of Fig. 3.39*b* is shown in Fig. 3.41*b*. The coupling coefficients and the quality factors can be determined from the measurement of the insertion loss T at resonant frequency and the 3-dB bandwidth Δf using the following well-known relations [54]:

$$T = \frac{2\sqrt{\beta_1 \beta_2}}{1 + \beta_1 + \beta_2} \tag{3.81}$$

$$Q_L = \frac{f_0}{\Delta f} \tag{3.82}$$

$$Q_u = Q_L(1 + \beta_1 + \beta_2) \tag{3.83}$$

REFERENCES

1. Collin, R. E., *Foundations of Microwave Engineering*, McGraw-Hill, New York, 1966, Chap. 7.
2. Ramo, S., J. Whinnery, and T. VanDuzer, *Fields and Waves in Communication Electronics*, Wiley, New York, 1965, Chap. 10.
3. Ghose, R. N., *Microwave Circuit Theory and Analysis*, McGraw-Hill, New York, 1963, Chap. 8.
4. Moreno, T., *Microwave Transmission Design Data*, Dover, New York, 1958, Chap. 18.
5. Liao, S. Y., *Microwave Devices and Circuits*, Prentice-Hall, Englewood Cliffs, NJ, 1980, Chap. 4.
6. Kretzschmar, J. G., "Wave Propagation in Hollow Conducting Elliptic Waveguide," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-18, Sept. 1970, pp. 547–554.
7. Kretzschmar, J. G., "Mode Chart for Elliptical Resonant Cavities," *Electron. Lett.*, Vol. 6, July 1970, pp. 432–433.
8. Rengarajan, S. R., and J. E. Lewis, "Quality Factor of Elliptical Cylindrical Resonant Cavities," *J. Microwave Power*, Vol. 15, Jan. 1980, pp. 53–57.
9. Itoh, T., R. Mittra, and R. D. Wards, "A Method for Computing Edge Capacitance of Finite and Semi-Finite Microstrip Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-20, Dec. 1972, pp. 847–849.
10. James, D. S., and S. H. Tse, "Microstrip End Effects," *Electron. Lett.*, Vol. 8, Jan. 1972, pp. 46–47.
11. Garg, R., and I. J. Bahl, "Microstrip Discontinuities," *Int. J. Electron.*, Vol. 45, Jan. 1978, pp. 81–87.
12. Carlile, D., T. Itoh, and R. Mittra, "A Study of Rectangular Microstrip Resonators," *Archiv für Elektronik und Ubertragungstechnik*, Vol. 30, Jan. 1976, pp. 38–41.
13. Itoh, T. "Analysis of Microstrip Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, Nov. 1974, pp. 946–952.
14. Itoh, T., and R. Mittra, "Analysis of Microstrip Disk Resonators," *AEU*, Vol. 27, Nov. 1973, pp. 456–458.
15. Sharma, A. K., and B. Bhat, "Influence of Shielding on the Capacitance of Microstrip Disk and Ring Structures," *AEU*, Vol. 34, Jan. 1980, pp. 41–44.
16. Wolff, I., and N. Knoppik, "Rectangular and Circular Microstrip Disk Capacitors and Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, Oct. 1974, pp. 857–864.
17. Watkins, J., "Circulator Resonator Structures in Microstrip," *Electron. Lett.*, Vol. 5, Oct. 1969, pp. 524–525.
18. Entschladen, H., "Simple Design of Microstrip Disk Resonators," *AEU*, Vol. 29, 1975, pp. 184–185.
19. Troughton, P., "Measurement Techniques in Microstrip," *Electron. Lett.*, Vol. 5, Jan. 1969, pp. 25–26.
20. Wolff, I., and N. Knoppik, "Microstrip Ring Resonator and Dispersion Measurement on Microstrip Line," *Electron. Lett.*, Vol. 7, 1971, pp. 779–781.

21. Sharma, A. K., and B. Bhat, "Spectral Domain Analysis of Microstrip Ring Resonators," *AEU*, Vol. 33, Mar. 1979, pp. 130–132.

22. Wu, Y. S., and F. J. Rosenbaum, "Mode Chart for Microstrip Ring Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-21, July 1973, pp. 487–489.

23. Helszajn, J., and D. S. James, "Planar Triangular Resonators with Magnetic Walls," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-26, Feb. 1978, pp. 95–100.

24. Helszajn, J., D. S. James, and W. T. Nisbet, "Circulators Using Planar Triangular Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, Feb. 1979, pp. 188–193.

25. Cuhaci, M., and D. S. James, "Radiation from Triangular and Circular Resonators in Microstrip," paper presented at the IEEE International Microwave Symposium, Dig., 1977, pp. 438–441.

26. Schelkunoff, S. A., *Electromagnetic Waves*, Van Nostrand, New York, 1943, p. 393.

27. Ng, F. L., "Tabulation of Method for the Numerical Solution of the Hollow Waveguide Problem," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, Mar. 1974, pp. 322–329.

28. Thomas, D. T., "Functional Approximation for Solving Boundary Value Problems by Computer," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-17, Aug. 1969, pp. 447–454.

29. Sharma, A. K., "Spectral Domain Analysis of Microstrip Resonant Structures," Ph.D. Thesis, Indian Institute of Technology, Delhi, Dec. 1979.

30. Sharma, A. K., and B. Bhat, "Analysis of Triangular Microstrip Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Nov. 1982, pp. 2029–2031.

31. Hagesawa, T., et al., "Characteristics of Valley Microstrip Lines for Use in Multilayer MMICs," *IEEE Microwaves Guided Wave Lett.*, Vol. 1, Oct. 1991, pp. 275–277.

32. Sun, Y., J. L. Tauritz, and R. G. F. Baets, "Micromachined RF Passive Components and Their Applications in MMICs," *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 9, July 1999, pp. 310–325.

33. Chi, C. Y., and G. M. Rebeiz, "Planar Microwave and Millimeter-Wave Lumped Elements and Coupled Line Filters Using Micro-machining Techniques," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-43, Apr. 1995, pp. 730–738.

34. Weller, T. M., L. P. B. Katehi, and G. Rebeiz, "High Performance Microshield Line Passive Components," Vol. MTT-43, Mar. 1995, pp. 534–543.

35. Schnieder, F., R. Doerner, and W. Heinrich, "High Impedance Coplanar Waveguides with Low Attenuation," *IEEE Microwave Guided Wave Lett.*, Vol. 6, Mar. 1996, pp. 117–119.

36. Bahl, I. J., E. L. Griffin, J. Dilley, and M. Balzan, "Low Loss Multilayer Microstrip Line for Monolithic Microwave Integrated Circuits Applications," *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 8, Dec. 1998, pp. 441–454.

37. Bhartia, P., and I. J. Bahl, "Frequency Agile Microstrip Antennas," *Microwave J.*, Oct. 1982, pp. 67–70.

38. Chang, K., *Microwave Ring Circuits and Antennas*, Wiley, New York, 1998, Chap. 4.

39. Navarro, J. A., and K. Chang, "Varactor Tunable Uniplanar Ring Resonators," *IEEE Trans. Microwave Theory and Tech.*, Vol. MTT-41, May 1993, pp. 760–766.
40. Plourde, J. K., and C. L. Ren, "Application of Dielectric Resonators in Microwave Components," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-29, Aug. 1981, pp. 754–770.
41. Fiedziuszko, S. J., and S. Holme, "Dielectric Resonators, Raise Your High-Q," *IEEE Microwave Mag.*, Vol. 2, Sep. 2001, pp. 51–60.
42. Ready, D. W., et al., "Microwave High Dielectric Constant Materials," Quarterly Repts. 1–4, Final Rep., Contract DAAB07-69-C-0455, Raytheon Company, Waltham, MA, Nov. 1969–Feb. 1971.
43. Masse, D. J., et al., "A New Low-Loss High-K Temperature Compensated Dielectric for Microwave Applications," *Proc. IEEE*, Vol. 5, Nov. 1971, pp. 1628–1629.
44. Simonet, W., and J. C. Mage, "(Zr, Sn) TiO_4 Dielectric Materials with Adjustable Temperature Coefficients and Low Microwave Loss," paper presented at the American Ceramic Society Meet., Chicago, IL, Apr. 29, 1980.
45. Montvala, A. J., "Microwave Dielectric Constant Materials," Quarterly Repts. 1–3, Contract DAAB07-69-C-0402, Illinois Research Institute, Chicago, IL, July 1969–Jan. 1970.
46. Kajfez, D., and P. Guillon, *Dielectric Resonators*, Artech House, Dedham, MA, 1986.
47. Guillon, P., et al., "Microstrip Bandstop Filter Using a Dielectric Resonator," *Proc. IEE*, Vol. 128, No. 3, June 1981, pp. 151–154.
48. Khanna, A. P. S., and Y. Garault, "Determination of Loaded, Unloaded and External Quality Factors of a Dielectric Resonator Coupled to a Microstrip Line," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-31, Mar. 1983, pp. 261–264.
49. Guillon, P., IRCOM, Uny. of Limoges, France, Apr. 1982, private communication.
50. Gil, F. H., and J. P. Martinez, "Analysis of Dielectric Resonators with Tuning Screw and Supporting Structure," *IEEE Trans. Microwave Theory Tech.*, Vol. MIT-33, Dec. 1985, pp. 1453–1457.
51. Osbrink, N. K., "YIG-Tuned Oscillator Fundamentals," *Microwave Syst. News*, Vol. 13, Nov. 1983, pp. 207–225.
52. Ollivier, P. M., "Microwave YIG-Tuned Transistor Oscillator Amplifier Design Application to C-Band," *IEEE J. Solid-State Circuits*, Vol. SC-7, Feb. 1972, pp. 54–60.
53. Ginzton, E. L., *Microwave Measurements*, McGraw-Hill, New York, 1957, Chap. 9.
54. Sucher, M., and J. Fox, *Handbook of Microwave Measurements*, 3rd ed., Vol. 2, Wiley, New York, 1963, Chap. 8.

PROBLEMS

3.1 Identify the degenerate modes of a rectangular waveguide resonator when (a) all sides are unequal, (b) two sides are equal, and (c) all sides are equal.

3.2 Determine the lowest frequency at which degenerate modes exist in a cylindrical waveguide cavity.

3.3 Calculate the resonant frequency of a radial reentrant coaxial cavity having inner and outer radii of 0.5 and 1.2 cm and length of 10 cm when the center post is 5 cm from the end wall. Determine the tuning range when d is varied from 2 to 6 cm.

3.4 Find the resonant frequency and Q for a TE_{111} mode cylindrical cavity (copper) of radius 2 cm and length 10 cm. Repeat this for the case when the cavity is filled with material of dielectric constant 4.0.

3.5 Calculate the lowest order resonant frequency of a microstrip resonator having $\epsilon_r = 10$, $h = 0.6$ mm, $t = 6$ μm, $W = 2$ mm, and $L = 1$ cm.

3.6 Derive expressions for Q factors of rectangular and circular patch microstrip resonators.

3.7 Calculate and compare the quasi-static resonant frequencies of circular, triangular, and hexagonal microstrip resonators with identical dimension a and substrate thickness h.

3.8 Design a $TE_{01\delta}$ mode cylindrical dielectric resonator at 35 GHz when placed in the shielded MIC package shown in Fig. 3.29 using the following parameters.

Substrate: $\epsilon_s = 9.9$, height $h = 0.25$ mm.
Dielectric resonator: $\epsilon_r = 36$, $D/H = 2.5 \pm 0.1$.
Distance: d_s: 1 mm.

3.9 In Problem 3.8, find the change in the resonant frequency if the dielectric constant of the dielectric resonator is changed to 38 from 36.

3.10 The equivalent circuit of a dielectric resonator coupled to a microstrip line is shown in Figs. 3.31 and 3.32. Find the relations, in terms of S_{110} and S_{210}, to draw the loci of the unloaded, loaded, and external quality factor determination points (Q circles) in the S_{11} and S_{21} planes at PP'. Draw the Q circles on a Smith chart and a Polar chart from the S_{11} and S_{21} planes, respectively.

3.11 In the case of a parallel-tuned resonant circuit, 40% of the incident power is reflected at the resonant frequency. The real part (R) equals the imaginary part (X) on the impedance locus at 9230 and 9240 MHz. Calculate the resonant frequency, the coupling coefficient, and the loaded, unloaded, and external quality factors for an undercoupled case and an overcoupled case.

4

IMPEDANCE TRANSFORMATION TECHNIQUES

Prasad N. Shastry

4.1 INTRODUCTION

Impedance transformation or matching is an essential part of the design of RF and microwave devices or systems. Impedance matching is important for (a) maximum transfer of power from a generator to a device, from a device to a load, or between devices; (b) improving the sensitivity of receivers; (c) reducing the amplitude and phase imbalances in power distribution networks; (d) obtaining optimum gain, output power, efficiency, and dynamic range in amplifiers; and (e) minimizing power loss in feed lines. There are a variety of matching networks that can be synthesized to match two given impedances. The important factors to be considered in the selection of a matching network are (a) complexity, (b) bandwidth, (c) frequency response, and (d) ease of implementation.

In this chapter, impedance matching techniques for narrow-band and wide-band applications are presented in Sections 4.2 and 4.3, respectively. The techniques for synthesizing lumped as well as distributed matching networks are illustrated by means of examples. The impedance matching techniques presented here are in general applicable to systems using parallel wire lines, coaxial lines, waveguides, or planar transmission lines such as strip lines, microstrip lines, and coplanar waveguides.

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

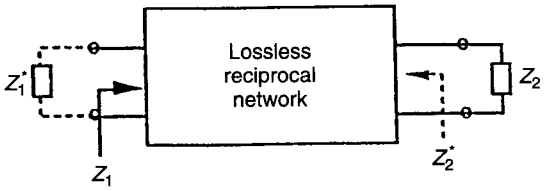


Figure 4.1 Impedance transformation by lossless reciprocal network.

The extent of published literature on impedance matching is vast. In this chapter the author has made an attempt to cover only the important and commonly used techniques. For further reading on this subject, several references have been given.

4.2 NARROW-BAND TRANSFORMATION TECHNIQUES

In this section, the techniques of narrow-band matching will be illustrated with examples. The matching is accomplished perfectly only at the desired frequency. A bandwidth around the design frequency can be defined if the tolerable reflection coefficient in the band is specified.

The relationship between the impedances at the two ports of a lossless, reciprocal matching network designed to transform Z_2 to Z_1 is summarized in Fig. 4.1. If an impedance, Z_2, is connected at port 2 of the network, the impedance measured at port 1 is Z_1. Then if an impedance, Z_1^*, is placed at port 1, the impedance measured at port 2 will be Z_2^*. Therefore, if Z_1^* represents the internal impedance of a source, then maximum power transfer to the load impedance Z_2 is ensured.

The foregoing property of a lossless, reciprocal network implies that the same network is capable of transforming Z_2 to Z_1 as well as Z_1^* to Z_2^* and hence offers the choice from two design approaches: (a) transformation of Z_2 to Z_1 or (b) transformation of Z_1^* to Z_2^*. This property of the matching network may be verified in the design examples presented in this section.

The design of lossless matching networks consisting of lumped as well as distributed elements will be presented in the following sections. The graphical design procedures making use of the Smith chart as a tool are illustrated by means of examples. The theory and applications of the Smith chart are covered in detail in the literature [1–4]. Analytical solutions to matching circuit design problems can be found elsewhere [2, 3, 5]. Smith chart solutions are sufficiently accurate for most applications. However, if better accuracy is needed or for development of a computer program for the design, one may resort to analytical solutions. Therefore, analytical solutions have been provided here only for selected design techniques.

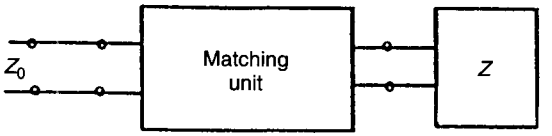


Figure 4.2 Matching an impedance Z (complex) to Z_0 (real).

4.2.1 Distributed-Element Techniques

The matching network design techniques that make use of lossless transmission lines as network elements will be presented here. Typical matching problems in practice involve matching a complex impedance (input or output impedance of a device) to a real impedance (source, load, or characteristic impedance of a line), as shown in Fig. 4.2, or transformation between complex impedances, as shown in Fig. 4.1.

Transmission-Line Transformer. A lossless transmission line having characteristic impedance Z_0 and length l can be used in the transformation of $Z_1 = (R_1 \pm jX_1)$ to $Z_2 = (R_2 \pm jX_2)$. The values of Z_0 and l needed for the transformation can be determined analytically [6] or graphically (Smith chart). However, practically realizable design usually involves a tedious iterative process. Further, a realizable solution may not always exist.

Single-Stub Matching

Principle of Shunt Stub Matching. The basic single-shunt-stub matching system is shown in Fig. 4.3. The normalized load admittance Y_L' (normalized with respect to the characteristic admittance Y_0 of the line to which it is connected) is transformed to the unit conductance circle on the Smith chart by the trans-

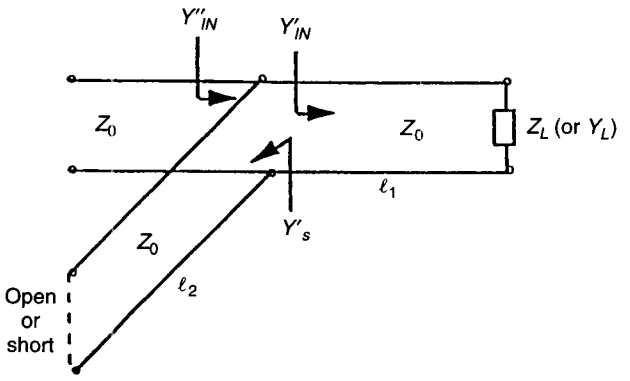


Figure 4.3 Single-stub matching system.

mission line of length l_1. The normalized admittance looking toward the load is, then, $Y'_{IN} = 1 \pm jB'$. Now a length l_2 of an open or short-circuited stub can be chosen such that the admittance Y'_S $(\mp jB'_s)$ of the stub equals $\pm jB'$, thereby canceling the susceptance provided by Y'_{IN}. Then the admittance Y''_{IN} looking toward the load from just left of the stub is $1 + j0$. Thus the match is accomplished.

Design Example. Consider the problem of matching $Z_L = 100 + j100\ \Omega$ to a line having a characteristic impedance $Z_0 = 50\ \Omega$. The single-stub matching is illustrated on the Smith chart in Fig. 4.4 for this example. The two solutions

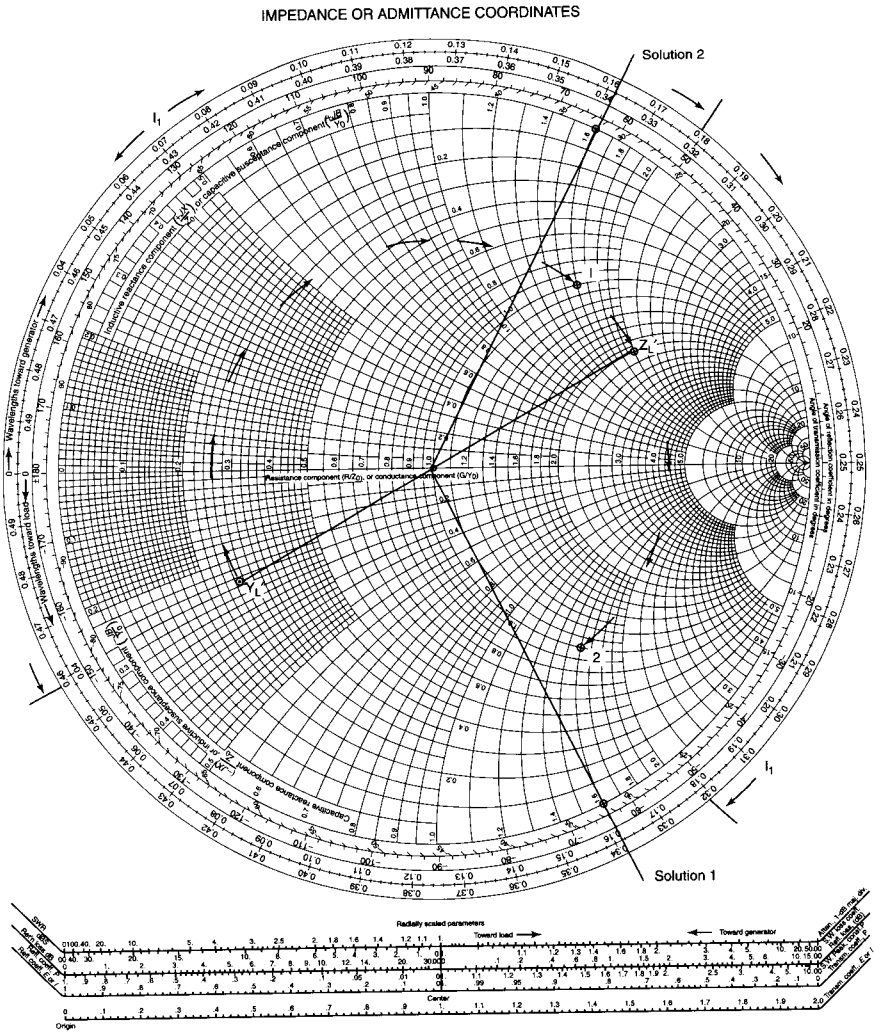


Figure 4.4 Solutions to single-stub matching example problem.

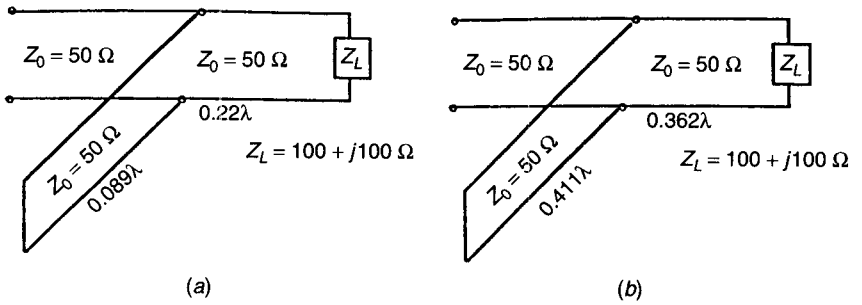


Figure 4.5 Solutions to single-stub design problem.

(using short-circuited stubs) to the design problem are shown in Figs. 4.5*a* and *b*.

The matching solution in Fig. 4.5*a* is preferable because it has a greater bandwidth and uses shorter lengths of transmission lines, and consequently the losses are smaller. The bandwidth of the matching unit is an important selection criterion. It should be noted that the addition of $\frac{1}{2}n\lambda$ (where n is an integer) lengths of lines in the designs in Fig. 4.5 will not affect the match.

In this example the line and the stub have the same characteristic impedance. If different characteristic impedances are used for the line and the stub, the impedances (or admittances) have to be appropriately normalized while using the Smith chart. The disadvantage of a single-stub system is that it requires an adjustable length of line of constant characteristic impedance in order to match to an arbitrary load. Such lines are difficult to construct in practice.

Principle of Series Stub Matching. The schematic of a series stub matching system is shown in Fig. 4.6. The distance d is selected so that the impedance, Z_{IN}, seen looking into the line at distance d from the load is of the form

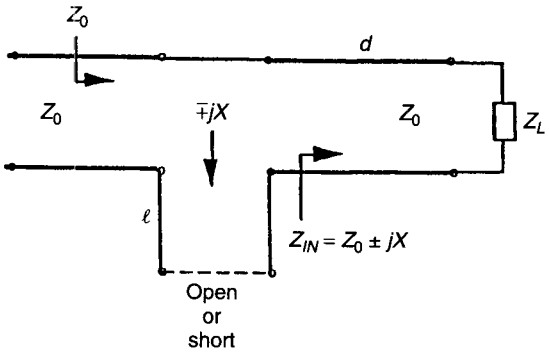


Figure 4.6 Series stub matching system.

$Z_0 \pm jX$. Then the stub reactance is chosen as $\mp jX$, resulting in a match to the feed-line impedance Z_0. The line length d and the stub length l can be determined using a Smith chart. The reader is referred to the design example by Pozar [2]. The series stub is difficult to fabricate in coaxial, strip-line, and microstrip-line systems, and hence this technique is not commonly used in practice. However, series stubs can be constructed easily in coplanar waveguide systems in which the center conductor and the ground plane are located on the same side of the circuit board.

Double-Stub Matching

Principle of Double-Stub Matching. A double-stub matching system is shown in Fig. 4.7. The advantage of a double-stub system is that it uses only adjustable-length stubs and a fixed length of line, unlike a single-stub tuner. However, a double-stub tuner cannot match an arbitrary load impedance, unlike a single-stub tuner [1].

The length l_1 of the shunt stub at the load (Fig. 4.7) must be chosen such that the admittance of the load and stub (open or shorted) is transformed to the unit conductance circle on the Smith chart by the fixed length d of the line. Then the normalized admittance Y'_{IN} seen looking toward the load is $1 \pm jB'$. The length l_2 of the second stub (open or shorted) must then be chosen such that its admittance Y'_s ($\mp jB'_s$) cancels $\pm jB'$. The normalized admittance Y''_{IN} looking toward the load from just left of the second stub is then $1 + j0$, and hence the match to the feed-line impedance Z_0 is obtained.

In order to determine the length (l_1) of the first stub, the unity conductance circle is rotated on the Smith chart toward the load through a distance d. Then the susceptance needed from the first stub is determined from the point of intersection of the load conductance contour and the rotated unity conductance circle. The susceptance needed from the second stub can then be determined from the corresponding point on the unity conductance circle.

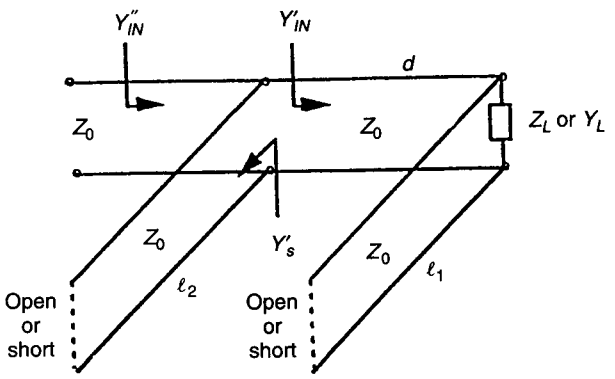


Figure 4.7 Double-stub matching system.

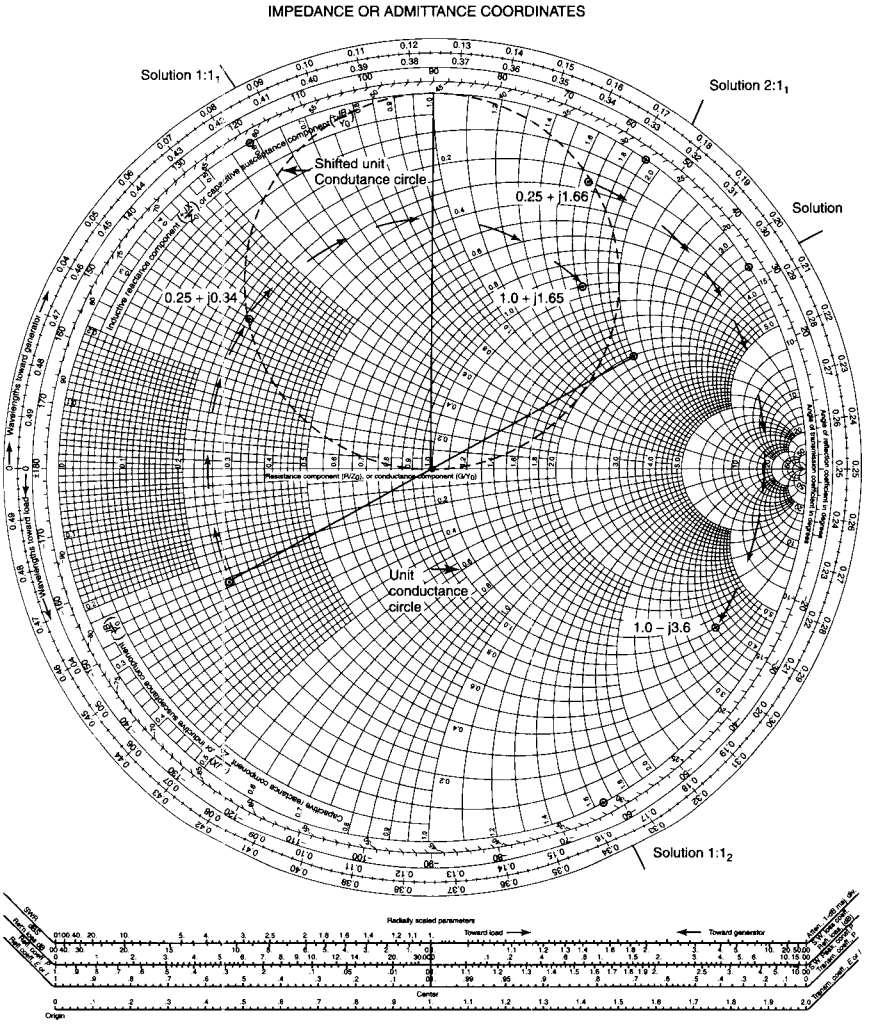


Figure 4.8 Solutions to double-stub design problem.

Design Example. Consider the problem of matching $Z_L = 100 + 100\ \Omega$ to $Z_0 = 50\ \Omega$; $d = \frac{1}{8}\lambda$. The double-stub matching is illustrated on the Smith chart in Fig. 4.8. The two solutions (using shorted stubs) are depicted in Figs. 4.9*a* and *b*. The design procedures are given in the following:

(a) Locate Z_L' and determine Y_L' $(0.25 - j0.25)$ as shown in Fig. 4.8.

(b) Draw the "shifted unity conductance circle" by rotating the "normal unity conductance circle" toward load by a distance d $(=\frac{1}{8}\lambda)$, as shown in Fig. 4.8.

(c) Determine the points of intersection of the load conductance (0.25) contour and the shifted unity conductance circle as shown in Fig. 4.8. The points of intersection are located at $0.25 + j0.34$ and $0.25 + j1.66$. From the difference in susceptance at the intersection point and the load susceptance, the length (l_1) of the first stub can be determined. The two solutions for l_1 are shown in Fig. 4.8.

(d) Move the points $0.25 + j0.34$ and $0.25 + j1.66$ on the shifted unity conductance circle to points $1 + j1.65$ and $1 - j3.6$, respectively, on the normal unity conductance circle. This movement involves traveling on constant standing wave ratio (SWR) circles, toward the generator, through a distance d $(=\frac{1}{8}\lambda)$, as shown in Fig. 4.8.

(e) Determine the length (l_2) of the second stub needed to cancel the susceptances at the points on the normal unity conductance circle. The two solutions for l_2 are shown in Fig. 4.8, and the two solutions to the design problem are shown in Fig. 4.9.

In this example problem, there is no appreciable difference between the solutions with regard to losses in the lines. But the solution in Fig. 4.9*a* has a greater bandwidth. The distance d between the stubs must be chosen such that the conductance contour of the load admittance intersects the rotated unity conductance circle. Stub spacings of $\frac{1}{8}\lambda$ and $\frac{3}{8}\lambda$ have been found to be reasonable choices in practice as far as matching range and critical adjustments are concerned.

If the load should happen to fall in the portion of the chart such that matching is not possible using a specified distance between the stubs, then it is necessary to add a fixed length of line between the load and the first stub in order to adjust the load to a value that can be matched. In this design example, the lines and the stubs have the same characteristic impedance. If different characteristic impedances are used for the line and the stubs, the impedances (or admittances) have to be appropriately normalized while using the Smith chart.

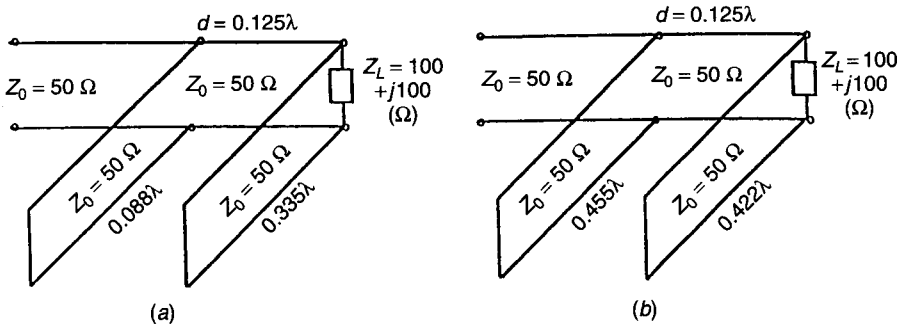


Figure 4.9 Solutions to double-stub design problem.

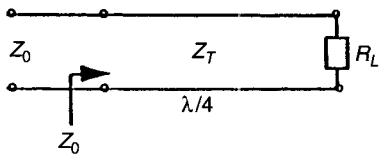


Figure 4.10 Quarter-wave impedance transformer.

In practice, when stub matching is used in devices containing active components that need DC bias, the short circuits in the stubs are replaced by capacitors in order to provide DC bias through the stubs. The stub lengths in these cases must be calculated taking the capacitive reactance into account. Open stubs can be used only if designed properly to prevent radiation (e.g., open-ended microstrip lines). The fringing capacitance at the open end must be accounted for in the design of the stubs.

Quarter-Wave Transformer. A load impedance (Z_L) having only a real part (R_L) can be matched to a transmission line having a characteristic impedance Z_0 by means of a quarter-wavelength ($\frac{1}{4}\lambda$) of a transmission line, as shown in Fig. 4.10. Here, λ is the wavelength in the quarter-wave section at the design frequency. The characteristic impedance of the quarter-wave section is given by

$$Z_T = \sqrt{R_L Z_0} \tag{4.1}$$

If the load impedance is complex ($R_L + jX_L$), one must convert it into a real impedance (R_L') by means of an additional length (l) of line, as shown in Fig. 4.11*a*, or by tuning out the reactive part by means of a stub, as shown in Fig. 4.11*b*, before designing the quarter-wave transformer. Then,

$$Z_T = \sqrt{R_L' Z_0} \tag{4.2}$$

Quarter-Wave and $\frac{1}{8}\lambda$ Transformer. The complex load impedance (Z_L) can also be converted into a real impedance (R_L') by means of a $\frac{1}{8}n\lambda$ (where n is an odd integer) length of line having a characteristic impedance equal to the magnitude of the load impedance ($|Z_L|$). The load impedance Z_L, however, must

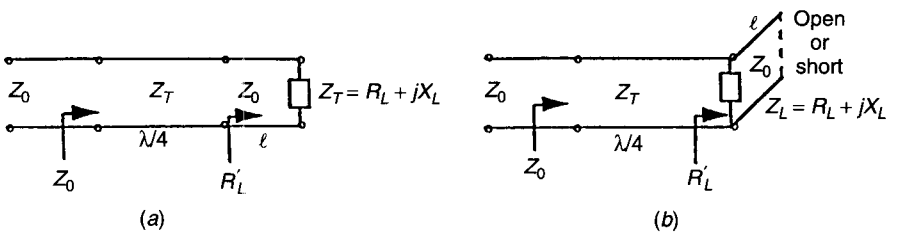


Figure 4.11 Matching complex load to real impedance by means of quarter-wave transformer.

be normalized with respect to $|Z_L|$ before using the Smith chart [7]. Using this technique, complex load and source impedances can be matched with the help of two $\frac{1}{8}\lambda$ transformers and a quarter-wave transformer [6].

The fractional bandwidth of a quarter-wave transformer is given by

$$\frac{\Delta f}{f_0} = 2 - \frac{4}{\pi}\cos^{-1}\left(\frac{\rho_m}{\sqrt{1-\rho_m^2}}\frac{2\sqrt{Z_0 R_L}}{|R_L - Z_0|}\right), \tag{4.3}$$

where ρ_m is the magnitude of the tolerable reflection coefficient in the band Δf and f_0 is the band center frequency. Equation (4.3) is valid for TEM lines only. In non-TEM lines, the phase velocity and wave impedance are frequency dependent, and hence the analysis based on which (4.3) has been derived is not valid unless the bandwidth of operation is small. Discontinuities in the cross-sectional dimensions of the lines due to the step change in impedance produce junction capacitances. This effect can be compensated for by making an adjustment in the length of the transformer.

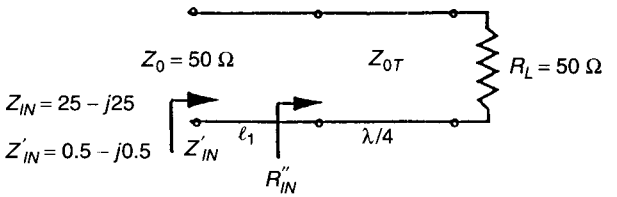
Real to Complex Impedance Transformation. Real to complex impedance transformation can be achieved using the techniques already discussed in this section. The need for such a transformation arises, for example, in transistor amplifiers in which the real load and source impedances have to be transformed to appropriate complex impedances at the transistor output and input terminals, respectively. The matching network design is illustrated here with an example.

Design Example. Consider the design of a matching network to transform R_L (50 Ω) to Z_{IN} ($25 - j25$ Ω). The technique of impedance transformation using a quarter-wave transformer and a transmission line ($Z_0 = 50$ Ω) is illustrated in Fig. 4.12*a*.

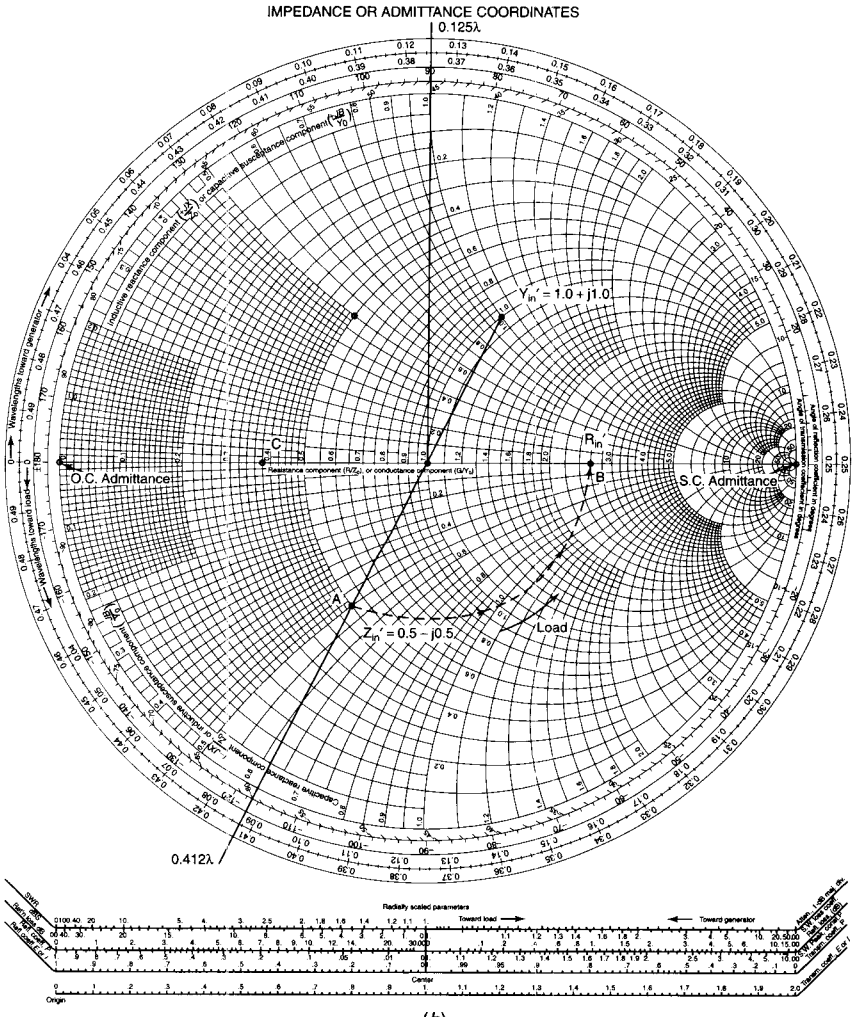
The reference impedance in this problem is 50 Ω. Therefore, $Z'_{IN} = Z_{IN}/50$. Here, Z'_{IN} is located at point A on the Smith chart in Fig. 4.12*b*. By moving toward the load on a constant SWR circle on the Smith chart, point B on the real axis is located. At point B the normalized impedance R''_{IN} (2.6) is purely real. The length l_1 of the 50-Ω transmission line needed to accomplish this is 0.162λ. Then R_{IN} ($R''_{IN} \times 50$) can be transformed to 50 Ω by means of a quarter-wave ($\frac{1}{4}\lambda$) transformer having a characteristic impedance Z_{0T} ($\sqrt{2.6 \times 50 \times 50} = 80.62$ Ω).

The second solution to this matching problem is obtained by going along the constant SWR circle toward the load on the Smith chart and intersecting the real axis at point C to the left of the chart center. The length l_1 of the line needed is 0.412λ and $Z_{0T} = 30.82$ Ω.

The matching networks in this example could have been designed by transforming Z^*_{IN} to R_L, which is evident, for example, in Fig. 4.12*b*. This is due to the property of a reciprocal lossless matching network, illustrated in Fig. 4.1.



(*a*)



(*b*)

Figure 4.12 (*a*) Matching circuit example. (*b*) Solutions to design example; quarter-wave transformer and transmission line.

4.2.2 Lumped-Element Techniques

In this section, impedance transformation techniques using lossless lumped reactive elements will be presented. The lumped elements (inductors or capacitors) could be discrete components or they can be realized in microwave ICs as electrically short lengths ($< \frac{1}{8}\lambda$) of microstrip lines (high Z_0 for an inductor and low Z_0 for a shunt capacitor [8]), spiral inductors, or thin-film capacitors. The resonant frequency, quality factor (Q factor), and size of the lumped components are the important factors to be considered in the implementation of the lumped-element matching networks.

Matching networks are required to transform real to complex, complex to complex, and complex to real impedances. A matching network designed to convert Z_L to Z_{IN} also enables one to transform Z_{IN}^* to Z_L^*, as illustrated in Fig. 4.1.

Approaches to the design of lumped-element matching networks are described in the following sections.

L-Network Approach. Two basic topologies of L-section matching networks are shown in Figs. 4.13*a* and *b*. These networks consist of purely reactive elements. The impedances Z_{SE} and Z_{SH} of the series and shunt elements are represented respectively by jX and jB. Consider a design problem where the complex load impedance Z_L $(= R_L + jX_L)$ is required to be transformed to a real impedance Z_{IN} $(= Z_0)$.

If Z_L is outside the unity resistance circle of the Smith chart, then the circuit in Fig. 4.13*a* should be used. Then the series reactance and shunt susceptance are given by [2]

$$X = \pm\sqrt{R_L(Z_0 - R_L)} - X_L \tag{4.4a}$$

$$B = \pm\frac{\sqrt{(Z_0 - R_L)/R_L}}{Z_0} \tag{4.4b}$$

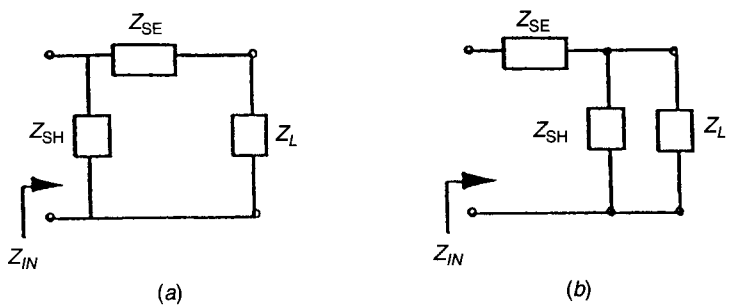


Figure 4.13 Basic topologies of lumped-element matching networks: Z_{SE} = impedance of series element, Z_{SH} = impedance of shunt element (inductive or capacitive).

Since $R_L < Z_0$ (Z_L is outside the unity resistance circle of the Smith chart), the arguments in the square roots are always positive. It is evident that there are two possible solutions.

If Z_L is inside the unity resistance circle of the Smith chart, then the circuit in Fig. 4.13*b* should be used. Then the shunt susceptance and series reactance are given by [2]

$$B = \frac{X_L \pm \sqrt{R_L/Z_0}\sqrt{R_L^2 + X_L^2 - Z_0 R_L}}{R_L^2 + X_L^2} \tag{4.5a}$$

$$X = \frac{1}{B} + \frac{X_L Z_0}{R_L} - \frac{Z_0}{B R_L} \tag{4.5b}$$

Since $R_L > Z_0$ (Z_L is inside the unity resistance circle of Smith chart), the arguments in the square roots are always positive. Again, it is clear that there are two possible solutions.

Once the values of X and B are determined, one could proceed to calculate the values of the lumped elements needed at the design frequency. Positive X implies an inductor and negative X implies a capacitor, while positive B implies a capacitor and negative B implies an inductor.

The design of L networks is illustrated in the following sections using the Smith chart. The graphical design approach gives quick and accurate solutions. Further, it also enables one to gain an insight into the matching network design process.

ZY Smith Chart. It is easier to design L networks using the ZY Smith chart. In this chart, the normal Smith chart (Z chart) is rotated by 180° and superimposed on the original chart. The rotated chart is an admittance chart (Y chart); as a result, the conversion from impedance to admittance coordinates is obtained easily. On a Smith chart, the effect of adding a reactance in series constitutes a movement along a constant resistance circle, and the effect of adding a shunt susceptance results in a movement along a constant conductance circle. Each motion gives the value of the component to be added.

On the basis of the matching network topologies shown in Figs. 4.13*a* and *b*, one can conclude that, depending on the combinations of lumped elements chosen, eight different matching networks are available, as shown in Fig. 4.14. In a given matching network design problem, more than one of the solutions in Fig. 4.14 are possible, and some of the matching networks may not yield the desired match, as will be shown later in this section. In general, the networks shown in Figs. 4.14*a–d* yield the desired match if Z_L' is outside the normalized unity resistance circle on the Smith chart and the networks shown in Figs. 4.14*e–h* give the required match for Z_L' inside the normalized unity resistance circle. Among the possible solutions, the solution in which the lumped elements can be realized easily and that gives the required bandwidth and frequency

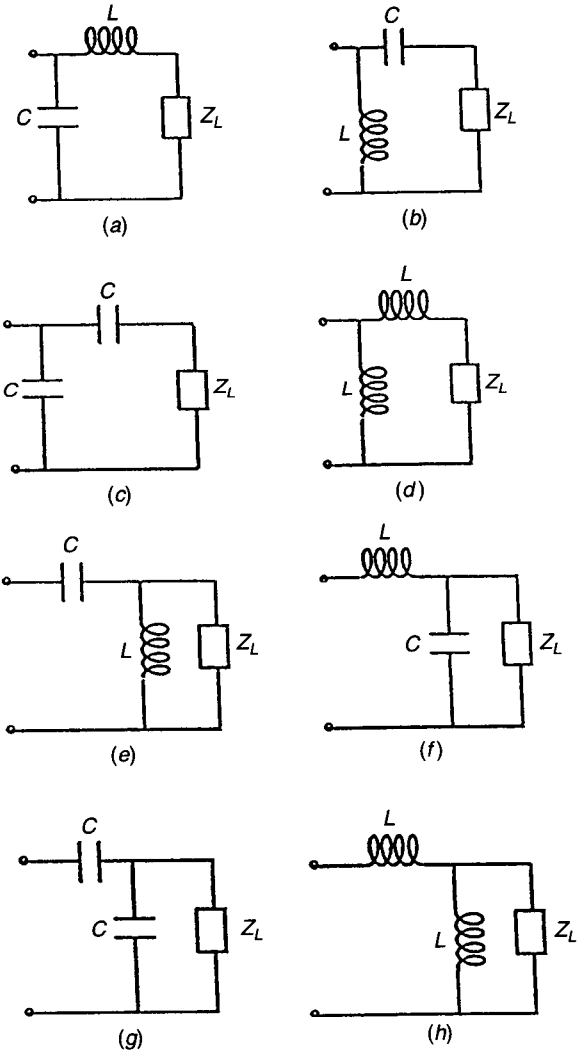


Figure 4.14 Matching networks.

response must be selected. The physical size and resonant frequency of the lumped component are also important considerations in the design of the matching network.

Complex to real and real to complex impedance transformations using lumped-element matching networks are illustrated by means of examples to follow. The admittances and impedances are normalized to Y_0 $(= \frac{1}{50}\ \Omega^{-1})$ and Z_0 $(= 50\ \Omega)$ references, respectively, and represented by primed symbols.

Complex to Real Impedance Transformation

Design Example 1. Consider the transformation of $Y_L = (8 - j12) \times 10^{-3}\ \Omega^{-1}$ to $Z_{IN} = 50\ \Omega$ at 1 GHz. Here Y_L' (Y_L normalized to $\frac{1}{50}\ \Omega^{-1}$) $= 0.4 - j0.6$; $Z_L' = 0.77 + j1.15$.

The load impedance (Z_L') is located at point A on the ZY Smith chart in Fig. 4.15*a*. The load impedance cannot be matched to 50 Ω by means of the network shown in Fig. 4.15*b* because the Z_L' (although outside the normalized unity resistance circle) is in the forbidden region on the Smith chart for this network! The region shown in Fig. 4.15*a* is a forbidden region for this network because adding L in series produces motion in a clockwise direction, away from the unity normalized-conductance circle that passes through the origin.

The design of the matching network shown in Fig. 4.15*c* will be described below using the Smith chart in Fig. 4.15*a*:

$$Y_L' = 0.4 - j0.6 \qquad Z_L' = 0.77 + j1.15 \quad \text{at point } A$$

We need to move to point B on the unity conductance circle, where

$$Y' = 1 + j0.52 \qquad Z' = 0.77 - j0.42 \quad \text{at point } B$$

Therefore, the reactance (X_C) needed to reach point B is $-j78.5\ \Omega$ ($= -j1.57 \times 50\ \Omega$).

The value of the series capacitive element is

$$C = \frac{1}{2\pi f \times 78.5}$$

Since $f = 1$ GHz, $C = 2.027$ pF.

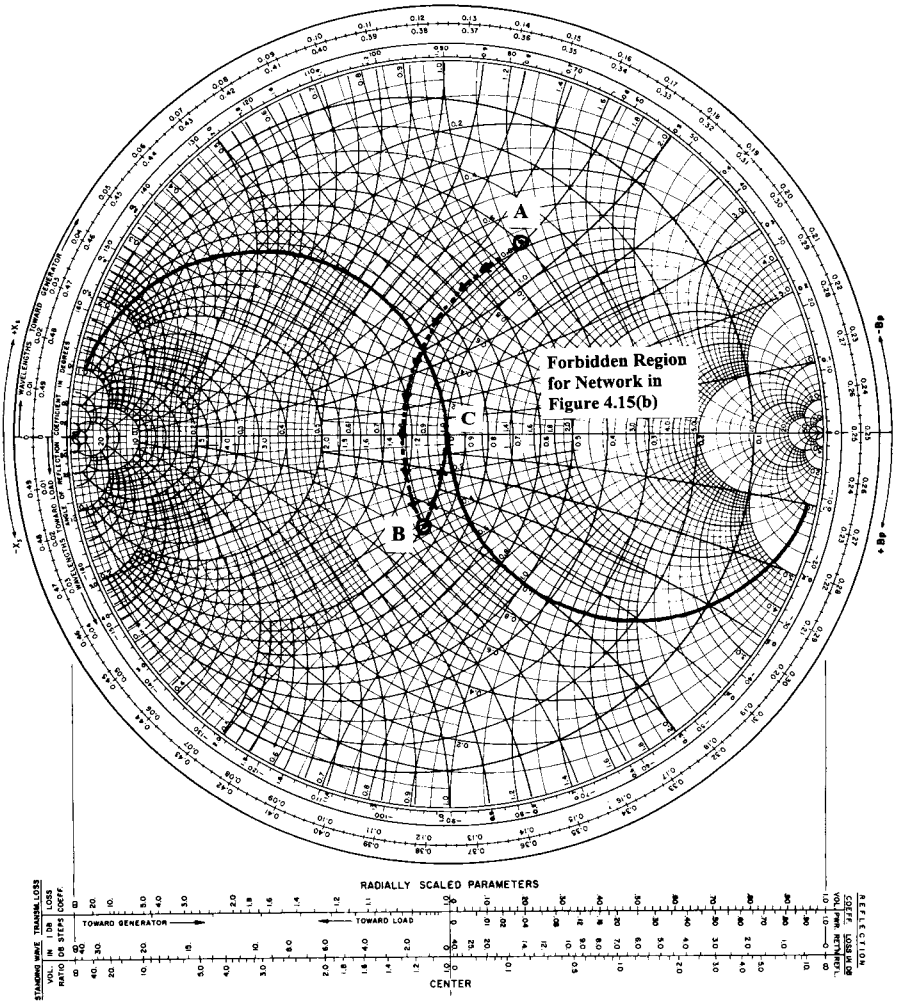
The reactance (X_L) of the shunt inductor needed to reach the center (C) of the chart is $j96.15\ \Omega$ ($= j1.923 \times 50\ \Omega$). Therefore, the value of the shunt inductor is

$$L = \frac{96.15}{2\pi \times 10^9} = 15.3 \text{ nH}$$

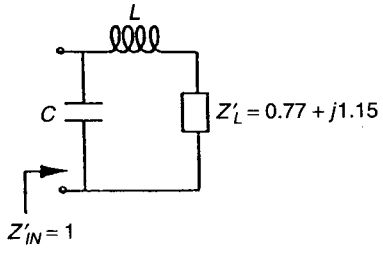
Thus the network in Fig. 4.15*c* transforms the given Y_L to the specified $Z_{IN}(50\ \Omega)$.

Real to Complex Impedance Transformation

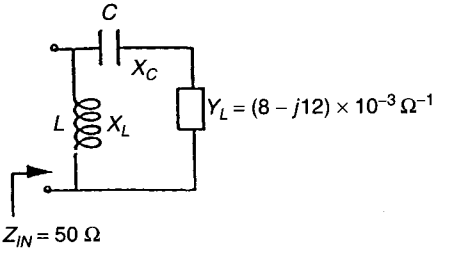
Design Example 2. In this example the design of a matching network, shown in Fig. 4.16*a*, to transform a real impedance to a complex impedance will be illustrated. It is required to transform 50 Ω resistance to an admittance, Y_L, at 700 MHz. Here, $Y_L = (4 - j4) \times 10^{-3}\ \Omega^{-1}$. The procedures are illustrated on the Smith chart in Fig. 4.16*b*.



(*a*)

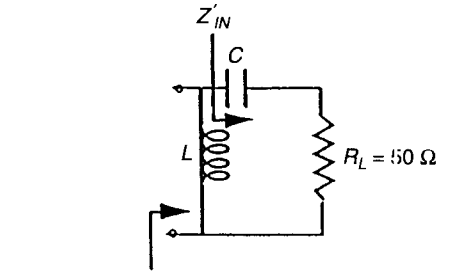


(*b*)



(*c*)

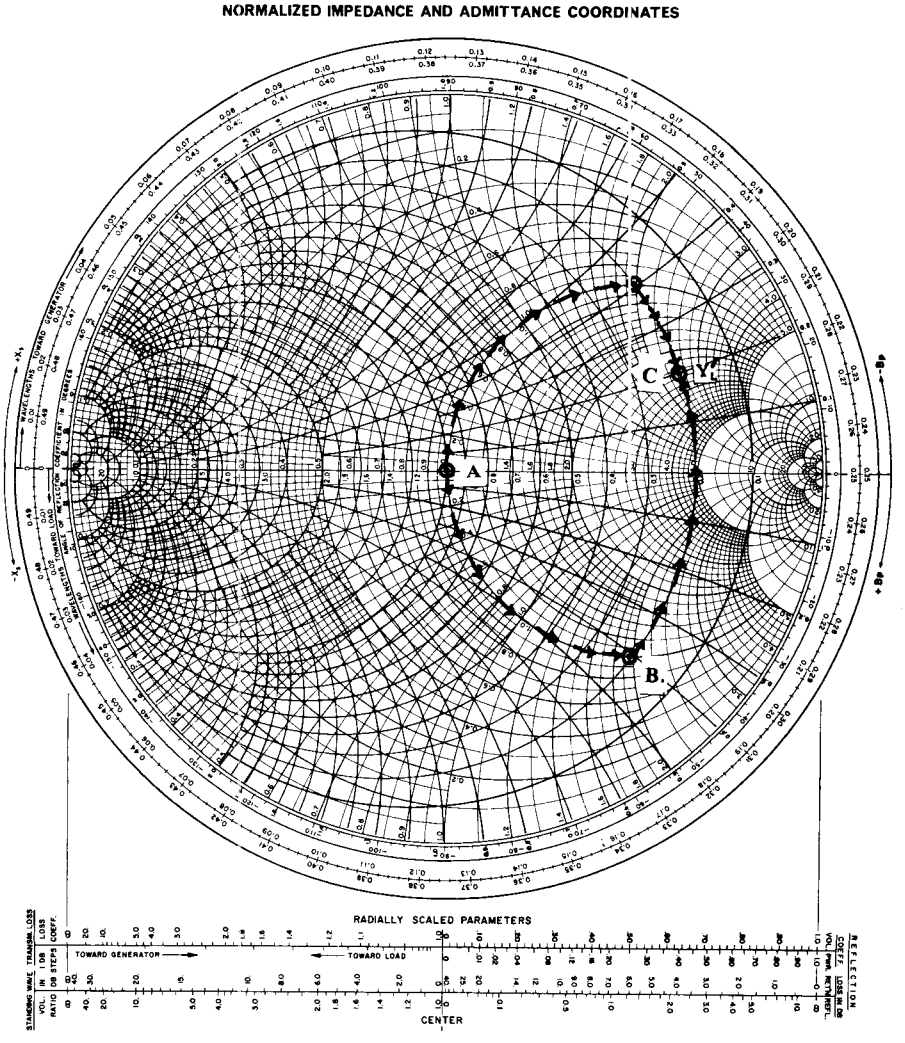
Figure 4.15 (*a*) Solution to Example 1 problem. (*b–c*) Matching networks considered in Example 1.



$Y_L = (4 - j4) \times 10^{-3}\ \Omega^{-1}$

$Y'_L = 0.2 - j0.2$

(*a*)



(*b*)

Figure 4.16 (*a*) Matching network to transform R_L to Y_L in Example 2. (*b*) Solutions to Example 2 problem.

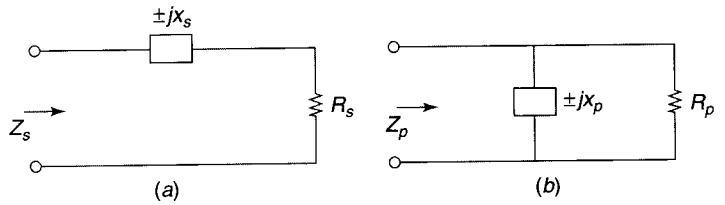


Figure 4.17 Network having resistance and reactance: (*a*) in series; (*b*) in parallel.

The normalized load resistance is located at the center (point A) on the Smith chart. A movement along the constant-resistance circle to point B transforms the load resistance to a normalized impedance, Z'_{IN} at B. Here, $Z'_{IN} = 1 - j1.98$. This requires a series reactance of $-j99.0\ \Omega$. The value of C required to accomplish this is 2.297 pF. At B $Y'_{IN} = 0.2 + j0.4$. We need to reach point C, where Y'_L, the normalized admittance to which the 50 Ω resistance needs to be transformed, is $0.2 - j0.2$. We need to move along the 0.2 constant conductance circle to reach point C. This is accomplished by a shunt inductive reactance of $j5\ \Omega$. At 700 MHz this reactance is obtained by an inductance (L) equal to 18.94 nH.

Network Conversion Approach. This design approach is based on the fact that a series combination of resistance and reactance can be converted to an equivalent parallel combination of similar elements and vice versa.

Consider the two networks shown in Fig. 4.17.

At a given frequency, the two networks in Fig. 4.17 are equivalent if $Z_S = Z_P$. Therefore, one can convert a "parallel network" in Fig. 4.17*b* to a "series network" in Fig. 4.17*a* using the following equations [5]:

$$R_S = \frac{R_P X_P^2}{R_P^2 + X_P^2} \tag{4.6a}$$

$$X_S = \frac{X_P R_P^2}{R_P^2 + X_P^2} \tag{4.6b}$$

Similarly, a series network can be converted into a parallel network using the following equations [5]:

$$R_P = \frac{R_S^2 + X_S^2}{R_S} \tag{4.7a}$$

$$X_P = \frac{R_S^2 + X_S^2}{X_S} \tag{4.7b}$$

It is seen in the foregoing equations that the reactances in the series and parallel networks will be of the same type. The design of impedance matching

networks using the network conversion approach will be illustrated using the following examples.

Design Example 1. Consider the design of a matching network to transform $Z_L = 10 + j25.1\ \Omega$ to 50 Ω [5]. Converting the impedance $10 + j25.1\ \Omega$ into an equivalent parallel network, we get

$$R_P = \frac{10^2 + 25.1^2}{10} \cong 73\ \Omega$$

Since R_P is greater than 50 Ω, we need to determine the value of X_S that gives $R_P = 50\ \Omega$. Therefore, we have to solve for X_S from the equation

$$50 = \frac{10^2 + X_S^2}{10}$$

Solving the above equation for X_s, we obtain $X_S = 20\ \Omega$. In order to obtain $X_S = 20\ \Omega$, we must reduce the effective X_S in the load impedance (Z_L) by connecting a capacitor having $-j5.1\ \Omega$ in series with the load. If this is done, then the equivalent parallel resistance is 50 Ω, and the equivalent parallel reactance is given by

$$jX_P = j\frac{10^2 + 20^2}{20} = j25\ \Omega$$

The inductive parallel reactance of $j25\ \Omega$ must then be canceled by adding a capacitive reactance of $-j25\ \Omega$ in parallel. The complete circuit is shown in Fig. 4.18. The values of the inductance and capacitance in the matching network can be calculated by using the design frequency.

Design Example 2. Consider the design of a matching network to transform $Z_L = 100 + j25.1\ \Omega$ to 50 Ω [5]. From the previous example we know that a series-to-parallel network conversion results in a larger parallel resistance. Therefore, here one possible solution is to cancel $j25.1\ \Omega$ by connecting a $-j25.1\ \Omega$ in series with it and then connecting a reactance (X_P) in parallel with

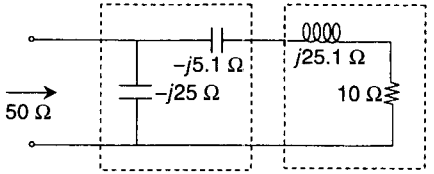


Figure 4.18 Matching network solution; Design Example 1 [5].

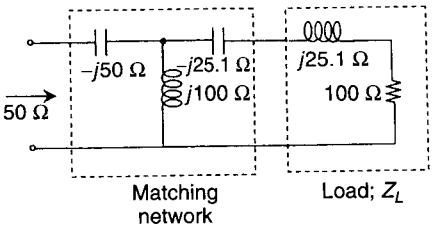


Figure 4.19 Matching network solution; Design Example 2 [5].

the resultant resistance (100 Ω) to obtain a series resistance (R_S) equal to 50 Ω. The resulting series reactance must then be canceled by connecting an opposite reactance in series with the network. The reactance X_P can be calculated by solving the equation

$$R_S = 50\ \Omega = \frac{100X_P^2}{100^2 + X_P^2}$$

Solving the above equation, we get $X_P = \pm 100\ \Omega$. Here, one could either choose an inductive reactance ($j100\ \Omega$) or a capacitive reactance ($-j100\ \Omega$). If we choose an inductive reactance, then the resulting equivalent series reactance and resistance due to the parallel combination of 100 Ω and $j100\ \Omega$ are given by

$$X_S = \frac{100^3}{100^2 + 100^2} = 50\ \Omega \qquad R_S = \frac{100^3}{100^2 + 100^2} = 50\ \Omega$$

The resulting series inductive reactance must then be canceled by a capacitor having a reactance of $-j50\ \Omega$. Then the input impedance of the network is just the series resistance of 50 Ω. The resulting network in this design is shown in Fig. 4.19.

The matching network component values can be determined in the usual manner by using the design frequency. The matching network in Fig. 4.19 has a high-pass frequency response. If we had chosen a capacitive reactance for X_P, then we would have a network with a low-pass frequency response.

4.2.3 Combined Lumped- and Distributed-Element Techniques

In the previous two sections, impedance matching techniques by means of exclusively distributed and lumped elements were described. In this section, techniques that make use of both distributed and lumped elements in the design of impedance matching networks will be illustrated with the help of examples.

Complex to Real Impedance Transformation

Design Example 1. Consider the problem of matching the load impedance $Z_L = 100 + j100\ \Omega$ to a 50-Ω transmission line. The impedance transformation

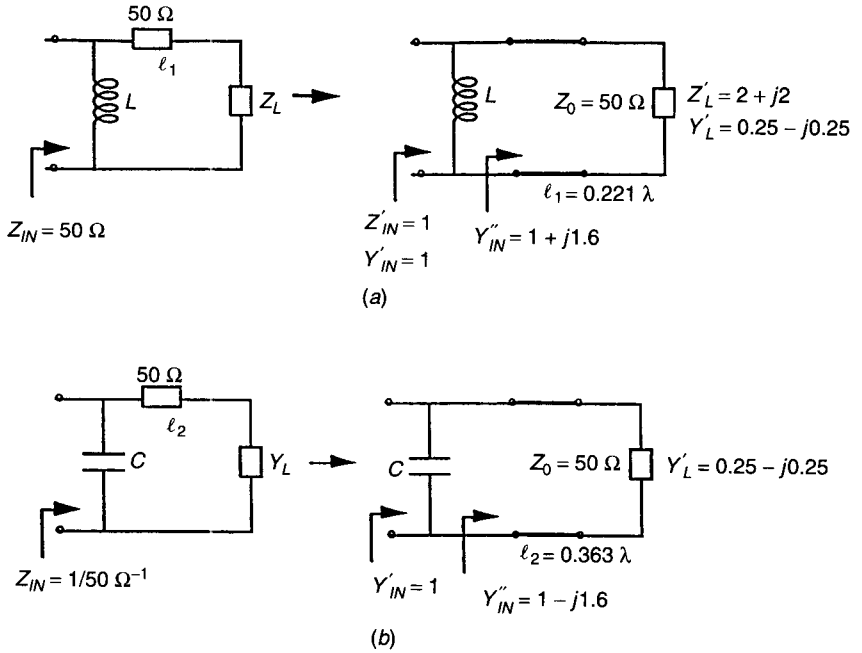


Figure 4.20 (*a, b*) Matching networks considered in Example 1. (*c*) Solutions to Example 1 problem.

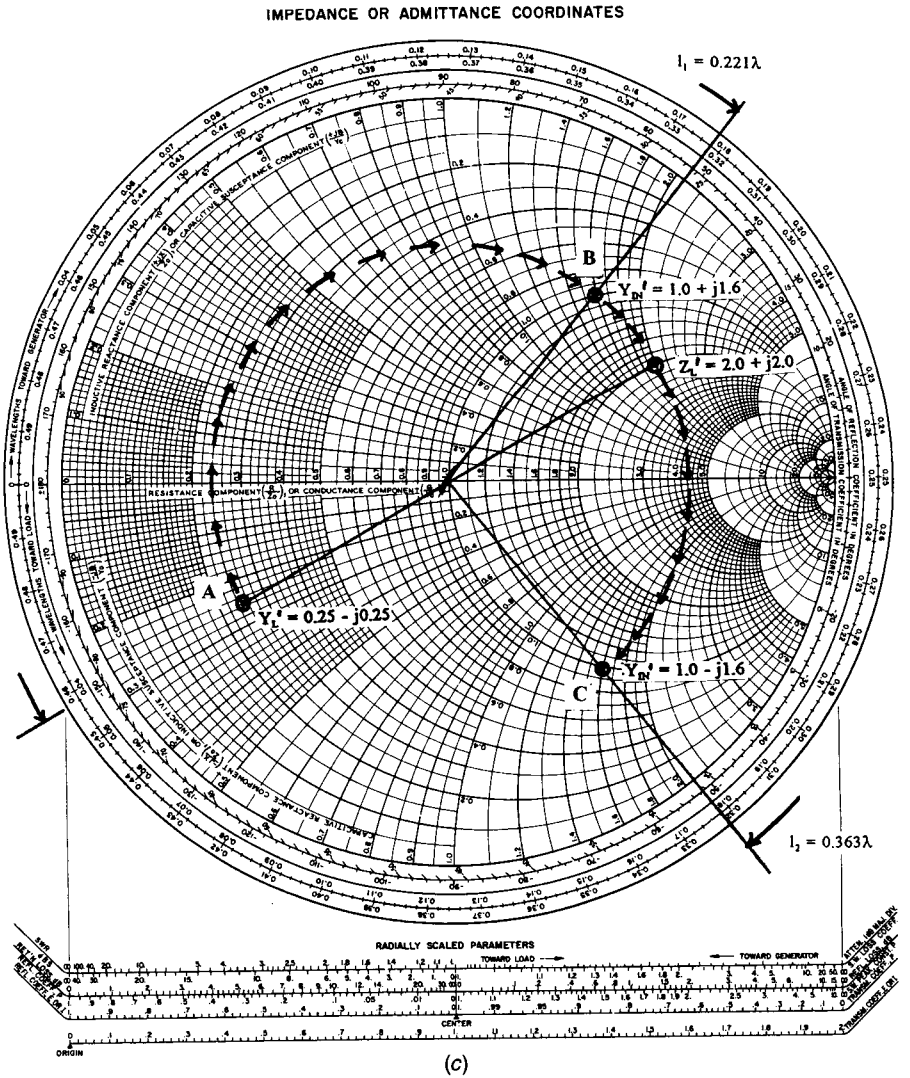
can be accomplished by the networks shown in Figs. 4.20*a* and *b* and as illustrated on the Smith chart in Fig. 4.20*c*.

At point A, Y'_L is first transformed to Y''_{IN} $(1 + j1.6)$ at point B on the Smith chart. Point B lies on the unit conductance circle. Point B is reached by traveling along the constant SWR circle toward the generator. This transformation is accomplished by means of a length l_1 (0.221λ) of a transmission line ($Z_0 = 50\ \Omega$). The shunt susceptance needed to cancel jB''_{IN} $(+j1.6)$ is obtained by a lumped inductor having an inductance $L = 50/2\pi f \times 1.6$ H, where f is the design frequency (in hertz). Now Y'_{IN} is unity.

If Y'_L is transformed to Y''_{IN} $(1 - j1.6)$ at point C on the Smith chart, then the match can be accomplished by connecting a lumped capacitor in shunt (to produce a susceptance of $+j1.6$), as shown in Fig. 4.20*b*. The capacitance C is given by $C = 1.6 \times 50/2\pi f$ farads. The length l_2 of the transmission line needed is 0.363λ.

Real to Complex Impedance Transformation

Design Example 2. Consider the problem of transforming R_L $(50\ \Omega)$ to Z_{IN} $(25 - j25\ \Omega)$. The design of the matching network using a distributed and



(*c*)

Figure 4.20 *(Continued)*

a lumped element is illustrated in Fig. 4.21*a*. The impedance transformation is accomplished by means of a quarter-wave transformer and a lumped capacitor in series.

A second design solution to the impedance transformation problem is shown in Fig. 4.21*b*. In this network, a 50-Ω transmission line of length l_1 and a lumped capacitor in shunt are used.

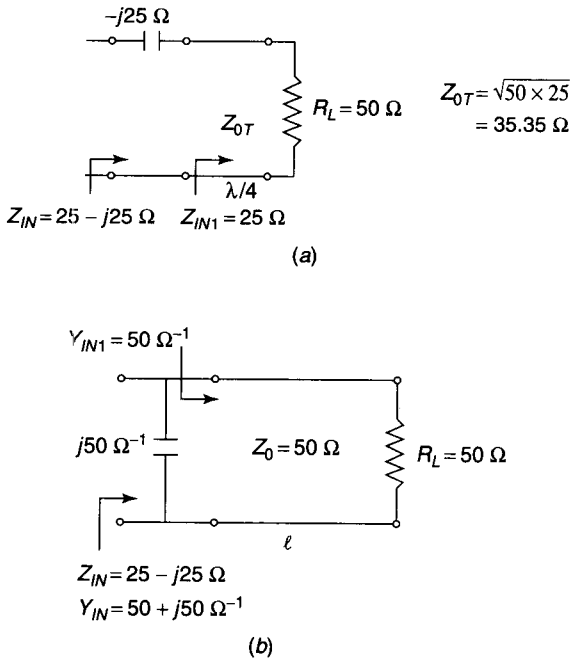


Figure 4.21 Matching network design using (*a*) quarter-wavelength transmission line and lumped capacitor in series (Design Example 2) and (*b*) transmission line and shunt capacitor (Design Example 2).

4.2.4 T- and π-Network Techniques

The impedance matching techniques presented in Sections 4.2.2 and 4.2.3 have dealt with *L*-section networks. When matching with only two elements, it is difficult to obtain a high loaded Q factor. Higher values of Q factor can be obtained by adding a third element to the network. The addition of the third element results in a T or π network, as shown in Fig. 4.22. The elements of the T and π networks can be realized in the form of (1) lumped reactive elements, (2) transmission-line elements, or (3) combination of lumped reactive and transmission-line elements. The procedures to design T and π matching net-

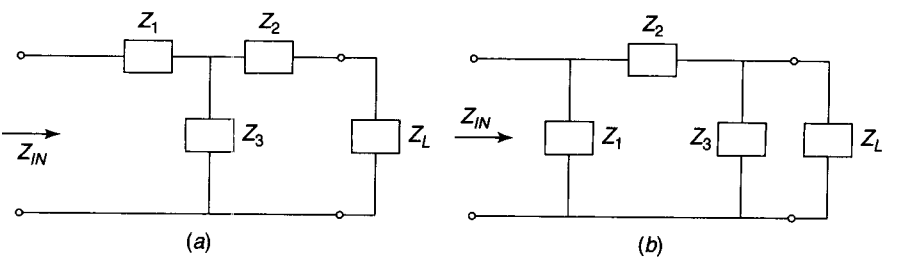


Figure 4.22 (*a*) T network; (*b*) π network.

works having a specified Q factor to transform Z_L to Z_{IN} (Fig. 4.22) are outlined elsewhere [4, 6].

4.3 WIDE-BAND TRANSFORMATION TECHNIQUES

In the previous section narrow-band impedance transformation techniques were presented. These techniques provided perfect impedance match (zero reflection coefficient) at a single frequency. The bandwidth of a matching network was defined based on the tolerable reflection coefficient away from the design frequency. The bandwidth can be increased using the techniques presented in Section 4.2 by transforming the impedances gradually by means of a small number of cascaded networks. However, in this section the classical wide-band impedance matching network design techniques will be presented with examples.

4.3.1 Bode–Fano Criterion

Before embarking on the study of wide-band impedance matching networks, it is advisable to understand the theoretical limits that constrain the performance of an impedance matching network. Consider the circuits shown in Fig. 4.23, where lossless networks are used to match arbitrary complex loads. A designer would raise the following general questions before starting to design a wide-band impedance matching network:

1. Is it possible to achieve perfect match (zero reflection coefficient) over a specified bandwidth?
2. If the answer to the previous question is negative, what is the optimum solution?
3. What is the trade-off between Γ_m, the maximum allowable reflection coefficient in the passband, and the bandwidth?

The foregoing questions can be answered by the Bode–Fano criterion [2, 9, 10], which gives for certain types of load impedances a theoretical limit on the minimum reflection coefficient magnitude that can be obtained with an arti-brary matching network. The Bode–Fano criterion thus gives the optimum solution that can be ideally achieved, even though such a solution may only be approximated in practice. Such an optimum solution is important because it gives the designer the upper limit of performance and provides a benchmark for comparisons of practical designs.

The Bode–Fano criterion will be explained here by considering Fig. 4.23*a*, in which a lossless network is used to match a parallel RC load impedance. The Bode–Fano criterion for this problem states that

$$\int_0^\infty \ln \frac{1}{|\Gamma(\omega)|}\, d\omega \leq \frac{\pi}{RC} \tag{4.8}$$

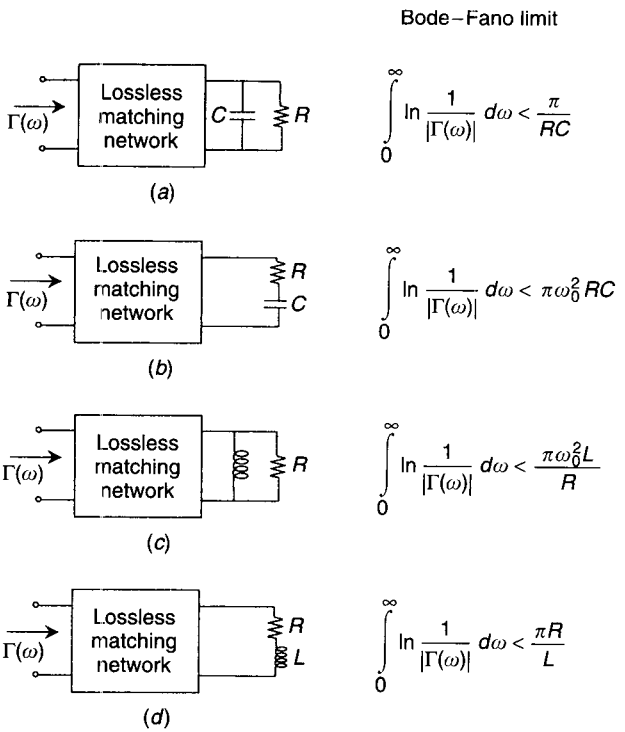


Figure 4.23 Bode–Fano limits for *RC* and *RL* loads: (*a*) parallel *RC*; (*b*) series *RC*; (*c*) parallel *RL*; (*d*) series *RL* [2].

where $\Gamma(\omega)$ is the reflection coefficient seen looking into the arbitrary matching network.

Assume that a designer wishes to synthesize a matching network with a reflection coefficient response as shown in Fig. 4.24. Substituting this function in (4.8) gives [2]

$$\int_0^\infty \ln \frac{1}{|\Gamma|}\, d\omega = \int_{\Delta\omega} \frac{1}{\Gamma_m}\, d\omega = \Delta\omega \ln \frac{1}{\Gamma_m} \leq \frac{\pi}{RC} \tag{4.9}$$

From (4.9) the following conclusions can be arrived at [2]:

(a) For a given load (fixed *RC* product), a wider bandwidth ($\Delta\omega$) can be obtained only at the expense of a higher reflection coefficient (Γ_m) in the passband.

(b) The passband reflection coefficient (Γ_m) cannot be zero unless the bandwidth ($\Delta\omega$) is zero. Thus a perfect match (zero reflection coefficient) can be accomplished only at a finite number of frequencies.

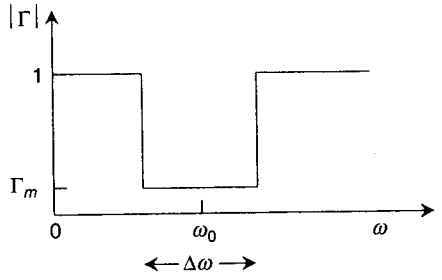


Figure 4.24 Illustration of Bode–Fano criterion; optimum response: $\Delta\omega$ = passband; Γ_m = maximum allowable reflection coefficient in passband; ω_0 = center frequency of passband [2].

(c) As R or C increases, the quality of match ($\Delta\omega$ or $1/\Gamma_m$) must decrease. Thus higher Q circuits are more difficult to match than are lower Q circuits.

Since $\ln(1/|\Gamma|)$ in (4.8) is proportional to the return loss (in decibels), it is evident from (4.8) that the area between the return loss curve and the $|\Gamma| = 1$ (0 dB return loss) line in Fig. 4.24 must be less than or equal to a constant [2]. The optimization of the matching network design then requires the adjustment of the return loss curve in Fig. 4.24 such that $|\Gamma| = \Gamma_m$ over the passband and $|\Gamma| = 1$ elsewhere. The response shown in Fig. 4.24 is therefore the optimum response, which can only be approached in practice.

The ideal response in Fig. 4.24 cannot be obtained in practice because it would require an infinite number of elements in the matching network. It can be approximated, however, with a reasonably small number of elements as alluded to in the introduction to Section 4.3. The response of a Chebyshev transformer described in the following section can be viewed as a close approximation to the ideal response of Fig. 4.24 when the ripple of the Chebyshev response is made equal to Γ_m. In a Chebyshev response, perfect match (zero reflection coefficient) is achieved at a finite number of frequencies in the passband, as illustrated in the following section.

Application of Bode–Fano Criterion. The Bode–Fano limits for four canonical types of loads are listed in Fig. 4.23. In order to apply the Bode–Fano criterion and determine the upper limit of performance of the matching network, the designer must first determine the equivalent circuit model of the load impedance to be matched. The equivalent circuit model of the load must conform to one of the four types of loads shown in Fig. 4.23. The elements in the equivalent circuit model of the load (the input or output impedance of a device) may exhibit frequency dependency. In such a situation the optimal reflection coefficient magnitude response derived by applying the Bode–Fano limit will also be frequency dependent.

After obtaining the equivalent circuit model of the load, the designer must apply the appropriate Bode–Fano limit (Fig. 4.23) for a specified center frequency (ω_0), bandwidth ($\Delta\omega$), and maximum reflection coefficient in the pass-

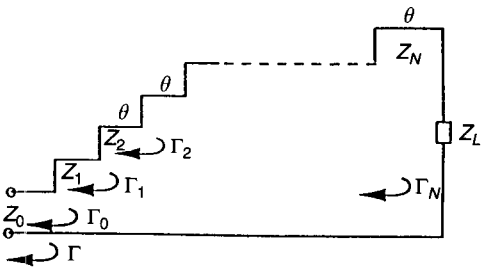


Figure 4.25 Multisection quarter-wave transformer: N = number of sections.

band (Γ_m) and determine the optimal reflection coefficient magnitude versus frequency response as shown before. This response then serves as a goal, an upper limit of performance, to be reached by optimizing the matching network design.

4.3.2 Multisection Quarter-Wave Transformers

The schematic of a multisection quarter-wave transformer is shown in Fig. 4.25, in which the load impedance Z_L is transformed to Z_0, the characteristic impedance of the feed line. The load Z_L is assumed to be real and may be greater or smaller than Z_0. The transformer consists of discrete sections of lines having different characteristic impedances but the same electrical length, $\beta l = \theta$. The electrical length will be a quarter wavelength at the band center frequency f_0. The transformer is symmetrical; therefore, the reflection coefficients at the junctions of the sections are related as $|\Gamma_0| = |\Gamma_N|$, $|\Gamma_1| = |\Gamma_{N-1}|, \ldots$.

In this section, two widely known multisection quarter-wave transformers, namely, binomial and Chebyshev, will be presented. The design equations and data based on approximate as well as exact theories of multisection quarter-wave transformers will be given. The design procedures will be illustrated by means of examples.

Binomial Transformer. In a binomial transformer, the function $\Gamma(\theta)$ is chosen such that $|\Gamma| = \rho$ and the first $N - 1$ derivatives (N is the number of sections) with respect to frequency (or θ) vanish at the band center frequency f_0, where $\theta = \frac{1}{2}\pi$. Therefore, a maximally flat passband characteristic for the transformer is obtained. The maximally flat passband characteristic of a binomial transformer is shown in Fig. 4.26. The maximum tolerable reflection coefficient in the passband is given as ρ_m. The transformer sections will be a quarter-wavelength long at the band center frequency. The parameter θ_m (Fig. 4.26) is given by

$$\theta_m = \cos^{-1}\left|\frac{2\rho_m}{\ln(Z_L/Z_0)}\right|^{1/N} \tag{4.10}$$

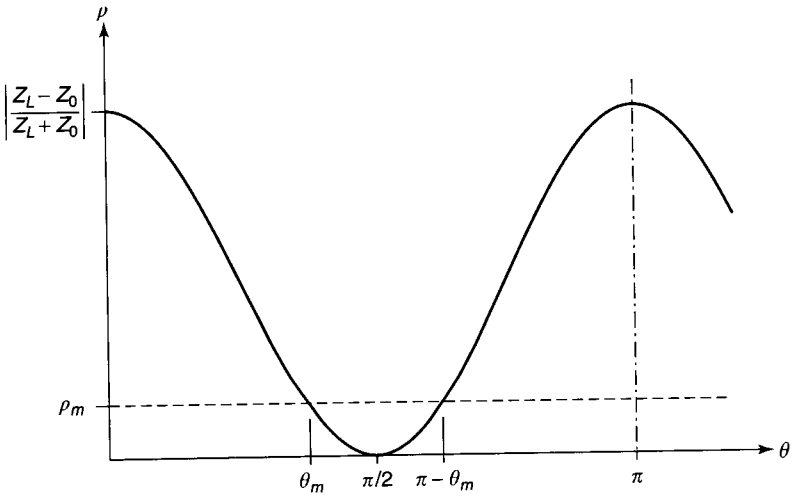


Figure 4.26 Characteristic of binomial transformer.

In the case of transmission-line sections, $\theta = \pi f/2f_0$, and therefore the fractional bandwidth is given by

$$\frac{\Delta f}{f_0} = \frac{2(f_0 - f_m)}{f_0} = 2 - \frac{4}{\pi} \cos^{-1} \left| \frac{2\rho_m}{\ln(Z_L/Z_0)} \right|^{1/N} \tag{4.11}$$

since $\theta_m = \pi f_m/2f_0$. Note that the solution to (4.10) is chosen such that $\theta_m < \frac{1}{2}\pi$. The maximum out-of-band reflection coefficient occurs for $\theta = 0\pi$, and is given by

$$\rho_{\max} = \left| \frac{Z_L - Z_0}{Z_L + Z_0} \right| \tag{4.12}$$

In (4.12), $\rho_{\max}$ is the reflection coefficient when Z_L is directly connected to the transmission line having the characteristic impedance Z_0.

Design Based on Approximate Theory. The impedances of the sections may be calculated using the equations

$$\ln \frac{Z_{n+1}}{Z_n} = 2\rho_n = 2^{-N} C_n^N \ln \frac{Z_L}{Z_0} \tag{4.13a}$$

$$C_n^N = \frac{N!}{(N-n)!n!} \tag{4.13b}$$

where Z_n and Z_{n+1} are the impedances of the nth and $(n+1)$th sections,

respectively, ρ_n is the reflection coefficient at the junction between Z_n and Z_{n+1}, and C_n^N are the binomial coefficients. Since the theory upon which the equations are based in this section is approximate, the range of Z_L is limited to about $0.5Z_0 < Z_L < 2Z_0$.

Design Example. Consider the design of a two-section transformer to match $Z_L = 100\ \Omega$ to $Z_0 = 50\ \Omega$. The required ρ_m is 0.05. The transformer is to be realized in the form of air-filled coaxial lines. The band center frequency is 4 GHz. From (4.13a),

$$\ln \frac{Z_1}{Z_0} = \frac{1}{4} \ln \frac{Z_L}{Z_0} \quad \text{since } C_0^2 = 1$$

$$\ln \frac{Z^2}{Z_1} = \frac{1}{2} \ln \frac{Z_L}{Z_0} \quad \text{since } C_1^2 = 2$$

Therefore, we have

$$Z_1 = Z_L^{1/4} Z_0^{3/4} = (100)^{1/4}(50)^{3/4} = 59.45\ \Omega$$

$$Z_2 = Z_L^{3/4} Z_0^{1/4} = (100)^{3/4}(50)^{1/4} = 84.09\ \Omega$$

Since $\rho_m = 0.05$, from (4.11), the fractional bandwidth $(\Delta f/f_0)$ is 0.496 and the absolute bandwidth is 1.98 GHz. From (4.12), $\rho_{\max} = 0.33$. The length of each quarter-wavelength section of the transformer is 18.75 mm. A schematic diagram of the transformer is shown in Fig. 4.27.

Design Based on Exact Theory. Exact results can be obtained by solving transmission-line equations for each section numerically. Table 4.1 lists the exact impedance data for two, three, and four sections [3]. The table lists data for $Z_L/Z_0 > 1$. For $Z_L/Z_0 < 1$, the results for Z_0/Z_L should be used, with Z_1 beginning from the load end. This is because of the symmetry in the transformer; therefore, the transformer that matches Z_L to Z_0 can be reversed and used to match Z_0 to Z_L.

Chebyshev Transformer. In a Chebyshev transformer, ρ is allowed to vary between 0 and ρ_m in a periodic manner in the passband. Thus the transformer has an equal-ripple characteristic, as shown in Fig. 4.28. A transformer of this type provides a considerable improvement in bandwidth over the binomial

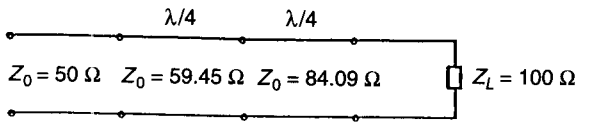


Figure 4.27 Two-section binomial transformer.

Table 4.1 Binomial Transformer Design Data

	$N = 2$		$N = 3$			$N = 4$			
Z_L/Z_0	Z_1/Z_0	Z_2/Z_0	Z_1/Z_0	Z_2/Z_0	Z_3/Z_0	Z_1/Z_0	Z_2/Z_0	Z_3/Z_0	Z_4/Z_0
1.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.5	1.1067	1.3554	1.0520	1.2247	1.4259	1.0257	1.1351	1.3215	1.4624
2.0	1.1892	1.6818	1.0907	1.4142	1.8337	1.0444	1.2421	1.6102	1.9150
3.0	1.3161	2.2795	1.1479	1.7321	2.6135	1.0718	1.4105	2.1269	2.7990
4.0	1.4142	2.8285	1.1707	2.0000	3.3594	1.0919	1.5442	2.5903	3.6633
6.0	1.5651	3.8336	1.2544	2.4495	4.7832	1.1215	1.7553	3.4182	5.3500
8.0	1.6818	4.7568	1.3022	2.8284	6.1434	1.1436	1.9232	4.1597	6.9955
10.0	1.7783	5.6233	1.3409	3.1623	7.4577	1.1613	2.0651	4.8424	8.6110

transformer. Further, it should be noted that a Chebyshev transformer provides perfect match (zero reflection coefficient) at finite number of frequencies in the passband (Fig. 4.28), as was stated in Section 4.3.1. The ripple in the passband is equal to the maximum allowable reflection coefficient (ρ_m). Hence the Chebyshev response can be considered as a close approximation to the ideal response (based on the Bode–Fano criterion) shown in Fig. 4.24.

The equal-ripple characteristic is obtained by forcing ρ to behave according to a Chebyshev polynomial, thus the name Chebyshev transformer. The coefficient ρ can be made to vanish at as many frequencies in the passband as there are sections of transformer.

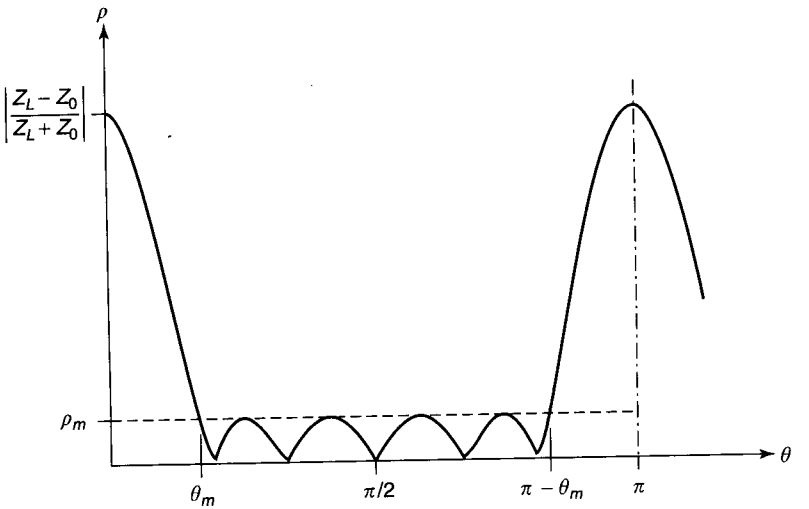


Figure 4.28 Characteristic of Chebyshev transformer.

The reflection coefficient Γ is given by

$$\Gamma = 2e^{-jN\theta}[\rho_0 \cos N\theta + \rho_1 \cos(N-2)\theta + \cdots + \rho_n \cos(N-2n)\theta + \cdots]$$
$$= Ae^{-jN\theta} T_N(\sec \theta_m \cos \theta) \tag{4.14}$$

where $\rho_0, \rho_1, \ldots, \rho_N$ are the magnitudes of the reflection coefficients at the junctions (see Fig. 4.25) and A is a constant to be determined; T_N (sec θ_m cos θ) is a Chebyshev polynomial of the Nth degree. The last term in the series in (4.14) is $\rho_{(N-1)/2} \cos \theta$ for N odd and $\frac{1}{2}\rho_{N/2}$ for N even. The electrical length of the transformer sections is $\frac{1}{2}\pi$ at the band center frequency f_0. Therefore, the length of the transformer sections is $\frac{1}{4}\lambda$ at the band center frequency.

Design Based on Approximate Theory. The design equations based on approximate theoretical analysis are given below. These equations are valid if Z_L is restricted to about $0.5Z_0 < Z_L < 2Z_0$:

$$\Gamma = \frac{1}{2} e^{-jN\theta} \ln \frac{Z_L}{Z_0} \frac{T_N(\sec \theta_m \cos \theta)}{T_N(\sec \theta_m)} \tag{4.15}$$

$$\rho_m = \frac{\ln(Z_L/Z_0)}{2T_N(\sec \theta_m)} \tag{4.16a}$$

$$\sec \theta_m = \cos\left[\frac{1}{N} \cos^{-1} \frac{\ln(Z_L/Z_0)}{2\rho_m}\right] \tag{4.16b}$$

$$\ln \frac{Z_{n+1}}{Z_n} = 2\rho_n \tag{4.17}$$

It is evident from (4.16) that if the passband, and hence θ_m, is specified, the passband tolerance ρ_m is fixed and vice versa.

The fractional bandwidth of the transformer is given by

$$\frac{\Delta\theta}{\pi/2} = \frac{\Delta f}{f_0} = \frac{4}{\pi}\left(\frac{\pi}{2} - \theta_m\right) \tag{4.18}$$

The maximum out-of-band reflection coefficient, $\rho_{\max}$, occurs at $\theta = 0, \pi$ and is given by

$$\rho_{\max} = \left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| \tag{4.19}$$

The expressions for $T_n(\sec \theta_m \cos \theta)$ useful in designing transformers up to four sections are given below:

$$T_1(\sec \theta_m \cos \theta) = \sec \theta_m \cos \theta \tag{4.20a}$$

$$T_2(\sec \theta_m \cos \theta) = \sec^2 \theta_m(1 + \cos 2\theta) - 1 \tag{4.20b}$$

$$T_3(\sec\theta_m \cos\theta) = \sec^3\theta_m(\cos 3\theta + 3\cos\theta) - 3\sec\theta_m\cos\theta \tag{4.20c}$$

$$T_4(\sec\theta_m \cos\theta) = \sec^4\theta_m(\cos 4\theta + 4\cos 2\theta + 3) - 4\sec^2\theta_m(\cos 2\theta + 1) + 1 \tag{4.20d}$$

Expressions for $T_n(\sec\theta_m \cos\theta)$ for degrees greater than 4 can be derived using the following equations:

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x) \tag{4.21a}$$

$$T_n(x) = \begin{cases} \cos(n\cos^{-1}x) & \text{for } |x| < 1 \quad (4.21b) \\ \cosh(n\cosh^{-1}x) & \text{for } |x| > 1 \quad (4.21c) \end{cases}$$

$$(\cos\theta)^n = 2^{-n}e^{-jn\theta}(1 + e^{2j\theta})^n = 2^{-n}e^{-jn\theta}\sum_{m=0}^{n} C_m^n e^{jm\theta}$$
$$= 2^{-n+1}[C_0^n\cos n\theta + C_1^n\cos(n-2)\theta + \cdots + C_m^n\cos(n-2m)\theta + \cdots] \tag{4.21d}$$

The last term in (4.21d) is $\frac{1}{2}C_{n/2}^n$ for n even and $C_{(n-1)/2}^n\cos\theta$ for n odd. It can be shown that the constant A in (4.14) is equal to ρ_m. For a given ρ_m, Z_L, and Z_0, $T_N(\sec\theta_m)$ can be found from (4.16a). In (4.14), ρ_n can be determined by using the expressions for $T_N(\sec\theta_m\cos\theta)$ in (4.20) and equating similar terms of the form $\cos(N-2n)\theta$ in (4.14). Then Z_n can be found using (4.17).

Alternatively, $T_N(\sec\theta_m)$ and θ_m can be determined using the equations

$$T_N(\sec\theta_m) = \frac{1}{\rho_m}\left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| \tag{4.22a}$$

$$\sec\theta_m = \cosh\left[\frac{1}{N}\cosh^{-1}\left(\frac{1}{\rho_m}\left|\frac{Z_L - Z_0}{Z_L + Z_0}\right|\right)\right] \tag{4.22b}$$

Then ρ_n in (4.14) can be determined by using expressions for $T_N(\sec\theta_m\cos\theta)$ in (4.20) and equating similar terms of the form $\cos(N-2n)\theta$ in (4.14). Then Z_n can be found using the equations

$$\rho_0 = \frac{Z_1 - Z_0}{Z_1 + Z_0} \tag{4.22c}$$

$$\rho_n = \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} \tag{4.22d}$$

$$\rho_N = \frac{Z_L - Z_N}{Z_L + Z_N} \tag{4.22e}$$

Design Example. Consider the design of a two-section Chebyshev transformer to match impedance $Z_0 = 50\ \Omega$ to $Z_L = 100\ \Omega$. The maximum passband tolerance required is $\rho_m = 0.05$.

Using (4.16) and (4.20b), we obtain

$$T_2(\sec\theta_m) = 2\sec^2\theta_m - 1 = \frac{\ln(100/50)}{2(0.05)} = 6.93$$

Therefore, $\sec\theta_m = 1.99$ and $\theta_m = 1.045$. From (4.18) the fractional bandwidth is

$$\frac{\Delta f}{f_0} = 0.669$$

From (4.14), (4.15), and (4.20b) we obtain

$$\begin{aligned} 2\rho_0\cos 2\theta + \rho_1 &= \rho_m T_2(\sec\theta_m\cos\theta) \\ &= \rho_m\sec^2\theta_m\cos\theta + \rho_m(\sec^2\theta_m - 1) \end{aligned}$$

Therefore, we have

$$\rho_0 = \tfrac{1}{2}\rho_m\sec^2\theta_m = \rho_2 = 0.099 \qquad \rho_1 = \rho_m(\sec^2\theta_m - 1) = 0.148$$

The impedances of the sections can be calculated using (4.17). Thus,

$$Z_1 = e^{2\rho_0}Z_0 = 60.95\ \Omega \qquad Z_2 = e^{2\rho_1}Z_1 = 81.95\ \Omega$$

The schematic of the transformer is shown in Fig. 4.29.

Design Based on Exact Theory. The equations and data necessary to design Chebyshev transformers based on exact theory are given here. The design data for a three-section transformer are given. The power loss ratio P_{LR} for the transformer is given by [3]

$$P_{LR} = \frac{1}{1-\rho^2} \tag{4.23}$$

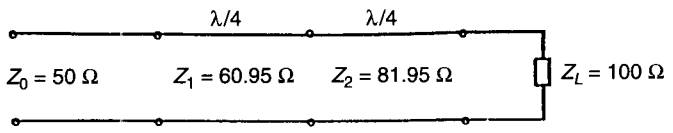


Figure 4.29 Two-section Chebyshev transformer.

where ρ is the magnitude of the reflection coefficient and P_{LR} can be expressed in the form

$$P_{LR} = 1 + Q_{2N}(\cos\theta) \tag{4.24}$$

where $Q_{2N}(\cos\theta)$ is an even polynomial of degree $2N$ in $\cos\theta$ having coefficients that are functions of the impedances Z_n. For a Chebyshev transformer, P_{LR} is specified to be

$$P_{LR} = 1 + k^2 T_N^2(\sec\theta_m \cos\theta) \tag{4.25}$$

where k^2 is the passband tolerance on P_{LR}. The maximum value of P_{LR} in the passband is $1 + k^2$.

By equating (4.24) and (4.25), the equations that can be solved for the various characteristic impedances are obtained. The passband tolerance k^2 and maximum tolerable passband ripple ρ_m are related by

$$\rho_m = \left(\frac{k^2}{1+k^2}\right)^{1/2} \tag{4.26}$$

From (4.26) for a specified k^2, ρ_m can be determined and vice versa. For a specified value of k^2 and the number of sections in the transformer, θ_m can be determined. The fractional bandwidth of the transformer is given by

$$\frac{\Delta f}{f_0} = \frac{2(\pi/2 - \theta_m)}{\pi/2} \tag{4.27}$$

The Chebyshev quarter-wave transformer design data for three sections are given in Table 4.2 [3]. More extensive tables can be found elsewhere [11]. Using Table 4.2, for a given Z_L/Z_0, k^2, and bandwidth, the impedances of the sections can be determined. The design procedures are illustrated by the following example.

Table 4.2 Chebyshev Transformer Design Data

	$\Delta f/f_0 = 0.2$		$\Delta f/f_0 = 0.4$		$\Delta f/f_0 = 0.6$	
Z_L/Z_0	Z_1/Z_0	k^2	Z_1/Z_0	k^2	Z_1/Z_0	k^2
2	1.09247	1.19×10^{-7}	1.09908	7.98×10^{-6}	1.1083	9.57×10^{-5}
4	1.19474	5.35×10^{-7}	1.20746	3.55×10^{-5}	1.23087	4.31×10^{-4}
10	1.349	1.92×10^{-7}	1.37482	1.28×10^{-4}	1.4232	1.55×10^{-3}
20	1.48359	4.29×10^{-7}	1.52371	2.85×10^{-4}	1.60023	3.45×10^{-3}
100	1.87411	2.33×10^{-6}	1.975	1.55×10^{-3}	2.17928	1.87×10^{-2}

Note: $Z_2 = \sqrt{Z_L Z_0}$; $Z_3 = Z_L Z_0 / Z_1$.

Design Example. Consider the design of a three-section Chebyshev transformer to match $Z_L = 100\ \Omega$ to $Z_0 = 50\ \Omega$ over a fractional bandwidth of 0.6 and having a maximum voltage standing-wave ratio (VSWR) of 1.02 in the passband. From the given maximum VSWR, $\rho_m = (1.02 - 1)/(1.02 + 1) = 9.9 \times 10^{-3}$, and from (4.26),

$$k^2 = \frac{\rho_m^2}{1 - \rho_m^2} = 9.8 \times 10^{-5}$$

In this design problem, $Z_L/Z_0 = 2$. From Table 4.2 for $Z_L/Z_0 = 2$ and $\Delta f/f_0 = 0.6$, $k^2 = 9.57 \times 10^{-5}$. This value of k^2 is less than what we need and hence can be accepted. If an acceptable value of k^2 cannot be obtained for the chosen number of sections, the number of sections must be increased.

From Table 4.2, we get $Z_1/Z_0 = 1.1083$. Therefore,

$$Z_1 = 1.1083 \times 50 = 55.42\ \Omega$$

$$Z_2 = \sqrt{Z_L Z_0} = \sqrt{100 \times 50} = 70.71\ \Omega$$

$$Z_3 = Z_L Z_0 / Z_1 = \frac{100 \times 50}{55.42} = 90.22\ \Omega$$

The schematic of the transformer is shown in Fig. 4.30. The sections are a quarter-wavelength long at the band center frequency. The guide wavelength in the sections may be different as in microstrip transformers. In a microstrip transmission line the impedance and wavelength are functions of the microstrip width–substrate height ratio. The maximum out-of-band reflection coefficient for this transformer is

$$\rho_{\max} = \left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| = \left|\frac{100 - 50}{100 + 50}\right| = 0.33$$

Additional Remarks. In the realization of multisection quarter-wave transformers, the designer has to contend with the discontinuities in the cross-sectional dimensions of the sections. These discontinuities have to be modeled, and the lengths of the sections should be appropriately adjusted. If the discontinuities are not compensated for in the design, the frequency response of the reflection coefficient will be modified. For example, at an abrupt change in

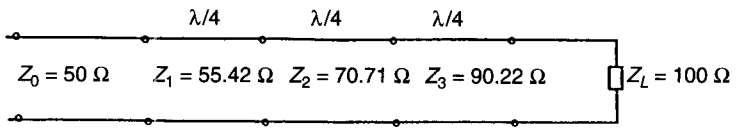


Figure 4.30 Schematic of three-section Chebyshev transformer.

the microstrip width, there will be additional fringing electric fields. The effect of the fringing field can be modeled as a shunt capacitance at the junction. The equivalent circuit of a step in the width of a microstrip line consists of an ideal impedance step and a shunt capacitive susceptance. The junction capacitance can be compensated for by changing the length of each section. The junction capacitance for a step change in the width of a microstrip line can be evaluated [3].

Another important fact that modifies the performance of the transformer is the dispersion in the lines constituting the sections. The effects of dispersion must be taken into account in the design of the transformer.

Finally, it should be noted that for a given number of sections and a tolerable value of ρ (i.e., ρ_m) in the passband, the fractional bandwidth provided by the Chebyshev transformer is greater than that of the binomial transformer. A binomial design gives a maximally flat frequency response with a zero reflection coefficient at the band center frequency irrespective of the number of sections, unlike a Chebyshev transformer. A Chebyshev transformer will have fewer sections than a binomial transformer for a given ρ_m and fractional bandwidth. Therefore, the Chebyshev design is considered to be the optimum multisection quarter-wave impedance transformer design. The bandwidths of the Chebyshev and binomial designs are compared in Pozar [2] and Collin [3].

4.3.3 Tapered Transmission-Line Transformers

In multisection quarter-wave transformers, the impedance transformation is obtained by changing the impedance in discrete steps. An alternative technique is to change the impedance level by using a tapered transition in which the characteristic impedance of the line varies continuously.

The schematic of a tapered transmission-line matching unit is shown in Fig. 4.31*a*, in which the normalized load impedance Z_L' is matched to a feed line having a normalized impedance, $Z_c' = 1$. The impedance of the taper $Z'(z)$ varies with distance z along the impedance taper. Figure 4.31*b* illustrates that the taper can be considered to be made up of a very large number of sections, each having an infinitesimally small length, dz, and an impedance step, dZ', from section to section.

From the approximate theory (small reflections) of multisection quarter-wave transformers, the input reflection coefficient (Γ_{in}) can be shown to be [3]

$$\Gamma_{in} = \frac{1}{2}\int_0^L e^{-j\beta z}\frac{d}{dz}(\ln Z')\,dz \tag{4.28}$$

where L is the length of the taper.

For a given variation of Z' with z, Γ_{in} can be evaluated using (4.28). Three types of practical tapers will be presented in the following sections.

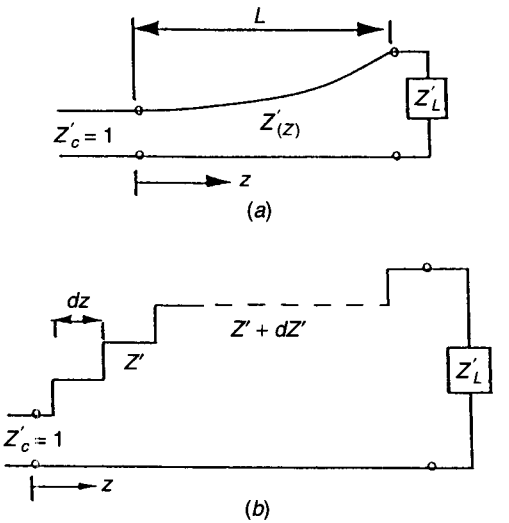


Figure 4.31 Tapered transmission-line transformer.

Exponential Taper. In exponential tapers, Z' varies exponentially with z and is given by

$$Z' = e^{(z/L)\ln Z'_L} \tag{4.29}$$

and Γ_{in} is given by

$$\Gamma_{in} = \tfrac{1}{2} e^{-j\beta L} \ln Z'_L \frac{\sin \beta L}{\beta L} \tag{4.30}$$

where it is assumed that the transmission line is a TEM line in which β is not a function of z. A plot of ρ_{in} ($|\Gamma_{in}|$) as a function of βL is shown in Fig. 4.32. For a given length of taper, Fig. 4.32 depicts the plot of ρ_{in} versus frequency because in a TEM line β is proportional to frequency.

Triangular Distribution Taper. The value $d(\ln Z')/dz$ is chosen as a triangular function of z in a triangular distribution taper. Therefore,

$$\frac{d(\ln Z')}{dz} = \begin{cases} \dfrac{4z}{L^2} \ln Z'_L & 0 \le z \le \dfrac{L}{2} \\ \dfrac{4}{L^2}(L - z) \ln Z' & \dfrac{L}{2} \le z \le L \end{cases} \tag{4.31}$$

and Γ_{in} is given by,

$$\Gamma_{in} = \frac{1}{2} \ln Z'_L \left[\frac{\sin(\beta L/2)}{\beta L/2}\right]^2 e^{-j\beta L}. \tag{4.32}$$

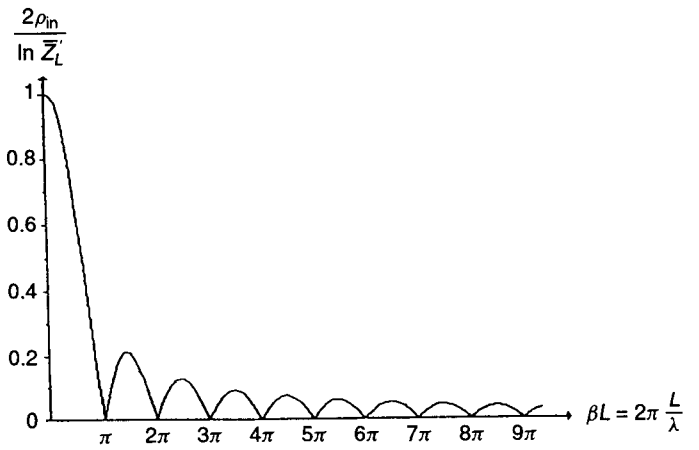


Figure 4.32 Reflection coefficient characteristic of exponential taper.

A plot of ρ_{in} versus βL is depicted in Fig. 4.33. It is evident, by comparison with Fig. 4.32, that the triangular distribution taper has a first minor lobe peak value less than that for the exponential taper. However, this minor lobe peak occurs for a length of nearly $\frac{3}{2}\lambda$.

Chebyshev Taper. The Chebyshev taper is obtained by increasing the number of sections in a Chebyshev transformer indefinitely while keeping the length L of the transformer fixed. The input reflection coefficient is given by

$$\Gamma_{in} = \tfrac{1}{2} e^{-j\beta L} \ln Z_L' \frac{\cos L\sqrt{\beta^2 - \beta_0^2}}{\cosh \beta_0 L} \tag{4.33}$$

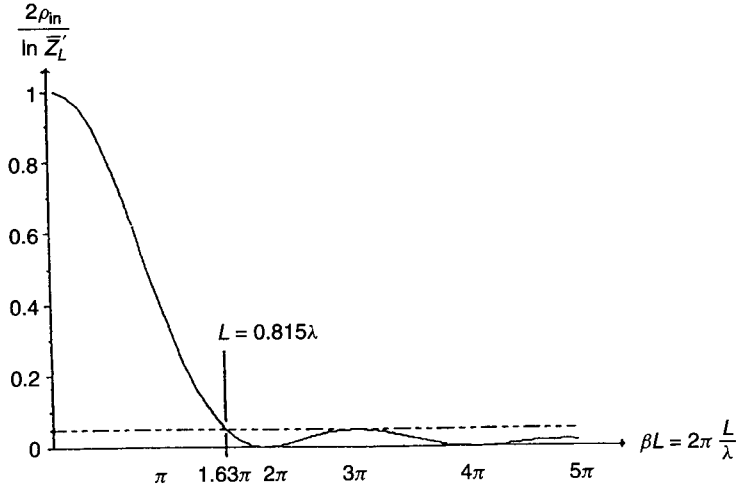


Figure 4.33 Reflection coefficient characteristic of triangular taper.

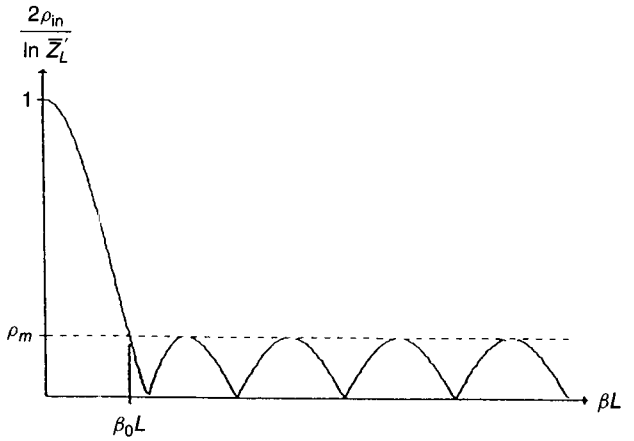


Figure 4.34 Reflection coefficient characteristic of Chebyshev taper transformer.

where β_0 is the value of β at the lower edge of the passband, as shown in Fig. 4.34. The Chebyshev taper has an equal-ripple characteristic in the passband and is an optimum design because it gives the smallest ripple amplitude for a fixed taper length, and, conversely, for a specified ripple amplitude it has the shortest length.

The exact theory of tapered transmission lines and their syntheses procedures can be found elsewhere [3]. In Pozar [2] the designs of three types of tapered transmission line transformers are illustrated by means of examples. The variation of impedance with distance along the taper as well as the frequency response of the reflection coefficient are also presented in Pozar [2].

4.3.4 Lumped- and Distributed-Element Matching Networks

The synthesis of wide-band matching networks (distributed as well as lumped element) having a prescribed frequency response of the transmission coefficient (S_{21}) will be presented in this section. The need for such matching networks arises in wide-band amplifier designs.

Consider a two-stage microwave FET amplifier schematic, shown in Fig. 4.35. By a proper choice of the transistors and matching network characteristics, the amplifier may be designed for (a) high output power, (b) low noise, (c) high gain, or (d) a combination of these requirements over a specified frequency range. The output, input, and interstage matching networks may also be required to have a prescribed gain (S_{21}) slope over the given frequency range. The following requirements in the design are commonly encountered:

1. Z_L (usually 50 Ω) to Z_{LP} (or Γ_{LP}) for high power—real to complex transformation;
2. Z_{I2} (or Γ_{I2}) to Z_{01} (Γ_{01}) for maximum power transfer—complex to complex transformation;

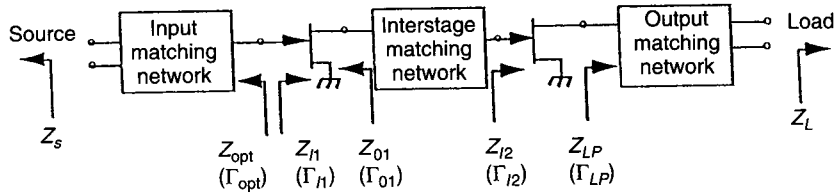


Figure 4.35 Two-stage microwave amplifier schematic.

3. Z_{I1} (Γ_{I1}) to Z_S (usually 50 Ω) for maximum power transfer—complex to real transformation; and
4. Z_S to Z_{opt} (Γ_{opt}) for low-noise figure—real to complex transformation.

The matching networks may also be required to have a prescribed $|S_{21}|$-versus-frequency characteristic; either a zero slope or an appropriate positive slope (to compensate for the transistor gain roll-off). Generally a zero-gain slope is used to obtain low VSWR at input and output. A positive-gain slope for an interstage network is necessary to obtain a flat gain versus frequency response for the amplifier. The design of wide-band amplifiers and the criteria for selection of appropriate matching network characteristics are discussed extensively elsewhere [12]. The matching networks in Fig. 4.35 can be designed using lumped elements, distributed elements, or a combination of lumped and distributed elements.

Extensive work has been done in the area of synthesis of wide-band lumped and distributed matching networks. A method known as the real-frequency technique has been developed [13, 14, 15]. Synthesis of networks having equal-ripple characteristics has been presented [16]. A method known as the transformation-Q approach has also been proposed [17, 18]. Here, the real-frequency and transformation-Q techniques will be briefly reviewed.

Real-Frequency Technique. This is a numerical method and only utilizes real-frequency (e.g. measured) load impedance data. Neither a model or analytic impedance function for the load nor a matching network topology or analytic form of the network transfer function is required. The real-frequency technique is further divided into two approaches: (a) line segment approach [13] and (b) reflection coefficient approach [15]. The line segment approach is a single-matching technique that involves synthesis of networks to transform complex to real impedances and vice versa; the reflection coefficient approach is a double-matching technique that enables transformation of complex to complex impedances.

A commercial matching network synthesis software known as MATCHNET [19] uses the real-frequency technique to synthesize low-pass, high-pass, and bandpass lumped- as well as distributed-element matching networks having a prescribed gain ($|S_{21}|$) slope. Figures 4.36*a* and *b* show schematics of

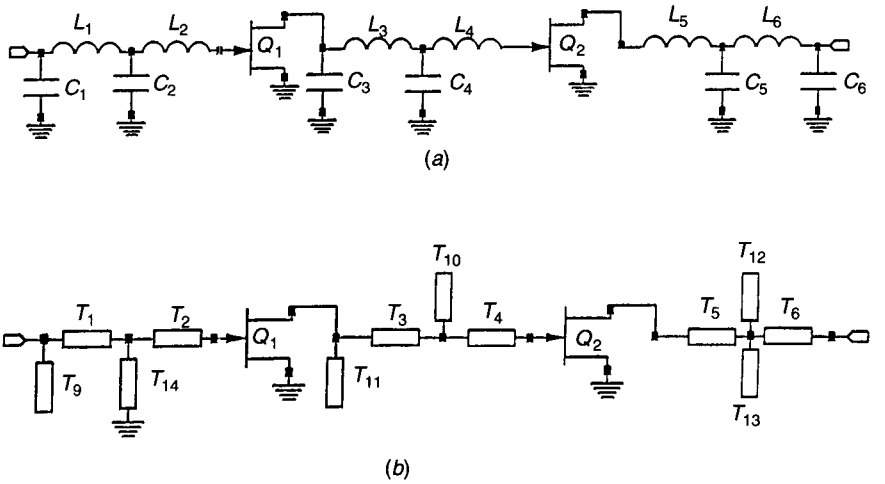


Figure 4.36 Wide-band power amplifier schematic: (*a*) lumped-element matching networks; (*b*) distributed-element matching networks; T = microstrip lines.

wide-band (2- to 18-GHz) amplifiers. The matching networks in these amplifiers were designed using MATCHNET. Figure 4.36*a* depicts lumped-element matching networks, and Fig. 4.36*b* shows distributed-element matching networks. The distributed elements in Fig. 4.40*b* are lengths of transmission lines and open or shorted stubs.

Transformation-Q Approach. The transformation-Q approach [17, 18] is a double-matching technique that enables the transformation between two arbitrary complex impedances. Though similar to the real-frequency reflection coefficient approach, this technique relies on extending the bandwidth of a narrow-band matching network by systematically optimizing the number of elements and their Q factors in the matching network. A matching network synthesis software based on this technique is available [17]. The technique can also be implemented as a routine in commercially available computer-aided design (CAD) software such as ADS (Advanced Design System) [20].

Additional Remarks. The matching networks designed using the real-frequency technique or transformation-Q technique can be further optimized using the optimization tools available in CAD software such as ADS [20]. This optimization can be used to further improve the matching bandwidth and frequency response of the network transmission coefficient ($|S_{21}|$). The Bode–Fano limit can also be implemented within the CAD software. The passband response based on the Bode–Fano limit then serves as an optimal solution for comparison with the solutions obtained using the real-frequency and transformation-Q approaches.

It should also be noted that whereas the software MATCHNET provides both lumped- and distributed-element solutions, the software based on the transformation-Q approach [17] gives only the lumped-element solution. The lumped-element network has to be converted into a distributed-element network and optimized further if a distributed-element solution is desired.

The other commercially available software tool for matching network synthesis is E-syn which is part of the software ADS [20].

4.3.5 Image Impedance Terminations

It is well known that iterative and image properties of networks can be used in the design of attenuators and filters [21]. Filters and attenuators having single or cascade connection of T or π networks can be designed by the image parameter method [2, 21]. The attenuators and filters designed by the image parameter method need to be terminated in their image impedances. The image impedance of an attenuator network consisting of purely resistive elements is frequency independent, whereas the image impedance of a filter network consisting of purely reactive elements is frequency dependent but real in the passband. Therefore, it is necessary to synthesize networks that transform frequency-independent loads to impedances that vary with frequency in a prescribed manner. The designs of image impedance terminations for low-pass and bandpass filter networks are illustrated in the following examples.

Low-Pass Network. In Fig. 4.37 a wide-band low-pass distributed amplifier [22] schematic is shown. The input and output lines are low-pass filters designed on an image parameter basis. In Fig. 4.37, I is an m-derived half-section filter [2] that transforms R_{01} and R_{02} (both real) to T-section image impedance in the passband. This termination network provides an image impedance having low-

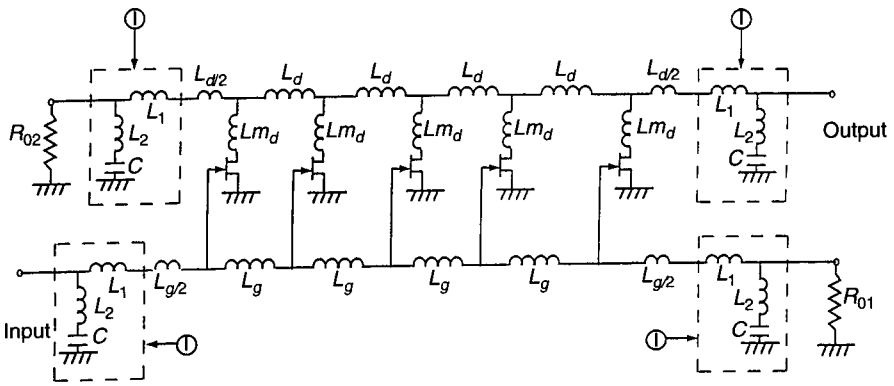
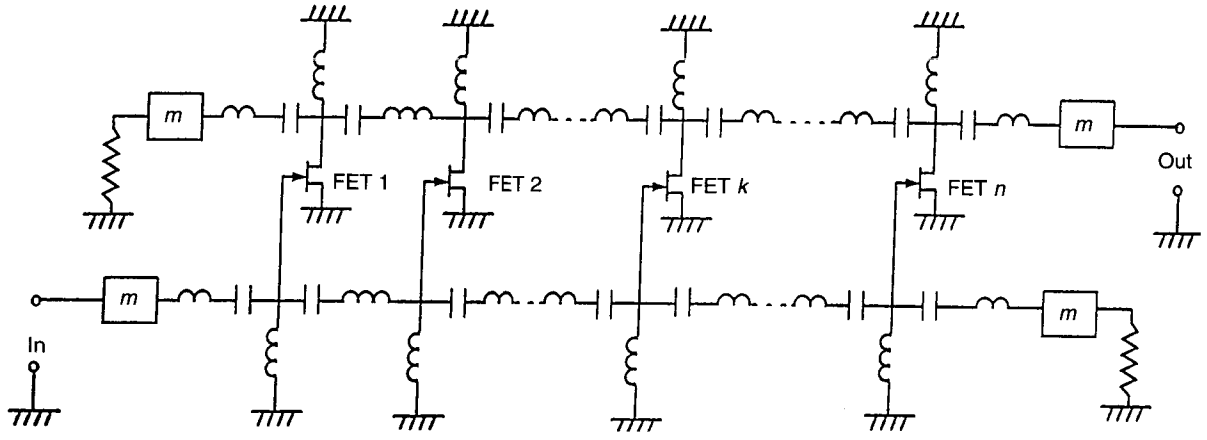
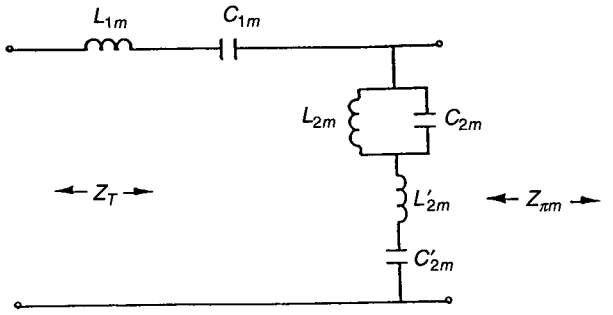


Figure 4.37 Low-pass distributed amplifier schematic; I: m = derived half-section filter.



(*a*)

Figure 4.38 (*a*) Bandpass distributed amplifier schematic; $m = m$-derived half-section Filter. (*b*) Image impedance termination network.



$$L_{1m} = \frac{m}{2} L_{1k}$$

$$C_{1m} = \frac{2C_{1k}}{m}$$

$$L_{2m} = \frac{2}{m} L_{2k}$$

$$C_{2m} = \frac{m}{2} C_{2k}$$

$$L'_{2m} = \frac{1-m^2}{2m} L_{1k}$$

$$C'_{2m} = \frac{2m}{1-m^2} C_{1k}$$

$$m \cong 0.6$$

(*b*)

Figure 4.38 *(Continued)*

pass frequency response in the passband. The equations needed for the design are

$$L_1 = \tfrac{1}{2} m L_g \qquad L_2 = \left(\frac{1-m^2}{2m}\right) L_g$$

$$C = \tfrac{1}{2} m C_{gs} \qquad m = 0.6$$

where C_{gs} is the FET gate–source capacitance.

Bandpass Network. A wide-band bandpass distributed amplifier [23] schematic is shown in Fig. 4.38*a*. The input and output lines are designed to be bandpass filter structures based on the image parameter approach. The image terminations (m in the diagram) are m-derived half-section filters [24]. In Fig. 4.38*b* the image impedance termination network and the design equations are given. This network provides an image impedance having bandpass frequency response in the passband of the filter. The topologies of other bandpass termination networks can be found in the literature [24].

REFERENCES

1. Brown, R. G., R. A. Sharpe, and W. L. Hughes, *Lines, Waves and Antennas*, Ronald, New York, 1961.
2. Pozar, D. M., *Microwave Engineering*, Wiley, New York, 1998.
3. Collin, R. E., *Foundations for Microwave Engineering*, 2nd ed., McGraw-Hill, New York, 1992.
4. Gonzalez, G., *Microwave Transistor Amplifier Analysis and Design*, Prentice-Hall, Upper Saddle River, NJ, 1997.
5. Smith, J., *Modern Communication Circuits*, McGraw-Hill, New York, 1997.
6. Grebennikov, A. V., "Create Transmission-Line Matching Circuits for Power Amplifiers," *Microwaves & RF*, Oct. 2000, p. 113.
7. Steinbrecher, D. H., "An Interesting Impedance Matching Network," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, June 1967, p. 382.
8. Edwards, T. C., *Foundations for Microstrip Circuit Design*, Wiley, New York, 1981.
9. Bode, H. W., *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945.
10. Fano, R. M., "Theoretical Limitations on the Broad-Band Matching of Arbitrary Impedances," *Journal of Franklin Inst.*, Vol. 249, Jan. 1950, pp. 57–83, and Feb. 1950, pp. 139–154.
11. Matthei, G. L., L. Young, and E. M. T. Jones, *Microwaves Filters, Impedance Matching Networks and Coupling Structures*, Artech House, Dedham, MA, 1980.
12. Vendelin, G., *Design of Amplifiers and Oscillators by the S-Parameter Method*, Wiley, New York, 1981.
13. Carlin, H. J., "A New Approach to Gain-Bandwidth Problems," *IEEE Trans. Circuits Syst.*, Vol. CAS-24, Apr. 1977, pp. 170–175.
14. Carlin, H. J., and P. P. Civalleri, *Wideband Circuit Design*, CRC Press, New York, 1998.
15. Carlin, H. J., and B. S. Yarman, "The Double Matching Problem: Analytic and Real Frequency Solutions," *IEEE Trans. Circuits Syst.*, Vol. CS-30, Jan. 1983, pp. 15–28.
16. Carlin, H. J., and P. Amstutz, "On Optimum Broad-Band Matching," *IEEE Trans. Circuits Syst.*, Vol. CAS-28, May 1981, pp. 401–405.
17. Abrie, P. L. D., *Design of RF and Microwave Amplifiers and Oscillators*, Artech House, Boston, MA, 1999.
18. Abrie, P. L. D., *The Design of Impedance Matching Networks for Radio-Frequency and Microwave Amplifiers*, Artech House, Boston, MA, 1985.
19. Sussman-Fort, S. E., *MATCHNET: Microwave Matching Network Synthesis Software and User's Manual*, Artech House, Dedham, MA, 1991.
20. *Advanced Design System®*, Agilent Technologies, Westlake Village, CA, 2002.
21. Guillemin, E. A., *Communication Networks*, Vol. II, Wiley, New York, 1946.
22. Prasad, S. N., J. B. Beyar, and I. S. Chang, "Power-Bandwidth Considerations in the Design of MESFET Distributed Amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-36, July 1988, pp. 1117–1123.

23. Prasad, S. N., and J. B. Beyer, "Band-Pass Distributed Amplifiers," *Microwave Opt. Technol. Lett.*, Vol. 2, Oct. 1989, pp. 349–354.

24. *Reference Data for Radio Engineers*, Howard W. Sams & Co., New York, 1982.

FURTHER READING

Chen, W. K., "Synthesis of Optimum Butterworth and Chebyshev Broadband Impedance Matching Networks," *IEEE Trans. Circuits Syst.*, Vol. CAS-24, Apr. 1977, pp. 152–169.

Cohn, S. B., "Optimum Design of Stepped Transmission Line Transformers," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-3, Apr. 1955, pp. 16–21.

Collin, R. E., "Theory and Design of Wide Band Multisection Quarter-Wave Transformers," *Proc. IRE*, Vol. 43, Feb. 1955, pp. 179–185.

Collin, R. E., "The Optimum Tapered Transmission Line Matching Section," *Proc. IRE*, Vol. 44, Apr. 1956, pp. 539–548.

Karakash, J. J., *Transmission Lines and Filter Networks*, Macmillan, New York, 1950.

Klopfenstein, R. W., "A Transmission Line Taper of Improved Design," *Proc. IRE*, Vol. 44, Jan. 1956, pp. 31–35.

Ku, W. H., and W. C. Peterson, "Optimum Gain-Bandwidth Limitations of Transistor Amplifiers as Reactively Constrained Active Two-Port Networks," *IEEE Trans. Circuits Syst.*, Vol. CAS-22, June 1975, pp. 523–533.

Mellor, D. J., and J. G. Linvill, "Synthesis of Interstage Networks of Prescribed Gain Versus Frequency Slopes," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-23, Dec. 1975, pp. 1013–1020.

Riblet, H. J., "General Synthesis of Quarter-Wave Impedance Transformers," *IRE Trans. Microwave Theory Tech.*, MTT-5, Jan. 1957, pp. 36–43.

Solymar, L., "Some Notes on the Optimum Design of Stepped Transmission Line Transformers," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-6, Oct. 1958, pp. 374–378.

Wang, Y., "New Method for Tapered Transmission Line Design," *Electron. Lett.*, Vol. 27, No. 25, Dec. 1991, pp. 2396–2397.

Youla, D. C., "A New Theory of Broadband Matching," *IEEE Trans. Circuit Theory*, Vol. CT-11, Mar. 1964, pp. 30–50.

Young, L., "Tables for Cascaded Homogeneous Quarter-Wave Transformers," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-7, Apr. 1959, pp. 233–237; see also Vol. MTT-8, 1960, pp. 243–244, for corrections.

PROBLEMS

4.1 Design a single-stub matching unit to match a load impedance, $Z_L = 75 + j100\ \Omega$ to a transmission line having $Z_0 = 50\ \Omega$ using a short-circuited stub. All transmission lines in the matching unit should have $Z_0 = 50\ \Omega$. Give only the solution, which has the shorter length of line from load to stub.

4.2 A load impedance, $Z_L = 100 + j50\ \Omega$ is to be matched to a transmission line having $Z_0 = 50\ \Omega$. Design a single-stub matching units using short-circuited stubs having $Z_0 = 100\ \Omega$. Give two possible solutions.

4.3 Design a single-stub matching unit at 500 MHz to transform a load consisting of a 100 Ω resistor in series with a inductor of $50/\pi$ nH, to a transmission line having $Z_0 = 50\ \Omega$. Give only the solution having the shorter length (d) of line between load and stub. Use an open-circuited stub. Determine the lengths in cm of d and l (the length of stub) in the solution, if the phase velocity on the lines is 3×10^8 m/sec. All transmission lines in the matching unit have $Z_0 = 50\ \Omega$.

4.4 Design double-stub matching units to match a load impedance, $Z_L = 50 + j50\ \Omega$ to a transmission line having $Z_0 = 50\ \Omega$, using short-circuited stubs. The distance between the stubs must be $3\lambda/8$. All lines including the stubs should have $Z_0 = 50\ \Omega$. Give two possible solutions.

4.5 Design double-stub matching units consisting of open-circuited stubs separated by $\frac{1}{8}\lambda$, to match a load admittance, $Y_L = (1.4 + j2)Y_0$ to a transmission line having $Z_0 = 50\ \Omega$. Give two possible solutions. All lines including the stubs should have $Z_0 = 50\ \Omega$.

4.6 A load resistance of 100 Ω is to be matched to a transmission line having $Z_0 = 50\ \Omega$ at 500 MHz, using a quarter-wave transformer. Determine the following: (a) characteristic impedance, Z_{0T} of the quarter-wave transformer, (b) physical length (in meters) of the transformer if the ϵ_r of the line is 2.25, and (c) the fractional bandwidth obtained if the maximum VSWR that can be tolerated is 1.5.

4.7 A load impedance, $Z_L = 25 - j25\ \Omega$ is to be matched to a transmission line having $Z_0 = 50\ \Omega$. Design matching units consisting of a transmission line ($Z_0 = 50\ \Omega$) and a quarter-wave transformer. Give two possible solutions.

4.8 A load impedance, $Z_L = 25 + j25\ \Omega$ is to be transformed to an impedance of 25 Ω. Design a matching unit consisting of an open-circuited stub ($Z_0 = 50\ \Omega$) and a quarter-wave transformer to accomplish the transformation.

4.9 Design a matching unit to transform a load impedance, $Z_L = 25 - j25\ \Omega$ to 50 Ω using a $\frac{1}{8}\lambda$ line having $Z_0 = |Z_L|$ and a quarter-wave transformer.

4.10 Design a matching unit consisting of a quarter-wave transformer and a transmission line ($Z_0 = 50\ \Omega$) to transform a load impedance of 50 Ω to $25 + j25\ \Omega$. Give the solution having the shorter length of transmission line.

4.11 Design a matching unit consisting of a quarter-wave transformer and a short-circuited stub to transform a load impedance, $Z_L = 50\ \Omega$ to $20 - j40\ \Omega$.

4.12 Design a "series-L" and "shunt-C" lumped element matching network using L-network approach, to transform the load impedance, $Z_L = 15 + j15\ \Omega$ to $50\ \Omega$ at 500 MHz. Determine the element values.

4.13 Design a lumped element impedance transformation network using a L-network approach to transform $50\ \Omega$ to $100 - j100\ \Omega$ at 1 GHz. Determine the element values. Give all possible solutions.

4.14 Design lumped element impedance matching networks using a L-network approach to transform $50 - j50\ \Omega$ to $50\ \Omega$ at 2 GHz. Determine the element values. Give all possible solutions.

4.15 Design lumped element matching networks to match $Z_L = 10 + j40\ \Omega$ to $50\ \Omega$ at 2 GHz, using a L-network approach. Determine the element values. Give all possible solutions.

4.16 Design lumped element matching networks using a network conversion approach to transform $400\ \Omega$ to $75\ \Omega$ at 900 MHz. Give all possible solutions. Determine the element values.

4.17 A load impedance, $Z_L = 100 + j100\ \Omega$ is to be matched to $50\ \Omega$, using a combined lumped-distributed Element Technique. Design matching networks at 2.4 GHz using a transmission line ($Z_0 = 50\ \Omega$) connected to the load, and then followed by a lumped element in series. Calculate the lumped element values. Give two possible solutions.

4.18 Design a three-section binomial transformer to transform a load impedance of $100\ \Omega$ to transmission line having $Z_0 = 50\ \Omega$ if the maximum reflection coefficient (ρ_m) that can be tolerated is 0.05. Determine the fractional bandwidth obtained.

4.19 Design a two-section Chebyshev transformer to match a load impedance of $100\ \Omega$ to a transmission line having $Z_0 = 50\ \Omega$ if the maximum reflection coefficient (ρ_m) that can be tolerated is 0.05. Determine the fractional bandwidth obtained.

4.20 Design a three-section Chebyshev transformer using exact theory to transform a load impedance of $100\ \Omega$ to a transmission line having $Z_0 = 50\ \Omega$ if the maximum reflection coefficient (ρ_m) that can be tolerated is 0.05. Determine the fractional bandwidth obtained. Compare the bandwidth obtained to those in Problems 4.18 and 4.19.

5

HYBRIDS AND COUPLERS

P. Bhartia and P. Pramanick

5.1 INTRODUCTION

Hybrids and couplers form an indispensable component group in modern MIC technology. With the inventions of new planar transmission lines like strip line, microstrip line, fin line, dielectric image line, and their derivatives, hybrid and coupler technology has undergone a substantial change over the past decade due to the rapidly growing applications of MICs in the electronic-warfare, communications, and radar industries.

Despite the fact that the basic philosophy behind the operation of such couplers remains the same as in couplers designed using conventional two-wire transmission line, their analyses and syntheses are quite involved. This is because most of these lines support hybrid modes due to inhomogeneity in configuration. However, the present-day analysis and synthesis techniques for such hybrids and couplers are believed to have gained maturity.

This chapter describes the design aspects of planar hybrids and couplers in as self-contained a presentation as possible within a limited space. In what follows in this section, we present the basics of hybrids and couplers and discuss different types of hybrids and couplers and their applications. The next section describes the design of matched hybrid Ts, hybrid rings, and 90° hybrids. This is followed by a section on coupled-line couplers, both the TEM and the distributed types, as well as other miscellaneous types of couplers. The final section includes various aspects of coupler design, such as losses and improvement of directivity.

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

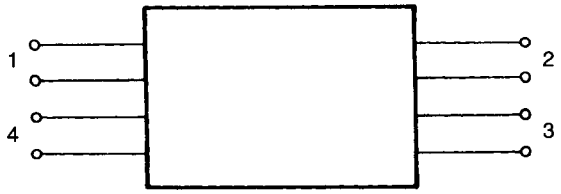


Figure 5.1 Four-port network.

5.1.1 Basics of Hybrids and Couplers

A hybrid or a directional coupler can in principle be represented as a multiport network. In such a network the port into which the electrical power is fed is called the incident port. The ports through which the desired amounts of coupled power are extracted are called coupled ports, while the rest of the ports are called isolated ports. Although hybrids and couplers having up to six ports find applications in many systems, we will mostly restrict ourselves to the discussion of four-port networks without loss of generality.

Consider the four-port network shown in Fig. 5.1. If P_1 is the power fed into port 1 (which is matched to the generator impedance) and P_2, P_3, and P_4 are the powers available at the ports 2, 3, and 4, respectively (while each of the ports is terminated by its image impedance), the coupling coefficient is defined as

$$C = -10 \log \left| \frac{P_n}{P_1} \right| \qquad n = 2, 3, 4 \tag{5.1}$$

If port 3 happens to be the desired coupled port, the coupling coefficient is given by

$$C = -10 \log \left| \frac{P_3}{P_1} \right| \qquad \text{dB} \tag{5.2}$$

If port 4 is the desired uncoupled port, the desired isolation is given by

$$I = -10 \log \left| \frac{P_4}{P_1} \right| \qquad \text{dB} \tag{5.3}$$

The transmission to the primary port 2 is given by

$$T = 10 \log \left| \frac{P_2}{P_1} \right| \qquad \text{dB} \tag{5.4}$$

The measure of directivity between the coupled and the uncoupled ports is

given by

$$D = I - C \tag{5.5}$$

As a general practice, the performance of a hybrid or a directional coupler is specified in terms of its coupling, directivity, and the characteristic impedance at the center frequency of its band of operation. These data enable the circuit designer to calculate the structural parameters of the coupler.

5.1.2 Types of Hybrids and Couplers

Hybrids are a direct-coupled type of circuit. Couplers can be direct coupled, parallel coupled, or aperture coupled.

Figure 5.2*a* shows the simplest direct-coupled hybrid. This is a branch-line coupler, consisting of two main lines coupled by two $\frac{1}{4}\lambda$ line sections spaced $\frac{1}{4}\lambda$ apart, where λ is the wavelength. Such a branch-line coupler can also be in circular form as shown in Fig. 5.2*b*. In either case, the total length of all the lines is one wavelength.

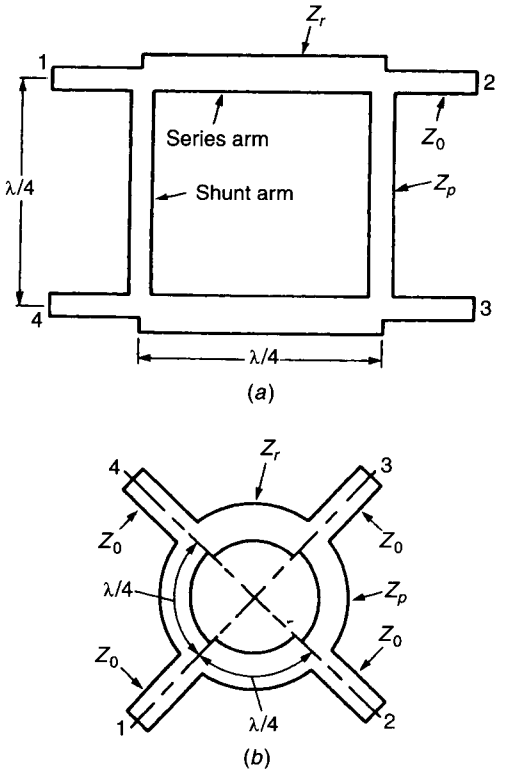


Figure 5.2 (*a*) Single-section branch-line coupler, (*b*) circular form of branch-line coupler.

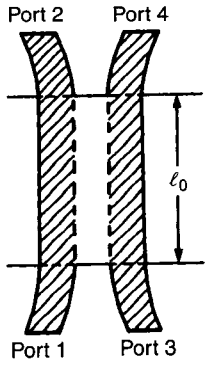


Figure 5.3 Direct-coupled coupler.

Figure 5.3 shows a direct-coupled coupler. It consists of two directly coupled transmission lines. The lines can have any form depending upon the application.

A parallel-coupled line coupler (Fig. 5.4) can be of two types, that is, the TEM type or the distributed type. In the former, the coupled transmission lines support a pure TEM or a quasi-TEM mode. In the latter, the mode supported by the coupled lines are non-TEM in nature. While all of these types of couplers are realized using planar IC technology, there can be another type of coupler using aperture coupling, through some common ground plane (as used in conventional aperture-coupled waveguide couplers as shown in Fig. 5.5) with two similar or dissimilar types of planar transmission lines.

To achieve tighter coupling over a wider bandwidth multisections of the previously mentioned couplers can be cascaded in tandem (to be discussed in a later section).

5.1.3 Applications

Virtually all kinds of microwave circuits use hybrids or couplers in one form or the other. In general, the areas of application can be divided into two parts.

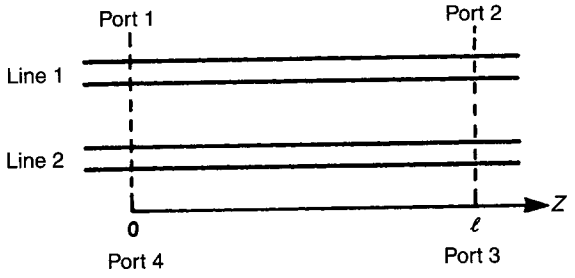


Figure 5.4 Parallel-coupled line coupler.

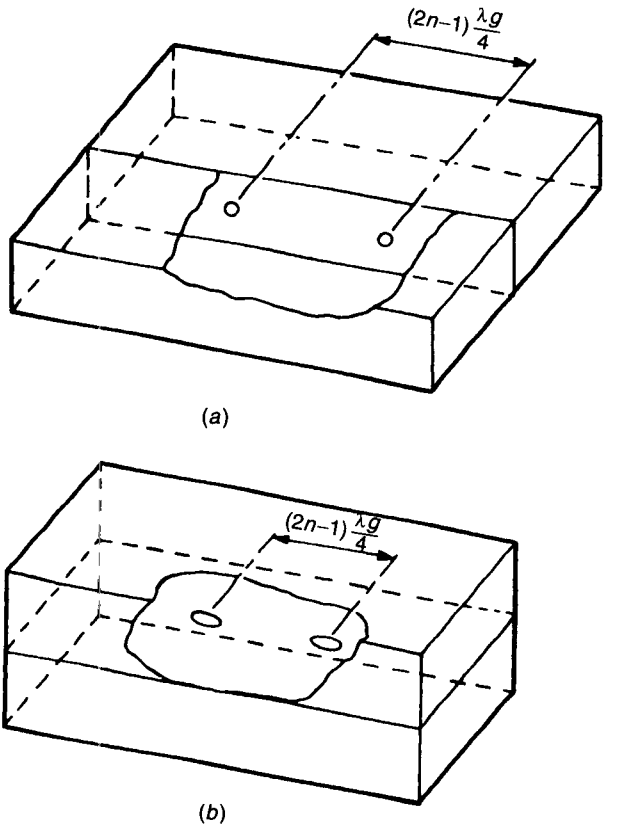


Figure 5.5 Aperture-coupled couplers: (*a*) narrow wall coupling and (*b*) broad wall coupling.

1. *Passive.* Tuners, delay lines, filters, and matching networks. Sometimes an array of couplers may be needed for a desired performance of the network.
2. *Active Circuits.* Mainly used as directional couplers in balanced amplifiers, mixers, attenuators, modulators, discriminators, and phase shifters.

5.2 DESIGN OF HYBRIDS

5.2.1 90° Hybrid

The simplest 90° hybrid is the branch-line coupler shown in Fig. 5.2*a*. For a certain input power at port 1, under matched condition, there will be 90° phase difference between the waves appearing at ports 2 and 3, at the center frequency at which each arm is exactly a quarter wavelength long. This 90° phase difference varies over $\pm 5°$ for a 10% change in frequency around the center

frequency. The coupling bandwidth is 20%, but its usable bandwidth is limited to 10% due to an unacceptable change in the isolation over a bandwidth exceeding 10%. Ideally such couplers can be designed for 3–9 dB coupling.

There are three main arm losses in branch-line couplers. They are losses due to the portion of power coupled to the secondary arm and the power dissipated due to dielectric and copper losses.

Branch-line couplers can be realized using virtually all kinds of planar transmission lines, for example, strip line, microstrip line, slot line, fin line, and image line. However, the basic design principle is extremely simple and the same in all cases.

The coupling factor is determined by the ratio of the impedance of the shunt and series arms and is optimized to meet the proper match over the required bandwidth.

For 90° hybrids the following conditions hold good (see Fig. 5.2*a*).

$$\frac{P_2}{P_3} = \left(\frac{Z_0}{Z_p}\right)^2 \tag{5.6}$$

$$\left(\frac{Z_0}{Z_r}\right)^2 = \left(\frac{Z_0}{Z_p}\right)^2 + 1 \tag{5.7}$$

Although the two-branch coupler is the most fundamental structure, it has a very narrow bandwidth. This disadvantage is overcome by using multisection couplers. But in most cases planar transmission lines require too wide a range of impedances for this purpose. These may sometimes be difficult to realize physically. On the other hand, very wide linewidths may require unreasonable aspect ratios at higher frequencies due to shorter wavelengths. The physically unrealizable high-impedance line can be avoided by using a modified hybrid ring [1].

Complete analytical design techniques for such hybrids using Chebyshev and Zolotarev [2, 3] functions are available. However, such techniques may sometimes be unsuitable for the design of planar circuits because of the wide impedance range problem, as previously mentioned. Muraguchi et al. [4] presented a computer-aided design technique that is most suitable for an optimum design. The method is as follows.

Let us consider the lossless reciprocal four-port branch-line hybrid that has a twofold symmetry around the x,y planes, as shown in Fig. 5.6*a*. The S matrix of the hybrid can be written as

$$[S] = \begin{bmatrix} S_{11} & S_{21} & S_{31} & S_{41} \\ S_{21} & S_{11} & S_{41} & S_{31} \\ S_{31} & S_{41} & S_{11} & S_{21} \\ S_{41} & S_{31} & S_{21} & S_{11} \end{bmatrix} \tag{5.8}$$

together with the unitary relationship

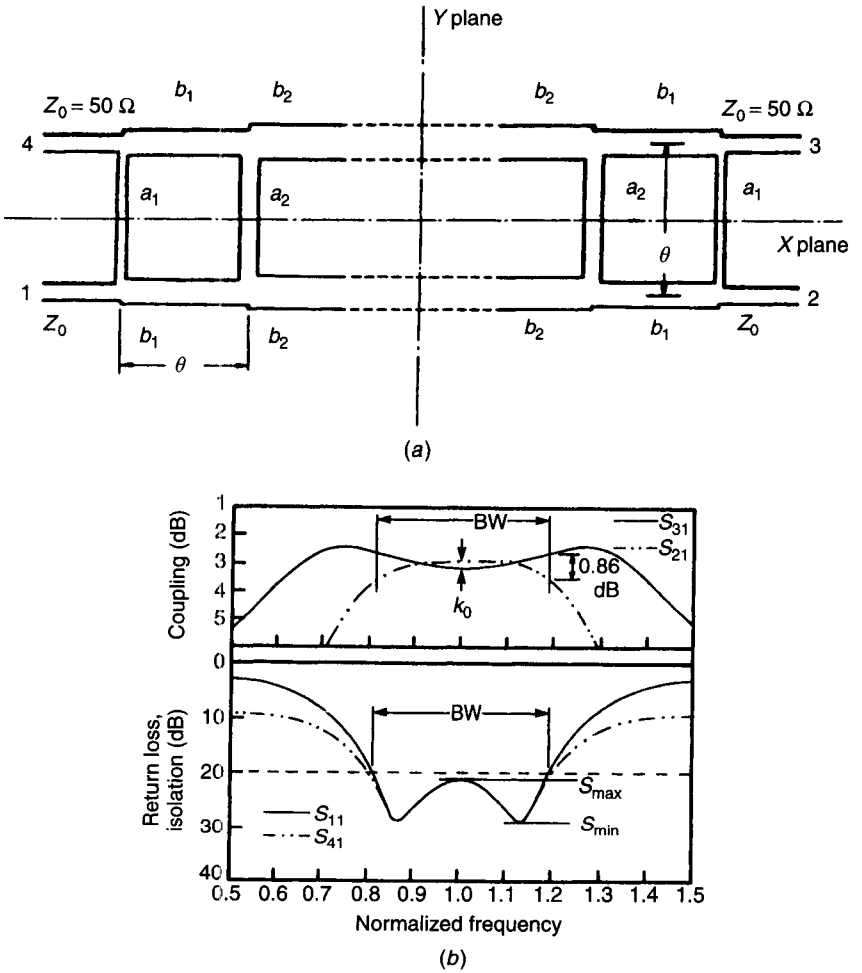


Figure 5.6 (*a*) Multisection branch-line coupler. (*b*) Frequency response of a three-branch optimized branch-line coupler. (After Muraguchi et al. [4]. Reprinted with permission of IEEE.)

$$|S_{11}|^2 + |S_{21}|^2 + |S_{31}|^2 + |S_{41}|^2 = 1 \tag{5.9}$$

Other conditions for the return loss, the power-coupling coefficient C, and the isolation are

$$|S_{11}|^2 = 0 \qquad |S_{21}|^2 = C \qquad |S_{31}|^2 = 1 - C \qquad |S_{41}|^2 = 0$$

The previous quantities are required to be within certain tolerance limits over a broad frequency band. Such tolerance limits depend upon the design requirements. For instance, for a two-branch 3-dB microstrip hybrid reported

by Muraguchi et al. [4], assuming less than ±0.43 dB variation in coupling and better than −20 dB isolation, a bandwidth of 30% was realized about the 1-GHz center frequency.

Having defined the tolerance limits, the following penalty function is formed

$$F(a_1, \ldots, a_n, b_1, \ldots, b_m) = \sum_{j=1}^{4} g_j \tag{5.10}$$

$$g_1 = \sum_{i=1}^{N} |S_{11}(f_i)|^2 \qquad g_2 = \sum_{i=1}^{N} [|S_{21}(f_i)|^2 - C]$$

$$g_3 = \sum_{i=1}^{N} [|S_{31}(f_i)|^2 - (1 - C)] \qquad g_4 = \sum_{i=1}^{N} |S_{41}(f_i)|^2$$

where

$$f_i = f_0\left(1 + \frac{i-1}{D}\right) \qquad i = 1, \ldots, N \tag{5.11}$$

and N is the number of sampling points. The f_i are the corresponding sampling frequencies, and f_0/D is the sampling interval. All four scattering parameters S_{11}, S_{21}, S_{31}, and S_{41} are considered despite the fact that they are not independent of each other for a lossless hybrid. The parameters $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ are obtained numerically in order that the penalty function F may be minimized by a suitable search method. The optimization process has the following steps:

Step 1: The first computation is performed without any restrictions on the line impedances by changing the sampling interval $1/D$ only.

Step 2: If some undesirably low and/or high impedance is encountered in the result of the first computation, the second computation is performed after one of their impedance values is changed to an appropriate fixed value.

Step 3: If there still happens to be undesirably low- and/or high-impedance lines in the result of the second computation, the third computation is performed after two or three impedance values are held constant.

The computation process is repeated until the prescribed tolerance limits are exceeded.

Because of the two-dimensional nature of a planar transmission line, it is impossible to define the circuit uniquely. Therefore a transmission-line or one-dimensional approach is used. For hybrids with large impedance steps, junction reactance effects are added after the optimization has been carried out. A final

experimental correction of the electrical length is also needed. Typical frequency characteristics of an optimized three-branch coupler are shown in Fig. 5.6*b*.

5.2.2 Ring Form of Branch-Line Hybrid

This is a circular ring version of the square 90° hybrid described in the previous section. Therefore, the discussion of the previous section is equally valid for hybrid rings. The configuration is shown in Fig. 5.2*b* and is particularly advantageous in the realization of microstrip phase detectors and balanced mixers, with all ports matched.

5.2.3 Matched Hybrid T (Rat-Race Hybrid)

A matched hybrid T is a special kind of ring form of the branch-line coupler in which the circumference is an odd multiple of $\frac{3}{2}\lambda$. As a result, the phase response is 0°/180°. The simplest version of a matched hybrid T is shown in Fig. 5.7. The ports A–B, B–C, and C–D are separated by 90°, and port A and port D are three quarter-wavelengths away from each other.

Because of the impedance and the phase relationships shown in the structure, any power fed into port C splits equally into two parts that add up in phase at ports B and D and out of phase of port A. As a result port A is isolated from the input. Similarly, power fed at port A divides equally between ports B and D with 180° phase difference, and port C remains isolated.

The frequency response of a typical $\frac{3}{2}$ wavelength hybrid T is shown in Figs. 5.8*a* and *b*. It is observed that it offers around a 20% bandwidth in terms of matching, split, and isolation. Moreover, the 180° phase relationship is much more frequency sensitive than the 0° phase relationship. Still, such a response is quite adequate for applications in mixers, SSB generators, and so on.

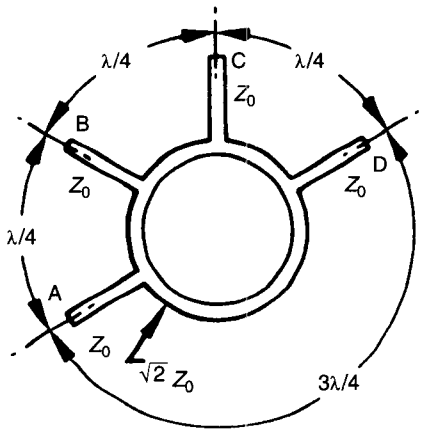


Figure 5.7 Hybrid ring.

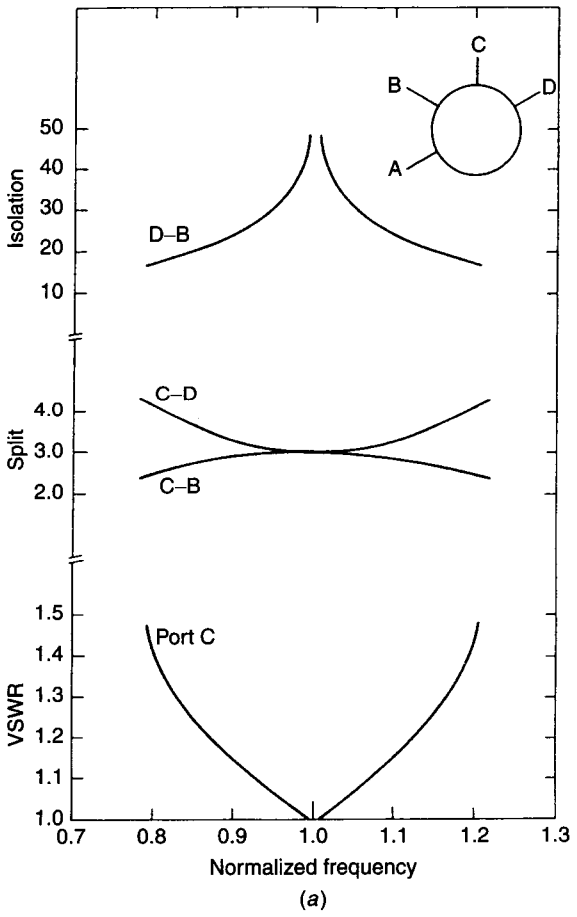


Figure 5.8 (*a*) Response curves for $\left(\frac{3}{2}\right)$ rat-race magic T. Power split and isolation are express in decibels. (*b*) Phase response curves for $\left(\frac{3}{2}\right)$ rat-race magic T.

For high-frequency applications, a circumferential length of $\frac{3}{2}$ wavelengths may pose fabricational difficulties due to unrealizable aspect ratios of the transmission lines. This problem is overcome by having a ring of circumferential length of two wavelengths, where port A and port D are located $\frac{5}{4}\lambda$ away on the ring. The bandwidth of the ring length is around 20%.

The idealized S matrix of a matched hybrid T is given by

$$S = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \tag{5.12}$$

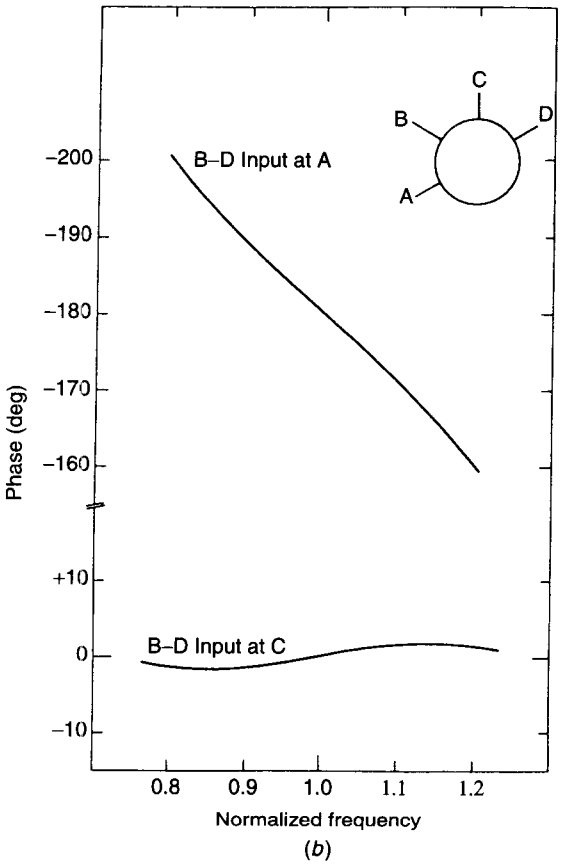
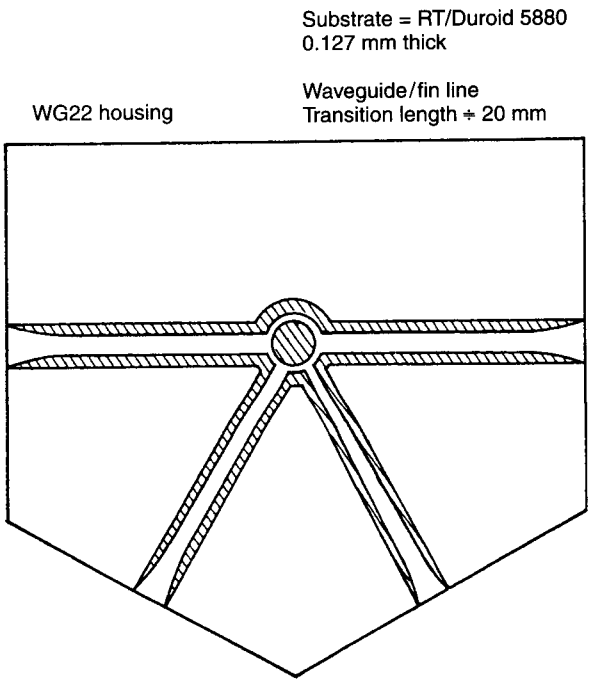


Figure 5.8 *(Continued)*

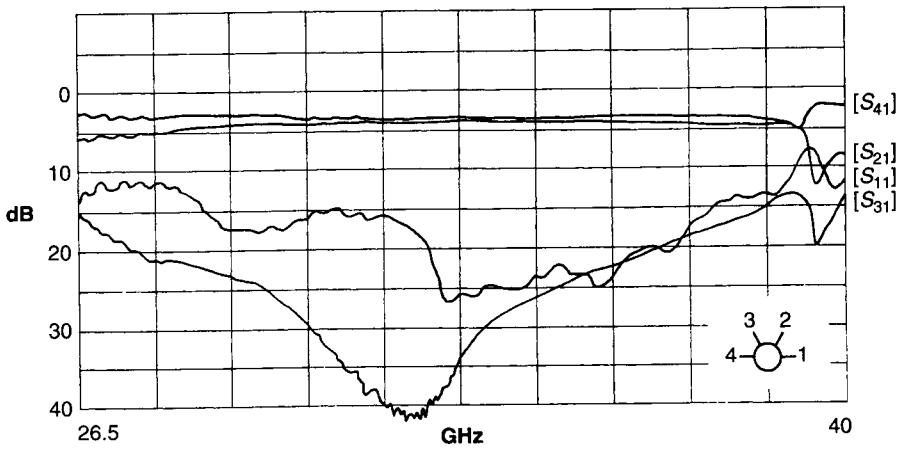
The design of a hybrid amounts to realizing the required transmission-line sections with proper phase velocities and characteristic impedances. Direct synthesis equations can be used in cases of strip line [5], microstrip [6], suspended microstrip, and inverted microstrip [7], which support pure TEM or quasi-TEM modes of propagation.

Realizations of the hybrids using non-TEM transmission lines like slot line, fin line, or image line use iterative techniques with the help of accurate analysis equations [7–9] depending upon the mode of propagation. However, closed-form design equations are also available.

Microstrip and strip-line hybrids have been realized and successfully used in commercially available microwave-balanced mixers and other circuit components. Development of fin-line and image-line hybrids were researched [10]. Figure 5.9*a* shows a unilateral fin-line matched hybrid T, and Fig. 5.9*b* shows its frequency response in the *Ka* band [11].



(*a*)



(*b*)

Figure 5.9 (*a*) Fin-line hybrid ring. (*b*) Frequency response of fin-line hybrid ring. (After Rycroft [11]. Reprinted with permission of Marconi Research Center.)

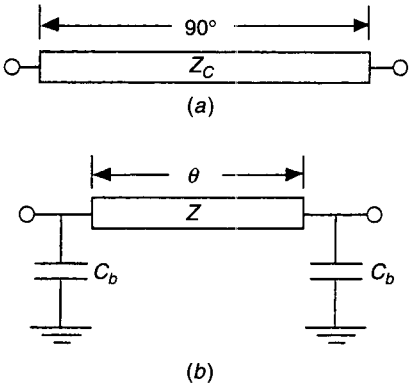


Figure 5.10 (*a*) Quarter-wave long transmission section. (*b*) Reduced-size circuit equivalent to quarter-wave section.

5.2.4 Reduced-Size Quasi-Lumped Quadrate Hybrid [12]

The basic building block of a distributed-line branch-line hybrid is a 90° transmission line as shown in Fig. 5.10*a*. The *ABCD* matrix of the line is given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 0 & jZ_c \\ \dfrac{j}{Z_c} & 0 \end{bmatrix} \tag{5.13}$$

The same 90° transmission line can be replaced by combination of a shorter transmission line cascaded with two shunt capacitances as shown in Fig. 5.10*b*. The *ABCD* matrix of the network in Fig. 5.10*b* is given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ j\omega C_b & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & jZ\sin\theta \\ \dfrac{j\sin\theta}{Z} & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ j\omega C_b & 1 \end{bmatrix} \tag{5.14}$$

Comparing Eqs. (5.13) and (5.14) it can be shown that θ and C_b are obtained from

$$\theta = \sin^{-1}\left(\frac{Z_c}{Z}\right) \tag{5.15}$$

$$C_b = \frac{\cos\theta}{\omega Z_c} \tag{5.16}$$

It is apparent from the above two equations that one may choose either θ or Z and calculate the other from Eq. (5.15) and subsequently obtain C_b from Eq. (5.16). Let us consider a typical distributed-line 3-dB branch-line hybrid. The line impedances in such a network are Z_0 and $Z_0/\sqrt{2}$ of the main and the shunt are arms, respectively. If the impedances of the main and the shunt arms are

made equal and replaced by a value $Z > Z_0$, then both arms will have their electrical lengths less than 90°. It can be shown that in that case the lengths of the main and the shunt arms are given by

$$\theta_{\text{main}} = \sin^{-1}(v) \tag{5.17}$$

$$\theta_{\text{shunt}} = \sin^{-1}\left(\frac{v}{\sqrt{2}}\right) \tag{5.18}$$

and

$$\omega Z_0 C_b = \sqrt{1 - v^2} + \sqrt{2 - v^2} \tag{5.19}$$

where

$$v = \frac{Z_0}{Z} \tag{5.20}$$

In MMIC technology, C_b can be realized in a very small size. At the same time, the electrical lengths of the main and the shunt arms can be made much smaller than 90°, choosing an appropriate value of Z. For example, if we choose $Z = \sqrt{Z_0}$, then the main and the shunt arms will have the lengths $\frac{1}{8}\lambda$ and $\frac{1}{12}\lambda$, respectively. Once Z is known, C_b can be obtained from Eq. (5.19). Figure 5.11

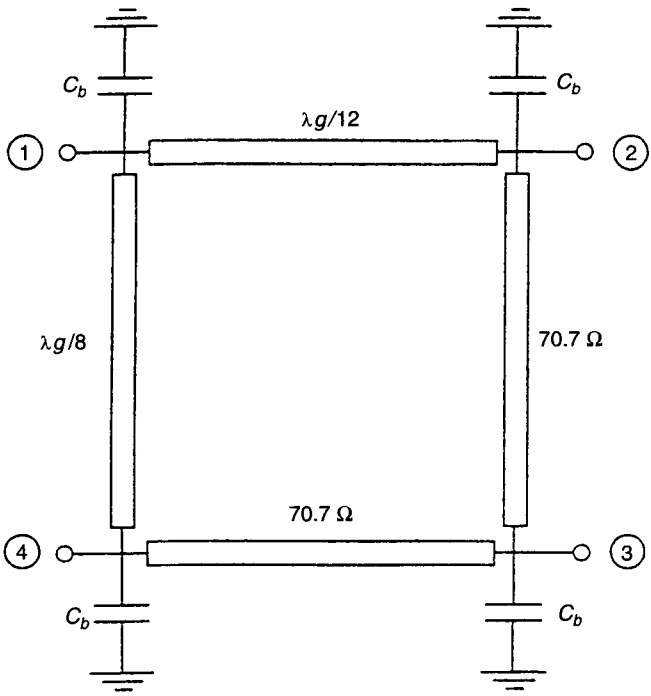


Figure 5.11 Reduced-size branch-line hybrid.

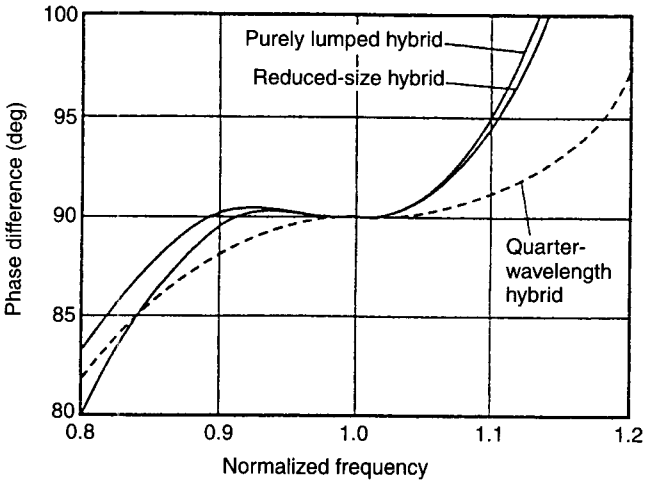


Figure 5.12 Calculated phase difference between S_{21} and S_{31} of the reduced-size hybrid, the conventional hybrid, and the purely lumped-element hybrid. (From Hirota et al., *IEEE-MTT*, Vol. 38, 1990, pp. 270–275. Reprinted with permission.)

shows the reduced size branch-line hybrid. The reduction in size of this branch-line hybrid, of course, comes with a price. Figure 5.12 shows that the distributed-line branch-line hybrid offers a wider bandwidth. However, the modified circuit needs groundings at two locations.

5.2.5 Modified Rat Race

The rat-race hybrid discussed in the previous section offers a narrow bandwidth of nearly 20%. However, a modified form of conventional rat-race hybrid can offer a much larger bandwidth. The modified version is obtained by replacing the $\frac{3}{4}\lambda_g$ long section of the conventional rat race by a pair of $\frac{1}{4}\lambda_g$ long coupled lies as shown in Fig. 5.13.

The even- and the odd-mode impedances of the coupled lines are given by

$$Z_{0e} = 2.414Z_{01} \tag{5.21}$$

$$Z_{0o} = \frac{Z_{01}^2}{Z_{0e}} \tag{5.22}$$

Figure 5.14 compares the performances of a modified rat-race hybrid with that of a conventional hybrid. The modified rat-race hybrid not only offers a wider bandwidth but also a more compact size.

Reduced-Size Rat Race [12]. Let us consider the conventional rat-race hybrid shown in Fig. 5.7. A reduced size quasi-lumped rat-race hybrid is obtained

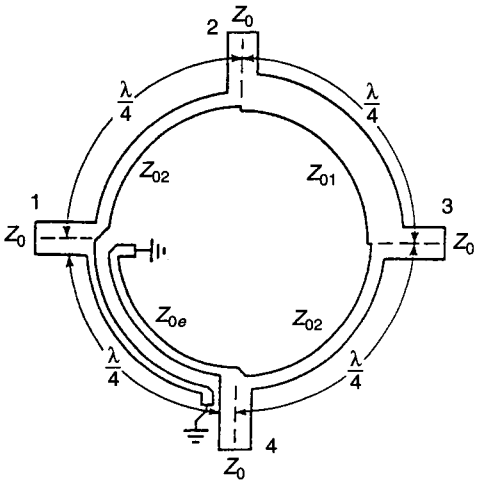


Figure 5.13 Broadband hybrid ring using a shorted parallel-coupled quarter-wave section.

by replacing the $\frac{3}{4}\lambda_g$ and the $\frac{1}{4}\lambda_g$ section of a conventional rat-race hybrid by the π networks shown in Figs. 5.15*a* and *b*, respectively.

Figure 5.16*a* shows the overall network. Since the parallel combination of L_a and C_b offers a very high impedance at the center frequency of the rat race, it is eliminated in the actual physical realization of the rat-race hybrid. Figure 5.16*b* shows the final form of the circuit.

Lumped-Element Rat Race [13]. The $\frac{1}{4}\lambda_g$ and the $\frac{3}{4}\lambda_g$ lines in a conventional distributed rat-race hybrid can be replaced by a low-pass and a high-pass network, respectively, as shown in Figs. 5.17*a* and *b*. The overall rat-race hybrid then assumes the completely lumped-element form as shown in Fig. 5.18. On equating the *ABCD* matrices of the corresponding distributed and lumped networks the values of the lumped elements are obtained as

$$C_1 = \frac{1}{\sqrt{2}Z_0\omega}\sqrt{\frac{1-\cos\theta_1}{1+\cos\theta_1}} \tag{5.23}$$

$$C_2 = \frac{1}{\sqrt{2}Z_0\omega}\sqrt{\frac{1-\cos\theta_2}{1+\cos\theta_2}} \tag{5.24}$$

$$L_1 = \frac{\sqrt{2}}{\omega}Z_0 \sin\theta_1 \tag{5.25}$$

$$L_2 = \frac{\sqrt{2}}{\omega \sin\theta_2}Z_0 \tag{5.26}$$

The element values of a rat for 50 Ω system impedance are $L_1 = L_2 = 11.25/f$ nanohenrys and $C_1 = C_2 = 2.25/f$ picofarads where f is in gigahertz.

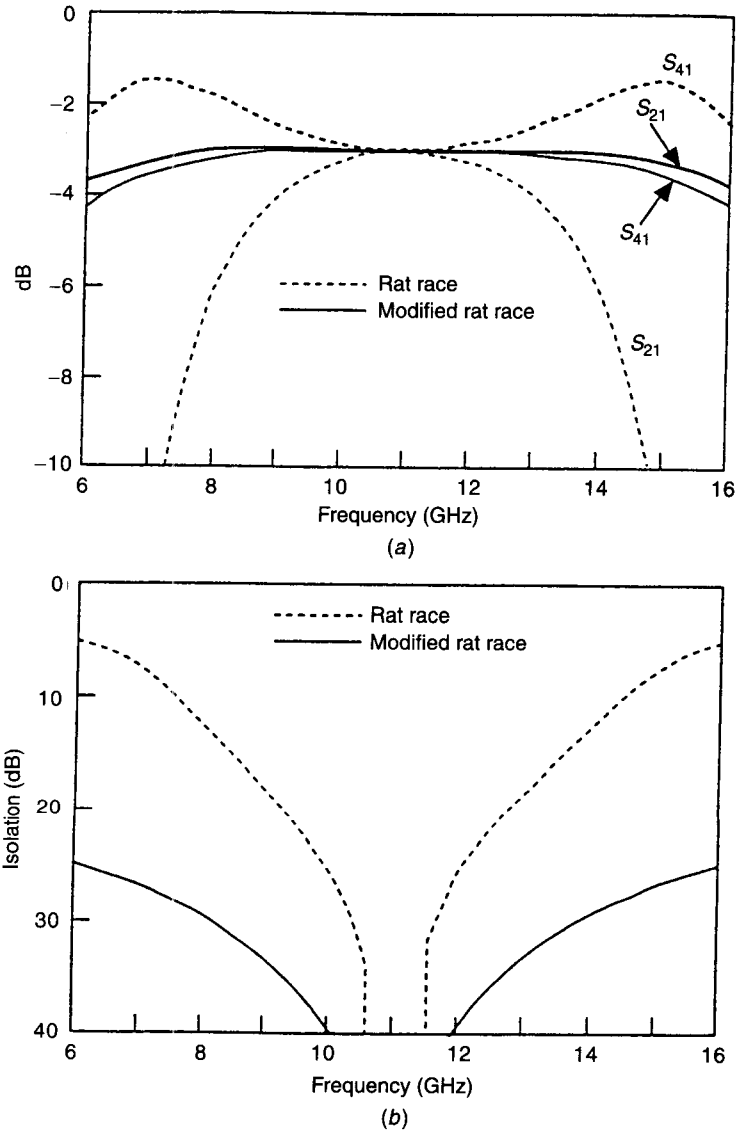


Figure 5.14 Broadband hybrid response (*a*) coupling, (*b*) isolation, and (*c*) return loss.

5.3 COUPLED-LINE DIRECTIONAL COUPLERS

The couplers described in the previous sections are inherently of narrow bandwidth. Broadbanding of microwave couplers is achieved in a number of different ways, depending upon the application.

Broadband couplers are either aperture coupled or parallel coupled. Aperture coupling is used very successfully in conventional waveguide techniques

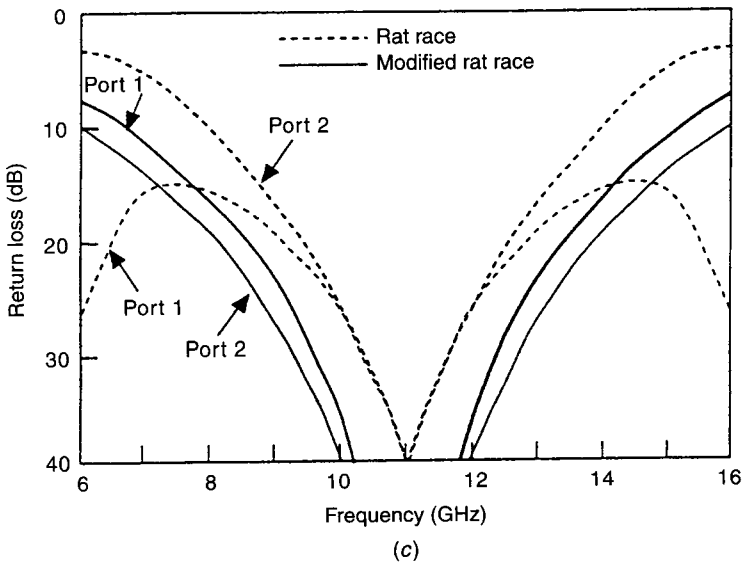


Figure 5.14 *(Continued)*

for realizing directional couplers with large bandwidths. Although aperture coupling is a convenient way of realizing directional couplers using two different planar transmission lines, directional couplers using the same kind of planar transmission lines are realized more efficiently using parallel coupling.

5.3.1 Directional Couplers Using Aperture-Coupled Lines

Figure 5.19 shows a simple two-hole directional coupler. Electromagnetic energy is coupled from the primary guide to the secondary guide due to the field radiated by the excited electric and magnetic dipoles generated at the holes

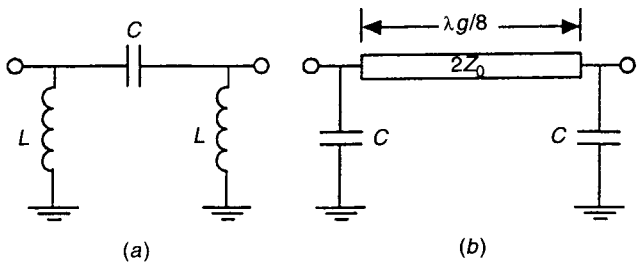
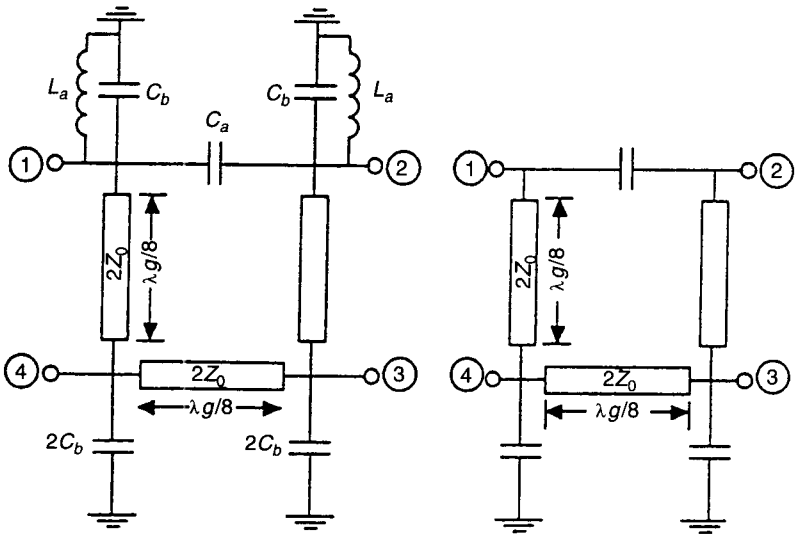


Figure 5.15 (*a*) Lumped-element circuit to replace $\frac{3}{4}\lambda g$ section and (*b*) reduced-size circuit to replace $\frac{1}{4}\lambda g$ section of a rat-race hybrid.



$$L_a = \frac{\sqrt{2}Z_0}{\omega_0} \qquad C_b = \frac{1}{2\omega_0 Z_0}$$

$$C_a = \frac{1}{\sqrt{2}\omega_0 Z_0}$$

Figure 5.16 (*a*) Reduce size rat-race hybrids; ω_0 is the center angular frequency. (*b*) Simplified version of reduced size rat-race hybrid.

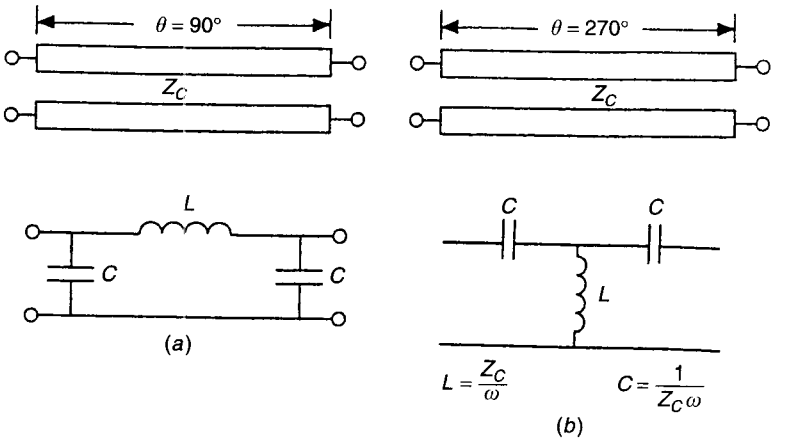


Figure 5.17 (*a*) Equivalent lumped-circuit of a transmission-line section of length $\frac{1}{4}\lambda g$ and (*b*) lumped-element circuit for $\frac{3}{4}\lambda g$.

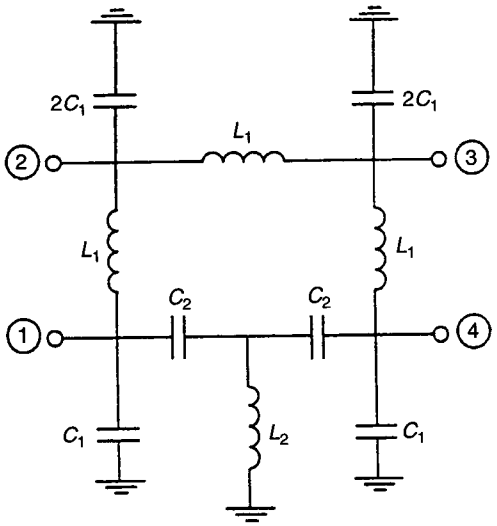


Figure 5.18 Lumped-element 180° hybrid.

by the propagating electromagnetic wave in the primary guide. The holes are spaced such that the round-trip phase shift of a wave through them should be 180°. Therefore the backward-traveling waves in the secondary guide will be completely out of phase to cancel each other at port 3. If the coupled lines have the same propagation constant, then the forward-traveling waves in the secondary guide will be of the same phase regardless of hole spacing and are added at port 4.

Such aperture-coupled directional couplers can be realized using various kinds of planar transmission-line combinations, for example (1) microstrip–microstrip, (2) microstrip–image line, (3) image line–image line, (4) image line–trough guide, or (5) trough guide–trough guide. Some possible combinations are shown in Fig. 5.20.

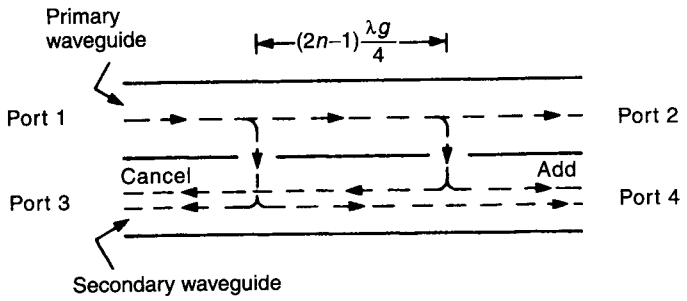


Figure 5.19 Schematic of a two-aperture directional coupler.

Expressions for Single-Aperture Coupling. The design of aperture-coupled directional couplers requires an evaluation of the coupling and the directivity of a simple single-aperture coupler. This is accomplished in the following way.

Let α_e and α_m be the electric and magnetic polarizabilities of the aperture, respectively. Then the fields radiated into the secondary guide by the electric dipole may be expressed as

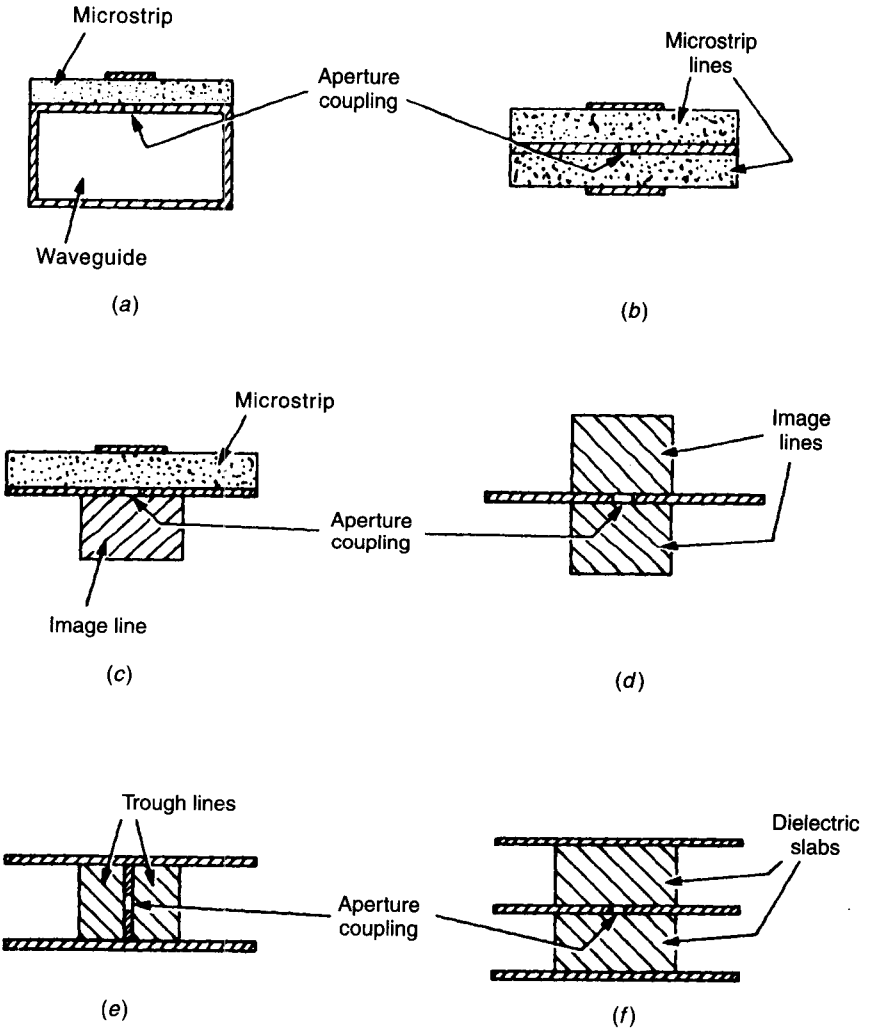


Figure 5.20 Possible combinations of various aperture-coupled lines: (*a*) microstrip–waveguide coupling, (*b*) rnicrostrip–microstrip line, (*c*) microstrip–image line, (*d*) image line–image line, (*e*) trough line--trough line, and (*f*) dielectric guide–dielectric guide.

$$\mathbf{E}_e = \begin{cases} A_1\mathbf{E}^+ & z \geq 0 \\ A_2\mathbf{E}^- & z \leq 0 \end{cases} \tag{5.27a}$$

$$\mathbf{H}_e = \begin{cases} A_1\mathbf{H}^+ & z \geq 0 \\ A_2\mathbf{H}^- & z \leq 0 \end{cases} \tag{5.27b}$$

whereas those radiated by the magnetic dipole are

$$\mathbf{E}_m = \begin{cases} B_1\mathbf{E}^+ & z \geq 0 \\ B_2\mathbf{E}^- & z \leq 0 \end{cases} \tag{5.28a}$$

$$\mathbf{H}_m = \begin{cases} B_1\mathbf{H}^+ & z \geq 0 \\ B_2\mathbf{H}^- & z \leq 0 \end{cases} \tag{5.28b}$$

where the superscript plus denotes propagation along the positive z direction and the superscript minus indicates propagation along the negative z direction. The electric and magnetic fields in the secondary guide are represented by $\mathbf{E}^+$, $\mathbf{E}^-$, $\mathbf{H}^+$, and $\mathbf{H}^-$. The subscripts e and m denote the electric and the magnetic components, respectively.

The expressions for the amplitude coefficients can be written as

$$\begin{aligned} A_1 = A_2 &= -\frac{j\omega}{P_n}\mathbf{E}\cdot\mathbf{P} \\ B_1 = -B_2 &= \frac{j\omega\mu_0}{P_n}\mathbf{H}\cdot\mathbf{M} \end{aligned} \tag{5.29}$$

where ω is the angular frequency, μ_0 is the permeability of free space, and P_n is defined at $z = 0$ as

$$P_n = 2\int_S\int \mathbf{E}^+ \times \mathbf{H}^+ \cdot \hat{z}\, ds \tag{5.30}$$

The electric and magnetic dipole moments $\mathbf{P}$ and $\mathbf{M}$, respectively, are given by

$$\mathbf{P} = -\epsilon_0\overline{\epsilon(f)}\alpha_e\mathbf{E} \tag{5.31a}$$

$$\mathbf{M} = -\alpha_m\mathbf{H} \tag{5.31b}$$

where $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic fields, respectively, in the primary guide and $\overline{\epsilon(f)}$ is the effective dielectric constant at the aperture. An expression for the effective dielectric constant is given as

$$\overline{\epsilon(f)} = \frac{\epsilon_{ep}(f)\epsilon_{es}(f)}{\epsilon_{ep}(f) + \epsilon_{es}(f)} \tag{5.32}$$

where $\epsilon_{ep}(f)$ and $\epsilon_{es}(f)$ are the frequency-dependent effective dielectric constants in the primary and the secondary guides, respectively.

The total field radiated into the secondary guide is given by

$$\mathbf{E} = \begin{cases} (A_1 + B_1)\mathbf{E}^+ & z \geq 0 \\ (A_2 + B_2)\mathbf{E}^- & z \leq 0 \end{cases} \tag{5.33a}$$

$$\mathbf{H} = \begin{cases} (A_1 + B_1)\mathbf{H}^+ & z \geq 0 \\ (A_2 + B_2)\mathbf{H}^- & z \leq 0 \end{cases} \tag{5.33b}$$

For all cases just described, it can be shown that $A_1 + B_1 \neq 0$. This implies that for the single-aperture coupler the forward-coupled wave is never equal to zero, that is, there is always radiation in the forward direction or toward port 4. For no radiation in the backward direction, that is, to port 3, the following condition needs to be satisfied [14, 15]:

$$\frac{\overline{\epsilon(f)}}{\sqrt{\epsilon_{ep}(f)\epsilon_{es}(f)}} = \frac{\alpha_m}{\alpha_e} \cos\theta \tag{5.34}$$

where θ is the angle between the longitudinal axes of the two guides.

For a circular aperture

$$\alpha_e = -\tfrac{2}{3}r^3 \tag{5.35a}$$

$$\alpha_m = \tfrac{4}{3}r^3 \tag{5.35b}$$

where r is the radius of the aperture.

For rectangular slots

$$\bar{\alpha}_e = \frac{\pi d^2 l}{12}\hat{y} \tag{5.36a}$$

$$\bar{\alpha}_m = (0.233dl^2 + 0.044l^3)\hat{x} + \frac{\pi d^2 l}{16}\hat{z} \tag{5.36b}$$

where $d \times l$ are the dimensions of the slot.

The forward coupling coefficient is given by

$$C_f = -20 \log|(A_1 + B_1 \cos\theta)| \tag{5.37a}$$

and the backward coupling coefficient is given by

$$C_b = -20 \log|(A_2 + B_2 \cos\theta)| \tag{5.37b}$$

Using the preceding analysis, expressions for the coupling coefficients for various couplers are shown in Table 5.1.

Table 5.1 Expressions for the Coupling Coefficients for Couplers

For Image Guide to Image Guide[a]

$$C_f = -20\log|(A_1 + B_1\cos\theta)|$$

$$= -20\log\frac{\frac{4}{3}\omega r^3\cos^2(k_x x_0)\left[\frac{1}{2}\epsilon_0\epsilon_r\eta_0 + (\mu_0\cos\theta/\eta_0)\right]}{\{a + [\sin(2k_x a)/2k_x]\}\{b + [\sin(2k_y b)/2k_y]\}}$$

$$C_b = -20\log|(A_2 + B_2\cos\theta)|$$

$$= -20\log\frac{\frac{4}{3}\omega r^3\cos^2(k_x x_0)\left[\frac{1}{2}\epsilon_0\epsilon_r\eta_0 - (\mu_0\cos\theta/\eta_0)\right]}{\{a + [\sin(2k_x a)/2k_x]\}\{b + [\sin(2k_y b)/2k_y]\}}$$

For Microstrip Line to Image Line[b]

$$C_f = -20\log\left|\frac{\frac{4}{3}r^3\omega\left[\frac{1}{2}\epsilon_0\epsilon_r\eta_0 + (\mu_0\epsilon_e(f)/\eta_0)\right]}{a_1 b_1\{1 + [\sin(2k_x a_1)/2k_x a_1]\}\{1 + [\sin(2k_y b_1)/2k_y b_1]\}}\right|$$

$$C_b = -20\log\left|\frac{\frac{4}{3}r^3\omega\left[\frac{1}{2}\epsilon_0\epsilon_r\eta_0 - (\mu_0\epsilon_e(f)/\eta_0)\right]}{a_1 b_1\{1 + [\sin(2k_x a_1)/2k_x a_1]\}\{1 + [\sin(2k_y b_1)/2k_y b_1]\}}\right|$$

$$C_f = -20\log\left|\frac{\frac{4}{3}r^3\omega\left[\frac{1}{2}\epsilon_0\epsilon_r\eta_0 + (\mu_0\epsilon_e(f)/\eta_0)\right]}{hW_e(f)\sqrt{\epsilon_e(f)}}\right|$$

$$C_b = -20\log\left|\frac{\frac{4}{3}r^3\omega\left[\frac{1}{2}\epsilon_0\epsilon_r\eta_0 - (\mu_0\epsilon_e(f)/\eta_0)\right]}{hW_e(f)\sqrt{\epsilon_e(f)}}\right|$$

For Waveguide to Image Line[a]

$$C_f = -20\log\left|\frac{\frac{4}{3}r^3\omega\cos(k_x d)\cos(\pi d/2a)[(\omega/2)\epsilon_0\epsilon_r\eta_0 + \beta\cos\theta]}{\{a_1 + [\sin(2k_x a_1)/2k_x]\}\{b_1 + [\sin(2k_y b_1)/2k_y]\}}\right|$$

$$C_b = -20\log\left|\frac{\frac{4}{3}r^3\omega\cos(k_x d)\cos(\pi d/2a)[(\omega/2)\epsilon_0\epsilon_r\eta_0 - \beta\cos\theta]}{\{a_1 + [\sin(2k_x a_1)/2k_x]\}\{b_1 + [\sin(2k_y b_1)/2k_y]\}}\right|$$

For Image Line to Waveguide[c]

$$C_f = -20\log\left|\frac{\frac{4}{3}r^3\omega\cos(k_x d)\cos(\pi d/2a)[(\epsilon_0\epsilon_r/2) + \mu_0 Y_w\cos\theta/\eta_0]}{2ab\,Y_w}\right|$$

$$C_b = -20\log\left|\frac{\frac{4}{3}r^3\omega\cos(k_x d)\cos(\pi d/2a)[(\epsilon_0\epsilon_r/2) - \mu_0 Y_w\cos\theta/\eta_0]}{2ab\,Y_w}\right|.$$

[a] For definition of parameters, see Lange [21].

[b] For definition of parameters, see [13].

[c] For definition of parameters, see [15].

Effects of Large-Size Aperture and Finite Conductor Thickness. The preceding expressions for coupling ignored the effects of the finite common ground plane thickness and large aperture size. It is assumed that the aperture's major dimension is considerably smaller than a quarter wavelength. The correction factor for both these coupling factors can be expressed as [14, 16]

$$C_F = \frac{\exp\{(-2\pi tA/\lambda_c)\sqrt{[1-(\lambda_c/\lambda_0)^2]}\}}{1-(\lambda_c/\lambda_0)^2} \tag{5.38}$$

where t is the thickness of the ground plane, λ_c is the cutoff wavelength of the aperture for the appropriate mode of operation excited, λ_0 is the operating wavelength, and factor A accounts for the interactions of local fields on either side of the aperture. Factor A has been determined empirically to be 3 for a narrow slot and unity for a circular aperture. Cutoff wavelengths λ_c for rectangular and circular apertures are given in Table 5.2. The proper values for the coupling factors with the thickness of ground plane and larger aperture effect taken into account are obtained by multiplying A_1, B_1, A_2, and B_2 by C_F with λ_c related to the type of coupling involved.

The coupling characteristics of single-hole couplers using image line–image line reported by Bahl and Bhartia [16] are shown in Fig. 5.21*a*.

Design of Directional Coupler. The design of a directional coupler requires the use of an array of apertures to obtain good directivity and coupling. For $N+1$ apertures the forward and backward coupled waves are given by [15]

$$C_f = Ae^{-jN\beta_1 d}\sum_{n=0}^{N} C_n e^{-jn(\beta_2-\beta_1)d} \tag{5.39a}$$

$$C_b = A\sum_{n=0}^{N} D_n e^{-jn(\beta_1+\beta_2)d} \tag{5.39b}$$

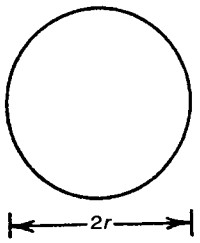
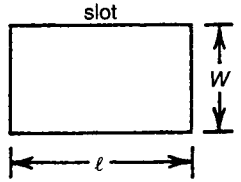
where $\beta_1 = \sqrt{\epsilon_{ep}(f)}k_0$, $\beta_2 = \sqrt{\epsilon_{es}(f)}k_0$ ($k_0 = 2\pi/\lambda_0$, and β_1 and β_2 are the propagation constants in the primary and the secondary guide, respectively), and A is a constant.

For a symmetrical array of apertures $r_i = r_{N-i}$ and expressing $C_n = d_n\tau_f$ and $D_n = d_n\tau_b$ where d_n is a frequency-independent term, the forward and backward coupling coefficients are given by

$$\mathrm{CF} = -20\log|\tau_f| - 20\log\left|\sum_{n=0}^{M} 2d_n\cos(N-2n)(\theta-\beta_1 d)\right| \tag{5.40a}$$

$$\mathrm{CB} = -20\log|\tau_b| - 20\log\left|\sum_{n=0}^{M} 2d_n\cos(N-2n)\theta\right| \tag{5.40b}$$

Table 5.2 Cutoff Wavelengths for Circular and Rectangular Apertures

Aperture	Cutoff Mode	λ_c Magnetic Coupling	λ_c Electric Coupling
Circular hole 	TE_{11}	$3.41\sqrt{\epsilon_r}r$	
	TM_{01}		$2.61\sqrt{\epsilon_r}r$
Rectangular slot 	TE_{10}	$2\sqrt{\epsilon_r}l$	
	TM_{11}		$\dfrac{\sqrt{\epsilon_r}}{[(\pi/l)^2 + (\pi/W)^2]^{1/2}}$

where

$$\theta = \tfrac{1}{2}(\beta_1 + \beta_2)d \tag{5.40c}$$

Hence, the directivity is given by

$$D = \text{CB} - \text{CF} = -20 \log\left|\frac{\tau_b}{2\tau_f}\right| - 20 \log\left|\frac{\sum_{n=0}^{M} 2d_n \cos(N-2n)\theta}{\sum_{n=0}^{M} 2d_n \cos(N-2n)(\theta - \beta_1 d)}\right| \tag{5.41}$$

where CB, the total backward coupled wave, is zero for $\theta = \frac{1}{2}\pi$ at the center frequency, and hence $M = \frac{1}{2}(N-1)$ for N odd and $M = \frac{1}{2}N$ for N even, and the aperture spacing

$$d = \pi(\beta_1 + \beta_2)^{-1} \tag{5.42}$$

An examination of (5.41) shows that unless β_1 and β_2 are widely different, the numerator of the right-hand side has a much faster variation than the denominator. Hence an approximate broadband coupler design is possible by equating

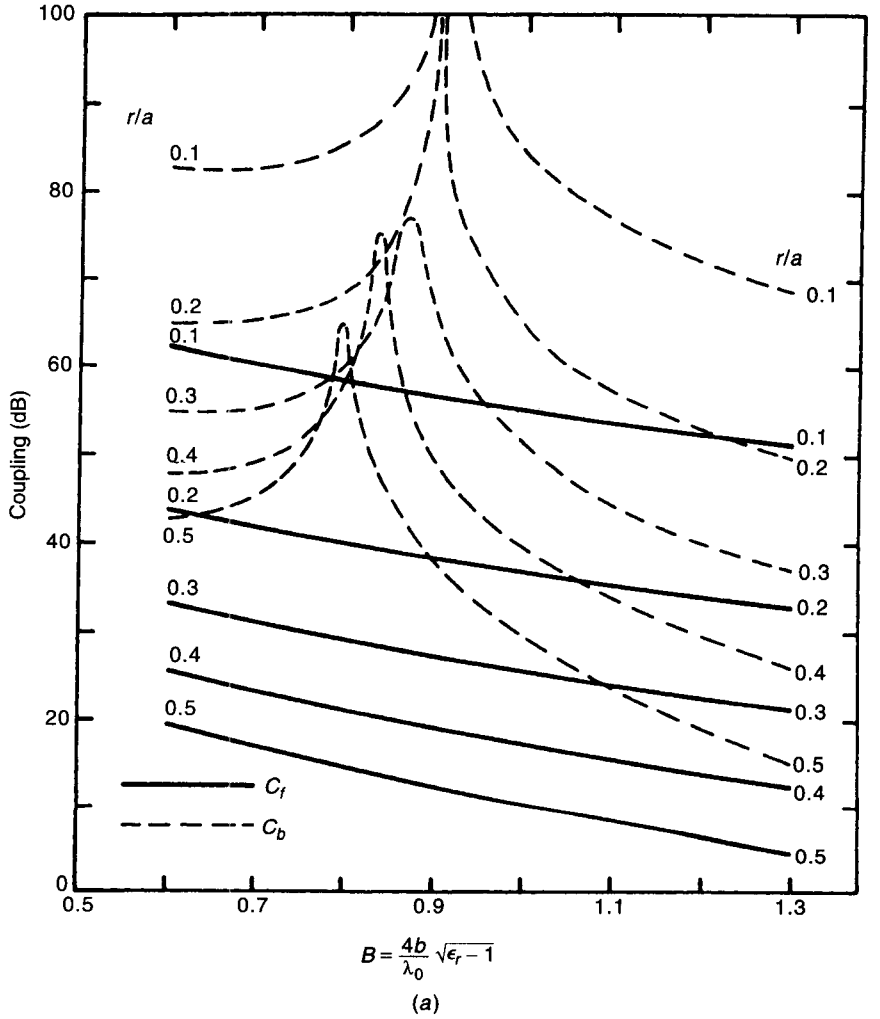
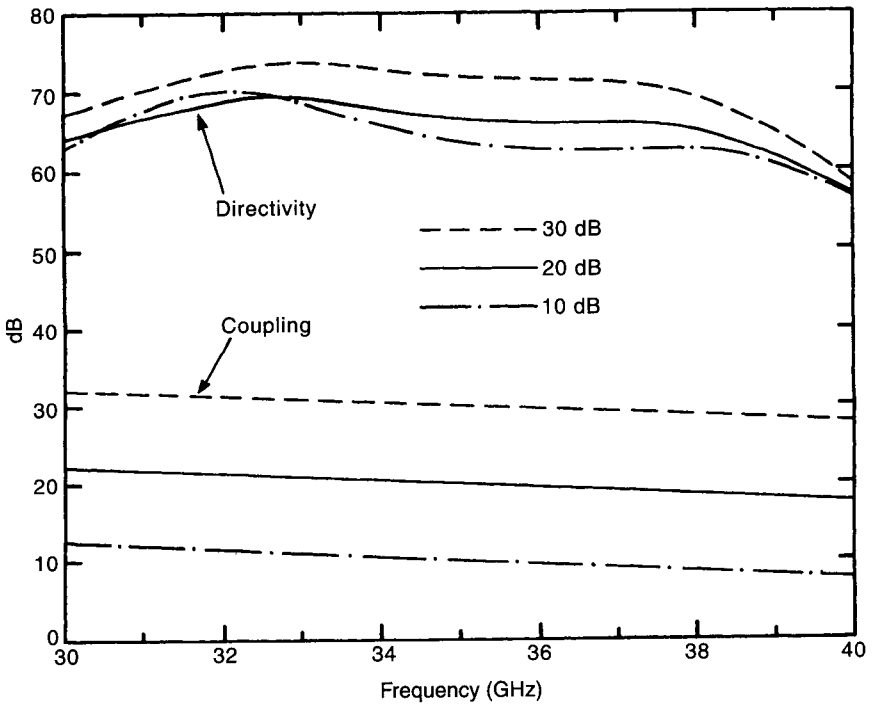


Figure 5.21 (*a*) Characteristics of single-aperture image-guide coupler. (*b*) Frequency response of aperture-coupled image-guide coupler.

the numerator to an appropriate polynomial according to the desired frequency response of the directivity. We illustrate the method for an approximate Chebyshev-type response. Only the design steps are given. For a detailed derivation of the method, the reader is referred to Ref. [17].

We write from (5.41)

$$F \approx \left| \sum_{n=0}^{M} 2d_n \cos(N - 2n)\theta \right| = K|T_N(\sec \theta_m \cos \theta)| \tag{5.43}$$



(*b*)

Figure 5.21 *(Continued)*

The minimum value of directivity D_m in the passband as contributed by the array factor F is given by

$$D_m = 20 \log|T_N(\sec \theta_m)| \tag{5.44}$$

The Chebyshev polynomial, $T_N(x)$, is of order N with an argument x, and θ_m is the value at the upper and lower edges of the passband. According to (5.44), therefore, if D_m is specified, then θ_m automatically gets fixed and vice versa. Or in other words, the bandwidth and the minimum passband directivity fix each other.

The constant K is chosen to give the desired value of coupling C at the center frequency. This is done in the following way. The coupling at the center frequency is given by

$$C = -20 \log[K|\tau_f|\,|T_N(\sec \theta_m)|] \tag{5.45}$$

Polynomial τ_f is obtained by dividing C_f by the frequency-independent term in C_f. For example, for a circular aperture, τ_f is obtained from

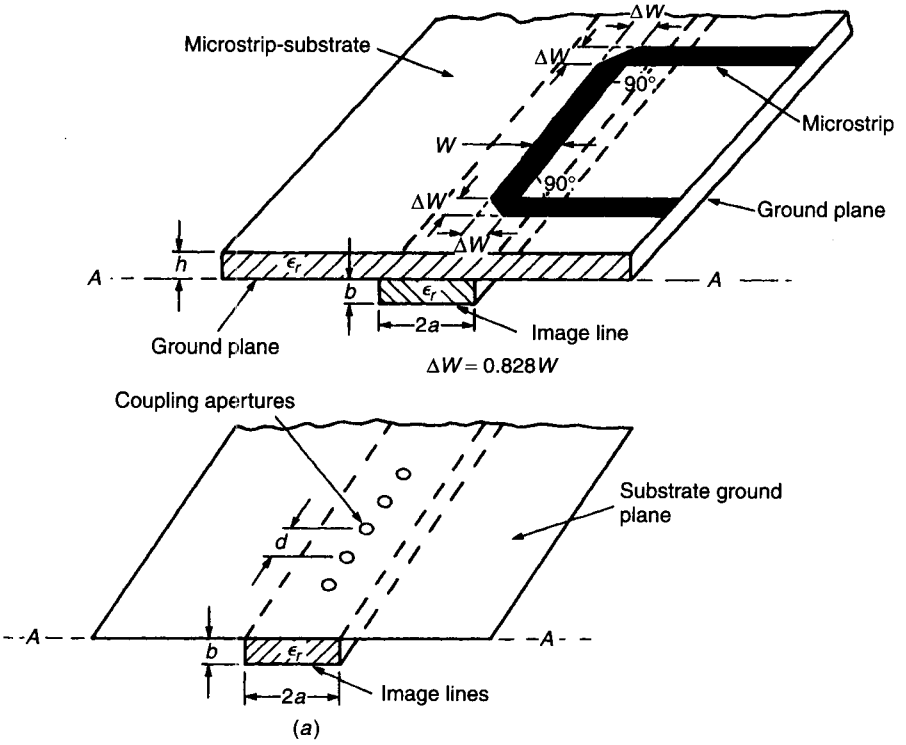


Figure 5.22 (*a*) Conceptual design drawings of microstrip–image guide coupler. (*b*) Frequency response of microstrip–image-line coupler.

$$\tau_f = \frac{10^{-C_f/20}}{r^3} \tag{5.46}$$

Knowing C and τ_f, K is obtained from (5.45). Having obtained K, the aperture dimensions are obtained from (5.43) by equating the coefficients of $\cos n\theta$ $(n = 0, \ldots, M)$.

Using the preceding procedure, an image guide–image guide broadband coupler was designed for operation in the *Ka* band. The computed response is shown in Fig. 5.21*b* for 10-, 20-, and 30-dB couplers.

A conceptual design drawing and the characteristics are shown in Figs. 5.22*a* and *b*, respectively, for a *Ka*-band microstrip–image guide aperture-coupled coupler.

5.3.2 TEM Line Directional Couplers

When the center conductors of two coaxial lines supporting a pure TEM mode of propagation are brought in close proximity of each other, electromagnetic

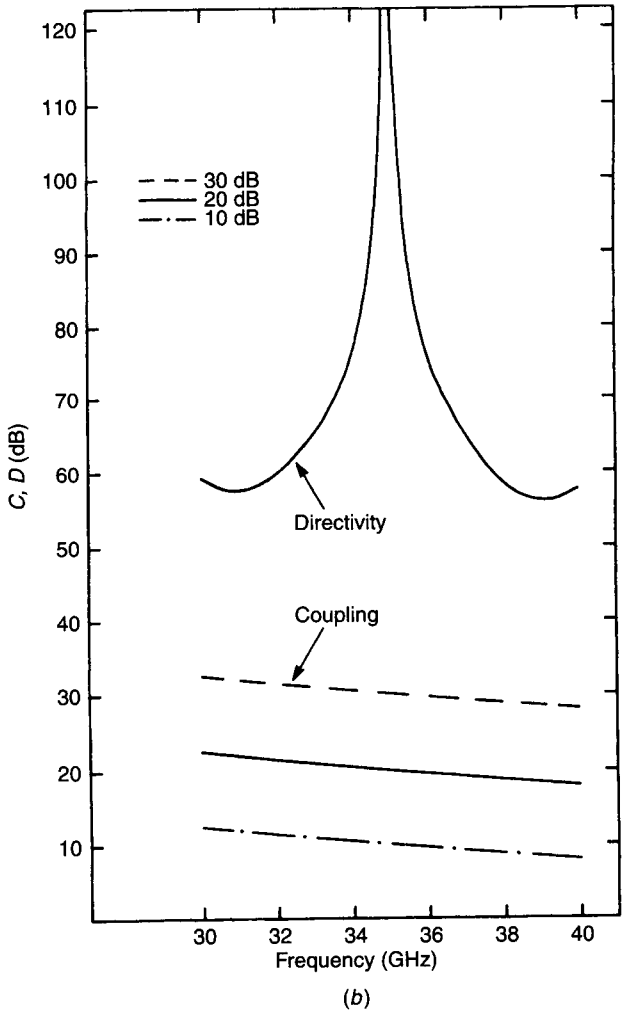


Figure 5.22 *(Continued)*

energy is coupled from one line to the other. This property has given rise to a class of broadband planar directional couplers. Most such couplers use stripline or microstrip transmission lines that support pure TEM or quasi-TEM modes.

Planer TEM line directional couplers can be either edge coupled or broadside coupled as shown in Figs. 5.23*a* and *b*.

In general, such coupled TEM lines support two modes that interact to give rise to the coupling. These are the even and odd modes. Properties of the cou-

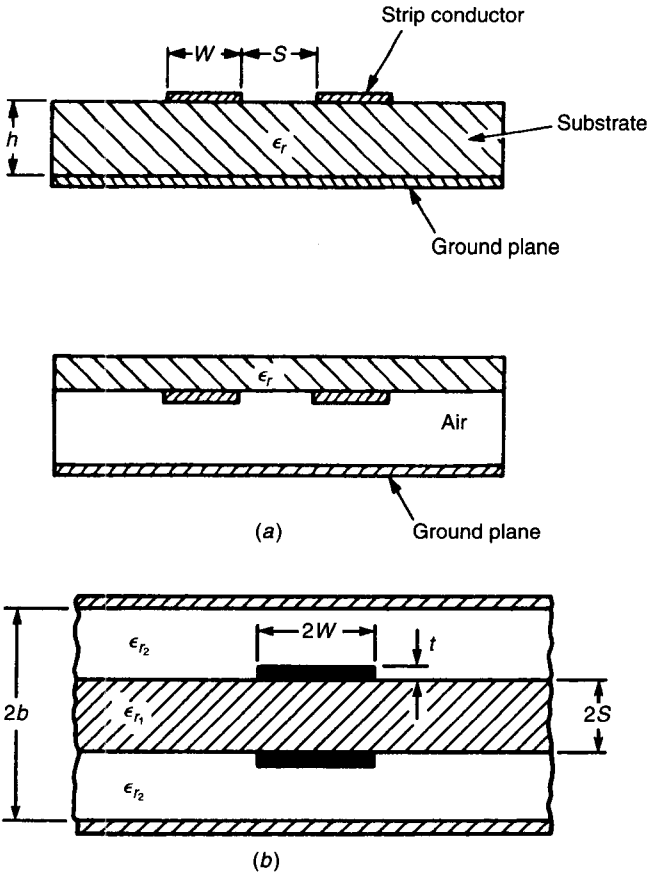


Figure 5.23 (*a*) Edge-coupled microstrip lines. (*b*) Broadside-coupled microstrip line.

pled lines can be evaluated by suitable linear combinations of even and odd modes.

TEM line couplers can be reduced to a four-port network as shown in Fig. 5.24*a*. There is a plane of symmetry that becomes a perfect magnetic wall for the incident signals of equal amplitude and same phase at ports 1 and 4.

The plane of symmetry becomes a perfect electric wall for the incident signals of equal amplitude but of exactly opposite phases at the same ports. Therefore each mode corresponds to a two-port network as shown in Figs. 5.24*b* and *c* for even and odd modes, respectively. Analysis of the directional coupler is accomplished by analyzing these two networks and superimposing the responses as shown in Fig. 5.24*d*.

The *S*-matrix equation of a symmetrical and reciprocal network can be

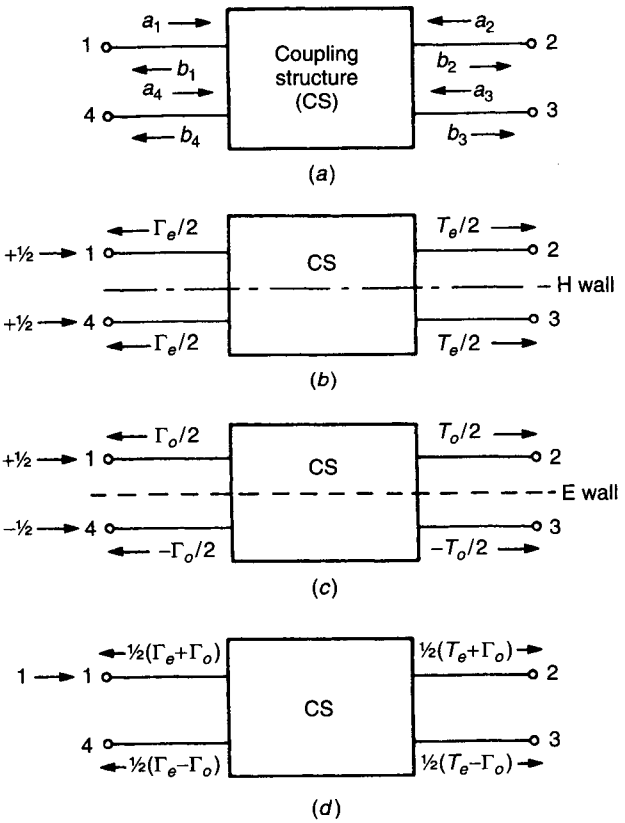


Figure 5.24 Schematic of directional coupler: (*a*) wave parameters, (*b*) even-mode excitation, (*c*) odd-mode excitation, and (*d*) input excitation.

written as

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{12} & S_{11} & S_{14} & S_{13} \\ S_{13} & S_{14} & S_{11} & S_{12} \\ S_{14} & S_{13} & S_{12} & S_{11} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \tag{5.47}$$

For the even-mode excitation $a_1 = a_4 = \frac{1}{2}$. The corresponding reflection coefficient Γ_e and the transmission coefficient T_e are given by

$$\Gamma_e = \frac{b_1}{a_1} = \frac{b_4}{a_4} = S_{11} + S_{14} \tag{5.48a}$$

$$T_e = \frac{b_2}{a_1} = \frac{b_3}{a_4} = S_{12} + S_{13} \tag{5.48b}$$

For the odd-mode excitation $a_4 = -a_1 = -\frac{1}{2}$. The reflection and the transmission coefficients may be written as

$$\Gamma_o = \frac{b_1}{a_1} = \frac{b_4}{a_4} = S_{11} - S_{14} \tag{5.49a}$$

$$T_o = \frac{b_2}{a_1} = \frac{b_3}{a_4} = S_{12} - S_{13} \tag{5.49b}$$

Superposition of the preceding solution gives

$$S_{11} = \tfrac{1}{2}(\Gamma_e + \Gamma_o) \tag{5.50a}$$

$$S_{12} = \tfrac{1}{2}(T_e + T_o) \tag{5.50b}$$

$$S_{13} = \tfrac{1}{2}(T_e - T_o) \tag{5.50c}$$

$$S_{14} = \tfrac{1}{2}(\Gamma_e - \Gamma_o) \tag{5.50d}$$

Evaluation of the even- and odd-mode reflection and transmission coefficients is done from the corresponding effective dielectric constants and the characteristic impedances:

$$\Gamma_i = \frac{A_i + B_i/Z_0 - C_i Z_0 - D_i}{A_i + B_i/Z_0 + C_i Z_0 + D_i} \tag{5.51a}$$

$$T_i = \frac{2}{A_i + B_i/Z_0 + C_i Z_0 + D_i} \tag{5.51b}$$

where the transmission matrix

$$\begin{bmatrix} A_i & B_i \\ C_i & D_i \end{bmatrix} = \begin{bmatrix} \cos\theta_i & jZ_{0i}\sin\theta_i \\ \dfrac{j\sin\theta_i}{Z_{0i}} & \cos\theta_i \end{bmatrix} \qquad i = e, o \tag{5.52}$$

Knowing $Z_{0e}, \theta_e, Z_{0o}, \theta_o$, and Z_0, and the system characteristic equation, the performance of the directional coupler can be calculated by using (5.50) and (5.51). For the computation of the even- and the odd-mode characteristic impedance and phase velocity in terms of the physical parameters, the reader is referred to Section 2.2.

Coupled TEM Line. For coupling between the purely TEM lines shown in Fig. 5.25, we have the following special case:

$$\theta_e = \theta_o = \theta \tag{5.53}$$

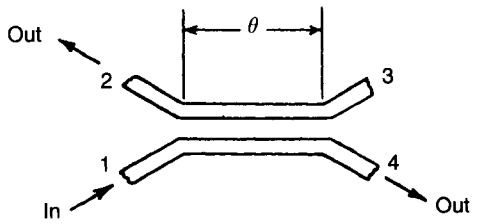


Figure 5.25 Coupling between two TEM lines.

which means equal, even-, and odd-mode phase velocities, which gives

$$Z_0 = \sqrt{Z_{0e}Z_{0o}} \tag{5.54}$$

$$\Gamma_e = -\Gamma_o = \frac{j[(Z_{0e}/Z_{0o})^{1/2} - (Z_{0o}/Z_{0e})^{1/2}] \sin\theta}{\Sigma} \tag{5.55}$$

where

$$\Sigma = 2\cos\theta + j\left[\left(\frac{Z_{0e}}{Z_{0o}}\right)^{1/2} + \left(\frac{Z_{0o}}{Z_{0e}}\right)^{1/2}\right]\sin\theta \tag{5.56}$$

Substitution of (5.55) and (5.56) into (5.50) gives

$$S_{11} = 0 \tag{5.57a}$$

$$S_{12} = T_e \tag{5.57b}$$

$$S_{13} = 0 \tag{5.57c}$$

$$S_{14} = T_e \tag{5.57d}$$

Such couplers are known as backward wave couplers, and are in general a quarter wavelength long at the center frequency or $\theta = \frac{1}{2}\pi$.

From (5.54) and (5.55) the coupling coefficient is given by

$$C = -20\log|S_{14}| = -20\log\left|\frac{Z_{0e} - Z_{0o}}{Z_{0e} + Z_{0o}}\right| \quad \text{dB} \tag{5.58}$$

Moreover, for a system impedance Z_0 of 50 Ω, the matching condition (condition for a perfect input match) gives

$$Z_0^2 = Z_{0e}Z_{0o} \tag{5.59}$$

Therefore, for a specified coupling C the design equations become

$$Z_{0e} = Z_0 \left[\frac{1 + 10^{-C/20}}{1 - 10^{-C/20}}\right]^{1/2} \tag{5.60a}$$

$$Z_{0o} = Z_0 \left[\frac{1 - 10^{-C/20}}{1 + 10^{-C/20}}\right]^{1/2} \tag{5.60b}$$

Once Z_{0e} and Z_{0o} are known, the physical dimensions of the coupler can be obtained using the equations for coupled TEM transmission lines.

Coupled Quasi-TEM Line. In coupled quasi-TEM lines, for example, in microstrip, the odd-mode phase velocity is different from the even-mode phase velocity. Therefore, the condition given by (5.53) does not hold good. However, for weak couplings, (5.53) can be assumed to be approximately true. Hence the designer may go ahead and determine the initial design using (5.58)–(5.60). However, as the coupling gets tighter, the previous equations tend to be less valid. In such a case, the condition for input matching becomes

$$Z_0 = \left(\frac{Z_{0e} \sin \theta_e + Z_{0o} \sin \theta_o}{Z_{0e} \sin \theta_o + Z_{0o} \sin \theta_e}\right)^{1/2} \sqrt{Z_{0o} Z_{0e}} \tag{5.61}$$

and the electrical length at the center frequency is

$$\theta = \frac{1}{2}(\theta_e + \theta_o) = \frac{2\pi}{\lambda_0} \frac{\sqrt{\epsilon_{ee}} + \sqrt{\epsilon_{eo}}}{2} l = 90^\circ \tag{5.62}$$

where l is the physical length of the coupler.

Single-section quarter-wave parallel-coupled line couplers are used extensively in many applications. They are usually of narrow bandwidth of approximately one octave. To obtain the desired coupling at the band edges, the coupler has to be designed for overcoupling at the center frequency.

Frequency Response of a Single-Section Coupler. Using the analysis equations (5.50)–(5.58), it can be shown that the frequency response of the coupling coefficient is given by

$$C(\theta) = \frac{jC \sin \theta}{\sqrt{1 - C^2} \cos \theta + j \sin \theta} \tag{5.63a}$$

where C is the midband coupling for a matched, loosely coupled coupler. The approximate frequency response of a quasi-TEM coupler is shown in Fig. 5.26. The response is exact for a TEM coupler.

The general expression for the directivity or the undesired coupling is given

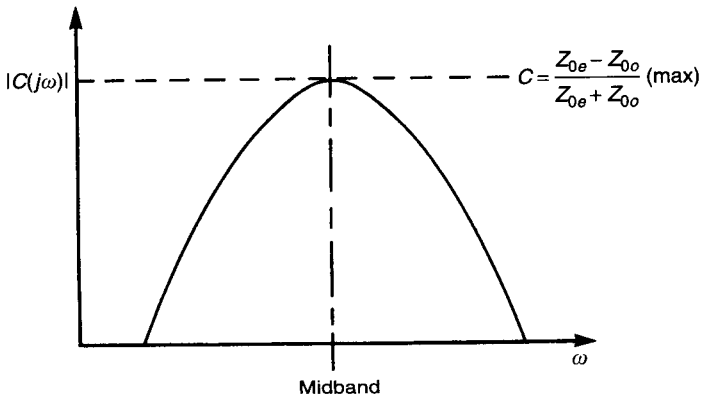


Figure 5.26 Approximate frequency response of a quasi-TEM line.

by [18]

$$D = \left[\frac{\pi\Delta(1 - |\xi|^2)}{4|\xi|}\right]^2 \tag{5.63b}$$

where

$$\Delta = \frac{\beta_e - \beta_o}{\beta_o}$$

$$\xi = \left(\frac{\rho_e}{1 + \rho_e^2}\right) - \left(\frac{\rho_o}{1 + \rho_o^2}\right)$$

$$\rho_e = \frac{Z_{0e} - Z_0}{Z_{0e} + Z_0} \qquad \rho_o = \frac{Z_{0o} - Z_0}{Z_{0o} + Z_0}$$

and β_e and β_o are the even- and odd-mode propagation constants, respectively, $D = 0$ for $\beta_e = \beta_o$, that is, for a TEM coupler.

Multisection Couplers for Wider Bandwidth. For many applications, the single-section coupler proves to be of inadequate bandwidth. Therefore, the designer should have recourse to a multisection design. A multisection coupler is a cascaded combination of more than one single-section coupler, each being a quarter wavelength long at the center frequency of the band. The number of sections to be used depends upon the tolerable insertion loss, bandwidth, and the available physical space.

Multisection couplers can be either symmetric or asymmetric around the center section, as shown in Fig. 5.27.

Symmetric Coupler. The symmetric coupler gives 90° phase difference between the direct and the coupled output ports under matched conditions. In what follows we present a direct synthesis technique.

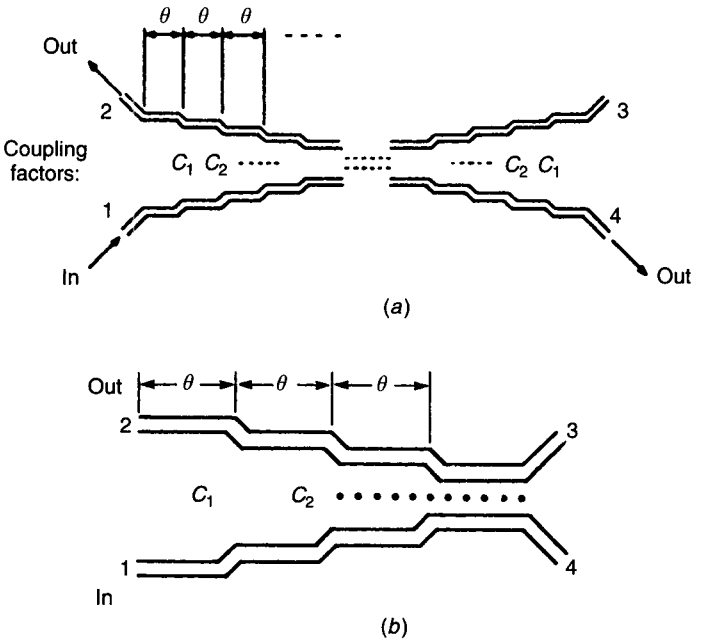


Figure 5.27 Multisection (*a*) symmetrical and (*b*) asymmetrical couplers.

Figure 5.27*a* shows an n-section symmetric coupler. The coupling factor for such an n-section symmetric coupler with weak coupling can be written as [19]

$$\begin{aligned} C(\theta) &= \left|\frac{V_2}{V_1}\right| \\ &= C_1 \sin(n\theta) + (C_2 - C_1)\sin[(n-2)\theta] + \cdots \\ &\quad + (C_i - C_{i-1})\sin[(n-2i+2)\theta] \\ &\quad + (C_{[(n+1)/2]} - C_{[(n-1)/2]})\sin\theta \end{aligned} \tag{5.64}$$

If the desired coupling response is maximally flat, then C_i must satisfy a set of $\frac{1}{2}(n-1)$ linear equations obtained from

$$\left[\frac{d^r C(\theta)}{d\theta^r}\right]_{\theta=\pi/2} = 0 \qquad r = 2, 4, 6, \ldots, n-1 \tag{5.65}$$

Note that n is always an odd integer.

The design concept is based on the fact that the backward coupled wave of a TEM coupler corresponds to the reflected wave of a quarter-wave filter. Therefore the designer of a TEM wave coupler has to synthesize only a two

port in place of a four port with the reflection coefficient response the same as the desired coupling coefficient response of the four-port directional coupler.

In terms of the midband VSWR R of the quarter-wave filter, the coupling coefficient is given by [19]

$$C_0 = \frac{R-1}{R+1} \tag{5.66}$$

Design of Three-Section Maximally Flat Coupler. From (5.66) we obtain

$$R = \frac{1+C_0}{1-C_0} \tag{5.67}$$

The step VSWRs, V_1 and V_2, are calculated from

$$V_1 = 1.1592 - 0.01666C_0 + 0.000474C_0^2 \tag{5.68}$$

where C_0 is in decibels and $V_2 = V_1\sqrt{R}$.

The step impedances of the quarter-wave filter are determined as

$$Z_1 = V_1 \tag{5.69a}$$

$$Z_2 = V_1 V_2 \tag{5.69b}$$

Couplings of the individual sections are given by

$$C_1 = \frac{Z_1^2 - 1}{Z_1^2 + 1} \tag{5.70a}$$

$$C_2 = \frac{Z_2^2 - 1}{Z_2^2 + 1} \tag{5.70b}$$

The odd- and even-mode impedances are obtained from

$$(Z_{0e})_i = Z_0\left[\frac{1+C_i}{1-C_i}\right]^{1/2} \tag{5.71a}$$

$$(Z_{0o})_i = Z_0\left[\frac{1-C_i}{1+C_i}\right]^{1/2} \qquad i = 1,2 \tag{5.71b}$$

Once $(Z_{0e})_i$ and $(Z_{0o})_i$ are known, the physical dimensions can be obtained depending upon the type of transmission line. The preceding design method holds exactly for purely TEM lines. However, a multisection symmetric micro-

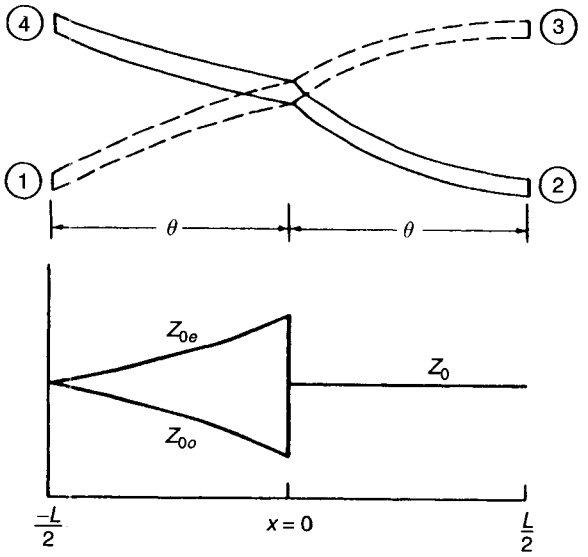


Figure 5.28 Asymmetric coupler.

strip coupler can be designed approximately with the help of the previous method. Then the design can be improved using optimization techniques.

Asymmetric Coupler. The asymmetric coupler provides a phase difference of 0° or 180° between the coupled and the direct output ports over a wide frequency band. In such couplers the two transmission lines become more lightly coupled gradually or in steps until they reach the center of the coupler, where they become uncoupled, as shown in Fig. 5.28. The scattering matrix for the asymmetric coupler is obtained in the following way.

If ports 2 and 3 are excited by a pair of even- and odd-mode generators, then analogous to (5.50), we obtain the scattering parameters at ports 2 and 3 as

$$S_{22} = S_{33} = \tfrac{1}{2}(\Gamma_{2e} + \Gamma_{2o}) \tag{5.72a}$$

$$S_{23} = S_{32} = \tfrac{1}{2}(\Gamma_{2e} - \Gamma_{2o}) \tag{5.72b}$$

Now if one assumes that the coupler is matched, so that

$$C = \frac{Z_0}{Z_{0e}(0)} = \frac{Z_{0o}(0)}{Z_0} \tag{5.73}$$

then the maximum coupling is at $x = 0$, where there is a discontinuity due to an abrupt change from either Z_{0e} or Z_{0o} to Z_0. Therefore the even- and the odd-

mode reflection coefficients at input port 1 are given by [20]

$$\Gamma_{1e} = \frac{C-1}{C+1} e^{-j2\theta} \tag{5.74a}$$

$$\Gamma_{1o} = -\Gamma_{1e} = \frac{1-C}{1+C} e^{-j2\theta} \tag{5.74b}$$

Looking from port 2, a transmission line of characteristic impedance Z_0 is terminated either by Z_{0e} or Z_{0o} at $x = 0$. Therefore the even-mode reflection coefficient at port 2 is given by

$$\Gamma_{2e} = -\Gamma_{2o} = \frac{1-C}{1+C} e^{-j2\theta} \tag{5.74c}$$

For the energy conservation principle and the electrical length of the coupler, the transmission coefficients are given by

$$T = \sqrt{1 - |\Gamma|^2} e^{-j2\theta} \tag{5.75a}$$

$$T_{1e} = T_{1o} = \frac{2\sqrt{C}}{1+C} e^{-j2\theta} \tag{5.75b}$$

Therefore the scattering matrix is given by

$$[S] = \begin{bmatrix} 0 & p & 0 & -q \\ p & 0 & q & 0 \\ 0 & q & 0 & p \\ -q & 0 & p & 0 \end{bmatrix} \tag{5.76}$$

where

$$p = \frac{2\sqrt{C}}{1+C} e^{-j2\theta} \tag{5.77a}$$

$$q = \frac{1-C}{1+C} e^{-j2\theta} \tag{5.77b}$$

From (5.76) the relative amplitudes of the waves in the coupled and uncoupled ports are given by coupling response C_∞ at infinite frequency.

The design of the tapered asymmetric coupler can be obtained using optimization programs [21]. But the Klopfenstein [22] taper seems to be the easiest to realize for the best performance.

Three main parameters of an asymmetric coupler are the cutoff frequency f_0, the sidelobe level in the passband, and the physical length of the coupler. Out of these, any two can be chosen by the designer.

The even-mode impedance distribution is obtained from

$$\ln Z(x) = \tfrac{1}{2}\ln(Z_1 Z_2) + \frac{A^2 \ln(Z_2/Z_1)}{2\cosh(A)}\Phi\left(\frac{2x}{L}, A\right) \tag{5.78}$$

where Z_1 and Z_2 are the even-mode impedances at the ends of the asymmetric coupler, and

$$\Phi(x, A) = \sum_{n=0}^{\infty} a_n b_n \tag{5.79}$$

$$a_0 = 1 \qquad b_0 = \tfrac{1}{2}x$$

$$a_n = \frac{A^2}{4n(n+1)} a_{n-1} \tag{5.80a}$$

$$b_n = \frac{(x/2)(1-x^2)^n + 2nb_{n-1}}{2n+1} \tag{5.80b}$$

Usually $Z_1 = 1$ and Z_2 is obtained from the desired coupling response C_∞ at infinite frequency.

$$Z_2 = \frac{Z_{0e}(0)}{Z_0} = \frac{1 + |C_\infty|}{1 - |C_\infty|} \tag{5.81}$$

Once Z_{0e} has been obtained, Z_{0o} is obtained from (5.59). Suitable taper distribution is used for optimum performance [21, 22]. For each set of Z_{0e} and Z_{0o} the required physical dimensions of the coupler are obtained using a suitable synthesis technique depending upon the type of coupled transmission lines.

5.3.3 Multiconductor Couplers

The interdigital coupler, or multiconductor coupler, invented by Lange [23], has always been a popular component in planar circuits. Figure 5.29 shows a four-element interdigital coupler, although in certain applications the number of elements may be greater than four. The coupler is usually designed for 3 dB coupling and the output phases are in quadrature. Obviously, the best realization is in the microstrip form.

An interdigital coupler has advantages because of its small size and relatively large line separation when compared with the two-coupled line device and has a much larger bandwidth when compared with branch-line couplers.

Interdigital couplers are used for balanced MIC amplifiers, balanced mixers, and binary power divider trees.

Design of Interdigital Couplers. Kajfez et al. [24] have described a simplified design technique for the interdigital coupler. The proposed technique

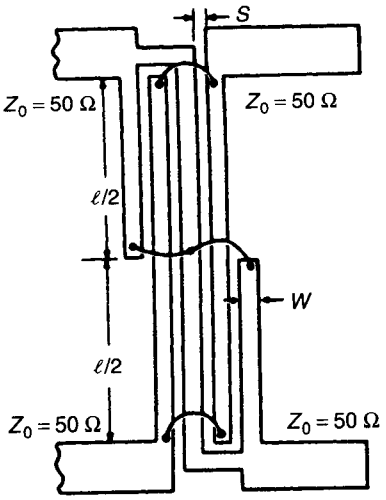


Figure 5.29 Lange coupler.

serves many practical purposes but seems to be inadequate for an accurate design. So far the method due to Presser [25] has been found to be the most accurate and simplest. Moreover, it has the provision for finite strip-thickness correction.

Consider the layout shown in Fig. 5.29. The designer is supplied with the desired coupling coefficient, C, and the system characteristic impedance Z_0. The length of the coupled region l has to be a quarter wavelength at the center of the band.

The main design equations for an N element (N even) coupler are written as

$$R = \frac{Z_{0o}}{Z_{0e}} \tag{5.82a}$$

$$C = \frac{(N-1)(1-R^2)}{(N-1)(1+R^2)+2R} \tag{5.82b}$$

$$Z = \frac{Z_{0o}}{Z_0}\frac{\sqrt{R[(N-1)+R][(N-1)R+1]}}{1+R} \tag{5.82c}$$

Figure 5.30 shows the impedance ratio R as a function of the coupling coefficient, while Fig. 5.31 shows the normalized odd-mode impedance as a function of the coupling coefficient. Equations (5.82a)–(5.82c) can be solved for Z_{0o} and Z_{0e}. Knowing Z_{0o} and Z_{0e}, the physical shape ratios can be obtained using the design equations of Garg and Bahl [26].

When the finite thickness of conductors is neglected, the designed coupler shows an overcoupled characteristic. This is because the conductor thickness considerably influences the coupling between the lines. The overcoupling can be

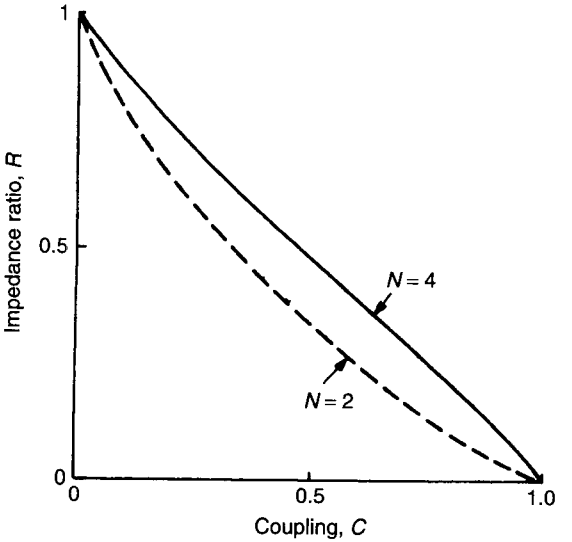


Figure 5.30 Impedance ratio vs. coupling coefficient of a Lange coupler.

corrected by adding an extra gap ΔS to S expressed as

$$\frac{\Delta S}{h} = \frac{t/h}{\pi\sqrt{\epsilon_{eo}}}\left(1 + \ln\frac{4\pi W/h}{t/h}\right) \tag{5.83}$$

where t/h is the actual normalized thickness of the metallization and ϵ_{eo} is the odd-mode effective dielectric constant of the coupled lines.

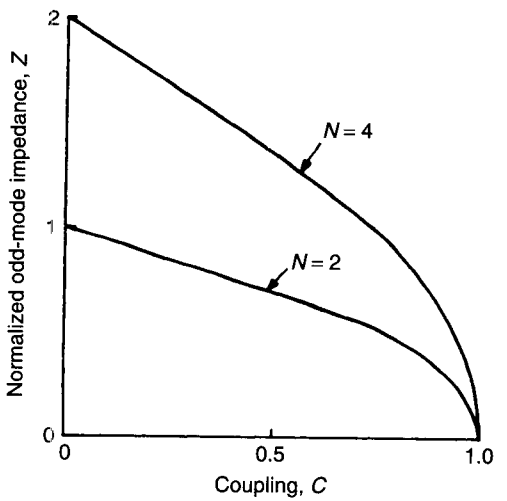


Figure 5.31 Normalized odd-mode impedance vs. coupling coefficient of a Lange coupler.

5.3.4 Distributed-Type Couplers

Distributed-type coupling takes place between two adjacent transmission lines supporting purely non-TEM modes. For example, distributed-type couplers can be realized using two open dielectric waveguides (or image guides) or fin lines.

In general, two distributed-type coupled lines can be represented as shown in Fig. 5.32*a*. Under the assumption that all four ports are matched and the coupling structure has the length l, the ratio of the fields in the two lines can be shown to be [14]

$$\frac{E_b(l)}{E_a(l)} = \tan\left((\beta_e - \beta_o)\frac{l}{2}\right) \tag{5.84}$$

where β_e and β_o are the even- and the odd-mode phase constants, respectively.

Therefore for complete transfer of power from line a to line b requires

$$(\beta_e - \beta_o)\frac{L}{2} = \frac{\pi}{2} \tag{5.85}$$

or

$$L = \frac{\pi}{\beta_e - \beta_o} \tag{5.86}$$

The scattering coefficients of the structure can be written as a function of l normalized by L as

$$|S_{12}| = \left|\cos\left(\frac{\pi}{2}\frac{l}{L}\right)\right| \tag{5.87a}$$

$$|S_{13}| = \left|\sin\left(\frac{\pi}{2}\frac{l}{L}\right)\right| \tag{5.87b}$$

Equation (5.87b) shows that the required length for 3 dB coupling is one half of L.

The preceding equations are based on the assumptions that the bent portion of the guides have no effect on coupling. Branch junction effects near the bends are negligible, but the assumed uncoupled lines $z \leq 0$ and $z \geq z_0$ do get coupled. This extra coupling can be taken into consideration by defining the effective length of the coupler as

$$l_{\text{eff}} = l + \frac{2L}{\pi}\int_{z_0}^{z'} [\beta_e(z) - \beta_o(z)]\, dz \tag{5.88}$$

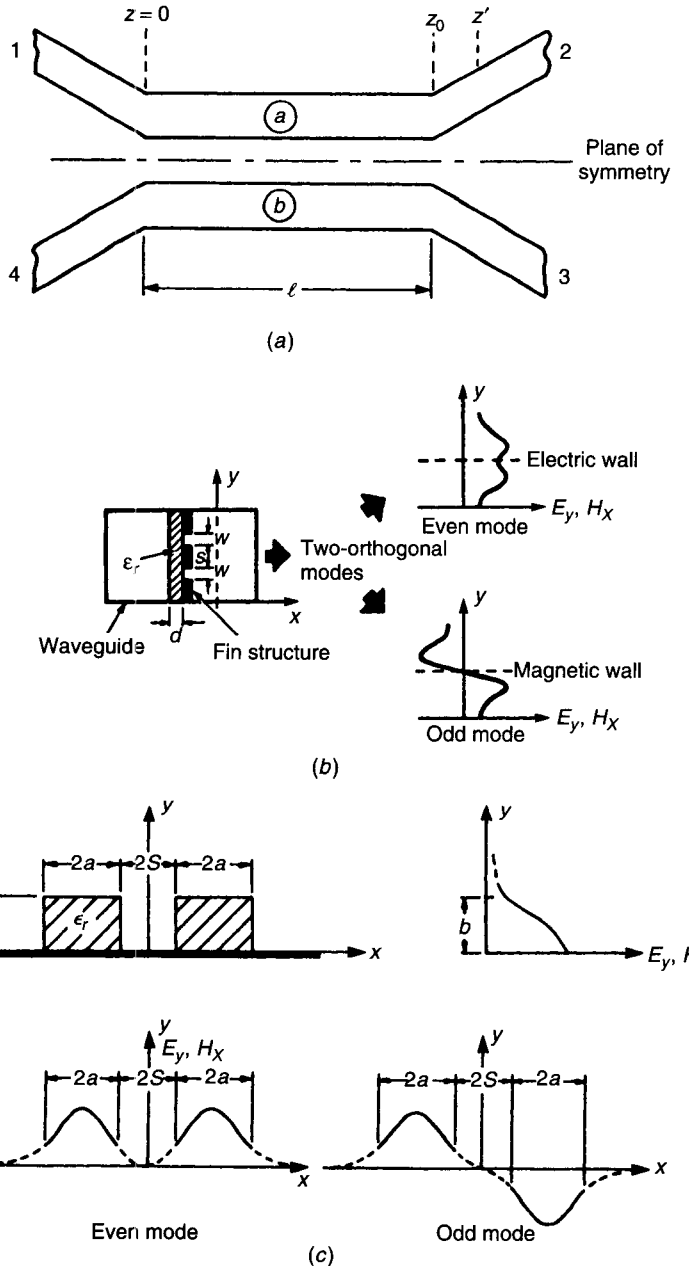


Figure 5.32 (*a*) Schematic of distributed coupler. (*b*) Coupled fin lines and the field distributions. (*c*) Coupled image lines and the field distributions.

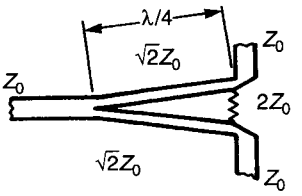


Figure 5.33 Single-section Wilkinson coupler.

The integration limit z' is chosen to be the point at which the coupling is practically negligible.

From the preceding discussions it appears that the design of any distributed coupler requires a precise knowledge of the even- and the odd-mode phase constants of the coupled lines. Figures 5.32*b* and *c* show the cross sections of a coupled fin lines and image lines, respectively, as examples of two commonly used distributed couplers.

5.3.5 Wilkinson Couplers, Power Dividers, and Combiners

A Wilkinson coupler [27, 28] offers broad bandwidth and equal phase characteristics at each of its output ports. Figure 5.33 shows the schematic diagram of a Wilkinson coupler. The output port isolation is obtained by series terminating the output port. Each of the quarter-wave lines has the characteristic impedance of $\sqrt{2}Z_0$ and the output is terminated by a resistor of $2Z_0$ ohms, Z_0 being the system impedance.

A Wilkinson power divider offers a bandwidth of about one octave. The typical frequency response is shown in Fig. 5.34. An adequately flat response is

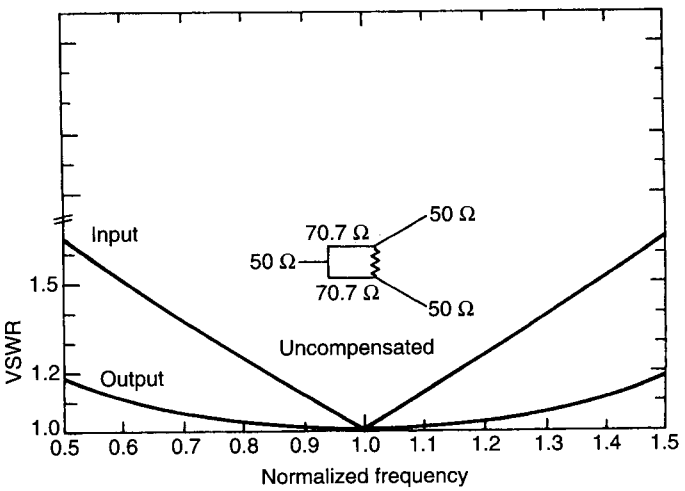


Figure 5.34 Frequency response of Wilkinson coupler. (After Howe [5]. Reprinted with permission of Artech House.)

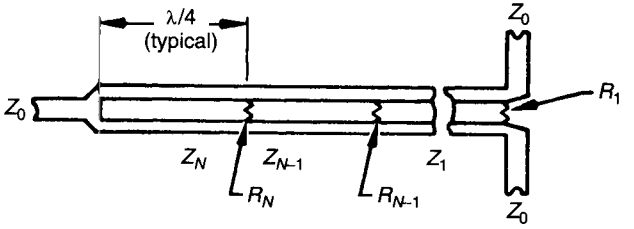


Figure 5.35 Multisection Wilkinson coupler.

obtained over more than one octave band. But at the band edges the isolation is affected by the load impedance.

The performance of a Wilkinson coupler can be further improved, depending upon the availability of space, by the addition of a $\frac{1}{4}\lambda$ transformer in front of the power division step. The input VSWR of the uncompensated coupler is better than the input VSWR of the compensated circuit.

Multisection Wilkinson Coupler. The octave bandwidth of a single-section coupler proves to be inadequate in many applications. Therefore Cohn [28] proposed the use of multisections for bandwidth expansion. The use of multisections makes it possible to obtain a decade bandwidth.

A multisection Wilkinson power-divider coupler consists of a number of quarter-wave sections with resistive terminations at the end of every section, as shown schematically in Fig. 5.35. Larger bandwidth and greater isolation are obtained when a larger number of sections are used.

The characteristic impedances of the sections are obtained from the normalized impedances for $\frac{1}{4}\lambda$ transformer sections for a 2:1 transformer. This can be done with the help of Fig. 5.36, which presents design curves up to four sections. Similar figures are also available for higher numbers of sections. Having obtained the impedance of each section, values of the terminating resistors for each section can be obtained.

For a two-section divider the values of the terminating resistors are given by

$$R_2 = \frac{2Z_1Z_2}{[(Z_1 + Z_2)(Z_2 - Z_1 \cot^2 \phi)]^{1/2}} \tag{5.89a}$$

$$R_1 = \frac{2R_2(Z_1 + Z_2)}{R_2(Z_1 + Z_2) - 2Z_2} \tag{5.89b}$$

where

$$\phi = \frac{\pi}{2}\left[1 - 0.707\left(\frac{f_2 - f_1}{f_2 + f_1}\right)\right] \tag{5.89c}$$

and f_1, f_2 are the upper and lower band edge frequencies of operation.

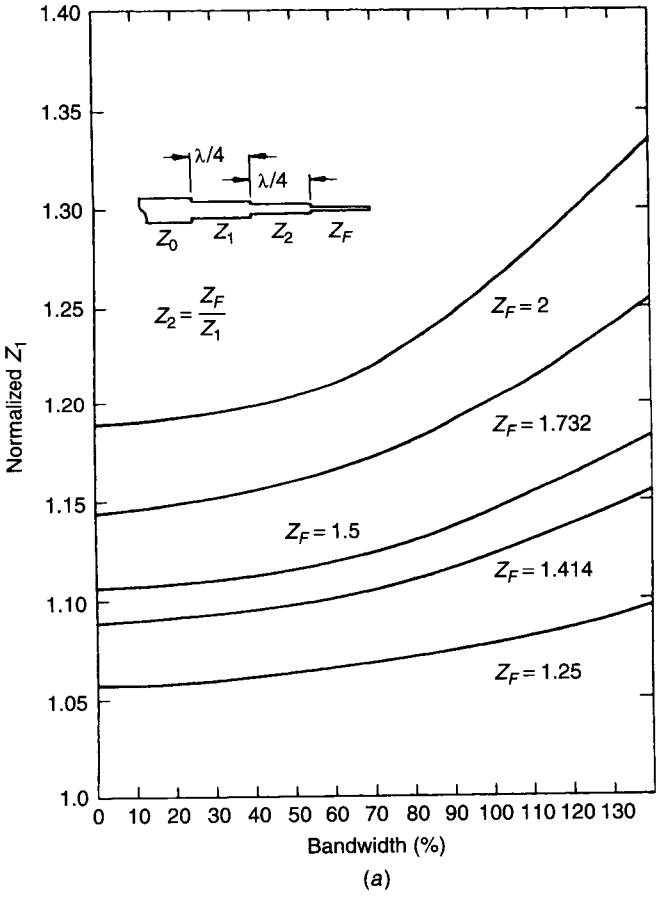


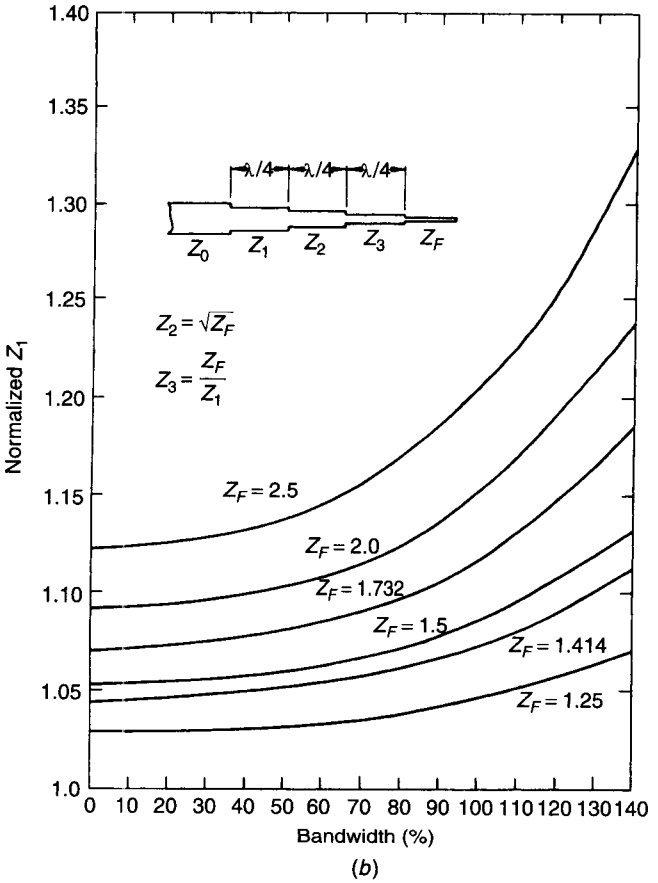
Figure 5.36 Design curves for impedance transformer: (*a*) two section, (*b*) three section, and (*c*) four section. (After Howe [5]. Reprinted with permission of Artech House.)

Although the design of a two-section coupler is straightforward, that of more than two sections is cumbersome. The characteristic admittance of each section is chosen as it is in the case of a two-section coupler. The output resistor is obtained from

$$\frac{1}{R_1} = 1 - Y_1 \tag{5.90a}$$

The intermediate resistor is obtained from

$$\frac{1}{R_K} = G_K = \frac{Y_{K-1} - Y_K}{Y_{K-1} T_1 T_2 \ldots T_{K-1}} \quad \text{for } K = 2, \ldots, N-1 \tag{5.90b}$$



(*b*)

Figure 5.36 *(Continued)*

$$T_K = \frac{4Y_{K-1}Y_K}{(Y_{K-1} + Y_K + 2G_K)^2} \qquad \text{for } K = 1, \ldots, N \tag{5.90c}$$

$$\frac{1}{R_n} = G_n$$

$$= \cfrac{0.5Y_{n-1}^2}{-2G_{n-1} + \cfrac{Y_{n-2}^2}{-2G_{n-2} + \cfrac{Y_{n-3}^2}{\ddots \cfrac{}{-2G_2 + \cfrac{Y_1^2}{-2G_1 + 1 + 0.7(S_{e90} - 1)}}}}} \tag{5.90d}$$

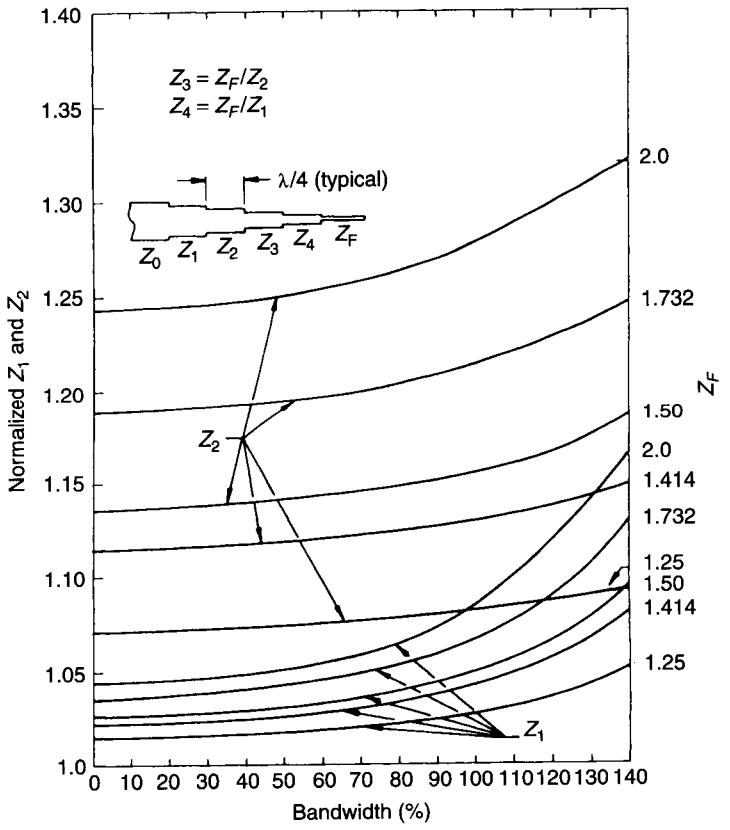


Figure 5.36 *(Continued)*

where

$$S_{e90}\begin{cases} = 1 & \text{for } N \text{ odd} \\ = S_{em} & \text{for } N \text{ even} \end{cases}$$

Use of (5.90d) requires the knowledge of the maximum ripple VSWR for the prototype transformer sections chosen for the design. These data are obtained from Fig. 5.37.

The minimum isolation in decibels is computed from

$$I \cong 20 \log\left(\frac{2.35}{S_{em} - 1}\right) \tag{5.91}$$

where S_{em} is the maximum VSWR ripple.

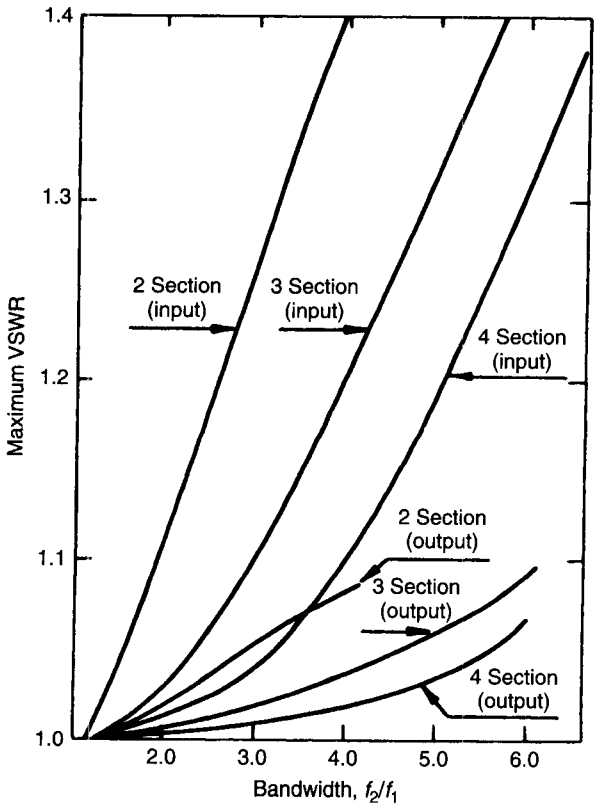


Figure 5.37 Maximum input–output VSWR vs. bandwidth for a multisection Wilkinson coupler. (After Howe [5]. Reprinted with permission of Artech House.)

Theoretically speaking, the performance of a multisection Wilkinson power-divider or combiner is the same as that of a single section except for the equi-ripple VSWR characteristics. In reality, much depends upon the fabrication and the tolerance in the values of the terminating resistors. Smaller degradation in performance is observed over the L through the S band. However, in the X or Ka band the degradation may be severe, lowering the minimum isolation by even 40%. Table 5.3 gives the values of R's and Z's for $N = 3, 4, 7$ [28].

Unequal Split Wilkinson Power Dividers. Figure 5.38 shows the schematic of the unequal split Wilkinson divider [29]. As can be seen from the figure, output impedance transformers are also required, in contrast to the equal power split case.

The design equations for the compensated case are as follows:

Table 5.3 Parameters of Power Divider and Common Design Parameters for Both Wilkinson and Improved Version Power Dividers

Number	R	Z	Width (mm)
Number of Expected Section = 3, *Epsilon* = 2.150, *Thickness of Sub* = 1.600 mm			
1	400.000	57.485	2.506
2	211.460	70.710	1.729
3	107.180	86.980	1.096
Expected Section = 4, *Epsilon* = 2.150, *Thickness of Sub* = 1.600 mm			
1	482.160	55.785	2.547
2	291.630	64.785	1.973
3	172.620	77.175	1.401
4	103.165	89.630	0.985
Expected Section = 7, *Epsilon* = 2.150, *Thickness of Sub* = 1.600 mm			
1	442.480	56.370	2.415
2	616.145	60.255	2.161
3	446.230	65.085	1.887
4	319.900	70.710	1.615
5	217.580	76.820	1.365
6	129.620	82.985	1.150
7	248.260	88.700	0.978

$$
\begin{aligned}
K^2 &= \frac{P_b}{P_a} \\
Z_1 &= Z_0\left(\frac{K}{1+K^2}\right)^{1/4} \qquad Z_2 = Z_0[K^{3/4}(1+K^2)^{1/4}] \\
Z_3 &= Z_0\left(\frac{(1+K^2)^{1/4}}{K^{5/4}}\right) \qquad Z_4 = Z_0\sqrt{K} \qquad Z_5 = \frac{Z_0}{\sqrt{K}} \\
R &= Z_0\left(\frac{1+K^2}{K}\right)
\end{aligned}
\tag{5.92}
$$

while those for the uncompensated case are available elsewhere [30].

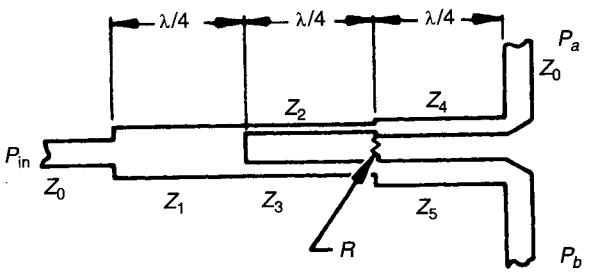


Figure 5.38 Schematic of unequal power split Wilkinson coupler.

Example. If one wishes to design a split-T power combiner, with one output port having three times the power of the other, and if the input–output line impedances are 50 Ω, then $K^2 = 3$, and substituting into the preceding equations,

$$Z_1 = 40.56 \qquad Z_2 = 106.77 \qquad Z_3 = 35.59$$
$$Z_4 = 66.0 \qquad Z_5 = 37.88 \qquad R = 115.47$$

5.3.6 Other Couplers

Tandem Coupler. The design of a multisection coupler with tight coupling over a broad bandwidth requires some of its sections to have tighter coupling than the overall coupling. This invariably leads to physically unrealizable spacings between the two conductors or severely reduced directivity due to significant mechanical discontinuities in the sections. To solve this problem, in a restricted physical space, various combinations of symmetric and asymmetric couplers are tandemed [5, 20]. Since in the majority of applications, the tightest coupling may be 3 dB, two couplers may be connected in tandem to achieve the goal. Figure 5.39*a* shows the symmetric tandem of two 8.34-dB couplers, while Fig. 5.39*b* shows the symmetric tandem of two asymmetric couplers. Each of the couplers of 8.34 dB when tandemed gives an overall coupling of 3 dB. This configuration offers high power-handling capability and often represents a good choice, provided the particular application does not require maximum bandwidth with very low loss.

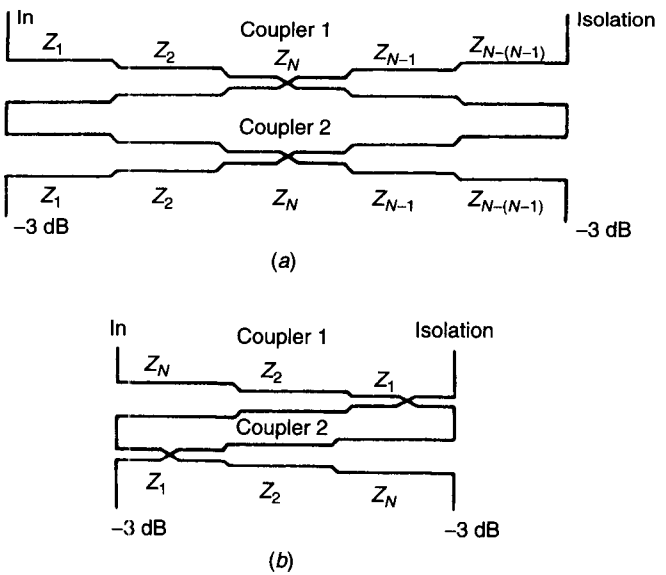


Figure 5.39 Tandem coupler: (*a*) Symmetric tandem of symmetric couplers; (*b*) symmetric tandem of asymmetric couplers.

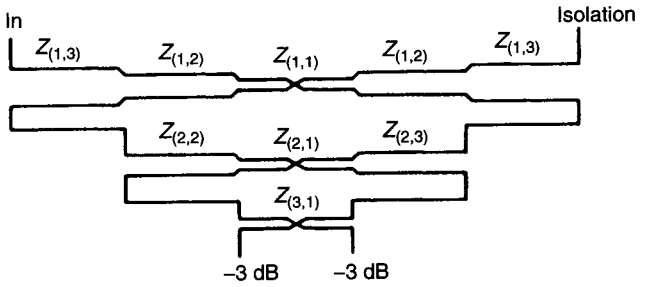


Figure 5.40 Asymmetric tandem of symmetric coupler.

Figure 5.40 shows the configuration of an asymmetric tandem of symmetric couplers. As in the symmetric tandem of symmetric couplers, the quadrature phase shift is maintained between the outputs.

As is apparent from the interconnections, the 90° phase relationship is maintained in the symmetric case, but in the asymmetric case, the phase relation depends upon the number of sections.

De Ronde Coupler. The De Ronde coupler is suitable for realization of 3-dB hybrids in MICs. The structure, shown in Fig. 5.41, was first proposed by De Ronde [31] and has been named after him. Figure 5.41 shows that the hybrid consists of a coupled section of a microstrip and a slot line connected by four microstrip lines of width W_p. The effective length l of the coupled section is defined by the reference planes T_1 and T_2. The slot line of length $l + 2l_s$ is terminated on either side by a disc-shaped slot, D, forming an open circuit. The effects of the parasitics due to the field perturbation occurring at the junctions at each end are nullified by proper choice of the distance d between the two reference planes T_1 and T_2.

As can be seen in Fig. 5.41, the hybrid has a double symmetry with respect to planes P_1 and P_2, as shown in Fig. 5.42, and it is a reciprocal four-port network. The scattering parameters of the network can be written in terms of the even- and the odd-mode reflection coefficients as in [32]

$$S_{11} = \tfrac{1}{4}(\Gamma_{em} + \Gamma_{ee} + \Gamma_{om} + \Gamma_{oe}) \tag{5.93a}$$

$$S_{21} = \tfrac{1}{4}(\Gamma_{em} - \Gamma_{ee} + \Gamma_{om} - \Gamma_{oe}) \tag{5.93b}$$

$$S_{31} = \tfrac{1}{4}(\Gamma_{em} + \Gamma_{ee} - \Gamma_{om} - \Gamma_{oe}) \tag{5.93c}$$

$$S_{41} = \tfrac{1}{4}(\Gamma_{em} - \Gamma_{ee} - \Gamma_{om} + \Gamma_{oe}) \tag{5.93d}$$

The reflection coefficients $\Gamma_{em}, \Gamma_{ee}, \Gamma_{om}$, and Γ_{oe} are referenced to Z_0 (the characteristic impedance of the connecting section) appearing at each of the ports on application of certain combinations of perfect electric (pec) and perfect magnetic (pmc) planes at the planes of symmetry P_1 and P_2.

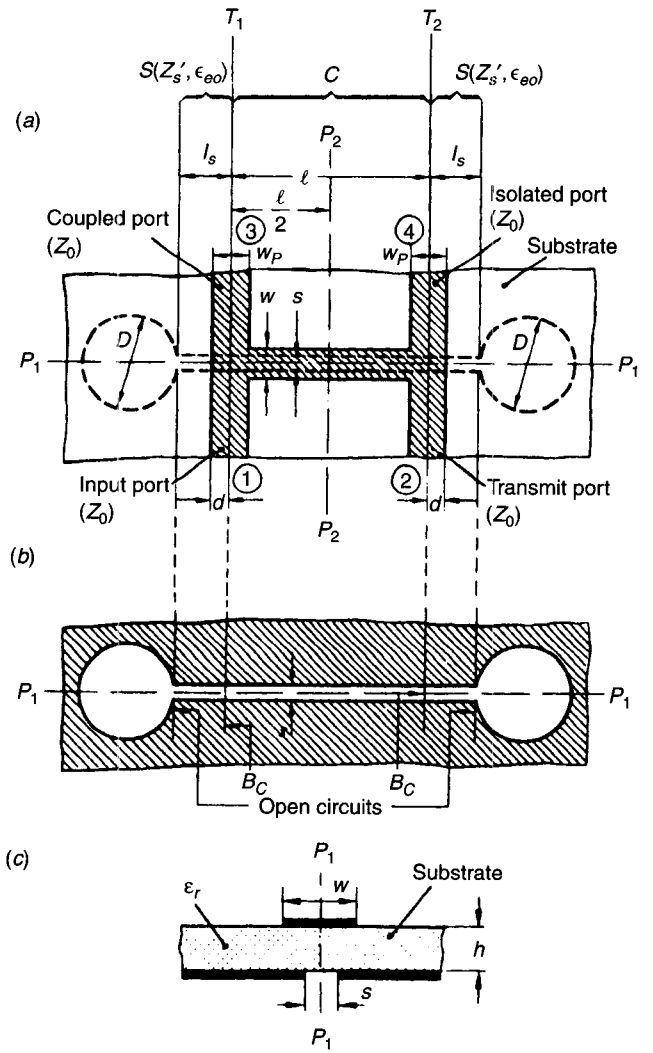


Figure 5.41 De Ronde rnicrostrip slot coupler: (*a*) Upperside of substrate, (*b*) bottom side of substrate, and (*c*) cross section.

A pmc at P_1 (Fig. 5.43*a*) corresponds to an even-mode excitation with terminal voltages of equal amplitude and the same phase. The corresponding quasi-microstrip field pattern is shown in Fig. 5.43*b*. This pattern divides the network shown in Fig. 5.43*b* into two identical transmission lines of characteristic impedance Z_e

$$Z_e = 2Z_M \qquad \epsilon_{re} = \epsilon_{ee} \tag{5.94}$$

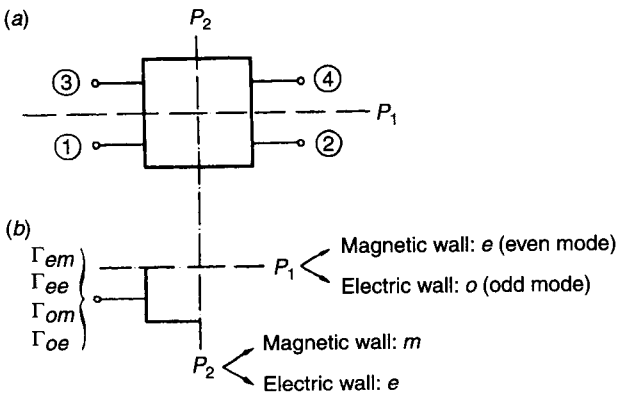


Figure 5.42 Common four-port network with double symmetry.

where Z_M is the characteristic impedance of the microstrip and $\epsilon_{ee}, \epsilon_{re}$ are the effective dielectric constants of the even mode of the coupled lines and the microstrip, respectively. The corresponding relation for the odd mode (pec at P_1) are given by (Fig. 5.43*c*)

$$Z_0 = \tfrac{1}{2}Z_s \qquad \epsilon_{ro} = \epsilon_{eo} \tag{5.95}$$

where Z_s and ϵ_{eo} are the characteristic impedance and the effective dielectric constant of the quasi-slot-mode due to the odd-mode excitation.

In terms of the preceding parameters, the reflection coefficients are given by

$$\Gamma_{em} = \exp\left\{-j2 \arctan\left[\frac{Z_0}{2Z_M} \tan\left(\frac{\theta_e}{2}\right)\right]\right\} \tag{5.96a}$$

$$\Gamma_{ee} = \exp\left\{j2 \arctan\left[\frac{Z_0}{2Z_M} \cot\left(\frac{\theta_o}{2}\right)\right]\right\} \tag{5.96b}$$

$$\Gamma_{om} = \exp\left\{-j2 \arctan\left[\frac{2Z_0}{Z_s} \tan\left(\frac{\theta_e}{2}\right) + 2B_c Z_0\right]\right\} \tag{5.96c}$$

$$\Gamma_{oe} = \exp\left\{-j2 \arctan\left[-\frac{2Z_0}{Z_s} \cot\left(\frac{\theta_o}{2}\right) + 2B_c Z_0\right]\right\} \tag{5.96d}$$

where

$$\theta_e = \beta_e l \tag{5.97a}$$

$$\theta_o = \beta_o l \tag{5.97b}$$

and $\beta_e = (2\pi/\lambda_0)\sqrt{\epsilon_{ee}}$ and $\beta_o = (2\pi/\lambda_0)\sqrt{\epsilon_{eo}}$ are the even- and odd-mode propagation constants, respectively.

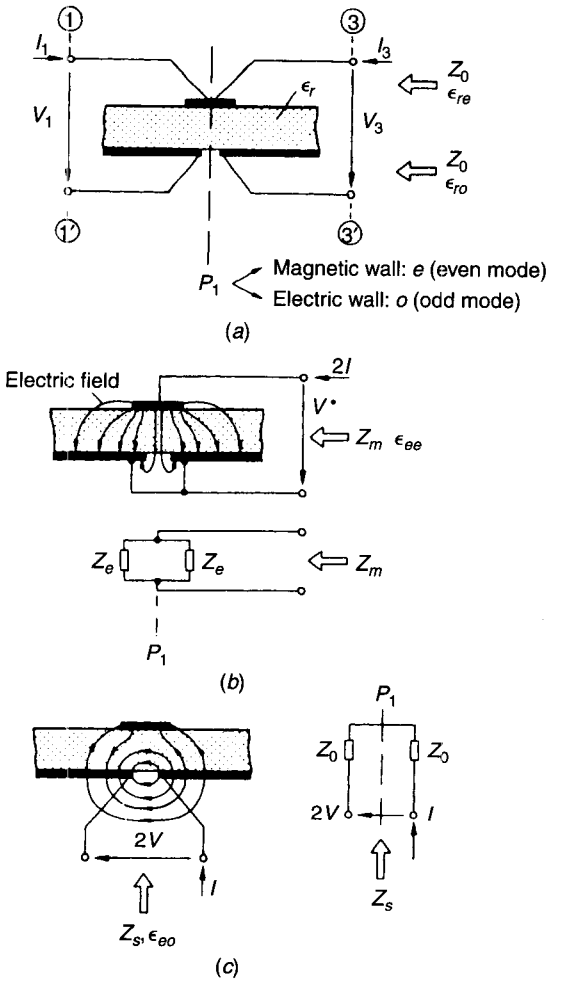


Figure 5.43 (*a*) Even-mode excitation of De Ronde coupler. (*b*) Quasi-microstrip field pattern corresponding to even-mode excitation. (*c*) Odd-mode excitation.

The term $2B_cZ_0$ accounts for the slot compensation length l_s (see Fig. 5.44):

$$B_c = \frac{1}{Z_s'} \tan \theta_s' = \omega C_c \tag{5.98a}$$

$$\theta_s' = \beta_s l_s' \qquad \beta_s = \frac{2\pi}{\lambda_0}\sqrt{\epsilon_{eo}'} \tag{5.98b}$$

where Z_s' and ϵ_{eo}' are the characteristic impedance and the effective dielectric constant, respectively, of the slot s. The compensation length l_s does not affect the even-mode excitation. The capacitance C_c effectively increases the odd-

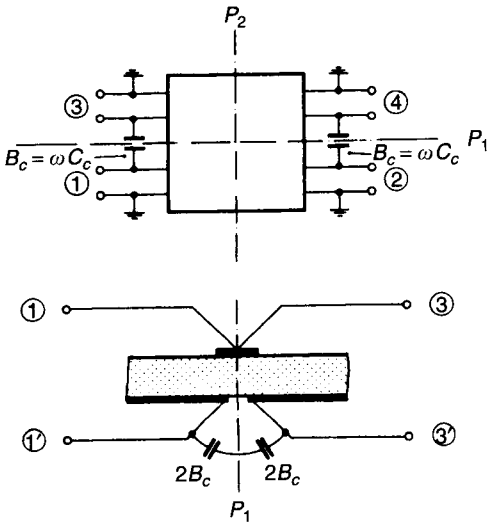


Figure 5.44 Equivalent circuit of compensated De Ronde coupler.

mode transmission phase between terminals 1 and 3, and 2 and 4. Therefore the compensation in terms of C_c leads to equal phase velocities and consequently infinite directivity.

Special Cases of De Ronde Coupler. *Ideal TEM.* Although the modes supported by the slot line and the microstrip are truly non-TEM in nature, one can think of an ideal coupler where either supports a purely TEM mode. Such a coupler is physically impossible to realize. However, it can be analyzed to derive the design equations for an ideal coupler.

For an ideal coupler that is matched, $S_{11} = S_{41} = 0$ over the entire useful band of frequencies. This gives, from (5.93),

$$\Gamma_{ee} = -\Gamma_{om} \qquad \Gamma_{em} = -\Gamma_{oo}$$

which requires equal, even-, and odd-mode effective dielectric constants, or

$$\epsilon_{re} = \epsilon_{ro} = \epsilon_e \qquad \epsilon_{ee} = \epsilon_{eo} = \epsilon_e$$

and

$$Z_0 = \sqrt{Z_e Z_o} = \sqrt{Z_M Z_s} \tag{5.99}$$

With the preceding conditions, the coupling and transmission coefficients are given by

$$S_{21}(f) = \frac{(1 - C^2)^{1/2}}{(1 - C^2)^{1/2} \cos\theta + j \sin\theta} \tag{5.100a}$$

$$S_{31}(f) = \frac{jC \sin\theta}{(1 - C^2)^{1/2} \cos\theta + j \sin\theta} \tag{5.100b}$$

where

$$\theta = \frac{2\pi}{\lambda_0} \sqrt{\epsilon_e} l$$

At the center frequency $\theta = \frac{1}{2}\pi$. These are the parameters of an ideal TEM coupler as shown in Section 5.3.2. Since

$$C = \frac{Z_{0e} - Z_{0o}}{Z_{0e} + Z_{0o}} \tag{5.101}$$

which gives

$$C = \frac{4Z_M - Z_s}{4Z_M + Z_s} \tag{5.102}$$

Combining (5.99) and (5.102) gives

$$Z_M = \frac{Z_0}{2} \sqrt{\frac{1 + C}{1 - C}} \tag{5.103a}$$

$$Z_s = 2Z_0 \sqrt{\frac{1 - C}{1 + C}} \tag{5.103b}$$

For a 3-dB coupler in a 50-Ω system, one obtains, using the previous equations,

$$Z_M = 60.35\ \Omega \qquad Z_s = 41.4\ \Omega$$

Real Uncompensated Coupler. In a real uncompensated coupler $l_s = 0$. For all couplers of conventional dimensions, $\epsilon_{eo} \leq \epsilon_{ee}$. This gives the odd-mode phase velocity greater than the even-mode phase velocity. However, the following condition is always satisfied

$$2\left(\frac{\epsilon_{ee} - \epsilon_{eo}}{\epsilon_{ee} + \epsilon_{eo}}\right) \leq 1 \tag{5.104}$$

Equation (5.104) implies that for feeding at port 1, the direct port 2, the coupled port 3, and the isolated port 4 remain unchanged.

In the preceding case certainly the directivity $\log D \leq \infty$ and $S_{11} \geq 0$. However, if the synthesis equation (5.103) is used for designing the coupler, the following conditions hold:

$$S_{11}, S_{41} \ll 1$$

and a first-order approximation yields

$$S_{11}(f_0) = -j\frac{\pi}{4}\left(\frac{\epsilon_{ee} - \epsilon_{eo}}{\epsilon_{ee} + \epsilon_{eo}}\right) C\sqrt{1 - C^2} \tag{5.105a}$$

$$D(f_0) = -20 \log\left|\frac{\pi}{4}\left(\frac{\epsilon_{ee} - \epsilon_{eo}}{\epsilon_{ee} + \epsilon_{eo}}\right)\frac{1 - C^2}{C}\right\}\Bigg| \tag{5.105b}$$

where f_0 is the center frequency.

Real Compensated Coupler. For compensation, the slot length is increased by an amount l_s and Z_s is changed to Z_s^*. Under the assumption that the compensation is realized at the center frequency referenced to the microstrip line and $Z_s^* \approx Z_s$, $\epsilon_{eo}^* \approx \epsilon_{eo}$, the required compensation parameters are given by

$$Z_s^* = \frac{Z_s}{2}\left[\cot\left(\frac{\pi}{4}\sqrt{\frac{\epsilon_{eo}}{\epsilon_{ee}}}\right) + \tan\left(\frac{\pi}{4}\sqrt{\frac{\epsilon_{eo}}{\epsilon_{ee}}}\right)\right] \tag{5.106a}$$

$$l_s = l\sqrt{\frac{\epsilon_{ee}}{\epsilon_{eo}}}\frac{2}{\pi}\tan^{-1}\left\{\frac{Z_s^*}{Z_s}\left[1 - \sin^2\left(\frac{\pi}{4}\sqrt{\frac{\epsilon_{eo}}{\epsilon_{ee}}}\right)\right]\right\} \tag{5.106b}$$

Computation of Even- and Odd-Mode Parameters. It is obvious that the coupled section of the device is not purely microstrip nor purely slot line either. An exact analysis of the structure therefore requires the help of numerical techniques. Such techniques are often difficult to carry out. On the other hand, accurate models for analysis of microstrip and slot line are available [6]. These models can be used to approximately realize the design parameters, and the approximations have been found to be valid for all practical purposes as long as the coupling coefficient is less than 6 dB over the 2- to 18-GHz range. The typical performance of an uncompensated coupler is shown in Figs. 5.45*a* and *b*, where a_{21} and a_{31} are the transmission loss and coupling loss, respectively.

5.4 DESIGN CONSIDERATIONS

5.4.1 Losses in Hybrids

The total loss in direct-coupled hybrids can be estimated from the combined dielectric and conductor losses in the individual lines. These are obtained using closed-form equations for microstrip- and strip-line-type transmissions lines

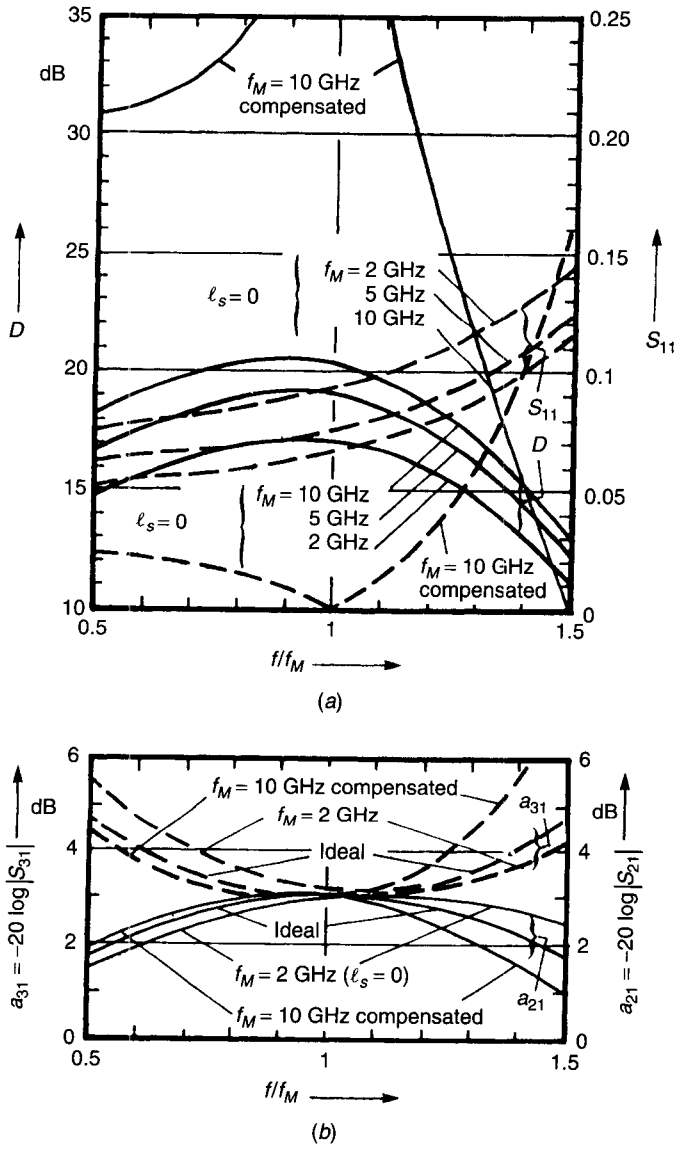


Figure 5.45 Performance of uncompensated coupler. (After Hoffmann and Siegel [32]. Reprinted with permission of IEEE.)

[26]. For other types of transmission lines, numerical techniques are used [33]. The loss in coupled lines was dealt with in Chapter 2.

The attenuation due to the even mode α_c^e is always less than that due to the odd mode α_c^o. In coupled lines the loss is given by the average of the losses due to even and odd modes. In almost all planar coupled lines the conductor loss greatly exceeds the dielectric loss [26].

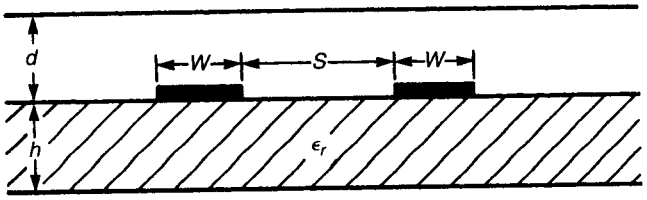


Figure 5.46 Parallel coupler microstrip with grounded shield.

As in strip lines and microstrip lines, the primary contributors to losses in dielectric-based planar waveguide complers are the dielectric loss and the metallic loss. Such losses are always computed numerically from the associated field equations [34].

5.4.2 Directivity Improvement

Because of the inhomogeneity in dielectric structure the directivity of microstrip couplers offers a poor bandwidth resulting from different odd- and even-mode phase velocities. There are several ways to equalize the phase velocities, which are as follows.

Use of a Shield. These structures nearly equalize even- and odd-mode phase velocities. The cross section of the structure is shown in Fig. 5.46. This configuration essentially redistributes the field with a substantial amount of the field in the air dielectric medium above the coupled strips. For $d = h$ the phase velocities are exactly equal and each is equal to

$$v_e = v_o = \frac{c}{\sqrt{(\epsilon_r + 1)/2}} \tag{5.107}$$

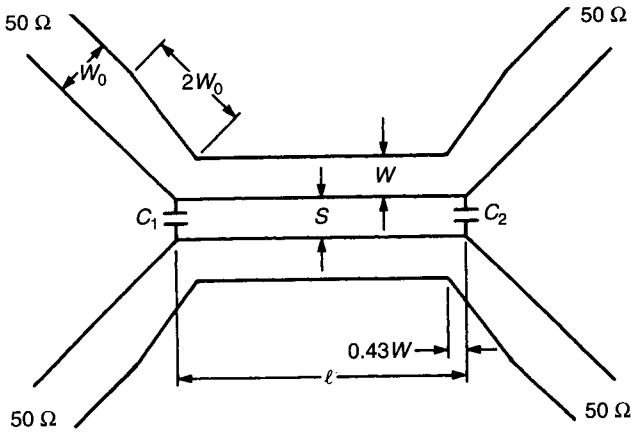


Figure 5.47 Lumped-capacitor compensation of microstrip coupler.

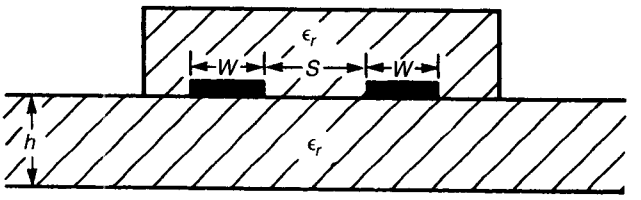


Figure 5.48 Parallel coupled microstrip with overlay compensation.

Strictly speaking improvement in directivity is obtained at the cost of manufacturing difficulties in this case.

Use of Lumped Capacitors. Using lumped capacitors at the ends of the coupled section is the simplest way of achieving equal phase velocities. The structure is shown in Fig. 5.47. This effectively increases the odd-mode phase angle by [14]

$$\Delta\theta_o = 2\pi f_0(C_1 + C_2)Z_{0o} \tag{5.108}$$

where f_0 is the center frequency of the coupled line.

Use of a Dielectric Overlay. The presence of another dielectric layer of the same permittivity as that of the substrate reduces the odd-mode phase velocity to a large extent without considerably affecting the even-mode phase velocity [35]. Thus controlling the thickness and width of the overlay, the even- and the odd-mode phase velocities can be equalized within 1% over quite a broad band. As shown in Fig. 5.48, the overlay covers the two strips where the main coupling takes place. The overlay is bonded with the help of some kind of epoxy. However, this method is practically cumbersome due to its poor repeatability. To overcome this problem a multilayer structure using different dielectric constants materials has been recently proposed [35].

REFERENCES

1. Agrawal, A. K., and G. F. Mikucki, "A Printed Circuit Hybrid Ring Directional Coupler for Arbitrary Power Divisions," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-34, Dec. 1986, pp. 1401–1407.
2. Levy, R., "Zolotarev Branch-Guide Couplers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-21, Feb. 1973, pp. 95–99.
3. Levy, R., and J. Helszajn, "Specific Equation for One or Two Section Quarter-Wave Matching Networks for Stub Resistor Loads," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Jan. 1982, pp. 55–62.
4. Muraguchi, O., M. Y. Takeshi, and Y. Naito, "Optimum Design of 3-dB Branch Line Couplers Using Microstrip Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-31, Aug. 1983, pp. 674–678.

5. Howe, H., *Stripline Circuit Design*, Artech House, Dedham, MA, 1974.
6. Gupta, K. C., R. Garg, and I. J. Bahl, *Microstrip Lines and Slotlines*, Artech House, Dedham, MA, 1979.
7. Pramanick, P., and P. Bhartia, "Analysis and Synthesis Equations for Suspended and Inverted Microstriplines," *Arch. Elek. Übertragung.*, Vol. 39, Sept. 1985, pp. 323–326.
8. Pramanick, P., and P. Bhartia, "Computer Aided Design Models for Millimeter Wave Finlines and Suspended Substrate Microstrip Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-33, Dec. 1985, pp. 1429–1435.
9. Bhartia, P., and I. J. Bahl, *Millimeter Wave Engineering and Applications*, Wiley, New York, 1984, pp. 300–346.
10. Koul, S. K., Center for Advanced Research in Electronics, Indian Institute of Technology, New Delhi, India, private communication, 1985.
11. Rycroft, C. M. D., "Finline Rat Race Hybrid Couplers," Interim Report ITM 84, 84/159, Chemsford, Essex, U.K., 1984.
12. Hirota, T., A. Minakawa, and M. Muraguchi, "Reduced-Size Branch-Line and Rat-Race Hybrids for Uniplanar MMICs," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-38, Mar. 1990, pp. 270–275.
13. Parisi, S. J., "A Lumped Element Rat-Race Coupler," *Applied Microwaves*, Vol. 1, Aug/Sept., 1989, pp. 84–93.
14. Bhartia, P., and I. J. Bahl, *Millimeter Wave Engineering and Applications*, Wiley, New York, 1984, pp. 358–382.
15. Pramanick, P., and P. Bhartia, "Design of Imageguide Microstrip Line Couplers," *Microwave J.*, Vol. 27, May 1984, pp. 247–255.
16. Bahl, I. J., and P. Bhartia, "Aperture Coupling between Dielectric Image Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-29, Sept. 1981, pp. 591–596.
17. Pramanick, P., and P. Bhartia, "Waveguide Imageline Directional Coupler," *Arch. Elek. Übertragung.*, Vol. 38, Jan. 1984, pp. 46–50.
18. Edwards, T. C., *Foundations for Microstrip Circuit Design*, Wiley, New York, 1981, pp. 147–148.
19. Matthaei, G. L., L. Young, and E. M. T. Jones, *Microwave Filters Impedance Matching Networks and Coupling Structures* (Reprinted), Artech House, Dedham, MA, 1980.
20. Davis, W. A., *Microwave Semiconductor Circuit Design*, Van Nostrand, New York, 1983.
21. Hoffmann, R. K., and G. Morler, "Directional Coupler Synthesis—Computer Program Discription," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, Jan. 1974, p. 77.
22. Klopfenstein, R. W., "A Transmission Line Taper of Improved Design," *Proc. IRE*, Vol. 44, Jan. 1956, pp. 31–35.
23. Lange, J., "Interdigited Stripline Quadrature Hybrid," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-17, Dec. 1969, pp. 1150–1151.
24. Kajfez, D., Z. Paunovic, and S. Paulin, "Simplified Design of Lange Coupler," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-26, Oct. 1978, pp. 806–808.
25. Presser, A., "Interdigited Microstrip Coupler Design," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-26, Oct. 1978, pp. 801–805.

26. Garg, R., and I. J. Bahl, "Characteristics of Coupled Microstrip," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, July 1979, pp. 700–705.
27. Wilkinson, E., "An N-Way Hybrid Power Divider," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-8, Jan. 1960, pp. 116–118.
28. Cohn, S. B., "A Class of Broadband 3-Port TEM Hybrids," *IEEE Trans. Microwave Theory Tech.*, Vol. MIT-16, Feb. 1968, pp. 110–118.
29. Howe, H., *Stripline Circuit Design*, Artech House, Dedham, MA, 1974, pp. 97–98.
30. Howe, H., *Stripline Circuit Design*, Artech House, Dedham, MA, 1974, pp. 164–172.
31. De Ronde, F. C., "A New Class of Microstrip Directional Couplers," *IEEE Int. Microwave Symp. Dig.*, 1974, pp. 184–186.
32. Hoffmann, R. K., and J. Siegel, "Microstrip-Slot Coupler Design—Parts I & II," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Aug. 1982, pp. 1205–1216.
33. Trinh, T. N., and R. Mittra, "Suspended H-waveguide and its mm-Wave Applications," *IEEE Int. Microwave Symp. Digest*, 1983, pp. 305–308.
34. Shelad, B., and B. E. Spielman, "Broadband (7–18 GHz) 10-dB Overlay Coupler for MIC Applications," *Electron Lett.*, Vol. 11, Apr. 1975, pp. 175–176.
35. Horno, M., and F. Medina, "Multilayer Planar Structures for High-Directivity Directional Coupler Design," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-34, Dec. 1986, pp. 1442–1449.

PROBLEMS

5.1 Determine the aperture radii of five-section Chebyshev rectangular waveguide to image-line and image-line to rectangular waveguide directional couplers for 20 dB coupling, with minimum directivity of 40 dB, a bandwidth of 14%, and operational frequency of 35 GHz. Choose suitable guide dimensions for the fundamental mode of operation. Assume that the image line has a dielectric constant of 2.54.

5.2 Determine the aperture radii of five-section Chebyshev image-line–microstrip and microstrip–image-line couplers for a required coupling of 10 dB, directivity of 50 dB, bandwidth of 16%, ϵ_r of 2.54, and center frequency 35 GHz.

5.3 Design a single-section microstrip directional coupler with the following specifications: Coupling 10 dB, substrate $\epsilon_r = 9.0$, substrate thickness 0.635 mm, system center frequency 4.0 GHz, impedance 50 Ω. Neglect dispersion.

5.4 **(a)** Find the impedances of a hybrid ring directional coupler for the following power split ratios:
1. 0 dB
2. 3 dB
3. 9 dB

(b) For a strip-line-type coupler, find the corresponding strip width dimensions for a substrate $\epsilon_r = 3.8$ and ground plane spacing of 2.5 mm.

5.5 Design an image-line edge-coupled directional coupler with 0 dB coupling at 33 GHz, return loss $\leq$25 dB, and directivity $\leq$30 dB over the 30- to 40-GHz frequency band. Assume $\epsilon_r = 2.22$.

5.6 A TEM-mode asymmetric directional coupler has $C_\infty = 3.01$ dB. Draw the even- and odd-mode impedance profile as a function of the electrical length of the coupled section, which is required to be of $\frac{1}{2}\lambda$ length. Repeat the plot for $C_\infty = 10$ dB.

5.7 Design an edge-coupled 10-dB microstrip coupler on 0.25-mm alumina substrate for operation over the 30- to 40-GHz band. The required return loss and the directivity are 30 and 45 dB, respectively. Use suitable capacitances for phase velocity compensation.

5.8 Design a 3-dB Lange coupler on a 0.38-mm fused quartz microstrip substrate with isolation and return loss $\geq$25 dB over the 8- to 12-GHz range. Redesign the circuit for operation over 11–15 GHz.

5.9 Design a four-way symmetric power divider in microstrip configuration with the following specifications:

Center frequency = 4 GHz
Power in the outermost arms = 20%
Power in the innermost arms = 30%
Input and output impedance = 50 Ω

The microstrip parameters are $\epsilon_r = 9.9$, $h = 0.63$ mm, and $t = 5$ μm.

5.10 Design a three-branch symmetric branch-line coupler in strip line having 3 dB coupling at 4 GHz. Compare its 3 ± 0.5 dB coupling bandwidth with a two-branch coupler. The strip-line parameters are $b = 1.2$ mm, $t = 0$, $\epsilon_r = 2.32$, and discontinuity effects may be ignored.

6

FILTERS

Edward Griffin and Inder Bahl

6.1 INTRODUCTION

Most microwave systems consist of many active and passive components that are difficult to design and manufacture with precise frequency characteristics. In contrast, microwave passive filters can be designed and manufactured with remarkably predictable performance. As a result, microwave systems are usually designed so that all of the troublesome components are relatively wide in frequency response to minimize their effect on the overall system. Filters are then incorporated to very precisely set the system frequency response. Since the filters are the narrowest bandwidth components in the system, it is usually the filters that limit such system parameters as gain and group delay flatness over frequency.

A passive microwave filter is a circuit component consisting of lumped elements (inductors, capacitors, and resistors) only or distributed elements (waveguide sections or microstrip or fin line or any other medium) or both arranged in a particular configuration (topology), so that desired signal frequencies are allowed to pass with minimum possible attenuation while undesired frequencies are attenuated. In the broadest sense, all microwave components can be considered as filters, since each will exhibit some band-limiting behavior when used in a system. Sometimes filters are also used for impedance matching, as described in Chapter 4.

Immense work has been done on filters, and it is not possible to include every aspect of filters in this book due to its limited scope. However, the intent of this chapter is to touch upon the important topics dealing with design of printed-circuit filters and provide references to facilitate more in-depth study.

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

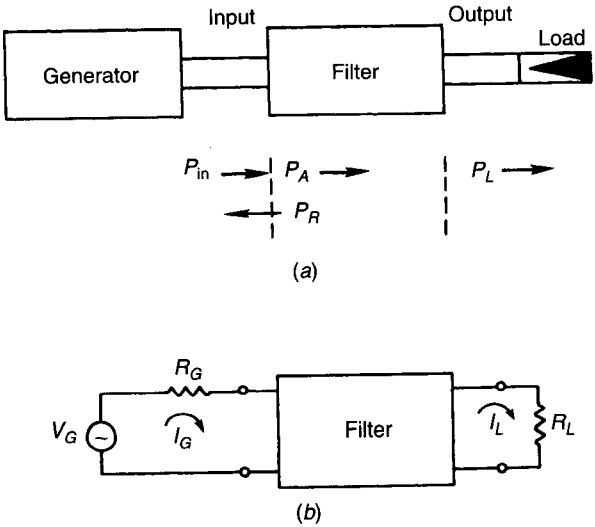


Figure 6.1 (*a*) General form of a filter network. (*b*) Equivalent circuit for power transfer calculations.

The microwave filters that are described in this chapter are two-port reciprocal, doubly terminated, passive, linear, reflective, and lossy. A filter may be viewed conceptually as shown in Fig. 6.1. *Two port* simply refers to the fact that the device has two electrical terminals: The input is driven by a signal generator while the output is connected to a load. A *reciprocal* device is one for which there is no preferred signal propagation direction between ports. This does not mean that a reciprocal device can have input and output interchanged without changing electrical performance, because the impedance levels of the two ports may be different. If the generator and load impedances are identical, a reciprocal device will have identical transmission characteristics when hooked up in either direction, but the reflection characteristics may be different. Practical filters are never precisely identical in reflection characteristics at the input and output due to manufacturing tolerances. Most microwave systems are designed to operate with resistive source and load impedances of 50 Ω. A filter designed to operate with resistive source and load impedances (not necessarily 50 Ω) is called a *doubly terminated* filter.

A *passive* device is one that has no internal energy sources. Incident signal power is reflected, absorbed, or transmitted but not amplified. A *linear* device is one for which all power levels (reflected, absorbed, transmitted) are proportional to the incident power. A linear device generates no new frequency components and no intermodulation products and does not compress the signal. Since there is no interaction between signals, problems with multiple transmission can be treated by superposition. For example, independent signals could be propagating both ways simultaneously through the filter.

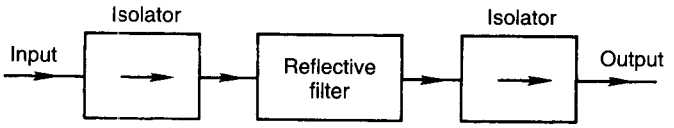


Figure 6.2 Dissipative filter.

No device is truly linear for all input powers, since at some power level, the device will either arc out or melt. Practical filters operating below their maximum rated power levels are generally well treated as linear devices. One exception to this rule occurs in sensitive systems with high power levels where passive low-level (typically 100 dB below the signal level) intermodulation generation can be a problem, particularly at joints and connections between dissimilar metals.

A *reflective* filter is one that achieves rejection by reflecting the incident power. Reflective filters can be thought of as transformers that impedance match the source and load at frequencies where minimum loss is desired and mismatch the source and load at frequencies where rejection is desired. Filters using capacitor and inductor lumped elements and low-loss distributed elements are of the reflective type. Filters that dissipate the rejected signal internally are called *dissipative* or *absorbtive filters*. A dissipative filter can also be realized using a reflective filter and two isolators, as shown in Fig. 6.2. A *lossy* filter is one for which the incident power is greater than the sum of the reflected and transmitted power; some power is always attenuated within the filter. All practical filters are lossy to varying degrees. Filters made using superconductors can be nearly lossless. Recent advances in high-temperature superconductors may cause much greater exploitation of superconducting elements in filters.

6.1.1 Filter Parameter Definition

We now go on to define some of the commonly used terms in the design of filters. When one is designing a filter, the important specifications one looks into are frequency range, bandwidth, insertion loss, stopband attenuation and frequencies, input and output impedance levels, voltage standing-wave ratio (VSWR), group delay, phase linearity, temperature range, and transient response.

As shown in Fig. 6.1*a*, the input is driven by a signal generator with the output passing to a load. At the input plane of the filter, the power may be broken into three components: P_{in}, the incident power from the generator; P_R, the power reflected back toward the generator; and P_A, the power absorbed by the filter. The power passed on to the load is P_L.

Conservation of energy demands that

$$P_{\text{in}} = P_R + P_A \qquad P_L \leq P_A \tag{6.1}$$

where $P_L = P_A$ if the filter is lossless and $P_L = P_{\text{in}}$ if the filter is lossless and there are no reflections.

The incident power is the same as the maximum available power from the generator. Then

$$P_{\text{in}} = \frac{V_G^2}{4R_G} \tag{6.2a}$$

$$P_L = |I_L|^2 R_L \tag{6.2b}$$

$$P_A = \text{Re}[I_G(V_G - I_G R_G)] \tag{6.2c}$$

$$P_R = \frac{V_G^2}{4R_G} - \text{Re}[I_G(V_G - I_G R_G)] \tag{6.2d}$$

where Re denotes the real part of the argument and other quantities have their usual meanings. Complex notation has been introduced because filters will usually impose phase shifts on the signals passed.

Insertion Loss. The insertion loss IL (in decibels) at a particular frequency is defined as

$$\text{IL} = -10 \log \frac{P_L}{P_{\text{in}}} \tag{6.3}$$

Thus if IL = 3 dB, only 50% of the available or incident power is delivered to the load. The filter rejection RJ (in decibels) at a given frequency is defined as

$$\text{RJ} = -10 \log \frac{P_L}{P_{\text{in}}} - \text{IL}_m \tag{6.4}$$

where IL_m is the midband, or sometimes the minimum insertion loss. For example, a filter that has 2 dB minimum insertion loss must reach 22 dB insertion loss to obtain 20 dB of rejection.

Return Loss. Three related parameters, the return loss (RL), VSWR, and reflection coefficient (ρ), are commonly used to characterize filter reflections. Return loss is used with filters because it is a sensitive parameter describing filter performance. The return loss is the ratio of the input to reflected power:

$$\begin{aligned} \text{RL} &= -10 \log \frac{P_R}{P_{\text{in}}} = -10 \log \left(\frac{\text{VSWR} - 1}{\text{VSWR} + 1} \right)^2 \\ &= -10 \log(|\rho|^2) \end{aligned} \tag{6.5}$$

Table 6.1 Return Loss, Insertion Loss, and Related Parameters

RL (dB)	ρ (V)	VSWR	IL (dB)	Transmitted Power (%)
0	1	∞	∞	0
0.5	0.9441	34.75	9.64	10.87
1	0.8913	17.39	6.87	20.6
2	0.7943	8.72	4.33	36.9
3	0.7079	5.85	3.02	49.9
4	0.6310	4.42	2.20	60.2
6	0.5012	3.01	1.26	74.9
8	0.3981	2.32	0.75	84.2
10	0.3162	1.93	0.46	90.0
12	0.2512	1.67	0.28	93.7
15	0.1778	1.43	0.14	96.8
20	0.1000	1.22	0.04	99.0
25	0.0560	1.12	0.01	99.7
30	0.0316	1.07	<0.01	99.9

Some typical values for RL, ρ, VSWR, IL, and percentage of transmitted power are given in Table 6.1.

In addition to the insertion loss and return loss, phase characteristics of a filter are often very important. The transmission phase ϕ_T (in radians) is given by

$$\phi_T = \arg(I_L) \tag{6.6a}$$

and the group delay τ_D (in seconds) is defined as

$$\tau_D = \frac{d\phi_T}{d\omega} = \frac{1}{2\pi}\frac{d\phi_T}{df} \tag{6.6b}$$

where ω is in radians per second.

Group delay is important in several ways. It is a measure of how long a signal takes to propagate through the filter. Deviation from constant group delay over a given frequency band will cause FM signals to become distorted. With constant group delay, components of multifrequency signals will travel at the same velocity through the device with the result that there will be no frequency dispersion; sharp pulses will remain sharp; and so on.

Typically, signal delay can be compensated for in a system, but one has to live with dispersion. There are two ways commonly used to define the limits of frequency dispersion. Most common is to specify maximum permissible group delay variation over frequency. Alternatively, the maximum deviation from linear phase (DLP) may be specified. The DLP over a given frequency range is

the maximum deviation between the device phase and a linear phase, that is,

$$\text{DLP} = \max(\phi_T - K\omega t) \tag{6.7}$$

where constant K has been chosen to minimize the deviation from linear phase. Note that if the group delay is constant (namely for an idealized transmission line), DLP is zero.

There are some pitfalls associated with equating the device group delay with the signal time delay through the device. When the group delay is calculated by most analytic techniques, the actual physical extent of the filter is ignored, and the transient response is incompletely considered. The group delay will be a reasonable estimate of how long the main body of the signal takes to go through the filter, but the situation is more complicated.

Every filter has both a "steady state" and a "transient" response that may be quite different. That means, for the pulse-modulated signals whose outputs are complex functions of time, the rejection of a filter may be seriously altered from the steady state for pulsed signals with pulse length on the same order as or shorter than the group delay. In a channelized receiver, accounting for a transient phenomenon through filters is one of the most challenging tasks in the design, especially if the receiver is designed to measure the most accurate frequency information on short pulses.

If a single-frequency signal within the passband of a filter is abruptly turned on at time $t = 0$, the envelope of the output signal will look something like that shown in Fig. 6.3. Signal packets that arrive before the main body of the signal are called *precursors*. Note that the signal will typically overshoot and then follow a damped oscillation about its final value.

A different manifestation of the same phenomenon is shown in Fig. 6.4, which compares the steady state and rapidly pulsed on and off insertion loss

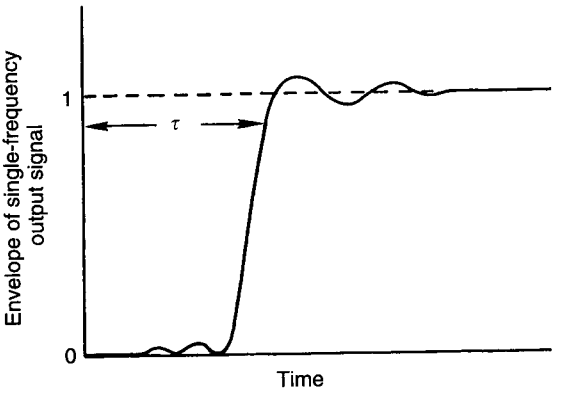


Figure 6.3 Typical output of a filter when a single-frequency input signal is abruptly switched on at $t = 0$.

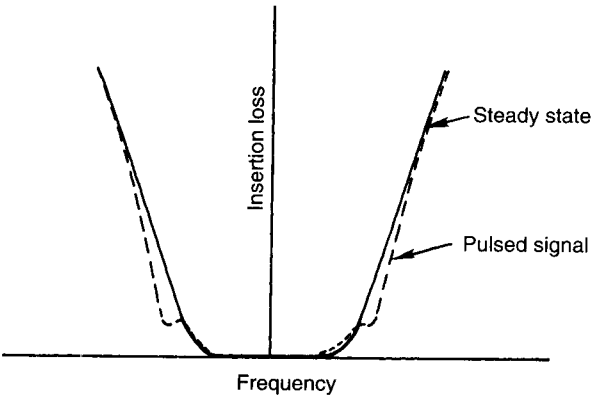


Figure 6.4 Qualitative curves comparing a bandpass filter response to pulsed and steady state signals.

response versus frequency. The two dips in the curve due to the transient response of the filter are called rabbit ears. In most applications, transient effects can be ignored if pulsewidths are longer than the group delays.

6.1.2 Basic Types

Basically there are four types of filters: low pass, bandpass, bandstop (also called band reject or notch) and high pass. Their frequency response is shown in Fig. 6.5. An ideal filter would have zero insertion loss and constant group delay across the desired frequency passbands and infinite rejection everywhere else. Practical filters deviate *substantially* from the ideal, however. In particular, no filter can operate over an unlimited frequency range. All filters exhibit spurious responses where they have rejection in the passband or regions of low loss where the rejection should be high. The best that can be accomplished is to have the filter perform well over frequencies of interest.

6.1.3 Applications

Over the past decade, the explosive growth in wireless personal communication and other portable receiver and transmitter applications has generated a significant market for low-loss, smaller size, lightweight, and lower cost filters. Various technologies such as ceramic block, low-temperature cofired ceramic (LTCC), micromachining, and high-temperature superconductors are being pursued to meet often stringent requirements for RF filters.

Almost all microwave receivers, transmitters, and test setups require filter action. The main filter functions are to reject undesirable signal frequencies

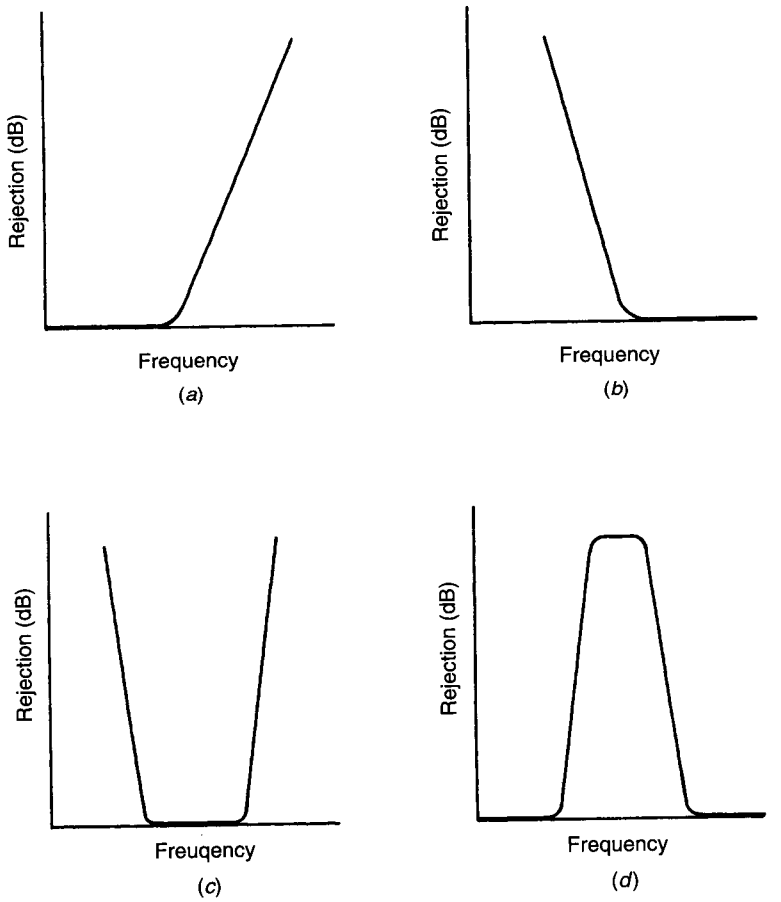


Figure 6.5 Basic types of filters: (*a*) low pass; (*b*) high pass; (*c*) bandpass; (*d*) bandstop.

outside the filter passband and to channelize or combine different frequency signals. Good examples for the former applications are mixers and multipliers. In mixers, a low-pass filter is required at the IF port to pass intermediate frequencies only while attenuating RF local oscillator, image, and other spurious frequencies. In multipliers, a passband filter is required to pass the desired harmonic while attenuating the fundamental and other harmonics. A well-described example for the latter application is the channelized receiver in which a bank of filters is used to separate input signals.

Specific applications include electronic support measure (ESM) receivers, satellite communications, mobile communications, direct broadcast satellite systems, pulse code modulation (PCM) communications, and microwave FM multiplexers.

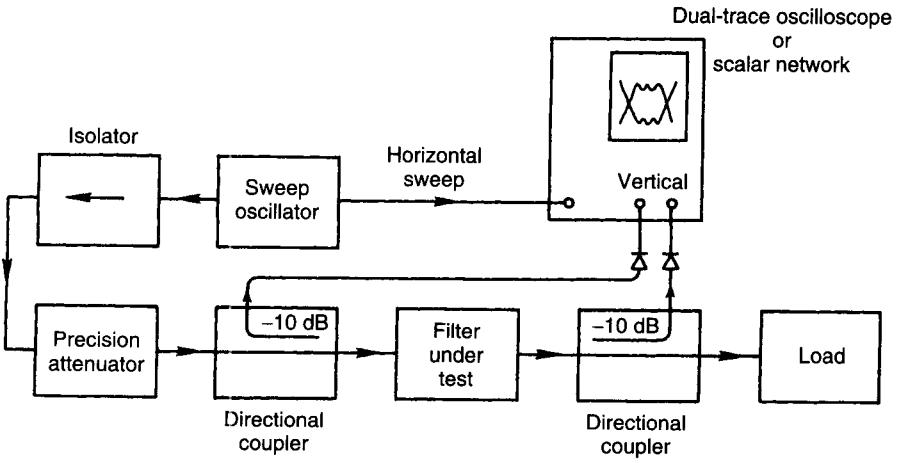


Figure 6.6 Measurement setup for insertion loss and return loss.

6.2 FILTER MEASUREMENTS

Insertion loss and return loss of filters can be measured by measuring signal magnitude only, while transmission phase and group delay measurements require a vector network. Filter measurements are briefly described in the following two sections.

6.2.1 Insertion Loss and Return Loss

A typical microwave setup for measuring insertion loss and return loss of a filter is shown in Fig. 6.6. A sweep oscillator is used as the signal source. An isolator minimizes reflections for signals flowing toward the generator. A precision attenuator is used to calibrate the detector. Directional couplers with detectors are used to monitor reflected and transmitted power. A dual-trace oscilloscope or scalar network analyzer with horizontal sweep synchronized to the sweep generator frequency ramp is used to simultaneously display the reflected and transmitted power. When designing a test setup for a particular filter, care must be taken to get the same impedance levels for the source and load as will be used in the filter operation, since filter performance is a strong function of source and load impedances.

6.2.2 *S* Parameters

Filters can be fully characterized by measuring *S* parameters by a setup using an automatic vector network analyzer, as shown in Fig. 6.7. Measurement of

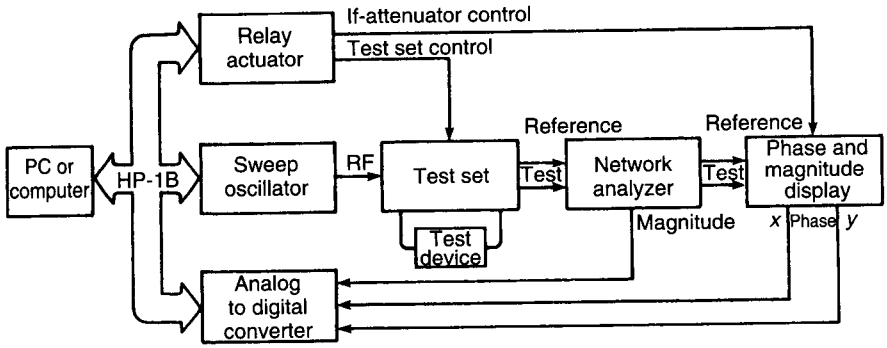


Figure 6.7 Automated *S*-parameter measurement setup.

group delay on the HP 8510 is accomplished using the phase–slope difference method. Here adjacent data buckets for S_{21} are manipulated to construct the group delay at each frequency increment, and then (6.6b) is used to determine group delay.

6.3 FILTER SYNTHESIS

Several methods are available for designing filters. Of these, the low-pass prototype filter synthesis and numerical method have been the most successful. Both methods depend critically on the judgment of the designer to choose between many possible solutions. The traditional design approach (low-pass prototype), which dates back to before computers, is to focus on special cases that allow the use of analytic solutions under highly idealized conditions. Despite drawbacks, this technique has been very successful and has been the basis for the vast majority of filter designs [1]. Furthermore, traditional design is the starting point for the second type of design using numerical methods. Therefore, we start with a discussion of the traditional approach [1–5].

6.3.1 Filter Design from Low-Pass Filter Synthesis

This method consists of the following steps:

1. design of a prototype low-pass filter with the desired passband characteristics,
2. transformation of this prototype network to the required type (low pass, high pass, bandpass, or bandstop) filter with the specified center and/or band-edge frequencies, and
3. realization of the network in terms of lumped and/or distributed circuit elements.

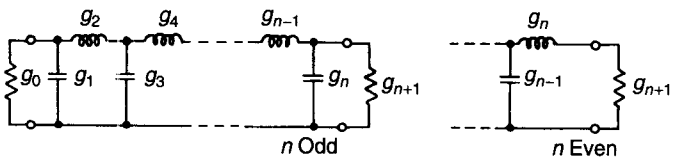


Figure 6.8 Prototype low-pass filter.

Low-pass prototype filter design using the insertion loss method is used extensively. In this method, the design of the filter starts with specifying the insertion loss as defined by (6.3) or the return loss for a lossless network as defined by (6.5) over the desired frequency band. After specifying the magnitude of the insertion loss as a function of frequency in the passband, a network that will give the desired insertion loss is then synthesized.

The combination of inductors and capacitors shown in Fig. 6.8 is obviously a low-pass circuit, since at high frequencies the series inductors open up and the shunt capacitor shorts out the signal, while at low frequencies, the series inductors short, the shunt capacitors open, and the input is connected without loss to the output. The immediate problem is how to choose element values.

There are virtually an unlimited number of different solutions to the low-pass prototype design. Here we focus on the two that have been applied the most: maximally flat (Butterworth) response, as shown in Fig. 6.9*a*, and equal-ripple (Chebyshev also often spelled Tchebycheff), as shown in Fig. 6.9*b*. Here ω_1' and ω_L' are the passband-edge and stopband-edge frequencies, respectively, and A_m is the allowed attenuation in the passband.

Maximally Flat Response. In a Butterworth low-pass prototype we want the low-pass insertion loss to be as flat as possible at zero frequency and then rise monotonically as fast as possible with increasing frequency. The attenuation loss (in decibels) for this response for $\omega_1' = 1$ may be expressed as

$$\text{IL} = 10 \log(1 + \omega'^{2n}) \tag{6.8}$$

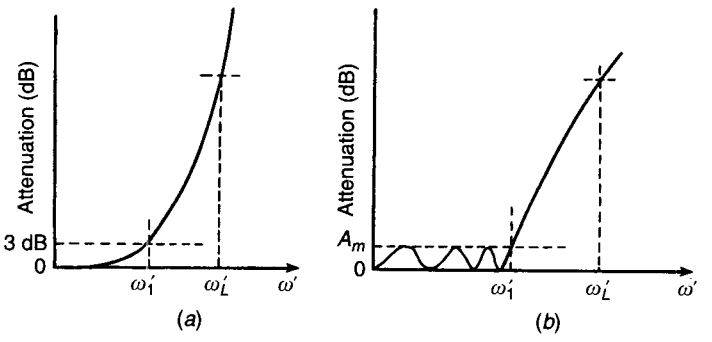


Figure 6.9 Typical attenuation response of (*a*) maximally flat and (*b*) Chebyshev filters.

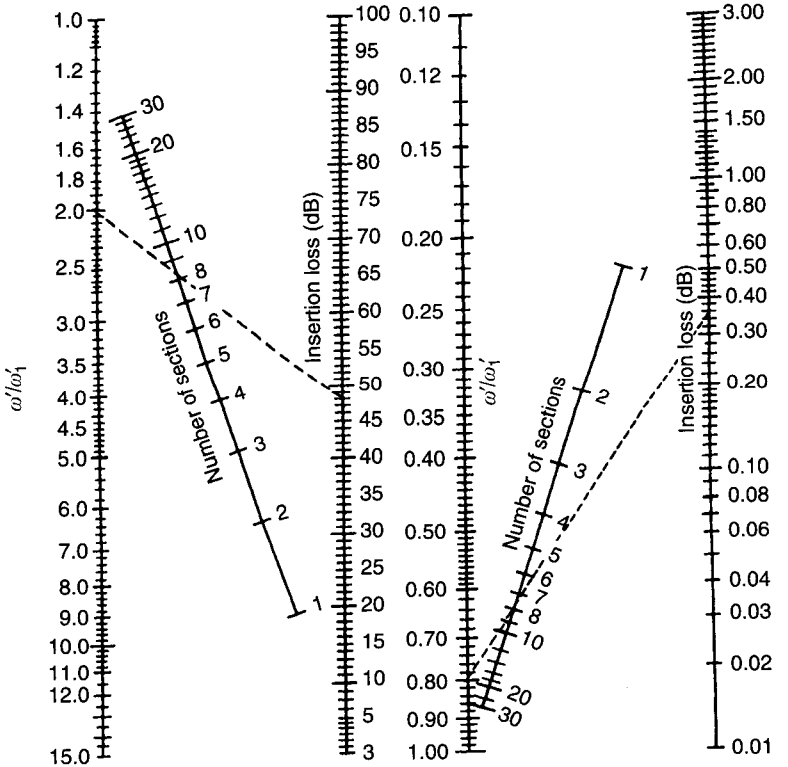


Figure 6.10 Nomograph for selecting number of sections of maximally flat filter for given insertion loss in stopband. This chart is separated into stopband (left-hand side) and passband (right-band side) regions. (*Microwaves & RF*, [6], 1985. Reprinted with permission.)

where n (order of the filter) is the number of reactive elements required to obtain the desired response. The characteristics such as stopband attenuation versus number of sections for the required bandwidth of maximally flat filters can be determined with the help of nomographs like those shown in Fig. 6.10 [6]. In most cases ω_1' is defined as the 3-dB band-edge point. The left-hand-side nomograph is used for $\omega'/\omega_1' \geq 1$ (stopband), and the right-hand-side nomograph for $\omega'/\omega_1' \leq 1$ (passband). For example, $\omega'/\omega_1' = 2.0$ for eight sections ($n = 8$) gives an attenuation of about 48 dB in the stopband, while for $\omega'/\omega_1' = 0.8$, the attenuation is about 0.35 dB in the passband.

For a Butterworth low-pass prototype, the element values g_k (shown in Fig. 6.8) may be calculated from the following equations:

$$g_0 = 1 \tag{6.9a}$$

$$g_k = 2 \sin\left[\frac{(2k-1)\pi}{2n}\right] \qquad k = 1, 2, \ldots, n \tag{6.9b}$$

$$g_{n+1} = 1 \quad \text{for all } n \tag{6.9c}$$

Table 6.2 gives element values for such filters having $n = 1, \ldots, 10$ reactive elements.

The maximally flat filter design is optimum in the sense that it provides an insertion loss response with a maximum number of frequency derivatives equal to zero at zero frequency with the rejection monotonically increasing as rapidly as possible with frequency. At microwave frequencies, maximally flat design is not as popular as the Chebyshev design.

Chebyshev Response. The Chebyshev solution addresses a slightly different problem: The insertion loss remains less than a specified level A_m up to a specified frequency ω_1' and then rises monotonically with frequency as fast as possible. For Chebyshev response, the attenuation loss (in decibels) for $\omega_1' = 1$ and $\omega' \leq 1$ may be expressed as

$$A = 10 \log[1 + (10^{A_m/10} - 1) \cos^2(n \cos^{-1} \omega')] \tag{6.10}$$

where n is the order of the filter, A_m is the ripple magnitude in decibels, and ω_1' is bandwidth over which the insertion loss has maximum ripple A_m. The Chebyshev filter nomograph [6] is shown in Fig. 6.11 illustrating (left to right) four variables: normalized frequency ω'/ω_1', number of sections n, stopband insertion loss, and passband ripple. For example, for passband ripple of 0.5 dB and $\omega'/\omega_1' = 4.6$, the insertion loss in the stopband for four sections is about 61 dB.

The g_k values for Chebyshev response may be calculated from the following equations:

$$g_0 = 1 \tag{6.11a}$$

$$g_1 = \frac{2a_1}{\gamma} \tag{6.11b}$$

$$g_k = \frac{4a_{k-1}a_k}{b_{k-1}g_{k-1}} \qquad k = 2, 3, \ldots, n \tag{6.11c}$$

$$g_{n+1} = 1 \qquad n \text{ odd} \tag{6.11d}$$

$$g_{n+1} = \coth^2 \frac{\beta}{4} \qquad n \text{ even} \tag{6.11e}$$

where

$$a_k = \sin \frac{(2k-1)\pi}{2n} \qquad k = 1, 2, \ldots, n \tag{6.12a}$$

$$b_k = \gamma^2 + \sin^2 \frac{k\pi}{n} \qquad k = 1, 2, \ldots, n \tag{6.12b}$$

$$\beta = \ln\left(\coth \frac{A_m}{17.37}\right) \tag{6.12c}$$

$$\gamma = \sinh \frac{\beta}{2n} \tag{6.12d}$$

Table 6.2 Element Values for Maximally Flat Low-Pass Prototype Filter Having $g_0 = 1$, $\omega_1' = 1$, and $n = 1,\ldots,10$

Value of n	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
1	2.000	1.000									
2	1.414	1.414	1.000								
3	1.000	2.000	1.000	1.000							
4	0.7654	1.848	1.848	0.7654	1.000						
5	0.6180	1.618	2.000	1.618	0.6180	1.000					
6	0.5176	1.414	1.932	1.932	1.414	0.5176	1.000				
7	0.4450	1.247	1.802	2.000	1.802	1.247	0.4450	1.000			
8	0.3902	1.111	1.663	1.962	1.962	1.663	1.111	0.3902	1.000		
9	0.3473	1.000	1.532	1.879	2.000	1.879	1.532	1.000	0.3473	1.000	
10	0.3129	0.908	1.414	1.782	1.975	1.975	1.782	1.414	0.9080	0.3129	1.000

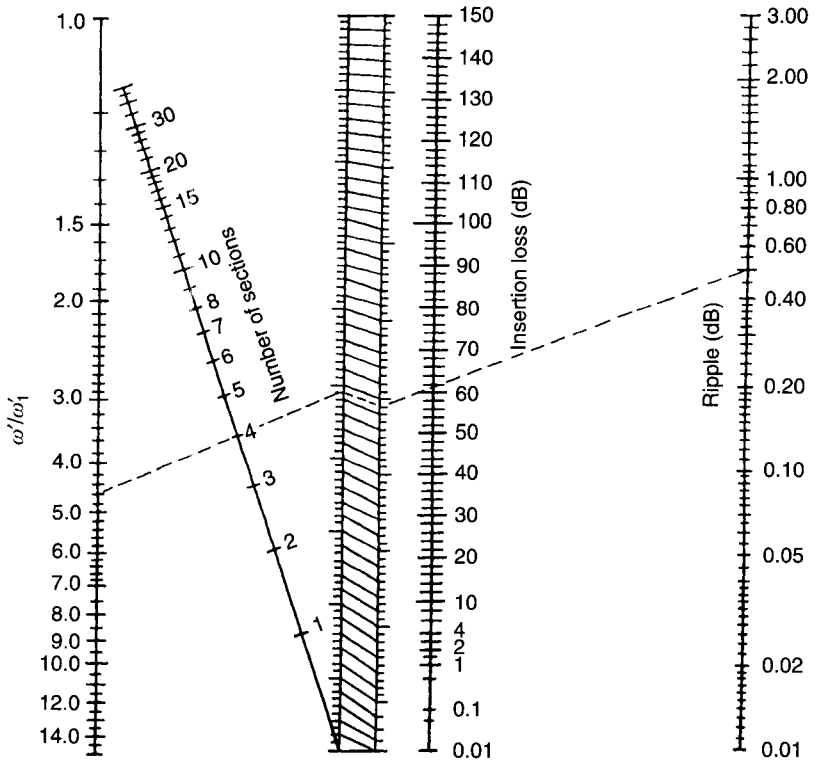


Figure 6.11 Nomograph for selecting number of sections of Chebyshev filter for given ripple and insertion loss in stopband. (*Microwaves & RF*, [6], 1985. Reprinted with permission.)

Table 6.3 gives element values for Chebyshev low-pass prototype filters having $g_0 = 1$, $\omega_1' = 1$, and $n = 1, \ldots, 10$ with various ripple values. Most microwave Chebyshev designs use a ripple value in the 0.01-dB (VSWR = 1:1.1) to 0.2-dB (VSWR = 1:1.54) range. Higher ripple values give rise to excessive ripples caused by interaction between cascaded components.

The Chebyshev filter design is optimum in the sense that for any n, and all possible element value choices, it provides the maximum possible stopband monotonic insertion loss for a specified maximum passband insertion loss ripple.

6.3.2 Special Response Filter Synthesis

The elliptic function response and generalized Chebyshev response are also frequently used for microwave filters. Bessel and Gaussian and many other similar responses realizable with a ladder prototype are seldom used at microwave frequencies because of beating effects caused by long phase lengths between cascaded components and high passband VSWRs of the preceding

Table 6.3 Element Values for Chebyshev Low-Pass Prototype Filter Having $g_0 = 1$, $\omega_1' = 1$, and $n = 1,...,10$ with Various Ripple Values

Value of n	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
					0.01 *dB Ripple*						
1	0.0960	1.0000									
2	0.4488	0.4077	1.1007								
3	0.6291	0.9702	0.6291	1.0000							
4	0.7128	1.2003	1.3212	0.6476	1.1007						
5	0.7563	1.3049	1.5773	1.3049	0.7563	1.0000					
6	0.7813	1.3600	1.6896	1.5350	1.4970	0.7098	1.1007				
7	0.7969	1.3924	1.7481	1.6331	1.7481	1.3924	0.7969	1.0000			
8	0.8072	1.4130	1.7824	1.6833	1.8529	1.6193	1.5554	0.7333	1.1007		
9	0.8144	1.4270	1.8043	1.7125	1.9057	1.7125	1.8043	1.4270	0.8144	1.0000	
10	0.8196	1.4369	1.8192	1.7311	1.9362	1.7590	1.9055	1.6527	1.5817	0.7446	1.1007
					0.1 *dB Ripple*						
1	0.3052	1.0000									
2	0.8430	0.6220	1.3554								
3	1.0315	1.1474	1.0315	1.0000							
4	1.1088	1.3061	1.7703	0.8180	1.3554						
5	1.1468	1.3712	1.9750	1.3712	1.1468	1.0000					
6	1.1681	1.4039	2.0562	1.5170	1.9029	0.8618	1.3554				
7	1.1811	1.4228	2.0966	1.5733	2.0966	1.4228	1.1811	1.0000			
8	1.1897	1.4346	2.1199	1.6010	2.1699	1.5640	1.9444	0.8778	1.3554		
9	1.1956	1.4425	2.1345	1.6167	2.2053	1.6167	2.1345	1.4425	1.1956	1.0000	
10	1.1999	1.4481	2.1444	1.6265	2.2253	1.6418	2.2046	1.5821	1.9628	0.8853	1.3554

0.2 *dB Ripple*

1	0.4342	1.0000									
2	1.0378	0.6745	1.5386								
3	1.2275	1.1525	1.2275	1.0000							
4	1.3028	1.2844	1.9761	0.8468	1.5386						
5	1.3394	1.3370	2.1660	1.3370	1.3394	1.0000					
6	1.3598	1.3632	2.2394	1.4555	2.0974	0.8838	1.5386				
7	1.3722	1.3781	2.2756	1.5001	2.2756	1.3781	1.3722	1.0000			
8	1.3804	1.3875	2.2963	1.5217	2.3413	1.4925	2.1349	0.8972	1.5386		
9	1.3860	1.3938	2.3093	1.5340	2.3728	1.5340	2.3093	1.3938	1.3860	1.0000	
10	1.3901	1.3983	2.3181	1.5417	2.3904	1.5536	2.3720	1.5066	2.1514	0.9034	1.5386

0.5 *dB Ripple*

1	0.6986	1.0000									
2	1.4029	0.7071	1.9841								
3	1.5963	1.0967	1.5963	1.0000							
4	1.6703	1.1926	2.3661	0.8419	1.9841						
5	1.7058	1.2296	2.5408	1.2296	1.7058	1.0000					
6	1.7254	1.2479	2.6064	1.3137	2.4758	0.8696	1.9841				
7	1.7372	1.2583	2.6381	1.3444	2.6381	1.2583	1.7372	1.0000			
8	1.7451	1.2647	2.6564	1.3590	2.6964	1.3389	2.5093	0.8796	1.9841		
9	1.7504	1.2690	2.6678	1.3673	2.7239	1.3673	2.6678	1.2690	1.7504	1.0000	
10	1.7543	1.2721	2.6754	1.3725	2.7392	1.3806	2.7231	1.3485	2.5239	0.8842	1.9841

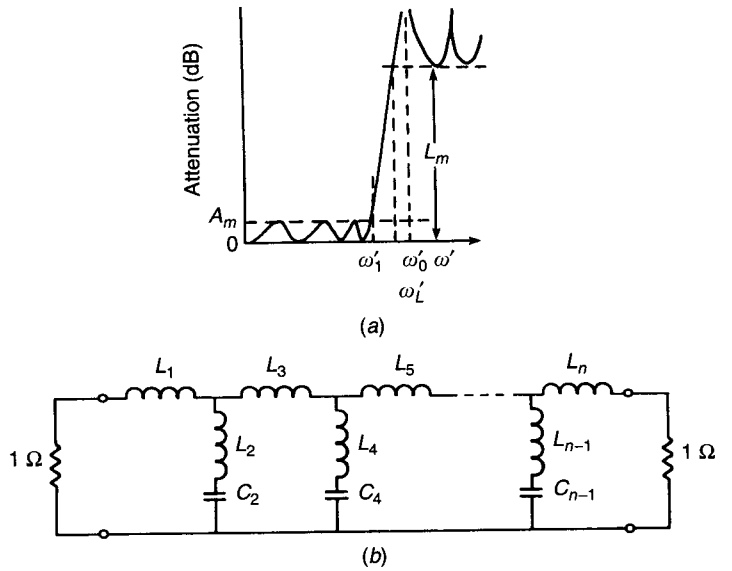


Figure 6.12 (*a*) Elliptic function response. (*b*) Prototype low-pass elliptic filter (n is odd integer).

type of designs. Synthesis of elliptic and generalized Chebyshev filters is briefly described below.

Elliptic Function Response. The elliptic filter stopband response has a series of peaks, corresponding to the number of sections, and a minimum attenuation level L_m (Fig. 6.12*a*) rather than monotonically increasing attenuation value provided by Butterworth and Chebyshev filters. The attenuation loss is expressible in several forms using elliptic functions or Chebyshev rational functions. However, these forms are not suitable for a simple calculation of attenuation loss. A complete discussion on elliptic filters is given in the literature [3–5]. Unlike the simple method of calculating g values for the Butterworth and Chebyshev filters previously described, the normalized L and C values of this class of filter as shown in Fig. 6.12*b* have been derived by synthesis [2–5]. Element values of some of the more useful cases for VSWRs up to 1.5 and $n = 3, 5, 7, 9$ sections is given by Howe [7]. In each case, the minimum attenuation L_m and the frequency ω_L' at which this attenuation occurs are also given.

The elliptic filter design is optimum in the sense that it provides a steeper stopband skirt for a given value of n with reduced passband insertion loss as compared to maximally flat and Chebyshev filters.

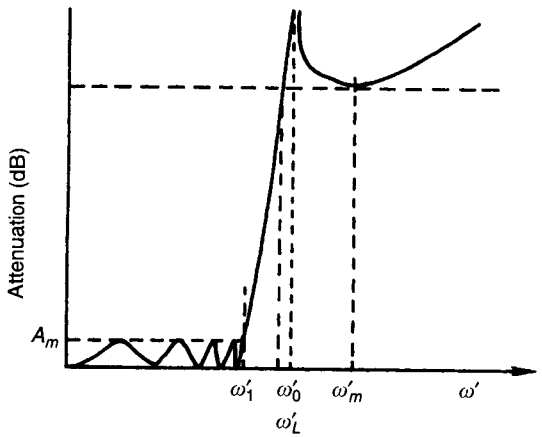
Generalized Chebyshev Response. Generalized Chebyshev response filters have been used [8, 9] to improve selectivity and physical realization in MICs.

The odd-degree generalized Chebyshev response having an equial-ripple passband, three transmission zeros at infinity, and an even multiple of transmission zeros close to the band edge has a selectivity nearly as good as the same-degree elliptic function prototype. Such filters are easier to realize physically in MICs, as impedance variation is typically less than 2:1. However, in elliptic filters, for normally required specifications, the impedance variation is as large as 10:1.

A typical insertion loss response for a generalized Chebyshev filter is shown in Fig. 6.13*a*, where ω_1', ω_L', ω_0', and ω_m' are the passband edge, stopband edge, transmission zeros, and minimum insertion loss frequencies, respectively. Element values for the doubly terminated low-pass prototype network shown in Fig. 6.13*b* are given in [8].

6.3.3 Filter Transformations

In principle, one could build low-pass prototype designs exactly, having the source and load resistance at 1 Ω and the upper edge of the passband at 1 rad/s. However, the utility of these prototypes arises from the fact that they



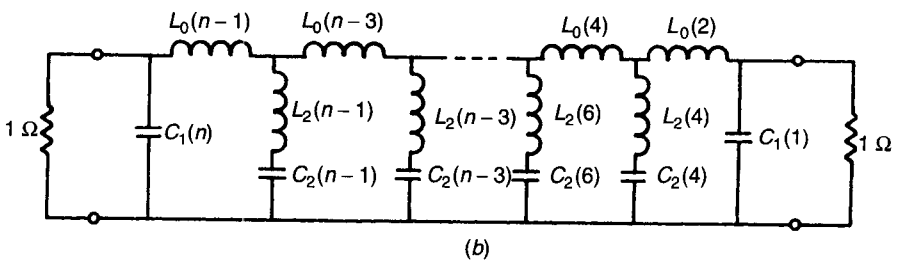


Figure 6.13 (*a*) Generalized Chebyshev insertion loss response. (*b*) Generalized Chebyshev low-pass prototype.

can be scaled to any frequency and transformed into any of the four filter types: low pass, high pass, bandpass, and bandstop.

The transformations of low-pass prototype filters with cutoff at $\omega_1' = 1$ and terminated in a 1-Ω source and load impedance into low-pass, high-pass, bandpass, and bandstop filters with arbitrary source and load impedance Z_0 are described in this section.

Low Pass. For low-pass filters, transformation of the low-pass prototype to the desired frequency band and impedance level is accomplished by multiplying g_k as follows:

$$L_k = g_k\left(\frac{Z_0}{\omega_{LP}}\right) \qquad \text{(for series inductors)} \tag{6.13a}$$

$$C_k = g_k\left(\frac{1}{\omega_{LP}Z_0}\right) \qquad \text{(for shunt capacitors)} \tag{6.13b}$$

where ω_{LP} is the required low-pass bandwidth. The following example describes step by step the design of a low-pass filter.

Example. Design a low-pass filter with a ripple factor of 0.2 dB having bandwidth of 2 GHz and attenuation at 2.4 GHz of 30 dB.

Step 1: Calculate the number of sections from Fig. 6.11:

$$\frac{\omega'}{\omega_1'} = \frac{2 \times \pi \times 2.4}{2 \times \pi \times 2.0} = 1.2$$

For $\omega'/\omega_1' = 1.2$, ripple is 0.2 dB and insertion loss is 30 dB; $n = 9$.

Step 2: Find the prototype element values from Table 6.3 for $n = 9$ and ripple of 0.2 dB:

$$g_1 = g_9 = 1.3860 \qquad g_2 = g_8 = 1.3938 \qquad g_3 = g_7 = 2.3093$$

$$g_4 = g_6 = 1.5340 \qquad g_5 = 2.3728$$

Step 3: Determine the lumped-element values from (6.13). For $Z_0 = 50\ \Omega$ and $\omega_{LP} = 2\pi \times 2 \times 10^9 = 12.57 \times 10^9$ rad/s,

$$L_1 = L_9 = 1.3860 \times \frac{50}{12.57 \times 10^9} = 5.513 \text{ nH}$$

$$C_2 = C_8 = 1.3938 \times \frac{1}{50 \times 12.57 \times 10^9} = 2.2 \text{ pF}$$

$$L_3 = L_7 = 9.1858 \text{ nH}$$

$$C_4 = C_6 = 2.4 \text{ pF}$$

$$L_5 = 9.4383 \text{ nH}$$

High Pass. The low-pass prototype network is transformed into a high-pass filter by transforming series inductances into series capacitances and shunt capacitances into shunt inductances. Using the frequency transformation gives

$$\frac{\omega'}{\omega_1'} = -\frac{\omega_{HP}}{\omega} \tag{6.14}$$

where ω_{HP} and ω are the band-edge and variable angular frequencies of the high-pass filter. The element values are obtained from

$$C_k = \frac{1}{g_k \omega_{HP} Z_0} \quad \text{(for series capacitors)} \tag{6.15a}$$

$$L_k = \frac{Z_0}{g_k \omega_{HP}} \quad \text{(for shunt inductors)} \tag{6.15b}$$

In order to illustrate the previously described transformation, we consider an example of a lumped-element high-pass filter with a passband ripple of 0.1 dB, band-edge frequency at 3 GHz, and 30 dB attenuation at 2 GHz.

Step 1: Calculate the number of sections from Fig. 6.11:

$$\left|\frac{\omega'}{\omega_1'}\right| = \frac{\omega_{HP}}{\omega} = \frac{2 \times \pi \times 3}{2 \times \pi \times 2} = 1.5$$

For $\omega'/\omega_1' = 1.5$, ripple is 0.1 dB and insertion loss is 25 dB; $n = 6$.

Step 2: Find the prototype element values from Table 6.3 for $n = 6$ and ripple of 0.1 dB:

$$g_1 = 1.1681 \quad g_2 = 1.4039 \quad g_3 = 2.0562 \quad g_4 = 1.5170$$
$$g_5 = 1.9029 \quad g_6 = 0.8618 \quad g_7 = 1.3554$$

Step 3: Determine the lumped-element values from (6.15). For $Z_0 = 50\ \Omega$ and $\omega_{HP} = 2\pi \times 3 \times 10^9 = 18.855 \times 10^9$ rad/s,

$$C_1 = \frac{1}{1.1681 \times 18.855 \times 10^9 \times 50} = 0.90834 \text{ pF}$$

$$L_2 = \frac{50}{1.4039 \times 18.855 \times 10^9} = 1.8894 \text{ nH}$$

$$C_3 = 0.51602 \text{ pF} \quad L_4 = 1.7485 \text{ nH} \quad C_5 = 0.5576 \text{ pF} \quad L_6 = 3.0779 \text{ nH}$$

$$R_7 = R_L = 50 \times 1.3554 = 67.77\ \Omega$$

For n even, $R_L \neq 50\ \Omega$, that is, input and output impedances are different. This can be avoided by selecting odd numbers of filter sections.

Bandpass. To map the low-pass prototype to a bandpass filter, the following transformation is used:

$$\frac{\omega'}{\omega_1'} = \frac{f_0}{\text{BW}}\left(\frac{f}{f_0} - \frac{f_0}{f}\right) \qquad f_0 = \sqrt{f_1 f_2} \qquad \text{BW} = f_2 - f_1 \tag{6.16}$$

where f_0, f, and BW are the center frequency ($\omega_0/2\pi$), variable frequency, and bandwidth, respectively, and f_1 and f_2 are the frequency band limits. The transformation does not result in a unique bandpass prototype. Several bandpass prototype networks are shown in Fig. 6.14. Note that the element values may be different in each circuit.

Applying the frequency transformation to series inductances and shunt

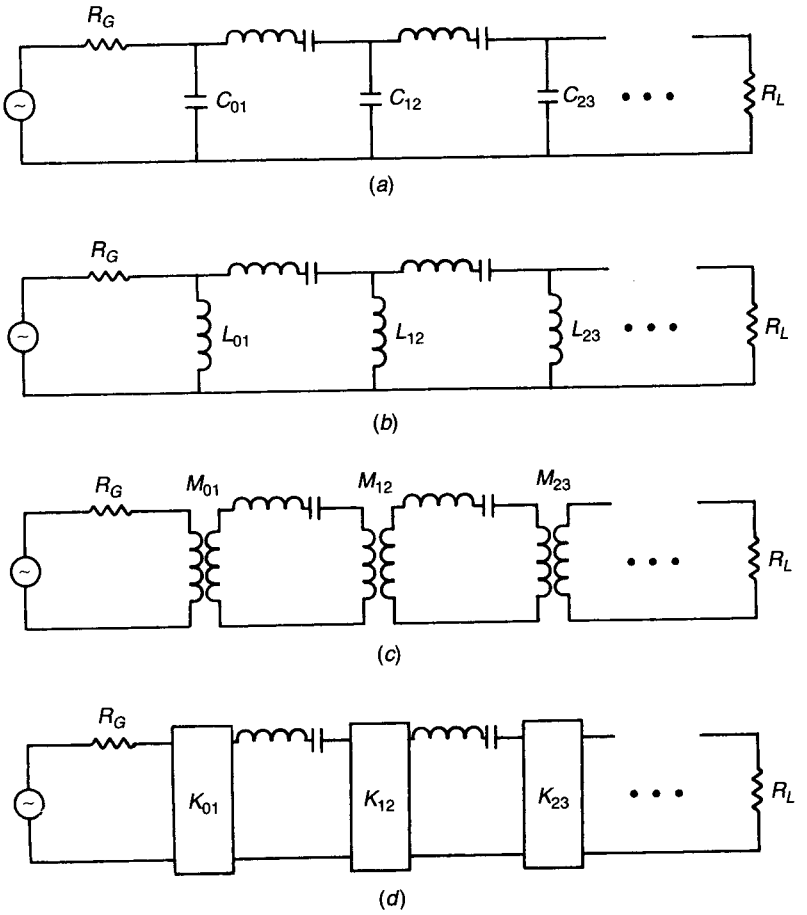


Figure 6.14 Several bandbass prototype networks: (*a*) capacitively coupled; (*b*) inductively coupled; (*c*) transformer coupled; (*d*) impedance inverted coupled.

capacitances of the low-pass prototype gives

$$L_k = g_k \frac{Z_0}{2\pi \text{ BW}} \qquad C_k = \frac{2\pi \text{ BW}}{g_k Z_0 \omega_0^2} \qquad \text{(series-tuned series elements)} \quad (6.17a)$$

$$L_k = \frac{2\pi \text{ BW } Z_0}{g_k \omega_0^2} \qquad C_k = \frac{g_k}{2\pi \text{ BW } Z_0} \qquad \text{(shunt-tuned shunt elements)} \quad (6.17b)$$

where

$$\omega_0^2 = \frac{1}{L_k C_k}$$

Example. Design a Chebyshev bandpass filter using LCs with 0.5 dB ripple and 5% bandwidth centered at 6 GHz. The minimum desirable rejection at 6 ± 1 GHz is 45 dB.

Step 1: Calculate the number of resonators from Fig. 6.11:

$$\frac{\omega'}{\omega_1'} = \frac{6}{0.3}\left(\frac{7}{6} - \frac{6}{7}\right) = 6.19$$

For $\omega'/\omega_1' = 6.19$, ripple is 0.5 dB and insertion loss is 45 dB; $n = 3$.

Step 2: Find the prototype element values from Table 6.3 for $n = 3$ and ripple of 0.5 dB:

$$g_0 = g_4 = 1.0 \qquad g_1 = g_3 = 1.5963 \qquad g_2 = 1.0967$$

Step 3: Determine the lumped-element values from (6.17). For $Z_0 = 50\ \Omega$, $2\pi \text{ BW} = 2\pi \times 0.3 \times 10^9 = 1.8850 \times 10^9$ rad/s, and $\omega_0 = 2\pi \times 6 \times 10^9 = 37.7 \times 10^9$ rad/s: For shunt-tuned shunt elements,

$$L_1 = L_3 = \frac{2\pi \text{ BW } Z_0}{g_1 \omega_0^2} = 0.0415 \text{ nH}$$

$$C_1 = C_3 = \frac{g_1}{2\pi \text{ BW } Z_0} = 16.94 \text{ pF}$$

and for series-tuned series elements,

$$L_2 = g_2 \frac{Z_0}{2\pi \text{ BW}} = 29.09 \text{ nH} \qquad C_2 = \frac{2\pi \text{ BW}}{g_2 Z_0 \omega_0^2} = 0.0242 \text{ pF}$$

Bandstop. The transformation from low-pass prototype to bandstop is given by

$$\frac{\omega_1'}{\omega'} = \frac{f_0}{\text{BW}}\left(\frac{f}{f_0} - \frac{f_0}{f}\right) \quad (6.18)$$

where all the quantities used in (6.18) were defined before. Here series inductance is mapped into a shunt-tuned circuit with element values

$$\omega_0 C_k = \frac{1}{\omega_0 L_k} = \frac{\omega_0}{2\pi \text{ BW } Z_0 g_k} \tag{6.19a}$$

and shunt capacitance into a series-tuned circuit with element values

$$\omega_0 L_k = \frac{1}{\omega_0 C_k} = \frac{\omega_0 Z_0}{2\pi \text{ BW } g_k} \tag{6.19b}$$

Example. Design a stopband filter with the following specifications:

Frequency of infinite attenuation, $f_0 = 6$ GHz

Bandwidth, BW = 300 MHz

Passband ripple = 0.5 dB

Minimum attenuation over 2% stopband = 20 dB

Step 1: Calculate the number of resonators from Fig. 6.11:

$$\frac{\omega_1'}{\omega'} = \frac{6}{0.3}\left(\frac{6.06}{6} - \frac{6}{6.06}\right) = 0.4$$

or

$$\frac{\omega'}{\omega_1'} = 2.5$$

For $\omega'/\omega_1' = 2.5$, ripple is 0.5 dB and insertion loss is 20 dB; $n = 3$.

Step 2: Find the prototype element values from Table 6.3 for $n = 3$ and ripple of 0.5 dB:

$$g_0 = g_4 = 1.0 \qquad g_1 = g_3 = 1.5963 \qquad g_2 = 1.0967$$

Step 3: Determine the lumped-element values from (6.19). For $Z_0 = 50\ \Omega$, $2\pi \text{ BW} = 2\pi \times 0.3 \times 10^9 = 1.8850 \times 10^9$ rad/s, and $\omega_0 = 2\pi \times 6 \times 10^9 = 37.7 \times 10^9$ rad/s: For shunt-tuned series elements,

$$L_1 = L_3 = \frac{2\pi \text{ BW } Z_0 g_{1,3}}{\omega_0^2} = 0.106 \text{ nH}$$

$$C_1 = C_3 = \frac{1}{2\pi \text{ BW } Z_0 g_{1,3}} = 6.65 \text{ pF}$$

and for series-tuned shunt elements,

$$L_2 = \frac{Z_0}{2\pi \text{ BW } g_2} = 24.19 \text{ nH} \qquad C_2 = \frac{2\pi \text{ BW } g_2}{\omega_0^2 Z_0} = 0.029 \text{ pF}$$

6.3.4 Impedance and Admittance Inverters

Impedance and admittance inverters play a very important role in classical filter design. Because of the inverting action, a series inductance (or shunt capacitance) with an inverter on each side looks like a shunt capacitance (or series inductance) from its external terminals. Making use of this property, low-pass filters can be realized with only one kind of reactance. Similarly, bandpass filters may be realized by series inductance–capacitance (LC), series resonant circuits separated by impedance inverters or shunt LC parallel resonant circuits separated by admittance inverters. Under ideal conditions, both networks, prototypes and circuits with inverters, will have identical transmission characteristics.

The impedance inverter K and admittance inverter J are defined in Fig. 6.15, and their simple realizations are a quarter-wavelength transmission line of characteristic impedance K and characteristic admittance J, respectively. Since in the design equations these inverters operate like quarter-wavelength lines at all frequencies, they are useful in filter designs having bandwidths up to 20%. This bandwidth limitation may be increased to about 40% if K and J inverters are used alternately.

The inverter derives its name from the impedance seen looking into a reactively terminated inverter:

$$Z' = \frac{K^2}{Z} \quad (K \text{ inverter}) \tag{6.20a}$$

$$Y' = \frac{J^2}{Y} \quad (J \text{ inverter}) \tag{6.20b}$$

The calculation of element values for bandpass filters using distributed resonators proceeds as follows:

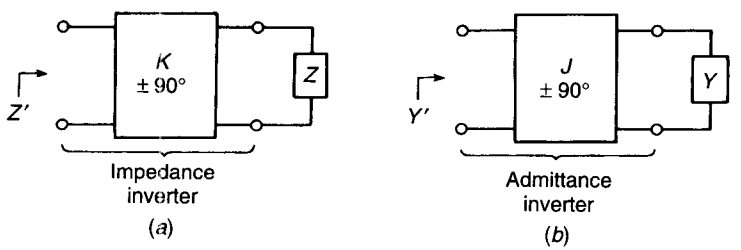


Figure 6.15 Definition of inverters: (*a*) impedance (K); (*b*) admittance (J).

$$\frac{K_{01}}{Z_0} = \frac{J_{01}}{Y_0} = \sqrt{\frac{\pi v}{2}\frac{\text{BW}}{f_0}}\frac{1}{\sqrt{g_0 g_1}} \tag{6.21a}$$

$$\frac{K_{r,r+1}}{Z_0} = \frac{J_{r,r+1}}{Y_0} = \frac{\pi v}{2}\frac{\text{BW}}{f_0}\frac{1}{\sqrt{g_r g_{r+1}}} \tag{6.21b}$$

$$\frac{K_{n,n+1}}{Z_0} = \frac{J_{n,n+1}}{Y_0} = \sqrt{\frac{\pi v}{2}\frac{\text{BW}}{f_0}}\frac{1}{\sqrt{g_n g_{n+1}}} \tag{6.21c}$$

where $g_0, g_1, \ldots, g_{n+1}$ are low-pass prototype values, $v = 1$ for half-wavelength-long resonators, and $v = \frac{1}{2}$ for quarter-wavelength-long resonators.

There are numerous other circuit elements that have good inverting properties over a much wider bandwidth than does a quarter-wavelength line. Figure 6.16 shows four inverting circuits for use as K inverters (i.e., inverters to be used with series resonators). They are particularly useful in circuits where the negative L or C can be absorbed into adjacent positive series elements. Here negative L and C means that the reactance X and susceptance B are negative, respectively, and ϕ negative or positive designates the transmission phase lags or leads, respectively. For Fig. 6.16*b*,

$$K = Z_0 \tan\left|\tfrac{1}{2}\phi\right| \qquad \Omega \tag{6.22}$$

$$\phi = -\tan 2\frac{X}{Z_0} \qquad \text{rad} \tag{6.23}$$

$$\left|\frac{X}{Z_0}\right| = \frac{K/Z_0}{1-(K/Z_0)^2} \tag{6.24}$$

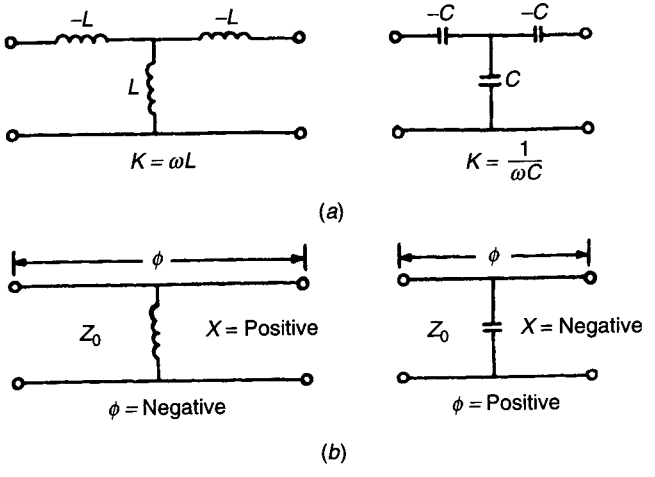


Figure 6.16 *K*-inverter circuit elements.

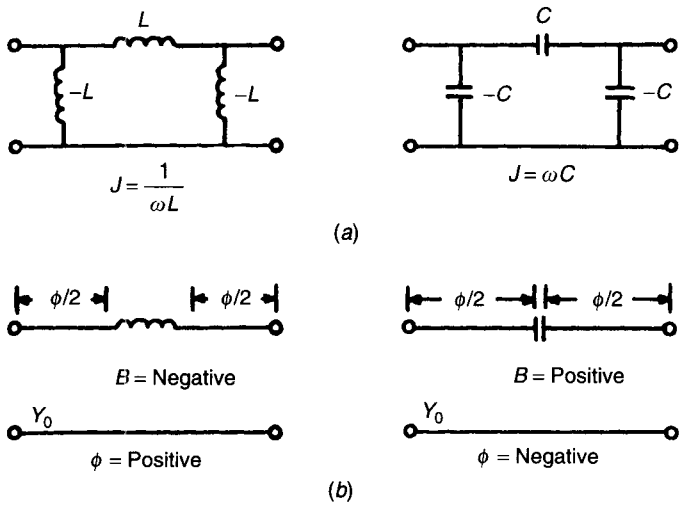


Figure 6.17 J-inverter circuit elements.

Figure 6.17 shows four inverting circuits that are useful as J inverters. In this case, for Fig. 6.17*b*

$$J = Y_0 \tan\left|\tfrac{1}{2}\phi\right| \quad \text{mho} \tag{6.25}$$

$$\phi = -\tan\frac{2B}{Y_0} \quad \text{rad} \tag{6.26}$$

$$\frac{B}{Y_0} = \frac{J/Y_0}{1-(J/Y_0)^2} \tag{6.27}$$

An open-circuited coupled line functioning as a K inverter is another example of a commonly used inverter circuit component in bandpass filters.

6.4 EXPERIMENTAL METHOD OF DESIGNING FILTERS

Bandpass filters that are physically symmetric can be realized as a simple combination of resonators. Narrow- to moderate-bandwidth bandpass filters using resonators can be designed by measuring the coupling coefficient between resonators and the external quality factor of the input and output resonators [10–13]. These measured values are then related to a normalized low-pass prototype value and can be used to realize all possible response shapes. This procedure is the most practical design method when the filter structure is complex or its equivalent circuit model is not readily available. The method can be used with microstrip, waveguide, or any other medium. For cases where accurate and

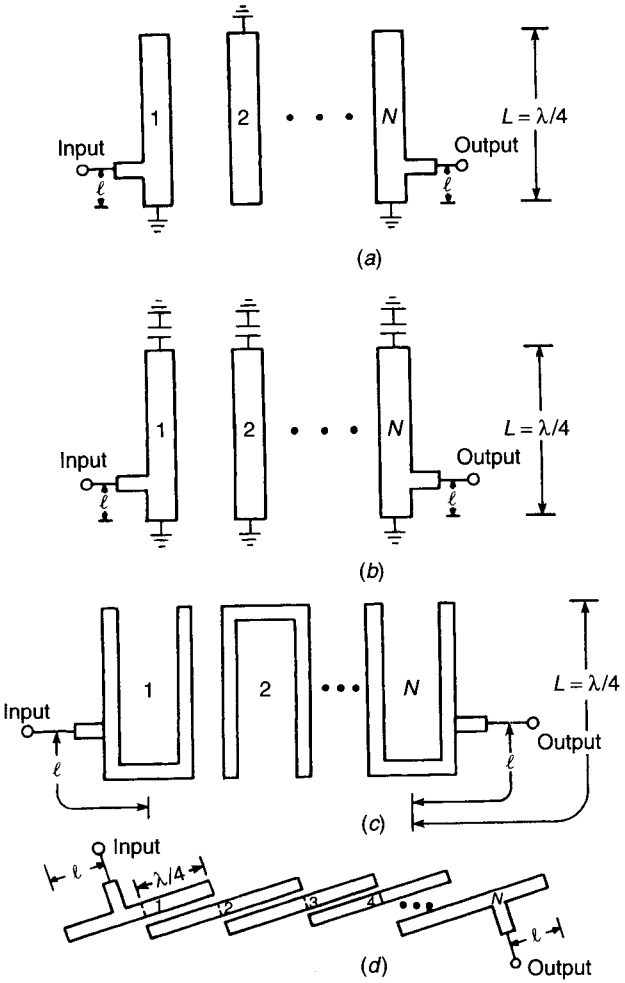


Figure 6.18 Tapped-line filter configurations: (*a*) interdigital; (*b*) combline; (*c*) hairpin line; (*d*) parallel coupled.

relatively simple equivalent circuits are available, the experimental method can be used to test the validity of the circuit models.

The experimental approach is very suitable for microstrip bandpass filters such as interdigital, combline, hairpin line, and parallel coupled. Tapped-line versions of these filters (Fig. 6.18) have more flexibility in terms of parallel coupling for the end sections and they offer space advantage. Since exact designs of tapped-line microstrip filters are not available [12, 13], experimental procedure and design curves for the sample filters are described in the following.

The first step in the filter design is to determine the coupling coefficients as a function of resonator spacing. This requires an assembly, as shown in Fig.

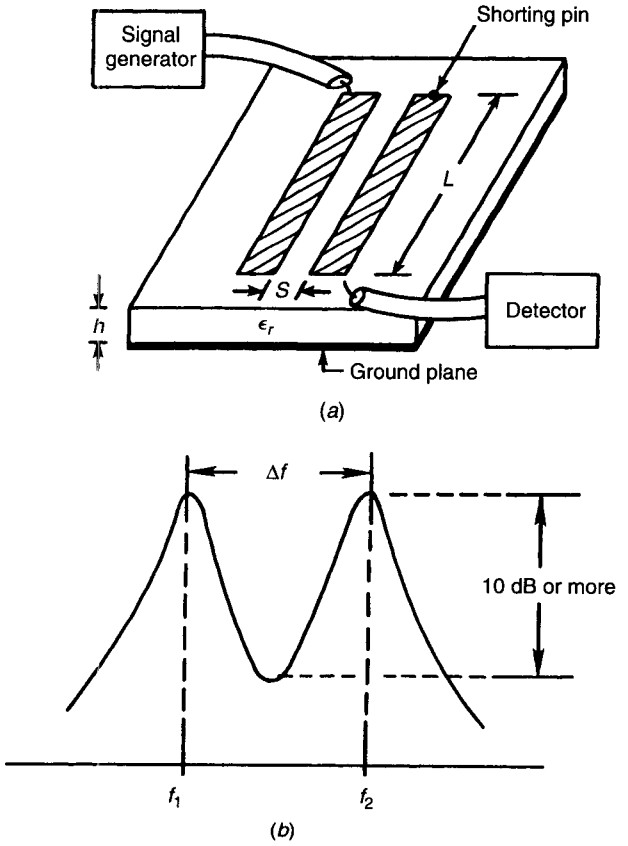


Figure 6.19 (*a*) Measurement of coupling coefficients. (*b*) Frequency response of a double-tuned resonator pair.

6.19*a*, for interdigital structure. Nonresonant probes are attached to a test fixture and both at the input and output the probes are loosely coupled to the electric field of the resonators. The widths of all resonators are identical with dimensions that give good Q and sufficient freedom from spurious responses [12]. When a source is connected to the first resonator and a detector to the second resonator, the pair of resonators become a double-tuned circuit. The pair is tuned so that two peaks are of equal height, and the valley is at least 10 dB down, as shown in Fig. 6.19*b*. The coupling coefficient K of the pair of resonators is related as

$$K = \frac{f_2 - f_1}{f_0} = \frac{\Delta f}{f_0} \tag{6.28}$$

where $f_0 = \frac{1}{2}(f_1 + f_2)$.

Thus, by selecting a number of pairs of resonators with physically realizable gaps between each pair (ranging from very narrow to very wide), a curve of K versus S/h can be experimentally obtained. Figure 6.20*a* shows the measured coupling coefficient as a function of S/h for $\epsilon_r = 2.22$ and $W/h = 1.8$, which corresponds to a single-strip impedance of approximately 70 Ω. A similar curve can be obtained for other linewidths and dielectric substrates. Figure 6.20*b* shows the measured coupling coefficient as a function of S/h for alumina substrate ($\epsilon_r = 9.8$) and $W/h = 0.7$, which corresponds to single-strip impedance of approximately 58 Ω.

The next step in the design procedure is finding the necessary normalized coupling coefficients in terms of low-pass prototype element values and design frequencies as follows:

$$K_{n,n+1} = \frac{\text{BW}}{f_0\sqrt{g_n g_{n+1}}} \tag{6.29}$$

where BW is the equal-ripple bandwidth for Chebyshev response or the 3-dB bandwidth for maximally flat response, f_0 is the center frequency of the proposed filter, and g_n are the low-pass prototype element values normalized to $\omega_1' = 1$ and $Z_0 = 1$.

The final step in the tapped filter is to measure the loaded Q of the first and the last resonators in order to obtain tap-point locations. With a simple test fixture as shown in Fig. 6.21*a*, where the signal generator is well matched and the nonresonant detector is loosely coupled, the frequency response of the first resonator is measured for various tap locations. In this case both ends of resonator 2 are short circuited, and resonator 1 becomes a single-tuned circuit whose frequency response is shown in Fig. 6.21*b*. The loaded Q is then calculated using

$$Q_L = \frac{f_0}{\text{BW}_{3\text{dB}}} \tag{6.30}$$

Variation of singly loaded Q for a tapped interdigital resonator on RT/duroid as described previously is plotted in Fig. 6.22 as a function of l/L for $Z_0 = 50$ Ω and $Z_{0I} = 70$ Ω, where Z_{0I} is the filter internal impedance.

In a filter design, the singly loaded Q is calculated from

$$Q_L = \frac{f_0}{\text{BW}} g_1 = \frac{f_0}{\text{BW}} g_{n+1} \tag{6.31}$$

where all these variables were defined previously. Once the singly loaded Q_L is known, the tap point l/L of Fig. 6.21*a* can be calculated from Fig. 6.22 or from the equation [12]

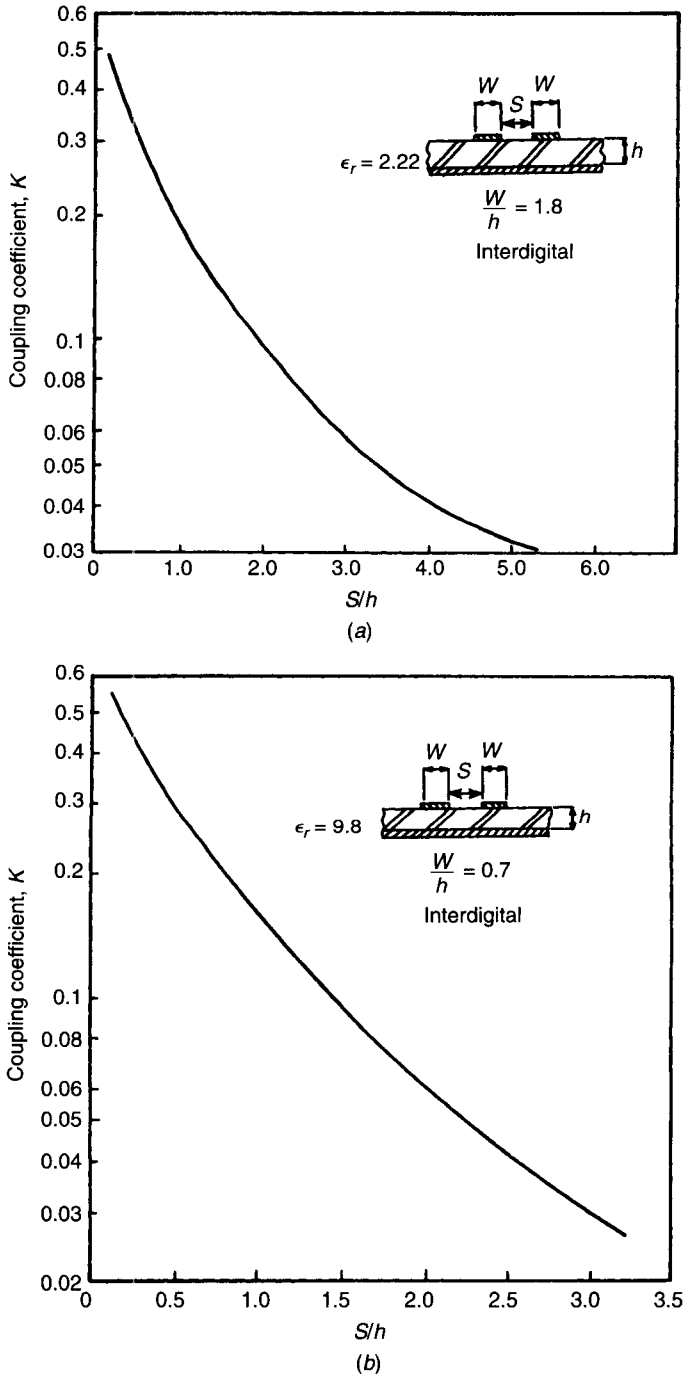


Figure 6.20 Measured combline coupling coefficient versus S/h for (*a*) RT/duroid and (*b*) alumina. (After Wong [12]. Reprinted with permission of IEEE.)

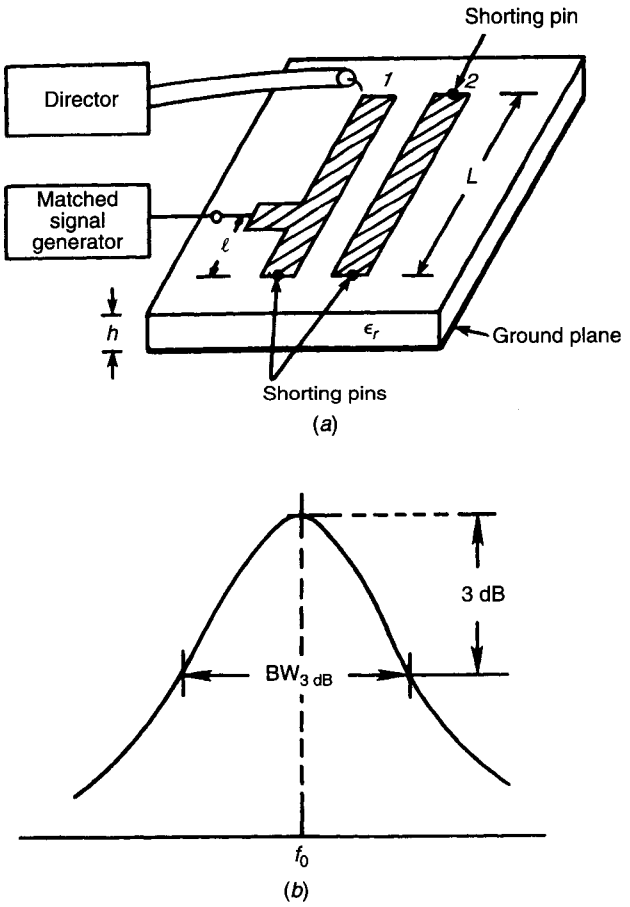


Figure 6.21 (*a*) Measurement of *Q*. (*b*) Frequency response of a singly tuned resonator.

$$\frac{Q_L}{Z_0/Z_{0I}} = \frac{\pi}{4\sin^2(\pi l/2L)} \tag{6.32}$$

The procedure just described is applicable to all types of coupled resonator filters whether realized in microstrip or in any other medium.

6.5 FILTER MODELING

6.5.1 Narrow-Band Approximation

In developing the prototype bandpass filter just presented, it has been assumed that couplings between resonators were frequency independent and that the resonator impedances were antisymmetric in frequency about the center fre-

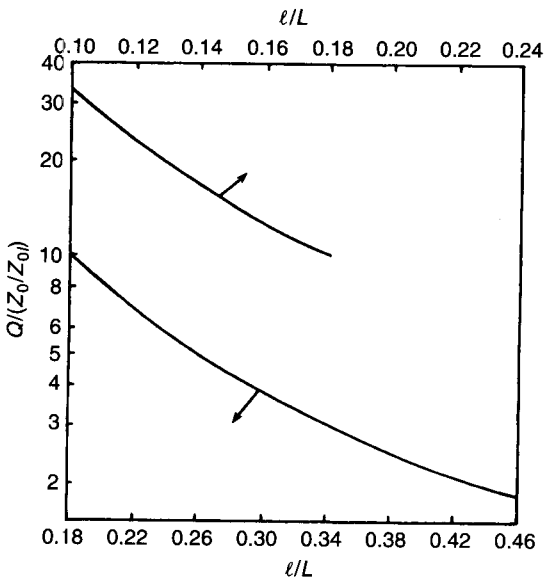


Figure 6.22 Singly loaded Q for tapped interdigital resonator. (After Wong [12]. Reprinted with permission of IEEE.)

quency (f_0). These two assumptions make up the *narrow-band approximation* and result in the insertion loss as well as the group delay being symmetric in frequency about f_0. This "approximation term" derives its name from the fact that the result will be accurate for any type of resonator or coupling provided that the percentage frequency range of interest is small enough. The narrow-band approximation typically works reasonably well for percentage bandwidths of up to 1–2% for waveguide filters and 10–15% for MIC filters.

In all cases the narrow-band approximation is a good design starting point for bandpass filters. Your intuition will at least tell you in what direction you are erring. For example, series capacitively coupled filters will have tighter coupling above f_0 than below, with the result that there will be relatively less rejection above f_0 than below. Series inductively coupled resonators are skewed the other way, however. To calculate the response more exactly, one can apply the techniques to be presented next.

6.5.2 Filter Analysis

Analytical formulations for the insertion loss and return loss for Butterworth filters, Chebyshev filters, and so on, were described in Section 6.3, but phase characteristics have not been calculated yet. Also, at microwave frequencies, all lumped elements have distributed effects (each element has to be characterized as a suitable combination of LRC; see Chapter 2) and resonators have finite Q. One can calculate phase information along with actual performance of the filter

structure by calculating the electrical response directly from the circuits instead of using the analytic formulation. The objective is to calculate I_G and I_L as shown in Fig. 6.1*b*, since once these currents are known, we have already shown in Section 6.2 how to calculate the parameters we are interested in. There are two techniques commonly used to solve filter circuits: the *ABCD* matrix and Kirchhoff's equations. These methods are briefly described below.

***ABCD*-Matrix Method.** The definition and properties of the *ABCD* matrix are given in Appendix C. From Figs. 6.1 and C.1,

$$I_G = I_1$$

$$I_L = I_2 = \frac{V_G}{AR_L + B + R_G(CR_L + D)} \tag{6.33}$$

Using (6.33) in (6.3) yields

$$\text{IL} = -10 \log \frac{4R_G R_L}{[AR_L + B + R_G(CR_L + D)]^2} \tag{6.34}$$

Similarly return loss and other electrical parameters of a filter can be expressed in terms of A, B, C, and D. If two-port devices are connected together, the overall *ABCD* matrix is just the matrix product of the individual matrices. One must remember that matrices are not multiplicatively transitive; one has to multiply the matrices in the right order. Thus the problem at hand is reduced to finding the *ABCD* matrix of the overall circuit.

A circuit that can be constructed as a series connection of two-port devices is called a *ladder network*. If one knows the *ABCD* matrix for each two-port circuit element in the ladder network, one can multiply the matrices together two by two to eventually get the overall *ABCD* matrix. This is an efficient way of doing the calculation that lends itself well to computer simulation. The *ABCD* matrices for a variety of elements are given in Table C.1 (in Appendix C). It is a simple matter to write a computer program to calculate the resultant response once the individual *ABCD* matrices are known.

The main limitation of the *ABCD*-matrix technique is that it cannot handle reentrant combinations (nonladder networks) of two ports that play a very important role in high-performance bandpass filters. This limitation can be overcome by solving Kirchhoff's equations or using nodal analysis.

Kirchhoff's Equations Method. To illustrate this method, an example of a prototype bandpass filter of order n is chosen. A typical two-resonator bandpass filter with all possible couplings included is shown in Fig. 6.23. The resultant equations for the n resonator network are given in Table 6.4. In this case, each resonator is coupled to every circuit node in the filter. The resonators that lie next to each other in the ladder network (Z_1, Z_2 or Z_7, Z_8) are called adja-

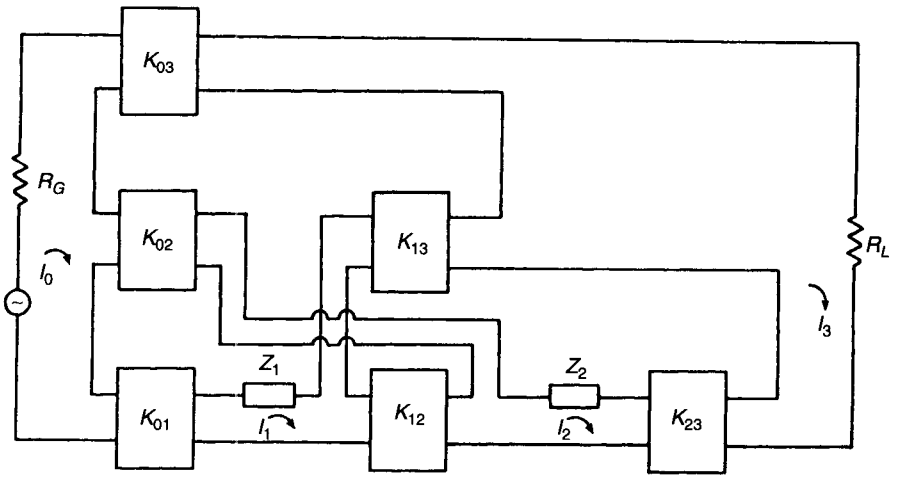


Figure 6.23 A two-resonator bandpass filter with all bridge couplings included.

cent resonators and couplings between them are called *mainline couplings*. Couplings between nonadjacent resonators are called *bridge couplings*. If all the bridge couplings are zero, the network reduces to the original ladder network.

The usual technique to solve a set of equations such as given in Table 6.4 is to perform Gaussian elimination with pivoting followed by back substitution as is described in most linear algebra texts. The calculation speed will likely not be an issue if all that is desired is to calculate the response, so a generalized linear equation solving subroutine should work fine. If speed is an issue, as may be the case in numerical design, the following comments may be useful:

1. If triangularization begins with the first term and proceeds down the main diagonal, good accuracy can usually be obtained without pivoting.
2. The matrix has symmetry about the main diagonal. About a factor of 2 speed up can be obtained by making use of this symmetry.

Table 6.4 Resultant Set of $n + 2$ Linear Equations for Resonator Bandpass Filter with all Possible Bridge Couplings

$$\begin{bmatrix} R_G & jK_{01} & jK_{02} & jK_{03} & \cdots & jK_{0,n+1} \\ jK_{01} & Z_1 & jK_{13} & jK_{13} & \cdots & jK_{1,n+1} \\ jK_{02} & jK_{12} & Z_2 & jK_{23} & \cdots & jK_{2,n+1} \\ jK_{03} & jK_{13} & jK_{23} & Z_3 & \cdots & jK_{3,n+1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & & Z_n & jK_{n,n+1} \\ jK_{0,n+1} & jK_{1,n+1} & jK_{2,n+1} & jK_{3,n+1} & jK_{n,n+1} & RL \end{bmatrix} \begin{bmatrix} I_0 \\ I_1 \\ I_2 \\ I_3 \\ \vdots \\ I_{n+1} \end{bmatrix} = \begin{bmatrix} V_G \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

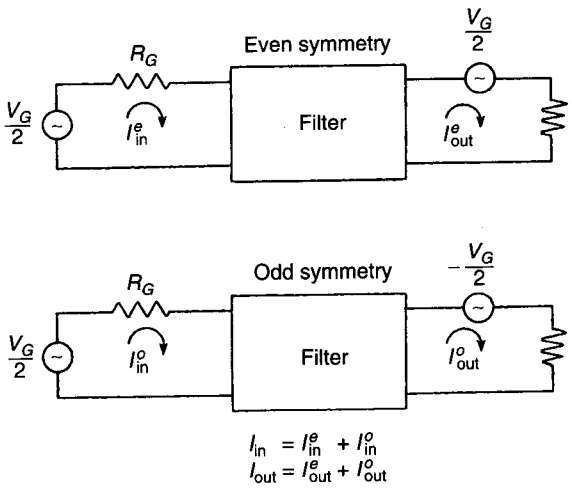


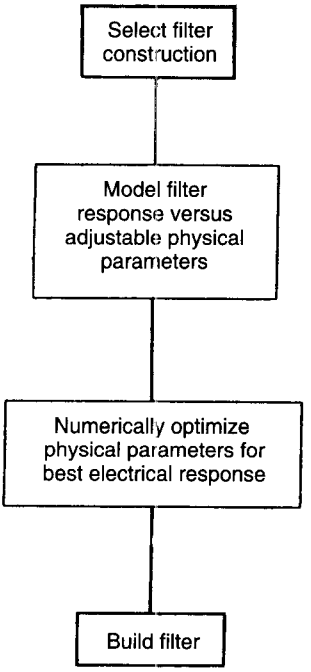
Figure 6.24 Even and odd symmetry excitations of a filter.

3. Finally, many designs, including Butterworth and Chebyshev designs, result in matrix symmetry about the minor axis as well as the major axis. These symmetries can be used to obtain a factor-of-4 computer speed up. One way to obtain this is to use an even–odd solution as shown in Fig. 6.24.

When the symmetrical filter is driven at both ends with in-phase signals (even symmetry), $I_0 = I_{n+1}$, $I_1 = I_n, \ldots$, by symmetry, so only $\frac{1}{2}(n+3)$ equations need to be solved. Similarly, when both ends are driven with out-of-phase signals (odd symmetry), $I_0 = -I_{n+1}$, $I_1 = -I_n, \ldots$, and again $\frac{1}{2}(n+3)$ equations must be solved. Since the time to solve n equations is roughly proportional to n^3, solving half as many equations twice results in a factor-of-4 savings.

6.5.3 Numerical Techniques

Microwave filters have traditionally been designed using the filter synthesis approach described in Section 6.3, in which the design is based on a previously derived transfer function. This technique is simple and accurate when applied to narrow-band and high-Q filters where nonideal effects can be ignored. Problems that arise when nonideal effects such as low Q and coupling dispersion become important are often manageable by designing with enough margin so that, although the filter response differs from the transfer function, the response is still acceptable. A fundamental concern remains. One would really prefer a response that is the best obtainable as limited by physical constraints, not a response limited by how well the filter approximates a transfer function that itself was an approximate solution to the design problem.



Advantages

1. Filter performance is limited by physical constraints.
2. Response obtained is as modeled.

Disadvantages

1. Special software is required.
2. Technique is computer intensive.

Figure 6.25 Numerical design techniques flowchart.

In this section, we introduce numerical methods that can be applied when the highest performance and accuracy are needed. These techniques are applied most frequently to high-performance waveguide filters; traditional design will normally suffice for MIC filters. However, it is important to be aware that more advanced techniques are available should the need arise.

The numerical design method, as summarized in Fig. 6.25, has been successfully applied to design filters that are close to truly optimal performance in the sense that the response is limited by physical constraints. The design procedure begins by selecting the construction to be used to meet the filter specifications. Next, one must model the filter response versus the physical parameters of the filter. At this point, accurate models (measured or analytical) for filter structure must be used. Once the response can be calculated, one must optimize the filter parameters to fit the application. This is done by first defining an error function that measures the difference between the calculated and the desired response. For filters, the error function chosen might be

$$\text{Error} = (\text{RL}_C - \text{RL}_O)^2 + (\text{RJ}_C - \text{RJ}_O)^2 + \cdots \tag{6.35}$$

where RL and RJ are return loss and rejection defined over the required bandwidth and subscripts C and O denote calculated and optimum, respectively. Next, a numerical optimization routine is used to minimize the error function.

6.6 ELECTROMAGNETIC SIMULATION

Passive components such as filters can be accurately simulated using an EM simulator also known as an EM field solver. In the past decade, outstanding progress made on personal computers and workstations in terms of speed, memory, and cost has led to the development of many commercial EM simulators. These simulators are commonly used to model circuit elements such as microstrip and coplanar waveguide structures, discontinuities, and coupling between transmission-line sections and discontinuities and structures using multilayer dielectric and plating.

6.6.1 Electromagnetic Simulation Methods

Several different field solver methods have been used and described in the literature [14, 15]. The most commonly used technique for planar structures is the method of moments (MoM), and for three-dimensional structures, the finite-element method (FEM) is usually used. Both these techniques perform EM analysis in the frequency domain. The FEM as compared to the MoM can analyze more complex structures but requires much more memory and longer computation time. There are several time-domain analysis techniques; among them the transmission-line matrix (TLM) and finite-difference time-domain (FDTD) methods are commonly used. Fast Fourier transformation is used to convert time-domain data into frequency-domain results. Typically, single time-domain analysis yields unlimited comprehensive frequency-based S parameters. An overview of commercially available EM simulators is given in Table 6.5. More comprehensive information on these tools can be found in recent publications [16–18].

In EM simulators the filter structure is divided into finite cells with two- and three-dimensional meshing. Maxwell's equations are solved in terms of electric and magnetic fields or current densities, which are in the form of integral-differential equations, by applying boundary conditions on these cells. Once the structure is analyzed and laid out, the input ports are excited by known sources (fields or currents), and the EM simulator solves numerically the integral-differential equations to determine unknown fields or induced current densities. Using the FEM, the six field components (three electric and three magnetic) in an enclosed three-dimensional space are determined while the MoM results in current distribution on the surface of metallic structures. Smaller cell sizes result in more accurate simulated data at the expense of longer simulation times.

All EM simulators are designed to solve arbitrarily shaped strip conductor structures and provide simulated single- or multiport S-parameter data that can be read in a circuit simulator. To perform an EM simulation, the structure to be simulated is defined in terms of dielectric and metal layers and their thicknesses and material properties. After creating the complete circuit/structure, the ports are defined and the layout file is saved as an input file for EM simulations. Then an EM simulation engine is used to perform EM analysis. After

Table 6.5 An Overview of Selected Electromagnetic Simulators Used for Passive Components

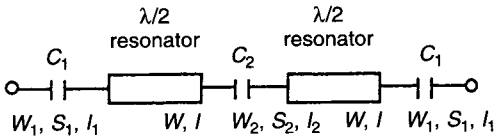
Company	Software Name	Type of Three-Dimensional Structure	Method of Analysis	Domain of Analysis
HP-EEsof	Momentum	Planar	FEM	Frequency
	HFSS	Arbitrary		
Sonnet Software	Em	Planar	MoM	Frequency
Jansen Microwave	Unisim	Planar	Spectral domain MoM	Frequency
	SFMIC	Planar		
Ansoft	Maxwell-Strata	Planar	MoM	Frequency
	Maxwell SI Eminence	Arbitrary	FEM	
Compact Software	Microwave Explorer	Planar	MoM	Frequency
MacNeal-Schwendler	MSC/EMAS	Arbitrary	FEM	Frequency
Zeland Software	IE3D	Arbitrary	MoM	Frequency
Kimberly Communications Consultants	Micro-Stripes	Arbitrary	TLM	Time
Remco	XFDTD	Arbitrary	FDTD	Time

the simulation is complete, the field or current information is converted into S parameters and saved to be used with other computer-aided design (CAD) tools.

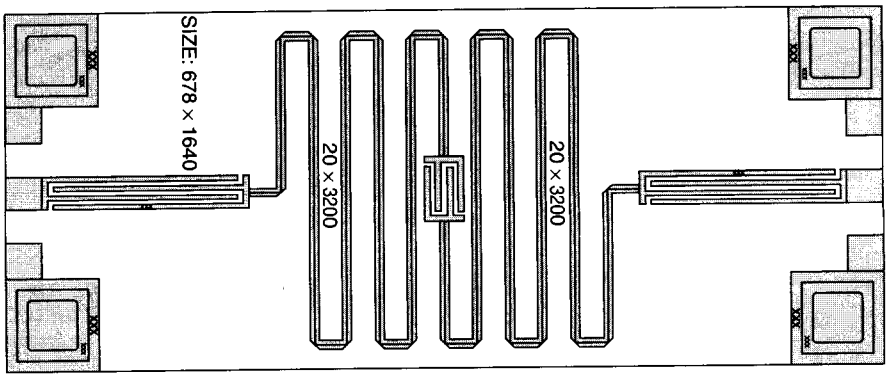
6.6.2 Filter Example

Electromagnetic simulation of several filter types has been described in the literature [19, 20]. Currently these simulators are extensively used in the design of RF/microwave filters. As an example we have selected a two-pole microstrip filter topology, as shown in Fig. 6.26*a*. It uses three interdigital capacitors as coupling elements and two $\frac{1}{2}\lambda$ high-impedance microstrip lines. The physical dimensions of the various components are given in Table 6.6. The filter is designed on a 75-μm-thick GaAs substrate. A 10-μm polyimide ($\epsilon_r = 3.2$) layer was placed between the filter conductors and the GaAs substrate. The physical layout of the bandpass filter is shown in Fig. 6.26*b*. The structure has ground–signal–ground pad configuration to make measurements on the wafer with CPW RF probes.

The filter was analyzed using the Electromagnetic Simulator from Sonnet Software. Different simulation conditions were employed in order to determine the appropriate grid size and layer thickness for greatest simulation accuracy. To improve simulation accuracy, the filter was divided into two parts. The first part is the input (and output) interdigital capacitor. This capacitor was simulated with a 10×2.5-μm grid size. The plating gold thickness (4.5 μm) effect



(a)



(b)

Figure 6.26 Two-pole filter: (*a*) Schematic and (*b*) physical layout. Dimensions in μm.

was taken into account by adding three 1.5-μm-thick metal layers. The internal part of the filter includes the two meandered transmission lines and a small interdigital capacitor in the center. This section was simulated using a 2.5 × 2.5-μm grid and one 4.5-μm layer. Figure 6.27 compares the EM simulated and measured performance and shows excellent agreement between them.

6.7 FILTER REALIZATIONS

A historical account of microwave filters is given by Levy and Cohn [21]. Filter developments include low pass, bandpass, and high pass in a variety of media such as waveguide, coaxial line, microstrip, and strip line as well as dielectric resonators. Filters can be realized in any medium depending on the desired application. The maximum useful frequency limits of various circuit elements are given in Table 6.7. Since the emphasis in this book is on integrated-type

Table 6.6 Summary of Two-Pole Filter Dimensions

Interdigital capacitor C_1: $W_1 = S_1 = 10$ μm, $l_1 = 350$ μm, $N = 4$
Interdigital capacitor C_2: $W_2 = S_2 = 10$ mm, $l_2 = 85$ μm, $N = 4$
Resonator microstrip line: $W = 20$ μm, $l = 3200$ μm measured along center line
Edge-to-edge spacing between microstrip conductors: 40 μm
Chamfer length $W' = \sqrt{2}W$

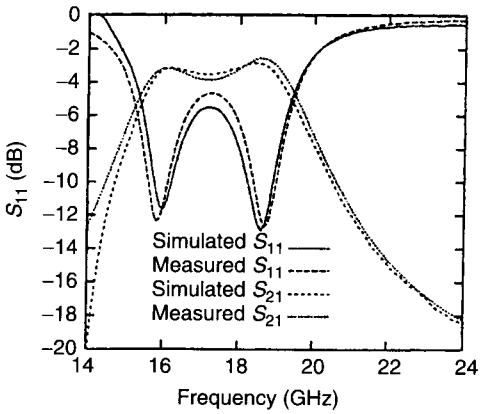


Figure 6.27 Measured and simulated performance of the two-pole filter with polyimide on a 75-μm thick GaAs substrate.

circuits, filter examples are restricted to microstrip-type and microstrip-coupled dielectric-resonator-type structures.

6.7.1 Printed-Circuit Filters

Satellite, airborne communications, and electronic warfare (EW) systems have requirements for small size, light weight, and low-cost filters. Microstrip- and strip-line filters are very suitable for wide-band applications and where the demand on selectivity is not severe. Various kinds of filters, as shown in Figs. 6.18 and 6.28, can be realized using microstrip-type structures (see Chapter 2 or Gupta et al. [22]). Note that f_{SPB} is the frequency of higher order resonances.

Table 6.7 Characteristics of Various Structures for Filters

Structure	Frequency Range (GHz)	Useful Bandwidth (%)	Q^a
Waveguide	1–100	0.1–20	≃5000
Coaxial	0.1–40	1–30	≃2000
Strip line	0.1–20	5-octave	≃150
Microstrip	0.1–100	5-octave	≃200
Suspended Microstrip	1–200	2–20	≃1000
Fin line	20–100	2–50	≃500
Dielectric resonator	1–40	0.2–20	≃10,000
Lumped elements	0.01–10 (hybrid)	20-octave	≃200
	0.1–60 (monolithic)	20-octave	≃100
Surface acoustic wave (SAW)	0.01–5		

[a] At X-band frequencies.

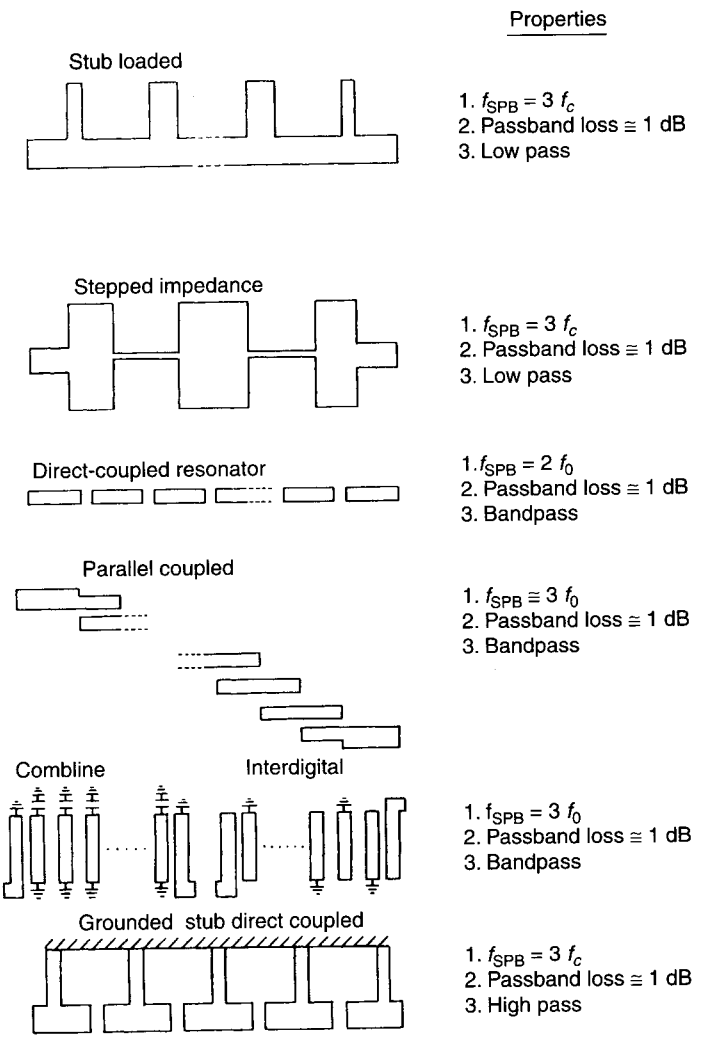


Figure 6.28 Microstrip-type filter configurations.

The suspended microstrip provides a higher Q than microstrip or strip line, as most of the energy is propagating in the air. This results in lower loss filters with sharper band edges. The wide range of impedance values achievable makes this medium particularly suitable for low-pass and broadband bandpass filters. An account of the development of many different types of filters in this medium was given in Rooney and Underkofler [23].

Filters can be realized using lumped elements, described in Chapter 2, or by employing microstrip sections. Kuroda's identities [24, 25] allow one to realize low-pass structures using shunt elements with the identical response. Richards's

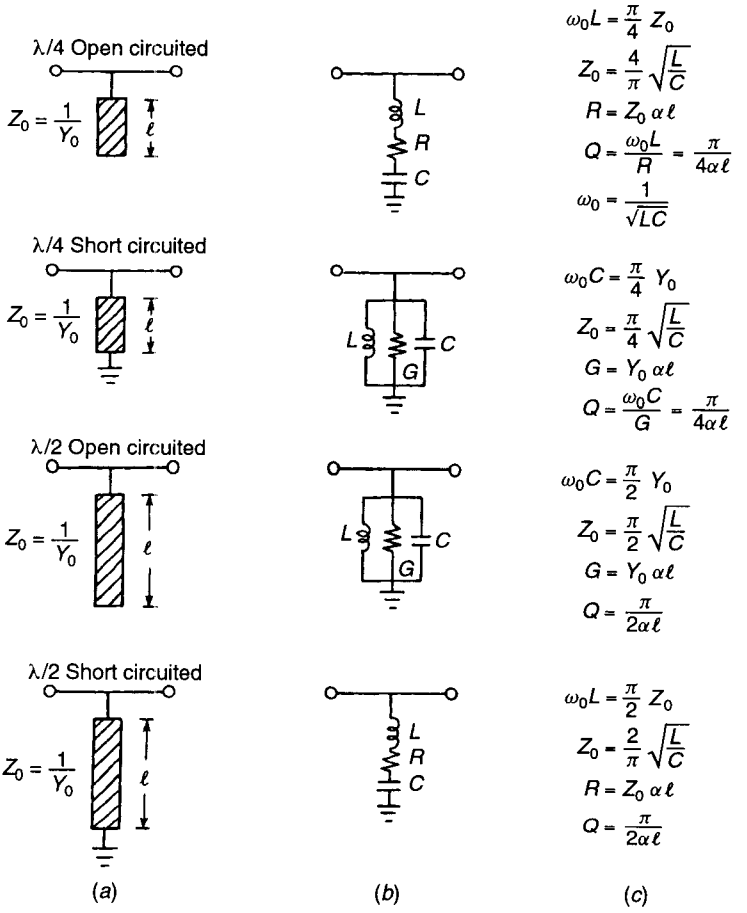


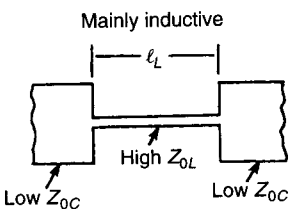
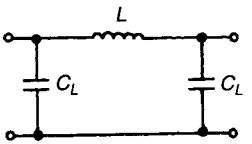
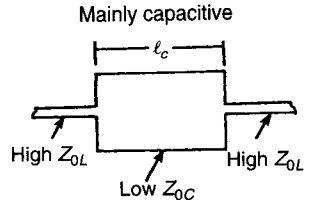
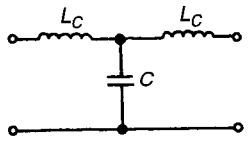
Figure 6.29 Equivalent circuits for TEM strip-line resonators: (*a*) stub configuration; (*b*) equivalent *LRC* network; (*c*) element values.

transformation [25, 26], which establishes a simple relationship between lumped and distributed circuit elements, enables one to design filters using distributed circuits.

In many microwave filter designs, a length of transmission line terminated in either an open circuit or a short circuit is often used as a resonator. Figure 6.29 illustrates four such resonators with their equivalent *LRC* networks, which were determined by equating slope parameters for both of these configurations at resonance $\omega = \omega_0$.

Low-Pass Filters. Low-pass filters in waveguide, coaxial, and strip-line form are very important components in microwave systems. A very good account of the early development is given in the classic book of Matthaei et al. [1]. Trans-

Table 6.8 Distributed Inductors and Capacitors

Distributed Element	Equivalent Circuit	Expressions	Approximate Expressions
Mainly inductive 		$L = \frac{Z_{0L}}{\omega} \sin\left(\frac{2\pi l_L}{\lambda_{gL}}\right)$ $C_L = \frac{1}{\omega Z_{0L}} \tan\left(\frac{\pi l_L}{\lambda_{gL}}\right)$	$l_L \simeq \frac{f\lambda_{gL} L}{Z_{0L}}$
Mainly capacitive 		$C = \frac{1}{\omega Z_{0C}} \sin\left(\frac{2\pi l_c}{\lambda_{gc}}\right)$ $L_c = \frac{Z_{0C}}{\omega} \tan\left(\frac{\pi l_c}{\lambda_{gc}}\right)$	$l_c = f\lambda_{gc} Z_{0C} C$

verse EM structures such as coaxial lines, strip lines, and microstrip lines are ideal for low-pass filters, and the design is approximated as nearly as possible to an idealized lumped-element circuit.

Consider an example for a microstrip low-pass filter with the following specifications:

Cutoff frequency	2 GHz
Response	Chebyshev with 0.2 dB ripple
30-dB attenuation frequency	3.5 GHz
Substrate material	Alumina, $\epsilon_r = 9.9$, $h = 0.63$ mm, $t = 6$ μm

The number of resonators required is 5. The prototype values are

$$g_0 = 1.0 \qquad g_1 = g_5 = 1.3394 \qquad g_2 = g_4 = 1.337 \qquad g_3 = 2.166$$

The lumped-element values are

$$C_1 = C_5 = 2.13 \text{ pF} \qquad L_2 = L_4 = 5.318 \text{ nH} \qquad C_3 = 3.446 \text{ pF}$$

The filter was realized using microstrip sections as described in Table 6.8. The microstrip layout of the filter is shown in Fig. 6.30. Various dimensions for

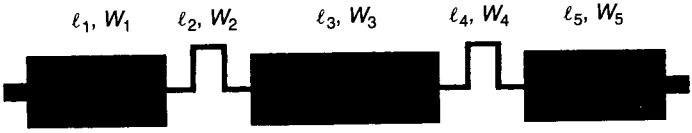


Figure 6.30 Layout of five-section low-pass filter.

Table 6.9 Design Parameters for Low-Pass Filter

	From Table 6.8		Optimized with Discontinuities
$l_{c1} = l_{c5}$ (mm)	4.85		5.15
$l_{L2} = l_{L4}$ (mm)	7.25		6.07
l_{c3} (mm)	8.99		6.92
$Z_0 = 50\ \Omega$	$W = 0.6$ mm	$\epsilon_e = 6.58$	$\lambda_g = 5.85$ cm at 2 GHz
$Z_{0C} = 20\ \Omega$	$W = 2.64$ mm	$\epsilon_e = 7.74$	$\lambda_{gc} = 5.39$ cm at 2 GHz
$Z_{0L} = 100\ \Omega$	$W = 0.075$ mm	$\epsilon_e = 5.82$	$\lambda_{gL} = 6.22$ cm at 2 GHz

microstrip sections are given in Table 6.9. Figure 6.31 shows insertion loss and return loss for this filter for ideal design, based on Table 6.8 and when optimized with discontinuities. No conductor and dielectric losses are included so far. Figure 6.32 shows a comparison between the measured and calculated responses (with losses).

Bandpass Filters. Popular configurations for printed-circuit filters (Figs. 6.18 and 6.28) are direct coupled, parallel coupled, interdigital, combline, and hairpin line. Direct-coupled resonator filters [1] have excessive length. The dimensions can be reduced by a factor of 2 with the introduction of parallel-coupled lines. Parallel coupling is much stronger than end coupling, so that realizable bandwidths could be much greater [24, 27]. In this configuration the first spurious response occurs at three times the center frequency and a much larger gap

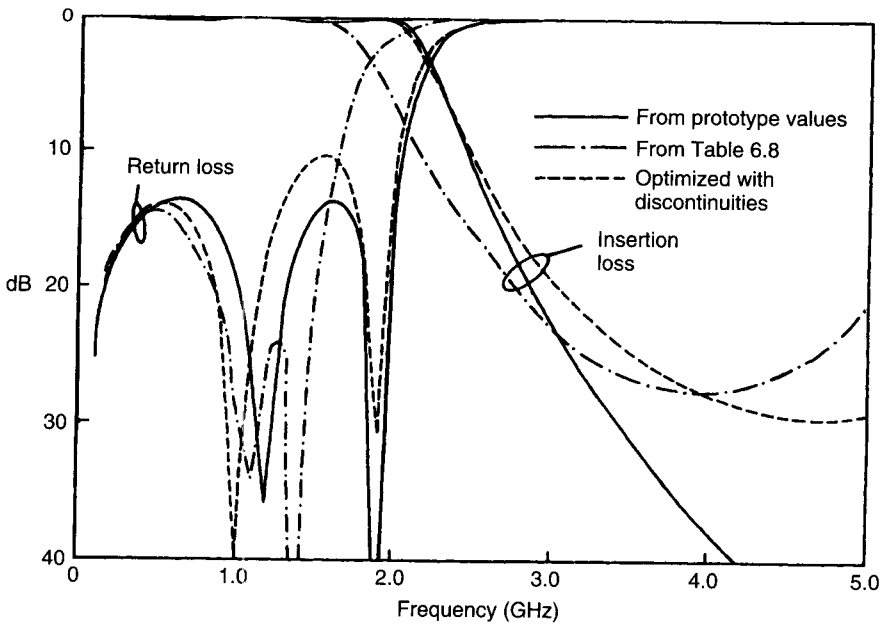


Figure 6.31 Calculated performance of five-section low-pass filter.

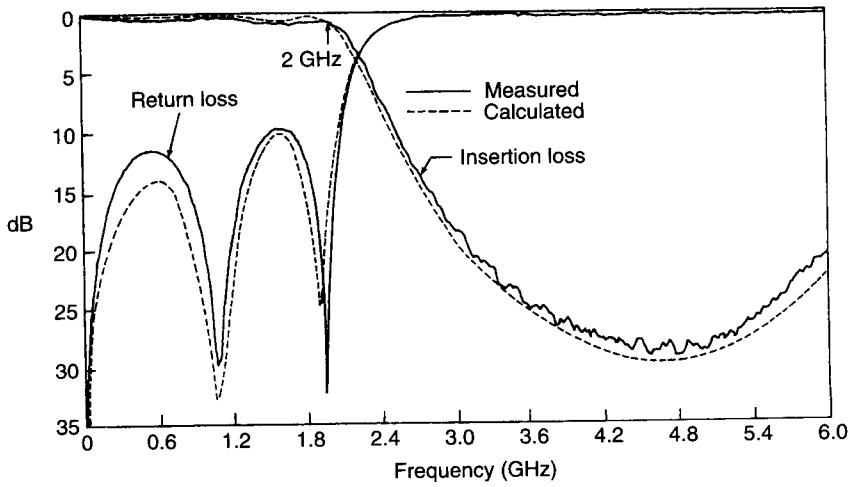


Figure 6.32 Comparison of calculated and measured responses of five-section low-pass filter realized using microstrip.

is permitted between parallel adjacent strips. The gap tolerance is also reduced, permitting a broader bandwidth for a given tolerance. Filters can be designed with reasonable accuracy using design information from the literature [1, 27–32]. In order to fabricate compact filters, resonators are placed side by side. Interdigital, combline, and hairpin-line filters are realized using this concept. Accurate design analysis of these filters, interdigital [1, 33, 34], combline [1, 35–37], and hairpin line [38, 39], are also available.

To illustrate a design example of an interdigital bandpass filter, the following specifications have been chosen:

Center frequency f_0	4 GHz
Response	Chebyshev with 0.2 dB ripple
Bandwidth	0.4 GHz
35-dB attenuation points	4 ± 0.4 GHz
Substrate	$\epsilon_r = 9.8$ and $h = 1.27$ mm

From Fig. 6.11, the number of resonators is 5. The prototype values are the same calculated for the low-pass filter example. Here $Z_0 = 50\ \Omega$ and $Z_{0I} = 58\ \Omega$ are selected for the tapped interdigital filter [12]. The coupling parameters from (6.29) and (6.31) are

$$Q_L = \frac{f_0}{\text{BW}} g_1 = 13.4$$

$$K_{12} = K_{45} = 0.0747$$

$$K_{23} = K_{34} = 0.0588$$

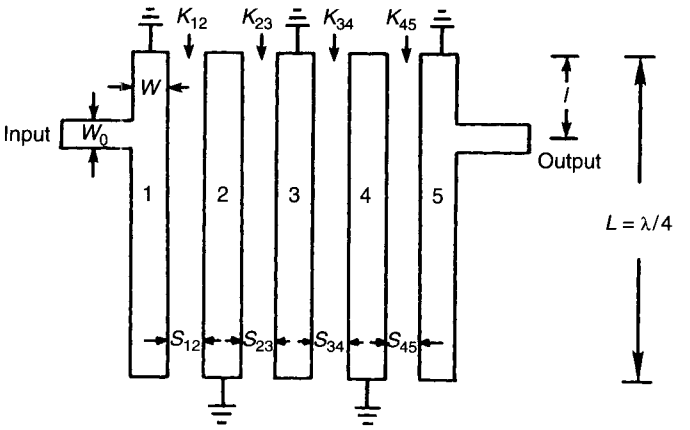


Figure 6.33 Five-section interdigital filter.

The dimensions of this filter, shown in Fig. 6.33, are determined from Figs. 6.20 and 6.22 and are given below:

$$W = 0.889 \text{ mm} \qquad L = 7.43 \text{ mm}$$

$$S_{12} = S_{45} = 2.2 \text{ mm} \qquad S_{23} = S_{34} = 2.52 \text{ mm}$$

$$\frac{l}{L} = 0.145 \qquad l = 1.077 \text{ mm}$$

Advanced RF/microwave bandpass filters such as microstrip cascaded quadruplet and trisection filters, as shown in Fig. 6.34, that meet stringent requirements for wireless communications systems have also been described in the literature [40]. These filters use microstrip open-loop resonators and are arranged in such a way that each resonator has one cross coupling. The cascaded quadruplet filter contains cascaded sections of four resonators while trisection filters consist of three resonators. In such filters, the cross couplings are adjusted to realize a pair of attenuation poles at finite frequencies to improve the insertion loss and the selectivity for a given number of resonators. Figure 6.35 shows the measured performance of a narrow-band trisection filter shown in Fig. 6.34*b* and fabricated on a 1.27-mm thick alumina substrate using copper conductors. The measured insertion loss is about 1.2 dB over the 860- to 910-MHz frequency range.

6.7.2 Dielectric Resonator Filters

Microwave IC structures have suffered from a lack of high-Q miniature elements, which are required to construct high-performance, highly stable, narrow-band filters. Filters such as bandpass and bandstop are frequently real-

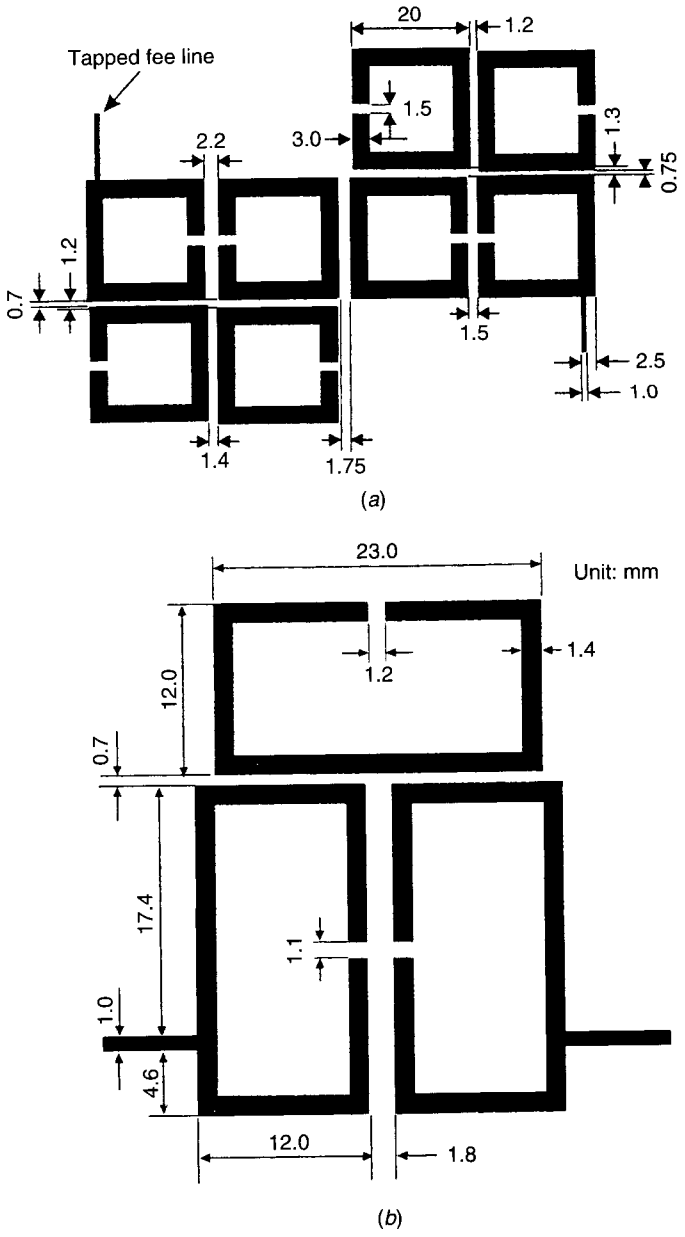


Figure 6.34 Physical layouts of cross-coupled microstrip filters: (*a*) cascaded quadruplet; (*b*) trisection. All dimensions are calculated for an alumina substrate $\epsilon_r = 10.8$ and thickness of 1.27 mm.

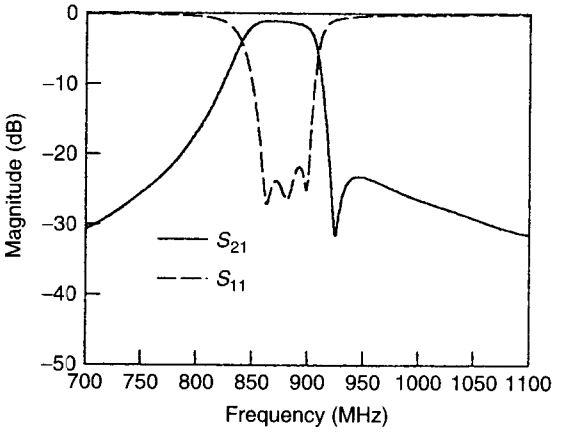


Figure 6.35 Measured performance of microstrip trisection filter.

ized using high-quality dielectric resonators. The dielectric constant is around 40, and resonator Q ($\simeq 1/\tan\delta$) lies between 5000 and 10,000 in the 2- to 7-GHz frequency range [41, 42]. The commercially available dielectric resonators have temperature stability as good as Invar these days. Filters may be constructed in all the common transmission media ranging from waveguide to microstrip. Such filters are small in size, lightweight, and low cost.

Most of the early work was carried on in the early 1960s as summarized by Cohn [43]. The basic bandpass filter topology consists of an evanescent-mode waveguide section (waveguide below cutoff) in which the dielectric resonators are housed. Figure 6.36 shows a microstrip-coupled dielectric resonator bandpass filter configuration. The most commonly used mode in the resonators is the $TE_{01\delta}$. The dual-mode or hybrid-mode $HE_{01\delta}$ resonators find applications in sophisticated elliptic function dielectric filters. Various kinds of dielectric resonator filters have been treated in the literature [44–49].

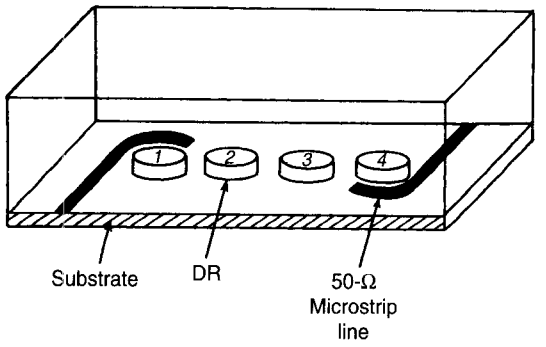


Figure 6.36 Microstrip-coupled dielectric resonator bandpass filter.

6.7.3 Ceramic Block Filters

Coaxial interdigital and combline filters using ceramic blocks are commonly used for 200- to 3000-MHz applications. These filters are based on high-dielectric-constant ceramic technology, in which filter size reduction is accomplished by a reduction in guide wavelength. A number of very high dielectric constant ($\epsilon_r \cong 40$–100) ceramic materials (e.g., barium titanates and zirconates) with very low loss (Q factor $\cong 5000$) are currently available. The filters using these materials have typically 2 dB insertion loss and have bandwidth from 1 to 20%. Ceramic block filters are temperature stable (3 ppm/°C) and their temperature range of operation is normally from −30 to +85°C. They are surface mountable, and in high volume their cost is \$2–\$5. Their specific applications include cellular radio, mobile radio, wireless local-area network (LAN), personal communication network (PCN), global positioning system (GPS), cable TV, and industrial, scientific, and medical (ISM) band.

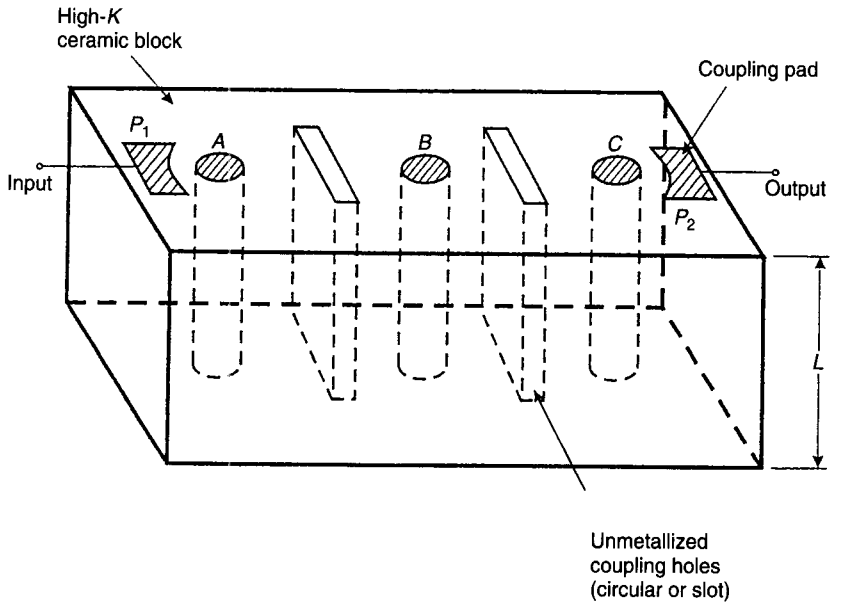
Figure 6.37*a* shows a schematic of a three-resonator combline ceramic block filter. The coupling between pairs of adjacent resonators is realized by circular/rectangular air holes. The inhomogeneous interface between the high-dielectric-constant ceramic and air hole gives rise to different phase velocities for the even and odd modes of the coupled lines. This difference provides the required coupling between the resonators to realize a filter and effectively reduces its size and cost. The design of such filters, although straightforward, requires numerical methods such as EM simulators to determine the coupling between the resonators. Normally filters are designed empirically and tuned after fabrication using ceramic grinders and metal scrapers. Analysis, design, and test results for various ceramic block filters have been discussed by many authors [50–55]. In this section, we briefly describe the design of such filters.

In Fig. 6.37*a*, A, B, and C are metallized center conductors of coaxial resonators. All resonators are short circuited at the bottom and open circuited at the top and are designed to be $\frac{1}{4}\lambda$ long at the operating center frequency. Resonators A, B, and C are coupled, to each other for filter action, through air holes between them. The first and last resonators are coupled to input and output ports, respectively, by coupling pads P_1 and P_2 located near them. The capacitive coupling between the filter and input and output ports is usually accomplished by a cut-and-try method. Figure 6.37*b* shows a lumped-element equivalent circuit for this filter; $\frac{1}{4}\lambda$ resonators are represented by parallel resonant circuits (C_i, L_i). Air holes provide magnetic/inductive coupling (L_{ij}) and the filter is connected to the input and output (usually 50 Ω) through capacitive coupling represented by C_{in} and C_{out}.

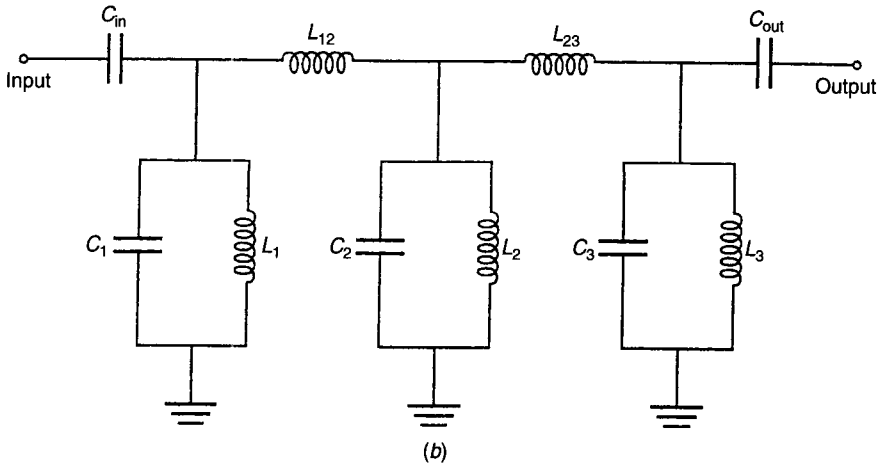
At the center frequency f_0, the length of each resonator is given by

$$L = 0.25\lambda = 0.25\frac{\lambda_0}{\sqrt{\epsilon_{re}}} = \frac{0.25c}{f_0\sqrt{\epsilon_{re}}} \tag{6.36}$$

where λ_0, c, and ϵ_{re} are the free-space wavelength, velocity of light, and effective dielectric constant, respectively and ϵ_{re} is obtained from



(a)



(b)

Figure 6.37 (*a*) High-*K* ceramic back comline bandpass filter. All surfaces are metallized except top surface; A, B, and C are metallized coaxial resonators. Metallized side walls of ceramic block act as outer conductors. (*b*) Lumped-element equivalent circuit for three-resonator filter.

$$\epsilon_{re} = \tfrac{1}{2}(\epsilon_{ree} + \epsilon_{reo}) \tag{6.37}$$

where ϵ_{ree} and ϵ_{reo} are respectively the even- and odd-mode effective dielectric constants of the medium in which coaxial resonators are embedded. Thus, the thickness of the ceramic block, L, determines the frequency of operation.

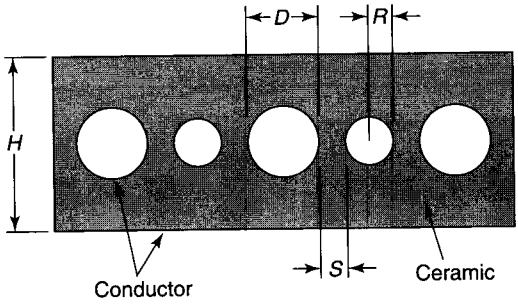


Figure 6.38 Cross-sectional view of three-resonator ceramic block combline.

The coupling coefficient is given by [52]

$$K = \frac{2(\sqrt{\epsilon_{ree}} - \sqrt{\epsilon_{reo}})}{\sqrt{\epsilon_{ree}} + \sqrt{\epsilon_{reo}}} \tag{6.38}$$

Figure 6.38 shows a cross-sectional view of an air hole coupled line structure with dimensions, and the calculated values of the even- and odd-mode effective dielectric constants are plotted in Figure 6.39. The FEM [52] was used to analyze the structure, with $\epsilon_r = 80$, $D = 2.4$ mm, $H = 6$ mm, and $S = 0.8$ mm. Figures 6.40*a* and *b* show the coupling coefficient versus air hole radius and

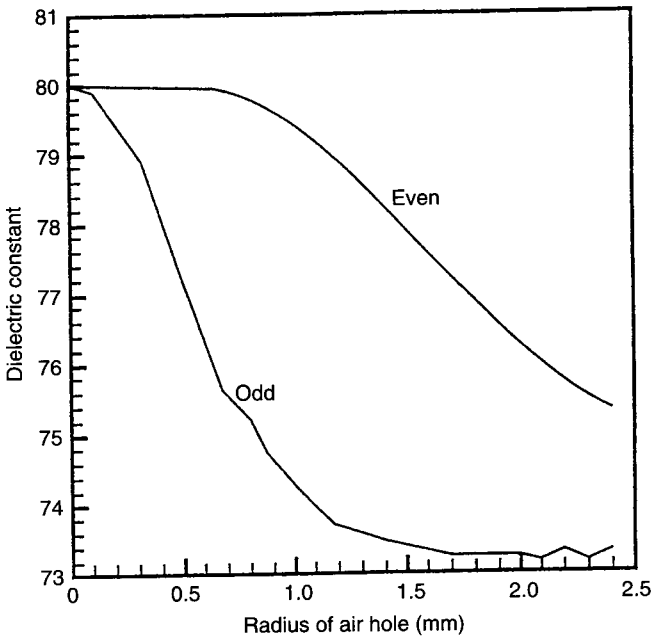


Figure 6.39 Effective dielectric constant for even and odd modes as function of air hole radius: $\epsilon_r = 80$, $D = 2.4$ mm, $H = 6$ mm, and $S = 0.8$ mm [52].

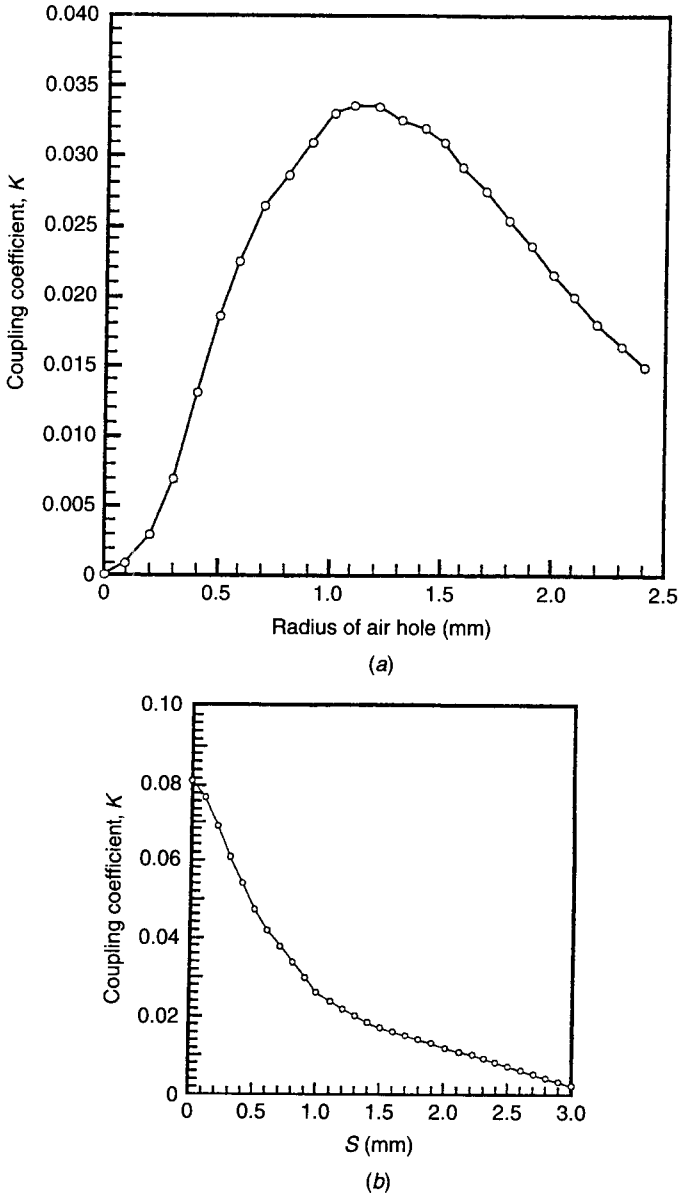


Figure 6.40 (*a*) Coupling coefficient as function of air hole radius: $\epsilon_r = 80$, $D = 2.4$ mm, $H = 6$ mm, and $S = 0.8$ mm. (*b*) Coupling coefficient as function of separation between metallized resonator and air hole: $\epsilon_r = 80$, $D = 2.4$ mm, $H = 6$ mm, and $R = 1.2$ mm. (From [52], *Microwave Journal*, 1994, reprinted with permission.)

separation between the resonator and air hole, respectively. More extensive data for coupling coefficient have also been published by Yao et al. [54, 55].

To illustrate a design example of a bandpass ceramic block filter, the following specifications are chosen:

Center frequency	900 MHz
Response	Chebyshev with 0.05 dB ripple
Bandwidth	30 MHz
Number of resonators	3

This filter can be designed by using normalized coupling coefficients in terms of low-pass prototype element values and design frequencies as given in Eqs. (6.29) and (6.30) The nomalized bandwidth is given as $\mathrm{BW}/f_0 = 30/900 = \frac{1}{30}$. From Matthaei et al. [1], the low-pass prototype element values are $g_0 = g_4 = 1$, $g_1 = g_3 = 0.8794$, and $g_2 = 1.1132$. From Eq. (6.29), the coupling coefficient $K_{12} = K_{23} = 0.0339$. For the structure shown in Fig. 6.38 and using the data in Figure 6.40, the filter design parameters are $\epsilon_r = 80$, $H = 6$ mm, $D = 2.4$, $R = 1.2$ mm, $S = 0.8$ mm, and $L = 9.55$ mm. The input and output impedance values are 50 Ω.

Figure 6.41 shows the measured frequency response of the three-resonator ceramic block filter where the input and output couplings were obtained by experiments. In 50 MHz bandwidth, the measured insertion loss was better than 1.5 dB. The ceramic materials loss tangent ($\tan\delta$) was 2.5×10^{-4} and temperature coefficient was 3 ppm/°C.

6.7.4 Compact Filters

Several techniques to design miniature planar bandpass filters at the L band have been used: meander or folded resonators [56, 57], high-K dielectric mate-

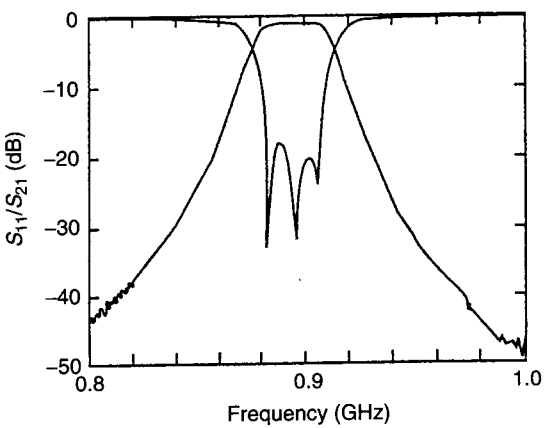


Figure 6.41 Measured frequency response of three-resonator ceramic block combline filter.

rials [58, 59], a multilayer dielectric and metallization circuit configuration [60], and stepped-impedance resonator circuits [61–64]. Pseudo-interdigital [65], hairpin resonator, and open-loop resonator [66] filters do not require any grounding of the resonators. Therefore, these filters are more suitable for low-cost and high-volume production, as a large number of filters can be simultaneously printed on a single substrate board. Their high-temperature superconducting versions are suitable for satellite applications [67]. Pramanick [59] provides a step-by-step design of hairpin resonator filters on high-K materials.

At RF frequencies, printed resonators using thin- and thick-film hybrid technologies have been successfully used to develop miniature and low-cost filters. The Q of microstrip resonators can be increased by 20–50% by employing a multilayer filter structure. To achieve lower insertion loss with narrower bandwidths and low cost, LTCC technology is preferred, a technology that, using high-K material such as BiCaNbO with a dielectric constant of about 60, is capable of producing low-loss (<2 dB), miniature, and narrow-band filters for mobile applications.

6.7.5 Lumped-Element Filters

At RF frequencies and the lower end of the microwave frequency band, filters have been realized using lumped elements (chip/coil inductors and chip capacitors), printed inductors, and discrete chip capacitors and employing multilayer printed-circuit technique such as LTCC or printed-circuit boards. Lumped-element filters can be implemented easily, and using currently available surface-mounted components, one can meet size and cost targets in high-volume production. Due to the low Q of inductors and capacitors, it is not possible to realize narrow-band filters using MIC or MMIC technologies for some wireless applications.

6.8 PRACTICAL CONSIDERATIONS

In the preceding sections, various design parameters of filters were described. The important design considerations are size, weight, cost, effect of finite Q, group delay, temperature effect, power-handling capability, and tuning of the filters.

6.8.1 Size, Weight, and Cost

Applications such as electronic countermeasure (ECM), airborne, satellite, and mobile communication demand small size, light weight, and less expensive filters. Filters constructed using strip line, microstrip, and fin line are limited to relatively wide band applications where the selectivity is not severe. Furthermore, for highly selective applications, these structures exhibit temperature instability and tuning difficulties. The suspended-substrate techniques provide a

higher Q (500–1500) than similar microstrip techniques. This results in lower loss filters with sharper band edges and with an excellent temperature stability. In certain cases, filters are available with temperature stabilities of $\pm 0.1\%$ over $-54°C$ to $+125°C$. Dielectric resonator filters have shown potential and are being developed from ultrahigh frequency (UHF) to millimeter-wave frequencies. Size can also be reduced using a high-dielectric-constant substrate, such as barium tetratitanate ($BaTi_4O_9$) or other compounds of barium.

Up to a couple of gigahertz surface acoustic wave (SAW) filters are commonly used to build channelized receivers that can include several hundred filters. Surface acoustic wave filters are very attractive in terms of size, weight, and cost at lower microwave frequencies.

6.8.2 Finite *Q*

So far no attempt has been made to consider the effect of losses of the filter circuits on their performance. In other words, what happens if the resonators have a finite Q? When the lossy line sections are used, the insertion loss in the passband increases. In the case of the equal-ripple response filter, the losses result in a suppression of the ripples, as shown in Fig. 6.42 for different unloaded Q values of the filter. It is assumed that all the five sections of the filters have the same unloaded Q and resonant frequency.

In practice the bandwidth of a filter is set by the loaded Q of the resonator. The loaded Q, Q_L, of a resonator depends on its losses and the external circuit connected to it and is given by

$$\frac{1}{Q_L} = \frac{1}{Q_u} + \frac{1}{Q_e} \tag{6.39}$$

where Q_u is the unloaded resonator quality factor and Q_e is the external Q, which accounts for the energy coupled into the external circuitry. When $Q_e \ll Q_u$ (or $\text{BW} \gg f_0/Q_u$), the bandwidth of the filter (BW) is almost independent of the unloaded Q values, but as Q_e approaches Q_u, the circuit gets very lossy. Thus unloaded Q sets a lower limit for the bandwidth of the filter. As Q_u approaches Q_e, it has three effects on the performance of filters: It broadens the bandwidth, introduces extra insertion loss, and reduces rejection in the stopband.

The insertion loss at the center of the passband of small-ripple Chebyshev bandpass filters is approximately given by

$$\text{IL (dB)} = \frac{4.343 f_0}{\text{BW}\, Q_u} \sum_{i=1}^{N} g_i \tag{6.40}$$

where BW/f_0 is the equal-ripple frequency bandwidth.

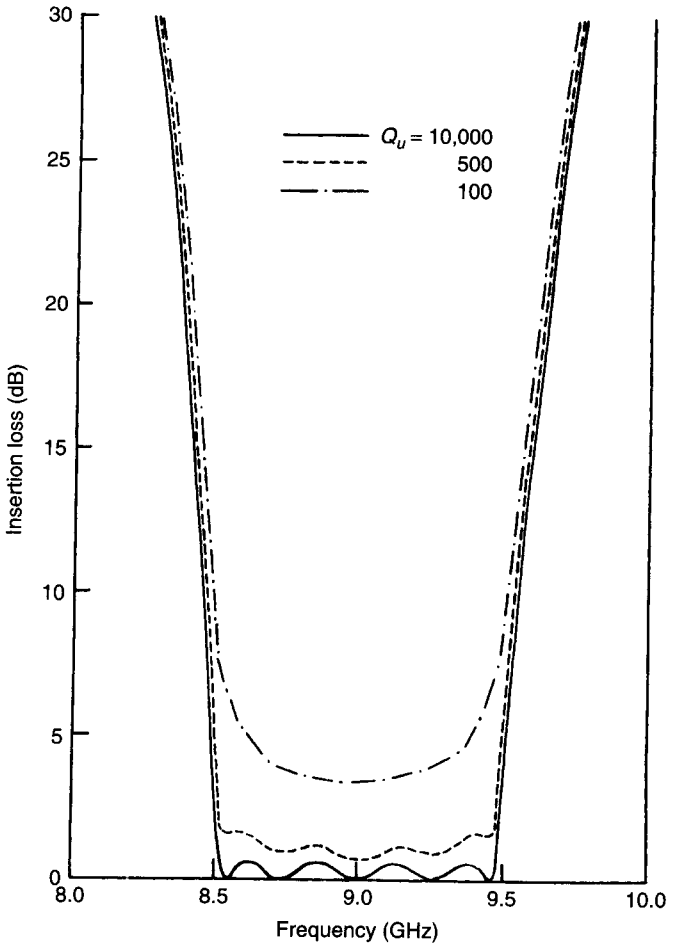


Figure 6.42 Bandpass response of direct-coupled filter for different Q values of filter: f_0 = 9 GHz, BW = 900 MHz, and A_m = 0.5 dB.

6.8.3 Power-Handling Capability

The amount of peak power or average power that a filter can handle without voltage breakdown or excessive heating depends on bandwidth, filter size, filter type, and operating environment and altitude. Printed-circuit filters can handle up to a few hundred watts, whereas for higher power applications, conventional waveguide and coaxial structures have to be used.

The power-handling capabilities of filters can be determined by knowing the power-handling capability of the transmission media [1, 12, 13] and filter structure geometries. Since it is not in the scope of this book to go into detail, readers are referred to the references. The larger gaps in the filter structure permit higher filter power rating.

For high-power applications (>10 kW) pressurized force-cooled waveguide filters are often the only choice. Coaxial filters are next, while strip-line, microstrip, lumped-element, and SAW filters are usually used for low-power applications. However, printed-circuit filters are capable of handling pulse powers on the order of a few kilowatts peak and up to 50 W average.

6.8.4 Temperature Effects

When the temperature of the environment surrounding the filter changes, the physical dimensions of the structure also change. Therefore, the electrical characteristics of the filter are modified. Since most solids expand linearly with an increase in temperature, a physical dimension l will expand by Δl when the temperature increases by ΔT, that is,

$$\frac{\Delta l}{l} = \alpha \, \Delta T \tag{6.41}$$

where α is the linear expansion coefficient. For commonly used materials its values are given in Table 6.10. For bandpass filters using $\frac{1}{2}\lambda$ resonators,

$$\frac{\Delta f}{f} = -\alpha \, \Delta T \tag{6.42}$$

Thus, the temperature coefficient of TEM resonators made of copper is -17 ppm/°C, while Invar resonators has -0.7 ppm/°C. The second effect of temperature on filters is due to the temperature dependence of the relative dielectric constant ϵ_r of dielectric materials. As a first approximation, the change is linear with temperature, that is,

$$\frac{\Delta \epsilon_r}{\epsilon_r} = \alpha_\epsilon \, \Delta T \tag{6.43}$$

where α_ϵ is the linear temperature coefficient of the dielectric constant. If the resonant frequency of resonators is inversely proportional to the square root of

Table 6.10 Linear Expansion Coefficients for Metals

Metal	α (per °C) $\times 10^{-6}$	Dielectric	α (per °C) $\times 10^{-6}$
Aluminum	23	Teflon	90
Brass	19	Glass	9
Copper	17	Alumina	7
Gold	15	RT/Duroid	6
Steel	11	Glass (Pyrex)	3
Invar	0.7	Fused quartz	0.6

Table 6.11 Properties of Dielectric Resonators

Material	Q	ϵ_r	α (ppm/°C)	α_ϵ (ppm/°C)	Company
(Zr, Sn)TiO_4 type C	15,000 (4 GHz)	37.3 ± 0.5	6.5	-13 ± 2	Murata
Ba(Zr, Zn, Ta)O_3 type C	10,000 (10 GHz)	28.6 ± 0.5	10.2	-20.4 ± 4	Murata
Zr/Sn titanate D-8515 type	10,000 (4 GHz)	36.0 ± 0.5	5.6	-6.9	Trans Tech
Barium tetra-titante D-8512 type	10,000 (4 GHz)	38.6 ± 0.6	9.5	-10.4	Trans Tech
(Zr, Zn)TiO_4 E-2036 type	4000 (10 GHz)	37.0 ± 0.4	5	-10 ± 1	Thomson-CSF

ϵ_r, then net change in resonant frequency is given by

$$\frac{\Delta f}{f} = -\alpha\,\Delta T - 0.5\alpha_\epsilon\,\Delta T \tag{6.44}$$

The previous equation suggests that by properly selecting α and α_ϵ, Δf can be made zero. Table 6.11 gives α and α_ϵ for commonly used materials in dielectric resonators. It is evident from this table that the temperature stability of most dielectric resonators is excellent. Temperature coefficients of other filter components are given in Table 6.12.

Consider an example of combline filters whose required performance over operating temperatures -50°C to 125°C is as follows:

f_0	2 GHz
3-dB points	1.8 GHz
	2.2 GHz
40-dB points	1.5 GHz
	2.5 GHz

Table 6.12 Temperature Coefficients of Filter Elements

Description	Center Frequency (GHz)	Temperature Coefficients (ppm/°C)
Standard lumped elements	>0.1	75
Stable lumped elements	0.1–16	10
Microstrip resonators (e.g., combline, interdigital)	0.5–18	25

If 25°C is the room temperature,

$$\Delta T \text{ (cold)} = -75°\text{C}$$
$$\Delta T \text{ (hot)} = 100°\text{C}$$
$$\Delta f \text{ (cold)} = -75 \times 2000 \times 25 \times 10^{-6} = -3.75 \text{ MHz}$$
$$\Delta f \text{ (hot)} = 100 \times 2000 \times 25 \times 10^{-6} = 5 \text{ MHz}$$

Specifications adjusted for temperature variation are as follows:

	$f_0 = 2$ GHz
3-dB points	$1.8 + (-0.00375) = 1.79625$ GHz
	$2.2 + (0.005) = 2.25$ GHz
40-dB points	$1.5 + (0.005) = 1.505$ GHz
	$2.5 + (-0.00375) = 2.49625$ GHz

In other words, the filter must be designed with about 9 MHz larger 3-dB bandwidth and about 9 MHz shorter 40-dB bandwidth to meet the specifications over the required temperature range.

6.8.5 Group Delay

The group delay of a filter is dependent on both the selected prototype and the number of sections used. Band-center group delay and group delay variation both increase with increasing number of sections. Many applications require constant group delay over the desired bandwidth.

The design of constant-group-delay filters has been described in the literature [1, 2, 68]. A general approach to design a constant group delay in the passband is to specify a function that describes the desired phase response, such as [68]

$$\psi_T(\omega) = A\omega\left[1 + B\left(\frac{\omega}{\omega_c}\right)^p\right] \tag{6.45}$$

and then to force the filter response to assume values that satisfy the above equation at a number of discrete frequencies while maintaining acceptable amplitude properties. Then the group delay τ_D is calculated using (6.6b).

6.8.6 Mechanical Tuning of Filters

The filters we have discussed are synchronously tuned. For ladder network filters, this means that all of the resonator frequencies would be at the same frequency (f_0) in the absence of coupling to the other resonators. The definition of synchronous is more complex for asymmetric bridge-coupled filters but the idea

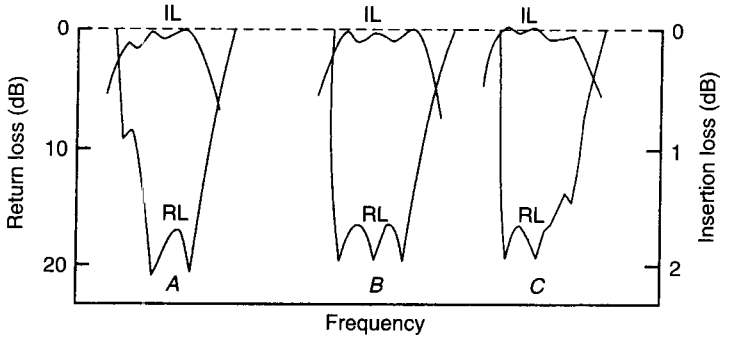


Figure 6.43 Mechanical tuning of a filter: *A*, pole tuned low; *B*, perfectly tuned; *C*, pole turned high.

is the same; the resonators comprise a coupled network resonator matched in the passbands.

On first construction, nearly all new filter designs must be tuned to account for design approximations and manufacturing tolerances. Tuning of high-performance filters (especially as the bandwidth narrows) is a routine part of the manufacturing process.

Tuning involves adjusting the resonator frequencies and sometimes the coupling values between the resonators. How this is accomplished depends on the filter construction. Waveguide and coaxial filters often incorporate tuning screws into the filter assembly. Microwave IC filters are often tuned by attaching dielectric chips to the resonators. Lumped filters may have provision for capacitance adjustment. Manual tuning of complex filters is an act that can be quite daunting to the uninitiated.

Filter tuning is frequently done using a test set that continuously displays return loss and insertion loss over frequency, as demonstrated by the response of the three-pole bandpass filter shown in Fig. 6.43. The resonator frequencies are seen as return loss poles with associated insertion loss minimums.

When a synchronous filter is properly tuned, the resonators form a coupled system such that the association between each return loss pole and each resonator frequency is lost. However, if a resonator frequency is high or low, an associated return loss pole will emerge from the coupled response, moving high or low in frequency with the resonator frequency. Describing how to tune the coupling values from the measured insertion loss and return loss is beyond the scope of this section, but it is done by recognizing characteristic patterns of under- or overcoupling.

6.9 ELECTRICALLY TUNED FILTERS

Electrically tunable filters have many applications in communication, EW, and ESM systems. System requirements necessitate that the filters exhibit high-Q

and broad tuning bandwidths with minimum degradation in passband performance. Yttrium–iron–garnet (YIG) spheres [69] and varactor diodes [70–79] are commonly used as tuning devices. Varactor Q factors are considerably lower than those of YIG spheres, and thus varactor-tuned filters are not generally used above 10 GHz. On the other hand, the tuning speed of YIG filters is severely limited by magnetic hysteresis effects, whereas in the case of varactor-tuned filters, it is limited by the time constant of the varactor bias circuitry.

A new design for a tunable bandpass filter having elliptic-type transfer function was developed using bridged networks [78]. The filter consists of microstrip sections, fixed metal–insulator–metal (MIM) capacitors, and tunable capacitors to get a compact size. The filter's performance was optimized in terms of insertion loss, bandpass bandwidth, and bandstop rejection. The tunable capacitors may be realized by high-Q varactors or Schottky diodes whose junction capacitance is varied by changing the junction voltage. These devices are small in size with tuning speed (limited by the time constant of the variable bias circuitry) on the order of 1 ns.

The schematic diagram of the filter is shown in Fig. 6.44. The microstrip transmission lines are designated by T_1, T_2, and T_3. The fixed MIM capacitor is given as C_b, and C_v is the variable capacitor. The ground connections are realized using via holes. The via hole model used consists of a series combination of R and L, where $R = 0.03\ \Omega$ and $L = 0.01$ nH. The filter response includes losses in the microstrip lines, MIM capacitors, and varactors. Measurement-

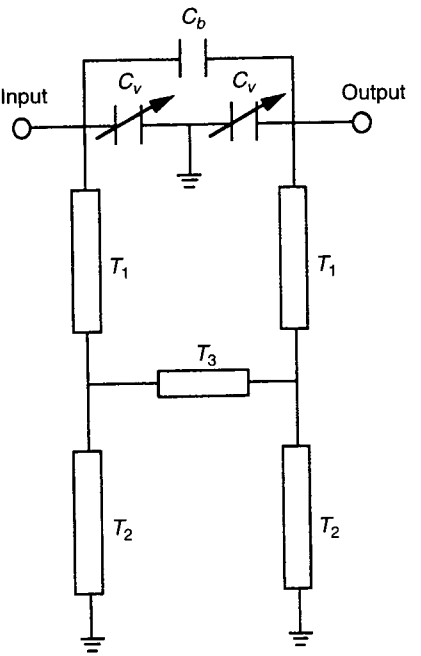
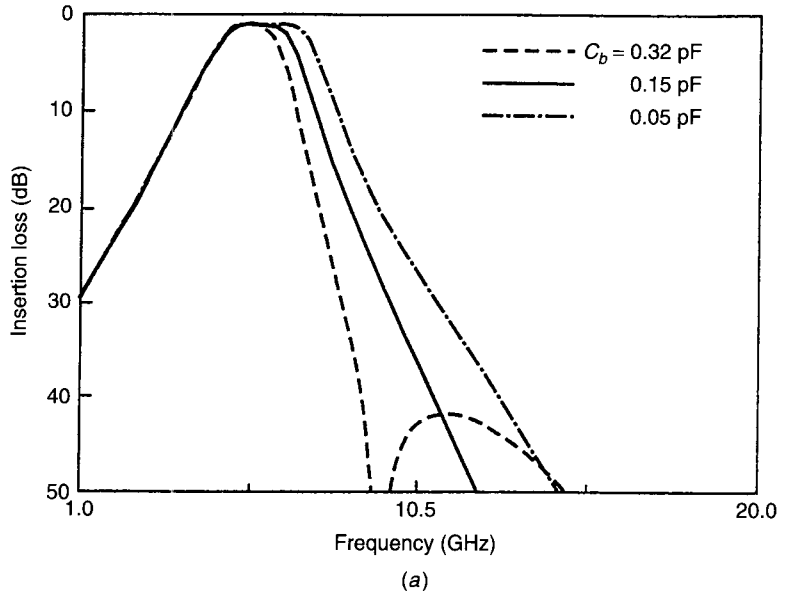


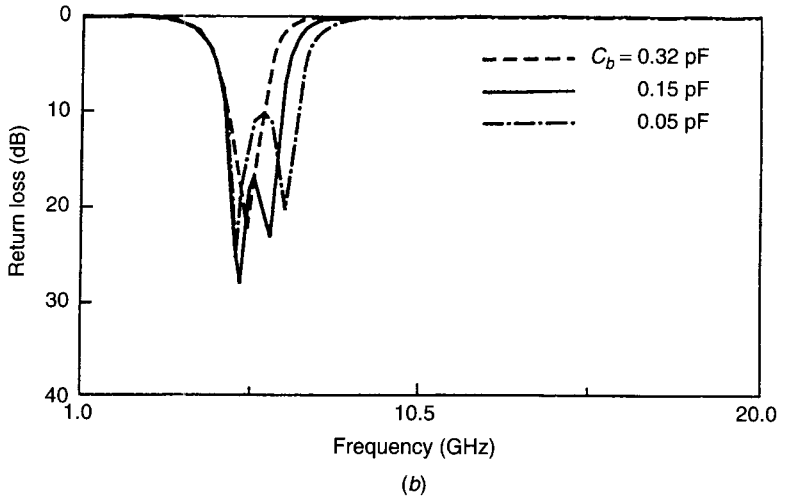
Figure 6.44 Schematic of elliptic-type filter.

based models for MIM capacitors [80] and a fixed value of Q of 200 for varactors were used in the design.

Substrate parameters used are $\epsilon_r = 12.9$, $h = 200$ μm, $t = 4.5$ μm, gold bulk conductivity, and $\tan\delta = 0.0005$. The microstrip sections have $W_1 = 50$, $L_1 = 200$, $W_2 = 50$, $L_2 = 580$, $W_3 = 30$, and $L_3 = 580$ μm.



(*a*)



(*b*)

Figure 6.45 Simulated response of variable-bandwidth elliptic-type filter: (*a*) insertion loss; (*b*) return loss.

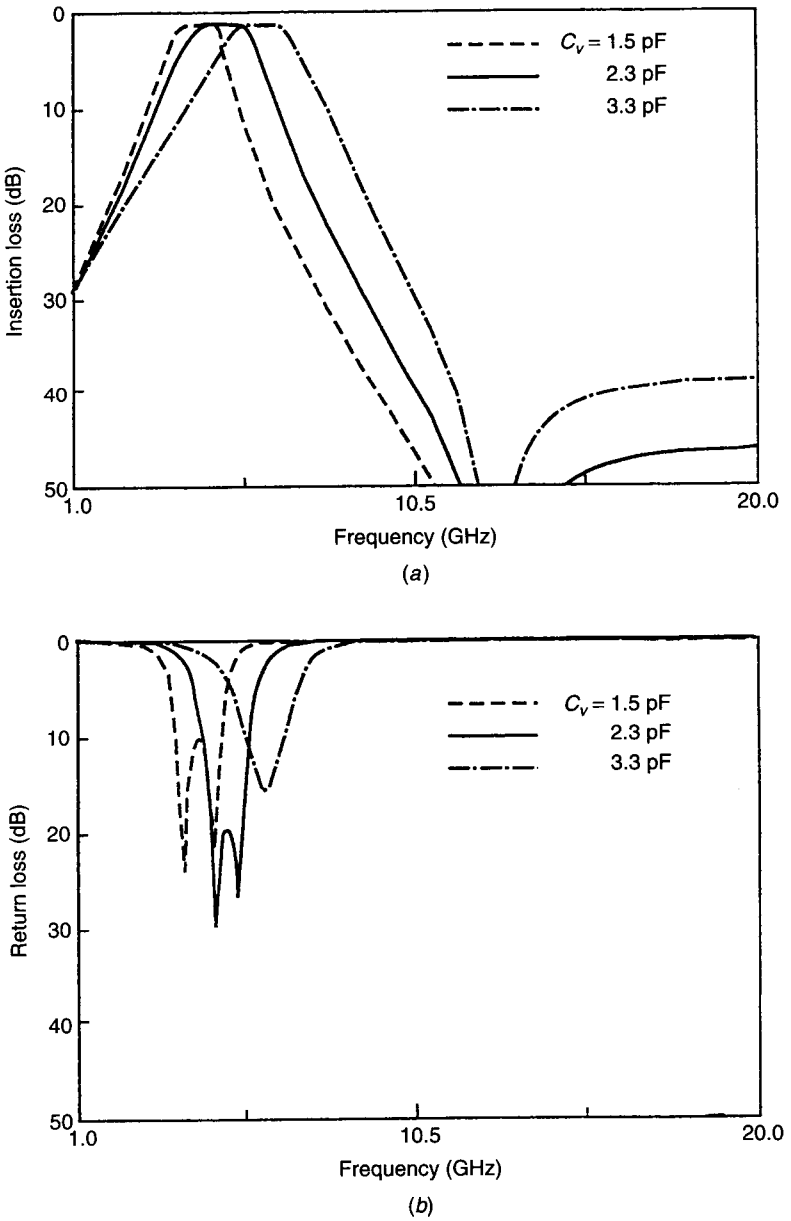


Figure 6.46 Simulated response of variable-center-frequency elliptic-type filter: (*a*) insertion loss; (*b*) return loss.

Simulated Results

Variable-Bandwidth Filter. In this case the shunt capacitance $C_v = 1.7$ pF is fixed. For various values of C_b, the filter response is shown in Fig. 6.45. When the value of C_b is varied from 0.05 to 0.32 pF, the bandwidth of the filter is increased from 12 to 30% and the insertion loss in the passband remains constant at 1.2 dB. The rejection was better than 50 dB in the stopband up to 20 GHz. Elliptic-type performance of this filter is also evident from Fig. 6.45*a*. When the value of the bridge capacitance C_b is lowered, the selectivity of the passband response at the higher frequency end degrades. Over the variable bandwidth, the return loss is better than 10 dB. In this configuration, when the Q of the varactor is lowered, it hardly changes the performance of the filter.

Variable-Center-Frequency Filter. In this case the bridge capacitance C_b is fixed at 0.15 pF and the varactor capacitance C_v is varied. For three values of C_v, the filter response is shown in Fig. 6.46. When the value of C_v is changed from 1.5 to 3.3 pF, the passband response is varied and the center frequency is shifted from 4.3 to 6.2 GHz while the passband bandwidth (in gigahertz) remains almost constant. Over this tunable frequency range the worst-case insertion loss, rejection, and return loss were better than 1.2, 38, and 10 dB, respectively. In this configuration, when the Q of varactors is lowered from 200 to 100, the insertion loss is increased by less than 0.2 dB.

REFERENCES

1. Matthaei, G. L., L. Young, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks and Coupling Structures*, McGraw-Hill, New York, 1964.
2. Rhodes, J. D., *Theory of Electrical Filters*, Wiley, New York, 1976.
3. Saal, R., *The Design of Filters using the Catalogue of Normalized Low-Pass Filters* (in German), Telefunken (GMBH), Backnang, W. Germany, 1961.
4. Skwirzynski, J. K., *Design Theory and Data for Electrical Filters*, Van Nostrand, London, 1965.
5. Zverev, A. I., *Handbook of Filter Synthesis*, Wiley, New York, 1967.
6. Milligan, T., "Nomographs Aid the Filter Designer," *Microwaves RF*, Vol. 24, Oct. 1985, pp. 103–107.
7. Howe, H., Jr., *Stripline Circuit Design*, Artech House, Dedham, MA, 1974.
8. Alseyab, S. A., "A Novel Class of Generalized Chebyshev Low-Pass Prototype for Suspended Substrate Stripline Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Sept. 1982, pp. 1341–1347.
9. Mobbs, C. I., and J. D. Rhodes, "A Generalized Chebyshev Suspended Substrate Stripline Bandpass Filter," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-31, May 1983, pp. 397–402.
10. Dishal, M., "Alignment and Adjustment of Synchronously Tuned Multiple-Resonant-Circuit Filters," *Proc. IRE*, Vol. 39, Nov. 1951, pp. 1448–1455.

11. Dishal, M., "A Simple Design Procedure for Small Percentage Bandwidth Round-Rod Interdigital Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-13, Sept. 1965, pp. 696–698.
12. Wong, J. S., "Microstrip Tapped-Line Filter Design," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, Jan. 1979, pp. 44–50.
13. Wenzel, R. J., "Lecture Notes—Microwave Circuit Design," UCLA Extension Course, University of California, Los Angeles.
14. Itoh, T. (Ed.), *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*, Wiley, New York, 1989.
15. Sorrentino, R. (Ed.), *Numerical Methods for Passive Microwave and Millimeter-Wave Structures*, Wiley, New York, 1989.
16. Special Issue on Engineering Applications of Electromagnetic Field Solvers, *Int. J. Microwave Millimeter-Wave Computer-Aided Eng.*, Vol. 5, Sept. 1995.
17. Special Issue on Automated Circuit Design Using Electromagnetic Simulators, *IEEE Trans. Microwave Theory Tech.*, Vol. 45, Part II, May 1997.
18. Conrad, A., and J. Browne, "EM Tools Enhance Simulation Accuracy," *Microwave RF*, Vol. 36, Nov. 1997, pp. 133–136.
19. "MIC Simulation Column," *Int. J. Microwave Millimeter Wave Computer-Aided Eng.*, Vol. 1-5, 1991–1995.
20. Swanson, D. G., Jr., "Using a Microstrip Bandpass Filter to Compare Different Circuit Analysis Techniques," *Int. J. Microwave Millimeter Wave Computer-Aided Eng.*, Vol. 5, Jan. 1995, pp. 4–12.
21. Levy, R., and S. B. Cohn, "A Brief History of Microwave Filter Research, Design, and Development," *IEEE Trans. Microwave Theory Tech. (Special Centennial Issue)*, Vol. MTT-32, Sept. 1984, pp. 1055–1067.
22. Gupta, K. C., R. Garg, I. J. Bahl, and P. Bhartia, *Microstrip Lines and Slotlines*, 2nd ed., Artech House, Norwood, MA, 1979.
23. Rooney, J. P., and L. M. Underkofler, "Printed Circuit Integration of MW Filters," *Microwave J.*, Vol. 21, Sept. 1978, pp. 68–73.
24. Ozaki, H., and J. Ishii, "Synthesis of a Class of Stripline Filters," *IRE Trans. Circuit Theory*, Vol. CT-5, June 1958, pp. 104–109.
25. Davis, W. A., *Microwave Semiconductor Circuit Design*, Van Nostrand-Reinhold, New York, 1984, Chap. 3.
26. Richards, P. I., "Resistor-Transmission-Line Circuits," *Proc. IRE*, Vol. 36, Feb. 1948, pp. 217–220.
27. Cohn, S. B., "Parallel-Coupled Transmission-Line Resonator Filters," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-6, Apr. 1958, pp. 223–231.
28. Makimoto, M., and S. Yamashita, "Bandpass Filters Using Parallel Coupled Stripline Stepped Impedance Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, Dec. 1980, pp. 1413–1417.
29. Minnis, B. J., "Printed Circuit Coupled-Line Filters for Bandwidths Up to and Greater Than an Octave," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-29, Mar. 1981, pp. 215–222.
30. Wenzel, R. J., and W. G. Erlinger, "Problems in Microstrip Filter Design," *IEEE Int. Microwave Symp. Dig.*, 1981, pp. 203–205.

31. Ho, C. Y., and J. H. Weidman, "Improved Design of Parallel Coupled Line Filters with Tapped Input/Output," *Microwave J.*, Vol. 26, Oct. 1983, pp. 127–130.
32. Bhartia, P., and I. J. Bahl, *Millimeter Wave Engineering and Applications*, Wiley, New York, 1984.
33. Matthaei, G. L., "Interdigital Band-Pass Filters," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-10, Nov. 1962, pp. 479–491.
34. Wenzel, R. J., "Exact Theory of Interdigital Bandpass Filters and Related Coupled Structures," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-13, Sept. 1965, pp. 558–575.
35. Matthaei, G. L., "Comb-Line Band-Pass Filters of Narrow or Moderate Bandwidth," *Microwave J.*, Vol. 6, Aug. 1963, pp. 82–91.
36. Wenzel, R. J., "Synthesis of Combline and Capacitively Loaded Interdigital Bandpass Filters of Arbitrary Bandwidth," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-19, Aug. 1971, pp. 678–686.
37. Vincze, A. D., "Practical Design Approach to Microstrip Combline-Type Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, Dec. 1974, pp. 1171–1181.
38. Cristal, E. G., and S. Frankel, "Hairpin Line/Half-Wave Parallel-Coupled-Line Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-20, Nov. 1972, pp. 719–728.
39. Gysel, U. H., "New Theory and Design for Hairpin-Line Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, May 1974, pp. 523–531.
40. Hong, J.-S., and M. J. Lancaster, *Microstrip Filters for RF/Microwave Applications*, Wiley, New York, 2001.
41. Plourde, J. K., and D. F. Linn, "Microwave Dielectric Resonator Filters Using $Ba_2Ti_9O_{20}$ Ceramics," *IEEE Int. Microwave Symp. Dig.*, 1977, pp. 290–293.
42. Ren, C. L., "Waveguide Bandstop Filter Utilizing $Ba_2Ti_9O_{20}$ Resonators," *IEEE Int. Microwave Symp. Dig.*, 1978, pp. 227–229.
43. Cohn, S. B., "Microwave Bandpass Filters Containing High-Q Dielectric Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-16, Apr. 1968, pp. 218–227.
44. Iveland, T. D., "Dielectric Resonator Filters for Application in Microwave Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-19, July 1971, pp. 643–652.
45. Mekerta, S., and P. Guillon, "Dielectric Resonator Bandstop Filters," *IEEE Int. Microwave Symp. Dig.*, 1981, pp. 170–172.
46. Atia, A. E., and R. R. Bonetti, "Generalized Dielectric Resonator Filters," *Comsat Tech. Rev.*, Vol. 11, Fall 1981, pp. 321–343.
47. Fiedziuszko, S. J., "Dual Mode Dielectric Resonator Loaded Cavity Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Sept. 1982, pp. 1311–1316.
48. Kajfez, D., and P. Guillon (Eds.), *Dielectric Resonators*, Artech House, Dedham, MA, 1986.
49. Guillon, P., B. Byzery, and M. Chaubet, "Coupling Parameters Between a Dielectric Resonator and a Microstrip Line," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-33, Mar. 1985, pp. 222–226.

50. Fukasawa, A., "Analysis and Composition of a New Microwave Filter Configuration with Inhomogeneous Dielectric Medium," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Sept. 1982, pp. 1367–1375.

51. Levy, R., "Simplified Analysis of Inhomogeneous Dielectric Block Combline Filters," *IEEE Int. Microwave Symp. Dig.*, 1990, pp. 135–138.

52. You, C.-C., C.-L., Huang, and C. C. Wei, "Single-Block Ceramic Microwave Bandpass Filters," *Microwave J.*, Vol. 37, Nov. 1994, pp. 24–35.

53. Hano, K., H. Kohriyama, and K.-I. Sawamoto, "A Direct Coupled $\lambda/4$-Coaxial Resonator Bandpass Filter for Land Mobile Communications," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-34, Sept. 1986, pp. 972–976.

54. Yao, H.-W., K. A. Zaki, A. E. Atia, and R. Hershtig, "Full Wave Modeling of Conducting Posts in Rectangular Waveguide and Its Applications to Slot Coupled Combline Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-43, Dec. 1995, pp. 2824–2829.

55. Yao, H. W., C. Wang, and K. A. Zaki, "Quarter Wavelength Combline Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. 44, Dec. 1996, pp. 2673–2679.

56. Hong, J. S., and M. J. Lancaster, "Development of New Microstrip Pseudo-Interdigital Bandpass Filters," *IEEE Microwave Guided Wave Lett.*, Vol. 5, Aug. 1995, pp. 261–263.

57. Korber, M., Jr., "New Microstrip Filter Topologies," *Microwave J.*, Vol. 40, July 1997, pp. 138–144.

58. Nishikawa, T., et al., "800 MHz Band Face Bonding Filter Using Dielectric B.D.L.S.," *IEEE Int. Microwave Symp. Dig.*, 1986, pp. 403–406.

59. Pramanick, P., "Compact 900-MHz Hairpin-Line Filters Using High Dielectric Constant Microstrip Line," *Int. J. Microwave Millimeter-Wave Computer-Aided Eng.*, Vol. 4, July 1994, pp. 272–281.

60. Cho, C., and K. C. Gupta, "A Methodology for Design of Multilayer Microwave Circuits," *Int. RF Microwave Computed Aided Eng.*, Vol. 8, Nov. 1998, pp. 455–473.

61. Makimoto, M., and S. Yamashita, "Bandpass Filters Using Parallel Coupled Stripline Stepped Impedance Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, Dec. 1980, pp. 1413–1417.

62. Makimoto, M., and S. Yamashita, "Compact Bandpass Filters Using Stepped Impedance resonators," *Proc. IEEE*, Vol. 67, Jan. 1979, pp. 16–19.

63. Rawat, B., R. Miller, and B. E. Pontius, "Bandpass Filters for Mobile Communications," *Microwave J.*, Vol. 27, Sept. 1984, pp. 146–152.

64. Ishizaki, T., and T. Uwano, "A Stepped Impedance Comb-Line Filter Fabricated by Using Ceramic Lamination Technique," *IEEE Int. Microwave Symp. Dig.*, 1994, pp. 617–620.

65. Sagawa, M., K. Takahashi, and M. Makimoto, "Miniaturized Hairpin Resonator Filters and Their Application to Receiver Front End MMIC's," *IEEE Trans. Microwave Theory Tech.*, Vol. 37, Dec. 1989, pp. 1991–1997.

66. Yu, C. C., and K. Chang, "Novel Compact Elliptic-Function Narrow-Band Bandpass Filters Using Microstrip Open-Loop Resonators with Coupled and Crossing Lines," *IEEE Trans. Microwave Theory Tech.*, Vol. 46, July 1998, pp. 952–958.

67. Matthaei, G. L., et al., "Hairpin-Comb Filters for HTS and Other Narrow-Band Applications," *IEEE Trans. Microwave Theory Tech.*, Vol. 45, Aug. 1997, pp. 1226–1231.

68. Malherbe, J. A. G., *Microwave Transmission Line Filters*, Artech House, Dedham, MA, 1979.

69. Keane, W. J., "Narrow-Band YIG Filters," *Microwaves*, Vol. 17, Sept. 1978, pp. 52–56.

70. Hunter, I. C., and J. D. Rhodes, "Varactor Tuned Microwave Bandpass Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Sept. 1982, pp. 1354–1360.

71. Hunter, I. C., and J. D. Rhodes, "Varactor Tuned Microwave Bandstop Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Sept. 1982, pp. 1361–1367.

72. Toyoda, S., "Quarter-Wavelength Coupled Variable Bandstop and Bandpass Filters Using Varactor Diodes," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Sept. 1982, pp. 1387–1389.

73. Makimoto, M., and M. Sagawa, "Varactor Tuned Bandpass Filters Using Microstrip-Line Ring Resonators," *IEEE Int. Microwave Symp. Dig.*, 1986, pp. 411–414.

74. Auffray, D., and J. L. Lacombe, "Electronically Tunable Bandstop Filter," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1988, pp. 439–442.

75. Martin, T. S., F. Wang, and K. Chang, "Theoretical and Experimental Investigation of Novel Varactor-Tuned Switchable Microstrip Ring-Resonator Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. 36, Dec. 1988, pp. 1733–1739.

76. Shu, Y.-H., J. A. Navarro, and K. Chang, "Electronically Switchable and Tunable Coplanar Waveguide-Slotline Band-Pass Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. 39, Mar. 1991, pp. 548–554.

77. Chang, C.-Y., and T. Itoh, "Microwave Active Filters Based on Coupled Negative Resistance Method," *IEEE Trans. Microwave Theory Tech.*, Vol. 38, Dec. 1990, pp. 1879–1884.

78. Bahl, I. J., and E. L. Griffin, "Low Loss Electronically Tunable Bandpass Filters on the GaAs Substrate," *Microwave Opt. Technol. Lett.*, Vol. 5, June 1992, pp. 328–330.

79. Kamitsuna, H., "Monolithically Integrated High-Gain and High-Sensitive Photoreceivers with Tunable Filtering Functions for Subcarrier Multiplexed Optical/Microwave Systems," *IEEE Trans. Microwave Theory Tech.*, Vol. 43, Sept. 1995, pp. 2351–2356.

80. Sadhir, V. K., I. J. Bahl, and D. A. Willems, "CAD Compatible Accurate Models of Microwave Passive Lumped Elements for MMIC Applications," *Int. J. Microwave Millimeter-Wave CAE*, Vol. 4, Apr. 1994, pp. 148–162.

PROBLEMS

6.1 Calculate low-pass prototype element values (g's) for maximally flat and Chebyshev response (0.1 dB ripple) when the number of sections is 3. If the edge frequency is 1 GHz, determine LC values for low- and high-pass filters.

6.2 Determine LC values for bandpass and bandstop filters when the center frequency is 2 GHz in Problem 6.1.

6.3 Design a low-pass strip-line filter having a ripple of 0.1 dB and stopband attenuation of 30 dB at 3 GHz. The filter may be realized using high- and low-impedance sections. The input and output impedances are 50 Ω.

6.4 Develop a low-pass filter using lumped elements with the specifications given in Problem 6.1.

6.5 Develop design equations for a parallel-coupled microstrip-line bandpass filter. Design a six-section bandpass filter at 4 GHz. For the microstrip, the parameters to be used are $\epsilon_r = 3.8$ and $h = 0.5$ mm. The input and output impedances are 50 Ω.

6.6 Develop design equations for an end-coupled microstrip-line bandpass filter. Describe limitations of this configuration with respect to the parallel-coupled microstrip-line filter. Suggest how these limitations can be overcome.

6.7 Suppose that a filter is desired with a 0.5-dB-ripple Chebyshev passband from 3 to 3.2 GHz and that 30 dB attenuation is required at 2.5 and 3.5 GHz. Determine the physical dimensions for an end-coupled strip-line filter. The input and output impedances are 50 Ω.

6.8 Design an eight-resonator, 0.1-dB-Chebyshev-ripple, and 25% bandwidth tapped-line interdigital bandpass filter on a 5880 RT/Duroid substrate using Figs. 6.20 and 6.22. The center frequency is 1 GHz. The input and output impedances are 50 Ω.

6.9 Design an LC high-pass filter with an f_c of 4 GHz and a minimum attenuation of 30 dB at 3 GHz by assuming that 0.1 dB passband ripple is tolerable. Calculate and compare performance using ideal LC components and monolithic LC components. The monolithic substrate is 200-μm-thick GaAs ($\epsilon_r = 12.9$) and MIM capacitance is 300 pF/mm^2. The input and output impedances are 50 Ω.

7

ACTIVE DEVICES

Robert J. Trew

7.1 INTRODUCTION

The development of microwave solid state circuits is directly dependent upon the availability of suitable active devices. Active devices are required as gain blocks in circuit applications and to allow a circuit to increase the RF energy at a desired frequency. For microwave circuit applications the microwave bipolar transistor, GaAs metal–semiconductor field-effect transistor (MESFET), and GaAs-based high-electron-mobility transistor (HEMT) are the most important and commonly used active elements. The basic operational principles of these devices are presented in this chapter. The emphasis upon basic physical phenomena and the development of physically based equivalent circuits should provide the information required as background for device design as well as analysis.

The chapter begins with a presentation of the basic equations that describe the operation of all semiconductor devices. Material parameters and their effect upon the operation of the devices are discussed. This is followed by sections discussing the two types of transistors.

7.2 BASIC SEMICONDUCTOR DEVICE EQUATIONS

The set of equations generally referred to as the *semiconductor equations* serves as the starting point for most device investigations. These equations consist of the current density equations, the continuity equations, Poisson's equation, and Faraday's law. In general, these equations must be solved simultaneously with

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

Faraday's law, and the appropriate boundary conditions in order to obtain an accurate representation of the device operation. For devices such as bipolar transistors where both electrons and holes are important to the operation of the device, two sets of equations are required, one for each type of charge carrier. These equations form the drift-diffusion approximation that is applicable to conditions where the mobile charge carriers are in thermal equilibrium with the crystal lattice. These conditions are generally valid when the device dimensions are large compared to the wavelength of the operation frequency or when the RF period is long compared to the charge carrier relaxation time. When these conditions are not applicable, nonequilibrium (hot-electron) effects can be significant and the drift-diffusion approximation is, in general, not valid. Under these conditions additional terms must be added to the semiconductor equations.

The current density equations for electrons and holes are given in Eqs. (7.1) and (7.2), respectively:

$$\mathbf{J}_n = q\mu_n n\mathbf{E} + qD_n \nabla n \tag{7.1}$$

$$\mathbf{J}_p = q\mu_p p\mathbf{E} - qD_p \nabla_p \tag{7.2}$$

The p and n are charge densities, q is the electronic charge, and μ_p and D_p are material transport parameters. The first term in these equations indicates that the mobile carriers drift in response to an applied electric field, whereas the second term indicates that a diffusion current also flows due to a gradient in the mobile charge density. These equations, as written, apply to relatively low electric field regions where linear charge transport is dominant. In this region the charge carrier velocity is linearly dependent upon the magnitude of the electric field, as shown in the expression

$$\mathbf{v} = \mu\mathbf{E} \tag{7.3}$$

The proportionality constant relating the velocity and electric field is called the mobility (units of $cm^2/V\text{-}s$) and is a characteristic of the particular semiconductor. In the linear region the drift terms can be written in terms of the semiconductor conductivity, where

$$\sigma = \sigma_n + \sigma_p = q(\mu_n n + \mu_p p) \tag{7.4}$$

The total semiconductor current density is composed of both the electron and hole current densities and is given by the expression

$$\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p \tag{7.5}$$

For nondegenerate semiconductors the diffusion coefficient (units of cm^2/s) can be determined from the mobility with use of the Einstein relation

$$D = \mu \frac{kT}{q} \tag{7.6}$$

Equations (7.1)–(7.6) apply to ohmic operation where the carrier mobility and diffusion coefficient are electric field independent. Many devices, however, are operated under high-field conditions. At high electric fields the carrier velocity saturates and the diffusion coefficient becomes field dependent. The current density equations must be modified. For high-field operation the equations can be modified by substituting the carrier velocity into the drift term and replacing the diffusion constant with a field-dependent equivalent according to the expressions

$$\mu E \rightarrow v \tag{7.7}$$

$$D \rightarrow D(E) \tag{7.8}$$

This substitution, of course, requires knowledge of the velocity–field and diffusion coefficient–field characteristics.

The continuity equations for electrons and holes are given in (7.9) and (7.10), respectively, as

$$\frac{\partial n}{\partial t} = G_n - U + \frac{1}{q} \nabla \cdot \mathbf{J}_n \tag{7.9}$$

$$\frac{\partial p}{\partial t} = G_p - U - \frac{1}{q} \nabla \cdot \mathbf{J}_p \tag{7.10}$$

These equations simply state that the time rate of change of charge within a volume is equal to the rate of flow of charge out of the volume. Charge within the volume may be generated by some mechanism, such as optical excitation or avalanche ionization, extinguished by recombination, or forced out of the volume by divergence. The recombination rate, represented by the U term, is given on a net basis and represents the decay of free charge in excess of thermal equilibrium within the semiconductor. It can be represented, in simple cases, as

$$U = \frac{\delta n}{\tau_n} \quad \text{or} \quad \frac{\delta p}{\tau_p} \tag{7.11}$$

where τ_n and τ_p are the carrier lifetimes and δn and δp are the mobile charge densities in excess of thermal equilibrium. They can be written in the form

$$\delta n \triangleq n - n_0 \qquad \delta p \triangleq p - p_0 \tag{7.12}$$

where p_0 and n_0 are the thermal equilibrium charge densities. The G_n and G_p terms in the continuity equations represent charge generation mechanisms other than thermal generation. Possible mechanisms include avalanche multi-

plication, tunneling, or generation by radiation sources such as light and X rays.

Poisson's equation and Faraday's law complete the basic set of equations required for device analyses. Poisson's equation is

$$\nabla \cdot \mathbf{E} = \frac{q}{\epsilon}(p - n + N_d^+ - N_a^-) \tag{7.13}$$

where N_d^+ and N_a^- represent the density of ionized donor and acceptor atoms, and Faraday's law is

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{7.14}$$

Generally, the semiconductor equations must be solved simultaneously in a self-consistent manner to accurately describe device performance.

7.3 MATERIAL PARAMETERS

The properties of semiconductor crystals are very important to the operation of semiconductor devices. Normally, it is the particular properties of certain types of crystals that permit a device to be fabricated. Ultimate performance limitations of the device are quite often determined by the properties of the crystal.

All semiconductor devices are only possible because of the unique property that the conductivity of certain crystals can be controlled by the addition of very small amounts of certain impurities. The conductivity is determined by the amount of free charge in the crystal and the transport characteristics of the charge. The operation of the device is determined by the ability to move, generate, and remove this free charge in a controlled manner.

The material characteristics vary widely in various semiconductors, and only certain materials are suitable for use in device fabrication. The most common materials in current use are Si and GaAs, although InP and various ternary and quaternary III–V compounds (e.g., GaInAs, GaInAsP, and AlGaAs) have been developed for device applications.

The requirement that the conductivity must be controllable with small impurity levels is probably the single most important material consideration. This implies that the intrinsic background concentration at the device operating temperature must be low, typically in the range of 10^{14}–10^{15} cm^{-3}. Since the intrinsic density in GaAs is less than that in Si, higher resistivity substrates are available in GaAs. High-resistivity substrates are desirable since they allow the active region of a device (e.g., the conducting channel in a GaAs MESFET) to be isolated. Also, for MMIC applications device isolation can be achieved without the use of current-blocking techniques, such as back-biased junctions. The availability of high-resistivity GaAs substrates is one reason that this material has found such widespread use for MESFET devices and for MMICs.

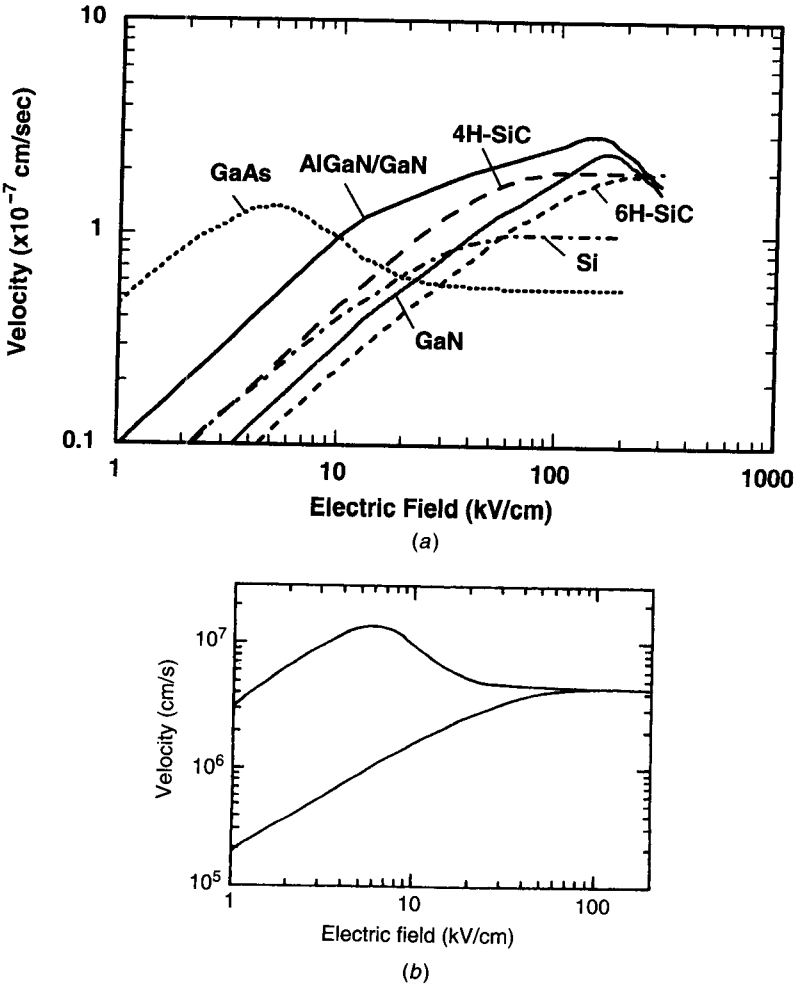


Figure 7.1 (*a*) Electron velocity–field characteristic for some common semiconductors at room temperature. (*b*) Velocity–field relationship for *n*-type GaAs at 10^{17} cm^{-3}. The upper curve is for electrons and the lower curve is for holes.

The ability to move charge is determined by the transport characteristics of the material. This information is most often presented in the charge carrier velocity–electric field characteristic. For example, the velocity-field characteristics for some commonly used semiconductors are shown in Fig. 7.1*a*, and the velocity-field characteristics for electrons and holes in GaAs are shown in Fig. 7.1*b*.

For low values of electric field the carrier velocity is linearly related to the electric field strength. The proportionality constant is the mobility, and this parameter is important in determining the low-field operation of a device. Par-

asitic resistances in a device are directly dependent upon the low-field mobility. The low-field mobility is also important in determining the RF noise characteristics of a semiconductor device. Generally, a high value of mobility is desired for optimum device performance. Since the mobility of electrons in GaAs ($\sim$6500 cm^2/V-s) is about six times that of Si ($\sim$1200 cm^2/V-s), GaAs is a more attractive material for high-RF and high-speed digital applications. This mobility advantage along with the availability of high-resistivity substrates makes GaAs the preferred and most widely used semiconductor material for these applications.

As the magnitude of the electric field increases, the carrier velocity saturates. Saturated velocity values for GaAs that have been reported in the literature for equilibrium conditions range from about 2×10^6 to 8×10^6 cm/s. For maximum high-frequency performance it is desirable to operate the device in a transit time mode at maximum velocity, which requires high fields. Of the semiconductors, Si has a $v_s \sim 10^7$ cm/s and would appear to be the material of choice. In fact, Si generally demonstrates higher operation frequencies than GaAs or other compound semiconductors only when certain devices are fabricated and compared. For example, fundamental-mode Si impact avalanche and transit time (IMPATT) oscillators have been operated as high as 341 GHz [1], whereas a GaAs device has produced RF power at about 248 GHz [2]. For three-terminal devices the compound semiconductors produce the highest frequency operation. These devices, have been operated to about 300 GHz [3].

The velocity–field transport characteristics are functions of various physical and operation parameters. An increase in doping produces a reduction in the low-field mobility and a decrease in the carrier velocity due to an increase in impurity scattering. The transport characteristics vary with temperature, and in the normal operating temperature range the mobility usually decreases with temperature. The saturated velocity also decreases with temperature, demonstrating a $v_s \sim 1/T$ relationship.

The behavior of the charge carrier diffusion coefficient with electric field is very important in the operation of certain solid state devices. Basically, the diffusion process will be important to a device that operates by the injection of a charge pulse into a high-field drift region and the extraction of the pulse from the opposite end. A high magnitude for the diffusion coefficient will result in a spreading out of the pulse and a degradation of information-transferring capability. The behavior of the diffusion coefficient with electric field for GaAs is shown in Fig. 7.2.

Thermal conductivity is important to devices intended for operation at higher power levels or in low-noise applications. Power dissipation requires that the device be adequately heat sinked. The thermal conductivity of compound semiconductors is generally much worse than for Si. For this reason Si transistors are often preferred for power applications. When compound semiconductors are used, careful heat sinking is required. Most heat sinks are fabricated from gold-plated copper. However, for high-power applications the use

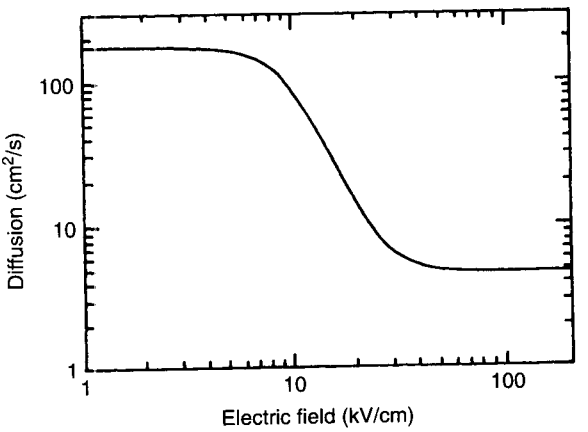


Figure 7.2 Electron diffusivity–field relationship for *n*-type GaAs at 10^{17} cm^{-3}.

of heat sink materials such as AlN, SiC, or type II diamond can improve the thermal conductivity over that obtainable from copper by about a factor of 6 at room temperature and a factor of 2 at 500 K.

7.4 BIPOLAR TRANSISTORS

The bipolar transistor was invented by Shockley, Bardeen, and Brittain in 1948. Since that time it has undergone continuous development and is currently one of the most widely used semiconductor devices [4]. For microwave applications the Si bipolar junction transistor (BJT) is useful for frequencies ranging from ultrahigh frequencies (UHF) (i.e., hundreds of megahertz) to about the *X* band (8–12 GHz), and the AlGaAs/GaAs heterojunction bipolar transistor (HBT) is useful to over 200 GHz. As the technology improves, the upper frequency limit for these devices is continuously being extended, and state-of-the-art Si devices now are capable of producing useful power through the *Ka* band and compound semiconductor HBTs can produce useful levels of RF power to well over 100 GHz. The majority of bipolar transistors are fabricated from Si- and GaAs-based epitaxial material. The Si BJT is inexpensive, durable, and integratable and offers higher gain than competing field-effect transistors. The bipolar transistor has moderate noise figure in RF applications and $1/f$ noise characteristics that are about 10–20 dB superior to field-effect transistors. For this reason it is often used for local oscillator applications. The bipolar transistor also has greater linearity than field-effect transistors, which makes it ideal for applications that require wide dynamic range.

The bipolar transistor is a *pn*-junction device and is formed from back-to-back junctions, as shown in Fig. 7.3. Since it is a three-terminal device it can be

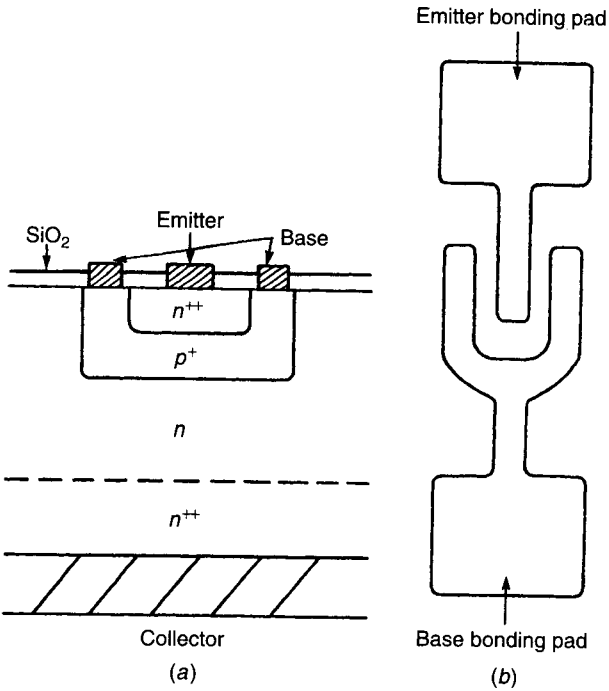


Figure 7.3 (*a*) Cross section for Si *npn* bipolar transistor. (*b*) Top view for Si *npn* bipolar transistor showing interdigitated planar geometry.

either *pnp* or *npn*. For high-frequency applications, the *npn* structure is preferred because the operation of the device is dependent upon the ability of minority carriers to diffuse across the base region. Since electrons generally have superior transport characteristics compared to holes, the *npn* structure is indicated.

Although there are many ways of fabricating a transistor, diffusion and ion implantation are generally used. For example, the structure in Fig. 7.3*a* would typically start with a lightly doped *n*-type epitaxial layer as the collector. The base region would be formed by counterdoping the base-region *p* type by diffusion. The emitter would be formed by a shallow heavily doped *n*-type diffusion or ion implantation. The emitter and base contacts are generally located on the semiconductor surface in an interdigitated, planar arrangement, as shown in Fig. 7.3*b*. The interdigitated geometry always provides for $n + 1$ base fingers, where n is the number of emitter fingers. The number of fingers varies with the application, with more fingers required as the output power capability of the transistor increases. Additional fingers, however, increase the device parasitics and degrade the noise and upper frequency capability of the device. An excellent review of microwave bipolar transistors is presented elsewhere [5].

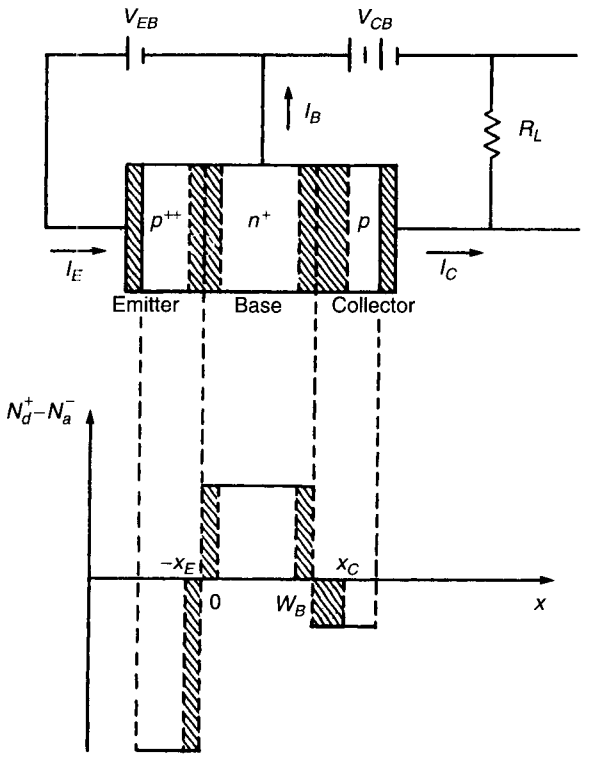


Figure 7.4 A *pnp* bipolar transistor connected in common-base configuration. Diagram defines terminology used in formulation given in the text.

7.4.1 Basic Transistor Operation

The basic physics responsible for the operation of bipolar transistors can be revealed through analysis of the common-base device shown in Fig. 7.4. For this analysis a *pnp* transistor is considered, although the arguments are analogous for an *npn* device. Under normal bias the emitter–base junction is forward biased and the collector–base junction is reverse biased. Holes are injected across the emitter–base junction, travel through the base region, and are extracted at the collector–base junction. The active characteristics of the device derive from the ability of a small potential applied across the base–emitter junction to control, in an efficient manner, a large emitter–collector current. This mechanism is the source for the name *transistor*, which is a unification of the words *transfer resistor*.

To derive the device I–V characteristics, the densities of free charges in each region of the device must be determined. From this information the device terminal currents can be calculated. The excess minority density in the base region $[p_B(x)]$ can be determined from a one-dimensional solution of the continuity equation. Using the geometry shown in Fig. 7.4 and the injected charge den-

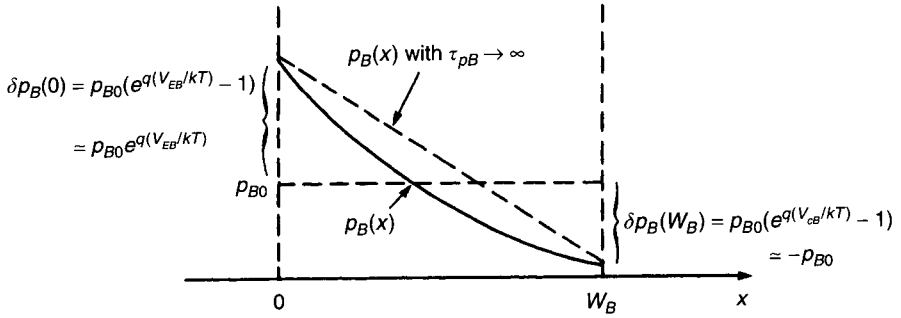


Figure 7.5 Minority-charge distribution in base region of bipolar transistor operated with normal biasing. The boundary conditions at the emitter and collector sides of the base region are indicated. The dotted line shows the situation for no minority-charge recombination.

sities for a biased *pn* junction, where

$$\delta p_B(0) = p_{B0}\left[\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right] \tag{7.15}$$

$$\delta p_B(W_B) = p_{B0}\left[\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right] \tag{7.16}$$

and p_{B0} is the thermal equilibrium density of minority-charge carriers in the base region, the excess minority density at the edges of the base region can be written for normal biasing (i.e., a forward-biased emitter–base junction and a reverse-biased collector–base junction) as

$$\delta p_B(0) = p_{B0}\left[\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right] \cong p_{B0}\exp\left(\frac{qV_{EB}}{kT}\right) \tag{7.17}$$

$$\delta p_B(W_B) \cong -p_{B0}$$

When these boundary conditions are applied to the solution, the result shown in Fig. 7.5 is obtained. The minority charge is injected into the base from the emitter and moves across the base region to be extracted at the collector. The operation of the transistor is therefore dependent upon the ability of the minority charge to travel the length of the base region. Effects that decrease the minority charge in the base will result in a degradation in transistor operation. For example, if some of the charge recombines before it reaches the collector, the gain capability of the transistor is degraded. For no minority-charge recombination in the base, the minority lifetime is infinite and the excess charge has a linear distribution. The minority-charge recombination in the base constitutes the base current, and high-performance transistors require a low base current.

The identical procedure can be applied to determine the minority-charge densities in the emitter and collector regions. The solutions applicable to these regions are

$$\delta n_E(x) = n_E(x) - n_{E0} = n_{E0}\left[\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right]\exp\left(\frac{x + x_E}{L_E}\right) \tag{7.18}$$

$$\delta n_C(x) = n_C(x) - n_{C0} = n_{C0}\left[\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right]\exp\left(\frac{-(x - x_C)}{L_C}\right) \tag{7.19}$$

where $L_E = \sqrt{D_{nE}\tau_E}$ and $L_C = \sqrt{D_{nC}\tau_C}$ are the minority-carrier diffusion lengths in the emitter and collector regions. The majority densities in each of the transistor regions are essentially equal to the background impurity densities.

Expressions for the emitter and collector currents as a function of the applied potentials can now be derived. Following basic *pn*-junction theory, it is customary to assume that the electron and hole currents are constant across the depletion regions. This assumption allows the depletion regions to be considered as a single node. The minority-carrier currents can be summed on each side of the depletion region to get the total current. If, in addition, the electric field in the bulk semiconductor regions is assumed to be very small, the total current can be determined from only diffusion currents. The emitter current density is

$$J_E = J_{pB}(0) + J_{nE}(-x_E)$$

or

$$J_E = -qD_{pB}\frac{\partial p_B(x)}{\partial x}\bigg|_0 + qD_{nE}\frac{\partial n_E(x)}{\partial x}\bigg|_{-x_E} \tag{7.20}$$

which when applied to a one-sided, abrupt p^+n junction yields

$$J_E \cong \frac{qD_{pB}p_{B0}}{L_B}\coth\left(\frac{W_B}{L_B}\right)\exp\left(\frac{qV_{EB}}{kT}\right) \tag{7.21}$$

In a similar manner the collector current density is

$$J_C \cong \frac{qD_{pB}p_{B0}}{L_B}\operatorname{csch}\left(\frac{W_B}{L_B}\right)\exp\left(\frac{qV_{EB}}{kT}\right) \tag{7.22}$$

The base current can be determined as the difference between the emitter and collector currents as

$$I_B = I_E - I_C \tag{7.23}$$

where

$$I_E \triangleq J_E A_{EB} \qquad I_C \triangleq J_C A_{CB}$$

where the term A is the area through which the current flows. Note that, in general, $A_{EB} \neq A_{CB}$.

If it is assumed, for simplicity, that the base–emitter and base–collector junctions have the same area A, the base current can be written as

$$I_B \cong qA \frac{W_B}{2\tau_{pB}} p_{B0} \exp\left(\frac{qV_{EB}}{kT}\right) \tag{7.24}$$

Note that the base current is directly proportional to W_B and inversely proportional to τ_{pB}. For no recombination in the base region $\tau_{pB} \rightarrow \infty$ and $I_B \rightarrow 0$. Also, for wide base regions, the minority charge has more opportunity to recombine and the base current increases.

7.4.2 Current Gain

The static common-base current gain is called alpha and is defined as

$$\alpha_0 \triangleq \frac{\Delta I_C}{\Delta I_E} \tag{7.25}$$

In the common-emitter configuration, the static current gain is called beta and is

$$\beta_0 \triangleq \frac{\Delta I_C}{\Delta I_B} \tag{7.26}$$

Since the base current is defined as the difference between the emitter and collector currents according to (7.23), it follows that the alpha and beta are related according to the expression

$$\beta_0 = \frac{\alpha_0}{1 - \alpha_0} \tag{7.27}$$

The alpha current gain can be defined as the product of two additional parameters as

$$\alpha_0 \triangleq B\gamma \tag{7.28}$$

where B is a base transport factor and γ is the emitter injection efficiency. The emitter injection efficiency is defined for a *pnp* transistor as

$$\gamma \triangleq \frac{\text{hole current injected from emitter into base}}{\text{total emitter current (hole and electron)}} \tag{7.29}$$

This can be written as

$$\gamma = \frac{1}{1 + J_{nE}(-x_E)/J_{pB}(0)} \tag{7.30}$$

and if the expressions previously derived are substituted, the emitter injection efficiency can be written as

$$\gamma \cong \left[1 + \frac{D_{nE}L_B N_B}{D_{pB}L_E N_E} \tanh\left(\frac{W_B}{L_B}\right)\right]^{-1} \tag{7.31}$$

where N_B and N_E are the base and emitter region dopings, respectively. For high injection efficiency the second term in the denominator must be minimized. This requires that the emitter doping be much greater than that in the base and also that the base-region length be much less than the minority-carrier diffusion length in the base.

The base transport factor is defined as

$$B \triangleq \frac{\text{hole current reaching collector}}{\text{total hole current injected into base}} \tag{7.32}$$

Substituting these expressions into the current gain expression, it follows that

$$B = \left[\cosh\left(\frac{W_B}{L_B}\right)\right]^{-1} \tag{7.33}$$

The alpha current gain can then be written in the form

$$\alpha_0 = B\gamma = \left[\cosh\left(\frac{W_B}{L_B}\right) + \frac{D_{nE}L_B N_B}{D_{pB}L_E N_E} \sinh\left(\frac{W_B}{L_B}\right)\right]^{-1} \tag{7.34}$$

The current gain expressed in (7.34) is the static, or DC, current gain. Under RF conditions the current gain will decrease, as shown in Fig. 7.6. From this figure the commonly used cutoff frequencies can be defined and are as follows:

1. ω_α—the alpha cutoff frequency is the frequency at which the common-base current gain is reduced to $0.707\alpha_0$ (3 dB down),
2. ω_β—the beta cutoff frequency is the frequency at which the common-emitter current gain is reduced to $0.707\beta_0$, and
3. ω_T—the gain–bandwidth product is the frequency at which the common-emitter current gain is reduced to unity.

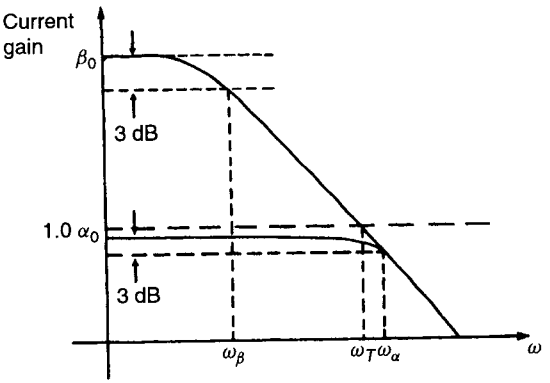


Figure 7.6 Current gains versus frequency, indicating the definitions for the various cutoff frequencies.

The frequency dependence of the current gain can be determined by repeating in the time domain the procedure used to determine the DC gain. For microwave transistors the current gain becomes

$$\alpha(\omega) \cong \operatorname{sech}\left(\frac{W_B}{L_B'}\right) \tag{7.35}$$

where L_B' is a frequency-dependent diffusion length. A simple one-pole model for α results as described by the expression

$$\alpha = \frac{\alpha_0}{1 + j(\omega/\omega_\alpha)} \tag{7.36}$$

7.4.3 Limitations and Second-Order Effects

There are several effects that limit the operation of bipolar transistors to restricted ranges of applied bias. These effects are examined in this section.

Collector Bias Variation (Early Effect). A changing collector–base bias causes a variation in the space-charge layer width at the base–collector junction and, consequently, in the width of the bulk base region. Base-width modulation by the collector–base voltage is called the Early effect [6]. An increase in V_{CB} reduces the base width, thereby increasing the gradient of the minority charge. A decrease in total minority charge in the base and an increase in collector current result. The base current is also decreased, since base current is proportional to base-charge storage.

Collector High-Injection Effects (Kirk Effect). Free carriers entering the base–collector depletion region modify the background charge in that region and thereby affect the electric field. For a constant collector–base voltage the

width of the depletion region changes to accommodate changes in the electric field. Under high injection the width of the depletion region tends to decrease, thereby increasing the effective neutral base region. This phenomenon is called the Kirk effect [7]. As the current increases, the neutral base region increases and the current gain is reduced.

The Kirk effect can have a significant effect upon device performance, since the neutral base region can increase by a large factor under high-injection conditions. The base transit time can be significantly increased under high current injection due to the added base length. Under these conditions, the high-frequency performance of the transistor will be degraded.

Base-Region High-Injection Effects. Under high-current-injection conditions, it is possible for the injected minority density in the base region to be on the same order of magnitude as the doping density. The large density of injected minority carriers causes an increase in majority carriers in order to satisfy charge neutrality conditions. The increase in majority carriers causes an effective increase in the base-region charge and a reduction in the alpha current gain. The reduction in α with I_C is known as the Webster effect [8].

7.4.4 Microwave Transistor

The state of the art in microwave bipolar transistors has developed rapidly, and these devices are now capable of operation approaching the *Ka* band. Much of the improvement in device performance is due to improvements in fine-line lithography where submicrometer emitter–base spacing and finger widths are now possible. Improvements in material growth and doping techniques have also contributed to the excellent performance now available from these devices. The Si bipolar transistor is the most widely used device in analog circuit design at microwave frequencies ranging from UHF to approximately the *S* band. In the *S* through *C* bands, the Si bipolar transistor and the GaAs MESFET are direct competitors.

The frequency performance of the bipolar transistor can be examined by consideration of the delay times encountered by a signal traveling through the device. For this analysis, it is convenient to consider the common-base configuration and normal biasing.

Four principal delay regions can be identified. The total emitter-to-collector delay time can be written as

$$\tau_{ec} = \tau_e + \tau_b + \tau_c + \tau_c' \tag{7.37}$$

where τ_{ec} = the total emitter-to-collector delay time
τ_e = the emitter–base junction capacitance charging time
τ_b = the base region transit time
τ_c = the collector-region depletion layer transit time
τ_c' = the collector capacitance charging time

Since a forward-biased emitter–base junction is assumed, the transit time associated with this depletion region can be ignored.

These delay times can be expressed as characteristic frequencies with the relationship

$$\omega = \frac{1}{\tau}$$

Therefore,

$$\frac{1}{\omega_T} = \frac{1}{\omega_e} + \frac{1}{\omega_b} + \frac{1}{\omega_c} + \frac{1}{\omega_c'} \tag{7.38}$$

where

$$\omega_T = 2\pi f_T$$

and f_T is the gain–bandwidth product.

Expressions for each of these delay terms can be determined in the following manner.

Base–Emitter Junction. Since the base–emitter junction is forward biased, an equivalent circuit is as shown in Fig. 7.7. The emitter current i_e entering the junction is divided into two paths. Due to the capacitance, a time delay results. The current that flows through r_e represents the charge carriers that are injected into the base region. This current is

$$i_e' = i_e \frac{1}{1 + j\omega r_e C_e} \tag{7.39}$$

By normalizing to the terminal current, a high-frequency emitter efficiency can be defined as

$$\frac{|\gamma_{hf}|}{\gamma_0} \triangleq \frac{\sqrt{2}}{2} \tag{7.40}$$

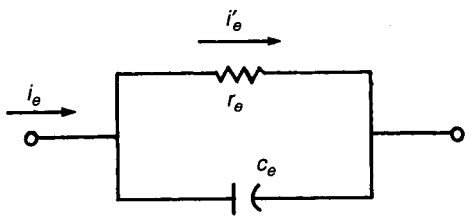


Figure 7.7 Equivalent circuit for the forward-biased base–emitter junction of a bipolar transistor.

and it follows that

$$\tau_e = \frac{1}{\omega_e} \cong r_e C_e \tag{7.41}$$

For microwave devices this expression must be modified due to the increased capacitance from the base–collector junction and other external capacitance connected to the base.

The expression is written as

$$\frac{1}{\omega_e} = r_e(C_e + C_c + C_p) \tag{7.42}$$

where C_c is the base–collector junction capacitance and C_p is any other capacitance connected to the base.

An expression can be derived for r_e from the relationship

$$r_e = \frac{dv}{di} = \frac{1}{di/dv} \tag{7.43}$$

where $i = I_E$, the current through the base–emitter junction. It follows that

$$r_e \cong \frac{kT}{qI_E} \tag{7.44}$$

$$\tau_e = \frac{1}{\omega_e} \cong \frac{kT}{qI_E}(C_e + C_c + C_p) \tag{7.45}$$

Base Region. The transit time across this region is

$$\tau_b = \frac{W_B^2}{\eta D_{pB}} \tag{7.46}$$

where η is a factor that includes the effects of a graded doping profile in the base region. For uniform doping $\eta = 2$.

Base–Collector Junction. Since the base–collector junction is reverse biased, it has a significant width. The transit time of carriers across the depletion region must be considered. The transit time across the depletion region is

$$\tau_c \triangleq \frac{\text{depletion width}}{\text{carrier velocity}} = \frac{X_C - W_B}{2v_s} \tag{7.47}$$

For a uniformly doped transistor and using the one-sided, abrupt junction approximation, it follows that

$$X_C - W_B \cong \left[\frac{2\epsilon(V_{CB} + \varphi_{bi})}{qN_C}\right]^{1/2} \tag{7.48}$$

where φ_{bi} is the base–collector junction built-in potential and N_c is the collector region doping. Therefore,

$$\tau_c \cong \frac{1}{2v_s}\left[\frac{2\epsilon(V_{CB} + \varphi_{bi})}{qN_c}\right]^{1/2} \tag{7.49}$$

Low doping and high V_{CB} cause increases in τ_c.

The base–collector capacitance charging time is calculated in a manner analogous to that for the emitter–base junction. The charging time is

$$\tau_c' = \frac{1}{\omega_c'} = r_c C_c \tag{7.50}$$

where r_c is the collector series resistance.

For typical microwave transistors the total emitter–collector delay time divides in the following manner:

$$\tau_e \sim 40\%\ \tau_{ec} \qquad \tau_b \sim 10\%\ \tau_{ec} \qquad \tau_c \sim 45\%\ \tau_{ec} \qquad \tau_c' \sim 5\%\ \tau_{ec}$$

The emitter–base charging time and collector depletion region transit time dominate (i.e., limit) the transistor frequency response.

The transistor gain–bandwidth product, f_T, is expressed as

$$f_T = \frac{1}{2\pi\tau_{ec}} \cong \left\{2\pi\left[\frac{kT(C_e + C_c + C_p)}{qI_C} + \frac{W_B^2}{\eta D_{pB}} + \frac{X_C - W_B}{2v_s}\right]\right\}^{-1} \tag{7.51}$$

assuming that $I_C = I_E$. For maximum frequency response the transistor should be operated at high collector current and low collector–base potential.

7.4.5 Equivalent Circuit

The Ebers–Moll mode [9] is generally considered to be the basic model for the bipolar transistor. It is a device physics model and is shown in both the common-base and common-emitter configurations in Fig. 7.8. The basic model consists of back-to-back diodes in shunt with current sources.

A useful measure of high-frequency performance is the unilateral gain (Mason's gain [10]), which can be defined from the y-parameter expression

$$U \triangleq \frac{|y_{21} - y_{12}|^2}{4\ \mathrm{Re}\ y_{11}\ \mathrm{Re}\ y_{22} - 4\ \mathrm{Re}\ y_{12}y_{21}} \tag{7.52}$$

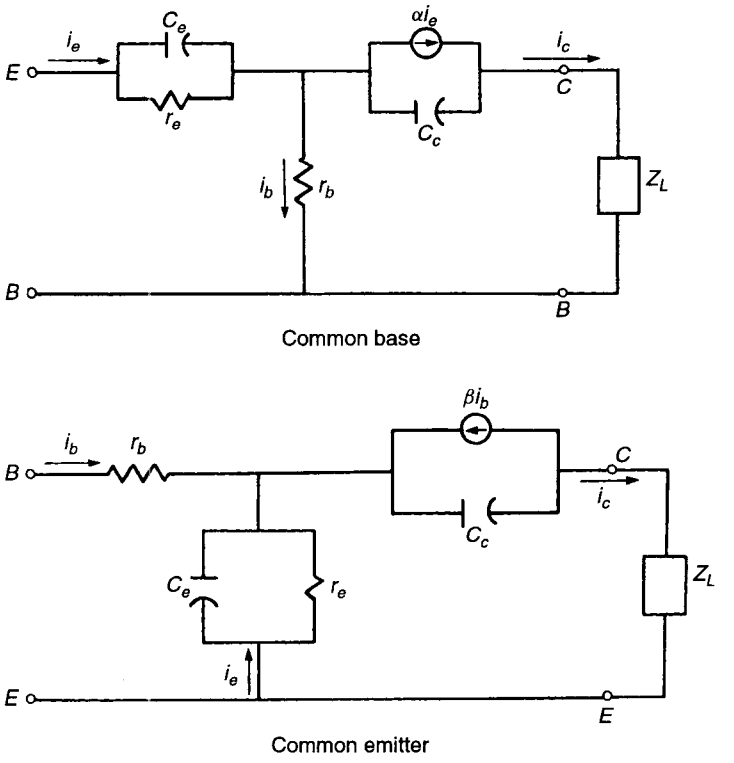


Figure 7.8 Ebers–Moll high-frequency equivalent circuits for a bipolar transistor connected in the common-base and common-emitter configurations.

where

$$y_{lm} = g_{lm} + jb_{lm}$$

and g and b are the device terminal conductance and susceptance, respectively. The unilateral gain is of interest because it excludes any effect of package parasitic reactances. It provides a useful figure of merit for the intrinsic device.

By applying the U definition to the common-base equivalent circuit, the transistor gain can be determined to be

$$U \cong \frac{\alpha_0}{16\pi^2 r_b C_c f^2(\tau_{ec} + r_e C_c/\alpha_0)} \tag{7.53}$$

The unilateral gain is therefore proportional to $1/f^2$, $1/r_b$, and $1/\tau_{ec}$. The $1/f^2$ dependence results in the 6-dB/octave gain rolloff, as shown in Fig. 7.9. The maximum frequency of oscillation, $f_{\max}$, is defined as the frequency at which U goes to unity. From the U expression it follows that

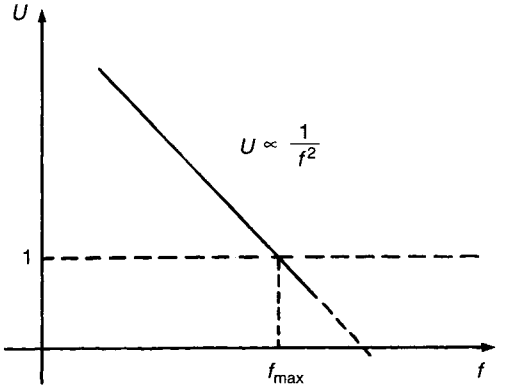


Figure 7.9 Unilateral gain vs. frequency for the single-pole equivalent circuit for a transistor.

$$f_{max} = \frac{1}{4\pi}\left[\frac{\alpha_0}{r_b C_c(\tau_{ec} + r_e C_c/\alpha_0)}\right]^{1/2} \tag{7.54}$$

Maximum oscillation frequency f_{max} can be expressed as a function of f_T as

$$f_{max} = \left(\frac{f_T}{8\pi r_b C_c}\right)^{1/2} \tag{7.55}$$

In general, $f_{max} \gg f_T$ for Si transistors. That is, the transistor is capable of producing power gain at frequencies above that at which the current gain is unity. In GaAs transistors, it is possible for f_T to be greater than f_{max} due to series losses and the effect the high mobility of GaAs has upon the relative sizes of the terms r_b, C_c, τ_{ec} and τ_{ec} in a particular transistor structure.

Maximum oscillation frequency f_{max} is a useful figure of merit, since it gives an upper limit to the frequency at which a given device should be capable of useful gain. In practice, parasitics and second-order effects will limit operation to frequencies below f_{max}.

For the analysis and design of practical microwave transistors the *T*-equivalent circuit shown in Fig. 7.10 is useful [11]. This circuit is essentially the Ebers–Moll circuit with a distributed *RC* network in the base to account for the capacitances and lateral base resistance elements associated with the planar transistor geometry. This circuit, when applied to the HP 505 microwave bipolar transistor, yields the element values shown in Table 7.1.

7.4.6 Noise Figure Analysis

Noise figure is a measure of the amount of noise added to a signal by a lossy device through which the signal passes. It is defined as the ratio of the input-to-output signal-to-noise ratios and is generally expressed in decibels according to

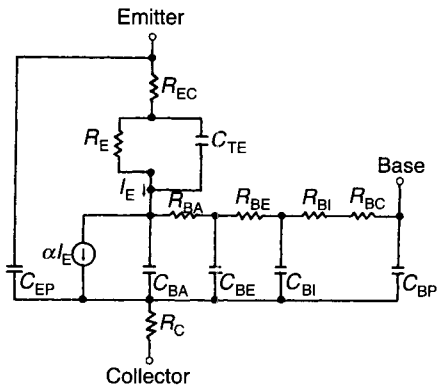


Figure 7.10 High-frequency equivalent circuit for a bipolar transistor including distributed base parameters.

the expression

$$F = 10 \log \frac{(S/N)_{\text{in}}}{(S/N)_{\text{out}}} \tag{7.56}$$

Note that a noise-free device would have equal input and output signal-to-noise ratios and the noise figure is therefore zero. Another parameter sometimes used to describe the noise of an amplifier is noise measure. The noise measure is defined in terms of noise figure and the gain of the device as

$$M = \frac{F' - 1}{1 - (1/G')} \tag{7.57}$$

where the primes indicate nondecibel magnitudes. Noise measure is also generally specified in decibels and is of interest because it combines the effects of noise figure and gain into a single parameter. In low-noise design, especially for front-end stages of receivers, it is often desirable to design for minimum noise measure, rather than simply minimum noise figure.

In terms of a two-port device, the noise figure can be calculated by considering a noise-free two-port to which noise generators are added. The noise added by the two-port can be represented as shunt current generators at the input and output ports or as a series noise voltage generator and shunt noise current generator at the two-port input [12]. In circuit design the amplifier noise figure calculations are generally performed with all noise sources referenced to the input.

The minimum noise figure for the bipolar transistor can be expressed in the form [13]

$$F_{\min} = a\frac{r_b + R_{\text{opt}}}{r_e} + \left(1 + \frac{f^2}{f_b^2}\right)\frac{1}{\alpha_0} \tag{7.58}$$

where the optimum source resistance is [14]

Table 7.1 HP 505 Equivalent-Circuit Parameters

Parameter	Symbol	Value	How Obtained
Active base resistance	R_{BA}	12.5 Ω	Calculated
External base resistance	R_{BE}	2.7 Ω	Calculated
Base insert resistance	R_{BI}	0.2 Ω	Calculated
Base contact resistance	R_{BC}	<0.2 Ω	Calculated
Emitter contact resistance	R_{EC}	<1 Ω	Measured
Collector resistance	R_C	5 Ω	Measured
Emitter–base dynamic resistance	r_e	13 Ω	kT/qI_E
Active base collector–base junction capacitance	C_{BA}	0.0040 pF	Calculated
External base collector–base junction capacitance	C_{BE}	0.0032 pF	Calculated
Base insert collector–base junction capacitance	C_{BI}	0.050 pF	Calculated
Base bonding pad capacitance	C_{BP}	0.05 pF	Calculated
Emitter bonding pad capacitance	C_{EP}	0.05 pF	Calculated
Emitter–base junction capacitance	C_{TE}	0.7 pF	τ_e/r_e
Emitter–base junction charging time	τ_e	9 ps	$\tau_T = (\tau_b + \tau_e + \tau_d)$
Neutral base delay time	τ_b	7 ps	From τ_T vs. I/I_e
Charging time of C_{BC} through R_C	τ_c	0.5 ps	R_cC_{BC}
Collector depletion region delay time	τ_d	10 ps	$W_d/2v_s$ ($v_s = 0.5 \times 10^6$ cm/s)
Low-frequency common-base gain	α_0	0.990	$n_{FE}/(1 + n_{FE})$
Total base resistance	R_B	15.6	$R_{BA} + R_{BE} + R_{BI} + R_{EC}$
Total base–collector capacitance	C_{BC}	0.107 pF	$C_{BA} + C_{BE} + C_{BI} + C_{BP}$
Total delay time	τ_T	26.5 ps	$(\tau_e + \tau_b + \tau_c + \tau_d) = 1/2\pi f_T$
Emitter–base junction cutoff frequency	f_e	17.7 GHz	$1/2\pi\tau_e$
Base cutoff frequency	f_b	22.7 Ghz	$1/2\pi\tau_b$
Common-emitter gain–bandwidth product	f_T	6 GHz	Measured

$$R_{\text{opt}} = \left[r_b^2 - X_{\text{opt}}^2 + \left(1 + \frac{f^2}{f_b^2}\right) \frac{r_e(2r_b + r_e)}{\alpha_0 a} \right]^{1/2} \tag{7.59}$$

and the optimum source reactance is

$$X_{\text{opt}} = \left(1 + \frac{f^2}{f_b^2}\right) \frac{2\pi f C_{\text{TE}} r_e^2}{\alpha_0 a} \tag{7.60}$$

where

$$a = \left[\left(1 + \frac{f^2}{f_b^2}\right) \left(1 + \frac{f^2}{f_e^2}\right) - \alpha_0 \right] \frac{1}{\alpha_0} \tag{7.61}$$

Here f_b and f_e are the base and emitter–base junction cutoff frequencies, respectively. These expressions were derived neglecting the parasitic bonding pad capacitances and the collector–base junction capacitance. When the parameters listed in Table 7.1 for the HP 505 transistor are used in the noise figure equations, the noise figure performance shown in Fig. 7.11 is obtained.

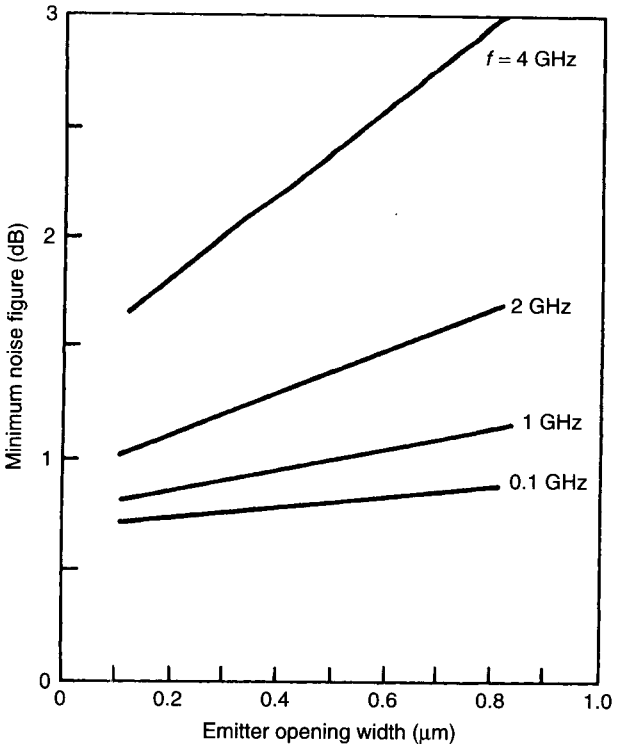


Figure 7.11 Minimum noise figure at 0.1, 1, 2, and 4 GHz vs. emitter opening width for a HP 505 transistor [14].

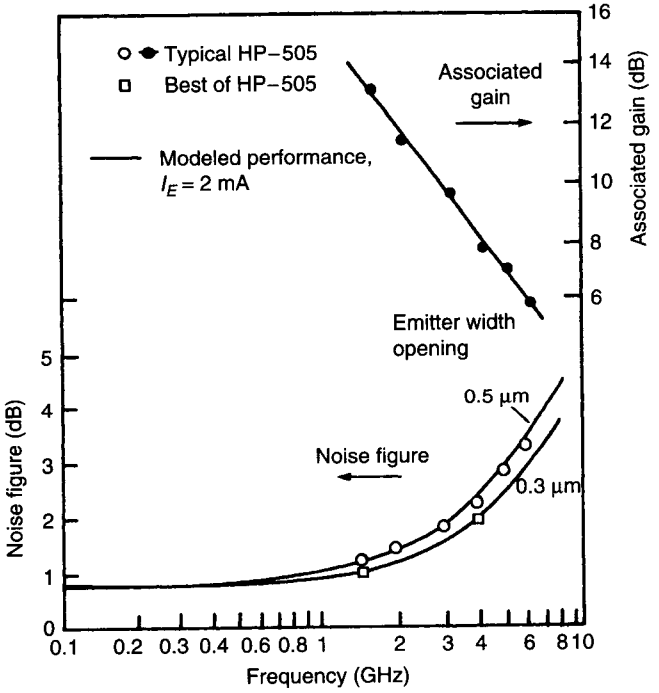


Figure 7.12 Measured and calculated noise figure for the HP 505 as a function of frequency. (After Hsu and Snapp [14]. Reprinted with permission of IEEE.)

The measured and calculated gain and noise figure performance for this transistor are compared in Fig. 7.12.

7.4.7 Heterojunction Bipolar Transistors

Since the bipolar transistor was invented by Shockly [49], and Bardeen and Brittain [50], in 1948 the HBT has undergone continued development and improvement and is now in wide use for microwave and millimeter-wave applications. Although the advantages of utilizing a wide-bandgap semiconductor for the emitter of a bipolar transistor were discussed by Shockley in his transistor patent, the modern HBT was proposed in 1957 by Kroemer [15], who also discussed the advantages of HBTs over conventional bipolar transistors. Significant development followed and promising results started to appear in the 1970s [16–18] with HBT development in III–V compound semiconductors based upon the AlGaAs/GaAs system. However, it was not until the development of MBE semiconductor growth technology that practical heterojunctions of sufficient quality for practical device application could be produced, and by the 1980s the technology of fabricating HBTs with excellent microwave performance was advancing [19–21]. By the early 1990s the current

gain frequency response of these devices had reached 200 GHz [4]. Submicrometer scaling has now pushed the f_T's to the range of 300 GHz [3]. The development of HBTs using SiGe as the base was reported in 1987 [22], and these devices advanced rapidly and now produce RF performance essentially equivalent to AlGaAs/GaAs HBTs. The SiGe material is used as the base region, and since SiGe has a smaller bandgap than Si, a device with the advantages of a wide-bandgap emitter is obtained. The SiGe/Si HBT has the advantage of being compatible with standard Si processing technology, which makes the device attractive from a cost perspective.

A HBT is fabricated by utilizing a semiconductor for the emitter that has a wider bandgap than for the base. The advantages of this structure were outlined by Shockly in his original bipolar transistor patent. Basically, in order to achieve good performance it is desirable for the emitter current injected into the base to consist essentially of minority current in the base region, as discussed earlier in this chapter. The currents in the base and emitter regions, however, consist of both minority and majority currents, and some majority current in the base is back injected into the emitter, where it degrades transistor performance. The back-injected current degrades the current injection efficiency and reduces transistor gain. In a standard bipolar transistor the back injection can only be minimized by fabricating the device with an emitter impurity concentration much larger than that in the base, typically by one to two orders of magnitude. However, by introducing a wide-bandgap semiconductor for the emitter, the back injection of current can be blocked by the energy band discontinuity. The emitter region doping can, therefore, be optimized for transistor performance without back-injection concerns. This is the advantage of the HBT noted by Shockley. The heterojunction bandgap can be chosen so that the HBT will have current gain independent of the base and emitter doping. This permits fabrication of a microwave transistor with a heavily doped base region and a lightly doped emitter region. Therefore, compared to a standard bipolar transistor, the HBT has reduced base resistance, output conductance, and emitter depletion capacitance and greatly improved high-frequency performance.

The ability to employ highly doped base regions permits low base-region resistance r_b to be obtained. Primarily for this reason HBTs have excellent high-frequency performance capability, as indicated in Eq. (7.54). Heterojunction bipolar transistors produce increased f_T compared to standard bipolar transistors and f_{max} values exceeding 200 GHz have been obtained with AlGaAs/GaAs devices and 300 GHz with InGaAs/InP devices. Good results have also been obtained from SiGe/Si HBTs. In these devices SiGe is used as the base region, and since SiGe has a reduced bandwidth compared to Si, a HBT can be fabricated. A comparison of gain–bandwidth products for similar GaAs and SiGe HBTs has been reported by Ning [4]. A GaAs HBT with a $0.6 \times 4.6\text{-}\mu m^2$ emitter produced an f_T of 140 GHz whereas a SiGe HBT with a $0.35 \times 3.55\text{-}\mu m^2$ emitter produced an f_T of 130 GHz. The peak f_T occurs at slightly lower collector current for the GaAs HBT compared to the SiGe HBT.

These results indicated that the two HBTs have comparable high-frequency performance. The SiGe/Si HBTs also have excellent low-noise performance due to high mobility in the SiGe material, which helps produce a low base resistance. Heterojunction bipolar transistors have improved performance and noise figures compared to standard Si bipolar transistors, and noise figures in the range of 2 dB at the K band can be obtained from AlGaAs/GaAs HBTs. Silicon–germanium HBTs produce noise figures just under 2 dB at the X band and in the range of 3 dB at 20 GHz.

7.5 FIELD-EFFECT TRANSISTORS

The basic idea for the field-effect transistor (FET) goes back at least as far as Lilienfeld's [51, 52] patents in 1930 and 1933. The concept for the FET predates the invention of the bipolar transistor that was presented in 1948. Development work on FETs was hindered in the early years by poor semiconductor quality, large feature size available with current lithography, and difficulty in obtaining suitable contacts. Most development work was temporarily interrupted by the invention of the bipolar transistor, as this device came under intense investigation and development.

The modern FET derives from the work of Stuetzer [23] in 1950 and Shockley [24] in 1952. Most of the early FET work was directed toward producing a high-frequency, high-power solid state analog to the vacuum triode. The basic structure was (and still is) similar to a triode in that two conducting electrodes are separated by some distance with a control electrode (grid) in between, as shown in Fig. 7.13. The control grid (called a gate) is biased negative with respect to the source so that a retarding potential barrier is presented

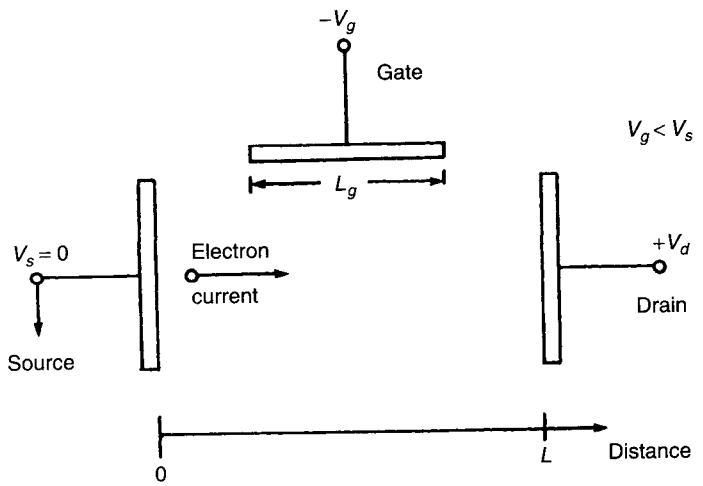


Figure 7.13 Basic electrode placement for a FET.

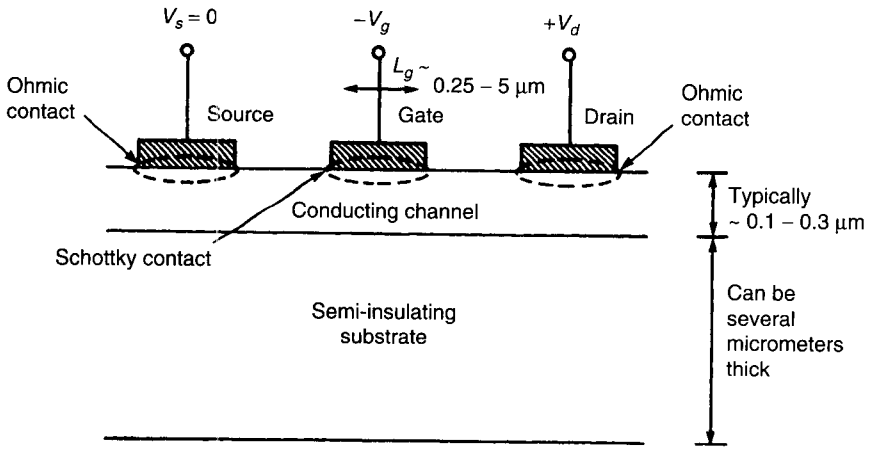


Figure 7.14 Typical cross-sectional structure of a microwave FET.

to electrons attempting to flow from the source to the drain. The FET depends upon only one type of charge carrier flow and is therefore called "unipolar."

7.5.1 Basic Operation Principles

The GaAs MESFET is one of the most commonly used and important active devices for use in microwave analog and high-speed digital ICs. The transistor consists of a highly conducting layer of high-quality semiconductor grown epitaxially upon a semi-insulating (i.e., high-resistivity) substrate, as shown in Fig. 7.14. The conducting channel is interfaced with external circuitry through two ohmic contacts (called the source and drain) separated by a distance and placed upon the semiconductor. The gate electrode is constructed by placing a rectifying (Schottky) contact between the two ohmic contacts. The conducting channel is very thin, typically on the order of 0.1–0.3 μm, so that the depletion region that forms under the Schottky contact (gate) can efficiently control the flow of current in the thin layer. The device therefore behaves as a voltage-controlled switch, capable of very high modulation rates.

The operation of the device can be understood by first considering a device without the gate, as shown in Fig. 7.15. As the drain–source potential is increased, a current will flow, as indicated by region *A*. The current–voltage behavior is linear and directly follows the velocity–field characteristic for the semiconductor from which the device is fabricated. As the drain–source voltage is increased to the magnitude at which the internal electric field achieves the saturation condition, deviations from linear performance occur (region *B*). The device is simply a resistor and the slope of the I–V characteristic defines the sum of the channel and contact resistances. Note in Fig. 7.15 that the parameter I_{dss} is defined for such a structure without a gate. The I_{dss} parameter is important in MESFET analyses and appears in the discussion that follows.

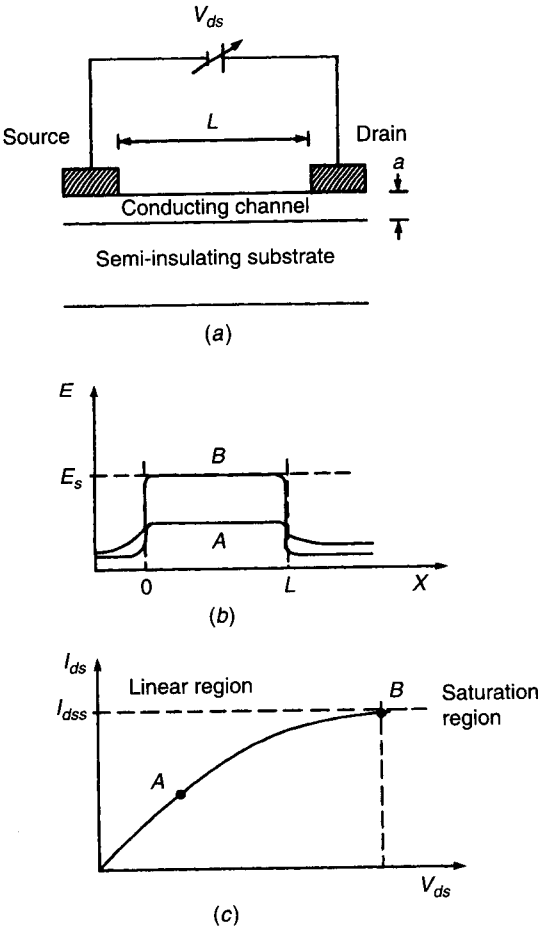


Figure 7.15 (*a*) Cross-sectional view of gateless FET. (*b*) Internal electric field for structure biased in linear (*A*) and saturation (*B*) regions. (*c*) The *I–V* characteristic for the device.

If a gate electrode is located between the source and drain contacts as shown in Fig. 7.16, a depletion region is formed under the gate metal. The depletion region affects the device *I–V* characteristic by constricting the cross-sectional area through which current can flow. This results in an increase in channel resistance [i.e., R_{ch} $(V_g = 0) > R_{ch}$ (no gate)]. The gate depletion region also causes the electric field under the gate to increase with distance toward the drain because of the potential drop along the channel. This, in turn, results in the saturation field being achieved first at the drain side of the gate. The drain voltage at which the maximum field in the channel achieves E_s is defined as $V_{d,sat}$. As the gate voltage is increased in the negative direction, the depletion region is moved deeper into the conducting channel, thereby reducing the cross-

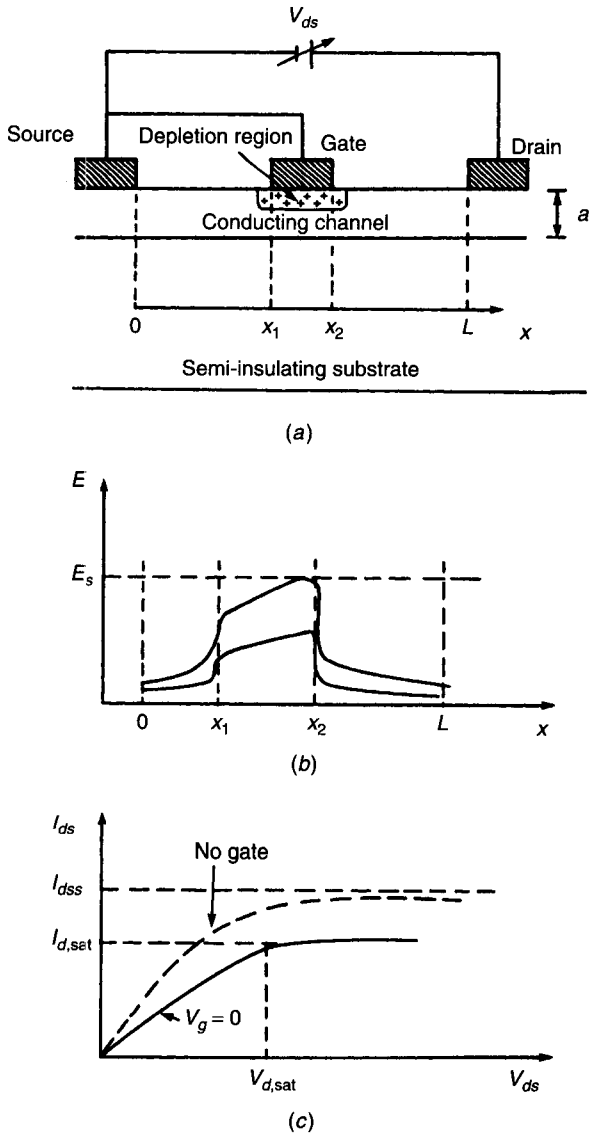


Figure 7.16 (*a*) Cross-sectional view of biased FET. (*b*) Internal electric field for device biased in linear and saturation regions. (*c*) The *I–V* characteristic for the device. Dashed line shows gateless device for comparison.

sectional area through which current can flow. The channel resistance increases and the terminal voltage at which saturation occurs is reduced.

Consider a biased condition and the region under the gate when the peak electric field exceeds the saturation field E_s. The current at all cross-sectional planes must be constant to satisfy current continuity requirements. The cross-

sectional area under the gate, however, is reduced in the direction of current flow. The channel current can be expressed as

$$I_{ds} = Zqn(x)v(x)b(x) \tag{7.62}$$

where $A = Zb(x)$, the channel cross-sectional area, Z is the channel width, and $b(x)$ is the channel opening. In the linear region $n(x) \cong N_d \neq f(x)$ for uniform doping. The reduction in $b(x)$ is compensated for by an increase in $E(x)$ and the corresponding increase in carrier velocity $v(x)$. The result is that I_{ds} is constant.

When $E = E_s$, however, $v(x)$ can no longer increase. The reduction in $b(x)$ can only be compensated for by an increase in $n(x)$. It follows that $n(x) > N_d$ and an accumulation region is created. In the region between the gate and drain $b(x)$ reaches a minimum and then starts to increase again as the depletion region is passed. Since the channel is still in velocity saturation, the increase in $b(x)$ is compensated for by a decrease in $n(x)$. A depletion of electrons [i.e., $n(x) < N_d$] is created, resulting in an electric field dipole in the gate–drain region [25]. This situation is illustrated in Fig. 7.17*a*.

As V_{ds} is increased beyond $V_{d,\text{sat}}$, the depletion region moves toward the drain. The point x_1 (Fig. 7.17*a*) moves toward the source and the potential at x_1 decreases. The channel opening at x_1 increases and more current is injected into the velocity saturation region. Therefore, an increasing V_{ds} produces an increase in I_{ds}, and a positive slope to the I_{ds}–V_{ds} characteristic is obtained in the saturation region. The channel has a finite positive resistance.

Note that the dipole formation is material independent, since it occurs because of geometry considerations in conjunction with velocity saturation. Dipoles can therefore form in Si as well as III–V materials. The negative differential mobility characteristics of materials such as GaAs, however, enhance the process and much larger dipoles are formed. This situation is illustrated in Fig. 7.17*b*. The existence of the dipole has significant influence upon the operation of field-effect-type devices, since it creates a feedback path from the output to the input. This feedback path affects the device gain and frequency performance. It is shown in the equivalent circuit analysis presented in Section 7.5.4 that the dipole results in a −12-dB/octave rolloff in the unilateral gain. The dipole also affects the output impedance of the device and is especially important in understanding the operation of devices designed for power and large-signal applications.

The operation of a MESFET is very sensitive to the channel thickness. Maximum performance is obtained when the device is designed so that the gate depletion region exercises optimum control over the current flow. This generally occurs for $L/a \cong 3$, where L is the gate length and a is the conducting channel thickness.

Since the operation of the FET is determined by the gating action of a reverse-biased rectifying junction, two modes of operation are possible. The device can be designed so that with zero gate bias a channel current flows. Such

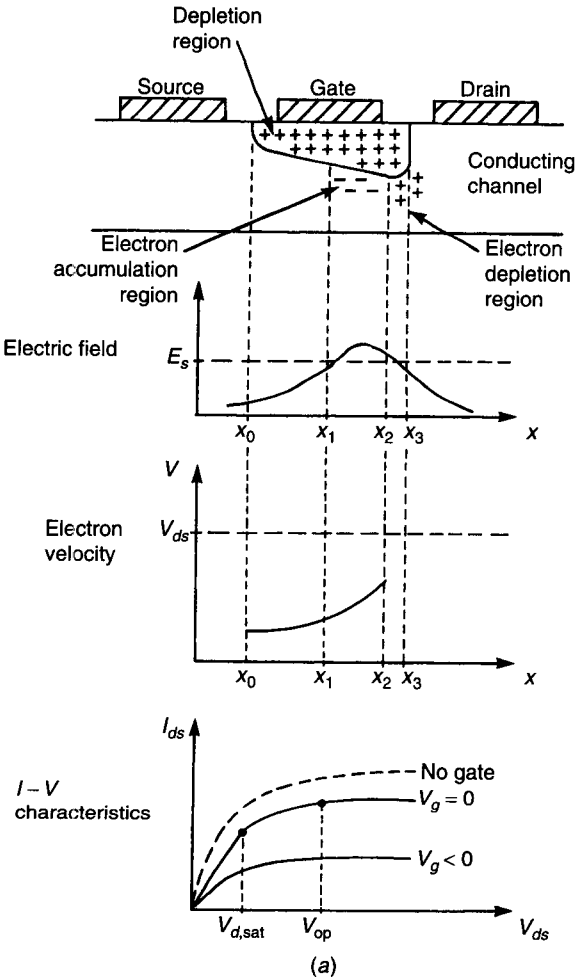
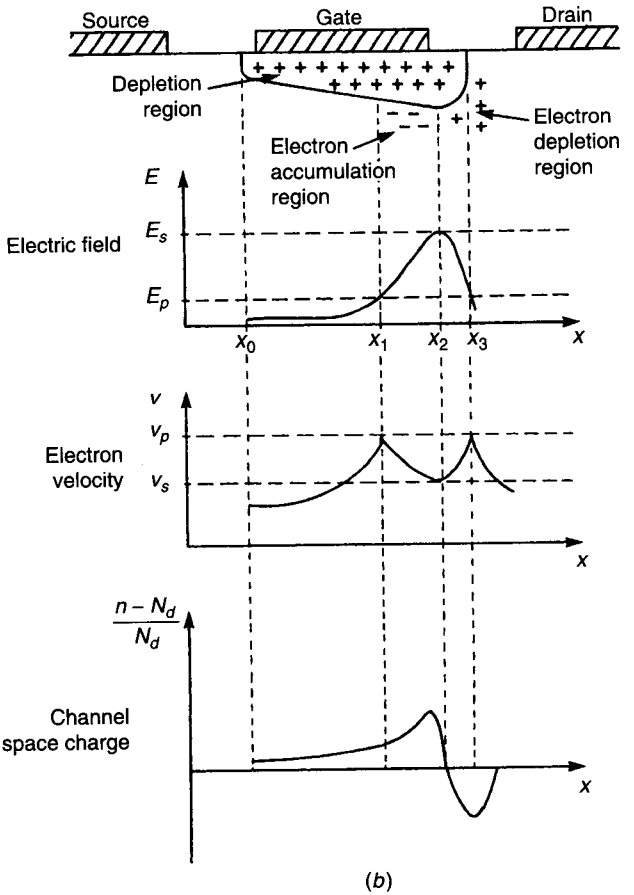


Figure 7.17 (*a*) Internal device physics for FET showing formation of gate–drain region electric dipole. (*b*) Internal device physics for GaAs FET showing formation of enhanced gate–drain region electric dipole due to negative differential mobility region in GaAs velocity–field characteristic (see Fig. 7.1).

operation occurs with relatively thick channels and is termed "normally on." The device can also be designed with a relatively narrow channel so that the zero bias depletion region is sufficient to pinch the channel off, resulting in no channel current. This situation is termed "normally off." In this mode of operation the gate must be forward biased to obtain current flow. The normally off mode is particularly interesting for high-speed logic applications, since very low power consumption occurs. The device will only dissipate energy when the channel conducts, and this only occurs for a fraction of an operation cycle. The



(*b*)

Figure 7.17 *(Continued)*

low power consumption is very attractive for high-density ICs, where large numbers of devices are located in a very small area. Heat dissipation is a fundamental limit to the level of integration that can be achieved.

7.5.2 MESFET Model

In a short-gate GaAs MESFET, part of the channel region under the gate will be dominated by linear current flow and part will be dominated by saturation. That is, electrons will be accelerated from a low velocity at the source end of the channel to their saturation velocity at the drain end of the gate region. The electric fields near the drain end of the gate region almost always exceed the saturation field in short-gate-length GaAs MESFETs. A two-region model that can satisfactorily account for these effects has been presented [26].

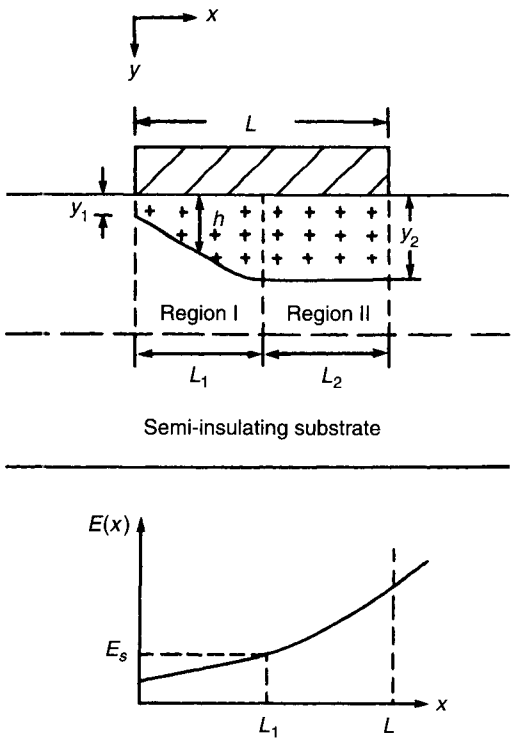


Figure 7.18 Field-effect geometry used in the formulation of the two-region model.

The channel region under the gate can be divided into linear and saturation parts, as shown in Fig. 7.18. In region I, the current is assumed dominated by linear behavior, and the electron velocity is linearly dependent upon the strength of the electric field. In region II, the carrier velocity is assumed constant and independent of field. The assumed velocity–field characteristic is shown in Fig. 7.19. Note that the saturation velocity is an "effective" parameter and not the true saturation velocity. An effective velocity is required since the two-region model is not capable of describing the negative differential mobility region of the GaAs velocity–field characteristic. The effective saturation velocity has values between the peak and true saturation values. An averaging method for determining an appropriate value for the effective saturation velocity has been presented [27].

The model is derived with the use of the depletion approximation and Shockley's gradual channel approximation. A two-sided device geometry is assumed, and only the region under the gate is simulated. The other regions of the device are considered as parasitic elements in the complete device equivalent circuit. The two-sided channel region eliminates the substrate from the analysis, thereby simplifying the derivation of the model. The chief advantage

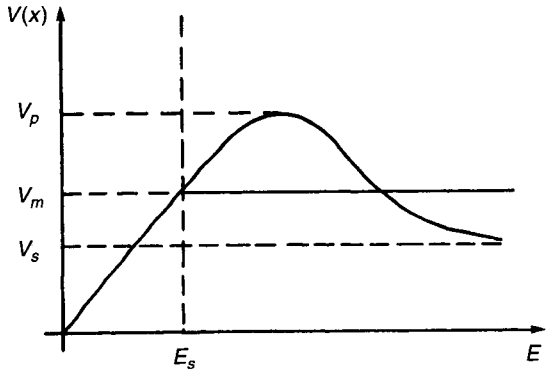


Figure 7.19 Piecewise linear two-region velocity–field characteristic used in the MESFET model.

in considering only the region under the gate is that the fringing fields that are associated with the edge regions of the gate do not need to be considered. Although this simplification allows solutions to be obtained in a direct manner, the simplification establishes limits to the model's validity. In fact, this model does not work well for submicrometer-gate-length devices where the fringing areas occupy a significant proportion of the gate depletion region. The model works well for 1-μm-gate-length devices and produces acceptable results, in certain instances, for gate lengths down to $\frac{1}{2}$ μm. It does not produce accurate results for shorter gate lengths.

The model is derived by writing a potential drop equation along the channel. The channel potential is written as

$$W(x) = V_{gs} + \varphi_{bi} - V(x) \tag{7.63}$$

where V_{gs} is the gate voltage, $V(x)$ is the potential across the depletion region, and φ_{bi} is the built-in potential. At the source end of the channel the potential is

$$W_s \triangleq V_{gs} + \varphi_{bi} \tag{7.64}$$

At the boundary between the linear and saturation regions the potential is

$$W_p \triangleq V_{gs} + \varphi_{bi} - V_p \tag{7.65}$$

and at the drain end of the channel the potential is

$$W_d \triangleq V_{gs} + \varphi_{bi} - V_{sd} \tag{7.66}$$

It is convenient to introduce normalized potentials defined as

$$s \triangleq \left(\frac{W_s}{W_{00}}\right)^{1/2} \tag{7.67a}$$

$$p \triangleq \left(\frac{W_p}{W_{00}}\right)^{1/2} \tag{7.67b}$$

$$d \triangleq \left(\frac{W_d}{W_{00}}\right)^{1/2} \tag{7.67c}$$

$$w(x) = \left[\frac{W(x)}{W_{00}}\right]^{1/2} \tag{7.68}$$

where $W_{00} \triangleq (qN_d/2\epsilon)a^2$, the channel pinchoff potential. The model is derived by considering the two channel regions separately and then invoking current continuity at the interface to obtain an overall channel current expression.

Region I. The channel opening is calculated by solving Poisson's equation in the depletion region to get

$$W(x) = W_{00}\left[1 - \frac{b(x)}{a}\right]^2 \tag{7.69}$$

where $b(x)$ is the channel opening under the depletion region. The drain current is calculated to be

$$I_d = J_d A = \sigma E_x A = \sigma\left(\frac{dW(x)}{dx}\right) Z[2b(x)] \tag{7.70}$$

where σ is the conductivity of the conducting channel. After some manipulation this can be written in terms of the normalized potentials as

$$I_d = \frac{g_0 Z W_{00}}{L_1}[p^2 - s^2 - \tfrac{2}{3}(p^3 - s^3)] \tag{7.71}$$

where $g_0 = 2a\sigma$ is the channel conductance and Z is the gate width.

Region II. In the saturation region the channel current is

$$I_d = J_d A = \sigma E_s(2b_p)Z \tag{7.72}$$

where $2b_p$ is the conducting channel opening at pinchoff. It can be shown that

$$b_p = a(1 - p) \tag{7.73}$$

$$I_d = g_0 Z E_s(1 - p) = I_s(1 - p) \tag{7.74}$$

where I_s is the saturation current (i.e., the maximum current that could flow if the channel were fully undepleted and the charge carriers were moving at the saturation velocity v_s).

By equating drain current expressions of regions I and II, the boundary L_1 between the linear and saturation regions can be determined. It follows that

$$L_1 = \frac{g_0 Z W_{00}}{I_s(1-p)} \left[p^2 - s^2 - \tfrac{2}{3}(p^3 - s^3)\right] \tag{7.75}$$

To define the drain current in terms of the applied potentials, the normalized potentials must be related to the applied bias potentials. This can be accomplished with the relationships

$$V_{sd} = -\int_0^L E_x \, dx \tag{7.76}$$

and

$$V_{ds} = -\underbrace{\int_0^{L_1} E_x \, dx}_{\text{Region I}} - \underbrace{\int_{L_1}^{L} E_x \, dx}_{\text{Region II}} \tag{7.77}$$

For region I, the integral can be written as

$$V_I = -[W(L_1) - W(0)] \tag{7.78}$$

and

$$V_1 = -(W_p - W_s) = -W_{00}(p^2 - s^2) \tag{7.79}$$

For region II, the potential is inherently two dimensional and, in general, is very difficult to determine in an analytic fashion. An approximation for the potential can be determined by using the superposition principle and an approach presented by Grebene and Ghandhi [28]. This approach results in an expression for the source–drain potential drop that has the form

$$V_{ds} = -W_{00}(p^2 - s^2) - \frac{2aE_s}{\pi} \sinh \frac{\pi L_2}{2a} \tag{7.80}$$

where L_2 is the length of the channel region in saturation. This equation, along with those for L_1 and I_d, allows the I–V characteristics for the device to be determined. A small-signal model can then be determined by taking various partial derivatives to define the critical equivalent-circuit elements.

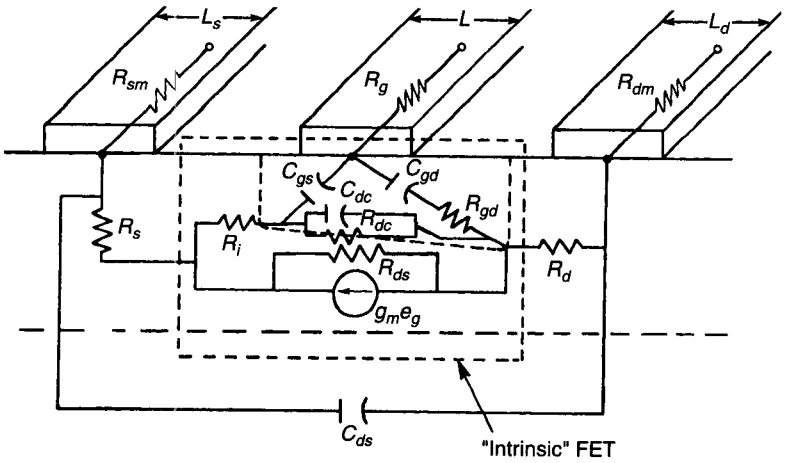


Figure 7.20 Equivalent circuit for MESFET superimposed upon device geometry to show physical basis for various elements.

7.5.3 Small-Signal Model

An equivalent circuit for the MESFET is superimposed upon the FET structure in the sketch in Fig. 7.20. The equivalent-circuit elements consist of both passive parasitic parameters and the parameters responsible for the active characteristics of the device. The sketch in Fig. 7.20 shows the physical origin of the various elements. The circuit is generally shown as a T-equivalent circuit, as depicted in Fig. 7.21. In this section expressions for some of the most important elements in the circuit are derived.

Transconductance. The critical elements can be determined by perturbation of the DC I–V characteristics. The device transconductance g_m is defined as

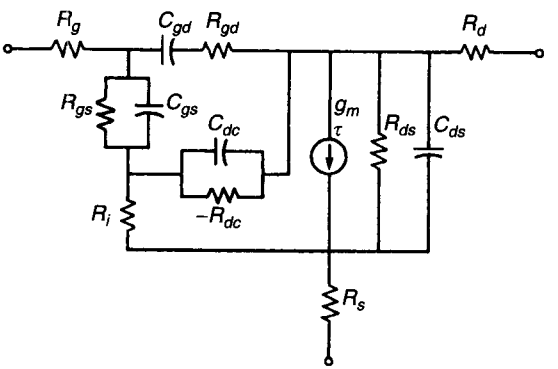


Figure 7.21 A T-equivalent circuit for MESFET shown in standard form.

$$g_m = -\left.\frac{\partial I_d}{\partial V_{gs}}\right|_{V_{ds}} = \frac{g_0 E_s Z}{2sW_{00}} \frac{\partial p}{\partial s} \tag{7.81}$$

When this definition is applied to the expressions previously derived, the transconductance is written for short-gate-length devices (i.e., $L \cong L_2$, $L_1 \to 0$, and $s \cong p$) as

$$g_m \cong \frac{I_s}{W_{00}(2p)} = \frac{I_s}{2W_{00}} \left(\frac{1}{1 - I_d/I_s}\right) \tag{7.82}$$

This approximation is quite accurate when $I_d/I_s > 0.1$. Note that the transconductance is proportional to the ratio I_d/I_s and also to the value of I_s.

Channel Resistance. The channel resistance R_{ds} (also commonly designated by R_0) is determined from

$$R_{ds} = -\left.\frac{\partial V_{ds}}{\partial I_d}\right|_{V_{gs}} \tag{7.83}$$

which, when applied to short-gate-length devices, results in the expression

$$R_{ds} \cong \frac{|V_{ds}|\pi W_{00} p}{aI_s E_s} = \pi \frac{W_{00}}{aE_s} \frac{|V_{ds}|}{I_s} \left(1 - \frac{I_d}{I_s}\right) \tag{7.84}$$

Note that R_{ds} is independent of gate length and is linearly dependent upon V_{ds}. Also, R_{ds} is linearly dependent upon I_d.

Gate–Source Capacitance. The gate–source capacitance C_{gs} is the rate of change of the free charge on the gate electrode with respect to the gate bias voltage when the drain potential is held constant. This definition actually includes the gate–source and gate–drain capacitances. If $C_{gd} \ll C_{gs}$ (typically $C_{gd} \cong 0.1 C_{gs}$ for microwave FETs), the definition for C_{gs} can be used. Therefore,

$$C_{gs} \cong \left.\frac{\partial Q_g}{\partial V_{gs}}\right|_{V_{ds}} \tag{7.85}$$

The gate charge Q_g can be determined from Poisson's equation by integrating the normal component of the electric field under the gate over the gate area. In region I, the $w(x)$ expression is integrated. In region II, the contributions from the drain and depletion regions must be considered. The normal components of the electric field are

$$E_{y\mathrm{I}}(x, a) = \frac{2W_{00}}{a}\left[1 - \frac{b(x)}{a}\right] \quad \text{(region I)} \tag{7.86}$$

$$E_{y\mathrm{II}}(x, a) = \left(\frac{2W_{00}}{a}\right)p + E_s \sinh\left[\frac{\pi(x - L_1)}{2a}\right] \quad \text{(region II)} \tag{7.87}$$

where

$$p = 1 - \frac{b_p}{a}$$

The gate charge is determined from

$$Q_g = 2\epsilon Z\left[\int_0^{L_1} E_{y\mathrm{I}}(x,a)\,dx + \int_{L_1}^{L} E_{y\mathrm{II}}(x,a)\,dx\right] \tag{7.88}$$

Integrating and substituting into the capacitance definition, the gate–source capacitance for short-gate-length FETs and bias voltages not close to W_{00} results in the expression

$$C_{gs} \cong 2\epsilon Z\left(\frac{L}{ap} + 1.56\right) = 2\epsilon Z\left[\frac{L}{a}\left(\frac{1}{1 - I_d/I_s}\right) + 1.56\right] \tag{7.89}$$

The 1.56 term in the expression accounts for fringing area capacitance [29]. The gate depletion region appears as a parallel-plate capacitor with the addition of the fringing area factor.

Drain–Gate and Source–Drain Capacitances. The capacitances C_{gd} and C_{ds} are considered parasitic parameters to the first order in this model. They are not intrinsic to the gain mechanism of the FET and are due to fringing fields of the contact electrodes. Expressions for these elements can be obtained from an analysis of strip conductors on a dielectric medium. Expressions for these capacitances have been determined [26] to be

$$C_{gd}, C_{ds} = (\epsilon_r + 1)\epsilon_0 Z \frac{K(1 - k^2)^{1/2}}{K(k)} \tag{7.90}$$

where $K(k)$ is the complete elliptic integral of the first kind and

$$k_{gd} = \left[\frac{L_{gd}}{L_{gd} + L}\right]^{1/2} \tag{7.91}$$

$$k_{ds} = \left[\frac{(2L_s + L_{ds})L_{ds}}{(L_s + L_{ds})^2}\right]^{1/2} \tag{7.92}$$

where L_{gd} and L_{ds} are the separation between gate–drain and drain–source electrodes, respectively. In these expressions it is assumed that $L_d \gg L$, $L_s = L_d$, and the contact areas are large compared to the electrode spacings. Note that C_{gd} and C_{ds} are constant and not functions of potential. This is, of course, an approximation and only accurate in certain applications. This

approximation is not generally valid when the switching characteristics of these devices are considered.

Gate–Source Capacitance Charging Resistor. The charging resistor R_i is difficult to relate to bulk material parameters in the device. It represents the time required to charge the gate depletion region capacitance. A suitable expression can be determined from time-constant arguments. For example, the gate–source capacitance charging time can be expressed as

$$\tau = R_i C_{gs} \tag{7.93}$$

For short-gate-length devices it is known that

$$C_{gs} = \epsilon Z \frac{L}{ap} \tag{7.94}$$

The time constant τ can be determined from the time required for the electrons to travel the length of the gate region. Therefore,

$$\tau = \int_0^L \frac{dx}{v(x)} = \int_0^{L_1} \frac{dx}{\mu E(x)} + \int_{L_1}^L \frac{dx}{v_s} \tag{7.95}$$

For short-gate devices at moderate to high drain potentials the first integral becomes small, since most of the gate region is in velocity saturation. For a first approximation, we can neglect the first integral. Therefore,

$$\tau \cong \frac{L - L_1}{v_s} \cong \frac{L}{v_s} \tag{7.96}$$

$$R_i = K' \frac{ap}{\epsilon Z v_s} \tag{7.97}$$

where K' is a proportionality constant chosen to match actual devices. For typical 1-μm-gate-length FETs

$$K' \sim 1.65 \times 10^{10}\ \Omega\text{-m/s}$$

7.5.4 Equivalent Circuit and Figures of Merit

The GaAs MESFET equivalent circuit is shown in Fig. 7.21. This circuit can be analyzed to determine the performance of the device. For purposes of analysis it is common to reduce the equivalent circuit to a simplified form, retaining only the most significant elements. The circuit reduces as shown in Fig. 7.22. Note that all elements that provide coupling from the output loop to the input loop have been removed. This simplification produces a circuit that is easily analyzed but does not necessarily accurately predict device performance, espe-

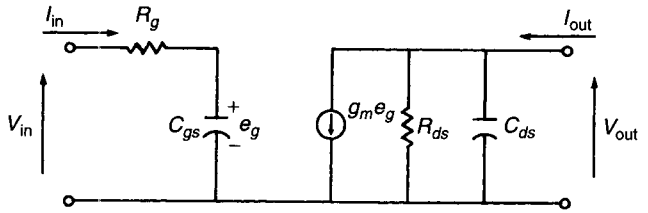


Figure 7.22 Simplified equivalent circuit for MESFET. Only elements of first-order importance are included.

cially at microwave frequencies. The simplified circuit produces the gain–bandwidth product f_T and the maximum frequency of oscillation $f_{\max}$, which are commonly used as figures of merit. These parameters will be derived and then the complete circuit will be analyzed to show the significance of the feedback elements.

The short-circuit current gain is calculated by placing a short circuit on the output and calculating the current gain from the expression

$$h_{21} = \frac{I_{\text{out}}}{I_{\text{in}}} = \frac{g_m}{j\omega C_{gs}} \tag{7.98}$$

The frequency at which the magnitude of h_{21} is reduced to unity is defined as the gain–bandwidth product,

$$f_T = \frac{g_m}{2\pi C_{gs}} \tag{7.99}$$

For short-gate-length devices, the transconductance and gate–source capacitance can be written as

$$g_m = \frac{I_s}{W_{00}(2p)} = \frac{2\epsilon Z v_s}{ap} \tag{7.100}$$

$$C_{gs} = 2\epsilon Z\left(\frac{L}{ap} + 1.56\right) \tag{7.101}$$

If the fringing-field term is neglected, these expressions can be substituted into the f_T expression to get

$$f_T = \frac{v_s}{2\pi L} \tag{7.102}$$

The f_T of the device is directly proportional to the saturated velocity of the charge carriers under the gate and inversely proportional to the gate length.

Since high f_T is desirable for high-frequency performance, short gate lengths and high-velocity semiconductors are required.

Power Gain and f_{max}. The power gain of the device can be calculated from the admittance matrix for the device,

$$\begin{bmatrix} I_{in} \\ I_{out} \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} V_{in} \\ V_{out} \end{bmatrix} \tag{7.103}$$

The device circuit couples energy from an external source to an external load. The unilateral power gain is defined as the ratio between the power delivered to the load over the power available from the source when the feedback network is tuned so that no feedback occurs and the input and output are simultaneously and conjugately matched.

The unilateral power gain expression can be applied to the simplified equivalent circuit in Fig. 7.22. It follows that the admittance matrix is

$$\begin{aligned} y_{11} &= \frac{1}{R_g + 1/(j\omega C_{gs})} \\ y_{21} &= \frac{g_m}{1 + j\omega C_{gs} R_g} \\ y_{22} &= \frac{1}{R_{ds}} + j\omega C_{ds} \\ y_{12} &= 0 \end{aligned} \tag{7.104}$$

Substituting into the unilateral gain definition, it follows that

$$U = \frac{1}{4}\left(\frac{f_T}{f}\right)^2 \frac{R_{ds}}{R_g} \tag{7.105}$$

The maximum frequency of oscillation, f_{max}, is defined as the frequency at which the unilateral power gain is reduced to unity and is given as

$$f_{max} = \frac{f_T}{2}\sqrt{\frac{R_{ds}}{R_g}} \tag{7.106}$$

The f_{max} frequency separates the active and passive regions for the network. Above f_{max} the network is passive and incapable of amplifying RF energy. The unilateral gain can be written in terms of f_{max} as

$$U = \left(\frac{f_{max}}{f}\right)^2 \tag{7.107}$$

and to increase the gain, it is necessary to maximize $f_{\max}$. Also, from the U expression the importance of a large channel resistance R_{ds} and a low input resistance R_g for obtaining optimum gain is seen.

This single-pole model predicts that the unilateral gain will decrease with frequency at a rate of 6 dB/octave or 20 dB/decade. At low frequencies (i.e., $f \ll f_{\max}$), this approximation is quite accurate. At higher frequencies, the other elements in the equivalent circuit become important and add complexity to the calculation. For example, when the complete equivalent circuit is analyzed, it is found that the unilateral gain decreases at a 6-dB/octave rate at low frequencies, as predicted from the simple analysis, but the rolloff rate increases to about 12 dB/octave at higher frequencies. It has been noted by Das [30] that parasitic common lead elements produce the increased unilateral gain rolloff. Steer and Trew [31] and Trew and Steer [32] have shown that internal feedback, caused by the charge dipole domain that forms in the channels of field-effect-type transistors, produces a complex pole in the unilateral gain, and above the pole frequency the gain decreases at a 12-dB/octave rate. If the complete equivalent circuit of Fig. 7.21 is analyzed, an approximation for the unilateral gain is found to be

$$U \cong \left[\frac{g_m^2 R_{ds}}{4C_{gs}R_i(C_{gs} - C_{dc}g_m R_{ds})}\right]\left[\frac{1}{\omega^2(1 - p^2\omega^2)}\right] \tag{7.108}$$

where

$$p^2 = \frac{(R_i^2 C_{gs})(C_{dc} + C_{gs})^2 + C_{dc}g_m R_{ds}(\tau^2/2)}{C_{dc}g_m R_{ds} - C_{gs}} \tag{7.109}$$

This calculation includes the elements listed in Table 7.2 and indicates the significance of the domain capacitance (C_{dc}) in establishing the additional pole.

Table 7.2 Equivalent Circuit Parameters for $\frac{1}{2}$-μm-Gate-Length Millimeter-Wave GaAs MESFET

Element	Value
R_s	4.55 Ω
R_g	1.46 Ω
R_d	6.7 Ω
R_i	2.69 Ω
R_{ds}	556 Ω
C_{gs}	0.071 pF
C_{gd}	0.001 pF
C_{ds}	0.025 pF
C_{dc}	0.011 pF
g_m	15.2 mS
τ	1.25 ps

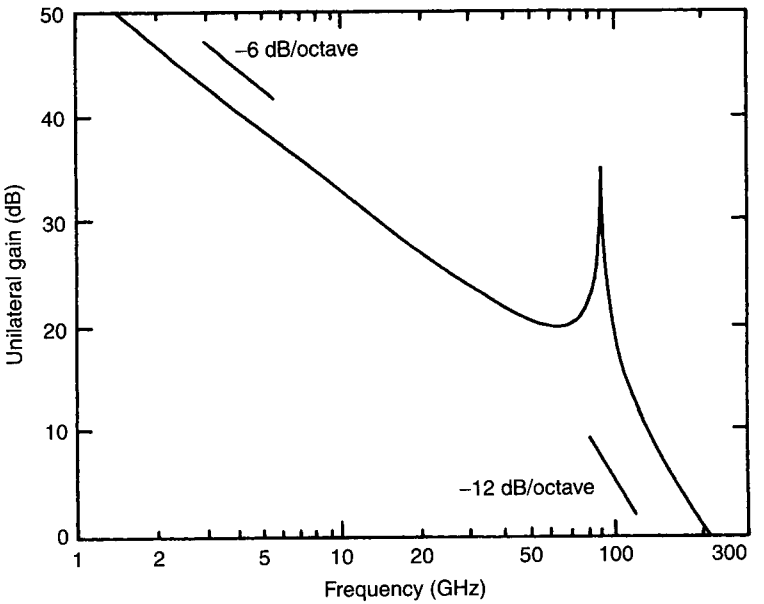


Figure 7.23 Unilateral gain calculated from complete equivalent circuit shown in Fig. 7.21.

Using the parameter values listed in Table 7.2 for a millimeter-wave GaAs MESFET, the unilateral gain performance shown in Fig. 7.23 results. Note the 12-dB/octave rolloff at the higher frequencies. If the domain capacitance is removed from the full circuit, the unilateral gain decreases at a 6-dB/octave rate throughout the frequency range. It should be noted that this analysis does not consider distributed effects, and these effects will result in even greater rolloff rates. When these effects and the effects of lead parasitics are included, rolloff rates greater than 12 dB/octave can occur.

7.5.5 Noise Figure Analysis

The noise figure performance of GaAs MESFETs is most readily calculated using the equivalent circuit shown in Fig. 7.24, where shunt current noise generators are located at the input and output ports. This representation is particularly meaningful for GaAs MESFETs, since the shunt generators represent noise sources at the gate, source, and drain electrodes. The noise source at the output represents the short-circuit drain–source channel noise. The drain current generator is defined in mean-square terms as [33]

$$\overline{i_{nd}^2} = 4kT_0\,\Delta f g_m\,P \tag{7.110}$$

where k is Boltzmann's constant, T_0 is the lattice temperature, Δf is a bandwidth, g_m is the device transconductance, and P is a dimensionless drain noise

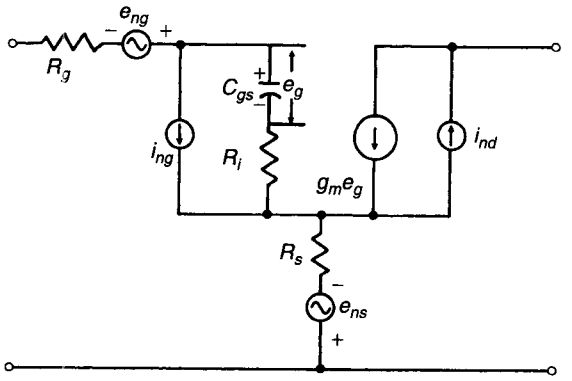


Figure 7.24 Noise equivalent circuit for GaAs MESFET. Noise sources i_{ng}, i_{nd}, e_{ng}, and e_{ns} represent induced gate noise, drain circuit noise, and thermal noise of gate and source resistances.

factor that depends upon device geometry and bias conditions. For zero drain–source voltage $P = (R_{ds}g_m)^{-1}$ and P usually has a numerical value between 1 and 3. The gate noise generator is defined as [34]

$$\overline{i_{ng}^2} = 4kT_0\,\Delta f\,\frac{\omega^2 C_{gs}^2 R}{g_m} \tag{7.111}$$

where R is a dimensionless gate noise factor that depends upon bias. For zero drain–source voltage, $R \sim g_m R_i$.

The gate and drain current generators are capacitively coupled and, therefore, partially correlated. The degree of correlation can be expressed by a factor C [34], where

$$jC = \frac{\overline{i_{ng}^* \cdot i_{nd}}}{\sqrt{\overline{i_{ng}^2} \cdot \overline{i_{nd}^2}}} \tag{7.112}$$

The star stands for the complex conjugate and j is the imaginary representation. The magnitude of C is generally about 0.8 or smaller.

The minimum noise figure for the intrinsic MESFET can be expressed [26] as

$$F_{\min} = 1 + 2\sqrt{PR(1-C^2)}\,\frac{f}{f_T} + 2g_m R_i P\left(1 - C\sqrt{\frac{P}{R}}\right)\left(\frac{f}{f_T}\right)^2 \tag{7.113}$$

This expression can be applied to the equivalent circuit in Fig. 7.24 to obtain

$$F_{\min} = 1 + 2\left(2\pi f\frac{C_{gs}}{g_m}\right)\{K_g[K_r + g_m(R_g + R_s)]\}^{1/2} + 2\left(2\pi f\frac{C_{gs}}{g_m}\right)[K_g g_m(R_g + R_s + K_c R_i)] \tag{7.114}$$

where

$$K_g = P\left[\left(1 - C\sqrt{\frac{R}{P}}\right)^2 + (1 - C^2)\frac{R}{P}\right]$$

$$K_r = \frac{R(1 - C^2)}{(1 - C\sqrt{R/P})^2 + (1 - C^2)R/P}$$

$$K_c = \frac{1 - C\sqrt{R/P}}{(1 - C\sqrt{R/P})^2 + (1 - C^2)R/P}$$

This expression has been simplified by Fukui [35], and for devices operating at frequencies below f_T the minimum noise figure can be approximated as

$$F_{\min} \cong 10 \log[1 + KfL\sqrt{g_m(R_g + R_s)}] \tag{7.115}$$

where L is the gate length and K is a fitting factor selected to match experimental data. Generally, for 1-μm-gate-length devices K varies from 0.25 to 0.3 and a value of 0.27 is found to be generally valid. This expression indicates that low minimum noise figures are obtained with short-gate-length devices that have low parasitic resistance. This expression agrees well with experimental data, as shown in Fig. 7.25, where the noise measure and noise figure are shown for devices of various gate lengths [36].

The minimum noise figure can be defined in terms of practical device parameters. The device transconductance can be expressed as [37]

$$g_m \cong KZ\left(\frac{N}{aL}\right)^{1/3} \tag{7.116}$$

where Z is the gate width in millimeters, a is the channel thickness in micrometers, N is the channel doping in 10^{16} cm^{-3}, and K is the fitting factor previously described. The gate resistance can be written as

$$R_g \cong 17\frac{z^2}{ZhL} \tag{7.117}$$

where z is a unit gate width in millimeters, Z is the actual gate width in millimeters, h is the gate metallization height in micrometers, and L is the gate length in micrometers.

For a recessed gate device the source resistance can be expressed as a sum of three component resistances as

$$R_s = R_1 + R_2 + R_3 \tag{7.118}$$

where R_1 represents the source contact resistance and R_2 and R_3 are resistances of the semiconductor between the source and gate electrodes. The resistance R_2 is calculated for the region outside the recessed area, and R_3 is calculated for the region under the recess. These resistances are given by the expressions

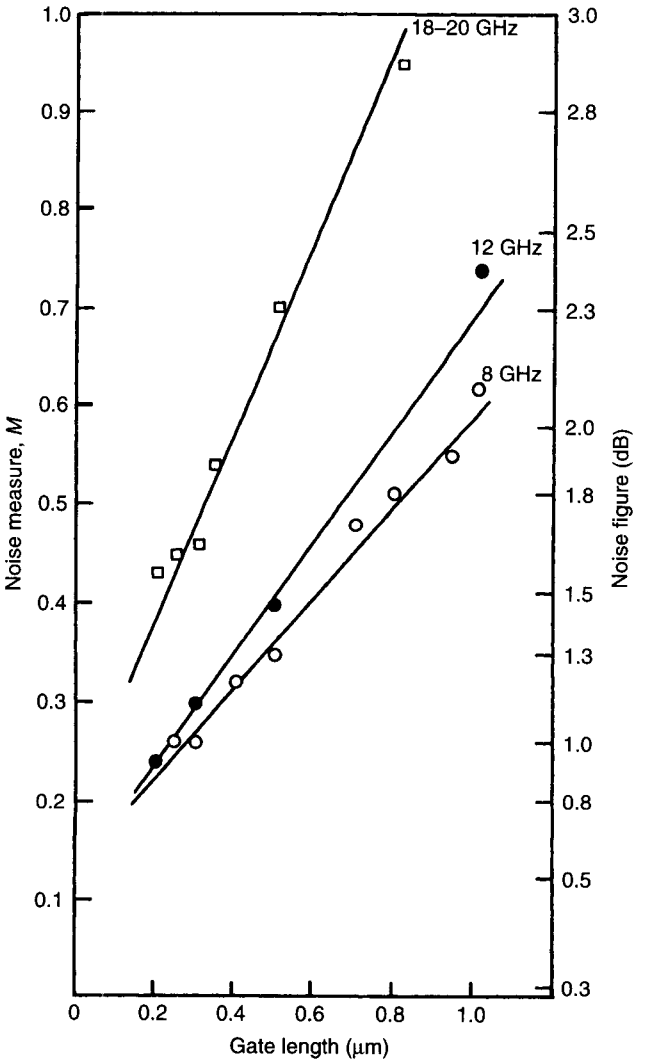


Figure 7.25 Noise measure and noise figure measured for GaAs MESFETs as a function of gate length (After Goronkin and Nair. Reprinted with permission of IEEE.)

$$R_1 \cong \frac{2.1}{Za_1^{0.5}N_1^{0.66}} \tag{7.119a}$$

$$R_2 \cong \frac{1.1L_2}{Za_2N_2^{0.82}} \tag{7.119b}$$

$$R_3 \cong \frac{1.1L_3}{Za_3N_3^{0.82}} \tag{7.119c}$$

where a_1 is the channel thickness under the source electrode in micrometers, N_1

is the doping density under the source electrode in 10^{16} cm^{-3}, L_2 and L_3 are the lengths of the respective fractions of the channel between the source and gate electrodes in micrometers, a_2 and a_3 are the thicknesses of their respective regions in micrometers, and N_2 and N_3 are the doping densities of their respective regions in 10^{16} cm^{-3}.

By defining the maximum unit gate width as the limit when R_g exceeds R_s, an expression for the minimum noise figure in a form suitable for practical device design is obtained. The expression is

$$F_{\min} = 1 + fK\left(\frac{NL^5}{a}\right)^{1/6} \times \left[\frac{17z^2}{hL}(1+s) + \frac{2.1}{a_1^{0.5}N_1^{0.66}} + \frac{1.1L_2}{a_2N_2^{0.82}} + \frac{1.1L_3}{a_3N_3^{0.82}}\right]^{1/2} \tag{7.120}$$

where $s = 0.08\sqrt{fhL}$ is a factor that accounts for the skin effect on the gate metal. In this expression a noise coefficient factor $K = 0.040$ is found to give good results.

From the circuit perspective it is useful to treat the device as a noisy two-port with the noise sources referenced to the input. Using this formulation, the minimum noise figure can be written in the form [38]

$$F = F_{\min} + \frac{R_n}{R_s}\left[\frac{(R_s - R_{\text{op}})^2 + (X_s - X_{\text{op}})^2}{R_{\text{op}}^2 + X_{\text{op}}^2}\right] \tag{7.121}$$

where R_n is an equivalent noise resistance, R_s and X_s are the signal source impedances, and R_{op} and X_{op} are the signal source impedances that produce the lowest noise figure. Note that when the source impedance is equal to the optimum value, the noise figure is equal to its minimum value. This equation is very useful for circuit design, since it shows the effect of input circuit mismatch upon noise figure.

The minimum noise figure is presented in (7.120) in a form useful for practical design. Fukui [39] has presented expressions for the other terms in a similar form. The expressions are

$$R_n = \frac{40}{Z}\left(\frac{aL}{N}\right)^{1/3} \tag{7.122}$$

$$R_{\text{op}} = 2.2Z^{-1}\left[12.5\left(\frac{aL}{N}\right)^{1/3} + \frac{17z^2}{hL} + \frac{2.1}{a^{0.5}N^{0.66}} + \frac{1.1L_{sg}}{(a-a_s)N^{0.82}}\right] \tag{7.123}$$

$$X_{\text{op}} = \frac{450}{fZ}\left(\frac{a}{NL^2}\right)^{1/3} \tag{7.124}$$

where L_{sg} is the distance between the source and gate electrodes in micrometers. The a_s term is the depletion depth in micrometers at the surface in the source–gate region.

Table 7.3 Comparison of Predicted Value of Optimal Noise Figure from Geometric and Material Parameters with Directly Measured Value for Sample GaAs MESFETs *A–E*

	Parameter Symbol	Unit	*A*	*B*	*C*	*D*	*E*
Predicted by (7.120)	F_{min}	dB	1.72	1.80	2.12	1.56	1.70
Measured directly	F_{min}	dB	1.75	1.76	2.22	1.51	1.74

As indicated by (7.121), the device noise figure is a sensitive function of input impedance matching. In order to decrease this sensitivity, it is necessary to decrease the magnitude of the equivalent noise resistance R_n. This is accomplished by designing the device with a short gate length and high channel doping. A comparison of five low-noise designs based upon these calculations is presented in Tables 7.3 and 7.4. The device dimensions are indicated in Fig. 7.26.

7.5.6 Arbitrary Doping Profile Model and Deep Levels

The two-region model discussed in Section 7.5.2 has been extended by Golio and Trew [40], so that arbitrary doping profile devices and the effects of deep levels in the conducting channel can be investigated. The noise analysis technique presented by Pucel et al. [26] has been applied to the general profile

Table 7.4 Design Parameters of Five Representative GaAs MESFETs (*a–e*) Used for Calculation of Optimal Noise Figure as Function of Frequency as Shown in Fig. 7.26

Parameter Symbol	Unit	*a*	*b*	*c*	*d*	*e*
L	μm	0.9	0.9	0.5	0.5	0.25
L_g	μm	0.9	1.2	0.8	0.8	0.4
L_2	μm	1.0	0.75	0.75	0.75	0.4
L_3	μm	0	0.4	0.3	0.3	0.2
h	μm	0.5	1.0	0.65	0.65	0.4
N	10^{16} cm^{-3}	7	4	8	8	18
N_1	10^{16} cm^{-3}	7	200	200	200	200
N_2	10^{16} cm^{-3}	7	200	200	200	200
N_3	10^{16} cm^{-3}	—	4	8	8	18
a	μm	0.3	0.27	0.15	0.15	0.1
a_1	μm	0.3	0.15	0.15	0.15	0.15
a_2	μm	0.17	0.12	0.12	0.12	0.12
a_3	μm	—	0.27	0.15	0.15	0.1
z	mm	0.25	0.25	0.25	0.1	0.065
z_m	mm	0.24	0.23	0.14	0.14	0.065

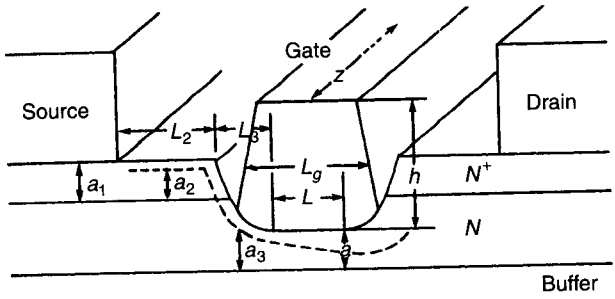
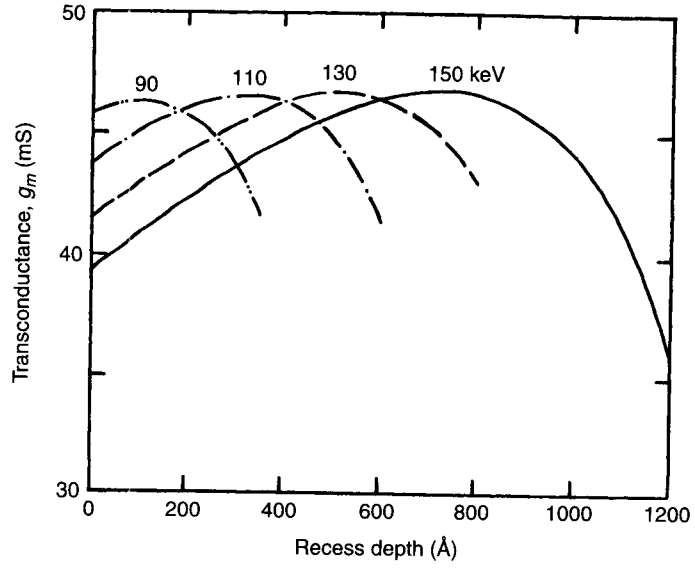


Figure 7.26 Device geometry for recessed-gate GaAs MESFET showing device dimensions used in noise figure calculations in text.

model by Trew et al. [41]. An investigation of recessed-gate, ion-implanted devices shows that an optimum recess depth exists for a given channel doping. For example, the device transconductance and gate–source capacitance are shown in Fig. 7.27 for devices fabricated with various implants and recess depths. Each implanted profile has the same peak doping density. The maximum transconductance is independent of implant energy but is obtained at a critical recess depth. The optimum recess depth occurs when the gate depletion region is able to exercise maximum control over the channel current, and this occurs with a gate length–channel thickness ratio of about 3. Thicker channels decrease the modulation efficiency, and thinner channels decrease the channel current.

Although deep levels generally charge and discharge with time constants too long to directly affect the RF performance of microwave devices, they influence the gain and noise performance of MESFETs through an indirect process. They degrade the charge carrier transport characteristics by introducing additional scattering centers. This effect is illustrated in Fig. 7.28 where the electron mobility for an ion-implanted channel is shown for various deep-level densities. The degree of degradation increases rapidly as the deep-level density approaches the channel donor density.

The gain and noise figure performances of GaAs MESFETs are degraded by deep levels, as shown in Fig. 7.29. This figure presents the associated gain and minimum noise figure calculated for devices fabricated with various channel implants. Each device has the same peak doping density and is recessed so that all devices have the same I_{dss}. The data are shown for a frequency of 8 GHz. This figure shows the associated gain to be independent of implant energy but to be reduced by the deep-level density. The minimum noise figure decreases with implant energy for low deep-level density but increases with implant energy for high deep-level density. This indicates that the lowest noise figures will be obtained from devices fabricated from high-quality material (i.e., low deep-level density) and relatively high implant energy. Deep recess depths are indicated.



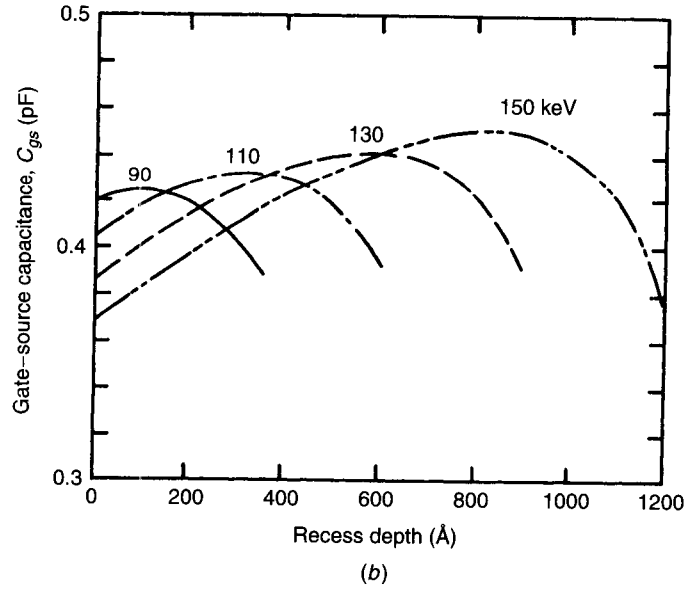


Figure 7.27 (*a*) Device transconductance vs. recess depth for devices fabricated with various implant energies. (*b*) Gate–source capacitance vs. recess depth for GaAs MESFETs fabricated with various implant energies.

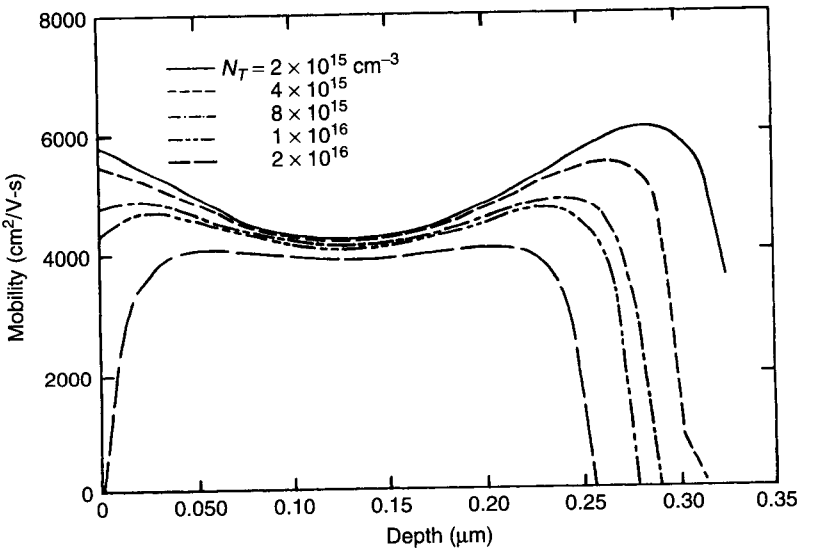


Figure 7.28 Channel electron mobility as function of depth and deep-level density. Depth is measured from channel–gate metal interface into conducting channel.

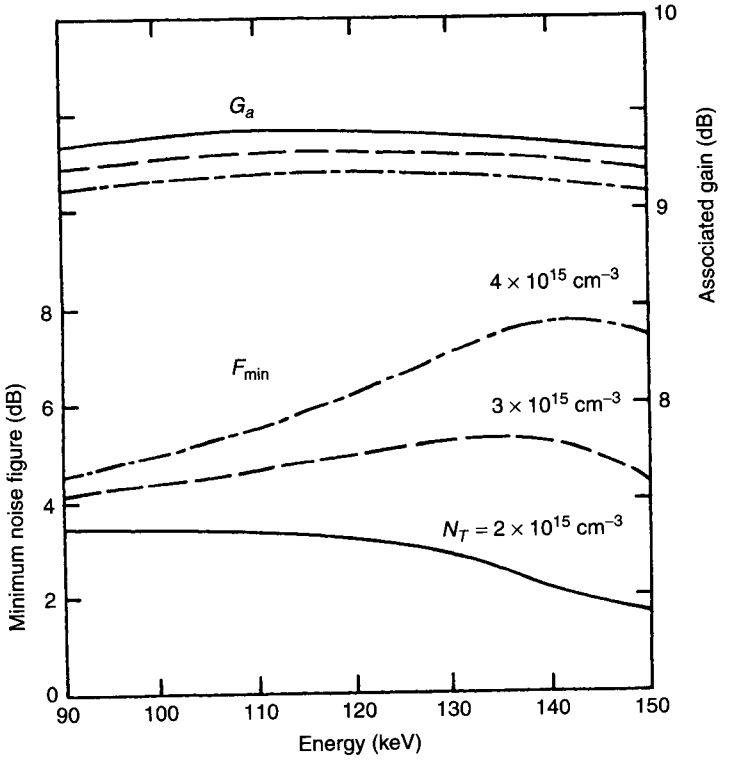
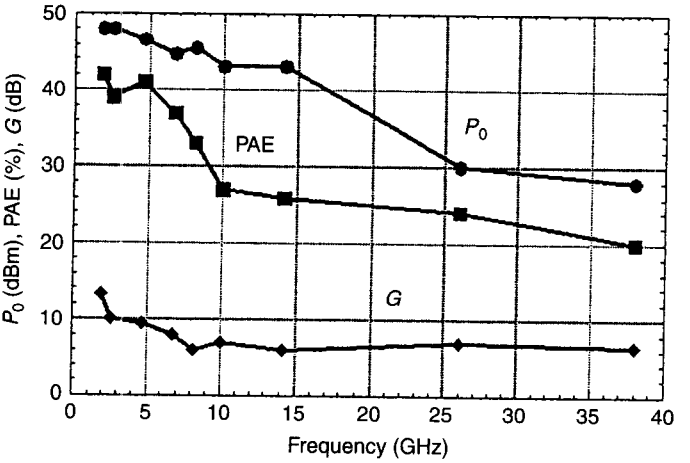


Figure 7.29 Gain and noise figure vs. implant energy. Peak doping for all devices is 2×10^{17} cm^{-3}, devices are biased at optimum low-noise bias, $I_{dss} = 30$ mA, and frequency is



8 GHz.

7.5.7 Power Field-Effect Transistors

Field-effect transistors are very useful as power devices. Transistors can be optimized for maximum RF output power or maximum power-added efficiency (PAE), although it is difficult to obtain both conditions simultaneously. The RF performance available from commercial FETs up to the *Ka* band is shown in Fig. 7.30. Power FETs producing 80 W RF output power at 3 GHz and on the order of 1 W at the *Ka* band are available. Power-added efficiency on the order of 40% is obtained at the *S* band. However, due to power cell combining losses and phase matching considerations, PAE is reduced to below 30% at the *X* band and 20% at the *Ka* band. Gain is typically on the order of 6 dB. The high RF output power shown in Fig. 7.30 is generally only possible over relatively narrow bandwidth, and it is difficult to broadband power devices due to low input impedance.

In order to obtain high output RF power, it is necessary to operate the FET at high drain–source voltage and current. This requirement has implications for the design of power devices. In order to support high drain bias voltage, it is necessary that the structure have a high gate–drain breakdown voltage. This suggests low channel doping density and large gate–drain spacing. Low channel doping density, however, is inconsistent with high channel current, and large gate–drain spacing increases parasitics and limits operation frequency. The spacing criterion is addressed with the use of recessed-gate structures. For example, the basic power FET structures are shown in Fig. 7.31. The flat channel device has a low breakdown voltage. A simple recess of the gate into the channel increases the gate–drain spacing and results in an increase in breakdown voltage [42]. Frensley [43] has presented a good analysis of the power-limiting effects of breakdown. The RF performance principles of power FETs, including the effects of RF breakdown, have been presented by Winslow and Trew [44].

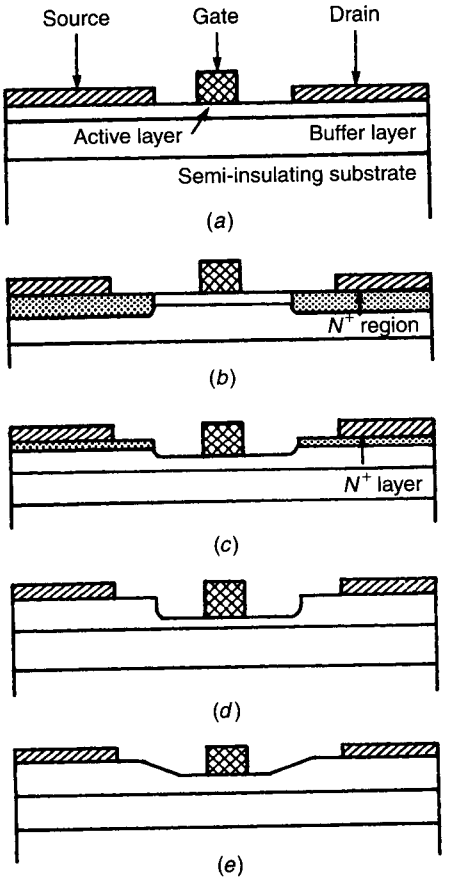


Figure 7.30 RF performance of commercial power FETs.

It is known that breakdown occurs in the high-field domain that forms in the gate–drain region of the channel. A smooth electric field transition in this region results in an increase in the breakdown voltage, and this situation can be achieved with the graded recess structure. Breakdown voltages in excess of 20 V can be obtained with this structure. Excellent performance has also been obtained with an n^+ contact ledge structure with a double recess. An investigation of this structure [45] shows that the gate recess should be as narrow as possible. The wide recess should be 0.5–0.7 μm wider than the gate and the gate recess depth should be about 0.05–0.09 μm. Devices of this type produce more than 1 W/mm of gate width at the X band with high gain (~10 dB) and efficiency (~40%).

The breakdown condition limits the extent to which doping can be increased to achieve high channel current. Other ways to increase the current are required. One technique is to increase the thickness of the channel. This method, however, decreases the modulation efficiency of the gate and the gain of the

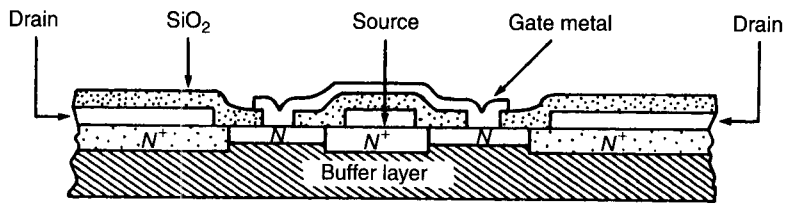


Figure 7.31 Basic power GaAs MESFET geometries.

device. Typical channel thickness is limited to the range of 0.2–0.4 μm at the *X* band, and the thickness must be scaled inversely with frequency. The most effective method for increasing channel current is to increase the gate width. At the *X* band, gate widths in the range of 0.4–4 mm are typically used, and gate widths in the range of 20–40 mm are possible. Gate widths of this magnitude are long relative to a wavelength, and in order to avoid distributed effects, multiple gate feeds and power-combining techniques are used. Generally, the gate width that an individual gate feeds should be less than a tenth of a wavelength, and at the *X* band this limits the gate width cell to the range of 150–300 μm. Higher frequency devices, of course, require reduced gate width per gate feed.

Various power-combining techniques have been used. The most commonly used technique involves the direct paralleling of multiple gate feeds, or fingers. This technique requires multiple bond wires or crossover structures, as shown in Fig. 7.32. The crossover is achieved with a metal that connects the various gate fingers. The metal must be isolated from the source contacts, and this is achieved by depositing the metal over a dielectric or air dielectric. The air bridge crossover is fabricated by depositing the metal over a photoresist layer that has been deposited over the source contact and channel region, excluding the gate contact. The air bridge results when the photoresist is dissolved and removed. The air bridge is preferred to the dielectric overlay due to reduced parasitic capacitance. From a circuit perspective, wide gate widths produce low input and output impedances, and this makes it difficult to design broadband power amplifiers.

Heat dissipation is important in power FET design, and adequate heat sinking must be provided. Various techniques have been investigated for reducing the thermal resistance of the device and mounting structure so that the temperature rise of the channel can be limited. The heat dissipation problem in power FETs is increased by the poor thermal conductivity of GaAs. This fundamental problem can be addressed by thinning the semi-insulating substrate as much as possible. The device can also be flip-chip mounted, so that the main heat flow path does not pass through the substrate. This technique is difficult to apply in many situations. The most effective technique for heat sinking involves the use of via holes with thin substrate chips. The via holes are fabricated by etching holes through the substrate under the source contacts and then filling

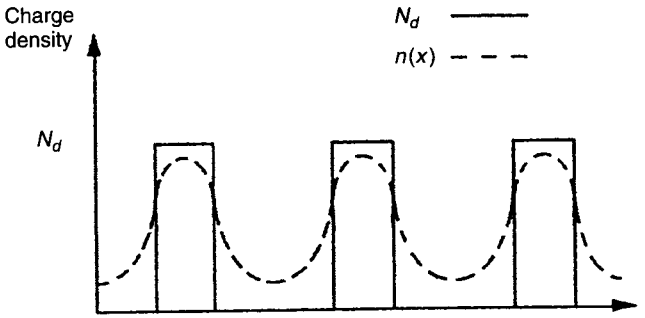


Figure 7.32 Power GaAs MESFET structure using overlays to connect gate fingers.

the holes with a metal. When used with plated heat sink technology this technique produces devices with low thermal resistance. It also produces low parasitic source resistance and inductance.

7.6 HEMTs

Field-effect transistors based upon heterojunctions can also be fabricated. These devices make use of the modulation doping principle proposed by Esaki and Tsu in 1969 [46], as shown in Fig. 7.33. Free charge from the high-doped regions diffuses into the low-doped regions where it is able to flow with high mobility due to the lack of impurity scattering in the low-doped regions. High current results. If the high-doped region is fabricated from a semiconductor with a wider bandgap than the low-doped regions and the discontinuity in the energy bands is restricted to the conduction band, a quantum well is created in the conduction bands at the interface between the two semiconductors. As electrons from the wide-bandgap semiconductor diffuse into the quantum well, a two-dimensional electron gas (2DEG) is created, as shown in Fig. 7.34.

The concept was demonstrated by Stormer et al. in 1979 [47]. The 2DEG can be used to form the channel region for a FET, as shown in Fig. 7.35. The resulting transistor is called a high-electron-mobility transistor (HEMT) and was demonstrated by Mimura et al. in 1980 [48]. HEMTs have extremely high frequency performance capability and very low noise performance, primarily due to the very high mobility characteristics of the 2DEG. These devices are also used for microwave and millimeter-wave power applications and above the X band are superior to MESFETs.

7.6.1 HEMT Model

A small-signal equivalent circuit for a HEMT can be derived in the same manner as for the MESFET. Starting from the basic semiconductor equations,

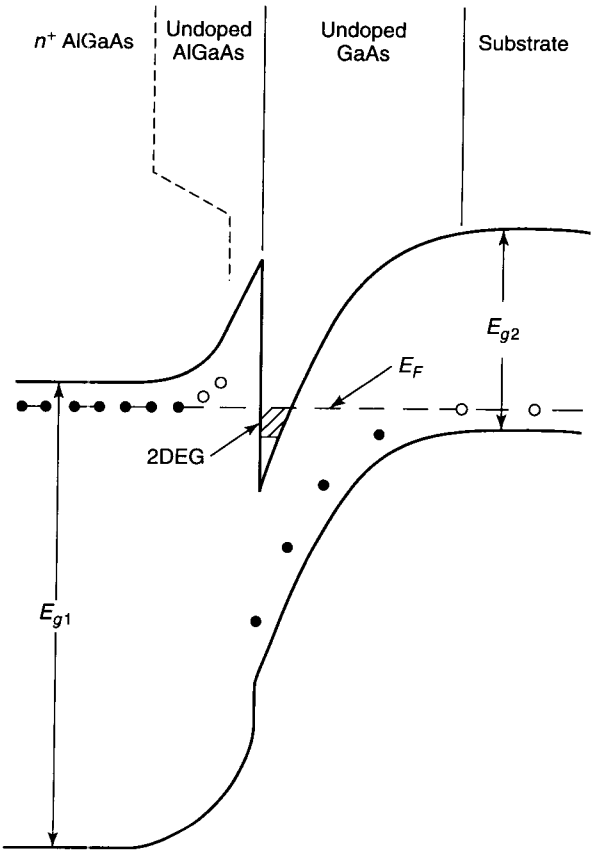


Figure 7.33 Modulation doping principle.

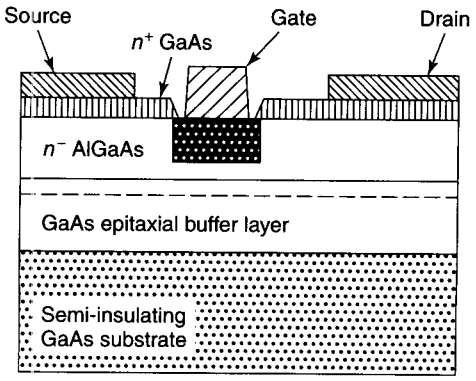


Figure 7.34 Formation of a 2DEG at hetero-interface between wide-bandgap and narrow-bandgap semiconductors.

the fundamental approach is to develop an expression for the channel current as a function of applied voltage. The current expression can then be appropriately differentiated to derive expressions for the equivalent circuit element values such as C_{gs}, g_m, and so on.

The 2DEG that forms at the interface between the AlGaAs and GaAs, as shown in Fig. 7.34, consists of electron charge confined in discrete energy levels. If a two-energy-level model is assumed, the charge in the 2DEG can be written as

$$Q_s = -q \int_0^d n_s \, dy \tag{7.125}$$

where d is the thickness of the AlGaAs layer and n_s is the charge at the AlGaAs–GaAs interface and is expressed as

$$n_s = \frac{DkT}{q} \ln[(1 + e^{(q/kT)(E_f - E_0)})(1 + e^{(q/kT)(E_f - E_1)})] \tag{7.126}$$

were E_0 and E_1 are the positions of the first two allowed energy levels in the interface charge quantum well and

$$D = \frac{qm^*}{\pi h^2} \qquad E_0 = \gamma_0 n_s^{2/3} \qquad E_1 = \gamma_1 n_s^{2/3} \tag{7.127}$$

The terms γ_0 and γ_1 are constants that can be theoretically derived or measured from Shubnikov–DeHaas measurements. Measurements indicate values of

$$\gamma_0 = 2.5 \times 10^{-12} \text{ V} \cdot \text{m}^{4/3} \qquad \gamma_1 = 3.2 \times 10^{-12} \text{ V} \cdot \text{m}^{4/3}$$

Also, the constant D has been measured to have a value

$$D = 3.24 \times 10^{17} \text{ m}^{-2}/\text{V}$$

The charge in the 2DEG arises from depletion from the AlGaAs and can be calculated from the expression

$$n_s \frac{\epsilon_2}{qd} \left[V_g - \left(\varphi_b - V_{P2} + \frac{E_{f1}}{q} - \frac{\Delta E_c}{q} \right) \right] \tag{7.128}$$

where

$$V_{P2} = \frac{qN_d d_d^2}{2\epsilon_2} \tag{7.129}$$

Simultaneous solution of these equations yields an expression for the 2DEG charge density as a function of distance along the channel. The resulting expression is

$$n_s(x) = \frac{\epsilon_2}{qd}[V_g - V_T - V(x)] \tag{7.130}$$

where $V(x)$ is the channel potential and

$$V_T = \varphi_b - \frac{\Delta E_c}{q} + \frac{E_{f1}}{q} - V_{P2} \tag{7.131}$$

The channel current is given by the expression

$$I_d = Wqn_s(x)v(x) \tag{7.132}$$

The channel current dependence upon channel voltage is contained within the $n_s(x)$ and $v(x)$ terms. After some manipulation, the following equation for current as a function of voltage is obtained:

$$I_d \cong \frac{W\mu_0\epsilon_2}{2(d+\Delta d)} \frac{[(V_g - V_{T0})^2 - (V_g - V_{T0} - V_d)^2]}{(L_g + V_d/E_c)} \tag{7.133}$$

where

$$V_{T0} = \varphi_b - \frac{\Delta E_c}{q} + \frac{\Delta E_{f1}(T)}{q} - V_{P2} \tag{7.134}$$

$$\Delta d = \frac{qa}{\epsilon_2} \tag{7.135}$$

The equivalent circuit element values as a function of gate–source voltage are shown in Fig. 7.36. The gate–source voltage can vary from pinch-off to forward bias. At pinch-off the negative voltage applied to the gate electrode is sufficient to force all charge from the 2DEG, thereby eliminating it. For forward bias more charge is injected into the 2DEG, thereby increasing its density. The channel current, represented by $I_{d,\text{sat}}$ in Fig. 7.36 follows the channel 2DEG density and varies from 0 at pinch-off to a maximum under forward bias. The transconductance has minimum values near pinch-off and for forward bias. Transconductance typically has a maximum for slight reverse bias on the gate. The gate–source capacitance tends to low values near pinch-off, with a broad region for moderate bias where C_{gs} is only weakly dependent upon gate–source voltage. For forward bias, C_{gs} increases due to the increased 2DEG charge density.

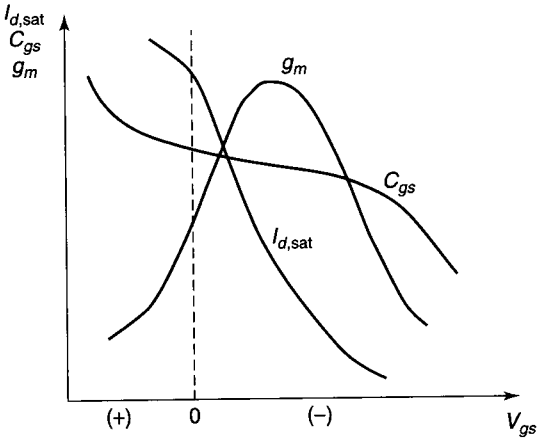


Figure 7.35 HEMT structure.

7.6.2 Noise Performance

High-electron-mobility transistors have excellent noise performance with very low noise figures. The excellent noise performance results primarily from low channel and source resistance due to the high 2DEG channel charge density and very high electron mobility. The HEMTs have the lowest noise figures of any transistor and are extensively used in front ends for low-noise receivers for satellite communications, cellular telephones, radiometers, and other applications.

The noise performance of HEMTs optimized for low-noise performance as a function of frequency is shown in Fig. 7.37. As shown, HEMTs produce noise figures on the order of 0.2–0.3 dB at the X band and noise figure approaching

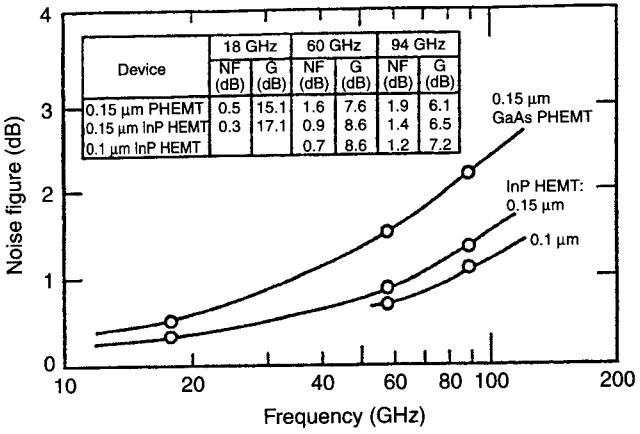


Figure 7.36 $I_{d,sat}$, C_{gs}, and g_m as a function of V_{gs} for a HEMT.

Figure 7.37 Noise figure performance for HEMTs.

slightly greater than 1 dB at 100 GHz. These devices produce excellent low-noise amplifiers for millimeter-wave applications.

7.7 COMPARISON OF BIPOLAR TRANSISTOR AND FET NOISE FIGURES

In low-noise applications FETs are preferred to bipolar transistors. The FET demonstrates a lower noise figure than the bipolar transistor throughout the microwave and millimeter-wave frequency range, and the advantage increases with frequency. This advantage is demonstrated by the comparison of the expressions for the minimum noise figure for the two devices. The bipolar transistor has a minimum noise figure of the form

$$F_{\min} \cong 1 + bf^2\left(1 + \sqrt{1 + \frac{2}{bf^2}}\right) \tag{7.136}$$

where

$$b = \frac{40 I_c r_r}{f_T^2} \tag{7.137}$$

where r_b is the parasitic base resistance and I_c is the collector current. The minimum noise figure for the FET is written in the form

$$F_{\min} \cong 1 + mf \tag{7.138}$$

where

$$m = \frac{2.5}{f_T}\sqrt{g_m(R_g + R_s)} \tag{7.139}$$

Comparing these expressions shows that the minimum noise figure increases with frequency quadratically for bipolar transistors and linearly for FETs. Therefore, the FET demonstrates increasingly superior noise figure performance as compared to bipolar transistors as the operating frequency is increased.

REFERENCES

1. Ino, M. T., T. Ishibashi, and M. Ohmori, "Submillimeter Wave Si p^+pn^+ IMPATT Diodes," *Jpn. J. Appl. Phys.*, Vol. 6, Suppl. 16-1, 1977, pp. 89–92.
2. Nishizawa, J. I., K. Motoya, and Y. Okuno, "GaAs TUNNETT Diodes," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-26, Dec. 1978, pp. 1029–1035.

3. Rodwell, M. J., M. Urteaga, T. Mathew, D. Scottt, D. Mensa, Q. Lee, J. Guthrie, Y. Betser, S. C. Martin, R. P. Smith, S. Jaganathan, S. Krishnan, S. I. Long, R. Pullela, B. Agarwal, U. Bhattacharya, L. Samoska, and M. Dahlstrom, "Submicron Scaling of HBT's," *IEEE Trans. Electron Dev.*, Vol. 48, Nov. 2001, pp. 2606–224.
4. Ning, T. H., "History and Future Perspective of the Modern Silicon Bipolar Transistor," *IEEE Trans. Electron Dev.*, Vol. 48, Nov. 2001, pp. 2485–2491.
5. Cooke, H. F., "Microwave Transistors: Theory and Design," *Proc. IEEE*, Vol. 59, Aug. 1971, pp. 1163–1181.
6. Early, J. M., "Effects of Space-Charge Layer Widening in Junction Transistors," *Proc. IRE*, Vol. 40, Nov. 1952, pp. 1401–1406.
7. Kirk, C. T., "A Theory of Transistor Cutoff (f_τ) Fall-Off at High Current Density," *IRE Trans. Electron Dev.*, Vol. ED-9, Mar. 1954, pp. 164–174.
8. Webster, W. M., "On the Variation of Junction Transistor Current Amplification with Emitter Current," *Proc. IRE*, Vol. 42, June 1954, pp. 914–920.
9. Ebers, J. J., and J. L. Moll, "Large Signal Behavior of Junction Transistors," *Proc. IRE*, Vol. 42, Dec. 1954, pp. 1761–1772.
10. Mason, S. J., "Power Gain in Feedback Amplifiers," *IRE Trans. Circuit Theory*, Vol. CT-1, June 1954, pp. 20–25.
11. Kronquist, R. L., J. Y. Forrier, J. P. Pestil, and M. E. Brilman, "Determination of a Microwave Transistor Model Based on an Experimental Study of Its Internal Structure," *Solid State Electron.*, Vol. 18, Nov. 1975, pp. 949–963.
12. Netzer, Y., "The Design of Low-Noise Amplifiers," *Proc. IEEE*, Vol. 69, June 1981, pp. 728–741.
13. Hawkins, R. J., "Limitations of Nielsen's and Related Noise Equations Applied to Microwave Bipolar Transistors, and a New Expression for the Frequency and Current Dependent Noise Figure," *Solid-State Electron.*, Vol. 20, Mar. 1977, pp. 191–196.
14. Hsu, T. H., and C. P. Snapp, "Low-Noise Microwave Bipolar Transistor with Sub-Half-Micrometer Emitter Width," *IEEE Trans. Electron Dev.*, Vol. ED-25, June 1978, pp. 723–730.
15. Kroemer, H., "Theory of a Wide-Gap Emitter for Transistors," *Proc. IRE*, Vol. 45, Nov. 1957, pp. 1535–1537.
16. Dumke, W. P., J. M. Woodall, and V. L. Rideout, "GaAs–GaAlAs Heterojunction Transistor for High Frequency Operation," *Solid-State Electron.*, Vol. 15, Dec. 1972, pp. 1339–1334.
17. Konagai, M., K. Katsukawa, and K. Takahashi, "(GaAl)As/GaAs Heterojunction Phototransistors with High Current Gain," *J. Appl. Phys.*, Vol. 48, Oct. 1977, pp. 4389–4394.
18. Asbeck, P. M., D. L. Miller, W. C. Petersen, and C. G. Kirkpatrick, "GaAs/GaAlAs Heterojunction Bipolar Transistors with Cutoff Frequencies Above 10 GHz," *IEEE Electron Dev. Lett.*, Vol. EDL-3, Dec. 1982, pp. 366–368.
19. Su, S. L., et al., "Double Heterojunction AlGaAs/GaAs Bipolar Transistors by MBE with a Current Gain of 1650," *IEEE Electron Dev. Lett.*, Vol. EDL-4, May 1983, pp. 130–132.
20. Malik, R. J., J. R. Hayes, F. Capasso, K. Alavi, and A. Y. Cho, "High-Gain AlInAs/GaInAs Transistors Grown by Molecular Beam Epitaxy," *IEEE Electron Dev., Lett.*, Vol. EDL-4, Apr. 1983, pp. 383–386.

21. Song, J. I., K. B. Chough, C. J. Palmstrom, B. P. Van der Gaar, and W. P. Hong, "Carbon-Doped Based InP/InGaAs HBT's with ft = 200 GHz," *IEEE Device Res. Conf. Dig.*, 1994. 52nd, pp. 97–98.
22. Iyer, S. S., G. L. Patton, S. S. Delage, S. Tiwari, and J. M. C. Stork, "Silicon-Germanium Base Heterojunction Bipolar Transistors by Molecular Beam Epitaxy," *Int. Electron Device Meet. Dig.*, 1987, pp. 74–76.
23. Stuetzer, O. M., "A Crystal Amplifier with High Input Impedance," *Proc. IRE*, Vol. 38, Aug. 1950, p. 868.
24. Shockley, W., "Unipolar 'Field Effect' Transistor," *Proc. IRE*, Vol. 40, Nov. 1952, p. 1365.
25. Liechti, C. A., "Microwave Field-Effect Transistors—1976," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-24, June 1976, pp. 279–300.
26. Pucel, R. A., H. Statz, and H. A. Haus, "Signal and Noise Properties of Gallium Arsenide Microwave Field-Effect Transistors," *Adv. Electron. Electron Phys.*, Vol. 38, 1975, pp. 195–265.
27. Golio, J. M., and R. J. Trew, "Compound Semiconductors for Low-Noise Microwave MESFET Applications," *IEEE Trans. Electron Dev.*, Vol. ED-27, July 1980, pp. 1256–1262.
28. Grebene, A. B., and S. K. Ghandhi, "General Theory for Pinched Operation of the Junction Gate FET," *Solid-State Electron.*, Vol. 12, July 1969, pp. 573–589.
29. Wasserstrom, G., and J. McKenna, "The Potential Due to a Charged Metallic Strip on a Semiconductor Surface," *Bell Sys. Tech. J.*, Vol. 49, May/June 1970, pp. 853–877.
30. Das, M. B., and P. Schmidt, "High-Frequency Limitations of Abrupt Junction FETs," *IEEE Trans. Electron Dev.*, Vol. ED-20, Sept. 1973, pp. 779–792.
31. Steer, M. B., and R. J. Trew, "High-Frequency Limits of Millimeter-Wave Transistors," *IEEE Electron Dev. Lett.*, Vol. EDL-7, Nov. 1986, pp. 640–642.
32. Trew, R. J., and M. B. Steer, "Millimeter-Wave Performance of State-of-the-Art MESFET and PBT Transistors," *Electron. Lett.*, Vol. 23, Feb. 1987, pp. 149–151.
33. Vander Ziel, A., "Thermal Noise in Field-Effect Transistors," *Proc. IRE*, Vol. 50, Aug. 1962, pp. 1808–1812.
34. Vander Ziel, A., "Gate Noise in Field-Effect Transistors at Moderately High Frequencies," *Proc. IEEE*, Vol. 51, Mar. 1963, pp. 461–467.
35. Fukui, H., "Determination of the Basic Device Parameters of a GaAs MESFET," *Bell Sys. Tech. J.*, Vol. 58, Mar. 1979, pp. 771–797.
36. Goronkin, H., and V. Nair, "Comparison of GaAs MESFET Noise Figures," *IEEE Electron Dev. Lett.*, Vol. EDL-6, Jan. 1985, pp. 47–49.
37. Fukui, H., "Optimal Noise Figure of Microwave GaAs MESFETs," *IEEE Trans. Electron Dev.*, Vol. ED-26, July 1979, pp. 1032–1037.
38. Rothe, H., and W. Dahlke, "Theory of Noisy Fourpoles," *Proc. IRE*, Vol. 44, June 1956, pp. 811–818.
39. Fukui, H., "Design of Microwave GaAs MESFETs for Broad-Band Low-Noise Amplifiers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, July 1979, pp. 643–650.
40. Golio, J. M., and R. J. Trew, "Profile Studies of Ion-Implanted MESFETs," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-31, Dec. 1983, pp. 1066–1071.

41. Trew, R. J., M. A. Khatibzadeh, and N. A. Masnari, "Deep-Level and Profile Effects Upon Low-Noise Ion-Implanted GaAs MESFETs," *IEEE Trans. Electron Dev.*, Vol. ED-32, May 1985, pp. 877–882.
42. Furutsuka, T., T. Tsuji, and F. Hasegawa, "Improvement of the Drain Breakdown Voltage of GaAs Power MESFETs by a Simple Recess Structure," *IEEE Trans. Electron Dev.*, Vol. ED-25, June 1978, pp. 563–567.
43. Frensley, W. R., "Power Limiting Breakdown Effects in GaAs MESFETs," *IEEE Trans. Electron Dev.*, Vol. ED-28, Aug. 1981, pp. 1019–1024.
44. Winslow, T. A., and R. J. Trew, "Principles of Large-Signal MESFET Operation," *IEEE Trans. Microwave Theory Tech.*, Vol. 42, June 1994, pp. 935–942.
45. Macksey, H. M., "Optimization of the n^+ Ledge Channel Structure for GaAs Power FETs," *IEEE Trans. Electron Dev.*, Vol. ED-33, Nov. 1986, pp. 1818–1824.
46. Esaki, L., and R. Tsu, "Superlattice and Negative Conductivity in Semiconductors," IBM Research Report RC-2418, 1969.
47. Stormer, H. L., R. Dingle, A. C. Gossard, W. Wiegmann, and M. D. Sturge, "Two-Dimensional Electron Gas at Differentially Doped GaAs–AlGaAs Heterojunction Interface," *J. Vac. Sci. Tech.*, Vol. 16, 1979, pp. 1517–1519.
48. Mimura, T., S. Hiyamizu, K. T. Fujii, and K. Nanbu, "A New Field Effect Transistor with Selectively Doped GaAs/*n*-AlGaAs Heterostructures," *Jpn. J. Appl. Phys.*, Vol. 19, 1980, pp. L225–L227.
49. Shockley, W., "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," *Bell Syst. Tech. J.*, Vol. 28, 1949, p. 435.
50. Bardsen, J., and Brattain, W. H., "The Transistor: A Semiconductor Triode," *Phys. Rev.*, Vol. 74, 1948, p. 230.
51. Lilienfeld, J. E., "Method and Appartus for Controlling Electric Currents," U.S. Patent 1,745,175, 1930.
52. Lilienfeld, J. E., "Device for Controlling Electric Current," U.S. Patent 1,900,018, 1933.

PROBLEMS

7.1 The one-dimensional time-dependent continuity equation for free holes in the base of a uniformly doped *pnp* transistor is

$$D_{pB}\frac{\partial^2(p_B - p_{B0})}{\partial x^2} - \frac{p_B - p_{B0}}{\tau_{pB}} = \frac{\partial(p_B - p_{B0})}{\partial t}$$

Assume a sinusoidal time dependence so that

$$p_B(x,t) - p_{B0} = U(x)e^{j\omega t}$$

(a) Derive a differential equation for $U(x)$.

(b) Determine a solution for $U(x)$ and write the solution in the form of a traveling wave (i.e., exponential form with real and imaginary arguments).

(c) Use the $U(x)$ solution to derive an expression for a frequency-dependent diffusion length L'_B.

(d) Show that for high frequency (i.e., $\omega\tau_{pB} \gg 1$) $L'_B \simeq 2D_{pB}/\omega$ and for low frequency (i.e., $\omega\tau_{pB} \ll 1$) $L'_B = L_B$.

7.2 Show that

$$\coth x - \operatorname{csch} x = \tanh \frac{x}{2}$$

This relationship was used to arrive at the expression we derived for the base current of a bipolar transistor.

7.3 Discuss the unity gain cutoff frequency (f_T) and the maximum frequency of oscillation ($f_{\max}$) for Si bipolar transistors and GaAs MESFETs. Compare these figures of merit for the two devices explaining their physical significance. Estimate (using calculations if possible) their upper frequency limits.

7.4 For a *pnp* transistor the collector current can be written as

$$I_C \simeq \frac{qD_{pB}n_i^2 A \exp(qV_{EB}/kT)}{\int_0^{W_B} n(x)\,dx}$$

Show that the change in collector current as a function of collector–base voltage can be written as

$$\frac{\partial I_C}{\partial V_{CB}} = -\frac{I_C}{V_A}$$

where V_A is the Early voltage. Write an expression for the Early voltage in terms of the Gummel number Q_B.

7.5 When considering the RF operation of a bipolar transistor, we derived an expression for the common-base current gain that had the form

$$\alpha(\omega) \simeq \operatorname{sech} \frac{W_B}{L'_B}$$

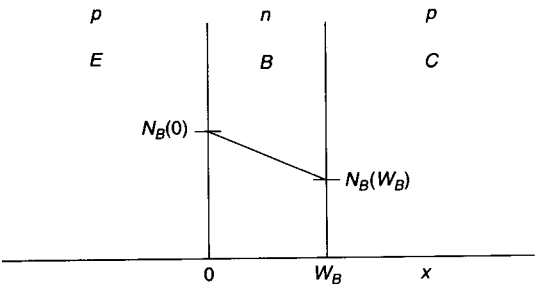
where W_B is the width of the undepleted base region and L'_B is the effective diffusion length for minority carriers in the base region. Show that this expression leads to a single-pole model for α and that α can be written as

$$\alpha(\omega) \simeq \frac{1}{1 + j(\omega/\omega_\alpha)}$$

where

$$\omega_\alpha \triangleq \frac{2D_{pB}}{W_B^2}$$

7.6 Consider a *pnp* bipolar transistor with exponential doping in the base region as shown in the sketch. Let $N_B(x) = N_B(0)\exp(-ax/W_B)$ where $a \triangleq \ln[N_B(0)/N_B(W_B)]$. Assume one dimension, no recombination in the base region, and steady state conditions.



(a) Derive an expression for the electric field that exists in the base region.

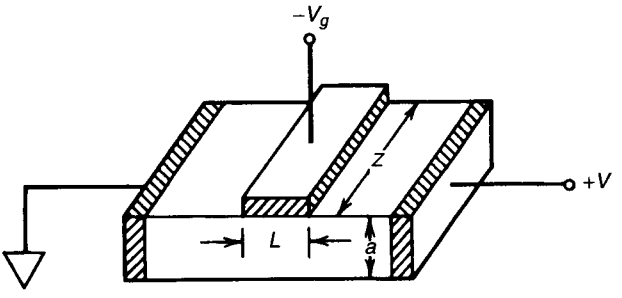
(b) Derive an expression for the minority-charge density in the base, $p_B(x)$. Use the boundary conditions that at $x = 0$, $p_B = p_B(0)$ and at $x = W_B$, $p_B \simeq 0$.

(c) Derive an expression for the base current, J_{pB}.

(d) Derive an expression for the minority-carrier transit time through the base region, τ_t.

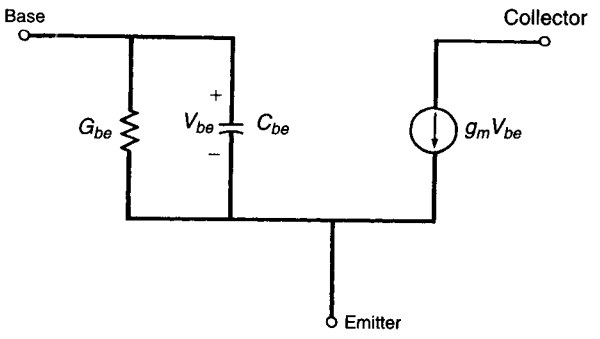
(e) Show that for uniformly doped devices $\tau_t = W_B^2/2D_{pB}$ and that $\tau_t \to 0$ as $a \to \infty$.

7.7 A semiconductor switch is fabricated by placing a Schottky contact on a bar of uniformly doped *n*-type semiconductor as shown in the sketch.



If $L \gg a$, derive an expression for the control electrode voltage V_g required to turn the switch off. State any assumptions or approximations you make.

7.8 A Si *npn* transistor is fabricated with uniform doping concentrations, and it is known that the doping densities in the emitter, base, and collector regions are $N_e = 10^{18}$ cm^{-3}, $N_b = 10^{16}$ cm^{-3}, and $N_c = 10^{14}$ cm^{-3}, respectively. The base–emitter and collector–base capacitances were measured at zero bias and were found to be $C_{be} = 5.16$ pF and $C_{cb} = 4.36$ pF. The transistor was then operated in the active mode with a base–emitter voltage of $V_{be} = -0.7$ V and a collector–base voltage of $V_{cb} = 6.0$ V. The collector current was $I_c = 50$ mA. Under these conditions the cutoff frequency for the transistor was determined to be $f_t = 2$ GHz.



(a) Determine the alpha cutoff frequency (f_α) for the transistor.

(b) A first-order, low-frequency equivalent circuit for the device is shown in the sketch. Determine values for the equivalent circuit elements. Use the following material parameters for Si:

$$T = 300 \text{ K}$$

$$\epsilon_r = 11.9$$

$$\epsilon_0 = 8.854(10)^{-14} \text{ F/cm}$$

$$n_i = 1.45(10)^{10} \text{ cm}^{-3}$$

$$\tau_n = \tau_p = 10^{-6} \text{ s (minority-carrier lifetime)}$$

Doping (cm^{-3})	μ_n (cm^2/V-s)	μ_p (cm^2/V-s)
10^{14}	1450	450
10^{16}	1200	410
10^{18}	270	160

8

PASSIVE DEVICES

Robert J. Trew

8.1 INTRODUCTION

The efficient control of electrical signals in RF circuits requires the use of nonlinear devices. These devices can be either active or passive. In hybrid circuit applications active devices (e.g., transistors or FETs) are generally used for signal generation and amplification, whereas passive devices (e.g., diodes) are often used for signal detection, frequency shifting (i.e., mixing), and control. Active devices are often used in place of passive devices in monolithic IC applications where, due to the fabrication process, it is not difficult to obtain large numbers of well-matched devices. In hybrid circuits it is generally preferrable to use passive devices where possible due to their low-loss and optimum performance characteristics.

The rectifying characteristics of certain semiconducting materials have been known for a long time. For example, in 1874 Braun* reported work on the asymmetric conduction properties of metal dots placed upon lead sulfide crystals. Point contact diodes were probably the first semiconductor devices to make use of this phenomenon and be extensively used in electronic equipment. These devices found widespread use as detectors in early radios and are still used for a variety of applications (e.g., millimeter-wave detectors).

Rectifying junctions are created by the establishment of an electrostatic barrier in the path of current flow. In order for the junction to conduct elec-

*Starting in 1874 Karl Ferdinand Braun (1850–1918) reported observations in which he noted the dependence of crystal electrical resistance on the polarity of the applied bias and on the detailed surface conditions. An excellent review of the historical record of rectifying contacts is presented in Henisch's book [1].

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

tricity, charge carriers must have enough energy to overcome the barrier. Although under certain conditions tunneling through the barrier is possible, the most common mechanism for charge conduction is thermionic emission over the barrier. By controlling the barrier with external bias, current control is achieved.

In this chapter the basic operating principles for various types of diode elements are discussed. The chapter begins with a description of the physics responsible for the operation of *pn* and Schottky barrier junctions. The remainder of the chapter is concerned with the utilization of the nonlinear characteristics of these rectifying junctions for producing circuit elements useful for controlling RF signals.

8.2 *pn* JUNCTIONS

In semiconductor crystals *pn* junctions are formed when the conductivity type changes from *n* type to *p* type. Free electrons and holes diffuse across the interface region where they recombine when they reach the opposite conductivity type material. The fixed donor and acceptor atoms that remain on each side of the junction create an electric dipole that results in an electrostatic potential barrier centered at the metallurgical junction between the *n*- and *p*-type material. The energy band diagrams for the formation of a *pn* junction are shown in Fig. 8.1.

The contact potential V_0 represents the barrier that must be overcome for charge to transfer from one side of the junction to the other. In terms of the semiconductor conduction band, the contact potential can be expressed as

$$qV_0 = \mathscr{E}_{cp} - \mathscr{E}_{cn} \tag{8.1}$$

where $\mathscr{E}_{cp}$ and $\mathscr{E}_{cn}$ are the *p*- and *n*-type conduction band energies and q is the electron charge. The contact potential depends upon the doping concentrations on each side of the junction according to the expression

$$V_0 = \frac{kT}{q} \ln \frac{p_{p0} n_{n0}}{n_i^2} \cong \frac{kT}{q} \ln \frac{N_a N_d}{n_i^2} \tag{8.2}$$

where p_{p0} and n_{n0} are the thermal equilibrium densities of majority carriers and N_a and N_d are the fixed acceptor and donor impurity densities. Note that the contact potential is also a function of the semiconductor material through the intrinsic density term n_i. Generally, wide-bandgap semiconductors have lower intrinsic concentrations and, consequently, higher contact potentials.

The contact potential expression given in (8.2) can be manipulated to obtain the minority-carrier densities on each side of the junction in terms of the majority (doping) concentrations on the other. For uniform doped junctions the majority-charge density on each side of the junction is essentially equal to

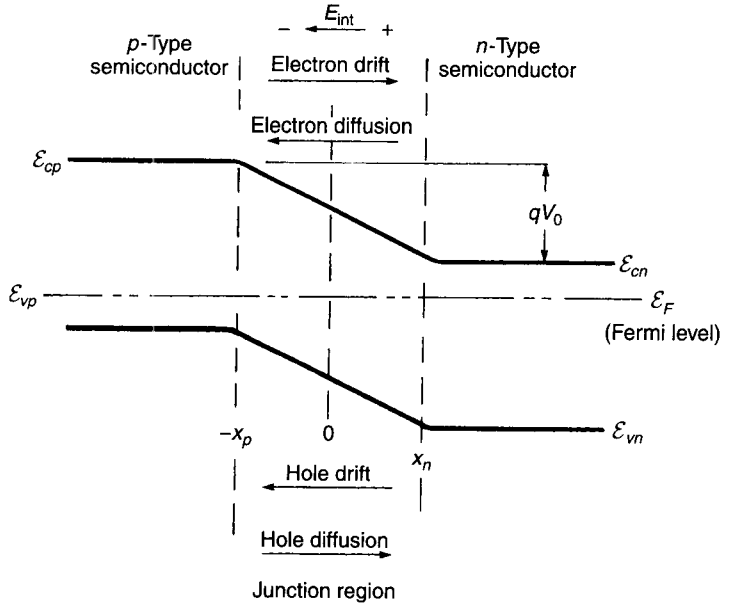


Figure 8.1 Energy band diagram for *pn*-junction diode.

that of the impurity doping. The resulting expression is sometimes called the "law of the junction" and is expressed as

$$\frac{p_p(-x_p)}{p_n(x_n)} = \frac{n_n(x_n)}{n_p(-x_p)} = e^{q(V_0 \pm V)/kT} \tag{8.3}$$

Equation (8.3) is written for bias situations and the contact potential V_0 has been replaced with the total junction voltage consisting of the contact potential and the applied bias ($V_0 \pm V$). The negative sign is used for forward bias and the positive sign is used for reverse bias.

In order to obtain expressions for the electric field and potential within the device, Poisson's equation can be solved. Poisson's equation is written as

$$\frac{dE}{dx} = -\frac{d^2\mathscr{V}}{dx^2} = \frac{q}{\epsilon}[p - n + N_d(x) - N_a(x)] \tag{8.4}$$

where p and n represent the free charge densities. The terms N_d and N_a are the fixed (doping) charge densities and ϵ is the permittivity of the material. To solve this equation, it is generally assumed that the semiconductor can be divided into quasi-neutral and space-charge regions. Applying the equation to the one-dimensional junction charge distribution shown in Fig. 8.2 yields for the region $-x_p \leq x \leq 0$ a potential distribution of the form

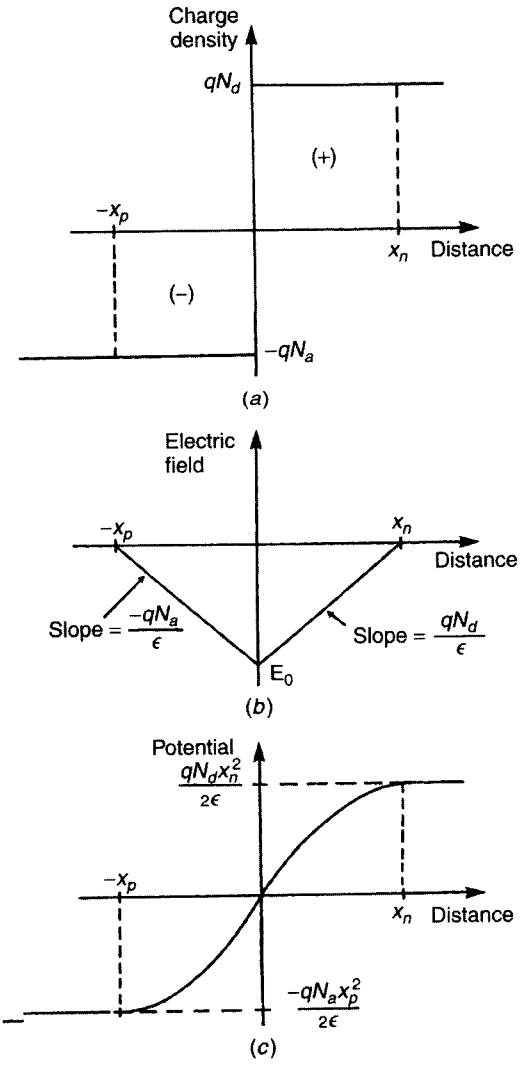


Figure 8.2 (*a*) Charge density, (*b*) electric field, and (*c*) potential distribution for uniformly doped *pn* junction.

$$\mathscr{V} = \frac{qN_a}{2\epsilon}[(x + x_p)^2 - x_p^2] \tag{8.5}$$

For the region $0 \leq x \leq x_n$, the potential is

$$\mathscr{V} = -\frac{qN_d}{2\epsilon}[(x - x_n)^2 - x_n^2] \tag{8.6}$$

Neglecting the finite resistivity of the neutral semiconductor regions, the potential in these regions will be constant with magnitudes: for $x \geq x_n$,

$$\mathscr{V} = \mathscr{V}(x_n) = \frac{qN_d x_n^2}{2\epsilon} \tag{8.7}$$

and for $x \leq -x_p$,

$$\mathscr{V} = \mathscr{V}(-x_p) = -\frac{qN_a x_p^2}{2\epsilon} \tag{8.8}$$

The electric field and potential throughout the junction area are shown in Figs. 8.2*b* and *c*.

For single-crystal semiconductors with no trapping states at the interface between the *p* and *n* regions, the electric field must be continuous. The electric field at the interface can be written as

$$E(0) = E_0 = -\frac{d\mathscr{V}(0)}{dx} = -\frac{qN_d x_n}{\epsilon} = -\frac{qN_a x_p}{\epsilon} \tag{8.9}$$

This equation yields the charge balance condition, which is expressed as

$$N_d x_n = N_a x_p \tag{8.10}$$

At thermal equilibrium with no applied bias voltage the total potential drop across the semiconductor is the contact potential and can be obtained from the potential expressions presented in (8.5) and (8.6). Therefore,

$$V_0 = \mathscr{V}(x_n) - \mathscr{V}(-x_p) = \frac{q}{2\epsilon}(N_d x_n^2 + N_a x_p^2) \tag{8.11}$$

This expression can be combined with the charge balance condition to obtain the dimensions of the depletion region. The depletion depths on the *n* and *p* side are

$$x_n = \left[\frac{2\epsilon V_0 N_a}{qN_d}\left(\frac{1}{N_a + N_d}\right)\right]^{1/2} \tag{8.12a}$$

$$x_p = \left[\frac{2\epsilon V_0 N_d}{qN_a}\left(\frac{1}{N_a + N_d}\right)\right]^{1/2} \tag{8.12b}$$

The total depletion region width is

$$W = x_n + x_p = \left[\frac{2\epsilon V_0}{q}\left(\frac{N_a + N_d}{N_a N_d}\right)\right]^{1/2} \tag{8.13}$$

Note that the depletion region penetrates most deeply into the less heavily doped side of the junction.

The maximum value for the electric field occurs at the position $x = 0$ and can be expressed as

$$E_{\max} = E_0 = -\left[\frac{2V_0 q N_a N_d}{\epsilon(N_a + N_d)}\right]^{1/2} = -\frac{q}{\epsilon}\frac{N_a N_d}{(N_a + N_d)}W \tag{8.14}$$

As the doping on both sides of the junction is increased, the maximum electric field increases. For very high doping densities it is possible for the electric field to reach and exceed the critical field for tunnel emission ($E_c \sim 10^6$ V/cm for Si and GaAs) and tunnel breakdown will occur. For lower doping levels the junction will break down by avalanche ionization, and this occurs for electric fields on the order of 500 kV/cm.

8.2.1 Ideal Diode Equation

When a potential is applied across the *pn* junction, a current will flow. The applied potential modifies the junction potential barrier, making it either easier or more difficult for mobile charge carriers to travel across the junction. The energy band diagrams for forward and reverse bias are shown in Fig. 8.3.

The voltage–current characteristic for the junction can be derived from the continuity equation, which can be written for the region $x \leq -x_p$ in the simplified one-dimensional form

$$\frac{d^2(n_p - n_{p0})}{dx^2} - \frac{(n_p - n_{p0})}{L_n^2} = 0 \tag{8.15}$$

where $L_n = \sqrt{D_p \tau_p}$ is the electron diffusion length in the *p*-type material and D_p and τ_p are the hole diffusion coefficient and lifetime, respectively. In writing (8.15) electric field effects have been neglected. Boundary conditions are supplied by (8.3) and the minority-carrier density is

$$n_p(-x_p) = n_{p0}e^{qV/kT} \tag{8.16}$$

A similar formulation applies to the *n*-type side of the junction in the region $x \geq x_n$. The continuity equation yields expressions for the minority-charge densities in these regions on each side of the junction. The solutions are

$$n_p - n_{p0} = n_{p0}(e^{qV/kT} - 1)e^{(x+x_p)/L_n} \tag{8.17}$$

$$p_n - p_{n0} = p_{n0}(e^{qV/kT} - 1)e^{-(x-x_n)/L_p} \tag{8.18}$$

Considering only minority-carrier diffusion in the bulk regions, the minority-carrier currents are

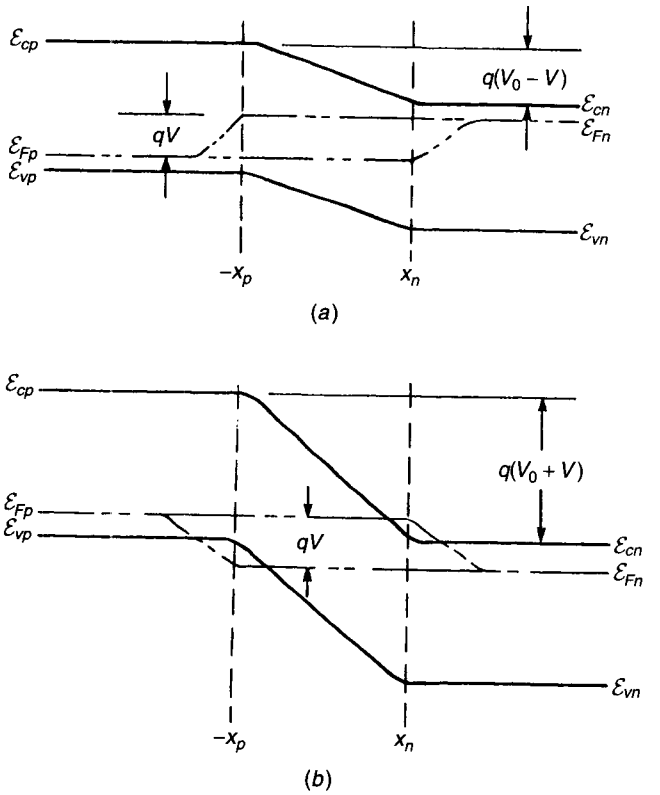


Figure 8.3 The *pn*-junction energy bands for (*a*) forward and (*b*) reverse bias.

$$J_n = qD_n \frac{dn_p}{dx}\bigg|_{-x_p} = \frac{qn_{p0}D_n}{L_n}(e^{qV/kT} - 1) \tag{8.19}$$

$$J_p = -qD_p \frac{dp_n}{dx}\bigg|_{x_n} = \frac{qp_{n0}D_p}{L_p}(e^{qV/kT} - 1) \tag{8.20}$$

The total current density is the sum of (8.19) and (8.20) and is written as

$$J = J_n|_{-x_p} + J_p|_{x_n} = \left(\frac{qp_{n0}D_p}{L_p} + \frac{qn_{p0}D_n}{L_n}\right)(e^{qV/kT} - 1) \tag{8.21}$$

If the diode cross-sectional area is A, (8.21) is written as the ideal diode equation

$$I = I_0(e^{qV/kT} - 1) \tag{8.22}$$

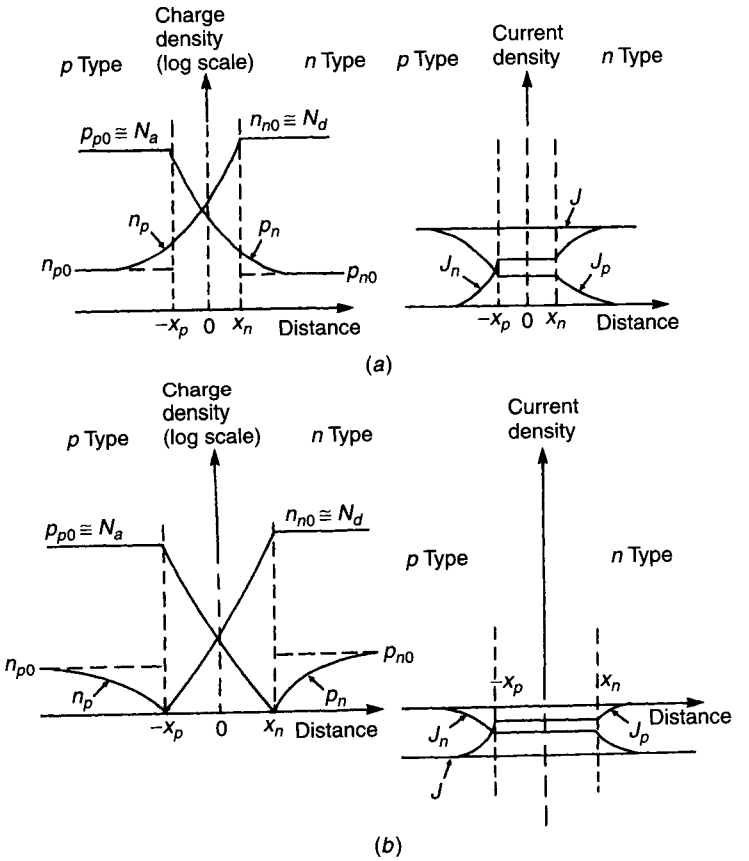


Figure 8.4 Charge density and current distribution throughout *pn* junction under (*a*) forward and (*b*) reverse bias.

where

$$I_0 \cong qn_i^2 A\left(\frac{D_p}{N_d L_p} + \frac{D_n}{N_a L_n}\right)$$

is the saturation current.

The charge and current densities in the diode under both forward and reverse bias are shown in Fig. 8.4.

8.2.2 Deviations from Ideal Diode Equation

The derivation of the ideal diode equation considers only low injection conditions and ignores effects that, in practice, result in deviations from the I–V characteristics predicted by the equation. The most significant deviations are

due to

1. high-level injection conditions,
2. recombination and generation in the depletion region, and
3. ohmic effects.

These effects are discussed in this section.

High-Level Injection Conditions. The largest forward bias that can be applied to a *pn* junction is the contact potential. Application of this voltage to the junction would result in a flat-band condition (i.e., no potential barrier). The ideal diode equation, however, indicates no limit to the allowed forward bias. The problem results from ignoring the majority-carrier density changes at the junction. Under large forward-bias voltages the magnitude of the injected minority density can become significant compared to the background impurity density. Since space-charge neutrality conditions are maintained, the majority free charge density will redistribute and take on a profile essentially equal to that of the injected minority charge. When this effect is included in the derivation of the diode equation, a modified equation is obtained that has the form

$$I = qA\left[\frac{e^{qV/kT}-1}{1-e^{-2q(V_0-V)/kT}}\right]\left[\frac{D_p p_{n0}}{L_p}\left(1+\frac{n_i^2}{p_{p0}^2}e^{qV/kT}\right)+\frac{D_n n_{p0}}{L_n}\left(1+\frac{n_i^2}{n_{n0}^2}e^{qV/kT}\right)\right] \tag{8.23}$$

Note that in this equation the contact potential V_0 is the limiting forward voltage for the junction.

Depletion Region Generation and Recombination. In the derivation of the ideal diode equation the generation and recombination of free charge carriers in the depletion region is ignored. Actually, a finite density of recombination centers will exist in the depletion region and there will be a net loss of charge carriers. This loss can be interpreted as a "recombination" current and can be expressed as

$$J_r \cong \frac{qWn_i}{2\tau}e^{qV/2kT} \tag{8.24}$$

where τ is the minority-carrier lifetime in the depletion region. Note the factor of 2 that appears in the denominator of the exponential.

As indicated by (8.24), recombination current is most significant for

1. low bias voltages,
2. large-bandgap semiconductors,
3. short-lifetime semiconductors,

4. high-impurity doping, and
5. high-temperature operation.

For practical diodes, the diode equation is generally written in the form

$$J = J_0(e^{qV/nkT} - 1) \tag{8.25}$$

where n is the ideality factor. An ideal diode will have $n = 1$, and diodes that are dominated by recombination currents will have $n = 2$. Generally, a good diode will have an ideality factor that approaches unity.

Under reverse bias the width of the depletion region increases and the charge generated will be swept from the depletion region and contribute to the reverse saturation current. Since the depletion region increases with reverse bias, the reverse saturation current will not saturate but will show some sort of bias dependence. For example, in an abrupt junction diode the depletion region width is proportional to the square root of the applied voltage, and it follows that the generation current is also proportional to the square root of the voltage. That is,

$$W \propto V^{1/2}$$

and, therefore,

$$J_r \propto V^{1/2}$$

Ohmic Effects. In the derivation of the ideal diode equation, it is assumed that all the applied bias voltage appears across the junction. The finite resistivity of the bulk semiconductor regions was ignored. Actually, the finite resistivity of these regions will produce a voltage drop that will reduce the fraction of the applied voltage that actually appears across the junction. The voltage that appears across the junction is

$$V = V_{\text{appl}} - I[R_p(I) - R_n(I)] \tag{8.26}$$

where R_p and R_n are the resistances of the bulk p- and n-type regions and are functions of current, since they are dependent upon the conductivity of the material. When the ohmic effects become significant, the current shows a saturation tendency.

The I–V characteristics for a pn junction are sketched in Fig. 8.5 where the various regions of deviation from the ideal diode equation are indicated.

8.2.3 Junction Capacitance

Capacitance is a measure of the energy storage in an electric field. It can be defined from

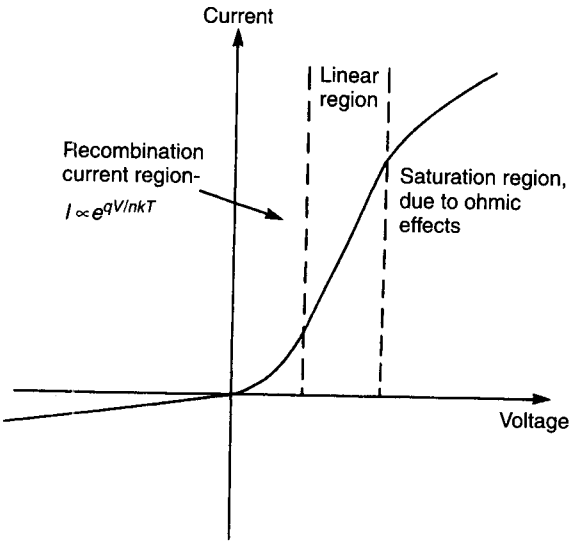


Figure 8.5 The *I–V* characteristic for *pn* junction showing regions of deviation from the ideal.

$$C \triangleq \frac{dq/dt}{dv/dt} = \frac{dq}{dv} \tag{8.27}$$

There are two types of capacitance associated with *pn* junctions. One is associated with the fixed space charge of the impurity atoms. This capacitance is called depletion layer capacitance and it dominates under reverse bias. The other capacitance is associated with the charge storage of free charge carriers. This is generally termed diffusion capacitance and dominates under forward bias.

Depletion Capacitance. The depletion layer capacitance results from the electric dipole that is created by the fixed donor and acceptor atoms of the *pn* junction. The capacitance can be calculated from the expression.

$$C_j \triangleq \left|\frac{dQ}{dV}\right| \tag{8.28}$$

where Q is the stored charge on one side of the junction. For a junction of area A, $Q = AE_{\max}$. For a one-sided p^+n diode with donor density given by the expression

$$N_d(x) = ax^b \tag{8.29}$$

where a and b are constants, the depletion region width is

$$W = \left[\frac{\epsilon(b+2)(V_0 - V)}{qa}\right]^{1/(b+2)} \tag{8.30}$$

The diode capacitance is calculated to be

$$C_j = A\left(\frac{qa\epsilon^{b+1}}{b+2}\right)^{1/(b+2)} (V_0 - V)^{-1/(b+2)} \tag{8.31}$$

This is more commonly written as

$$C_j = \frac{C_0}{(1 - V/V_0)^{\gamma}} \tag{8.32}$$

where $\gamma = 1/(b+2)$ and C_0 is the zero-bias capacitance. For an abrupt junction $\gamma = \frac{1}{2}$, for a graded junction $\gamma = \frac{1}{3}$, and for hyperabrupt junctions $\gamma \to 1$–2. Hyperabrupt junctions are very important in linear tuning varactor applications.

Diffusion Capacitance. The forward-biased diffusion capacitance is due to injected minority charge. It is important for time-varying situations (i.e., transient or RF operation). By applying the continuity equation to a forward-biased junction, it can be shown that the minority-charge storage has two components, one due to a recombination term in which the excess charge distribution is replaced on an average of every minority-carrier lifetime and a charge buildup term where this term can indicate either increasing or decreasing charge storage. For steady state the second term is zero. An analysis (e.g., [2]) of the forward-bias condition shows that the junction can be described by the parallel combination of a conductance with magnitude

$$G_d = \frac{q}{kT} I_{\text{dc}} \tag{8.33}$$

and a capacitance with magnitude

$$C_d = \frac{q\tau}{kT} I_{\text{dc}} \tag{8.34}$$

where I_{dc} is the diode current.

The depletion layer and diffusion capacitances are sketched in Fig. 8.6. The diffusion capacitance can severely restrict the high-frequency and switching performance of *pn* junctions. Note that the diffusion capacitance is directly proportional to the minority-carrier lifetime. Short lifetimes are desirable for minimizing charge storage effects. It will be shown that Schottky barrier diodes are inherently faster than *pn* junctions because they do not have any minority-carrier storage.

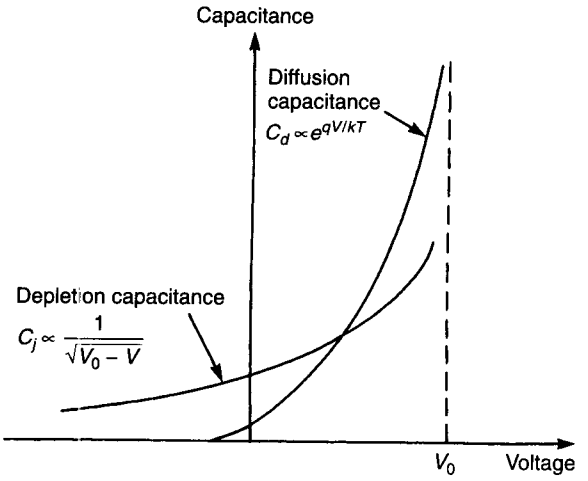


Figure 8.6 Capacitance–voltage characteristic for *pn* junction showing comparison between depletion and diffusion capacitance.

8.3 SCHOTTKY BARRIER JUNCTIONS

Schottky barrier junctions are formed when a metal is placed upon a semiconductor. The rectifying properties of a metal–semiconductor contact derive from the presence of an electrostatic barrier between the metal and the semiconductor. An argument for the existence of this barrier can be formulated in the following manner. Consider the energy band diagrams for a metal and an *n*-type semiconductor as shown in Fig. 8.7. The metal has all of its allowed energy states occupied up to the Fermi energy, $\mathscr{E}_F$. The metal work function is shown in Fig. 8.7 as $q\phi_m$ and is defined as the energy required to remove an electron at the Fermi level to a reference energy level (assumed to be a vacuum). That is, the work function is the energy required to free an electron from the surface of the metal into a vacuum just outside the metal. Likewise, a work function can be defined for the semiconductor. In general, electrons will not exist at the Fermi energy level in the semiconductor, since it lies within the energy gap. An electron affinity (given as the symbol $q\chi_s$) is defined in a similar manner as the energy required to remove an electron at the conduction band edge to the reference level.

A rectifying barrier can be formed if the work function for the metal is greater than that for the semiconductor, that is, if

$$q\phi_m > q\phi_s \tag{8.35}$$

As the metal and semiconductor are placed in physical contact, the energy requirements dictate that at equilibrium the Fermi levels must align. This

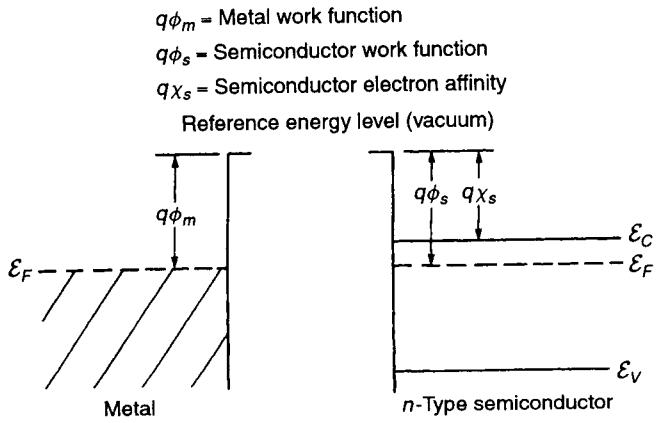


Figure 8.7 Energy band diagrams for metal and semiconductor.

argument assumes that surface states do not exist. Electrons in the semiconductor are more energetic than those in the metal, and there is a net flow of electrons into the metal. A net positive space charge is left in the semiconductor in the region adjacent to the metal, as illustrated in Fig. 8.8. The space charge produces a distortion (upward bending) of the energy band edges and an electrostatic barrier is created in the semiconductor. The presence of this electrostatic barrier impedes the movement of electrons from the semiconductor into the metal. Any electron attempting to move into the metal must have energy greater than qV_0, the built-in or contact potential. For movement from the metal into the semiconductor an electron must have energy greater than the barrier height, defined as

$$q\phi_b \triangleq q(\phi_m - \chi_s) \tag{8.36}$$

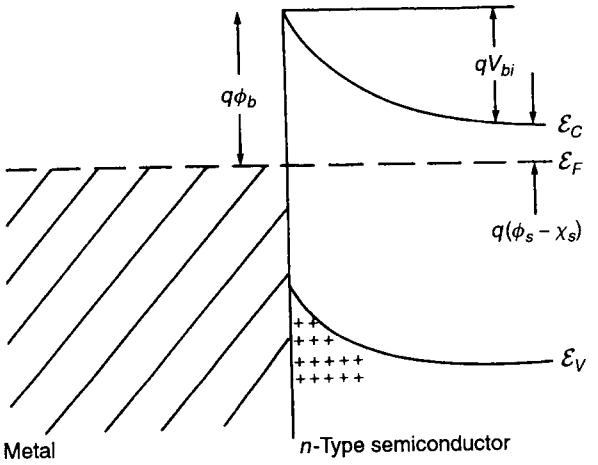


Figure 8.8 Energy band diagram for ideal Schottky barrier junction.

This expression is called the Mott limit [3] and the barrier it defines is dependent upon both the metal and the semiconductor.

8.3.1 Surface Effects

In experiments involving the fabrication of rectifying metal–semiconductor contacts, it is often observed that when various metals are deposited upon certain semiconductors, the junctions that result do not obey the Mott formulation. The barrier height is often independent of the metal.

Bardeen [4] was the first to recognize the importance of surface states in determining the barrier height. Surface states have their origin in the discontinuous crystal structure at the crystal surface and result from unsatisfied covalent bonds. The surface atoms do not have the required four neighboring atoms to form the covalent structure, and a number of electrons are loosely bound to their atoms. The atoms with the uncompleted covalent bonds can (a) give up an electron, thereby becoming a donor, or (b) accept an electron, thereby becoming an acceptor. A distortion in the lattice structure is created that produces additional allowed energy levels within the forbidden region at the semiconductor surface.

The surface states are usually continuously distributed in energy and are characterized with a "neutral level" ϕ_0 such that the states are occupied up to ϕ_0 and empty above ϕ_0. If the Fermi level lies above ϕ_0, the surface states possess a net negative charge (i.e., they are acceptorlike), and if the Fermi level lies below ϕ_0, the states possess a net positive charge (i.e., they are donorlike). This latter situation is shown in Fig. 8.9, where it is noted that the net positive charge causes an upward bending of the semiconductor energy bands.

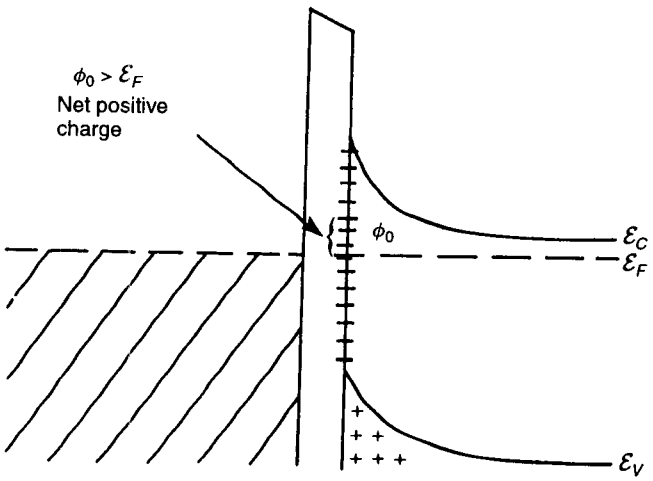


Figure 8.9 Energy band diagram for metal–insulator–semiconductor junction with surface states.

Some of the electric field lines terminating on the metal are due to the surface states rather than the ionized donors in the crystal. The depletion region and the barrier height are reduced. The surface states provide a negative feedback that acts to "push" ϕ_0 down toward the Fermi level. This feedback always tends to reduce the deviation of ϕ_0 from the Fermi level.

If the number of surface states is high, the Fermi level becomes "pinned" to ϕ_0 and the barrier height becomes essentially independent of the metal. The barrier height that results can be defined if ϕ_0 is measured from the top of the valence band. The barrier height is called the Bardeen limit and is defined as

$$\phi_b \triangleq E_g - \phi_0 \tag{8.37}$$

where E_g is the semiconductor energy gap. Note that the barrier height, as defined by the Bardeen limit, is not a function of the metal. In fact, many metal–semiconductor contacts obey the Bardeen limit. The properties of many such junctions are listed in Table 8.1.

8.3.2 Image Force Lowering

Any charged body close to a conducting plane produces an electrostatic field as if there were an equal and oppositely charged particle located at the mirror image. Since this image has the opposite charge, the total potential energy of the system is reduced. Consider the situation sketched in Fig. 8.10. An electron located at distance x in a vacuum from the metal will experience an attractive force between it and its image located at a location $-x$. The attractive force is given by the expression

$$F = \frac{-q^2}{4\pi(2x)^2\epsilon_0} = \frac{-q^2}{16\pi\epsilon_0 x^2} \tag{8.38}$$

The potential energy of the electron at distance x is calculated from the work done by an electron in moving from infinity to point x and is

$$\mathrm{PE}'(x) = \int_{\infty}^{x} F\,dx = \frac{q^2}{16\pi\epsilon_0 x} \tag{8.39}$$

When an external electric field is also applied, the total potential energy of the system is given as

$$\mathrm{PE}(x) = \frac{q^2}{16\pi\epsilon_0 x} + qEx \tag{8.40}$$

The amount of barrier lowering due to the image force can be calculated from

$$\frac{d[\mathrm{PE}(x)]}{dx} = 0 \tag{8.41}$$

Table 8.1 Measured Schottky Barrier Heights (Volts at 300 K)

Semiconductor	Type	E_g (eV)	Ag	Al	Au	Cr	Cu	Hf	In	Mg	Mo	Ni	Pb	Pd	Pt	Ta	Ti	W
Diamond	*p*	5.47			1.71													
Ge	*n*	0.66	0.54	0.48	0.59		0.52		0.64			0.49	0.38					0.48
	p		0.50		0.30				0.55									
Si	*n*	1.12	0.78	0.72	0.80	0.61	0.58	0.58		0.40	0.68	0.61		0.81	0.90		0.50	0.67
	p		0.54	0.58	0.34	0.50	0.46				0.42	0.51	0.55				0.61	0.45
SiC	*n*	3.00		2.00	1.95													
AlAs	*n*	2.16			1.20										1.00			
AlSb	*p*	1.63			0.55													
BN	*p*	7.50			3.10													
BP	*p*	6.00			0.87													
GaSb	*n*	0.67			0.60													
GaAs	*n*	1.42	0.88	0.80	0.90		0.82	0.72							0.84	0.85		0.80
	p		0.63		0.42			0.68										
GaP	*n*	2.24	1.20	1.07	1.30	1.06	1.20	1.84		1.04	1.13	1.27			1.45		1.12	
	p				0.72													
InSb	*n*	0.16	0.18[a]		0.17[a]													
InAs	*p*	0.33			0.47[a]													
InP	*n*	1.29	0.54		0.52													
	p				0.76													
CdS	*n*	2.43	0.56	Ohmic	0.78		0.50					0.45	0.59	0.62	1.10		0.84	
CdSe	*n*	1.70	0.43		0.49		0.33								0.37			
CdTe	*n*		0.81	0.76	0.71										0.76			
ZnO	*n*	3.20		0.68	0.65		0.45		0.30					0.68	0.75	0.30		
ZnS	*n*	3.60	1.65	0.80	2.00		1.75		1.50	0.82				1.87	1.84	1.10		
ZnSc	*n*		1.21	0.76	1.36		1.10		0.91				1.16		1.40			
PbO	*n*		0.95						0.93			0.96	0.95					

[a] 77 K.

Source: Sze [2].

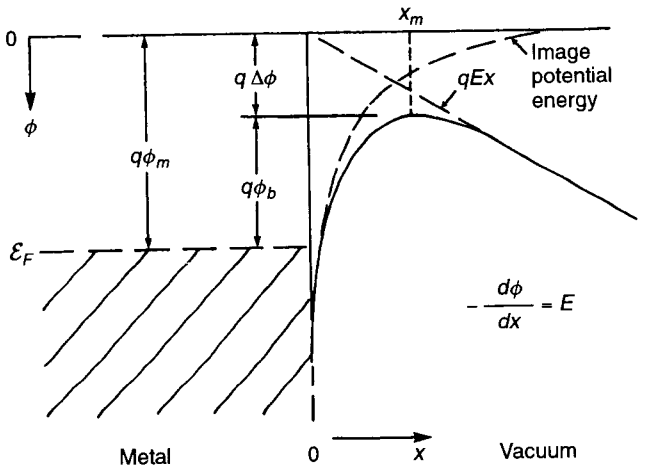


Figure 8.10 Energy band diagram for Schottky barrier junction showing effects of image force lowering.

The calculation indicates that the barrier is reduced by an amount

$$\Delta\phi = \sqrt{\frac{qE}{4\pi\epsilon_0}} = 2Ex_m \tag{8.42}$$

where the distance

$$x_m = \left(\frac{q}{16\pi\epsilon_0 E}\right)^{1/2} \tag{8.43}$$

indicates the location of the potential maximum from the metal–semiconductor interface.

This result can be easily applied to situations where the electron is in a semiconductor by simply replacing the free-space permittivity with that for the semiconductor. Note that for high-magnitude electric fields the barrier height, $q\phi_b$, is significantly reduced and the effective metal work function for thermionic emission is reduced.

8.3.3 Schottky Model

Although the rectifying properties of metal–semiconductor contacts were known and studied as far back as 1874 with the work of Braun, most of the modern work on these contacts is based, at least in part, upon the models of Schottky [5] and Mott [3], which were presented in 1938 and 1939, respectively. The theoretical background presented in this early work clearly established the role of space charge as the basis for the built-in potential barrier. The Schottky

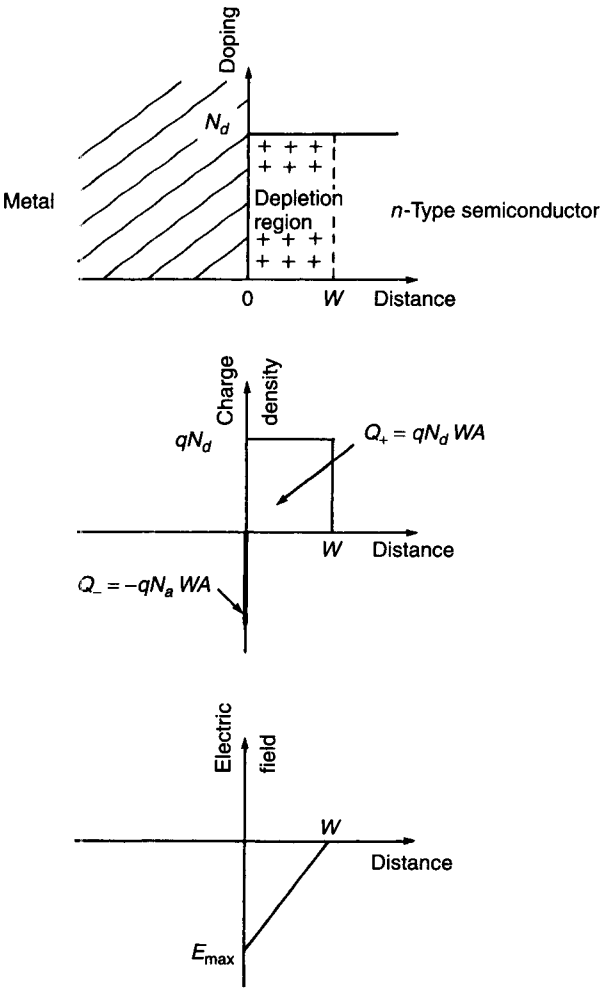


Figure 8.11 Schottky model formulation for metal–semiconductor junction.

model, in particular, describes how an electrostatic barrier is created when a metal is placed in contact with a semiconductor.

The model predicts the creation of a potential barrier at the metal–semiconductor interface, as shown in Fig. 8.11. If the depletion approximation is employed, the region of the semiconductor directly adjacent to the metal will be depleted of charge. This produces a charge distribution and an electric field in the interface region. By applying Poisson's equation to the junction, the maximum electric field will occur at the interface and is given by

$$E_{max} = -\frac{qN_d W}{\epsilon} \tag{8.44}$$

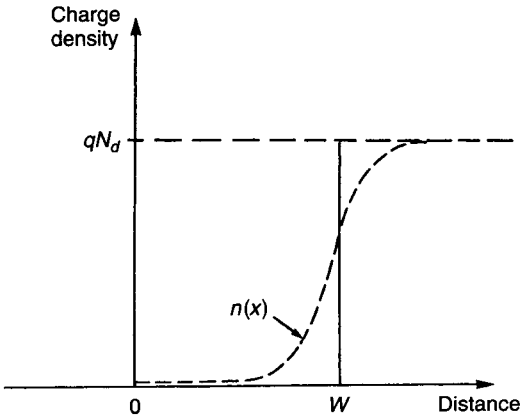


Figure 8.12 Charge distribution in metal–semiconductor junction (dashed line indicates free charge).

The voltage across the depletion region is the built-in or contact potential and is

$$V_0 = -\frac{1}{2}E_{\max}W = \frac{1}{2}\frac{qN_dW^2}{\epsilon} \tag{8.45}$$

From these expressions the space charge in the semiconductor depletion region of junction area A can be calculated and is

$$Q = qAN_dW = A(2q\epsilon N_dV_0)^{1/2} \tag{8.46}$$

The depletion approximation is actually not valid for many applications and leads to errors. The free electron concentration will decrease from a magnitude equal to the donor density to a low value at the interface, as shown by the dashed line in Fig. 8.12. The free electron density will modify (reduce) the background space charge. The modification required in the theory to account for the nonzero electron density results in an additional term of kT/q in (8.46), so that it is written as

$$Q = A\left[2q\epsilon N_d\left(V_0 \pm V - \frac{kT}{q}\right)\right]^{1/2} \tag{8.47}$$

where a biasing voltage V is introduced with the minus sign indicating forward bias and the plus sign indicating reverse bias. The bias situations are shown in Fig. 8.13.

For semiconductors with relatively high mobility (most common semiconductors) and moderate doping, the current across the junction will be

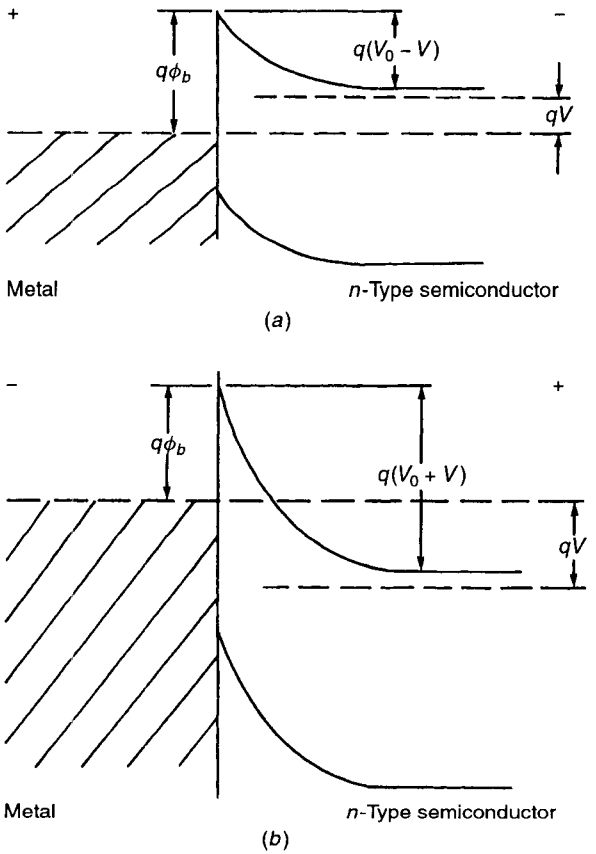


Figure 8.13 Energy band diagrams for Schottky barrier junction under (*a*) forward and (*b*) reverse bias.

dominated by thermionic emission [6]. Under these conditions the junction current is written as a function of voltage in the form

$$J = J_0(e^{qV/kT} - 1) \tag{8.48}$$

where

$$J_0 = A^* T^2 e^{-q\phi_b/kT} \tag{8.49}$$

and

$$A^* = \frac{4\pi m^* q k^2}{h^3}$$

is the effective Richardson constant. The term m^* is the electron effective mass, h is Planck's constant, and k is Boltzmann's constant.

The barrier height ϕ_b is a function of applied bias. The bias dependence of the barrier height can be modeled in a convenient manner by assuming a linear relationship,

$$\phi_b = \phi_{b0} + \beta V \tag{8.50}$$

where the coefficient β will always be positive because the barrier height always increases with forward voltage.

The current equation can be rewritten in the form

$$J = J_0 e^{qV/nkT}(1 - e^{-qV/kT}) \tag{8.51}$$

where

$$\frac{1}{n} = 1 - \beta = 1 - \frac{\partial \phi_b}{\partial V} \tag{8.52}$$

If $\partial \phi_b / \partial V$ is constant, then n is a constant. For $V > 3kT/q$ the current equation can be approximated by the commonly used form

$$J \cong J_0 e^{qV/nkT} \tag{8.53}$$

The n term is the ideality factor and its magnitude depends upon semiconductor doping and the current transport mechanism. For barrier current consisting of pure thermionic emission $n = 1$, and for current transport dominated by generation recombination $n = 2$. Generally, thermionic emission will dominate the current flow in good-quality diodes and a low magnitude for n is desired. Good diodes normally have n values ranging from about 1.02 to 1.2.

8.3.4 Junction Capacitance

Since Schottky barrier junctions are majority-carrier devices, minority-carrier storage is not a factor in their operation. Therefore, the diffusion capacitance that dominates under forward bias in *pn* junctions is, for most RF applications of Schottky diodes, not important. When minority-charge storage must be considered, as in switching applications, the minority-charge storage time can be calculated from the expression

$$\tau_s = \frac{q n_i^2 \sqrt{D_p \tau_p}}{N_d J_0} \tag{8.54}$$

The elimination of the diffusion capacitance provides Schottky barrier devices with an inherent speed advantage compared to *pn*-junction diodes.

The depletion layer capacitance can be calculated in a manner analogous to that used for *pn*-junction diodes. The device can be considered to be a parallel-plate capacitor with a voltage-dependent space-charge width given by the expression

$$W = \left[\frac{2\epsilon}{qN_d}\left(V_0 - V - \frac{kT}{q}\right)\right]^{1/2} \tag{8.55}$$

The capacitance per unit surface area is determined to be

$$C_j = \frac{\epsilon}{W} = \left[\frac{q\epsilon N_d}{2[V_0 - V - (kT/q)]}\right]^{1/2} \tag{8.56}$$

The application of (8.56) to a typical microwave diode with $N_d = 2 \times 10^{17}$ cm^{-3}, $V_0 = 0.8$ V, $\epsilon = \epsilon_r\epsilon_0 = (12.9)8.854 \times 10^{-14}$ F/cm, and $q = 1.6 \times 10^{-19}$ C gives a junction capacitance of $C_j = 1.51A$ fF, where Schottky area A is in square micrometers.

8.3.5 Rectifying Contact Materials

When selecting materials for rectifying contacts, it is necessary to consider a wide range of chemical, metallurgical, and physical as well as electrical characteristics. For example, the metal films suitable for use as contacts should have the following characteristics:

1. correct barrier height,
2. good adherence to the semiconductor,
3. different etch rate from the semiconductor,
4. easy bondability,
5. resistance to metallurgical reactions,
6. resistance to diffusion,
7. resistance to oxidation and corrosion, and
8. resistance to electromigration.

A variety of metals have been successfully used for Schottky junctions. Typical metals include aluminum (Al), gold (Au), tungsten (W), and titanium/tungsten (TiW). Tungsten and titanium/tungsten are refractory metals and can be used at high temperatures. For this reason they are sometimes used for devices designed to operate at high power levels where elevated temperatures are likely to be encountered.

In order to achieve contacts with the desired characteristics, it is often necessary to use combinations of metals. For example, titanium/tung sten/gold can be used to achieve a satisfactory rectifying contact. The TiW provides for a suitable barrier height and good physical characteristics, since on Si and GaAs

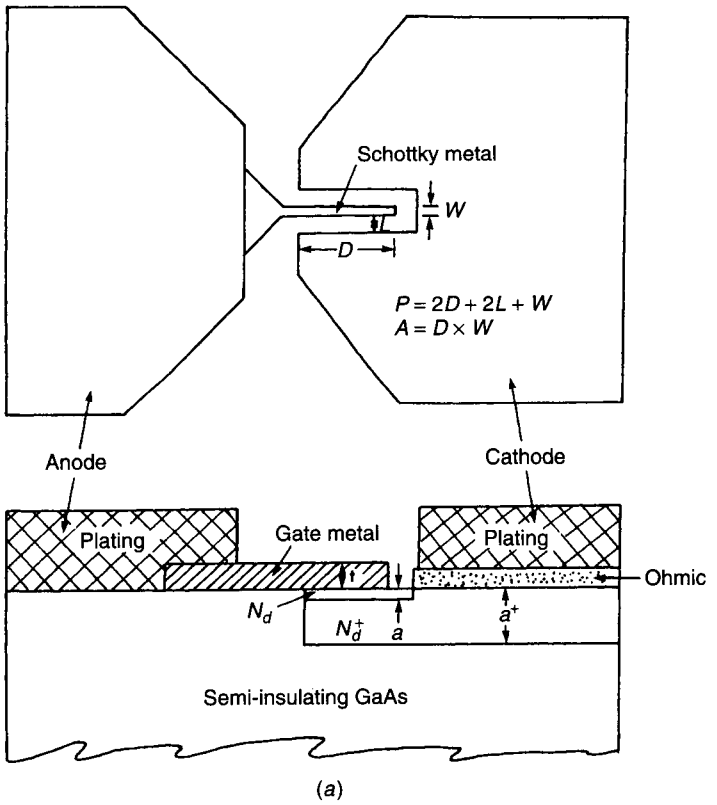


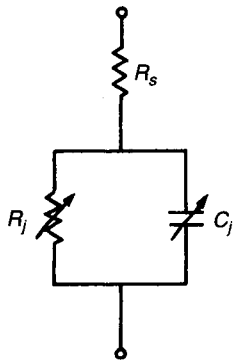
Figure 8.14 (*a*) Top side view of Schottky diode. (*b*) Equivalent circuit for Schottky diode. (*c*) Schottky diode characteristics as function of active layer doping.

it is nonreactive, adheres well, and resists electromigration. The gold provides for good bonding capability and low electrical resistance. The TiW layers are very thin (in the range of hundreds of angstroms), whereas the gold layer is relatively thick (on the order of micrometers). The TiW layers would typically be placed upon a cleaned semiconductor surface by sputtering or evaporation and the gold layer would be electroplated to the desired thickness.

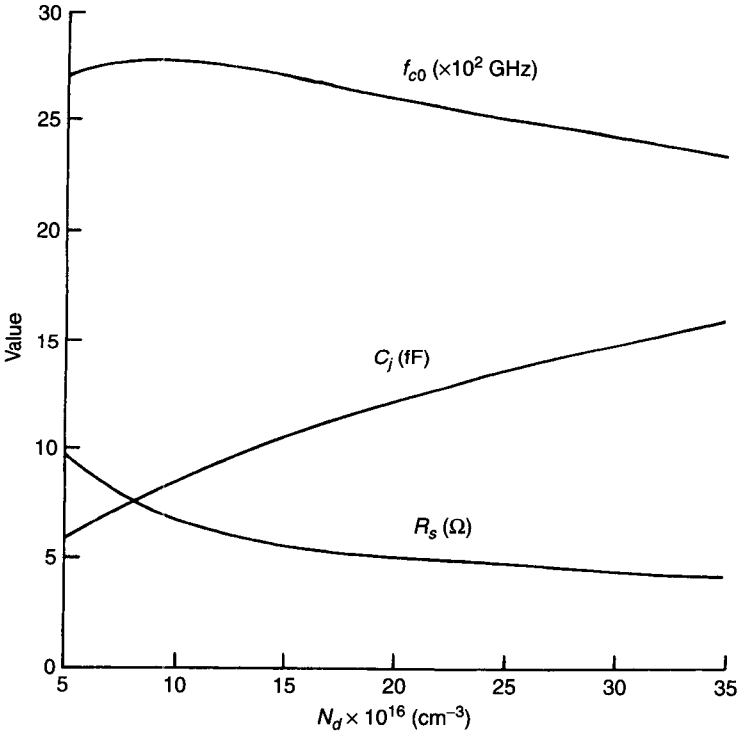
8.3.6 Series Resistance

Series resistance of a Schottky diode (shown in Fig. 8.14*a*) is comprised of three parts: contact resistance (R_c), channel resistance (R_{ch}), and resistance of the Schottky junction metal (R_g). The resistance is written in the form

$$R_s = R_c + R_{ch} + R_g \tag{8.57a}$$



(*b*)



(*c*)

Figure 8.14 *(Continued)*

An approximate expression is

$$R_s = \frac{2100}{P\sqrt{a}(N_d^+)^{0.66}} + \frac{R_{ch}(R_{ch}^+ + R'_{ch})}{R_{ch} + R_{ch}^+ + R'_{ch}} + \frac{1}{3}\frac{\rho_m D}{tW} \tag{8.57b}$$

where

$$R_{ch} = \frac{1100}{aN_d^{0.82}}\frac{L}{A} \qquad R_{ch}^+ = \frac{1100}{a^+(N_d^+)^{0.82}}\frac{L}{A} \qquad R'_{ch} = \frac{1100}{N_d^{0.82}}\frac{a}{WD}$$

and ρ_m is the bulk specific resistivity of the gate metal ($\cong 2.5 \times 10^{-6}$ Ω-cm for gold). The terms N_d and N_d^+ are expressed in units of 10^{16} and the other physical dimensions are in micrometers and are defined in Fig. 8.14*a*.

8.3.7 Equivalent Circuit

The diode equivalent circuit consists of nonlinear junction resistance R_j and capacitance C_j and series resistance R_s, as shown in Fig. 8.14*b*. The values of R_j and C_j depend upon the DC voltage and/or RF voltage across the diode. Typical values for good Schottky diodes are $R_j = 50$–$200\ \Omega$, $C_j \cong 0.02$ pF, and $R_S = 5\ \Omega$.

8.3.8 Figure of Merit

A figure of merit is a useful predictor of how a particular device might be expected to perform in a given circuit. The figure of merit is generally useful when comparing different device types and is often used as a factor in the selection process when a particular circuit is being designed. The figure of merit for Schottky diodes is the zero cutoff frequency, f_{c0}, defined as

$$f_{c0} = \frac{1}{2\pi R_s C_{j0}} \tag{8.58}$$

where C_{j0} is the junction capacitance at 0-V bias. Table 8.2 lists f_{c0} values for various physical and electrical parameters of a beam lead diode. Figure 8.14*c* depicts the variation of R_s, C_j and f_{c0} as a function of N_d for $a = 0.2$ μm, $a^+ = 2$ μm, $N_d^+ = 2 \times 10^{18}$ cm^{-3}, $V_0 = 0.8$ V, $D = 8$ μm, $L = 1.5$ μm, and $W = 1$ μm. These calculations do not include parasitic capacitances.

8.4 VARACTOR DIODES

A varactor diode is a semiconductor diode that is used as a variable-reactance circuit element. The variable-reactance characteristics derive from the variation

Table 8.2 Schottky Diode Characteristics

a (μm)	N_d (10^{16} cm^{-3})	N_d^+ (10^{16} cm^{-3})	R_s (Ω)	C_j (fF)	f_{c0} (GHz)
0.15	5	70	11.1	6	2372
0.15	5	270	7.7	6	3431
0.2	5	270	9.2	6	2853
0.15	7	270	6.4	7.1	3481
0.15	11	270	5.1	9.0	3450

Note: $t = 0.5$ μm, $D = 8$ μm, $L = 1.5$ μm, $W = 1$ μm, $a^+ = 2$ μm, $V_0 = 0.8$ V.

of the diode depletion layer capacitance with applied voltage (either DC or low-frequency RF). For common applications the diode is reverse biased so that the depletion layer capacitance dominates the device characteristics. An excellent review of varactor diodes is presented elsewhere [7].

A varactor diode is a nonlinear element that can produce three fundamentally different circuit functions consisting of

1. microwave signal tuning or modulation,
2. harmonic generation, and
3. parametric amplification and/or up conversion.

In the first two applications a bias voltage and RF signal are applied to the diode, whereas in the third application a bias voltage and multiple RF signals that may be harmonically or nonharmonically related are applied to the diode.

Varactors are commonly and extensively used in many microwave applications, including voltage-controlled oscillators (VCOs), parametric amplifiers (paramps), and frequency multipliers. The VCO application is probably one of the most useful applications for varactors, and these devices have been constructed using a wide range of active devices, including bipolar transistors, GaAs MESFETs, transferred electron devices, and IMPATT diodes. In principle, a VCO can be constructed with a varactor and any active device. Parametric amplifiers are capable of extremely low noise figure operation and, in the past, were extensively employed as front-end amplifiers in high-sensitivity receivers, such as are used in space applications. In recent years, however, the advances achieved in low-noise GaAs MESFET and HEMT technology have resulted in amplifiers constructed with these devices replacing paramps in many low-noise applications. Varactor frequency multipliers are generally used in doubler and tripler circuits. Such circuits have application for producing low-phase-noise sources of microwave and millimeter-wave energy. In particular, these circuits can produce RF signals at frequencies above that obtainable from a fundamental-mode source and for this reason are often used in millimeter-wave applications. Also, since varactors are capable of low-phase-noise performance, they are used in microwave signal synthesizers.

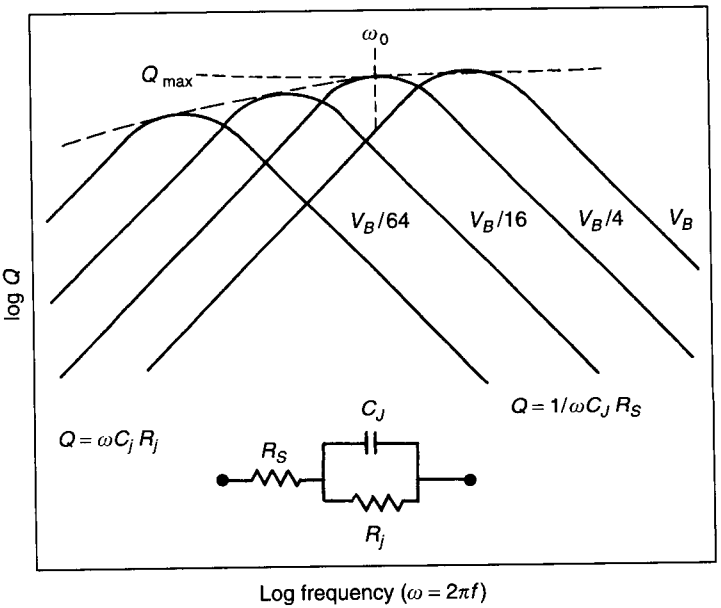


Figure 8.15 Qualitative dependence of Q vs. angular frequency for various reverse-bias voltages up to breakdown voltage V_B. (After Norwood and Shatz [7]. Reprinted with permission of IEEE.)

8.4.1 Equivalent Circuit

The equivalent circuit for a microwave varactor chip is shown in the inset in Fig. 8.15. Since the diode is generally operated in a reverse-biased mode, the equivalent circuit consists of the depletion layer capacitance as defined in (8.32) in shunt with the junction resistance. The depletion layer capacitance is repeated here for convenience and is

$$C_j = \frac{C_0}{(1 - V/V_0)^\gamma} \tag{8.59}$$

For reverse bias the junction resistance R_j is generally high (e.g., on the order of tens of megaohms) and can be ignored. Both the junction capacitance and resistance are functions of bias. The series resistance R_s represents the resistance of the bulk semiconductor regions extrinsic to the junction region and also the resistance due to the metal contacts. The series resistance is related to the Q for the diode, which can be defined as

$$Q \cong \frac{\omega C_j R_j}{1 + \omega^2 C_j^2 R_j R_s} \tag{8.60}$$

The diode Q is an indicator of the efficiency of the device. The variation of diode Q with frequency is shown in Fig. 8.15. The diode breakdown voltage is

indicated as V_B in the figure. The Q expression can be differentiated to determine the frequency at which maximum Q occurs. The frequency is

$$\omega_0 \cong \frac{1}{C_j(R_jR_s)^{1/2}} \tag{8.61a}$$

and

$$Q_{\max} \cong \left(\frac{R_j}{4R_s}\right)^{1/2} \tag{8.61b}$$

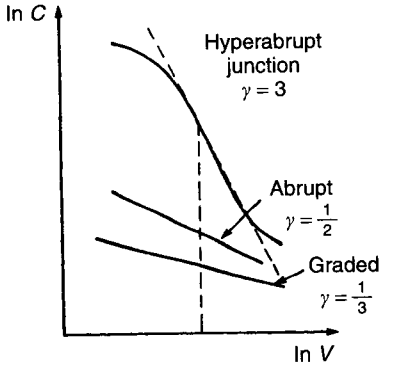
Generally, GaAs varactors have higher Q values than Si devices and are preferred in many high-efficiency and low-noise applications. An exception to this is low-phase-noise applications. The GaAs devices suffer from high baseband $1/f$ noise due to surface states, deep levels, and other trapping effects. There is currently no known passivation that is capable of completely eliminating these effects. The Si devices, however, have the advantage of native oxide passivation that significantly reduces the low-frequency $1/f$ noise. In general, Si devices have a $1/f$ spectrum that is about 10–20 dB lower in magnitude compared to that from GaAs devices. The $1/f$ corner frequencies where the $1/f$ effects can no longer be seen are in the hundreds of kilohertz range for Si devices and in the hundreds of megahertz for GaAs devices.

The capacitance variation of the varactor with voltage, as indicated in (8.59), is dependent upon the magnitude of the exponent γ. The exponent γ, in turn, is directly dependent upon the device doping profile. For devices with abrupt doping profiles $\gamma = \frac{1}{2}$, for those with linearly graded profiles $\gamma = \frac{1}{3}$, and for hyperabrupt profiles $\gamma \to 1$–5. The capacitance–voltage relationships for these diodes are shown in Fig. 8.16. The hyperabrupt diodes are particularly interesting for linear tuning applications, since it is possible to obtain a capacitance variation that is inversely proportional to the magnitude of the applied bias voltage. Such devices, when used in VCOs, are capable of producing oscillators that tune linearly with voltage. Greater than octave frequency coverage can be obtained with these oscillators. The hyperabrupt diode, however, has a lower Q than an abrupt junction diode with the same breakdown voltage and capacitance.

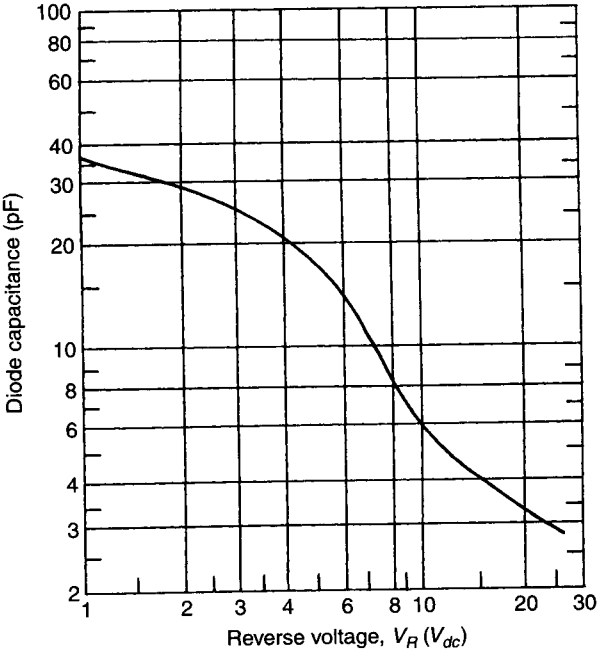
The equivalent circuit shown in the inset in Fig. 8.15 includes only the elements intrinsic to the varactor chip. When the device is bonded into a circuit, either in chip form or in a package, additional parasitic elements must be added to the circuit. The equivalent circuit that results is shown in Fig. 8.17. The parallel capacitance C_f accounts for fringing capacitance due to the mounting structure, and the series inductance L_s accounts for the inductance of the bonding leads.

8.4.2 Figure of Merit

Two types of figures of merit [8], static and dynamic, have been defined for varactor diodes. The static figures of merit include



(*a*)



(*b*)

Figure 8.16 (*a*) Capacitance vs. voltage for diodes with various doping profiles. (*b*) Capacitance vs. voltage for typical hyperabrupt diode.

1. The cutoff frequency, defined as

$$f_{cV} \triangleq \frac{1}{2\pi R_s C_{jV}} \tag{8.62}$$

where C_{jV} is the device capacitance at a specified bias, usually zero or -4 V.

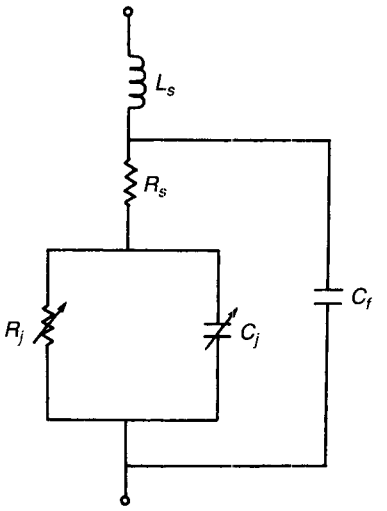


Figure 8.17 Equivalent circuit for packaged varactor diode.

2. The quality factor or device Q, where

$$Q_V \triangleq \frac{f_{cV}}{f} \tag{8.63}$$

and Q is defined for a given bias.

The Q factor in (8.63) is defined in terms of the series resistance in (8.60).
The dynamic figures of merit include

(a) A dynamic cutoff frequency defined as

$$f_c \triangleq \frac{(1/C_{j,\min}) - (1/C_{j,\max})}{2\pi R_s} \tag{8.64}$$

where the maximum and minimum capacitance values are defined as the capacitances at two bias voltages (e.g., zero and -15 V), these values being given in device data sheets.

(b) A dynamic quality factor defined as

$$\tilde{Q} \triangleq \frac{S_1}{\omega R_s} \tag{8.65}$$

where S_1 is the first Fourier component of the time-dependent elastance (i.e., the reciprocal of the capacitance), the diode elastance being written

as

$$S = S_0\left(1 - \frac{V}{V_0}\right)^{\gamma} \tag{8.66}$$

where S_0 is the zero-bias elastance.

An additional figure of merit that is useful in indicating the tuning range of the varactor, called the relative sensitivity, is defined as

$$s(V) \triangleq \frac{1}{\omega_0}\frac{d\omega}{dV} \tag{8.67a}$$

The parameter s indicates the percentage increase in frequency per volt, and a constant s indicates a linear variation in capacitance versus bias voltage. In terms of the diode C–V characteristic, the relative sensitivity can be written as

$$s(V) = \frac{d}{dV}\left[\frac{C_{j0}}{C_j(V)}\right]^{1/2} \tag{8.67b}$$

where C_{j0} is the zero-bias capacitance.

The static figures of merit cannot reflect the degree of capacitance variation of the varactor. The dynamic figures of merit, since they explicitly take the device nonlinearity into account, give a more accurate indication of how a given varactor should function. Since detailed capacitance variations are ignored, however, the indication is only approximate.

8.5 VARISTORS

A varistor is a diode that is designed to utilize the variation of the device resistance that occurs as a function of bias voltage. The variation of resistance with bias for point contact and Schottky barrier varistors is shown in Fig. 8.18 [9]. For reverse bias the resistance is high and the diode is a blocking contact. For forward-bias voltages, however, a range of useful resistance values can be obtained. The diode resistance can be determined from the I–V characteristic expression given for Schottky barrier diodes in (8.53). Since varistors are normally operated in the forward-bias region, *pn* junctions cannot be used. The diffusion capacitance that dominates in these devices under forward bias prevents the device from being used as a variable resistance. The relatively large diffusion capacitance would provide a parallel path for RF energy at high frequencies, thereby tending to short out the diode resistance. The RF resistance for the diode is defined as the slope of the I–V characteristic as shown in Fig. 8.19 and is given as

$$r = \frac{nkTA}{qI} \tag{8.68}$$

where n is the ideality factor previously defined.

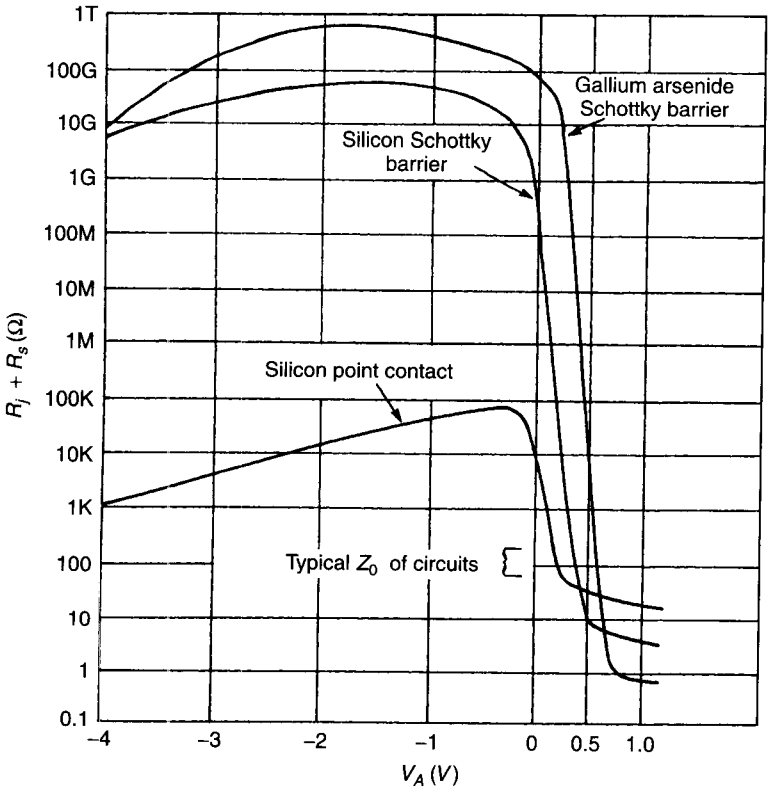


Figure 8.18 Variation of resistance with bias voltage for point contact and Schottky barrier varistors. (After Watson [9]. Reprinted with permission of McGraw-Hill.)

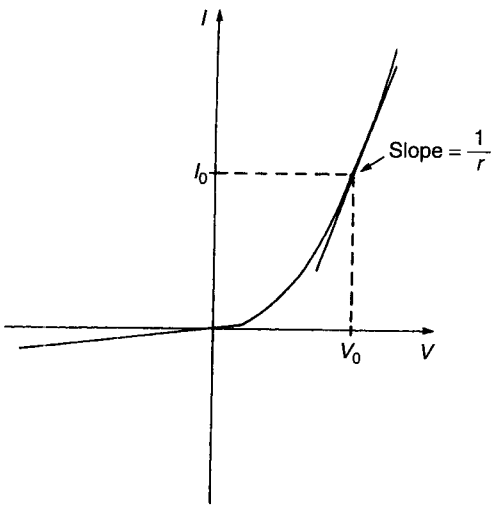


Figure 8.19 Current–voltage characteristic for diode showing RF resistance.

8.6 *pin* DIODES

A *pin* diode is a *pn*-junction device that has a very minimally doped or intrinsic region located between the *p*- and *n*-type contact regions. The addition of the intrinsic, or *i*, region results in characteristics that are very desirable for certain device applications. In reverse bias the *i*-region results in very high values for the diode breakdown voltage, whereas the device capacitance is reduced by the increased separation between the *p* and *n* regions. The diode is very useful, therefore, for high-frequency, high-power rectifier applications. In forward bias the conductivity of the intrinsic region is controlled or modulated by the injection of charge from the end regions. The diode is a bias-current-controlled resistor with excellent linearity and low distortion. The *pin* diodes are used extensively in microwave circuits for amplitude modulation, attenuation, and leveling functions. They also make excellent RF switches, phase shifters, and limiters.

8.6.1 Basic Device Physics

An analysis of *pin* diodes begins with the semiconductor equations presented in Section 7.2. Assuming the one-dimensional, abrupt junction geometry shown in Fig. 8.20, the electron continuity equation can be written in simplified form as

$$\frac{d^2n}{dx^2} + \frac{n}{L_a^2} = 0 \tag{8.69}$$

where $L_a = \sqrt{D_a \tau_a}$, the ambipolar diffusion length [10], and the free charge density is assumed to be much larger than the thermal equilibrium density. This equation is a simple second-order differential equation and has a solution

$$n(x) = A_1 e^{x/L_a} + A_2 e^{-x/L_a} \tag{8.70}$$

The double-injection condition for the diode under forward bias will result in $n(x) \cong p(x)$, and (8.70) can be considered as a general charge density equation. It follows that a minimum will occur at $x = 0$, that is,

$$\left.\frac{dn(x)}{dx}\right|_{x=0} = 0 \tag{8.71}$$

which requires that $A_1 = A_2$, and it follows that the solution can be written as

$$n(x) = 2A_1 \cosh\left(\frac{x}{L_a}\right) \tag{8.72}$$

The constant A_1 is found from the boundary condition determined from the device current. That is, the current density equations can be combined to yield

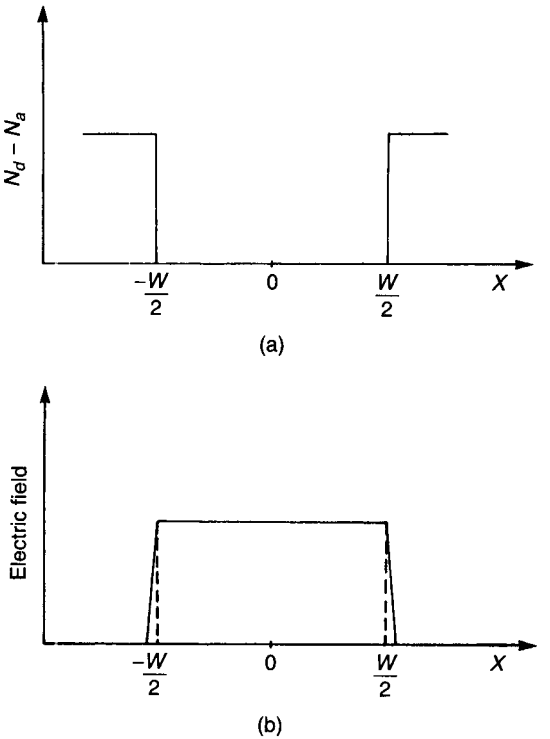


Figure 8.20 (*a*) Doping profile. (*b*) Electric field vs. distance for *pin* diode.

the condition

$$I_0 = J_n A|_{W/2} = 2qD_a A\frac{dn}{dx} \tag{8.73}$$

where A is the cross-sectional area of the device. Applying this condition to determine the constant A_1 gives the solution

$$p(x) = n(x) = \frac{I_0 L_a \cosh(x/L_a)}{2qD_a A \sinh(W/(2L_a))} \tag{8.74}$$

The i-region resistance can be calculated from the expression

$$R_i = \frac{1}{A}\int_{-W/2}^{W/2} \frac{dx}{\sigma(x)} \tag{8.75}$$

where

$$\sigma(x) = q[\mu_n n(x) + \mu_p p(x)] \cong 2q\bar{\mu}n(x)$$

and $\bar{\mu}$ is an effective mobility. Substituting and performing the integration produce an expression for the resistance that can be written for the condition that $W/2L_a \ll 1$ in the form

$$R_i \cong \frac{3(kT)W^2}{8qI_0L_a^2} \tag{8.76}$$

where W is the i-layer thickness (typically 10–100 μm). This expression shows that the resistance is proportional to $(W/L_a)^2$ and inversely proportional to the bias current. The resistance decreases with current and diffusion length (long carrier lifetimes) and increases with diode length.

The i region is shunted by a parallel-plate capacitor determined by the p- and n-region contacts. This capacitance appears in shunt with the junction resistance and can be approximated by the expression

$$C_i = \frac{\epsilon A}{W} \tag{8.77}$$

At forward bias the junction effects must be considered. The junction capacitance and resistance are derived in the same manner as for a *pn* junction and are

$$C_j = A\left[\frac{q\epsilon N_d}{2(V_0 - V)}\right]^{1/2} \tag{8.78}$$

$$R_j = \frac{nkTA}{qI_0} \tag{8.79}$$

At high frequencies and at forward bias the charge storage diffusion capacitance must also be considered. It can be shown that the frequency-dependent complex admittance for the diode is

$$Y = G + jB = \frac{qA}{kT}(J_p\sqrt{1 + j\omega\tau_p} + J_n\sqrt{1 + j\omega\tau_n}) \tag{8.80}$$

For high-frequency operation where $\omega\tau \gg 1$, this expression reduces to the form

$$Y = G + jB = \frac{qI_0}{kT}\sqrt{\frac{\omega\tau}{2}}(1 + j) \tag{8.81}$$

8.6.2 Switching Speed

The switching of a *pin* diode from a low-impedance to a high-impedance state is accomplished by switching the bias current from forward to reverse bias.

Under reverse bias a current will flow until the diode is depleted of charge. The switching time can be defined as the time required to remove the charge stored under forward bias less the charge that recombines in the i region. The switching or reverse recovery time is composed of two components: the time required to remove most of the charge (called the delay time) from the i region and the time during which the diode is changing from a low- to a high-impedance state (called the transition time). The transition time depends upon diode geometry and doping profile but is not sensitive to the magnitude of the forward or reverse current. The delay time is inversely proportional to the magnitude of the reverse-bias current and directly proportional to the charge carrier lifetime. Diodes with short carrier lifetimes have short delay times (i.e., fast switching speeds) but suffer from high values of forward-bias resistance and, therefore, high insertion loss.

8.6.3 Equivalent Circuit

The equivalent circuit for a *pin* diode is shown in Fig. 8.21. The diode basically consists of two main elements in series: a diode with diffusion capacitance C_d and junction parameters R_j and C_j all in parallel and the parallel combination of the undepleted i-region resistance and capacitance. The equivalent-circuit elements were defined in the previous section. The parameter L_s represents the lead inductance, C_f represents fringing capacitance from the top contact of the diode to the mounting structure, which can be neglected for first-order approximation, and C_p represents package and/or mounting structure capaci-

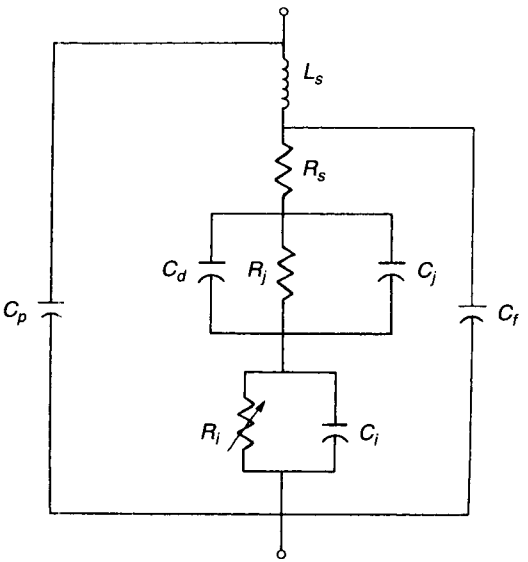


Figure 8.21 Equivalent circuit for *pin* diode.

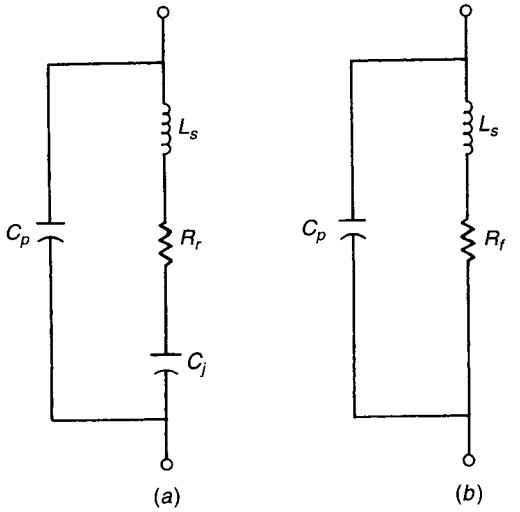


Figure 8.22 Simplified equivalent circuit for *pin* diode under (*a*) reverse and (*b*) forward bias.

tance. The parameter R_s represents the resistance of the bulk semiconductor regions plus the resistance of the contacts.

The equivalent circuit shown in Fig. 8.21 is modified under certain bias and high-frequency operating conditions. For example, under reverse bias the diffusion capacitance C_d vanishes and the junction resistance R_j becomes very large. Only C_j remains significant in the junction network. If the reverse bias is sufficient to completely deplete the i region, the R_i–C_i network will vanish. The equivalent circuit reduces to that shown in Fig. 8.22*a*, where R_r is the total series resistance for the reverse-bias state.

Under forward-bias and high-frequency operation the diffusion capacitance C_d will be large and will short out the junction parameters. The i region will be injected with charge carriers so that the R_i–C_i network is represented by R_i only. The equivalent circuit reduces to that shown in Fig. 8.22*b*, where R_f is the total series resistance for the forward-bias state.

8.6.4 Figure of Merit

Several different methods have been used for expressing the relative merit of two-terminal switching devices exhibiting two distinct impedance states. The ratio of the resistance in the high-impedance state to that in the low-impedance state can be used as a figure of merit of the device at lower frequencies where the device reactances are insignificant. Switching diodes have been traditionally characterized in terms of a frequency parameter known as switching cutoff frequency f_{cs}, which is defined in terms of series resistances in two impedance states and the series capacitance in the high-impedance state. Referring to the

equivalent circuit elements shown in Fig. 8.22, we have

$$f_{cs} = \frac{1}{2\pi C_j \sqrt{R_f R_r}} \tag{8.82}$$

The term switching cutoff is not intended to indicate that a particular device can be used for designing a switching circuit at frequencies as high as f_{cs}. In fact, *pin* diodes are used typically in a frequency range around one-hundredth to one-fiftieth of the switching cutoff frequency. Also it can be shown that losses in switching and phase-shifting circuits are proportional to the ratio of the switching cutoff frequency to the operating frequency.

A more general representation of the figure of merit for two-impedance-state switching devices has been proposed by Kurokawa and Schlosser [11]. When the device impedances in the two states are z_1 $(= r_1 + jx_1)$ and z_2 $(= r_2 + jx_2)$, respectively, the figure of merit Q may be written as

$$Q = \frac{\sqrt{(r_1 - r_2)^2 + (x_1 - x_2)^2}}{\sqrt{r_1 r_2}} = \frac{|z_1 - z_2|}{\sqrt{r_1 r_2}} \tag{8.83}$$

For an ideal diode, the two impedances z_1 and z_2 approach zero and infinity, respectively, and the figure of merit Q tends to infinity. Also for lossless devices $(r_1 = r_2 = 0)$, the factor Q tends to be infinite. An alternative expression for Q is written in terms of complex reflection coefficients Γ_1 and Γ_2 present on a lossless transmission line when terminated in the two-terminal switching device. We have [11]

$$Q' = \frac{|\Gamma_1 - \Gamma_2|^2}{(1 - |\Gamma_1|^2)(1 - |\Gamma_2|^2)} \tag{8.84}$$

This expression for Q' is more useful, as the complex reflection coefficients Γ_1 and Γ_2 can be measured conveniently (e.g., by using a network analyzer). Another major advantage of using Q' as a figure of merit is that Q' is invariant to lossless transformations. That is, the measured value of Q' will remain unchanged when we interpose a lossless two-port network between the switching device and the measurement system. This lossless two-port network can represent package reactances and/or any adapters (or connectors) and/or any bias network used in the measurement setup.

For *pin* diode chips (without any bonding wire), impedances z_1 and z_2 in two bias states may be written as

$$z_1 = R_f \tag{8.85a}$$

$$z_2 = R_r + \frac{1}{j\omega C_j} \tag{8.85b}$$

Table 8.3 Equivalent Circuit Parameters for Two Commercially Available *pin* Diodes

Parameter	MA47892-109	MA47899-030
C_j	1 pF	0.1 pF
R_f	0.4 Ω	1 Ω
R_r	0.5 Ω	4 Ω
L_{int}	0.3 nH	0.3 nH
C_p	0.08 pF	0.18 pF
τ	5 μs	0.5 μs
f_{cs}	350 GHz	800 GHz

and the figure of merit Q may be written as

$$Q = \frac{\sqrt{(R_f - R_r)^2 + (1/\omega C_j)^2}}{\sqrt{R_f R_r}} \simeq \frac{1}{2\pi f C_f \sqrt{R_r R_f}} \tag{8.86}$$

When $|R_f - R_r| \ll 1/\omega C_j$, (8.86) may be simplified as

$$Q \simeq \frac{f_{\text{cs}}}{f} \tag{8.87}$$

where f_{cs} is the switching cutoff frequency defined in (8.82). Values of Q at 3 GHz calculated from parameters in Table 8.3 for the MA47892 and MA47899 diodes are 118.6 and 265.3, respectively. The corresponding values for the switching cutoff frequency are 350 and 800 GHz, respectively.

8.7 STEP RECOVERY DIODES

A step recovery diode (SRD) is a diode that has a very nonlinear conduction characteristic [12, 13]. Although in principle SRDs can be fabricated from any diode structure, practical devices are usually based upon the *pin* structure. Step recovery diodes, which are also called snap-back diodes, are a type of charge storage diode and are useful for frequency multiplication applications where high-efficiency harmonic generation is desirable. Due to the very nonlinear switching characteristics of these diodes, harmonic generation with an efficiency approaching $1/n$, where n indicates the harmonic number, can be obtained. This is in contrast to varactor multipliers where the harmonic generation efficiency is on the order of $1/n^2$. Step recovery diode multipliers require no idler circuits, resulting in very simple and compact circuits. For these reasons SRDs are generally used where high-efficiency, high-order frequency multiplication is required, such as in frequency multipliers and comb generators.

8.7.1 Basic Device Physics

The operation of a SRD depends upon the charge storage characteristics of certain semiconductors. The minority-carrier lifetime in indirect bandgap materials, such as Si, can achieve magnitudes on the order of less than a microsecond to hundreds of milliseconds, depending upon the crystal quality and impurity distribution. Minority carriers in these materials can, therefore, remain in a free state without recombination for times on the order of RF cycle times at microwave frequencies. These materials are useful for fabrication of charge storage devices. Compound semiconductors such as GaAs, for comparison, have direct bandgaps and generally have relatively short minority-carrier lifetimes and are not of use for these applications. The important consideration in a SRD is that the charge extraction time, which is dependent upon the tran-

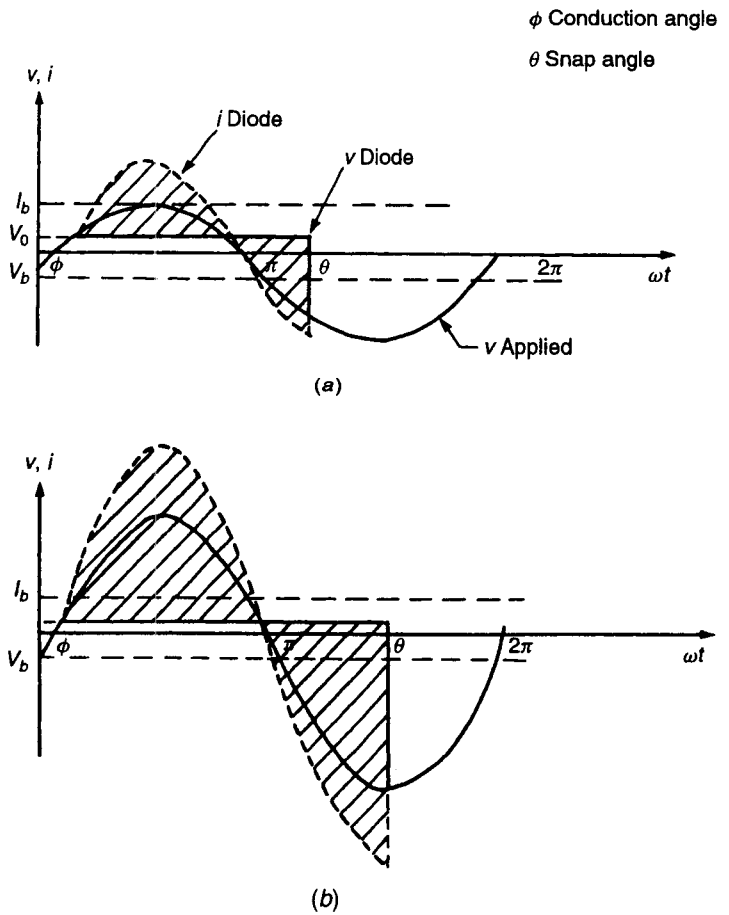
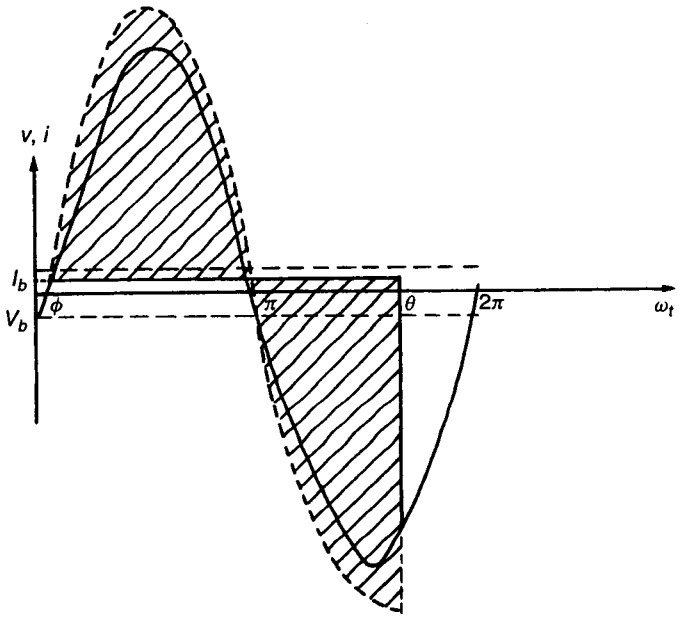


Figure 8.23 Terminal current and voltage waveforms for SRD: (*a*) $V_{RF} = 5$ V; (*b*) $V_{RF} = 10$ V; (*c*) $V_{RF} = 20$ V.



(c)

Figure 8.23 *(Continued)*

sit time of the charge carriers through the diode, be short relative to the minority-carrier lifetimes. For efficient operation the charge should be removed by the terminal bias and not be lost by recombination. Minority-charge recombination represents a loss mechanism that limits the harmonic generation efficiency of these devices.

The operation of a SRD can be qualitively understood with reference to the device terminal waveforms shown in Fig. 8.23. If the diode is biased with bias voltage V_b and a sinusoidal RF voltage is placed across the diode, the diode will conduct RF current when the sum of the DC and RF voltages is greater than the forward threshold voltage, which is typically in the range of $V_0 = 0.8$–1.0 V. The point in the RF cycle when conduction occurs is indicated as the conduction angle ϕ in Fig. 8.23. The RF current is dependent upon the RF voltage and the diode will conduct with the RF current waveform essentially following the RF voltage waveform. The charge injected from the contact regions will result in the i region of the diode becoming filled with an electron–hole plasma that, due to the long carrier lifetimes, will be stored in the diode. As the diode voltage drops below threshold and begins the negative portion of the RF cycle, current conduction will continue due to the charge stored in the structure. The terminal voltage will result in charge extraction from the diode and the device terminal current will continue to follow the diode voltage. The charge extraction time is represented as the angle θ in Fig. 8.23, and it repre-

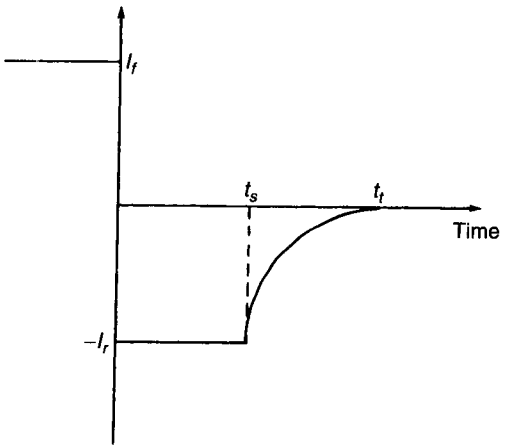


Figure 8.24 Current–time characteristic for *pn* junction diode under transient, switching conditions.

sents the time required to extract the free charges from the diode. The charge extraction time can be estimated from the switching characteristics for *pn*-junction diodes. For constant-amplitude waveforms, as shown in Fig. 8.24, the extraction time is approximated by the expression [14]

$$t_s = \tau \ln\left(1 + \frac{I_f}{I_r}\right) \tag{8.88}$$

where τ is the carrier lifetime, I_f is the forward current just before switching occurs, and I_r is the reverse current just after switching occurs.

Once the free charge has been removed from the diode, the current will be reduced to zero in a very short time. The current will "snap" from its reverse conduction magnitude (which can be quite substantial) to zero in a very short time. The transition time is represented by the angle θ in Fig. 8.23. The transition time for graded profile devices has been derived by Moll et al. [15] and is

$$t_t = \frac{1}{D_a}\left[\frac{Q_0/A}{4qb(N_0 b)}\right]^{2/3} \tag{8.89}$$

where D_a is the ambipolar diffusion coefficient [which is defined as $D_a = 2D_nD_p/(D_n + D_p)$], A is the diode area, Q_0 is the initial charge stored, and $N_0 b$ is the impurity gradient at the junction.

The rapid decrease in terminal current creates device waveforms that are very rich in harmonics. If a high-Q resonant circuit is placed across the diode, the resonant circuit will be "shocked" into a high-efficiency ringing mode. By tuning the resonant frequency of the circuit, selective harmonics can be extracted. Such a circuit should be parallel tuned so that all frequencies out of

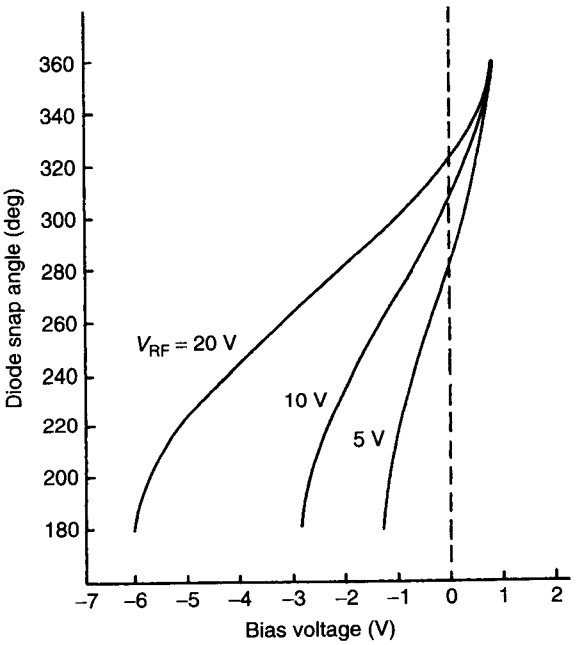


Figure 8.25 SRD snap angle vs. bias and RF voltages.

the resonator bandwidth will be terminated in a short circuit. All energy will be focused at the resonant frequency of the circuit and high-efficiency harmonic generation results. If the diode is operated in a low-Q circuit, the output signal will contain many harmonics. This type of operation is utilized in comb generators, which are useful for marker circuits.

For optimum operation, the snap angle should occur when the RF current is at a maximum. The snap angle is determined by the RF and DC voltages placed across the diode, as shown in Fig. 8.25. For a given RF voltage, an increase in negative bias voltage will cause the snap angle to occur sooner in the RF cycle. In typical harmonic generator circuits, the SRD can be either self-biased through a resistor or externally biased. Bias voltages are generally near zero and may be positive or negative. The bias voltage can be adjusted to achieve optimum harmonic generation efficiency.

8.7.2 Frequency Limits

The operation of the SRD is dependent upon charge recombination. If the stored charge remains in the device for relatively long periods compared to the RF cycle, recombination will occur and the device efficiency will be reduced. This consideration places a lower limit upon the frequency at which a given device may be expected to operate in a satisfactory mode. The lower frequency limit is defined as the inverse of the charge lifetime as

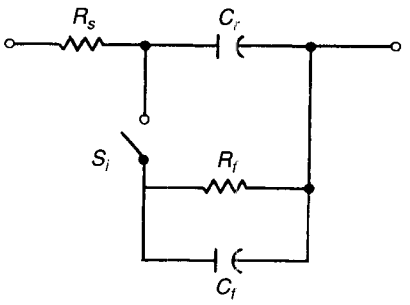


Figure 8.26 Full equivalent circuit for SRD.

$$f_{\text{low}} = \frac{1}{2\pi\tau} \tag{8.90}$$

The upper frequency limit results from the switching time considerations. An upper frequency at which the diode is capable of efficient harmonic generation is defined from the inverse of the transition time and is

$$f_{\text{high}} = \frac{1}{2\pi t_t} \tag{8.91}$$

8.7.3 Equivalent Circuit

The SRD is a form of electronic switch, and an equivalent circuit that is capable of describing its operation must include both forward- and reversebias operation. Under forward bias the equivalent circuit consists of a relatively large diffusion capacitance in shunt with a resistance. The diffusion capacitance accounts for the large charge storage in the device and the resistance accounts for the forward current that may flow. Under reverse bias very little steady state current flows. A reverse-bias equivalent circuit consists almost entirely of the depletion layer capacitance.

A complete equivalent circuit is obtained by combining the forward- and reverse-bias circuits with a switch, as shown in Fig. 8.26. The series resistance is included to account for the voltage drop across the bulk material. The switch is closed under forward bias and remains closed initially during reverse bias until the time at which all the charge is extracted from the capacitor C_f. A model based upon this equivalent circuit has been presented in the literature [16].

REFERENCES

1. Henisch, H. K., *Rectifying Semi-Conductor Contacts*, Clarendon, Oxford, 1957.
2. Sze, S. M., *Physics of Semiconductor Devices*, Wiley-Interscience, New York, 1981, pp. 94–96.

3. Mott, N. F., "The Theory of Crystal Rectifiers," *Proc. Roy. Soc.* (*London*), Vol. 171A, May 1939, pp. 27–38.
4. Bardeen, J., *Phys. Rev.*, Vol. 71, 1947, p. 717.
5. Schottky, W., *Naturwiss.*, Vol. 26, 1938, p. 843.
6. Howes, M. J., and D. V. Morgan, *Variable Impedance Devices*, Wiley-Interscience, New York, 1978, pp. 51–62.
7. Norwood, M. H., and E. Shatz, "Voltage-Variable Capacitor Tuning: A Review," *Proc. IEEE*, Vol. 56, May 1968, pp. 788–798.
8. Watson, H. A., *Microwave Semiconductor Devices and Their Circuit Applications*, McGraw-Hill, New York, 1969, Chap. 7.
9. Watson, H. A., *Microwave Semiconductor Devices and Their Circuit Applications*, McGraw-Hill, New York, 1969, p. 353.
10. McKelvey, J. P., *Solid-State and Semiconductor Physics*, Harper & Row, New York, 1966, Chap. 10.
11. Kurokawa, K., and W. O. Schlosser, "Quality Factor of Switching Diodes for Digital Modulation," *Proc. IEEE*, Vol. 58, Jan. 1970, pp. 180–181.
12. Boff, A. F., J. Moll, and R. Shen, "A New High-Speed Effect in Solid-State Diodes," *IEEE Solid-State Circuits Conf. Dig. Tech. Papers*, 1960, pp. 50–51.
13. Krakauer, S., "Harmonic Generation, Rectification and Lifetime Evaluation with the Step Recovery Diode," *Proc. IRE*, Vol. 50, July 1962, pp. 1665–1676.
14. Steetman, B. G., *Solid-State Electronic Devices*, Prentice-Hall, Englewood Cliffs, NJ, 1980, Chap. 5.
15. Moll, J. L., S. Krakauer, and R. Shen, "*P-N* Junction Charge-Storage Diodes," *Proc. IRE*, Vol. 50, Jan. 1962, pp. 43–53.
16. Kotzabue, K. L., "A Circuit Model of the Step Recovery Diode," *Proc. IEEE*, Dec. 1965, p. 2119.

PROBLEMS

8.1 A Schottky barrier diode is made by evaporating a dot of gold (Au) of area $A = 10^{-3}$ cm^2 onto an n-type Si crystal doped to a level of $N_d = 10^{14}$ cm^{-3}. For operation at room temperature:

(a) What current transport mechanism would you expect to dominate? Why?

(b) Calculate the magnitude of the "reverse saturation current," I_0, where

$$I = I_0(e^{qV/kT} - 1)$$

8.2 The electron concentration in a semiconductor can be expressed as a function of the quasi-Fermi level by the expression

$$n = n_i e^{(E_{Fn} - E_i)/kT}$$

Show that the electron current can be expressed as a function of the spa-

tial variation of the quasi-Fermi level, that is,

$$J_n \propto \frac{\partial E_{Fn}}{\partial x}$$

This calculation shows that the quasi-Fermi level must be constant with distance for zero net current flow. We used this condition in our discussion of junctions.

8.3 Two convenient devices that have rectification properties are the Si *pn*-junction diode and the GaAs Schottky barrier diode. Discuss which of these diodes you would use for

(a) high-frequency RF detection,

(b) high-speed switching, and

(c) high-power rectification.

Discuss fundamental properties and limitations. Be as complete as possible.

8.4 Consider a *pn* junction formed from uniformly (constant) doped *p*- and *n*-type semiconductor regions. Assume one dimension.

(a) Derive expressions for the minority-charge densities on each side of the junctions.

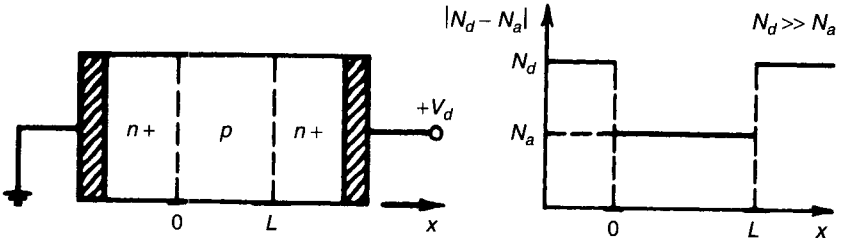
(b) Sketch $p(x)$ and $n(x)$ throughout the device for forward and reverse bias.

(c) Derive expressions for the minority-carrier currents on each side of the junction. State all approximations or assumptions you make.

(d) Sketch the hole, electron, and total currents that flow through the device for forward and reverse bias.

(e) Derive the ideal diode equation. State all assumptions or approximations you make.

8.5 An n^+pn^+ diode is fabricated as shown in the sketch.



(a) Derive an expression for the punch-through voltage (V_{pt}) defined as the drain voltage required to completely deplete the *p* region of mobile charge.

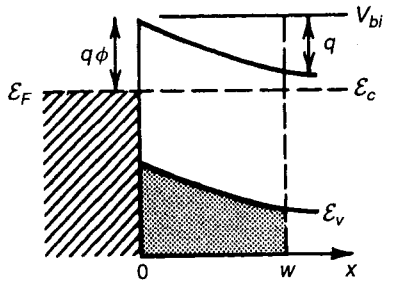
(b) Derive an expression for the drain voltage (V_{fb}) that would need to be applied to the device in order to completely eliminate any potential barrier internal to the device. Neglect any series resistance external to the p region.

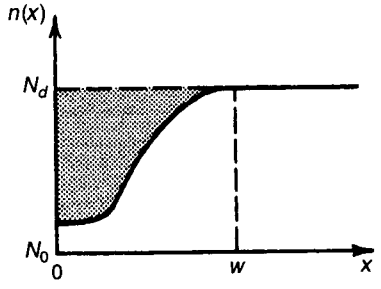
(c) Sketch the energy band diagrams and internal electric field versus distance for the three values of drain voltage $V_d = 0$, $V_d = V_{pt}$, and $V_d = V_{fb}$.

8.6 The limiting forward bias across a pn junction is equal to the contact potential, V_0. This effect is not predicted by the ideal diode equation. The reason for this discrepancy is the neglect of changes in the majority-carrier concentrations on each side of the junction at high injection levels. For the case of high-level injection and following a procedure similar to that used in deriving the ideal diode equation, show that a general diode equation can be derived that has the form

$$J = q\left(\frac{e^{qV/kT} - 1}{1 - e^{-2q(V_0 - V)/kT}}\right)\left[\frac{D_p p_{n0}}{L_p}\left(1 + \frac{n_i^2}{p_{p0}^2}e^{qV/kT}\right) + \frac{D_n n_{p0}}{L_n}\left(1 + \frac{n_i^2}{n_{n0}^2}e^{qV/kT}\right)\right]$$

8.7 Consider a Schottky barrier diode fabricated by placing a metal on a uniformly doped n-type semiconductor as shown in the sketch.





At $x = 0$	At $x = W$
$v = -v_{bi}$	$v \simeq 0$
$E = E_{max}$	$E \simeq 0$
$n = n_0$	$n \simeq N_d$
$\frac{dE}{dx} \simeq \frac{qN_d}{\epsilon}$	$\frac{dE}{dx} \simeq 0$

(a) Derive an expression for the space charge due to the positive donor atoms within the depletion region.

(b) Use your expression from part (a) to derive an expression for the capacitance of the junction where

$$C \triangleq \left|\frac{\partial Q}{\partial v}\right|$$

(c) Use your expression from part (b) to derive an expression for the width of the depletion region W (i.e., the space-charge region).

9

OSCILLATORS

A. P. S. Khanna

9.1 INTRODUCTION

Oscillators represent the basic microwave energy source for all microwave systems, such as radars, communications, navigations, or electronic warfare. They can be termed as DC-to-RF converters or infinite-gain amplifiers. A typical microwave oscillator essentially consists of an active device (a diode or a transistor) and a passive frequency-determining resonant element, such as a microstrip, SAW, cavity resonator, or dielectric resonator for fixed-tune oscillators and a varactor or YIG sphere for tunable oscillators. With the rapid advancement of microwave technology, there has been an increasing need for better performance of oscillators. The emphasis has been on low noise, small size, low cost, high efficiency, high-temperature stability, and reliability for all oscillators and additionally on wider bandwidths, better tuning linearity, and reduced setting time for the tunable oscillators.

Increased capability of modern technology to utilize the pure microwave signal and to measure precisely its characteristics has necessitated the development of highly stable signal sources having minimum FM and AM noise and maximum long- and short-term frequency stability with time and temperature. For application as a local oscillator in communication systems or radar, the frequency stability is often one of the important factors in determining the receiver intermediate-frequency (IF) bandwidth and for bandwidth allocations for communication systems. A narrower IF bandwidth improves the receiver performance by improving its noise figure. In Doppler radars, the short-term stability of the signal is of utmost importance for positive detection of moving

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

targets; low oscillator noise in a transmitter source ensures a high signal-to-noise ratio (SNR). When used in digital communication systems, oscillators generally require good phase noise performance. Poor phase noise can increase the bit error rate (BER) of communication systems based on digital modulation schemes including binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), and higher levels of quadrature amplitude modulation (QAM).

The speed of fiber-optic communication systems has been increasing steadily. Presently 10-Gb/s systems are in production while 40-Gb/s systems are being developed. An important application of oscillators is their use as a clock source in these high-speed fiber-optic communication systems. These oscillators are used to synchronize the MUX and DEMUX functions in a given transceiver. These oscillators need to meet low phase noise and are required to have high modulation bandwidth.

Quartz crystal oscillators represent highly stable sources, but their operation is limited to a few hundred megahertz. At microwave frequencies, stable sources have commonly been realized using frequency multiplication by a factor of N of a quartz crystal oscillator. This method increases the FM noise power by N, has a very low efficiency, and is very complex and expensive. Stable signals have also been generated in the past using huge metallic high-Q cavities in passive cavity stabilization systems or in complex and bulky frequency discriminator systems.

A dielectric resonator, due to its high Q, small size, and integration capability in MIC circuits, can be directly used as a frequency-determining element for realizing a stable MIC transistor oscillator; the transistor dielectric resonator node (TDRO) is a common choice for a vast number of applications of fixed-frequency oscillators.

Frequency-tunable oscillators find application in instrumentation, electronic warfare systems, communication systems, and phase-locked oscillator schemes. The YIG-tuned oscillators have the widest tuning bandwidths with 2–20 and 18–40 GHz tuning capability with single spheres and single transistors [1, 2, 3]. However, these oscillators are slow in tuning, bulky, and much less efficient compared to varactor-tuned oscillators. The YIG-tuned oscillators do present much better phase noise and tuning linearity compared to varactor-controlled oscillators (VCOs), which can provide up to an octave of bandwidth [4] and are commonly used in applications where size, fast tuning, and efficiency are the main considerations.

9.2 ACTIVE DEVICES FOR MICROWAVE OSCILLATORS

The origin of solid state microwave oscillators using Gunn and IMPATT diodes dates back to the late 1960s, before which microwave sources used to be massive klystron or magnetron tubes requiring huge power supplies. In less than two decades, solid state oscillators have come a long way. The extension

of the silicon bipolar transistor oscillator to microwave frequencies and the development of GaAs MESFET devices in the early 1970s have made available highly cost effective, miniature, reliable, and low-noise sources for use right up to the millimeter-wave frequency range.

Basic active elements that can be used for microwave solid state oscillators are Gunn diodes, IMPATT diodes, and transistors. While a Gunn oscillator has the advantage of low FM noise over an IMPATT oscillator, the latter had a higher efficiency and a higher power output as compared to the Gunn. Transistor oscillators, on the other hand, are low-noise as well as high-efficiency sources. Compared to Gunn oscillators, transistor oscillators also do not have the problems of threshold current, heat sinking, and the tendency to lock at spurious frequencies. The Gunn and IMPATT diodes are inherently negative-resistance devices, which means that just the application of the required DC bias is sufficient to produce negative resistance. The design of the oscillator is thus simplified to the design of the output matching circuit to deliver the desired power output. Application of a DC bias to the bipolar or the FET, on the other hand, requires a suitable series–parallel feedback mechanism to induce the negative resistance. The frequency range over which the negative resistance is present in the diodes is determined by the physical mechanisms in the device, while in the case of transistors this frequency range is also influenced by the chosen circuit topology. Compared to transistors, both Gunn and IMPATT devices are capable of generating higher power outputs at higher frequencies.

Transistor oscillators can be realized using bipolar or FET devices. Silicon bipolar and GaAs FET devices (see Chapter 7) have been used for oscillators for quite some time. Due to their low $1/f$ noise characteristics, silicon bipolar devices have been the preferred choice for oscillators. However, GaAs FET devices offer higher frequency operation and higher power output. Standard silicon bipolar devices can be used for oscillations up to about 20 GHz [2] while GaAs FET devices are usable up to greater than 60 GHz.

Newer devices, including SiGe-, GaAs-, and InP-based heterojunction bipolar devices, are now being used for applications as oscillators [5–8]. Heterojunction bipolar transistor (HBT) devices have the performance of bipolar transistors translated to higher frequencies. These technologies are expected to enable faster integration of high-performance oscillators as a part of complex multifunction monolithic ICs. The silicon–germanium (SiGe) HBT has shown excellent promise in oscillator applications for both frequency response and circuit delay. In the SiGe HBT, germanium is introduced into the base layer of an otherwise all-silicon bipolar transistor, creating significant improvements in operating frequency, current, noise, and power capabilities. Presently SiGe offers f_t of greater than 50 GHz and f_{max} of greater than 70 GHz, which combines the integration and cost benefits of silicon with the speed of more expensive technologies such as gallium–arsenide. Oscillator design approaches presented in this chapter are applicable to oscillators of any of the three terminal devices.

9.3 CONCEPT OF NEGATIVE RESISTANCE

The concept of negative resistance is as old as oscillators themselves. Contrary to a positive resistance, a negative resistance is considered as a source of electrical energy. A negative resistance implies that the device is active, while an active device does not imply a negative resistance. A two-terminal device like a Gunn or IMPATT diode inherently has negative resistance, while a three-terminal device like a transistor needs an appropriate impedance to be connected to one or more terminals to create the negative resistance. An important characteristic of the negative resistance is that it is a nonlinear function of the RF current through it. When a load R_l is connected to the negative resistance R_n, with $R_l < R_n$, RF current starts flowing through the circuit at a frequency at which the imaginary parts of both impedances cancel each other. This RF current causes the value of the negative resistance to change unless the oscillation condition of $-R_n = R_l$ is satisfied. The concept of negative resistance, though important to an understanding of the oscillation mechanism, is not very practical to deal with at microwave frequencies. It is not possible to make any significant measurements at these frequencies in terms of R, L, and C. The use of negative resistance for oscillator analysis can be misleading unless caution is exercised to use negative-conductance analysis when appropriate [9]. The most common approach is by using reflection–transmission coefficients, that is, S parameters. Negative resistance–conductance at any terminal can, however, be converted to a reflection coefficient (Γ) by using the following relations:

$$r_n = \frac{R_n}{Z_0} = \frac{1 - |\Gamma|^2}{1 - 2|\Gamma| \cos\theta + |\Gamma|^2} \tag{9.1}$$

$$g_n = \frac{G_n}{Y_0} = \frac{1 - |\Gamma|^2}{1 + 2|\Gamma| \cos\theta + |\Gamma|^2} \tag{9.2}$$

The extended Smith chart shown in Fig. 9.1 shows the negative-resistance region, that is, the area outside the normal Smith chart. The shaded area represents the normal Smith chart. This chart can be used to transform a negative impedance to a reflection coefficient and vice versa.

In view of the fact that the S-parameter approach is the most practical approach at microwave frequencies, it will be used for oscillator analysis and design here.

9.4 THREE-PORT S-PARAMETER CHARACTERIZATION OF TRANSISTORS

The transistor, a three-port device, is generally characterized by its two-port S parameters with one of its ports grounded. The resulting three different configurations in the case of a GaAs FET are shown in Fig. 9.2. Each configuration

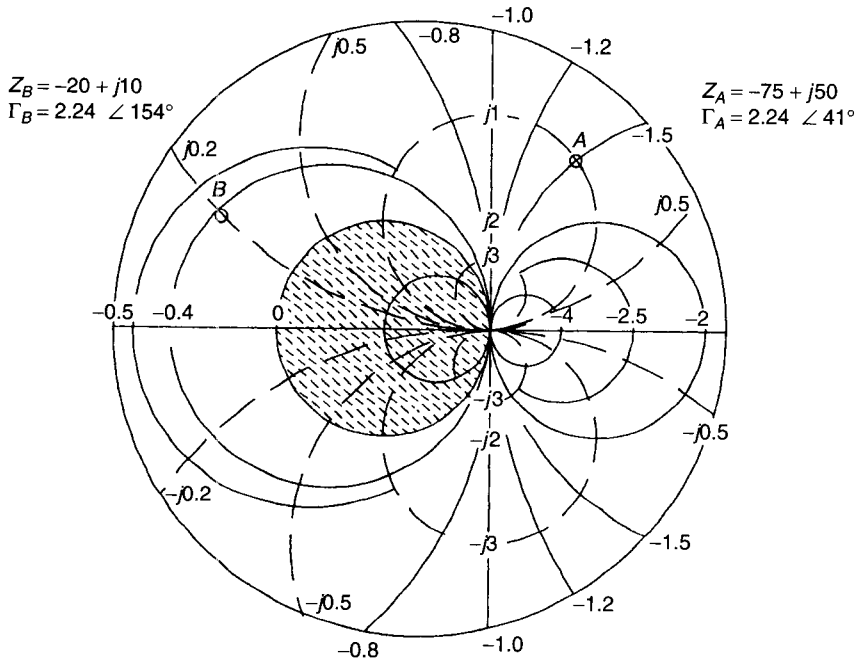


Figure 9.1 Extended Smith chart.

has its own advantages (e.g., the common-source configuration is used more often for amplifiers, common-gate configuration for wide-band oscillators, and common-drain configuration for medium-power oscillators). The use of three-port S parameters, although introduced quite sometime back [10], has not often been used due to the complexity of the analysis involved. The availability of desktop computers and CAD has now made their use practical. The use of three-port S parameters eliminates the otherwise necessary conversion to and from Z and Y parameters to analyze the series and parallel feedback effect, as

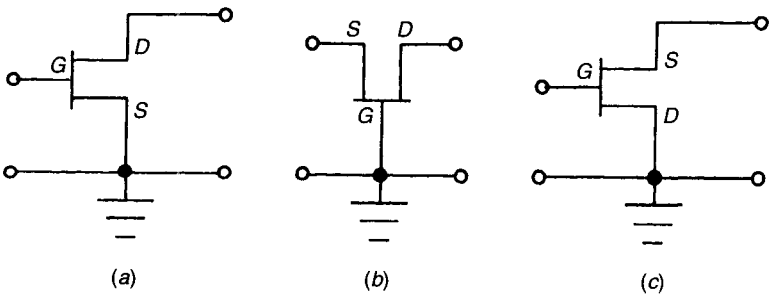


Figure 9.2 Three configurations of transistors: (*a*) common source; (*b*) common gate; (*c*) common drain.

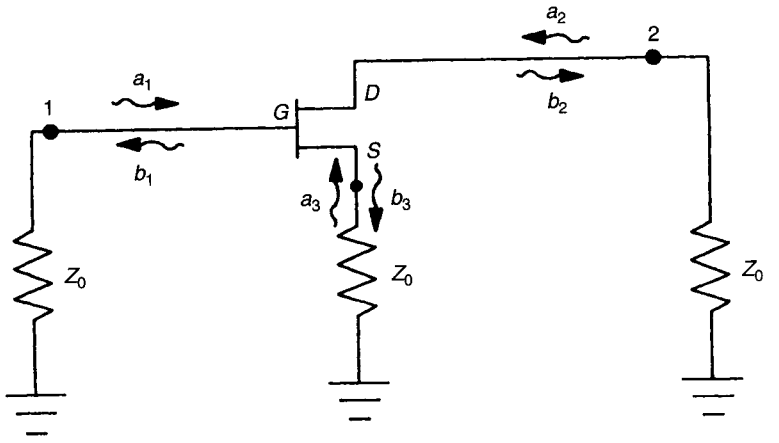


Figure 9.3 Transistor as a three-port device.

shown later in this chapter. The three-port indefinite S matrix of the transistor holds the property of having the sums of the rows and columns to be equal to 1, which helps in determining and eventually correcting the systematic errors in the measurement or analysis. The use of the three-port S parameters of the transistor sometimes become essential in the design of oscillators [11].

The transistor as a three-port device is shown in Fig. 9.3, and the S matrix of the incident and reflected waves is given by

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \tag{9.3}$$

The ports 1, 2, and 3 represent gate, drain, and source, respectively. This indefinite three-port S matrix satisfies the conditions

$$\sum_{j=1}^{3} S_{ij} = 1 \quad \text{for } i = 1, 2, 3 \tag{9.4a}$$

$$\sum_{i=1}^{3} S_{ij} = 1 \quad \text{for } j = 1, 2, 3 \tag{9.4b}$$

The three-port S parameters of a transistor can be measured or derived from commonly available two-port S parameters as discussed next. Equations (9.4a) and (9.4b) represent six equations in terms of nine three-port S parameters to be determined. Other equations necessary to determine all the nine S parameters of the three-port can be found as follows [12]: Using (9.3) we can write

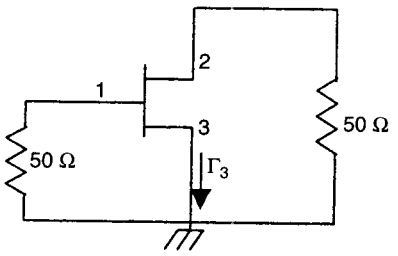


Figure 9.4 Common-source transistor configuration.

$$b_1 = S_{11}a_1 + S_{12}a_2 + S_{13}a_3 \tag{9.5a}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 + S_{23}a_3 \tag{9.5b}$$

$$b_3 = S_{31}a_1 + S_{32}a_2 + S_{33}a_3 \tag{9.5c}$$

If port 3 of the transistor shown in Figure 9.4 is connected to ground, making it a common-source configuration, we have

$$\Gamma_3 = \frac{a_3}{b_3} = -1 \tag{9.5d}$$

Using (9.9) in (9.6)–(9.8) and eliminating a_3 and b_3, we have:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} - \dfrac{S_{13}S_{31}}{1+S_{33}} & S_{12} - \dfrac{S_{13}S_{32}}{1+S_{33}} \\ S_{21} - \dfrac{S_{31}S_{23}}{1+S_{33}} & S_{22} - \dfrac{S_{23}S_{32}}{1+S_{33}} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{9.5e}$$

or

$$(b) = (S^T)(a) \tag{9.5f}$$

where S^T represents the reduced two-port S matrix of the transistor with terminal pair 3 connected to ground. Four equations represented by this relation when combined with 6 equations from (9.4a) and (9.4b) represent the necessary 10 equations to be solved to fully determine all nine three-port S parameters. Final relations thus obtained are given below in simplified form:

$$S_{33} = \frac{S_{11}^T + S_{12}^T + S_{21}^T + S_{22}^T}{4 - (S_{11}^T + S_{12}^T + S_{21}^T + S_{22}^T)} \tag{9.5g}$$

$$S_{32} = 0.5(1 + S_{33})(1 - S_{12}^T - S_{22}^T) \tag{9.5h}$$

$$S_{23} = 0.5(1 + S_{33})(1 - S_{21}^T - S_{22}^T) \tag{9.5i}$$

$$S_{22} = S_{22}^T + \frac{S_{23}S_{32}}{1 + S_{33}} \tag{9.5j}$$

$$S_{11} = 1 - S_{21} - S_{31} \tag{9.5k}$$

$$S_{13} = 1 - S_{23} - S_{33} \tag{9.5l}$$

$$S_{31} = 1 - S_{33} - S_{32} \tag{9.5m}$$

$$S_{12} = 1 - S_{22} - S_{32} \tag{9.5n}$$

$$S_{21} = 1 - S_{22} - S_{23} \tag{9.5o}$$

As an example, the two-port S parameters of a typical 0.5-μm GaAs FET at 10 GHz and the three-port S parameters calculated using these relations are given as

$$[S]_{2P} = \begin{bmatrix} 0.73/\underline{-102^\circ} & 0.1/\underline{42^\circ} \\ 2.23/\underline{96^\circ} & 0.54/\underline{-49^\circ} \end{bmatrix}$$

$$[S]_{3P} = \begin{bmatrix} 0.85/\underline{-56.6^\circ} & 0.29/\underline{66.4^\circ} & 0.61/\underline{47.3^\circ} \\ 1.26/\underline{104.7^\circ} & 0.83/\underline{-37.1^\circ} & 0.97/\underline{-47.7^\circ} \\ 0.99/\underline{-30.9^\circ} & 0.32/\underline{47.2^\circ} & 0.28/\underline{104.2^\circ} \end{bmatrix}$$

9.5 OSCILLATION AND STABILITY CONDITIONS [13]

Any oscillator can be represented in an arbitrary plane on the output line by a nonlinear impedance Z_{NL}, having a negative real part, in series with a load impedance Z_L (Fig. 9.5). We assume that the circuit has a sufficiently high Q factor to suppress the harmonic currents. Supposing that a current

$$i(t) = I_0 \cos(\omega_0 t) \tag{9.6}$$

exists in the circuit, we can apply the Kirchhoff voltage law and write in the plane PP':

$$[Z_{NL}(I_0, \omega_0) + Z_L(\omega_0)]I_0 = 0 \tag{9.7}$$

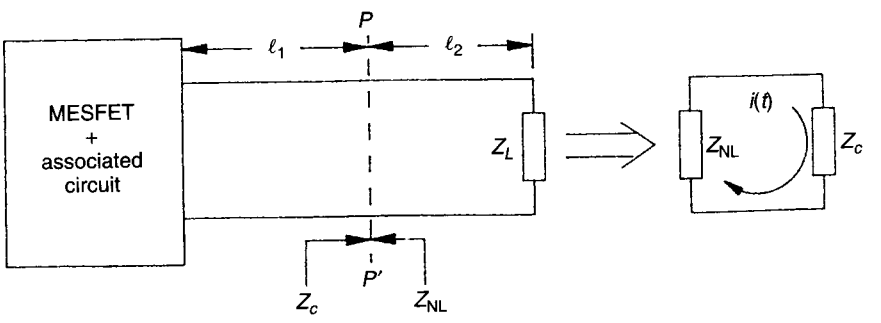


Figure 9.5 Nonlinear microwave oscillator.

Let

$$Z_{NL} + Z_L = Z_T = R_T + jX_T \tag{9.8}$$

Since I_0 is not equal to zero, (9.7) is satisfied by

$$R_T(I_0, \omega_0) = 0 \tag{9.9}$$

$$X_T(I_0, \omega_0) = 0 \tag{9.10}$$

Since $\mathrm{Re}(Z_L) > 0$, (9.9) implies that $\mathrm{Re}(Z_{NL}) < 0$. Hence, the device needs to present a negative resistance in order to be able to oscillate. The frequency of the oscillations is determined by (9.10), that is, by the requirement that the load reactance be equal and opposite to the device reactance.

Oscillators can also be represented by a nonlinear admittance Y_L. The oscillation conditions in this case can be determined in the same way to be

$$G_T(V_0, \omega_0) = 0 \tag{9.11}$$

$$B_T(V_0, \omega_0) = 0 \tag{9.12}$$

At microwave frequencies, it is more convenient to express (9.9)–(9.12) in terms of the corresponding reflection coefficients Γ_{NL} and Γ_L as

$$|\Gamma_{NL}| \cdot |\Gamma_L| = 1 \tag{9.13}$$

$$\underline{/\Gamma_{NL}} + \underline{/\Gamma_L} = 2\pi n \qquad n = 0, 1, 2, \ldots \tag{9.14}$$

Relation (9.13) implies that the device reflection coefficient Γ_{NL} modulus should be greater than unity.

An oscillator can be considered as a combination of an active multiport and a passive multiport (the embedding circuit), as shown in Fig. 9.6. With the active device and the embedding circuit characterized by their scattering

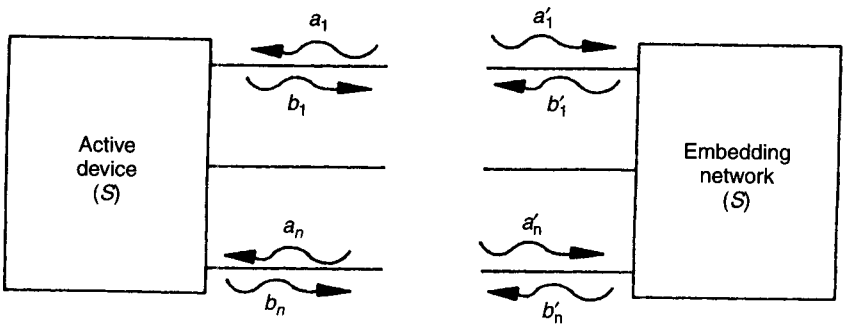


Figure 9.6 Generalized oscillator configuration.

matrix, we have, for the active device [14],

$$[b] = [S][a] \tag{9.15}$$

and for the embedding circuit,

$$[b'] = [S'][a'] \tag{9.16}$$

When the active device and the embedding network are connected together, we have, for the oscillation conditions,

$$[b'] = [a] \tag{9.17}$$

$$[b] = [a'] \tag{9.18}$$

From (9.15)–(9.18) we can write

$$[a'] = [S][S'][a'] \tag{9.19}$$

or

$$([S][S'] - [I])[a'] = 0 \tag{9.20}$$

where $[I]$ is an identity matrix. Since $[a'] \neq 0$, it follows that

$$[M] = [S][S'] - [I] \tag{9.21}$$

is a singular matrix or

$$\det[M] = 0 \tag{9.22}$$

which represents the generalized large-signal oscillation condition for an n-port oscillator.

In fact, the scattering matrix of the active device being defined at the small-signal level, the n-port oscillation condition at the small signal can be represented by

$$|\det([S][S'] - [I])| > 0 \tag{9.23}$$

$$\arg \det([S][S'] - [I]) = 0 \tag{9.24}$$

The oscillations can start as soon as the preceding relations are satisfied and go on building up until the device nonlinearities cause a steady state to be reached. As an example, consider an active two-port loaded by two passive impedances, as shown in Fig. 9.7. The active device is described by the scattering matrix

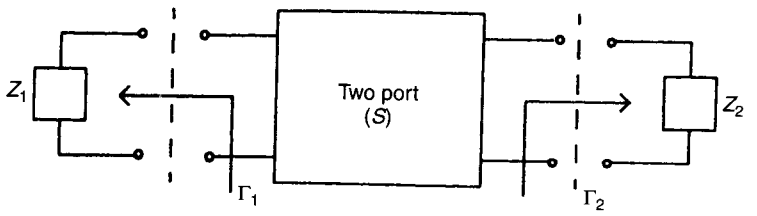


Figure 9.7 Two-port loaded with two impedances.

$$[S] = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \tag{9.25}$$

and the embedding circuit by

$$[S'] = \begin{bmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{bmatrix} \tag{9.26}$$

The oscillation condition from (9.22) is

$$\det[M] = \det \begin{bmatrix} S_{11}\Gamma_1 - 1 & S_{12}\Gamma_2 \\ S_{21}\Gamma_1 & S_{22}\Gamma_2 - 1 \end{bmatrix} = 0 \tag{9.27}$$

which gives

$$(S_{11}\Gamma_1 - 1)(S_{22}\Gamma_2 - 1) - S_{12}S_{21}\Gamma_1\Gamma_2 = 0 \tag{9.28}$$

From the preceding, we obtain the following two well-known conditions that are simultaneously satisfied for realizing oscillations [15]:

$$S_{11} + \frac{S_{12}S_{21}\Gamma_2}{1 - S_{22}\Gamma_2} = \frac{1}{\Gamma_1} \tag{9.29}$$

$$S_{22} + \frac{S_{12}S_{21}\Gamma_1}{1 - S_{11}\Gamma_1} = \frac{1}{\Gamma_2} \tag{9.30}$$

The oscillations are considered stable if any perturbation in the RF voltage or the RF current of the oscillator circuit at any instant decays itself, bringing the oscillator back to its point of equilibrium. The oscillator stability is analyzed [16] using a quasi-static approach by applying a small perturbation to the amplitude I_0 (Fig. 9.5). The impedance Z_T defined in (9.8), a function of I_0 and the complex frequency p, is developed in a Taylor series about I_0, ω_0. Since the perturbed current is nonzero, that is, the oscillations continue to exist after the perturbation, we should have in the plane PP'

$$Z_T(I_0, \omega_0) + \frac{\partial Z_T}{\partial p}\delta p + \frac{\partial Z_T}{\partial I_0}\delta I_0 = 0 \tag{9.31}$$

and since $Z_T(I_0, \omega_0) = 0$, we get

$$\delta p = -j \frac{(\partial Z_T/\partial I_0)(\partial Z_T^*/\partial \omega)}{|(\partial Z_T/\partial \omega)|^2} \delta I_0 \tag{9.32}$$

where δp can be decomposed into its real and imaginary parts

$$\delta p = \alpha + j\delta\omega$$

The oscillator will be stable if, for a positive variation of the current amplitude, the real part of the variation of the complex frequency is negative, that is, if α is negative, indicating a decreasing wave, returning to its point of equilibrium I_0.

From the previous expression of δp this condition is realized for

$$\frac{\partial R_T}{\partial I_0}\frac{\partial X_T}{\partial \omega} - \frac{\partial X_T}{\partial I_0}\frac{\partial R_T}{\partial \omega} > 0 \tag{9.33}$$

This relation represents the stability condition of an oscillator around an amplitude I_0 and angular frequency ω_0. Moreover, from (9.32), the imaginary part $\delta\omega$ vanishes if the condition

$$\frac{\partial R_T}{\partial I_0}\frac{\partial R_T}{\partial \omega} - \frac{\partial X_T}{\partial I_0}\frac{\partial X_T}{\partial \omega} > 0 \tag{9.34}$$

is fulfilled. This indicates that a variation of amplitude δI_0 will not result in a variation of the oscillator's real angular frequency ω_0. From (9.34) it can also be deduced that, for maximum stability, the device impedance $Z_T(I_0)$ and load impedance $Z_L(\omega)$ should intersect at right angles at the oscillation equilibrium point (I_0, ω_0) [17].

9.6 TRANSISTOR OSCILLATOR TYPES AND CONFIGURATIONS

Microwave oscillators can be classified in many ways. One of the common ways is to distinguish between oscillators based on resonator types (see Chapter 3). Examples are dielectric resonator oscillators (DROs), transmission-line resonator oscillators, YIG-tuned oscillators (YTOs), lumped-element oscillators, VCOs, and SAW oscillators.

From a design perspective oscillations can be divided into two main categories:

(a) Oscillators for fixed-frequency sources are used as local oscillators in communication, navigation, and radar systems as well as clock sources in fiber-optic communication systems. These oscillators are fixed fre-

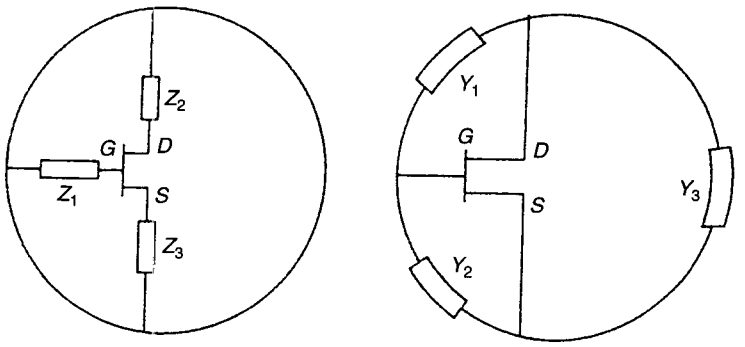


Figure 9.8 General transistor oscillator configurations: (*a*) series type; (*b*) parallel type.

quency or are capable of tuning over a narrow band, enough to keep locked to a reference source over operating conditions of temperature, bias, load, and time. Examples of these oscillators include DROs, SAW oscillators, lumped-element oscillators, and transmission-line oscillators.

(b) Oscillators for frequency-tunable sources are used in instrumentation (microwave sweepers, synthesizers, network and spectrum analyzers), electronic countermeasure (ECM) systems, and some wireless communication systems. Examples of these oscillators include YTOs and wideband VCOs.

A transistor microwave oscillator in its most general form can be represented as a series or a parallel multiport circuit, as shown in Fig. 9.8. An oscillator can be designed so that any of the immitances shown can contain the resistive output load. The other two immitances are generally reactive. In a fixed-frequency oscillator, all immitances are fixed while in tunable oscillators one or more immitances are made tunable by using varactor diodes or YIG resonators.

9.7 FIXED-FREQUENCY OSCILLATORS

Fixed-frequency oscillators are used for a variety of applications, including radar, communication transmitters, and receiver local oscillators for communications and navigation equipment. These oscillators generally require low frequency drift with temperature and low phase noise. To realize such an oscillator, an active device capable of generating negative resistance in the desired frequency range and a resonator at that frequency are required.

A number of different types of resonators can be used for fixed-frequency applications: low-Q transmission-line resonators (e.g., microstrip-line and coaxial-line resonators) and high-Q cavity resonators (e.g., metallic cavity and dielectric resonators). Microstrip-line or planar resonators offer a small

and compact size while oscillators based on high quality factor offer superior phase noise and frequency stability characteristics. Metal cavity resonators offer the highest quality factor, but their cost and size restrict their applications.

Dielectric resonators (DRs) are commonly used to realize-low noise, temperature-stable, fixed-frequency oscillators. Dielectric resonators are high-Q ceramic resonators conveniently sized for MIC applications (see Chapter 3). These devices not only are versatile and adaptable to various microwave structures and configurations but also fill the gap between waveguide and microstrip-line technologies. Dielectric resonators offer integration capability with MICs as well as quality factor and frequency stability approaching cavity resonators. Commonly used in a cylindrical shape, dielectric resonators sustain the same EM field modes of a metal cavity resonator. Due to the dielectric walls of DRs compared to metal walls of a metal cavity, the two behave as a dual of each other (E and H field interchanged). However, the finite dielectric constant of the material (typically between 20 and 80) causes the EM fields to extend beyond the bulk of the device. Normally, DRs are placed in a metal shield to prevent external interactions. A DR can be used to realize one or more of the immitances shown in Fig. 9.8.

Transistor DROs can be subdivided into two types: one using a DR as a series feedback element (Fig. 9.8*a*), and the other using a DR as a parallel feedback element (Fig. 9.8*b*). The analysis and design of both types [18] are discussed below.

9.7.1 Resonator as Series Feedback Element

Figure 9.9 illustrates various configurations using the dielectric resonator as a series feedback element. Figures 9.9*a–c* use the resonator at one terminal pair, while configuration 9.9*d* uses the DR as a series feedback element at two terminal pairs of the transistors. As an example, we now discuss the step-by-step procedure for the realization of the configuration shown in Fig. 9.9*a* using the three-port S parameters of a typical 0.5-μm GaAs FET. Comments on the design of other configurations are made where necessary.

As a first step in the design procedure, the impedance Z_3 in Fig. 9.10 is determined. With the impedance Z_3 connected to the source terminal, the reflection coefficient Γ_3 becomes

$$\Gamma_3 = \frac{Z_3 - Z_0}{Z_3 + Z_0} \tag{9.35}$$

Substituting into (9.3), we obtain

$$b_1 = S_{11}a_1 + S_{12}a_2 + S_{13}b_3\Gamma_3 \tag{9.36}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 + S_{23}b_3\Gamma_3 \tag{9.37}$$

$$b_3 = S_{31}a_1 + S_{32}a_2 + S_{33}b_3\Gamma_3 \tag{9.38}$$

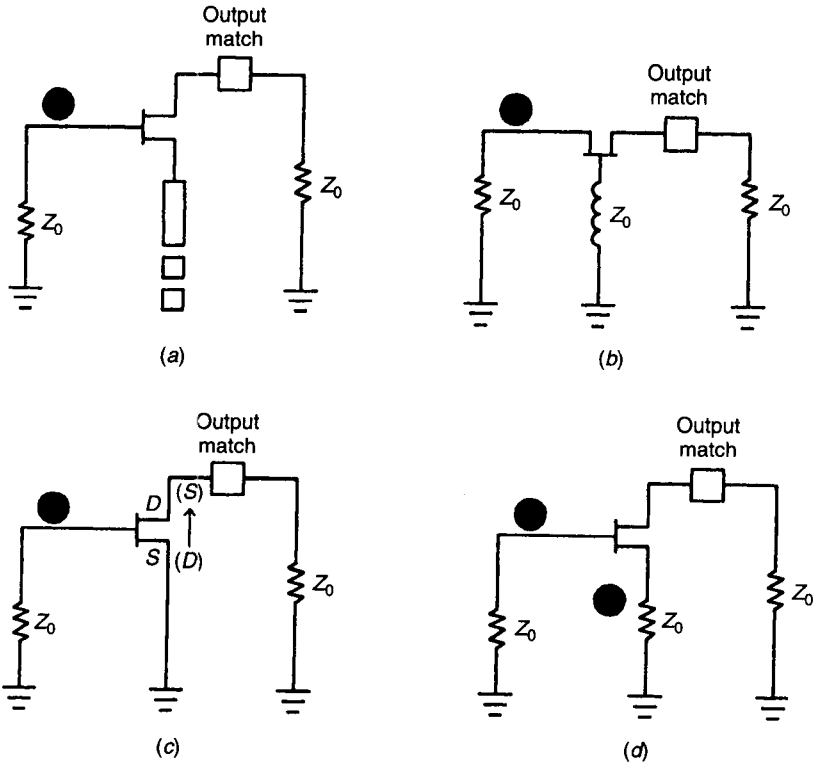


Figure 9.9 Different configurations for series feedback transistor DROs.

Eliminating b_3 from (9.36)–(9.38), the reduced two-port S matrix is given by

$$[S^T] = \begin{bmatrix} S_{11} + \dfrac{S_{31}S_{13}\Gamma_3}{1 - S_{33}\Gamma_3} & S_{12} + \dfrac{S_{13}S_{32}\Gamma_3}{1 - S_{33}\Gamma_3} \\ S_{21} + \dfrac{S_{31}S_{23}\Gamma_3}{1 - S_{33}\Gamma_3} & S_{22} + \dfrac{S_{23}S_{32}\Gamma_3}{1 - S_{33}\Gamma_3} \end{bmatrix} \tag{9.39}$$

The aim is to determine the value of the series feedback impedance Z_3 that will

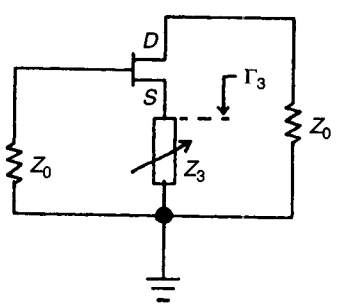
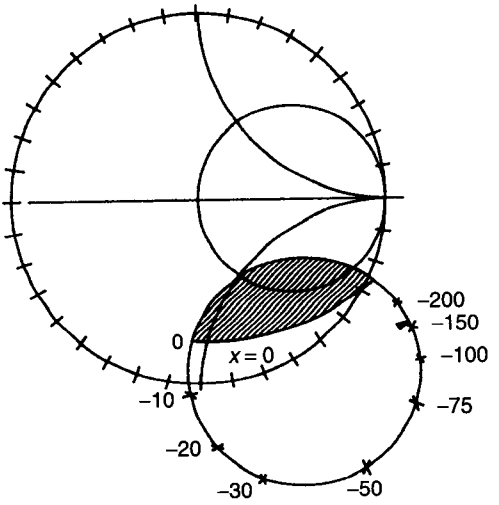
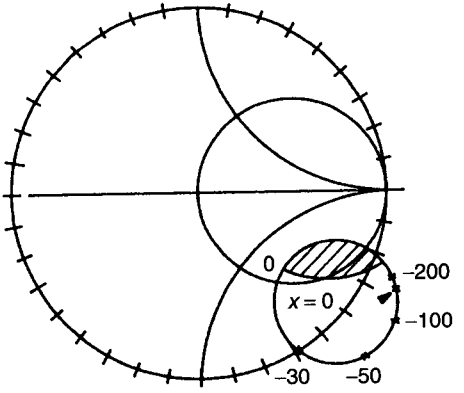


Figure 9.10 Determination of Z_3.



(a)



(b)

Figure 9.11 Mapping of $|\Gamma_3| = 1$ in (*a*) S_{11}^T and (*b*) S_{22}^T.

result in the modulus of S_{11}^T and/or S_{22}^T being greater than unity and, hence, create instability in the transistor. For the example considered in Fig. 9.9*a*, the open-circuited microstrip line represents a purely reactive impedance. The $|\Gamma_3| = 1$ plane determines a circle when mapped in the input and output reflection coefficient planes using well-known techniques [19]. Figure 9.11 shows the $|\Gamma_3| = 1$ plane mapped into the S_{11}^T and S_{22}^T planes for the half-micrometer GaAs FET at 10 GHz. The shaded area represents inductive impedance and the unshaded area represents the capacitive impedance in the source.

From Fig. 9.11 it may be noted that a negative reactance greater than $-j30$ Ω can be used to make both S_{11}^T and S_{22}^T greater than 1. A value of $-j159$ Ω is

selected that can be realized by an open-circuited stub $0.048\lambda_g$ long or by a 0.1-pF capacitor. The reduced two-port S matrix is now given by

$$[S^T] = \begin{bmatrix} 1.34/\underline{-33.5^\circ} & 0.49/\underline{70.4^\circ} \\ 0.49/\underline{156^\circ} & 1.16/\underline{-31.9^\circ} \end{bmatrix} \tag{9.40}$$

A dielectric resonator coupled to a microstrip line can also be used as impedance Z_3 in Fig. 9.10, as shown in Fig. 9.9*d*. The reflection coefficient Γ_3 in this case is a function of the coupling coefficient β_3 and the distance θ_3 between the transistor plane and the resonator plane as analyzed in Chapter 3. Using the matrix coefficients from (9.39), the Γ_3 plane in terms of β_3 and θ_3 can be mapped into any S parameter of the reduced two-port. Figure 9.12 shows, for example, mapping of the DR reflection coefficient plane into all four S parameters of an X-band FET at 8 GHz [20].

The nonconcentric circles shown in Fig. 9.12 are the constant coupling coefficient β_3 (proportional to the $|\Gamma_3|$) circles, while the radial arcs are the constant transmission electrical line length θ_3 (proportional to arg Γ_3) arcs. The relation between Γ_3, β_3, θ_3, and frequency is given by

$$\Gamma_3 = -\frac{\beta_3}{\sqrt{(1+\beta_3)^2 + \Delta^2}} \exp\left[-2j\left(\theta_3 + \tan^{-1}\frac{\Delta}{1+\beta_3}\right)\right] \tag{9.41}$$

with $\Delta = 2Q_u(f - f_0)/f_0$. Figure 9.12 can be used to determine the DR position in order to create the desired instability in the transistor.

Continuing the TDRO design example of Fig. 9.9*a*, we have already determined the impedance Z_3; the resulting two-port S parameters are given in (9.40). In the second step of the procedure, we will determine the value of reflection coefficient Γ_1 that maximizes the reflection coefficient Γ_d at the drain port (Fig. 9.13) using the relation

$$\Gamma_d = S_{22}^T + \frac{S_{12}^T S_{21}^T \Gamma_1}{1 - S_{11}^T \Gamma_1} \tag{9.42}$$

The reflection coefficient Γ_1 in this case is realized by a DR coupled to a microstrip line and is characterized by the coupling coefficient β_1 and the distance θ_1 between the transistor and the resonator plane.

In the present example, $|S_{11}^T|$ and $|S_{22}^T|$ being greater than unity, the mapping technique used in the first step cannot easily be used to determine the required Γ_1 and hence the position of the DR. Instead, we use the constant reflection coefficient circles approach, in which case the locus of constant reflection coefficient magnitude $|\Gamma_d|$ is drawn on the reflection coefficient plane Γ_1. From (9.42) the radius R and center Ω of the constant reflection coefficient circles can be determined to be

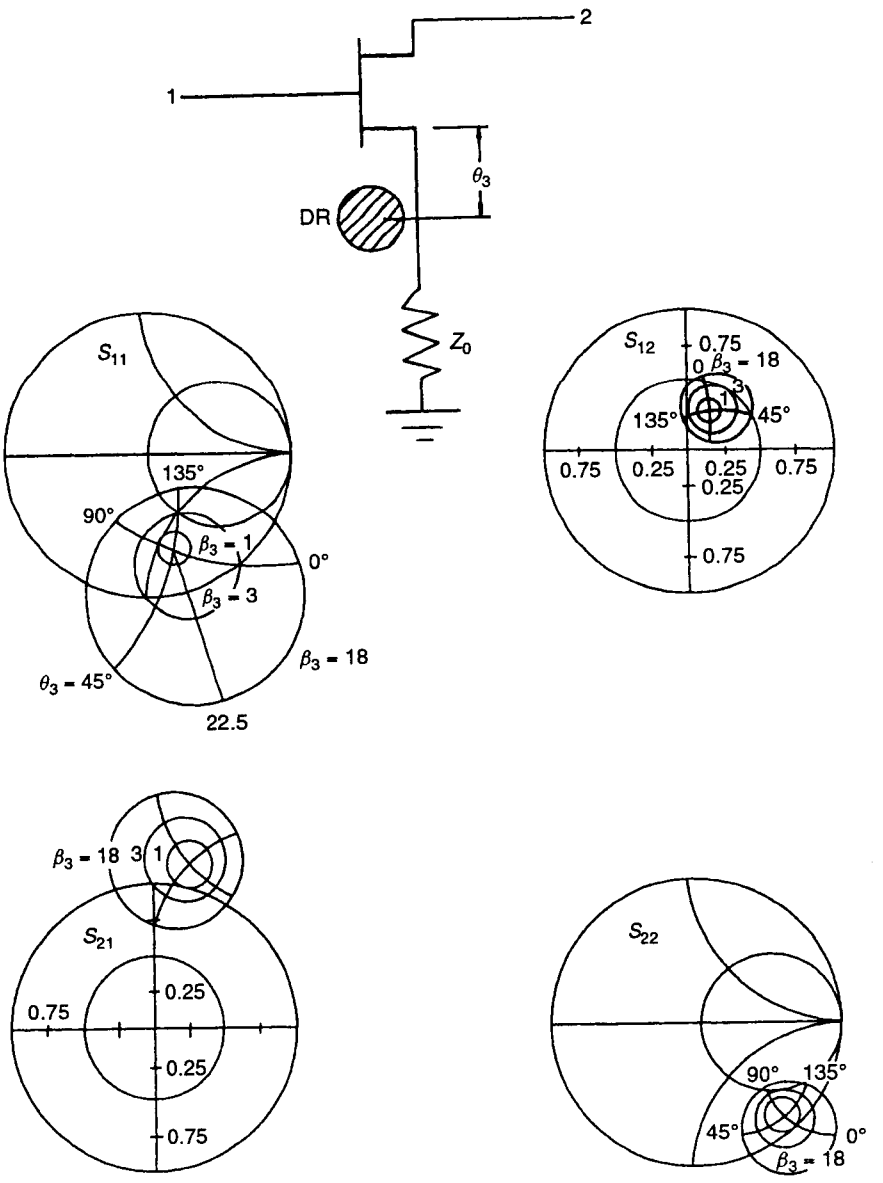


Figure 9.12 DR as a series feedback element in source of the FET oscillator.

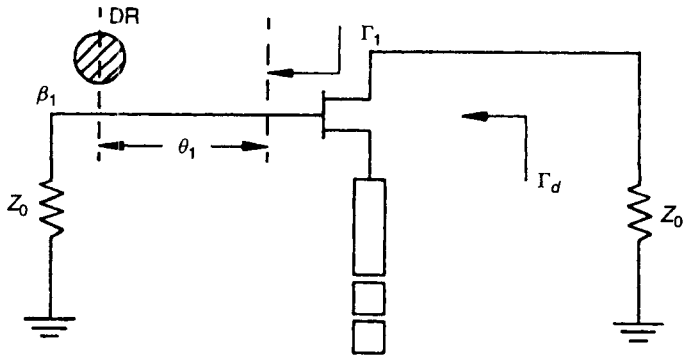


Figure 9.13 Determination of DR position in the gate.

$$R = \frac{|\Gamma_d S_{12}^T S_{21}^T|}{|\Gamma_d|^2 |S_{11}^T|^2 - |\Delta|^2} \tag{9.43}$$

$$\Omega = \frac{S_{11}^{T^*}[|\Gamma_d|^2 - |S_{22}^T|]^2 + S_{22}^T S_{12}^{T^*} S_{21}^{T^*}}{|\Gamma_d|^2 |S_{11}^T|^2 - |\Delta|^2} \tag{9.44}$$

where

$$\Delta = S_{11}^T S_{22}^T - S_{12}^T S_{21}^T \tag{9.45}$$

Figure 9.14 shows various $|\Gamma_d|$ = constant circles on the $\Gamma_1(\beta_1, \theta_1)$ plane for the example under consideration. The values of β_1 and θ_1 can now be determined for a high value of $|\Gamma_d|$ ($\gg 1$).

The first step and the second step used the small-signal S parameters to determine the impedances Z_1 and Z_3 to be connected at gate and source ports, respectively, in order to achieve a high value of reflection coefficient at the

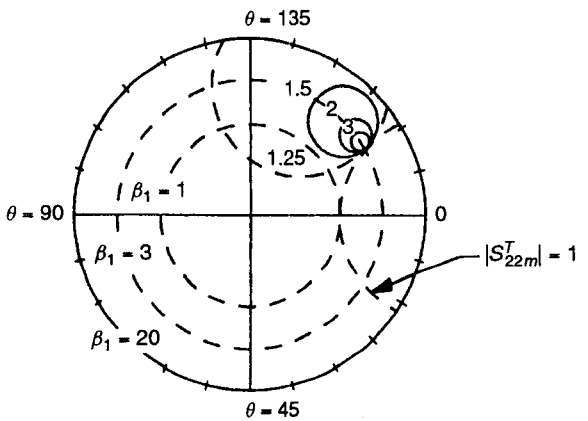


Figure 9.14 Constant $|S_{22m}^T|$ circles in DR reflection coefficient (β_1, θ_1) plane.

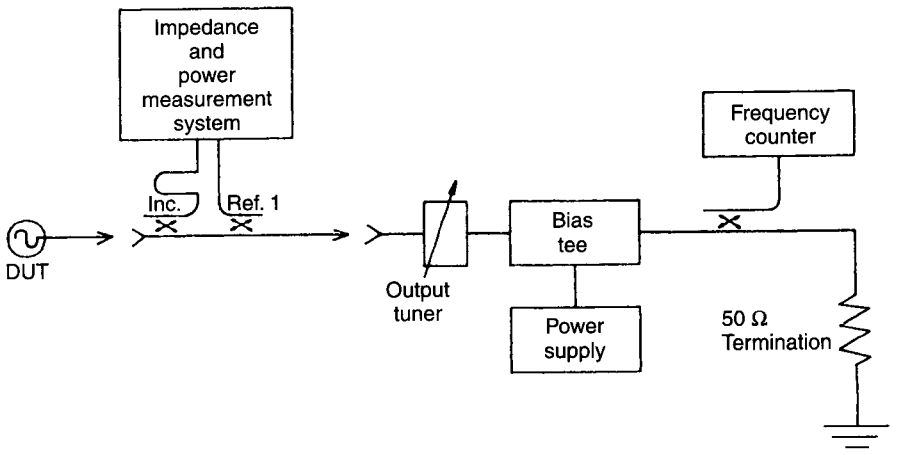


Figure 9.15 Load-pull setup for one-port oscillator.

drain port. A number of approaches exist to determine the load circuit impedance Z_2 in Fig. 9.8*a*. The two common ones, which assume the large-signal operation, are the device-line approach and the load-pull approach.

The device-line approach [21] is based on the measurement of the inverse reflection coefficient of the one-port and the return added power for different input power levels. The load impedance for maximum oscillator output power can be determined from the device line thus obtained. This approach, however, has the limitation that requires the source resistance to be greater than the modulus of the device resistance. Otherwise, oscillation takes place and the device line cannot be measured.

The test setup for the measurement of the load-pull approach [22] is shown in Fig. 9.15. The oscillator acts as the RF power source for the system. The drain port of the transistor circuit (with optimized Z_1 and Z_3 connected) is attached to the load-pull measurement system at the input port of the reflectometer through a 50-Ω line and powered up. The impedance displayed on the polar display will be the impedance presented to the output of the oscillator device. Using the output tuner, contours of constant output power can then be drawn on a Smith chart using the power readings from the output power meter and an x–y recorder connected to the polar display. Typical load-pull data are shown in Fig. 9.16. This load impedance chart can be used to design the output circuit for the transistor.

9.7.2 Resonator as Parallel Feedback Element

An alternate way of realizing a stable oscillator is by using the DR simultaneously coupled to two microstrip lines as a parallel feedback element for a transistor. In this case, the transistor can be used as a two- or three-port device, as shown in Figs. 9.17 and 9.18, respectively.

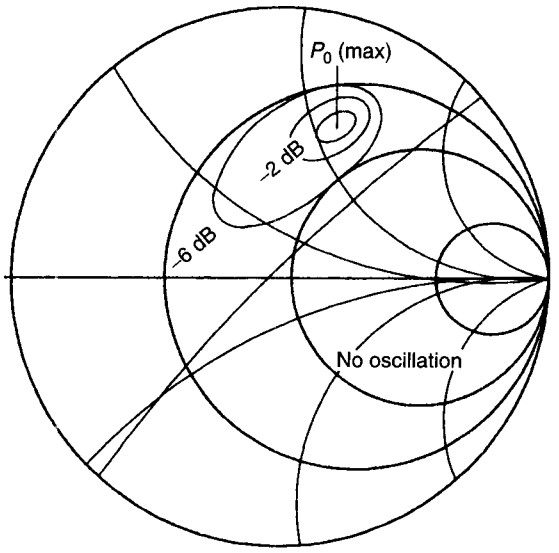


Figure 9.16 Typical load-pull data. (After Poulin [22]. Reprinted with permission from Penton Media Inc.)

In Fig. 9.17, the transistor is treated as a two-port [23]. In this case, the input and output matching circuits for a common-source transistor are designed for a maximum transducer gain amplifier around the oscillator frequency f_0. Highly selective positive feedback between the input and the output can be used to create stable oscillations. This is achieved by feeding back a part of the output signal into the input through the dielectric resonator transmission filter. The lengths l_1 and l_2 are adjusted to achieve the phase shift around the loop, con-

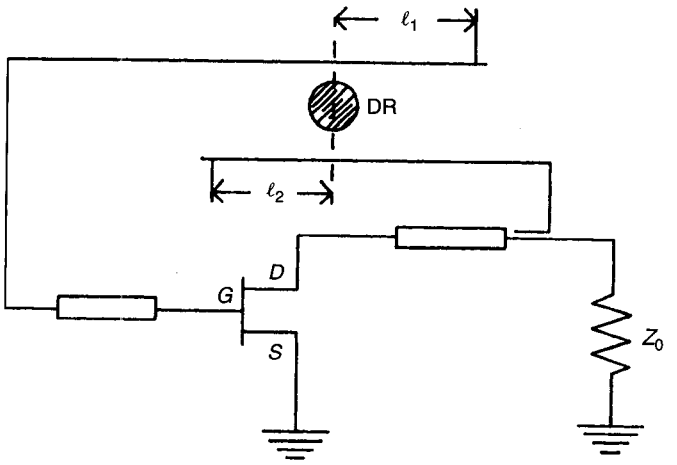


Figure 9.17 Parallel feedback transistor DRO using transistor as a two-port.

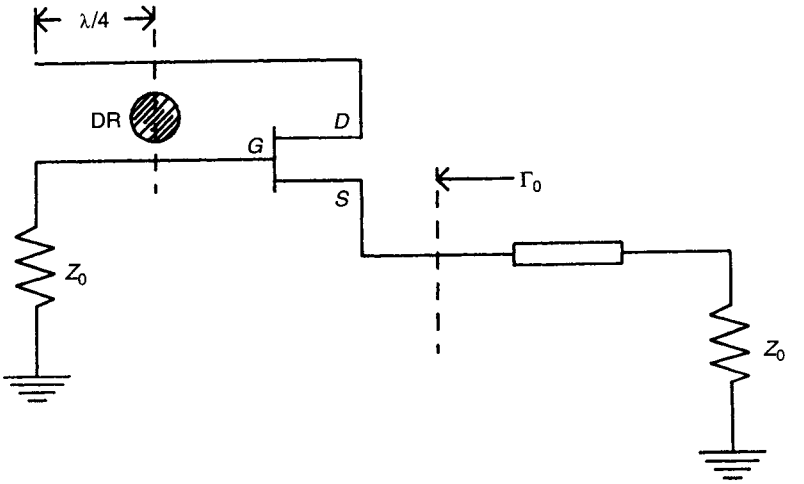


Figure 9.18 Parallel feedback transistor DRO using transistor as a three-port.

sisting of the amplifier and feedback circuit, equal to an integer multiple of 2π radians at f_0, that is,

$$\phi_A + \phi_R + \phi_C = 2\pi k \qquad k = 1, 2, 3 \tag{9.46}$$

where ϕ_A, ϕ_R, and ϕ_C are respective insertion phases of the amplifier, resonator, and remaining part of the feedback circuit at f_0. The other condition for the oscillations to take place is that the open-loop small-signal gain must exceed unity at f_0, that is,

$$G_A - L_R - L_C > 0 \text{ dB} \tag{9.47}$$

where G_A, L_R, and L_C are the amplifier gain, resonator filter loss, and loss in the other feedback components in decibels, respectively. The necessary relations for determining the resonator insertion phase ϕ_R and insertion loss L_R can be determined from Chapter 3:

$$\phi_R = \tan^{-1}\left(\frac{-j2Q_u\delta}{1+\beta_1+\beta_2}\right) \tag{9.48}$$

$$L_R = 10 \log \frac{(1+\beta_1+\beta_2)^2}{4\beta_1\beta_2} \tag{9.49}$$

The extent of the inequality given in (9.47) and, hence, the amplifier gain compression in the steady state oscillation condition affect the output power as well as the FM noise performance of the oscillator. Excessive gain compression can adversely affect the oscillator noise due to increased amplifier noise figure and AM-to-PM conversion.

Figure 9.18 shows another configuration of a parallel feedback DRO [11, 24]. In this case, the DR transmission filter is coupled between the two terminals of the transistor and the output is taken from the third. This configuration can be analyzed as a two-port containing the DR as a parallel feedback network to a three-port device, the transistor [11].

Feedback parameters (i.e., the position and coupling of the dielectric resonator to the microstrip lines) can be determined to maximize the reflection gain (>1) at the output terminal. The output matching circuit can now be designed using the load-pull approach described above.

9.7.3 Series versus Parallel Feedback

The series and parallel feedback approaches described above are both commonly used for designing TDROs. The series feedback has the advantage of ease of DR alignment due to its coupling to the microstrip on one side compared to the parallel feedback type, wherein the resonator is simultaneously coupled to both sides. This aspect also makes the series feedback configuration circuit operable over a much wider bandwidth compared to parallel feedback configuration. In the parallel feedback configuration use of a high-gain amplifier can allow significant decoupling of the resonator from the microstrip lines resulting in a higher loaded quality factor value of the DR and hence a lower phase noise oscillator. The temperature stability of the DR material, required for temperature compensation of the oscillator, is different in both cases due to the difference in resonator coupling configuration.

9.7.4 Maximum Oscillator Power Output

For a given transistor used as an amplifier the output power can be approximated as [25]

$$P_{\text{out}} = P_{\text{sat}}\left[1 - \exp\left(-\frac{G_0 P_{\text{in}}}{P_{\text{sat}}}\right)\right] \tag{9.50}$$

where P_{sat} is the saturated output power, G_0 is the small-signal gain, and P_{in} is the input power. For an oscillator, the maximum output power occurs at the point where $P_{\text{out}} - P_{\text{in}}$ is maximum or when $\partial P_{\text{out}}/\partial P_{\text{in}} = 1$. Thus from (9.50), the optimum input power can be calculated as

$$(P_{\text{in}})_{\text{opt}} = \frac{P_{\text{sat}} \log G_0}{G_0}$$

and then

$$(P_{\text{osc}})_{\text{max}} = P_{\text{sat}}\left(1 - \frac{1}{G_0} - \frac{\log G_0}{G_0}\right) \tag{9.51}$$

9.7.5 Temperature Stability of DROs

Microwave sources with high-temperature stabilities have been realized in the past using Invar cavities or by phase locking the source to a frequency-multiplied ovenized very high frequency (VHF) crystal oscillator. These approaches are cumbersome as well as expensive. With the advent of high-performance DRs, TDROs now offer a miniature, elegant, and inexpensive way of achieving high stabilities.

A free-running transistor oscillator is known to have a negative temperature coefficient. A DR having a positive temperature coefficient is thus required to compensate the frequency drift with temperature. Currently, the temperature coefficient of the DR can be controlled by varying its composition. A DR with coefficients equal to anywhere between +9 and −9 ppm/°C are commercially available, as discussed in Chapter 3.

The temperature stability of a TDRO can be analytically determined in terms of the coupling coefficient β, the Q factor, and the rate of change of transistor reflection phase with temperature [26].

In Fig. 9.19, the active device including the circuitry can be presented at its gate port as a temperature- and power-dependent reflection coefficient:

$$\Gamma_G(T,P) = |\Gamma_G(T,P)|e^{j\phi_G(T,P)} \tag{9.52}$$

with

$$|\Gamma_G(T,P)| > 1 \tag{9.53}$$

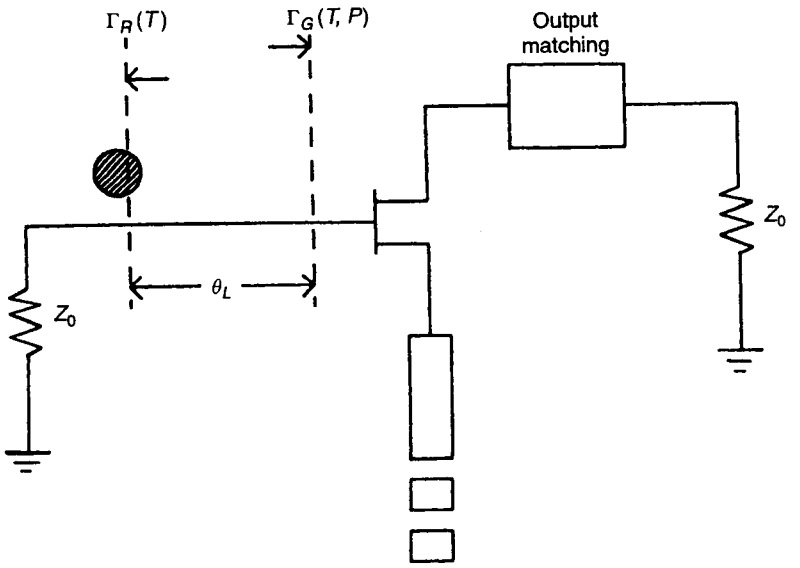


Figure 9.19 Circuit used for temperature variation analysis.

The stabilization circuit containing a 50-Ω microstrip line, loaded by a 50-Ω damping resistor and coupled with the dielectric resonators, can be represented at the resonator plane as a temperature-dependent reflection factor:

$$\Gamma_R(T) = \frac{\beta}{1+\beta+j2Q_u\delta(T)} \tag{9.54}$$

where

$$\delta(T) = \frac{f - f_0(T)}{f_0(T)} \tag{9.55}$$

β being the coupling coefficient between the resonator and microstrip line. Applying the oscillation condition in the plane of the gate port, we obtain

$$|\Gamma_R(T)|e^{-j2\theta_L} = \frac{1}{|\Gamma_G(T,P)|}e^{-j\phi_G(T,P)} \tag{9.56}$$

where θ_L is the inserted microstrip electrical transmission-line length, in order to adjust the phase for oscillation condition. This equation, split into real and imaginary parts, leads to the following relationships for oscillation frequency and power:

$$(\beta+1)\tan[\phi_G(T,P) - 2\theta_L] = \frac{f}{f_0(T)} - 1 \tag{9.57}$$

$$|\Gamma_G(T,P)| = \frac{1+\beta}{\beta} \tag{9.58}$$

where $\Gamma_G(T,P)$ is known to be a monotonous decreasing function of temperature and power. A decrease of Γ_G at higher temperatures will be compensated automatically by a degradation of output power P, thus keeping $|\Gamma_G|$ constant.

For determining the frequency stability with temperature, (9.57) can be differentiated with respect to temperature and leads after certain approximations to

$$\frac{df}{f\,dT} \simeq \frac{df_0}{f_0\,dT} + \frac{\beta+1}{2Q_u}\frac{\partial\phi_G}{\partial T} \tag{9.59}$$

where $\partial\phi_G/\partial T$ represents the active circuit phase drift with temperature.

The preceding relation can be used to help design temperature-stable DR oscillators. Phase drift $\partial\phi_G/\partial T$ is known to decrease linearly with temperature, thus necessitating a positive temperature coefficient of the resonator $\partial f_0/f_0\,dT$ varying linearly in the desired temperature range. Equation (9.59) can be written as

$$\tau_f = \tau_{f0} + \frac{1}{2Q_L}\frac{\partial \phi_G}{\partial T} \tag{9.60}$$

where τ_f, τ_{f0} are the temperature coefficients of the oscillator and resonator, respectively, and Q_L is the loaded Q of the resonator. For temperature-stable operation, one requires $\tau_f = 0$, which leads to the requirement

$$\tau_{f0} = -\frac{1}{2Q_L}\frac{\partial \phi_G}{\partial T} \tag{9.61}$$

This relation tells us how to make the oscillation frequency independent of temperature. For a given device, the value of $\partial \phi_G / \partial T$ is fixed, but Q_L may be selected by adjusting the coupling coefficient between the DR and the microstrip line. Different combinations of Q_L and τ_{f0} may be tried for optimizing other properties of the DRO, like tunability or power output. Using this approach and a suitable dielectric resonator material, the temperature stabilities of 0.1 ppm/°C have been reported from −20 to +80°C [26].

A digital compensation technique to realize high-temperature-stability TDROs has been reported [27]. In this case, a temperature sensor is mounted in the oscillator to detect the temperature changes. These data are digitized and are fed to an extended programmable read-only memory (EPROM) preprogrammed with temperature characteristics of the oscillator. The lookup table in the ROM provides the necessary digital temperature correction word that is converted to an analog signal (with proper synchronization between A/D and D/A) and applied to the DRO for frequency correction. The correction signal can be applied to the varactor in the case of a varactor-tuned DRO [27] or to control the phase shift in the feedback loop in case of a parallel feedback DRO. Using this digital compensation technique, frequency drift of better than ±20 ppm has been achieved from −55 to +85°C.

9.7.6 Tunable DROs

A TDRO is basically a fixed-frequency oscillator with its frequency determined by the resonator material permittivity, resonator dimensions, and shielding conditions, as discussed in Chapter 3. The oscillation frequency can, however, be tuned over a narrow frequency range using different approaches depending on the requirements. The frequency tuning of a TDRO can be accomplished mechanically or electrically, as discussed below.

Figure 9.20 shows the configuration used for mechanical frequency tuning of the TDRO. Use is made of the fact that the resonant frequency of the DR is highly sensitive to the shielding, that is, to the proximity of the ground plane. A tuning screw is inserted from the top cover of the package, right above the DR. The increase in the tuning screw depth d increases the resonant frequency of the DR in the commonly used mode $TE_{01\delta}$. Care should be taken to keep the distance h between the resonator and the tuning screw at least 0.5 times the reso-

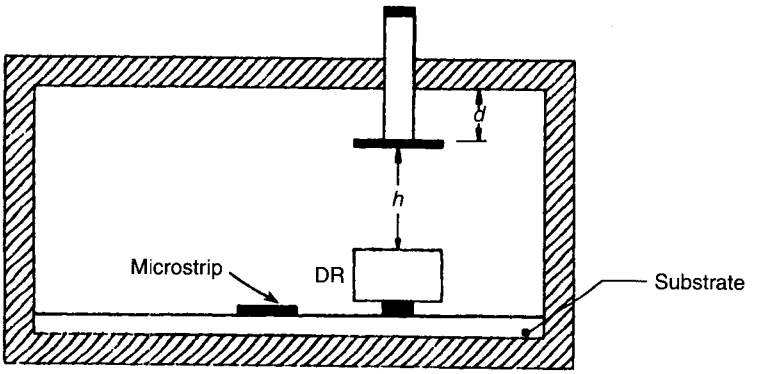


Figure 9.20 Mechanical tuning of a DR.

nator height in order to not degrade the DR quality factor. A mechanical frequency tuning range of ±2% can be obtained in a TDRO without noticeably affecting the FM noise, output power, and frequency stability over temperature.

The frequency of the TDRO can be tuned electrically by using a number of different approaches, such as varactor tuning, ferrite tuning, bias tuning, and optical tuning [28, 29]. Electrical tuning can be affected over a very small bandwidth without significantly affecting the oscillator performance. This fast tuning can be used for different applications, such as digital temperature compensation, low-deviation frequency-modulated sources, and phase locking. A brief description of the various approaches to electrically tune the TDRO follows.

In the varactor-tuned TDRO, a varactor in association with a microstrip line is made to resonate around the DR frequency. This resonant circuit is electromagnetically coupled to the dielectric resonator, forming a pair of mutually coupled resonant circuits. By varying the varactor capacitance with the bias voltage, the resonant frequency of the DR, coupled to varactor microstrip on one side and a 50-Ω microstrip line on the other, can now be tuned. Figure 9.21 shows a typical configuration for coupling the varactor and DR. Tighter coupling between the DR and varactor will result in a larger frequency control at the cost of lowered DR quality factor and, hence, the phase noise. Figure 9.22 shows the effect of varactor tuning on a low-noise TDRO phase noise for the varactor coupling adjusted for 0.1% frequency control at 11 GHz [29]. Using another configuration with two varactors on a quartz spacer placed directly above the resonator, a tuning bandwidth of 0.75% has been reported [30].

The DR can also be tuned by attaching a microwave ferrite on the resonator and applying a magnetic field to it. The magnetic field controls the magnetic properties of the ferrite and, hence, the field distributions in and around the DR, resulting in a shift in the resonant frequency. Tuning bandwidths of the order of 0.5% [31] and 1% [30] have been reported. This method, however,

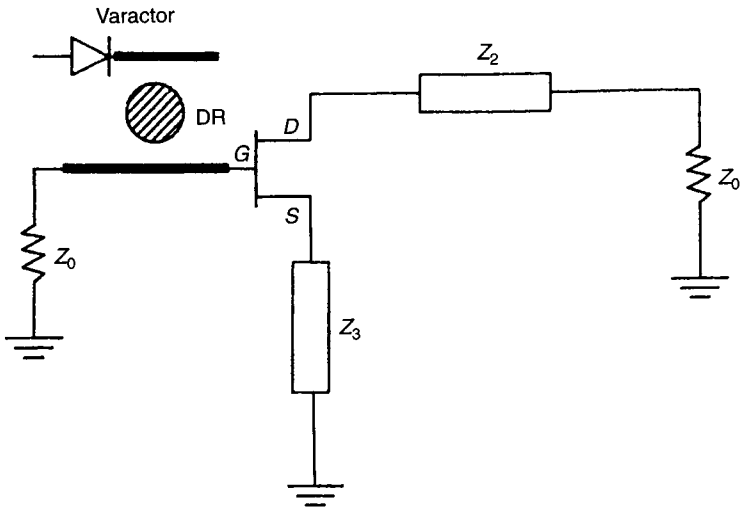


Figure 9.21 Varactor-tuned DRO.

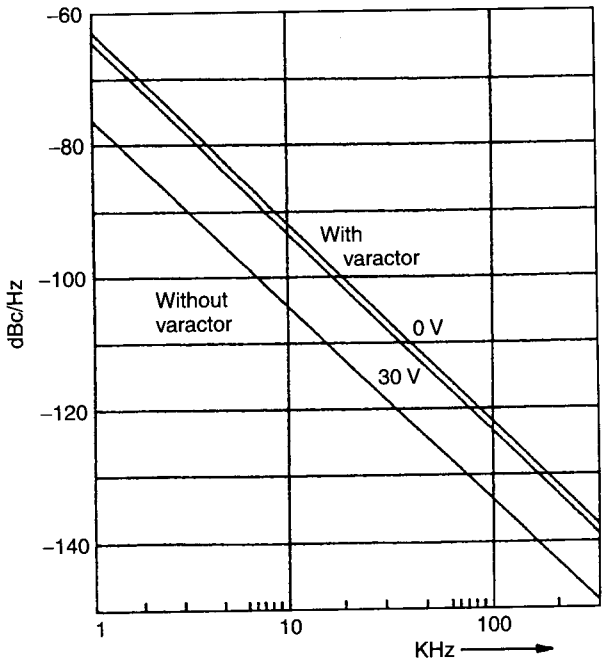


Figure 9.22 Effect of varactor tuning on DRO FM noise. (After Camiade et al. [29]. Reprinted with permission of Microwave Exhibitors & Publishers Ltd.)

cannot effectively be used in practice due to slow tuning speed, large size, and bulk of the electromagnet necessary to provide the variable magnetic field and excessive power consumption of the driving circuit.

Another method of electronic tuning is to use the bias voltage for this purpose. The frequency of any oscillator is known to be sensitive to the bias voltage described generally by the pushing figure. Unfortunately, the change in bias voltage also affects the output power, thus making it difficult to use for frequency modulation purposes. The bias circuit can, however, be designed in such a way as to minimize the output power variation with bias voltage. Using such a technique, a 15-MHz frequency-tuning range at 36 GHz with less than a 1-dB variation of oscillator power output has been reported [32]. A TDRO inherently has a high Q factor and, hence, a low pushing figure. This limits the bias-tuned frequency range that can only be increased by intentionally reducing the oscillator Q at the cost of degradation of other oscillator characteristics such as phase noise and temperature stability.

Optical control of microwave devices and subsystems is a rapidly growing field of research. The resonant frequency of the DR used in the TDRO can be optically modulated and tuned as shown in Fig. 9.23. A photosensitive material such as high-resistivity silicon is placed directly on the DR. Light from a laser or light-emitting diode (LED) is brought through an optical fiber to illuminate the photosensitive material, changing its conductivity and perturbing the EM field in and around the resonator. This perturbation results in a shift in the center frequency of the TDRO. Using this technique, tuning bandwidths of better than 0.1% have been achieved at the X band [33, 34].

9.7.7 Transmission-Line Resonator Oscillators

Planar transmission-line resonator oscillators represent an important configuration for oscillators when size, cost, and integration capabilities are important. Dielectric resonators offer excellent phase noise but call for accurate dimensions of the dielectric puck, controlled shielding, and precise placement on the substrate. Figure 9.24*a* represents a microstrip resonator voltage-controlled

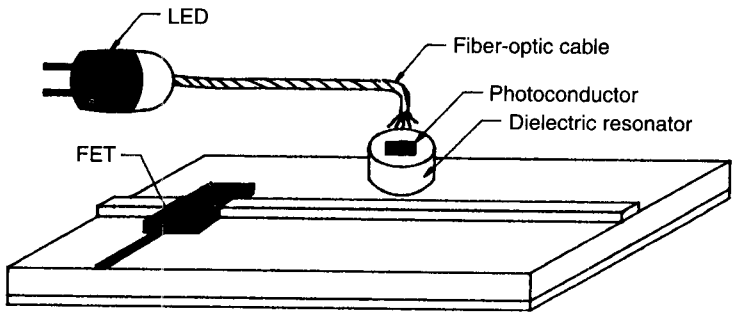
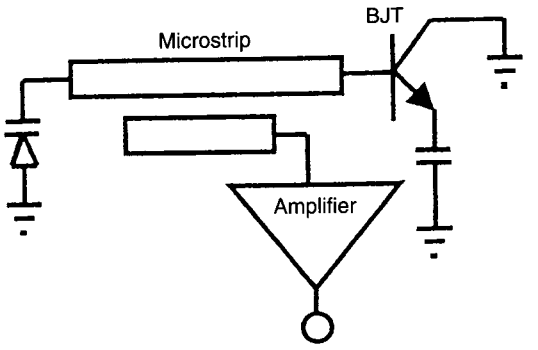
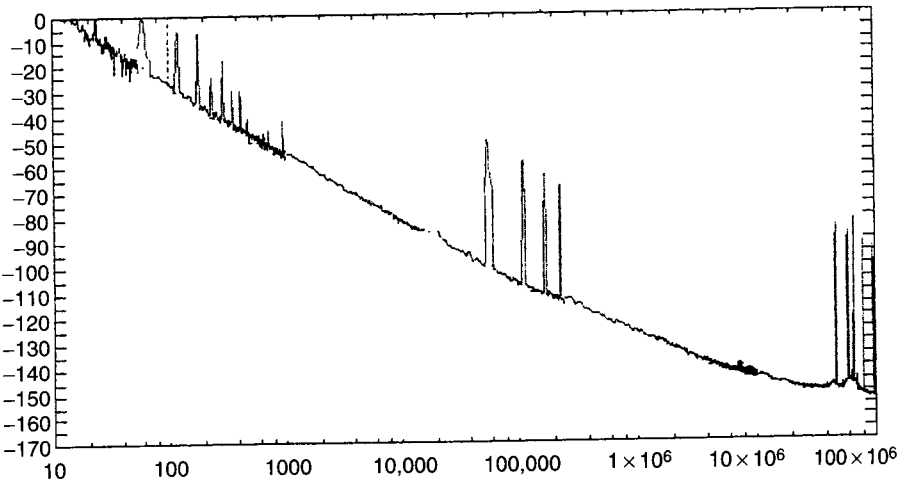


Figure 9.23 Optical tuning of DR. (After Hercsfeld [34]. Reprinted with permission of IEEE.)



(*a*)

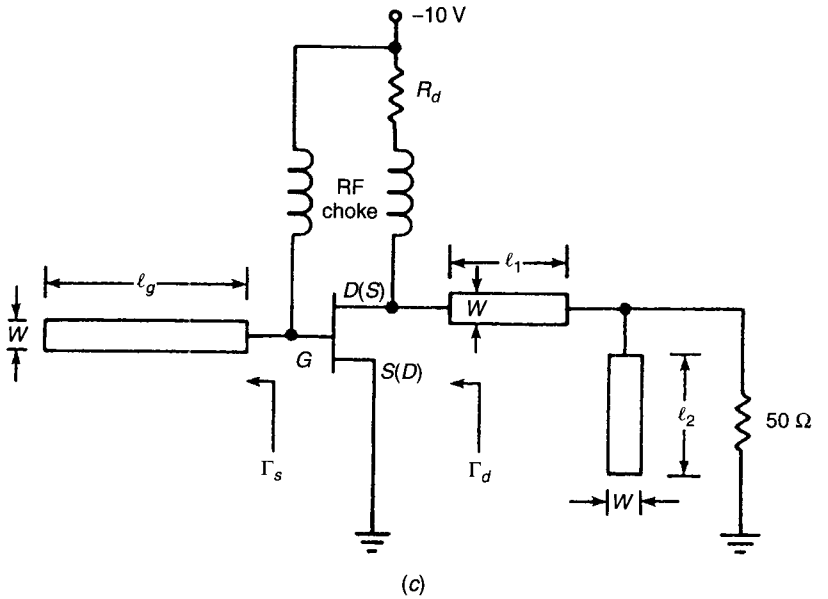


(*b*)

Figure 9.24 (*a*) Microstrip resonator oscillator. (*b*) Phase noise of 10.7-GHz transmission-line resonator. (*c*) Simple reverse-channel GaAs FET oscillator configuration using source as drain and vice versa.

oscillator tunable over a narrow band. A varactor diode can be designed to couple at either end of the resonator. The length of the resonator line required to meet the oscillation condition can be determined using an approach similar to the ones described in the previous section.

Increasing the resonator length in terms of half wavelengths in addition to the transmission-line length required to satisfy the oscillation conditions can increase the Q_L of the resonator. However, the length of the resonator is limited due to resulting increased losses. Use of a long transmission line also creates resonant conditions at multiple frequencies and requires precautions to allow the oscillation condition to be met at only the desired frequency. Micro-



(c)

Figure 9.24 *(Continued)*

strip resonator dimensions can be optimized for high Q_L (low noise) as well as for meeting the oscillation conditions. A varactor diode coupled to the resonator provides narrow-band tuning, enabling phase locking of the oscillator to the desired frequency, similar to a tunable TDRO. Low-noise VCOs have been reported [35] using low-noise silicon bipolar devices. Figure 9.24*b* shows that a phase noise of −112 dBc at 100 kHz can be achieved at 10 GHz.

A simple oscillator circuit using a packaged GaAs FET and a microstrip transmission-line resonator is described step by step in the next example.

Example. The small-signal two-port S parameters of a reverse-channel 0.5-μm GaAs FET at 9 GHz are given as $S_{11} = 0.87\angle{-155°}$, $S_{12} = 0.53\angle{-78°}$, $S_{21} = 0.62\angle{-143°}$, and $S_{22} = 0.25\angle{-100°}$. (Bias: $V_{DS} = -8$ V, $V_{GD} = -2$ V, and $I_D = 80$ mA.) Design a microstrip oscillator using an open stub on the gate and a two-section (series–shunt) matching circuit on the drain to match it to a 50-Ω load. For maximum power output assume that the load impedance Z_L is related to the transistor small-signal output impedance Z_{out} by the relations $R_L = -\frac{1}{3}R_{\text{out}}$ and $X_L = -X_{\text{out}}$. The characteristic impedance of the matching transmission lines is 50 Ω. Draw the RF circuit for a 0.63-mm-thick soft substrate with $\epsilon_r = 2.5$. Assuming a −10-V power supply, show the DC bias network.

For symmetrical channel GaAs FETs, a potentially unstable, reverse-channel configuration is obtained by reversing the drain–source voltage supply, thus reversing the drain–source channel [36]. In this configuration the drain

acts like a source and vice versa. A commonly available common-source packaged FET thus acts like an unstable common-drain configuration. This makes oscillations possible over a wide frequency range without any external series or parallel feedback. As shown in Fig. 9.24*c*, gate and source are both negative with respect to drain, and this configuration needs a single negative bias supply.

Instability of the transistor in this example can be verified by calculating the stabilization factor K. Using the given S parameters, $K = 0.488$ at 9 GHz. A stability factor $K < 1$ means that the transistor can be made to oscillate by properly selecting source and load impedances without series feedback in the source terminal. The oscillator configuration is shown in Fig. 9.24*c*.

Looking into the drain of the transistor (without the matching elements l_1 and l_2),

$$\Gamma_d = S_{22} + \frac{S_{12}S_{21}}{1/\Gamma_S - S_{11}} \tag{9.62}$$

where Γ_S represents the reflection coefficient of the microstrip open-circuited stub with an electrical length of θ_g (physical length of l_g) at the gate terminal.

Assuming a lossless open-circuited stub, Γ_S may be expressed as

$$\Gamma_S = 1\underline{/\theta_S} = 1\underline{/-2\theta_g} = 1\underline{/-4\pi l_g/\lambda_g}$$

The value of θ_s corresponding to $\Gamma_d > 1$ can be found using different approaches, including load mapping and constant reflection coefficient circles, as discussed before. Another possible approach is to solve Eq. (9.62) for a value of θ_s (phase of Γ_s) so that vector $S_{12} \cdot S_{21}/(1/\Gamma_s - S_{11})$ is in phase with vector S_{22}. Solving for θ_s to meet this condition gives [10]

$$\theta_S = \theta_{22} - (\theta_{12} + \theta_{21}) + \sin^{-1}[|S_{11}| \sin \theta_x]$$

with

$$\theta_x = (\theta_{12} + \theta_{21}) - (\theta_{11} + \theta_{22})$$
$$S_{ij} = |S_{ij}|\underline{/\theta_{ij}}$$

In this example, $\theta_S = 152°$, $\theta_g = 104°$, and $l_g = 6.6$ mm ($\lambda_g = 23$ mm).

The corresponding maximum value of $\Gamma_{d,\max} = 2.57\underline{/-78°}$. The small-signal output impedance of the FET with the open stub on the gate can be determined now to be

$$Z_{\text{out}} = \frac{1 + \Gamma_{d,\max}}{1 - \Gamma_{d,\max}} = -42.3 - j38.3\ \Omega$$

The required load impedance can be calculated as

$$R_L = -\tfrac{1}{3}R_{\text{out}} = 14.1\ \Omega \qquad X_L = -X_{\text{out}} = 38.3\ \Omega$$

Here, $Z_L = 14.1 + j38.3\ \Omega$ corresponds to $\Gamma_L = 0.66\underline{/105.6°}$.

A two-section matching circuit shown in Fig. 9.24*c* can be designed using a Smith chart or the following well-known relations:

$$l_2 = \frac{\lambda_g}{2\pi}\tan^{-1}\left(2\sqrt{\frac{|\Gamma_L|^2}{1-|\Gamma_L|^2}}\right) = 3.87\ \text{mm}$$

$$l_1 = \left(\frac{\theta_y - \theta_L}{720°}\right)\lambda_g = 3.91\ \text{mm}$$

where

$$\theta_y = 180° + \tan^{-1}\sqrt{\frac{1-|\Gamma_L|^2}{|\Gamma_L|^2}} = 228.3° \qquad \Gamma_L = |\Gamma_L|\underline{/\theta_L}$$

The width W for all the microstrip lines shown corresponds to 50 Ω and is 1.8 mm.

The bias resistance R_d shown in Fig. 9.24*c* can be calculated to be

$$R_d = \frac{2\ \text{V}}{80\ \text{mA}} = 25\ \Omega$$

9.8 WIDE-BAND TUNABLE OSCILLATORS

Wide-band tunable oscillators are used for ECM instrumentation as well as for wide-band communication transmitter applications. Important characteristics of an electronically tunable oscillator include frequency tuning range (bandwidth), speed or ability to rapidly change the frequency of the oscillator, and tuning linearity (slope of frequency vs. voltage curve) expressed in megahertz per volt or megahertz per milliampere depending on whether the oscillator is voltage or current tuned. There are two basic types of wide-band tunable oscillators commonly used: VCOs, which are voltage tuned, and YTOs, which are current tuned. In this section we will review design aspects of both types of oscillators.

The analysis and design of wide-band tunable oscillators are more complex than that of a fixed-frequency oscillator. It is important that the active element—for example, the transistor—can generate the negative resistance or the reflection gain >1 over the desired bandwidth, with the help of proper series or parallel feedback. The frequency-determining element—for example, a varactor diode or YIG sphere—should be tunable with voltage or current such that the oscillation conditions can be satisfied over the desired band [37]. The

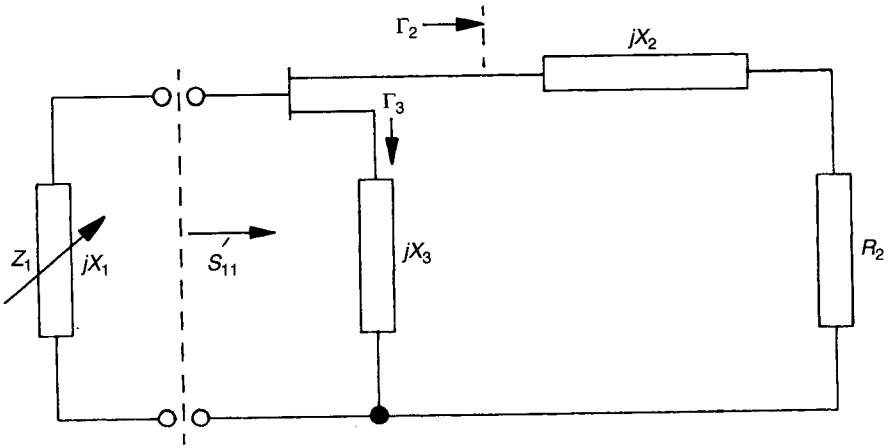


Figure 9.25 Equivalent-circuit representation of a tunable oscillator.

analysis using small-signal measured S parameters is sufficient to realize the negative resistance over the desired band. However, in order to optimize the power output, it is necessary to use one of the nonlinear techniques, for example, nonlinear model, device line, or load-pull method.

In Fig. 9.8 it can be shown that in order to obtain a maximum negative resistance in the oscillator circuit, the load impedance should be the only resistive element. In other words, the impedances connected to the terminals other than the output terminals should be low-loss reactive circuits. To realize an electronically tunable oscillator, it is also necessary that one or more of the reactances connected to the transistor be electronically tunable. Though any of the transistor terminals can be used to connect the useful load, the drain in FETs and collectors in bipolars are more commonly used. The electronic tuning element can be placed in either of the other two terminals. The tunable oscillator can now be represented as shown in Fig. 9.25.

The first step in the oscillator design consists of determining $X_2(\omega)$, $X_3(\omega)$, and R_2 so that the tuning element looks into an impedance having a negative real part in the desired frequency band. In wide-band tunable oscillators, the emphasis is more on the coverage of maximum frequency band rather than the power output. This calls for a design approach different than that used for fixed-frequency oscillators: Equation (9.39) already gives us the two-port S matrix in terms of the three-port S parameters. When a load (with reflection coefficient of Γ_2) is connected to the output terminal, the two-port can be converted to a one-port using

$$S'_{11} = S^T_{11} + \frac{S^T_{12}S^T_{21}}{1/\Gamma_2 - S^T_{22}} \tag{9.63}$$

where S^T_{ij} are the two-port S parameters given by (9.39). The preceding relation

can be represented in terms of three-port S parameters Γ_2, and Γ_3 to be

$$S'_{11} = S_{11} + v + \Gamma_2 \frac{S_{12}S_{21}(1 - S_{33}\Gamma_3) + S_{12}S_{31}S_{23}\Gamma_3 + S_{21}S_{13}S_{32}\Gamma_3 + vS_{23}S_{32}\Gamma_3}{(1 - S_{33}\Gamma_3)(1 - S_{22}\Gamma_2) - S_{23}S_{32}\Gamma_2\Gamma_3} \tag{9.64}$$

where

$$v = \frac{S_{31}S_{13}\Gamma_3}{1 - S_{33}\Gamma_3} \tag{9.65}$$

This relation can be used to analyze the negative resistance in a given device for any combination of Γ_2 and Γ_3. The same relation can also be used to optimize the impedances Γ_2 and Γ_3 for negative resistance. Figure 9.26 shows the reflection gain of a typical 0.5-μm GaAs FET optimized in the frequency range of 6–18 GHz.

Two types of tuning elements are commonly used as Z_1 (Fig. 9.25) to realize wide-band oscillators: YIG spheres and varactors. The YIG sphere, which has been discussed in detail in Chapter 3, is a ferrimagnetic resonant element magneticly tunable over a very wide bandwidth (more than a decade). The varactor is a voltage-controlled capacitance used to accomplish wide-band tunable oscillators with about an octave bandwidth. We now briefly discuss these oscillators before comparing their performances.

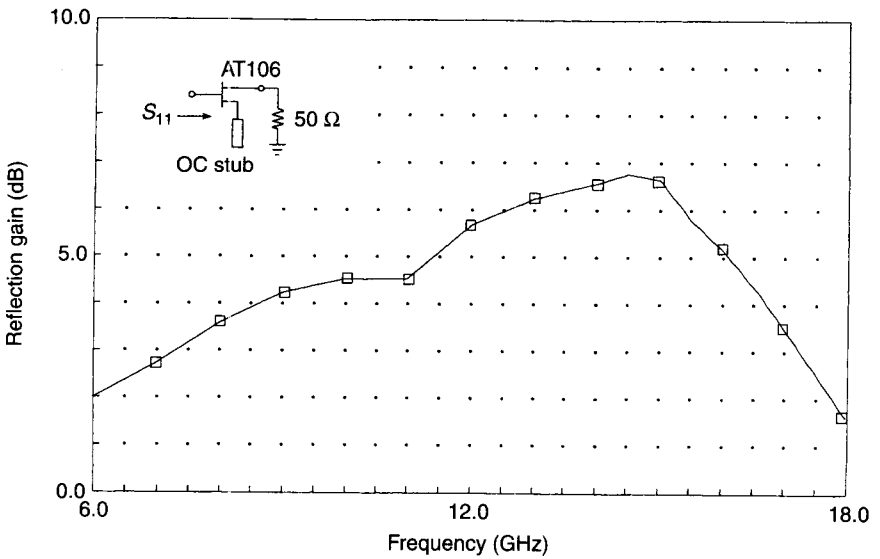


Figure 9.26 Reflection gain of a half-micrometer-gate-length GaAs FET.

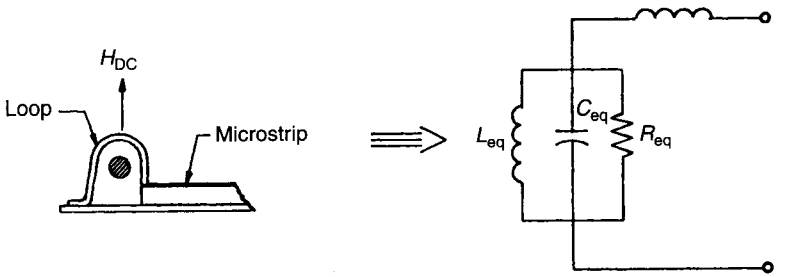


Figure 9.27 Loop-coupled YIG resonator and its *RLC* equivalent-circuit presentation.

9.8.1 YIG-Tuned Oscillators

YIG-tuned oscillators are commonly used as wide-band signal sources when bandwidth and low phase noise are simultaneously required. YIG-tuned Gunn oscillators were the early replacements of the backward-wave tube oscillators. GaAs FETs and microwave silicon bipolar transistors quickly replaced the Gunn devices up to 50 GHz. The main advantages offered by the transistors over Gunn diodes are better efficiency, wider bandwidth, and operation at lower oscillation frequencies. While Gunns do not oscillate below about 5 GHz, the transistor YTOs are realizable down to 0.5 GHz.

A YIG resonator coupled to a microstrip line can be modeled by a parallel *RLC* resonant circuit in series with a self-inductance, as shown in Fig. 9.27. The values of the elements depend on the magnetization, coupling between the sphere, and loop and resonance line width of the YIG sphere (Chapter 3). The resonant frequency ω_0 of the circuit is a linear function of the applied magnetic field to a first-order approximation. A schematic for a YTO using a GaAs FET and a bipolar is represented in Fig. 9.28*a*.

Because the YIG sphere is a very high Q tunable resonant circuit, the oscillation phase condition $\underline{/S'_{11}} = -\underline{/\Gamma}$ can be easily satisfied and stable oscillation achieved in the negative-resistance frequency band of the device. Under large-signal conditions, the relation (9.34) should also be satisfied. Once again, the high-Q characteristic or, in other words, a very high value of the reactance slope $dX/d\omega$ helps to meet this requirement.

A YIG sphere can be used either in the gate (base) or the source (emitter) as a high-Q tunable immitance. In a GaAs FET YTO, using the sphere in the source, the oscillator bandwidth is limited to about an octave due to the impracticality of the needed immitance on the gate [38, 39]. A YIG sphere used in the gate can generate wider bandwidths (8–18 GHz as an example) when a fixed capacitance is used in the source terminal [40]. A common YIG sphere coupled to both gate and source terminals has also been used to achieve 3.5–14 GHz [41] bandwidth.

Using two YIG spheres at both the gate and source terminals (Fig. 9.28*b*) can be used to obtain about a decade bandwidth (2–20 GHz) at the cost of

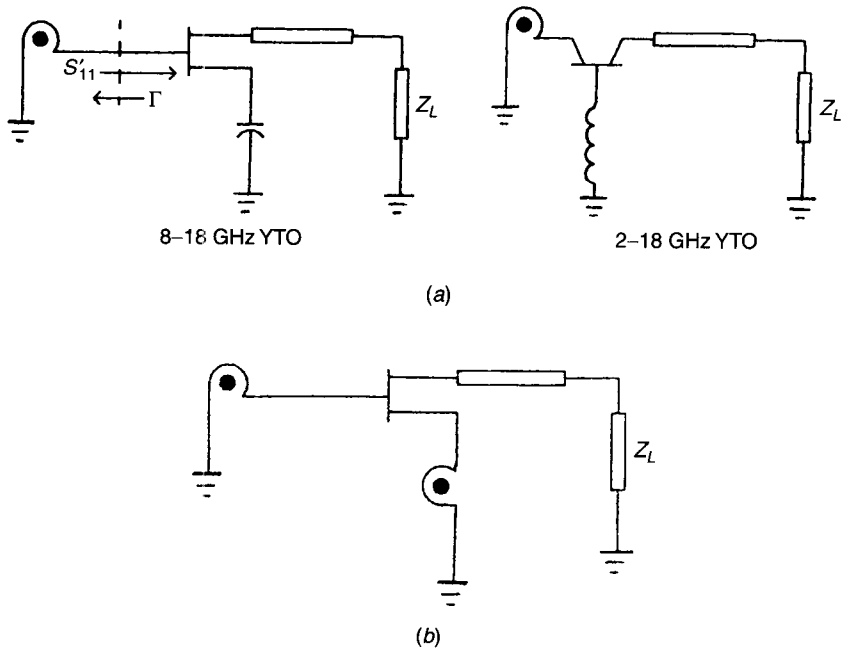


Figure 9.28 YTOs using GaAs FET and (*a*) Si bipolar and (*b*) two YIG spheres at both the gate and source terminals.

complexity of the oscillator circuit [42]. In order to satisfy oscillation conditions over a wide band, the YIG resonator in the source should offer a capacitive reactance while the sphere in the gate should offer an inductive reactance under the same magnetic field. The source sphere plays a dominant role at lower frequencies while the gate sphere determines the high end of the oscillator frequency range.

Silicon bipolar transistors offer lower $1/f$ noise and wider achievable bandwidths due to higher input capacitances. Using a low-noise silicon bipolar transistor with 0.5 μm emitter width and 2 μm emitter–emitter pitch and a single YIG sphere in the emitter (Fig. 9.29*a*), a frequency bandwidth of 2–18 GHz was achieved [1]. Superior phase noise characteristics of this oscillator are shown in Fig. 9.30.

Using a novel composite feedback architecture (Fig. 9.29*b*) with double coupling, the YIG sphere as a series feedback for higher frequencies and as a parallel feedback for lower frequencies achieved a tuning range of 2–22 GHz with a single sphere and single transistor [2].

Improvements in high-frequency GaAs FETs as well as YIG resonators were responsible for the development of YIG-tuned oscillators in the millimeter-wave frequency ranges. Using a single GaAs FET and a single YIG sphere, frequency ranges of 20–40 and 33–50 GHz were reported by engineers

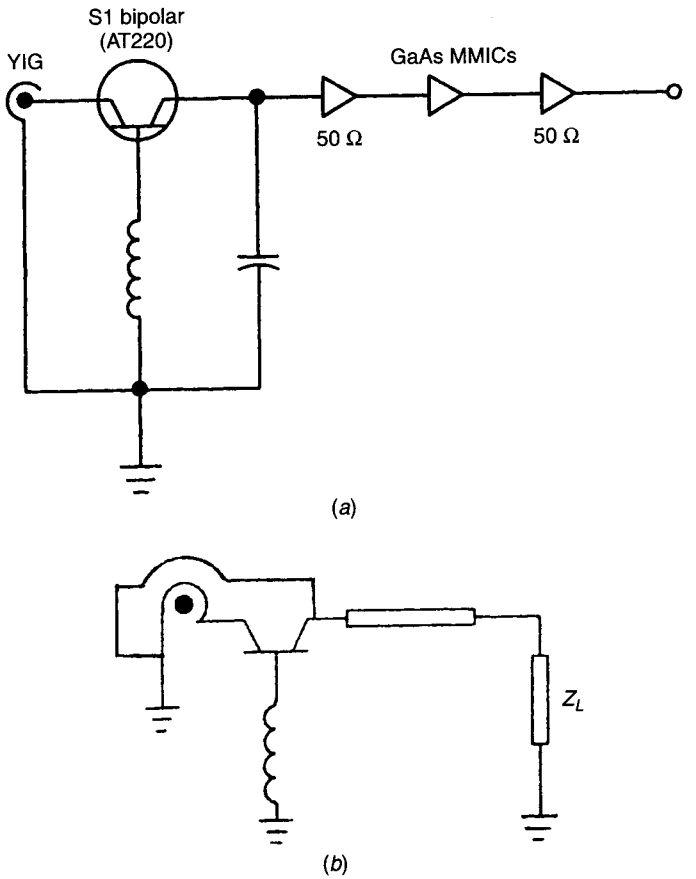


Figure 9.29 (*a*) YIG oscillator constructed by buffering oscillating output of transistor with a series of GaAs MMIC amplifiers. (After Leung [1]. Reprinted with permission of IEEE.) (*b*) YTO using Si bipolar and a composite feedback configuration.

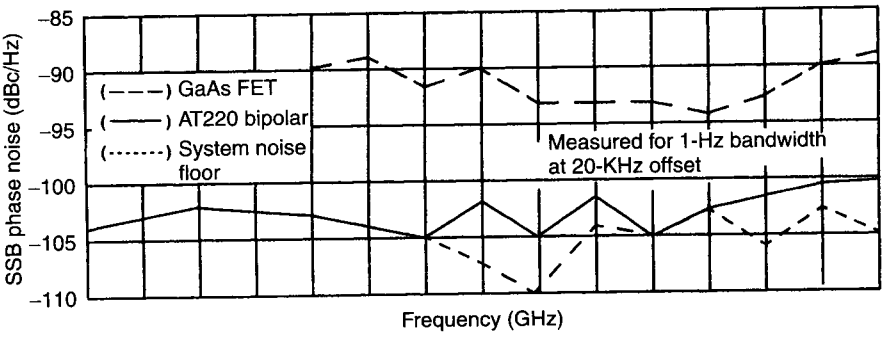


Figure 9.30 Single-sideband phase noise of AT220 bipolar oscillator. (After Leung [1]. Reprinted with permission of IEEE.)

from Avantek [3, 43]. Magnetic field saturation limits the fundamental YIG-tuned oscillators to about 60 GHz.

Parasitic and Spurious Oscillations. A YTO is susceptible to parasitic oscillations due to various reasons. There exist two types of parasitic/spurious oscillations: ones that are fixed in frequency and do not tune with the resonator and others that do tune and originate from the YIG sphere itself. The fixed-frequency oscillations result due to the coupling loop and can be noticed even when the magnetic field is not applied. These oscillations can be minimized by the proper design of the matching circuit and loop so that the fixed-frequency oscillations fall out of the desired frequency band and parasitic oscillations should not appear at all.

The tunable parasitic oscillations due to the YIG sphere are observed due to the higher order modes of the resonator. Thus, for certain applied magnetic field values, the oscillator can jump from the principal resonant curve to a parasitic curve and thereby provoke the phenomenon of tuning hysteresis. It is necessary to characterize the YIG sphere loop coupled to a microstrip line before using it in the oscillator circuit. The tuning hysteresis can also result from nonmonotone behavior of the phase of the active circuit with frequency.

Tuning Linearity. Yttrium–iron–garnet devices can be designed for excellent tuning linearity. It is important to assure that the high-permeability shell and center core pieces of the biasing electromagnet are never saturated in the desired tuning range to be covered. This unsaturated condition minimizes the tuning nonlinearity in a YTO. Another factor in YTO, which affects the tuning linearity, is the matching circuits. In the oscillator design approach explained herein, it may be noted that the YIG sphere does not work exactly at its resonant frequency. In fact, this component serves as a very high Q variable reactance that compensates the reactance presented by the transistor and the associated load and feedback circuits. Hence, deviations from linearity of the tuning curve do depend on the load circuit design. Obviously, the higher the quality factor, the higher the tuning linearity.

9.8.2 Voltage-Controlled Oscillators

Electronic tuning can be achieved using varactor diodes. A varactor-controlled transistor oscillator is an attractive solution for high-speed and moderate-tuning-range applications.

A varactor diode can be either a silicon or a GaAs diode. The main difference between silicon and GaAs is that a higher Q can be obtained from GaAs devices. This is due to the lower resistivity of GaAs for a given doping level N. The GaAs devices have a higher thermal resistance than silicon devices, resulting in significantly larger frequency settling times.

Figure 9.31 is a general representation of a varactor-tuned transistor oscillator. The tuning bandwidth of a YIG-tuned oscillator is limited by the

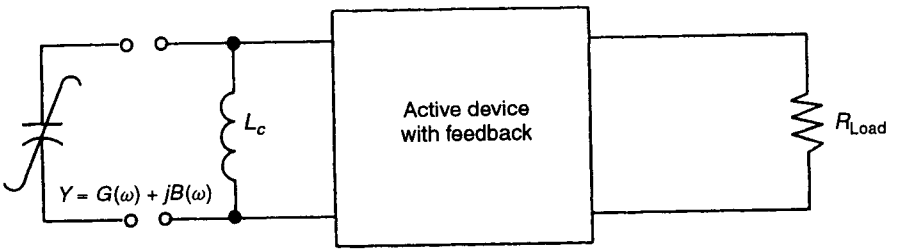


Figure 9.31 Reactance compensation in VCO.

negative-resistance bandwidth, whereas it is the susceptance ratio $C_{min}\omega_{max}/C_{max}\omega_{min}$ that limits the bandwidth of a VCO.

Let the admittance of the transistor and feedback circuit transferred in the plane of the varactor be $Y(\omega) = G(\omega) + jB(\omega)$ with both $G(\omega) < 0$ and $B(\omega) < 0$. Plotting $-B(\omega)$, $C_{min}\omega$, and $C_{max}\omega$ (Fig. 9.32), the electronic bandwidth is limited by the frequencies at which

$$C_{min}\omega_{max} = -B(\omega_{max}) \tag{9.66}$$

$$C_{max}\omega_{min} = -B(\omega_{min}) \tag{9.67}$$

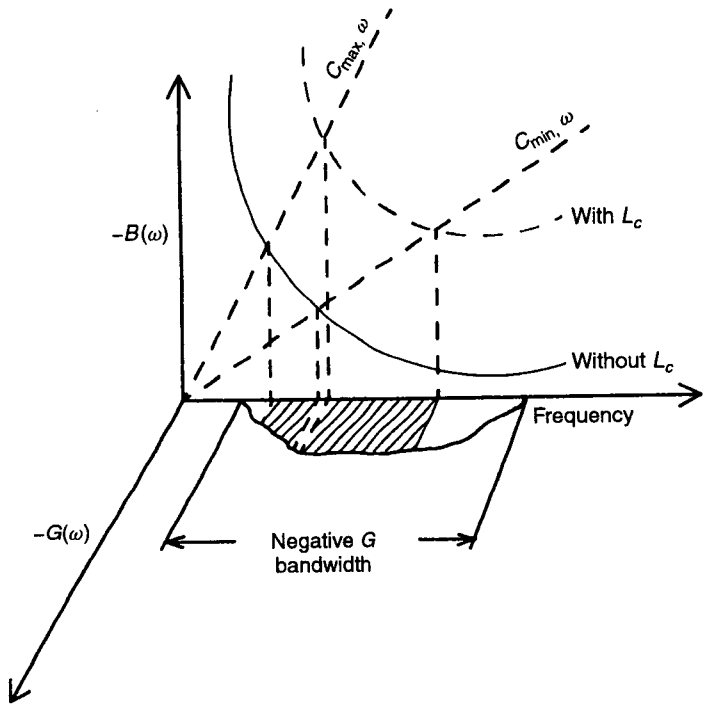


Figure 9.32 Frequency limits of varactor-tuned oscillator. (After Khanna [48]. Reprinted with permission of IEEE.)

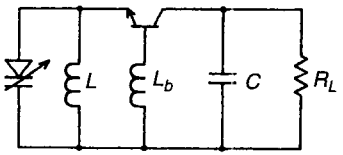


Figure 9.33 Varactor-tuned bipolar transistor oscillator.

This limitation holds even if the negative-conductance bandwidth is much wider, as shown in Fig. 9.32.

To increase the electronic tuning bandwidth for a given C_{max}/C_{min}, it is necessary to reduce the ratio $B(\omega_{min})/B(\omega_{max})$ and modify the variation of $B(\omega)$ to that of the susceptance of a lumped self-inductance. In order to obtain this, a self-inductance L_c is added in parallel to the transistor terminals (Fig. 9.31). In this case, the total susceptance ratio seen by the varactor is given by [44]

$$\frac{B_T(\omega_{min})}{B_T(\omega_{max})} = \frac{B(\omega_{min}) + 1/(L_c\omega_{min})}{B(\omega_{max}) + 1/(L_c\omega_{max})} \tag{9.68}$$

If L_c is small, this ratio is larger and approaches $\omega_{max}/\omega_{min}$. Equating this ratio to that of the susceptance ratio presented by the varactor, we get

$$\frac{\omega_{max}}{\omega_{min}} = \sqrt{\frac{C_{max}}{C_{min}}} \tag{9.69}$$

The value of L_c normally has to be optimized to maximize the bandwidth, due to the parasitic inductances present in the circuit.

Figure 9.33 shows a typical VCO circuit using a bipolar transistor and a varactor diode. This configuration using NEC 567 as the bipolar transistor and a GaAs hyperabrupt varactor covers greater than an octave bandwidth from 3 to 9 GHz [45]. Increased bandwidth can be realized by using two varactors, one in the source and another in the gate, as shown in Fig. 9.34 for a GaAs FET VCO.

Using the preceding configuration, bandwidths greater than an octave have been realized. A low-noise VCO configuration using a bipolar transistor and a silicon abrupt diode [46] is shown in Fig. 9.35. This oscillator has an 8% tuning bandwidth at 8 GHz and has excellent phase noise of 103 dBc at 100 kHz off the carrier.

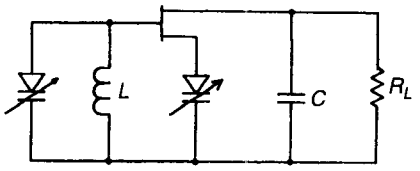


Figure 9.34 Varactor-tuned GaAs FET oscillator.

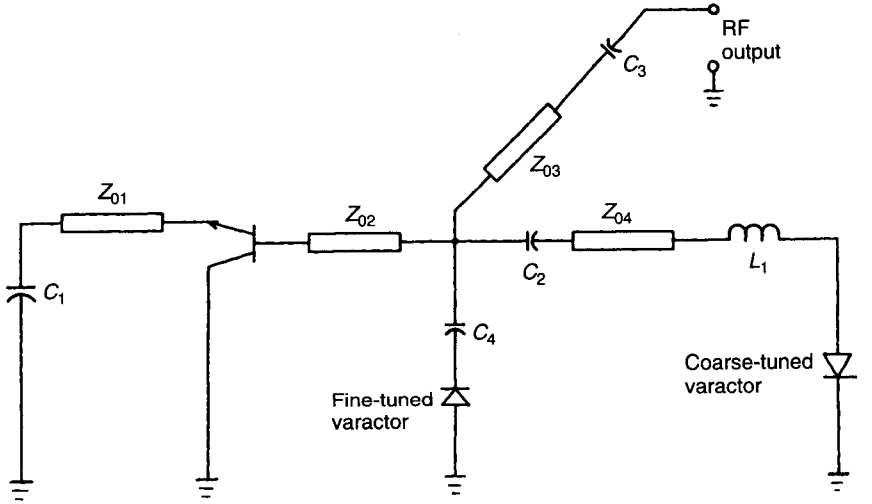


Figure 9.35 Wide-band varactor-tuned polar transistor oscillator. (After Niehenke and Hess [46]. Reprinted with permission of IEEE.)

Wide-band frequency operation of a transistor VCO can be extended in frequency by using a push–push type VCO configuration, as shown in Fig. 9.36. This circuit cancels the fundamental frequency and optimizes the second harmonic at the output point. Using hyperabrupt diodes with capacitance variation of 3.2–0.16 pF and a GaAs FET device, a VCO covering 20–30 GHz has been reported by Winch [47].

Settling Time and Posttuning Drift. In a frequency-tunable oscillator, settling time is defined as the interval between the time when the input tuning drive waveform reaches its final value and the time when the VCO frequency falls within a specified tolerance of a stated final value. Figure 9.37 defines the settling time and post tuning drift (PTD). The input drive waveform reaches its final value at t_0. Settling time is the interval between t_{st} and t_0. The post tuning drift is the frequency drift that occurs between two defined times t_1 and t_2, where t_1 may be specified typically as 10 μs–1 s after the tuning step has been applied to the VCO. Time t_2 is generally defined in the range of 10 μs–1 s after t_1.

Bias voltage drift and thermal effects are the primary contributors to post tuning drift. During the interval when a VCO is being tuned, the junction temperatures of both the transistor and the varactor are changing due to changes in RF circuit efficiency and loading. This causes impedance changes, which result in a frequency shift. The time interval during which this happens is dependent upon the thermal impedance of the devices. Silicon abrupt and hyperabrupt diodes give excellent settling time performance. Gallium arsenide abrupt has a medium and GaAs hyperabrupt has a large settling time compared to the silicon varactor diode.

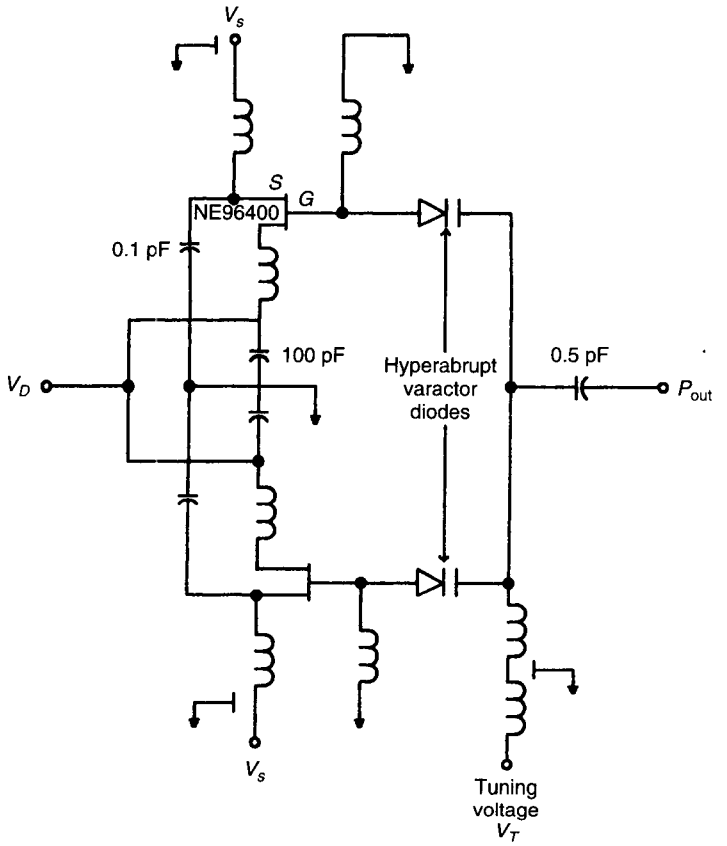


Figure 9.36 VCO circuit schematic diagram. (After Winch [47]. Reprinted with permission of *Electron. Lett.*)

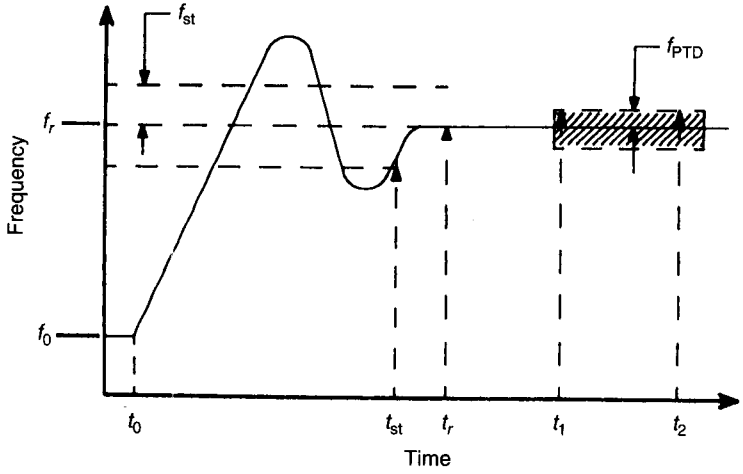


Figure 9.37 Definition of settling time and post tuning drift.

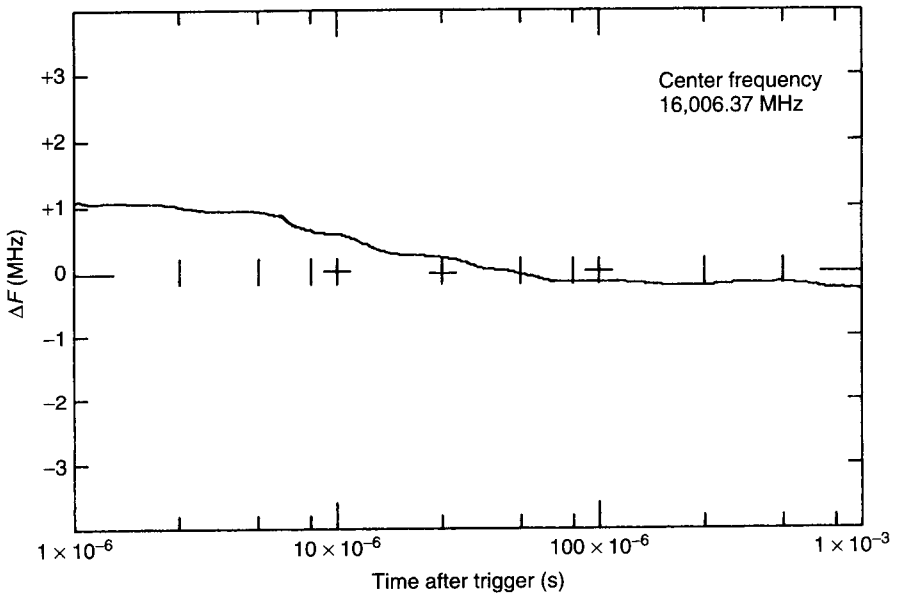


Figure 9.38 Typical settling time of *Ku*-band silicon bipolar VCO for 14.3–16 GHz.

Figure 9.38 presents typical settling times of a silicon bipolar *Ku*-band varactor-controlled oscillator [48]. The output frequency settles within 1 MHz in less than 2 μs.

The YIG- and varactor-tuning techniques previously discussed are significantly different from one another. Table 9.1 compares their important parameters. Figure 9.39 presents typical phase noise plots at 10 GHz for Agilent 2 to 20 GHz YTO, 9–11 GHz VCO, 10 GHz DRO (0.25% tuning bandwidth), and a 10-GHz microstrip transmission-line resonator oscillator (0.5% tuning bandwidth) using Agilent silicon bipolar transistors. In many cases more than one option is available for a given application. Careful trade-offs are generally required for the proper choice.

9.9 OSCILLATOR CHARACTERIZATION AND MEASUREMENTS

Accurate characterization of microwave components is important for their use in various applications. Important characteristics of a microwave oscillator are

(a) frequency range;
(b) power output;
(c) tuning sensitivity;
(d) modulation bandwidth;

Table 9.1 Varactor-Controlled Oscillator versus YIG-Tuned Oscillator

Parameter	YIG Tuning	Varactor Tuning
Bandwidth	Wide—octave and multi-octave	Narrow—octave or less
Phase noise	Excellent	Good performance in narrow bandwidth
Tuning	Current	Voltage
Linearity	Excellent (±1%)	Fair (±10–30%)
Slew rate	<1 MHz/ms	1–10 GHz/μs
Step response time	1–3 ms	<0.1 μs
Post tuning drift constant	Seconds	Microseconds to milliseconds
Temperature stability	50–100 ppm/°C	100–300 ppm/°C
Frequency pushing	Low	High
Oscillator Q	High (>1000)	Low (<100)
Power consumption	High (10 W maximum)	Low (1 W maximum)
Modulation bandwidth	Low (<3 MHz)	High (>30 MHz)
Control driver	Complicated current supply	Easy voltage supply
Volume and weight	Large	Small

(e) output VSWR;
(f) frequency and power variation over expected changes in temperature, power supply (pushing), load VSWR (pulling), and time (aging);
(g) phase noise or jitter;
(h) oscillator switching and settling times;
(i) power consumption;
(j) bias and tuning voltage/current required; and
(k) weight and size.

Many of these parameters are self-explanatory and are measured using well-known microwave measurement techniques. In this section we will discuss the

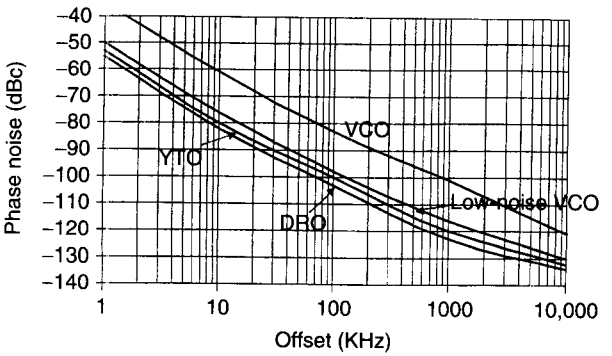


Figure 9.39 Phase noise comparison of various oscillators.

significance and the measurement of the following parameters: modulation bandwidth, frequency and power pulling, and phase noise and jitter.

9.9.1 Modulation Bandwidth

Modulation bandwidth represents the speed at which an oscillator can be modulated through the modulation port, which is commonly the frequency tuning port. High-modulation bandwidths are desired for clock recovery applications in high-speed communication applications and for radio systems requiring wide-band FM modulation capability. Modulation bandwidth is an important parameter to consider while designing a phase-lock loop circuit for a tunable oscillator. A frequency-tunable oscillator can be treated as a typical FM system in which the modulation bandwidth depends upon the input impedance of the tuning circuit. The series resistance and shunt capacitance of the input tuning impedance are used to control the modulation bandwidth. Modulation bandwidths greater than 20 MHz are easily achieved in X-band VCOs.

Figure 9.40 shows a setup commonly used for modulation bandwidth measurement. This measurement is based on the fact that the frequency deviation for a given voltage swing on the tuning port of a VCO decreases as its frequency increases. The frequency at which the frequency deviation reduces to 0.707 (or -3 dB) of the DC or low-frequency value represents the modulation bandwidth of the tuning port. In a typical FM system, the modulation index m is proportional to the ratio of the frequency deviation and the modulation frequency. With the frequency deviation being directly proportional to the modulating signal level, at low modulation frequencies, the modulation index remains constant when the level of the modulation signal is increased proportionally to the increase in the modulation frequency.

A low-level, low-frequency (typically less than 10% of the anticipated bandwidth) signal is used to modulate a given VCO. A convenient level of the FM sidebands (say 20 dB below carrier) is adjusted with the modulating voltage. The modulating frequency and signal level are simultaneously increased until the modulation bandwidth is decreased to 3 dB below the reference level. The modulation frequency at this point represents the modulation bandwidth of the

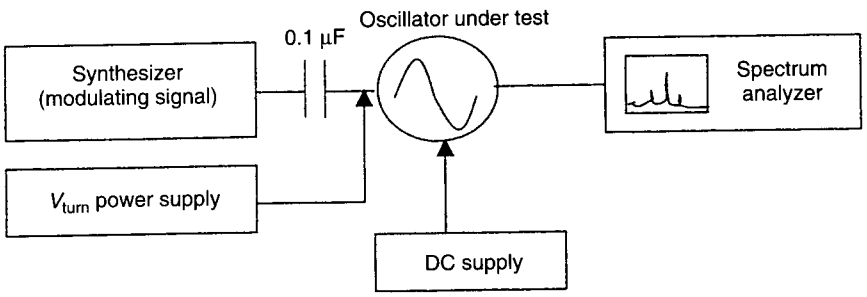


Figure 9.40 Modulation bandwidth measurement setup.

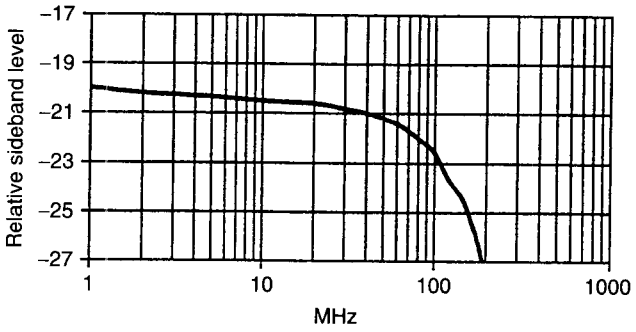


Figure 9.41 Modulation bandwidth of Agilent 10-GHz VCO.

VCO under test. Figure 9.41 shows a typical modulation bandwidth curve of an Agilent 10-GHz VCO.

9.9.2 Frequency and Power Pulling

Oscillator pulling characteristics represent variation of its frequency and power with change in load characteristics. Variations in both magnitude and phase of the load impedance (or reflection coefficient) affect the oscillator performance.

Frequency pulling is defined by the total variation in frequency due to 0–2π phase variation of a fixed load (typically one with 12 dB return loss). Figure 9.42 shows the experimental setup used for pulling measurements for a commonly used load VSWR of 1.67:1 (12 dB return loss). A simplified expression for frequency pulling can be written as [49]

$$\Delta F_{\text{total}} = \frac{f_0}{2Q_{\text{ext}}}\left(S - \frac{1}{S}\right) \tag{9.70}$$

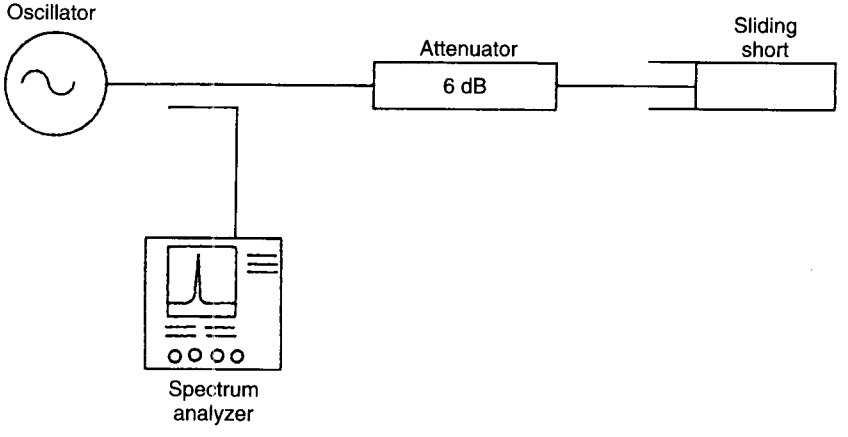


Figure 9.42 Measurement setup for oscillator pulling figure.

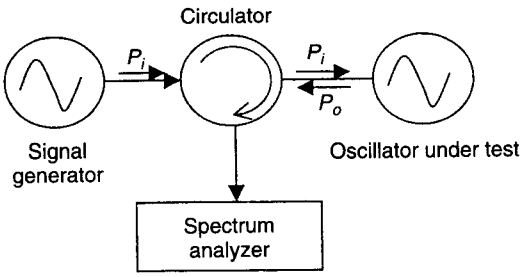


Figure 9.43 Injection locking technique to measure oscillator's pulling figure.

where S is the load VSWR, f_0 is the oscillator angular frequency, and Q_{ext} is external quality factor of the oscillator (another way to represent oscillator frequency dependence on load impedance). As an example, for a typical 10-GHz DRO, if the frequency pulling into a load VSWR of 1.67:1 is measured to be 4 MHz, the external quality factor can be calculated to be

$$Q_{ext} = 10{,}000\left(\tfrac{1}{2}\right)\left(\tfrac{1}{4}\right)(1.67 - 1/1.67) = 1339 \tag{9.71}$$

Injection locking of an oscillator is another technique used to determine oscillator frequency dependency on load. As shown in Fig. 9.43, using a frequency synthesizer, a small amount of power (generally 20 dB lower than the oscillator output power) at or near the oscillator frequency is injected in to the oscillator through a circulator or directional coupler. The frequency of the injected oscillator is varied and the spectrum analyzer is monitored. As the frequency of the synthesizer approaches that of the oscillator under test, injection sidebands start appearing until the oscillator gets locked to the injected signal. With the continuing change, the frequency oscillator remains locked over a certain bandwidth (called injection locking bandwidth $\Delta\omega_i$) before the injection locking sidebands reappear. The external quality factor Q_{ext} can be calculated using the well-known equation [50]

$$Q_{ext} = \frac{2\omega_o}{\Delta\omega_i} \cdot \sqrt{\frac{P_i}{P_o}} \tag{9.72}$$

where P_i is the injection locking signal power and P_o is the output power of the oscillator under test.

Power pulling of an oscillator represents the change in oscillator power with load impedance variation and can also be measured using the setup of Fig. 9.42 and replacing the spectrum analyzer with a power meter. Power pulling is representative of the output VSWR of the oscillator. The lower the power pulling, the lower the output VSWR. By changing the sliding short phase over 0–2π, the relative power change is measured as the power pulling for 12 dB return

loss. The following special precautions are important while measuring power pulling:

(a) Finite directivity of the directional coupler can result in a significant measurement error. A coupler with the highest directivity at the frequency of measurement should be selected for accurate measurement.
(b) A directional coupler's directivity depends also on the load. Using loads with less than 12 dB return loss can increase the error of measurement due to their impact on coupler directivity.

9.9.3 Phase Noise and Jitter

Communication system designers deal with jitter while component designers deal with phase noise. Phase noise and timing jitter are both measures of uncertainty in the oscillator output signal in the frequency and time domains, respectively. Figures 9.44*a* and *b* represent these uncertainties in frequency and

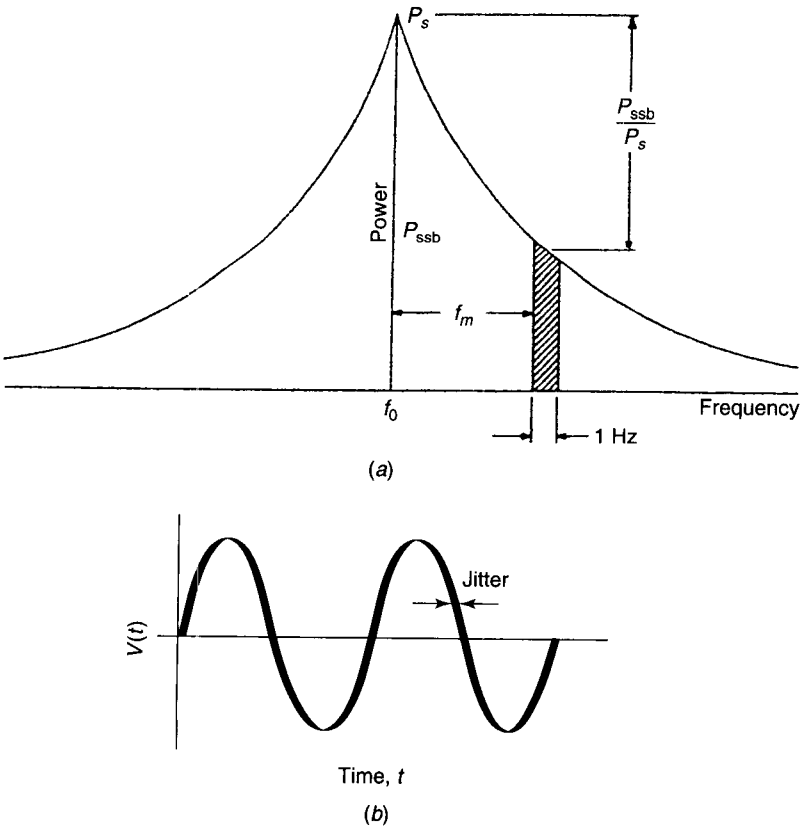


Figure 9.44 (*a*) Oscillator RF spectrum. (*b*) Time-domain phase noise jitter.

time domains. Clock jitter is the variation in timing of a critical instant in a periodic waveform with respect to a jitter-free reference. In other words, jitter represents short-term variations of the significant instants of a digital signal from their ideal positions in time. A "significant instant" is any convenient, easily identifiable point on the signal such as a rising or falling edge.

In the frequency domain, the oscillator power is not concentrated at the carrier frequency alone but rather is distributed around it: The spectral distributions on the opposite sides of the carrier are known as noise sidebands. Noise can be discussed as a modulation phenomenon, and these sidebands can be separated into amplitude and frequency modulation components. With the amplitude components being much smaller compared to the frequency modulation components, the oscillator noise close to the carrier almost always represents frequency modulation noise. The U.S. National Institute of Standards and Technology defines single-sideband phase noise $\pounds(f)$ as the ratio of the noise power in 1 Hz bandwidth at a frequency f away from the carrier to the signal power of the carrier:

$$\pounds(f) = 10 \log \frac{P_{ssb}}{P_c} \tag{9.73}$$

where $\pounds(f)$ is expressed in decibels relative to the carrier per hertz (dBc/Hz). A bandwidth of 1 Hz is used to allow the phase noise in other bandwidths to be easily calculated for comparison. The SSB phase noise at a carrier frequency is commonly graphically represented on a log-log plot. Using a log scale on the frequency axis, the phase noise can be conveniently displayed for a wide range of frequency offsets.

Another term representing one-sided spectral density of the phase fluctuations on a per-hertz basis, $S\phi$, is also commonly used to describe phase instability or phase noise of an oscillator. The term *spectral density* describes the energy distribution as a continuous function, expressed in units of energy within a specified bandwidth. Thus $S\phi$ is defined as

$$S\phi(f) = \frac{\Delta\phi^2_{\text{rms}}}{\text{BW used to measure}\Delta\phi_{\text{rms}}} \qquad \text{rad}^2/\text{Hz} \tag{9.74}$$

If the modulation sidebands are such that the total phase deviations are $\ll 1$ rad, then $\pounds(f)$ and $S\phi(f)$ are related by

$$S\phi(f) = 2\pounds(f) \tag{9.75}$$

Phase Noise Measurement. There are several methods of making phase noise measurements, each with its own set of advantages and disadvantages. A summary of some of the common methods follows.

(A) Direct Measurement. The power spectral density of a signal can be conveniently measured by directly connecting an oscillator through an appropriate attenuator to a spectrum analyzer, To measure phase noise at a given offset f hertz from the carrier, the noise level in dBc is measured in the predetermined resolution bandwidth. This number is then converted to phase noise per hertz by subtracting 10 log(resolution bandwidth in hertz). This measurement is limited in many ways and is valid if the following conditions are met:

(i) The spectrum analyzer SSB phase noise at the offset of interest is lower than the noise of the oscillator under test.

(ii) Since the spectrum analyzer measures total noise power, the amplitude noise of the oscillator under test must be significantly (>10 dB) lower than its phase noise.

(B) Using a Frequency Discriminator [10, 51]. Two types of frequency discriminators can be used to measure phase noise: a high-Q cavity discriminator and a delay-line discriminator. A block diagram of a simplified phase noise measurement setup using a high-Q cavity discriminator is shown in Fig. 9.45. This technique is based on the conversion of the frequency modulation signal into an AM signal using frequency response variation of a high-Q tuned circuit. The signal from the oscillator under test is divided into two paths: an LO path and an RF signal path. An adjustable phase shifter and an attenuator are placed in the LO path, while a high-Q carrier suppression filter, consisting of a tunable cavity and a circulator, is placed in the RF path. Using this cavity in this mode, a reduced RF power corresponding only to the FM sidebands of the oscillator signal is reflected toward the balanced mixer. The phase shifter in the LO path is adjusted in order to reintroduce the carrier phase advanced by 90°. Phase noise sidebands caused by frequency modulation are thus converted to AM sidebands, which are easily detected by the mixer diodes, while AM

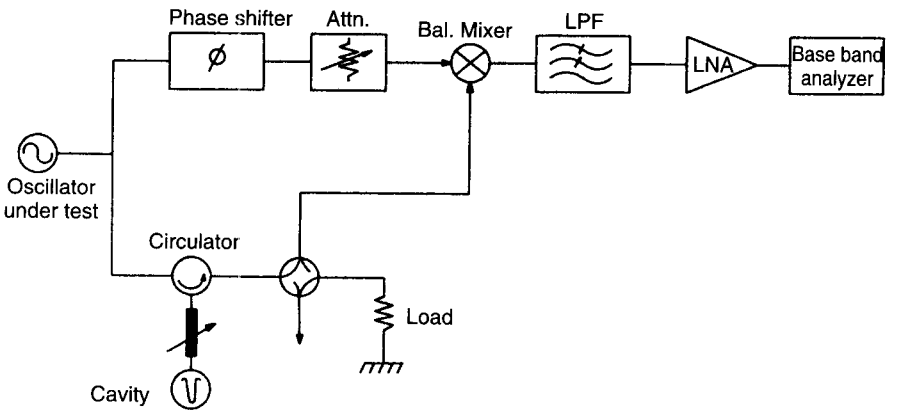


Figure 9.45 FM noise measurement setup using a cavity discriminator method.

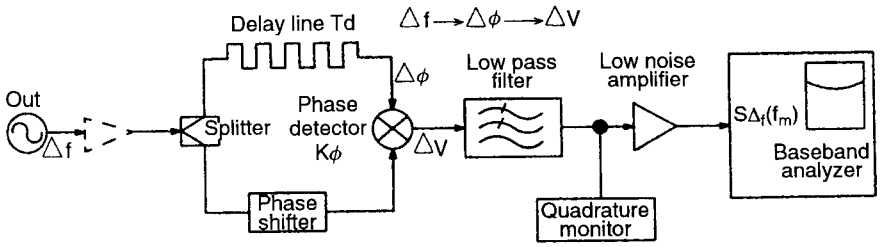


Figure 9.46 FM noise measurement setup using a delay-line discriminator method.

noise sidebands are converted to FM and are not detected by the mixer. The mixer output is detected by a highly sensitive, adjustable bandwidth-selective voltmeter.

The phase noise measurement system is calibrated by applying modulation to the oscillator under test. The modulation is set using a microwave spectrum analyzer at a known level (at modulation index of 2.405, for example, corresponding to a carrier null) and the baseband measurement is used to determine the system calibration constants.

A block diagram of a phase noise measurement system using a delay-line discriminator is shown in Fig. 9.46. The principle of operation of this approach is similar to that of a cavity discriminator approach. This method offers lower sensitivity but wider bandwidth and ease of automation compared to a cavity discriminator. The length of the delay line is adjusted to place zero crossing at the oscillator carrier frequency. In practice, this means setting up to obtain zero DC output from a balanced mixer used as a phase detector in order to minimize system sensitivity-to-amplitude variations. The length of the delay line controls the system sensitivity of the phase noise measurement and can be increased by increasing the delay-line length at the cost of bandwidth. The calibration procedure remains similar to that of the cavity discriminator.

(C) Phase Detector Method. Using two sources, this method yields overall the best sensitivity. Figure 9.47 shows the block diagram used in this technique. The basis of this method is the double-balanced mixer used as a phase detector.

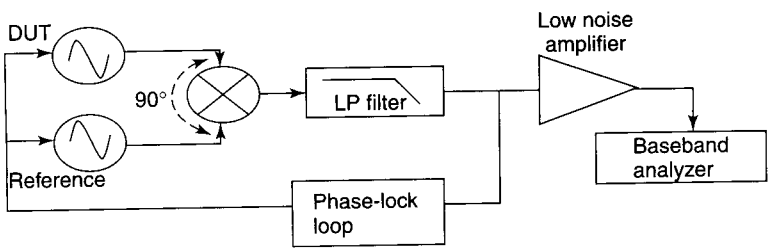


Figure 9.47 Phase detector method.

For the mixer to act as an accurate phase detector, it must remain within several degrees of quadrature. Two signals at identical frequencies and nominally in phase quadrature are input to the phase detector. At quadrature, the output of the phase detector is a difference frequency of 0 Hz and an average voltage output of 0 V. Phase noise in the oscillator causes phase deviation and hence a corresponding small fluctuating voltage at the phase detector output. For small phase deviations ($\ll 1$ rad, a good assumption for quality sources), this fluctuating voltage is proportional to the fluctuating phase difference between the two signals. This phase difference represents the combined phase modulation sidebands of the two input signals.

In other words, when the two input signals are identical in frequency and in phase quadrature, the output of the phase detector is a voltage directly proportional to the combined phase modulation sidebands of the two input signals.

Agilent Technologies E5504B is an automated phase noise measurement setup for use up to 26 GHz. Using the phase detector method, a phase noise plot of an Agilent low-noise VCO at 21.23 GHz is presented in Fig. 9.48.

Jitter Measurement. Jitter is an important performance parameter in digital circuits. Excessive jitter degrades the transmission performance of a digital circuit and prevents correct sampling of a digital signal by a recovered clock. It can cause eye closure and increase bit errors. Jitter is expressed in many different ways:

(a) *Unit Intervals (UI).* A single unit interval is one cycle of the clock frequency or the normalized clock period. Jitter is expressed as a decimal fraction of one unit interval (Fig. 9.49).

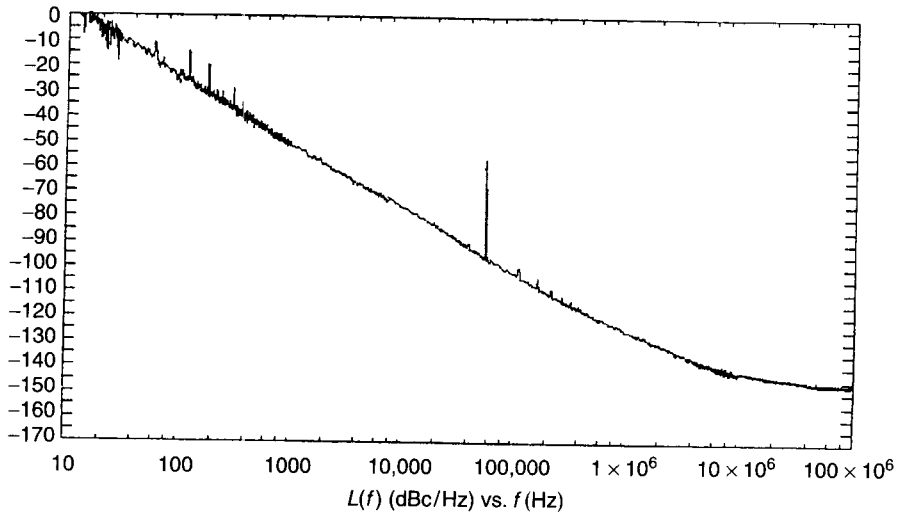


Figure 9.48 Phase noise plot of Agilent 21.4-GHz VCO using E5504B.

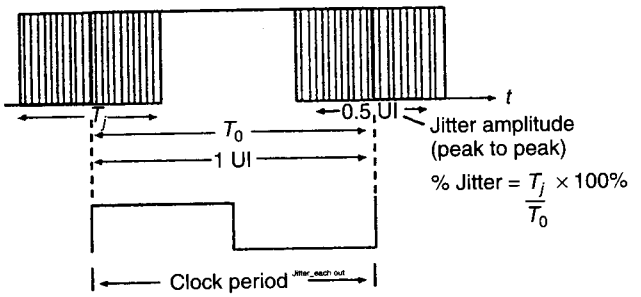


Figure 9.49 Jitter and unit interval.

(b) *Degrees.* Jitter is expressed in degrees where one cycle corresponds to 360 degrees.

(c) *Time.* Knowing the time period from frequency, degrees or UI can be easily converted to time. Jitter is commonly expressed in picoseconds.

(d) *Radians.* Jitter can be expressed in radians, where 2π radians equals one cycle or one time period.

Given phase deviation ϕ in radians,

$$\text{Jitter in unit intervals } U = \frac{\phi \text{ (rad)}}{2\pi} \tag{9.76}$$

$$\text{Time jitter in seconds } \Delta t = \frac{U \text{ (UI)}}{\nu_0} = \frac{\phi \text{ (rad)}}{2\pi\nu_0} \tag{9.77}$$

$$\text{Jitter in degrees} = 360 \times \frac{\phi \text{ (rad)}}{2\pi} \tag{9.78}$$

The jitter of an oscillator or a clock source can be measured using a number of different approaches. The simplest means of jitter measurement is by a high-speed digital sampling oscilloscope such as the Agilent 83480A. However, this technique is limited in high frequency and offers poor measurement sensitivity due to the large measurement bandwidth involved. Additionally, this technique does not provide any information about the jitter spectral characteristics. Another approach is using sampler-based instruments like Agilent 71501C, a jitter and eye diagram analyzer. These instruments operate by taking time samples of the data and analyzing those by using digital signal processing techniques. These instruments can be presently used to about 10 Gb/s.

A practical method of determining oscillator phase jitter for component designers is by measuring the phase noise plot and calculating jitter in the required offset bandwidth. This is possible because rms phase jitter is nothing but integrated rms phase noise in the specified bandwidth. In order to calculate jitter from oscillator phase noise data, the oscillator phase noise spectral den-

sity is integrated over the specified frequency band. The square root of this quantity represents jitter:

$$\Delta\varphi_{\text{rms}} = \sqrt{\int_{f_1}^{f_2} S_\varphi(f)\, df} \tag{9.79}$$

where f_1 and f_2 represent the specified bands for jitter measurement. Timing jitter can be found from the fact that 2π of the phase is equivalent to the signal period T_0 as follows:

$$\Delta t_{\text{rms}} = \frac{T_0}{2\pi}\Delta\varphi_{\text{rms}} \tag{9.80}$$

These equations can be used to estimate the phase noise corresponding to the rms time jitter in a frequency band and vice versa. Precise integration of the measured phase noise curve over the desired frequency band is required to accurately determine jitter. Using Agilent E5504B, the phase noise plot shown in Fig. 9.48 can be converted to jitter in any given bandwidth. In the 50-KHz to 80-MHz band the result for Agilent's 21.23-GHz VCO can be represented as

$$0.92 \text{ mUI rms} = 0.33^\circ = 43 \text{ fs} = 5.78 \times 10^{-3} \text{ rad rms} \tag{9.81}$$

In many cases reasonable estimates can also be made from phase noise data [52]. Jitter is a strong function of the highest phase noise point in the offset frequency band of interest. If the phase noise is assumed constant with frequency, using (9.79) and (9.80), we can write

$$\frac{2\pi}{T_0}\Delta t_{\text{rms}} = \sqrt{S\varphi_k(f_2 - f_1)} \tag{9.82}$$

or

$$S\varphi_k = \frac{4\pi^2}{f_2 - f_1}\left(\frac{\Delta t_{\text{rms}}}{T_0}\right)^2 \tag{9.83}$$

For a given jitter requirement (Δt_{rms}), desired phase noise can be calculated using 9.83. In reality, the phase noise plot of an oscillator always rolls down with frequency. This means that if the oscillator phase noise at f_1 is lower than the calculated value from 9.83, the oscillator will always meet the requirement. However, for accurate relationship between phase noise and jitter, detailed integration of the phase noise curve is necessary [52, 53]. For a given bandwidth, the rms value of jitter is usually well defined. The peak-to-peak value for jitter can however be 7–14 times the rms value.

REFERENCES

1. Leung, C., C. Snapp, and V. Grande, "A 0.5 μm Silicon Bipolar Transistor for Low Phase Noise Oscillator Applications Up to 20 GHz," *IEEE Int. Microwave Symp. Dig.*, June 1985, pp. 383–385.
2. Khanna, A. P. S., and J. Buenrostro, "2–22 GHz Low Phase Noise Silicon Bipolar YIG Tuned Oscillator Using Composite Feedback," *IEEE Int. Microwave Symp. Dig.*, 1992, pp. 1297–1299.
3. Khanna, A. P. S., and J. Hauptman, "18–40 GHz 13 dBm GaAs FET YIG Tuned Oscillator," *IEEE Int. Microwave Symp. Dig.*, 1991, pp. 209–212.
4. Camiade, M., and A. Bert, "Wide Tuning Bandwidth *Ku* band Varactor FET Oscillators," in *Proceedings of the Fifteenth European Microwave Conference*, Microwave Exhibitors & Publishers, Kent, U.K., 1985, p. 413.
5. Pluchart, J. O., B. U. Klepser, H. A. Ainspan, and M. Soyuer, "Fully Monolithic 3-V SiGe Differential Voltage Controlled Oscillator for 5 GHz and 17 GHz Wireless Applications," in *Proceeding of the European Solid State Circuits Conference*, Microwave Exhibitors & Publishers, Kent, U.K. Sept. 1998, pp. 332–335.
6. Rheinfelder, C. N., K. M. Strohm, L. Metzger, H. Kibbel, J.-F. Luy, and W. Heinrich, "47 GHz SiGe MMIC Oscillator," *IEEE Int. Microwave Symp. Dig.*, Vol. I, 1999, pp. 5–8.
7. Gruhle, A., C. Mahner, and K. Weidman, "Low Phase Noise 10 GHz DRO with Low $1/f$ Noise SiGe HBTs," *28th Eur. Microwave Proc.*, Microwave Exhibitors & Publishers, Kent, U.K., Vol. 1, Oct. 1998, pp. 391–394.
8. Hosoya, K., S. Tanaka, Y. Amamiya, T. Niwa, H. Shimakawi, and K. Honjo, "A Low Phase Noise 38-GHz HBT MMIC Oscillator Utilizing a Novel Transmission Line Resonator," *IEEE Int. Microwave Symp. Dig.*, 2000, pp. 47–50.
9. Boyles, J. W., "The Oscillator as a Reflection Amplifier: An Initiative Approach to Oscillator Design," *Microwave J.*, Vol. 29, June 1986, pp. 83–98.
10. Bodway, G. E., "Circuit Design and Characterization of Transistor by Means of Three-Port Scattering Parameters," *Microwave J.*, Vol. 11, May 1968, pp. 55–63.
11. Khanna, A. P. S., "Parallel Feedback FETDRO Using 3-Port *S*-Parameters," *IEEE Int. Microwave Symp. Dig.*, 1984, pp. 181–183.
12. Khanna, A. P. S., "Three Port *S*-Parameters Ease GaAs FET Designing," *Microwave RF*, Vol. 24, Nov. 1985, pp. 81–84.
13. Soares, R., J. Graffeuil, and J. Obregon, *Applications of GaAs MESFETs*, Artech House, Dedham, MA, 1983.
14. Khanna, A. P. S., and J. Obregon, "Microwave Oscillator Analysis," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-29, June 1981, pp. 606–607.
15. Basawaptna, G. R., and R. B. Stancliff, "A Unified Approach to the Design of Wide-Band Microwave Solid State Oscillators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, May 1979, pp. 379–385.
16. Khanna, A. P. S., "Oscillateurs microondes stables integres a transistors et resonateurs dielectriques," Ph.D. Thesis, University of Limoges, France, Sept. 1981.
17. Kurokawa, K., "Noise in Synchronized Oscillators," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-16, 1968, pp. 234–240.

18. Kajfez, D., and P. Guillon, *Dielectric Resonators*, Artech House, Dedham, MA, 1986, Chap. 10.

19. Golio, J. M., and C. M. Krowne, "New Approach for FET Oscillator Design," *Microwave J.*, Vol. 21, Oct. 1978, pp. 59–61.

20. Khanna, A. P. S., J. Obregon, and Y. Garault, "Efficient Low Noise Three Port *X*-Band FET Oscillator Using Two Dielectric Resonators," *IEEE Int. Microwave Symp. Dig.*, Dallas, 1982, pp. 277–279.

21. Wagner, W., "Oscillator Design by Device Line Measurement," *Microwave J.*, Vol. 22, Feb. 1979, pp. 43–48.

22. Poulin, D., "Load-Pull Measurements Help You Meet Your Match," *Microwaves*, Vol. 19, Nov. 1980, pp. 61–65.

23. Galani, Z., et al., "Analysis and Design of a Single-Resonator GaAsFET Oscillator with Noise Degeneration," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-32, Dec. 1984, pp. 1556–1565.

24. Ishihara, O., et al., "A Highly Stabilized GaAsFET Oscillator Using a Dielectric Resonator Feedback Circuit in 9–14 GHz," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, Aug. 1980, pp. 817–824.

25. Johnson, K. M., "Large Signal GaAs MESFET Oscillator Design," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, Mar. 1979, pp. 217–227.

26. Tsironis, C., and P. Lesartre, "Temperature Stabilization of GaAsFET Oscillators with Dielectric Resonators," *Proc. 12th European Microwave Conf.*, Microwave Exhibitors & Publishers, Kent, U.K., 1982, pp. 181–186.

27. Lee, J., et al., "Digital and Analog Frequency-Temperature Compensation of Dielectric Resonator Oscillators," *IEEE Int. Microwave Sym. Dig.*, Vol. 41, San Francisco, 1984, pp. 277–279.

28. Virdee, B. S., "Current Techniques for Tuning Dielectric Resonators," *Microwave J.*, Vol. 41, Oct. 1998, pp. 130–138.

29. Camiade, M., et al., "Low Noise Design of Dielectric Resonator FET Oscillators," *Proc. 13th European Microwave Conf.*, Microwave Exhibitors & Publishers, Kent, U.K., 1983, pp. 297–302.

30. Farr, A. N., et al., "Novel Technique for Electronic Tuning of Dielectric Resonators," *Proc. 13th European Microwave Conf.*, Microwave Exhibitors & Publishers, Kent, U.K., 1983, pp. 791–796.

31. Watanabe, S., et al., "Very High *Q* Dielectric Resonators Voltage Controllable Oscillator," *Proc. 8th European Microwave Conf.*, Microwave Exhibitors & Publishers, Kent, U.K., 1978, pp. 269–273.

32. Khanna, A. P. S., and E. Topacio, "A Highly-Stable 36 GHz GaAs FET DRO with Phase-Lock Capability," *Microwave J.*, Vol. 32, July 1989, pp. 117–122.

33. Hercsfeld, P. R., et al., "Optically Tuned and FM Modulated *X*-band Dielectric Resonator Oscillator," *Proc. 14th European Microwave Conf.*, Leige, 1984, pp. 268–273.

34. Hercsfeld, P. R., et al., "Optically Controlled Microwave Devices and Circuits," *IEEE MTT-S Int. Microwave Symp. Dig.*, St Louis, 1985, pp. 211–214.

35. Khanna, A. P. S., E. Topacio, E. Gane, and D. Elad, "Low Jitter Silicon Bipolar VCOs for Applications in High Speed Optical Communication Systems," *IEEE Int. Microwave Symp. Dig.*, 2001, pp. 1567–1570.

36. Wade, P. C., "Say Hello to Power FET Oscillators," *Microwaves*, Vol. 18, Apr. 1979, pp. 104–109.
37. Pengelly, R. S., *Microwave Field Effect Transistors—Theory Design and Applications*, Research Studies Press, Letchworth, Herts., England, 1982, Chap. 7.
38. Heyboer, T., and F. Emery, "YIG Tuned GaAsFET Oscillators," *IEEE Int. Microwave Symp. Dig.*, 1976, pp. 48–50.
39. Ruttan, T., "*X*-Band GaAs FET YIG Tuned Oscillators," *IEEE Int. Microwave Symp. Dig.*, 1977, p. 264.
40. Oyafuso, R., "An 8–18 GHz FET YIG Tuned Oscillator," *IEEE Int. Microwave Symp. Dig.*, 1979, pp. 183–184.
41. Le Tron, et al., "Multioctave FET Oscillators Double Tuned by a Single YIG," *ISSCC Dig. Tech. Papers*, 1979, pp. 162–163.
42. Obregon, J., et al., "Decade Bandwidth FET Functions," *IEEE Int. Microwave Symp. Dig.*, 1981, pp. 141–142.
43. Zensius, D., et al., "GaAs FET YIG Oscillator Tunes from 26 to 40 GHz," *Microwaves RF*, Vol. 22, Oct. 1983, pp. 129–139.
44. Marechal, P., and J. Obregon, "1.5 to 4.5 GHz Varactor-Tuned Transistor Oscillators," *Proc. 9th European Microwave Conf.*, Brighton, England, 1979, pp. 621–624.
45. Obregon, J., et al., "Ultrabroad Band Electronically Tunable Oscillators," *Proc. 11th Microwave Conf.*, 1981, pp. 475–479.
46. Niehenke, E., and R. Hess, "A Microstrip Low Noise *X*-band Voltage Controlled Oscillator," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, Dec. 1979, pp. 1075–1079.
47. Winch, R. G., "*K*-Band FET Doubling Oscillator," *Electron. Lett.*, Vol. 18, Oct. 1982, pp. 946–947.
48. Khanna, A. P. S., "Fast Settling Low Noise *Ku* Band Fundamental Bipolar VCO," *IEEE Int. Microwave Symp. Dig.*, 1987, pp. 579–581.
49. Obregon, J., and A. P. S. Khanna, "Exact Derivation of the Non-Linear Negative Resistance Oscillator Pulling Figure," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, July 1982, pp. 1109–1111.
50. Adler, R., "A Study of Locking Phenomenon in Oscillators," *Proc. IRE*, Vol. 34, June 1946, pp. 351–357.
51. Hewlett-Packard, *Phase Noise Characterization of Microwave Oscillators*, Product Note 11729C.
52. Drakhlis, B., "Calculate Oscillator Jitter by Using Phase Noise Analysis," *Microwaves RF*, Vol. 40, Jan. 2001, pp. 82–90, and Feb. 2001, pp. 109–119.
53. www.hfoscillators.com

PROBLEMS

9.1 Two-port *S* parameters of Agilent GaAs FET AT10600 in the common-source configuration are given in the table below. Calculate the two-port *S* parameters in the common-drain and common-gate configurations using the three-port *S*-parameter approach.

Frequency (GHz)	S_{11}		S_{21}			S_{12}			S_{22}	
	Mag	Ang	dB	Mag	Ang	dB	Mag	Ang	Mag	Ang
8	0.59	−128	8.4	2.64	91	−27.3	0.04	88	0.71	−3
9	0.56	−145	8.0	2.50	81	−26.2	0.05	88	0.70	−3
10	0.55	−164	7.5	2.37	73	−25.4	0.05	90	0.68	−5
11	0.56	178	7.1	2.27	63	−24.4	0.06	89	0.64	−8
12	0.58	161	6.5	2.11	52	−23.6	0.07	84	0.61	−12
13	0.62	150	5.7	1.92	44	−22.6	0.07	82	0.57	−18
14	0.68	142	5.4	1.86	36	−22.3	0.08	78	0.53	−24
15	0.70	133	4.7	1.71	25	−21.4	0.09	71	0.47	−37
16	0.74	128	4.4	1.65	18	−20.3	0.10	65	0.45	−50
17	0.76	117	3.7	1.52	6	−19.3	0.11	58	0.42	−65
18	0.75	105	3.1	1.43	−5	−19.0	0.11	49	0.40	−80

Bias = 4.5 V, 30.0 mA.

9.2 Using the common-gate S parameters at 12 GHz (determined in Problem 9.1), find the position of the DR in terms of electrical length from the GaAs FET (Fig. 9.8*b*) in order to maximize the reflection coefficient magnitude at the drain ($|S'_{22}|$). The DR unloaded quality factor and the coupling coefficient to the microstrip line are 3000 and 5, respectively. Plot the S'_{22} on a polar chart from 11 to 13 GHz using the interplotted S parameters for the intermediate frequencies. [*Hint:* Use Eq. (9.41).]

9.3 In the two-port network shown in Fig. 9.6, prove that if oscillation condition ($S'_{11}\Gamma_1 = 1$) is satisfied at the input port, it is automatically satisfied at the output port ($S'_{22}\Gamma_2 = 1$).

9.4 The Agilent GaAs FET AT10600 has $G_0 = 9$ dB with $P_{sat} = 20$ dBm at 12 GHz. Calculate maximum oscillator power obtainable using this device as an oscillator.

9.5 **(a)** The frequency pulling bandwidth of a 10-dBm-power-output oscillator at 12 GHz is measured to be 5 MHz into a 12-dB-return-loss load. Calculate the external quality factor neglecting the nonlinearities.

(b) If the load is changed to 6 dB return loss, find the frequency pulling bandwidth.

(c) The frequency drift of the oscillator is measured to be 4 MHz over −30 to +70°C. Find the minimum injection locking power required to keep the oscillator locked over the temperature range.

9.6 A modulation noise measurement system is found to have an effective gain of 940 and employs a detector with a sensitivity of 100 V/W. When operated as a direct detection system connected with a source producing 1.2 mW, the detector outputs are 0.14 and 0.105 mV in the unbal-

anced and balanced configurations. What is the AM noise of the source expressed as a power ratio?

A carrier suppression filter is now inserted and the system after calibration is found to have an rms sensitivity of 1.25 μW/Hz deviation. If the output from the detector is 0.24 mV with an unbalanced mixer and 0.091 mV with a balanced mixer configuration, what is the phase noise of the source?

10

AMPLIFIERS

Inder Bahl and Edward Griffin

10.1 INTRODUCTION

The introduction of radar during World War II provided the first significant application requiring amplification of microwave signals. In recent times, the wireless communication revolution has provided an explosion of microwave amplification applications. Early microwave amplifiers were the exclusive province of vacuum tube devices such as klystrons [1–3], traveling-wave tube (TWT) amplifiers [2–4], and magnetrons [2, 3]. Today, microwave amplification is dominated by solid state amplifiers except for applications at high output powers ($\gtrsim$100 W). The most common of today's vacuum tube applications is the 900-W microwave oven using 2.45-GHz magnetrons. The microwave oven magnetron, with a manufacturing cost of about $10, has no solid state competition in sight. Likewise, today's $0.50 1-W, 900-MHz transistor cell phone solid state amplifier has no tube competition.

Solid state amplifiers are of two general classes: those based on two-terminal negative-resistance diode devices and those based upon three-terminal transistors. Early solid state amplifiers were dominated by two terminal amplifiers because diodes are typically much easier to fabricate than transistors. Quite an array of two-terminal amplifier designs have been introduced, including parametric amplification (varactor diodes) [5–8], tunneling diodes [7–9], transferred electron (Gunn and limited space charge accumulation (LSA) diodes) [8, 10, 11], and avalanche transit-time diodes (IMPATT, TRAPATT, and BARITT) [8, 12]. Today solid state amplification is dominated by use of three-terminal transistors including MESFETs, psuedomorphic high electron mobility transistors (PHEMTs), and HBTs [13–21]. The switchover to three-terminal devices

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

was largely due to cost. Diodes are typically less expensive to manufacture than transistors, but the associated circuitry to achieve gain from a two-terminal device is much more expensive than that for a three-terminal device. In addition, design of two-terminal amplifiers for stable operation and routine high-yielding manufacturing is exceedingly difficult.

The balance of this chapter will focus on GaAs MESFET amplifiers. Most of the material presented is common to all transistor design (see Chapters 7 and 15 for a description of transistor types and their comparisons). The chapter will begin with a section on low-noise design and conclude with a section on power amplifier design.

10.2 AMPLIFIER CHARACTERIZATION

Although many characteristics must be considered when designing an amplifier, the most important of these are the power gain, noise figure, stability, input and output VSWR, power output, 1-dB compression point, intermodulation performance, and dynamic range. These parameters are briefly described in this section. The power-added efficiency, which is more important for power amplifiers, is treated in Section 10.5.

10.2.1 Power Gain

The power gain or transducer power gain (G_T) is defined as the ratio of the power delivered to the load to the power available from the source to the network. For a two-port network (shown in Fig. 10.1),

$$G_T = \frac{|S_{21}|^2(1-|\Gamma_S|^2)(1-|\Gamma_L|^2)}{|(1-S_{11}\Gamma_S)(1-S_{22}\Gamma_L)-S_{12}S_{21}\Gamma_S\Gamma_L|^2} \tag{10.1}$$

where

$$\Gamma_{L,S} = \frac{Z_{L,S}-Z_0}{Z_{L,S}+Z_0}$$

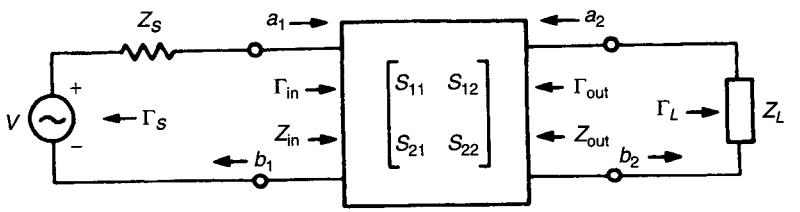


Figure 10.1 Two-port network of a transistor amplifier.

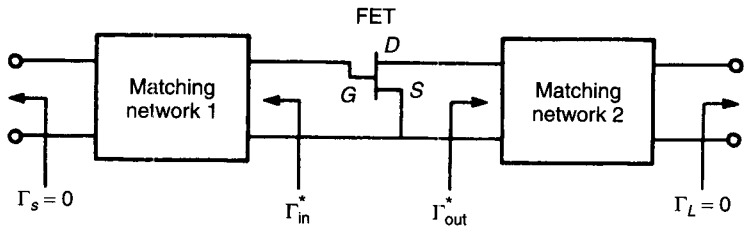


Figure 10.2 FET matched for maximum gain.

The S parameters are measured with usually 50 Ω as the input and output impedance, and arbitrary source impedance Z_S and load impedance Z_L are connected to the network.

For unilateral transducer power gain, the reverse power gain is set to zero (i.e., $|S_{12}|^2 = 0$), and (10.1) becomes

$$G_{TU} = \frac{|S_{21}|^2(1 - |\Gamma_S|^2)(1 - |\Gamma_L|^2)}{|(1 - S_{11}\Gamma_S)(1 - S_{22}\Gamma_L)|^2} \tag{10.2}$$

The maximum unilateral power gain is attained when $\Gamma_S = S_{11}^*$ and $\Gamma_L = S_{22}^*$, that is, when the network is conjugately matched at the input and output ports. Then the maximum unilateral power gain, called the maximum available gain, is given by

$$G_a = G_{TUm} = \frac{|S_{21}|^2}{(1 - |S_{11}|^2)(1 - |S_{22}|^2)} \tag{10.3}$$

Thus the maximum available gain is the product of the transistor transducer power gain $|S_{21}|^2$ between source and load impedances (usually 50 Ω) and the increase in gain due to matching the input port $[(1 - |S_{11}|^2)^{-1}]$ and matching the output port $[(1 - |S_{22}|^2)^{-1}]$. In other words, a single-stage amplifier design consists of (1) designing an input matching network to give $\Gamma_S \simeq 0$ and (2) designing an output matching network that simultaneously gives $\Gamma_L \simeq 0$, as shown in Fig. 10.2. As we will show in Section 10.2.3, the above-mentioned conditions fail if the solid state devices are unstable.

10.2.2 Noise Characterization

In a microwave amplifier, a small output power can be measured, even when there is no input signal. This is the amplifier noise power. The total output noise power consists of amplified input noise entering the amplifier plus the noise power generated in the amplifier itself.

The model of a noisy two-port microwave amplifier is shown in Fig. 10.3. The noise input power can be modeled by a noisy resistor. This noise is caused

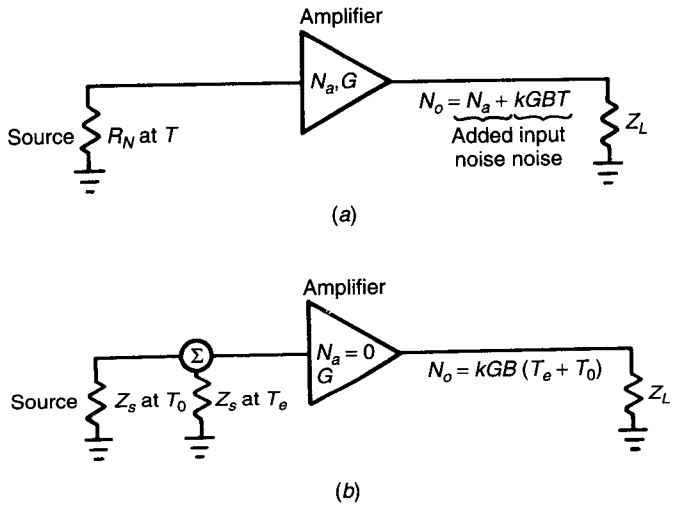


Figure 10.3 Equivalent noise power and noise temperature representation of an amplifier.

by random motion of electrons in the resistor due to thermal agitation and therefore is known as thermal or Johnson noise. The maximum available noise power N_R from R_N is

$$N_R = kTB \tag{10.4}$$

where k is Boltzmann's constant (i.e., $k = 1.38 \times 10^{-23}$ J/K), T is the resistance noise temperature expressed in Kelvin, and B is the noise bandwidth. At $B = 1$ GHz and $T = 290$ K, $kTB = -84$ dBm. Note that the available noise power is independent of the magnitude of the resistor value but, of course, the amount of noise power actually delivered to a load resistor will decrease as the ratio of the source and load resistor values varies from unity.

Equation (10.4) shows that the thermal noise power depends on the bandwidth but not on a given frequency. Such a distribution of noise is called "white noise." Obviously no true white noise sources exist because these sources would put out infinite noise power over an infinite bandwidth. Actually the formula in (10.4) breaks down at frequencies well above the millimeter-wave range, where the noise power drops. Over the microwave and millimeter-wave frequencies, most dissipative elements in electrical circuits are very well characterized as ideal kTB noise power sources.

In addition, most resistors exhibit increased noise at low frequencies. This increase is usually referred to as the $1/f$ or flicker noise. Typically $1/f$ noise does not actually have $1/f$ frequency dependence, but it does increase with decreasing frequency. The frequency at which the $1/f$ power equals kTB is called the $1/f$ knee frequency. Flicker noise is not usually important in the

design of microwave amplifiers as the knee frequency for transistors is below 100 MHz, but it is an important source of phase noise in microwave oscillators.

In microwave transistors, in addition to Johnson and $1/f$ noise, shot noise is generated by random passage of charges in the modulating channels. The shot noise is relatively constant from DC to 100 MHz. Noise in transistors is well treated in the literature [15].

Noise Figure. The noise figure of any linear two-port network can be defined as

$$F = \frac{\text{signal-to-noise ratio at input}}{\text{signal-to-noise ratio at output}} = \frac{\text{available noise power at output}}{\text{gain} \times \text{available noise power at input}} = \frac{N_o}{GkTB} \tag{10.5}$$

where N_o = available noise power at output

G = available gain of network over bandwidth B

If N_a is the noise power added by the amplifier, then

$$F = \frac{GkTB + N_a}{GkTB} = 1 + \frac{N_a}{GkTB} \tag{10.6}$$

An amplifier that contributes no noise to the circuit has $F = 1$. Now consider an amplifier having n stages cascaded in series with gain values $G_1, G_2, \ldots, G_n$ and noise figure values $F_1, F_2, \ldots, F_n$; then the total noise figure of an n-stage amplifier is given by

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_n - 1}{G_1 G_2 \cdots G_{n-1}} \tag{10.7}$$

When $F_1 = F_2 = \cdots = F$ and $G_1 = G_2 = \cdots = G$, as $n \to \infty$, the overall noise figure obtained is called the noise measure, F_M, and is given by

$$F_M = \frac{F - 1/G}{1 - 1/G} \tag{10.8}$$

For most applications, noise measure is a much better figure of merit for an amplifier than noise figure because noise measure accounts for the noise contribution of succeeding amplifier stages. An amplifier with low gain does not guarantee good system noise performance, no matter how low the amplifier noise figure may be.

Since Johnson noise is invariant against the resistance and proportional to absolute temperature, it is used to characterize the noise power in units of

temperature. Consider the amplifier shown in Fig. 10.3*b*. Typically, the source and load resistors are 50 Ω, but the following definitions apply to arbitrary impedances. The output noise power is given by

$$N_o = Gk(T_0 + T_e)B \tag{10.9}$$

where T_0 is the room temperature (usually $T_0 = 290$ K) and T_e is the equivalent noise temperature of the amplifier with $N_a = 0$. From (10.5) and (10.6), when the source impedance is at T_0,

$$F = 1 + \frac{T_e}{T_0} \tag{10.10}$$

The noise figure is usually expressed in decibels as

$$\text{NF} = 10 \log(F) = 10 \log\left(1 + \frac{T_e}{T_0}\right) \tag{10.11}$$

Example. Consider an example of two amplifiers cascaded in series. If the gain and noise figures are 9 and 2 dB for the first amplifier and 10 and 3 dB for the second amplifier, calculate the total noise figure of the cascaded amplifier. What happens to the noise figure if the gain of the first amplifier is 15 dB or higher?

The overall noise figure is given by (10.7), that is,

$$F = F_1 + \frac{F_2 - 1}{G_1} \tag{10.12}$$

Here

$$\text{NF}_1 = 2 \text{ dB} \qquad F_1 = 1.58$$

$$\text{NF}_2 = 3 \text{ dB} \qquad F_2 = 2$$

$$G_1 = 9 \text{ dB} = 7.94$$

$$G_2 = 10 \text{ dB} = 10$$

$$F = 1.58 + \frac{2 - 1}{7.94} = 1.7$$

or

$$\text{NF} = 10 \log(1.7) = 2.3 \text{ dB}$$

when

$$G_1 = 15 \text{ dB} = 31.6 \qquad F = 1.58 + \frac{2 - 1}{31.6} = 1.6 \qquad \text{NF} = 2.07 \text{ dB}$$

Thus, as the gain of first amplifier, G_1, gets larger, the relative importance of the noise contribution for the second amplifier becomes less and less.

Example. Consider an amplifier preceded by an attenuator when they are cascaded in series. If attenuation of the attenuator is L, the gain and noise figures of the amplifier are G_A and NF_A, and all are expressed in decibels, what is the overall noise figure of the assembly?

This example is very similar to the previous example, where

$$F_1 = 10^{L/10} \qquad F_2 = 10^{\mathrm{NF}_A/10}$$
$$G_1 = 10^{-L/10} \qquad G_2 = 10^{G_A/10}$$

From (10.12)

$$F = 10^{L/10} + \frac{10^{\mathrm{NF}_A/10} - 1}{10^{-L/10}}$$
$$F = 10^{L/10} \times 10^{\mathrm{NF}_A/10}$$
$$\mathrm{NF} = 10 \log F = L + \mathrm{NF}_A$$

In other words, a matched attenuator at room temperature connected in front of an amplifier increases the noise figure of the amplifier by an amount equal to its loss.

Noise Bandwidth. The total noise power from an amplifier is given by

$$\mathrm{NT} = \int_0^\infty T_A(\omega) G_A(\omega)\, d\omega \tag{10.13a}$$

where $T_A(\omega)$ and $G_A(\omega)$ are the noise temperature and power gain of the amplifier, respectively. It is often useful to treat the amplifier as having fixed T_A and G_A over a specific noise bandwidth and no gain elsewhere where the noise bandwidth (NBW) is chosen to make the total noise power correct, that is,

$$\mathrm{NBW}\, kT_A G_A = k \int_0^\infty T_A(\omega) G_A(\omega)\, d\omega$$

or

$$\mathrm{NBW} = \frac{\int_0^\infty T_A(\omega) G_A(\omega)\, d\omega}{T_A G_A} \tag{10.13b}$$

The range of ω is typically limited by other components in the system or by the gain response of the amplifier.

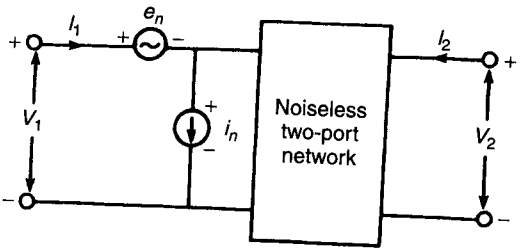


Figure 10.4 Noise equivalent circuit of two-port network.

Optimum Noise Match. In general, any noisy two-port network may be represented by a noise voltage and a noise current source connected at the input of a noiseless two-port network, as shown in Fig. 10.4. If the circuit has a dominant voltage noise, using a high source impedance will minimize the transmission of the noise signal, but if the current noise is dominant, connecting a low source impedance will minimize the transmission of the noise signal. When both noise sources are present, a minimum noise figure of the circuit results in a specific source admittance or impedance, known as the optimum source admittance. Circles of constant noise figure on the input admittance or impedance plane can be plotted using a Smith chart. How the noise figure increases from the minimum value is described by the following relation:

$$F = F_{\min} + \frac{R_n}{G_S}|Y_S - Y_O|^2$$

or

$$F = F_{\min} + \frac{R_n}{G_S}[(G_S - G_O)^2 + (B_S - B_O)^2] \tag{10.14}$$

where F = noise figure
Y_S = source admittance = $G_S + jB_S$
$F_{\min}$ = minimum noise figure
Y_O = optimum source admittance that gives minimum noise figure
= $G_O + jB_O$
R_n = equivalent noise resistance

The lower the equivalent noise resistance, the less the sensitivity of noise figure increase for a nonoptimum source. In the above equations, the device output has been assumed conjugately matched. However, if the device has poor isolation between its input and output terminals, the effect of load pull on the noise figure of the circuit must be considered.

We can express Y_S and Y_O in terms of the reflection coefficient Γ_S and Γ_O, and the resultant relation becomes

$$F = F_{\min} + \frac{4\bar{R}_n|\Gamma_S - \Gamma_O|^2}{(1 - |\Gamma_S|^2)|1 + \Gamma_O|^2} \tag{10.15}$$

where $\bar{R}_n = R_n/Z_0$ (Z_0 is usually 50 Ω). The quantities $F_{\min}$, R_n, and Γ_O are known as the noise parameters and are given by the manufacturer of the transistor or can be determined experimentally.

10.2.3 Stability

Any amplifier with power gain can be made to oscillate by applying external positive feedback (e.g., a high-gain MMIC amplifier in a plastic or ceramic package with poor isolation). At microwave frequencies unavoidable parasitics are often sufficient to cause oscillations if care is not taken in the design and fabrication of the amplifier. Any very abrupt change in the DC parameters of the amplifier (e.g., output power with no input power, circuits that are very sensitive to their surroundings) are typical indications of oscillations. Oscillations may occur at frequencies that do not propagate out the amplifier because they are filtered, blocked by bias capacitors, or below waveguide cutoff frequencies or at frequencies to which the test equipment is insensitive. It is not unusual for microwave amplifiers to oscillate anywhere between 1 MHz and 30 GHz or higher.

The best single test for oscillations involves using a combination of a broadband spectrum analyzer and sweep oscillator. By watching the broadband output spectrum as the input signal is swept, the absence of both fundamental and intermodulation oscillation signals is usually sufficient to assure stable operation. It must be emphasized that any variable change such as impedance levels, supply voltages, temperature, light intensity aging, and radiation may quench or create oscillations. Testing often must be done over a broad range of parameters if confidence is to be obtained that the amplifier is stable under those conditions. Fortunately design techniques can be applied to build in confidence of stability.

Stability against oscillations can be examined by using two-port S parameters [13–21]. The parameters S_{12} and S_{21} form a feedback loop that, depending on the source and load impedances, may support oscillations. In an ideal amplifier, S_{12} would be zero and the amplifier would be unconditionally stable. If $S_{12} \neq 0$, input reflection coefficient Γ_{in} with arbitrary Z_L and output reflection coefficient Γ_{out} with arbitrary Z_S can be expressed as

$$\Gamma_{\text{in}} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \tag{10.16}$$

$$\Gamma_{\text{out}} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S} \tag{10.17}$$

If the circuit is unconditionally stable, any source or load may be connected to the input or output of the circuit without oscillations. In terms of S parameters

of the FET, unconditional stability is assured if the following inequalities are simultaneously satisfied:

$$|S_{11}| < 1 \qquad |S_{22}| < 1$$
$$|\Gamma_{in}| < 1 \qquad |\Gamma_{out}| < 1$$

For $|\Gamma_S| < 1$ and $|\Gamma_L| < 1$, these conditions lead to the requirement

$$K = \frac{1 + |S_{11}S_{22} - S_{12}S_{21}|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{12}||S_{21}|} > 1 \tag{10.18}$$

The significance of the stability factor K is that a FET is unconditionally stable for all passive source and load terminations when $K > 1$.

If we set $|\Gamma_{in}|$ and $|\Gamma_{out}|$ equal to unity, a boundary is established beyond which the device is unstable. Each condition will give a solution of a circle on a complex reflection plane whose radius (r) and center (c) are given by

$$r_S = \frac{|S_{12}S_{21}|}{||S_{11}|^2 - |D|^2|} \qquad c_S = \frac{(S_{11} - DS_{22}^*)^*}{|S_{11}|^2 - |D^2|} \quad \text{(input)} \tag{10.19}$$

$$r_L = \frac{|S_{12}S_{21}|}{||S_{22}|^2 - |D|^2|} \qquad c_L = \frac{(S_{22} - DS_{11}^*)^*}{|S_{22}|^2 - |D|^2} \quad \text{(output)} \tag{10.20}$$

where $D = S_{11}S_{22} - S_{12}S_{21}$, S and L denote source and load, and the origin of the Smith chart is at $\Gamma_{in} = \Gamma_{out} = 0$. Figure 10.5 shows typical examples of the input plane of unconditionally stable and conditionally stable networks. The shaded area represents the area of the input plane in which instability occurs. If the circuit is potentially unstable ($K < 1$), the source and load impedances should be chosen so they will not fall into the unstable region (e.g., shaded) due to device parameter changes, fabrication variation, and changes in temperature. Under such conditions, the amplifier is said to be conditionally stable and will not oscillate. Stability analysis must be carried out from DC to above the frequency where the active devices have power gain.

If the circuit is unconditionally stable ($K > 1$), the conditions required to obtain maximum power gain are $\Gamma_{in} = \Gamma_S^*$ and $\Gamma_{out} = \Gamma_L^*$. Solving for simultaneous conjugate match, the reflection coefficients to be matched are denoted by Γ_{SM} and Γ_{LM} and are given by [14, 16]

$$\Gamma_{SM} = \frac{B_1 \pm \sqrt{B_1^2 - 4|C_1|^2}}{2C_1} \tag{10.21}$$

$$\Gamma_{LM} = \frac{B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}}{2C_2} \tag{10.22}$$

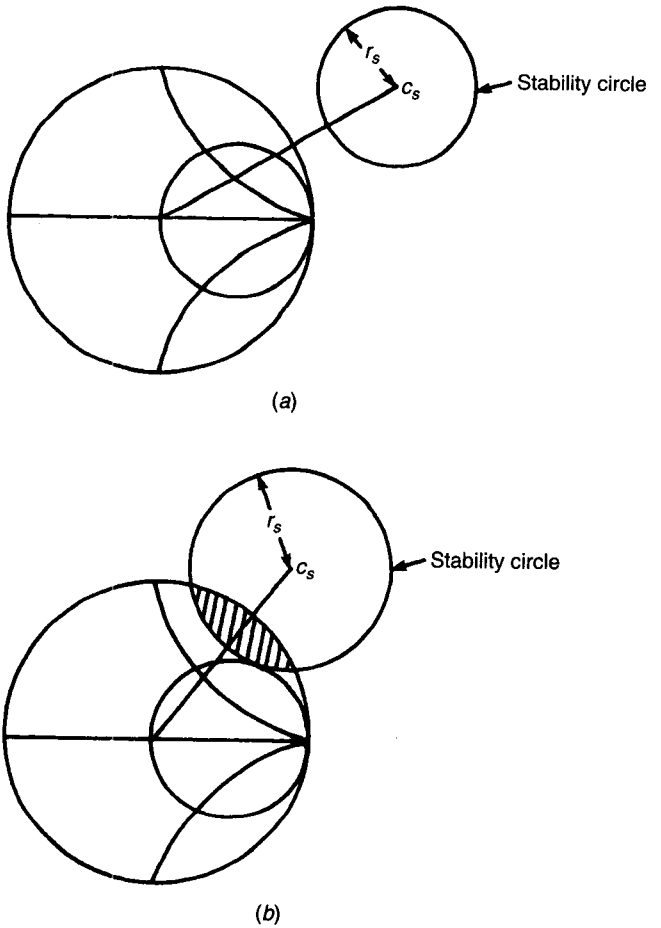


Figure 10.5 Examples of (*a*) unconditional stability and (*b*) conditional stability for $|S_{11}| < 1$.

where

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |D|^2$$

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |D|^2$$

$$C_1 = S_{11} - DS_{22}^* \qquad C_2 = S_{22} - DS_{11}^*$$

If $|B_1/2C_1| > 1$ and $B_1 > 0$, the solution with the minus sign produces $|\Gamma_{SM}| < 1$ and the solution with the plus sign produces $|\Gamma_{SM}| > 1$. If $|B_1/2C_1| > 1$ with $B_1 < 0$, the solution with the plus sign produces $|\Gamma_{SM}| < 1$ and the solution with the minus sign produces $|\Gamma_{SM}| > 1$. Similar considerations apply to Γ_{LM}.

The value of the matched gain (MG) is

$$\text{MG} = \left|\frac{S_{21}}{S_{12}}\right|(K - \sqrt{K^2 - 1}) \tag{10.23}$$

If the circuit is potentially unstable ($K < 1$), the maximum stable gain (MSG) is obtained by substituting $K = 1$ in (10.23):

$$\text{MSG} = \left|\frac{S_{21}}{S_{12}}\right| \tag{10.24}$$

Once a stability problem has been identified, it is usually not difficult to correct it, for example, by adding extra bias capacitance, putting resistances in the bias lines, or adding inductor, capacitor, resistor, or transmission-line terminations for problem frequencies. In most microwave system applications conditional stability is usually acceptable in the frequency band where component impedances are defined, but unconditional stability may be required out of band where there are usually no system specifications.

Ferrite isolators are often used to protect amplifiers from unknown impedances. Care must be taken, since stability problems may arise far outside the isolator frequency band where the isolator is neither matched nor unilateral.

The two-port stability analysis just described is known as even mode and applicable only to the sensitivity of the two ports to external impedances, not to internal oscillations and low-power linear amplifiers. For example, a multistage low-noise amplifier might oscillate even though its overall K factor is greater than unity. The stability analysis must be performed on all internal two-ports having active devices and on the complete circuit. A series of stable two-amplifier stages will be stable, but if additional feedback is introduced, the new circuit must also be analyzed.

The stability analysis of nonlinear power amplifiers is very complex. The oscillations in RF/microwave power amplifiers may be classified into five categories: even mode, odd mode, parametric, spurious parametric, and low frequency. These are discussed elsewhere [21].

10.2.4 Nonlinear Behavior

An amplifier is called linear when the output power increases linearly with the input power. The ratio of these two powers is the power gain G. As input power increases, the amplifier transfer function becomes nonlinear; that is, the output power is lower than predicted by the small-signal gain. This nonlinear behavior in amplifiers introduces distortion in the amplified signal. The output power at which the gain has dropped by 1 dB below the linear gain is called the *1-dB compression point*, $P_{1\text{ dB}}$. Typically, the gain will drop rapidly for powers above $P_{1\text{ dB}}$, reaching a maximum or fully saturated output power within 1–4 dB above $P_{1\text{ dB}}$, depending on the number of stages and load match conditions.

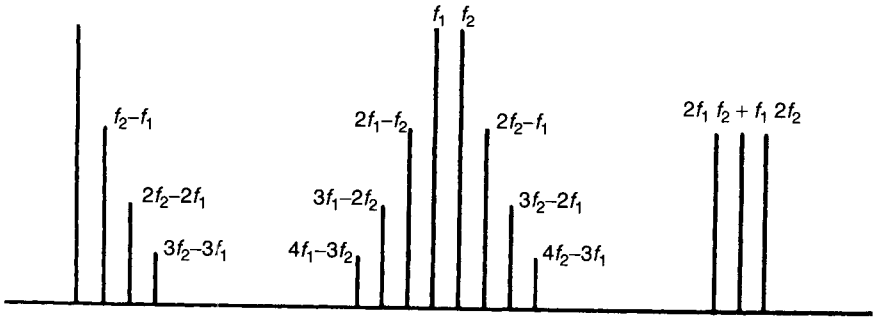


Figure 10.6 Schematic spectrum showing two signals at frequencies f_1 and f_2 and their IM products.

There are a number of different ways to measure the nonlinearity behavior of an amplifier. The simplest method is the measurement of the 1-dB compression power level $P_{1\,\mathrm{dB}}$. Another method that is very popular uses two closely spaced signals 5–10 MHz apart. When two signals at frequencies f_1 and f_2 are incident on an amplifier, the output of the amplifier contains these two signals as well as intermodulation (IM) products at frequencies $mf_1 + nf_2$, where $m + n$ is known as the order of the IM product. Figure 10.6 illustrates the IM products spectrum. The relative magnitude of the IM products depends on the details of how the amplifier saturates; however, third-order products that are closest in frequency to the fundamental tones usually dominate at moderate saturation levels, although second-order products can be important in multioctave amplifiers.

Figure 10.7 shows the transfer characteristics of a typical solid state amplifier having 10 dB signal gain. With proper filtering, one of the fundamental frequencies and the distortion products are measured separately. Output powers of the second- and third-order IM products are also shown in Fig. 10.7. Since second- and third-order IM products correspond to square and cubic nonlinearities, respectively, the output power for these products increases by 2 and 3 dB/dB at low power levels. Since the third-order terms normally dominate up to the point where distortion is very severe, IM distortion is often characterized by the third-order intercept, $P_{3\mathrm{rd}}$, as shown in Fig. 10.7. Intercept $P_{3\mathrm{rd}}$ is the point where the power in the third-order product and fundamental tone are equal if the amplifier is assumed to be linear. The third-order intercept power is typically 10 dB above the $P_{1\,\mathrm{dB}}$ and is a very useful parameter for calculating low-level IM effects.

Amplitude-modulated (AM) signals are strongly attenuated by saturated amplifiers, but frequency-modulated (FM) signals are often passed through power amplifiers running at the point of maximum efficiency, which is usually above $P_{1\,\mathrm{dB}}$. Although an FM signal is less subject to distortion by gain compression than an AM signal, both are subject to distortion due to harmonic generation in nonlinear devices.

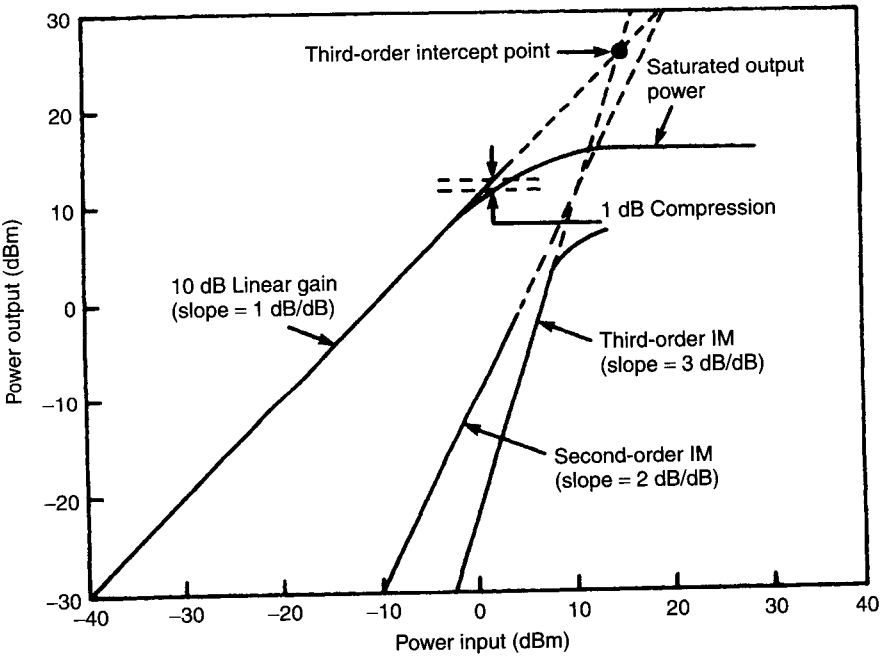


Figure 10.7 Variation of output power and IM products with input power for nonlinear amplifier.

The simplest form of distortion occurs when a sinusoidal input is converted to a square-wave output by power supply clipping. Under these conditions, the output signal will contain not only the fundamental frequency f but also the odd harmonics $(3f, 5f, \ldots)$ of the input signal. These harmonics will be more or less visible at the output, depending on the frequency response of the amplifier. If, for example, the output includes a low-pass filter that passes f but is cut off for $3f$, the output from a single frequency input may be compressed but still be purely sinusoidal. If clipping is unsymmetrical, even as well as odd harmonics may be generated, but if filtering is applied to attenuate the $2f$ signal, the output will remain purely sinusoidal.

Adjacent-channel power ratio (ACPR) is a commonly used figure of merit to evaluate the intermodulation distortion performance of RF power amplifiers designed for code division multiple access (CDMA) wireless communication systems. The ACPR is a measure of the spectral regrowth that appears in the signal sidebands and is analogous to IM3/IM5 for an analog RF amplifier.

The ACPR is defined as [22]

$$\text{ACPR} = \frac{\text{power spectral density in main channel 1}}{\text{power spectral density in offset channel 2 or 3}} \tag{10.25}$$

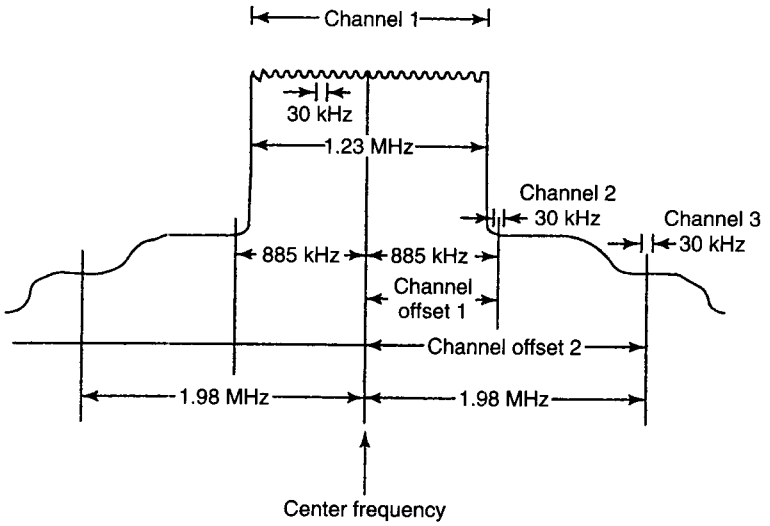


Figure 10.8 CDMA adjacent-channel power measurement frequency spectrum. (After Slovick [22]. Reprinted with permission of Microwave J.)

Here offset frequencies and measurement bandwidths vary with the system application. For example, as shown in Fig. 10.8, for CDMA power amplifiers, the two offsets are at 885 kHz and 1.98 MHz away from the channel 1 center frequency and the measurement bandwidth specified is 30 kHz. The ACPR requirements are −42 and −54 dBc at channel offsets 2 and 3, respectively.

10.2.5 Dynamic Range

The range of an input signal that can be detected by a receiver without much distortion is called the dynamic range (DR). The dynamic range of an amplifier is defined as the ratio of the 1-dB compressed power output to the amplified minimum detectable signal.

The output noise of a two-port device with noise figure F can be written from (10.5) as

$$N_o = FGkTB \tag{10.26}$$

If the minimum detectable input signal is X(dB) above the noise floor, then

$$P_{\text{in}}^{\min} = N_o + X \quad \text{(dB)} \tag{10.27a}$$

$$P_{\text{out}}^{\min} = P_{\text{in}}^{\min} + G \quad \text{(dB)} \tag{10.27b}$$

Dynamic range is defined as

$$\mathrm{DR} = P_{1\,\mathrm{dB}} - P_{\mathrm{out}}^{\mathrm{min}} \tag{10.28}$$

If a typical value of X is 3 dB, then from (10.25)–(10.28)

$$\mathrm{DR} = P_{1\,\mathrm{dB}} + 171 - 10\log(B) - F - G \quad (\mathrm{dB}) \tag{10.29}$$

where $P_{1\,\mathrm{dB}}$ is in decibels above 1 mW (0 dBm).

Example. Determine the dynamic range of a low-noise amplifier with a gain of 30 dB, a noise figure of 2 dB, a 1-dB compression point of 15 dBm, and a noise bandwidth of 1 GHz.

From (10.29),

$$\mathrm{DR} = 15 + 171 - 90 - 2 - 30 = 64\ \mathrm{dB}$$

The spurious free dynamic range (DR_f) of an amplifier is defined as the ratio of the fundamental signal power output to the third-order intermodulation product power output when the third-order intermodulation product is equal to the minimum detectable output signal. The spurious free dynamic range in decibels is given by [14]

$$\mathrm{DR}_f = \tfrac{2}{3}[P_{3\mathrm{rd}} - P_{\mathrm{out}}^{\mathrm{min}}] \tag{10.30}$$

10.3 BIASING NETWORKS

The application of FET gate and drain DC bias voltage is an important part of an amplifier design. The design considerations for biasing circuits are efficiency, noise, oscillation suppression, single source power supply, RF choking, and impedance matching. Five practical biasing configurations are given in Table 10.1. The circuits (a), (b), and (e) require two power supplies of opposite polarity. Since in these configurations the source is grounded with minimum possible source inductance, they provide maximum gain. In these circuits, gate voltage is applied before the drain voltage. Biasing circuits (c) and (d) with source resistors are widely used for small- and medium-power applications and require only one power supply. As the supply voltage is applied, the gate is simultaneously reverse biased with respect to the source by the series resistor R_s. The value of R_s is selected based on drain–source current I_{DS} and the operating bias. A 50-pF bypass capacitor is usually sufficient for RF grounding and transient protection. These circuits have lower amplifier efficiency due to DC power dissipation in source resistors. Thus for power amplifiers where $I_{DS} > 500$ mA,

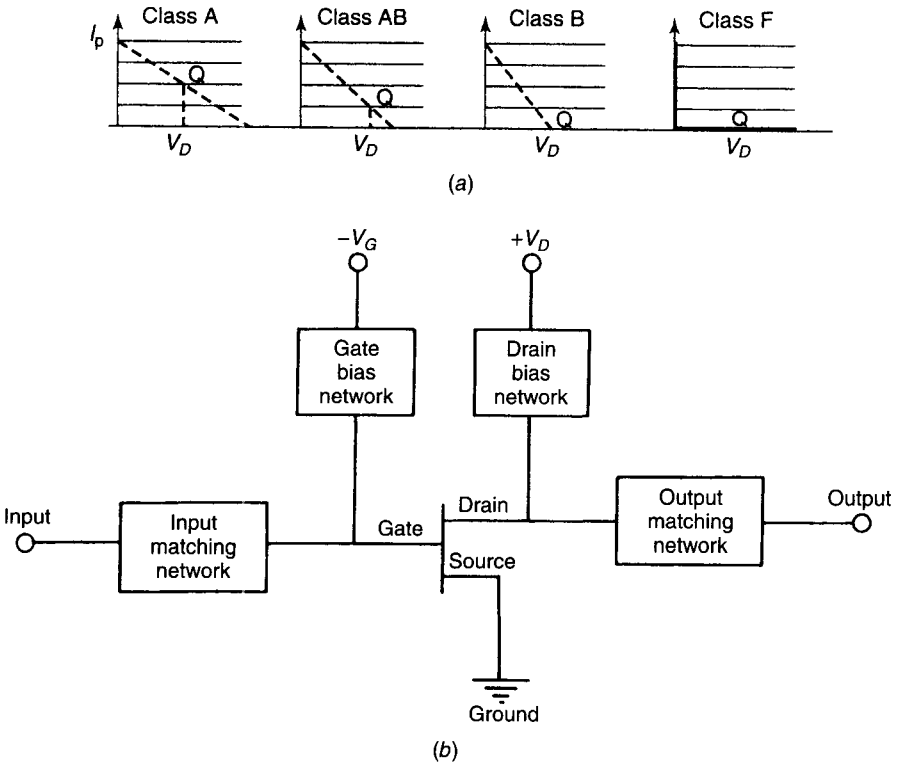


Figure 10.9 (*a*) Voltage-current relationships for various classes of operation for ideal FETs. (*b*) Single-stage common-source MESFET amplifier block diagram.

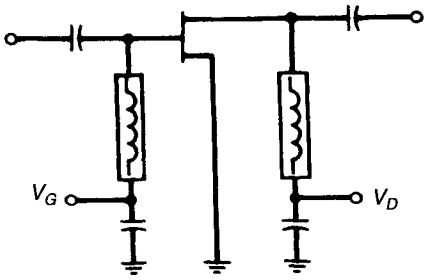
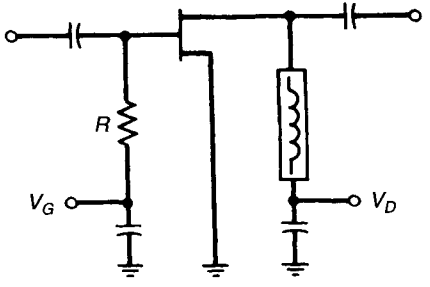
dual-polarity supplies are recommended with a sequencing circuit to bias the gate first before applying the drain bias. This sequencing is required because V_D applied at I_{DSS} (drain–source saturation current) may burn the device.

At RF/microwave frequencies power amplifers are defined to operate in class A, B, AB, C, D, E, and F. These classes are implemented by selecting bias conditions of the active device. Figure 10.9*a* shows voltage-current relationships for various classes of operation for ideal FETs.

10.4 SMALL-SIGNAL AMPLIFIER DESIGN

In this section we discuss aspects of the design of low-noise, maximum-gain GaAs FET amplifiers. Although there are many FET configurations that may be used for amplification, we concentrate on common-source amplifiers, as shown in Fig. 10.9*b*. The common-source configuration has high gain, low noise, and the best amplifier stability. Low-noise amplifiers are designed to

Table 10.1 Various Bias Schemes

Biasing Configuration	Typical Bias Voltages	Amplifier Characteristics	Other Comments
 (a)	$V_D = 3$ V, $V_G = -1$ V	Low noise, high gain, high power, high efficiency	Biasing networks are part of matching; insensitive to bias current
 (b)	$V_D = 3$ V, $V_G = -1$ V	Moderately low noise, high gain, high power, high efficiency	Biasing network is part of matching; insensitive to bias current; high value of R provides higher isolation between gate and power supply

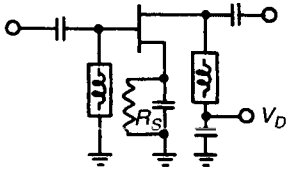
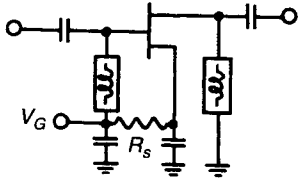
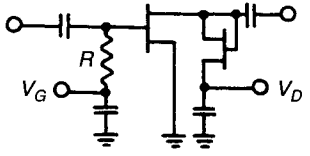
 (c)	$V_D = 4$ V, $I_{DS}R_S = 1$ V	Moderately low noise, high gain, medium power, low efficiency	R_S provides automatic transient protection; sensitive to bias current
 (d)	$V_G = -4$ V, $I_{DS}R_S = 1$ V	Moderately low noise, high gain, medium power, low efficiency	R_S provides automatic transient protection; sensitive to bias current
 (e)	$V_D = 5$ V, $V_G = -1$ V	Moderate noise, high gain, medium power, low efficiency	Broadband at lower frequencies; sensitive to bias voltage

Table 10.2 Typical Low Noise Performance of Transistors

Transistor	Gate Length (μm)	Noise Figure (dB)	Associated Gain (dB)	Measured Frequency (GHz)	Transistor Type	Company
TGF1350-SCC	0.3	1.5	11	10	MESFET	TriQuint
TGF4350-EPU	0.3	0.8	13	10	pHEMT	TriQuint
NFET300	0.4	0.7	9.7	10	MESFET	M/A-COM
ATF-36077	—	0.5	12.0	12	pHEMT	Agilent
MGF4319E	—	0.55	11.5	12	pHEMT	Mitsubishi
NE71300	0.3	1.6	9.5	12	MESFET	NEC
NE27200	0.2	0.45	12.5	12	HEMT	NEC
FHX45X	0.15	0.55	12	12	pHEMT	Fujitsu

increase signal levels while introducing a minimum amount of signal-to-noise degradation.

The design of an amplifier for minimum noise figure involves several considerations. First and foremost, one must select the proper FET. For low-noise applications, the FETs are biased at relatively low current levels (typically 20% of I_{DSS}, 3 V drain–source). At these bias levels, the FET's power-handling capability is much less than that at higher bias (typically $P_{1\,dB} = 10$–12 dBm for a 300-μm-wide gate FET). The input and output matching networks, when designed for best noise performance, match the output (FET drain) to the load but introduce considerable mismatch at the input (FET gate), with the result that the gain is usually considerably less than the maximum available gain. Finally, for best noise (also known as optimum noise) performance, it is critical to minimize circuit loss on the input side, since loss adds directly to the amplifier noise figure.

10.4.1 FET Selection

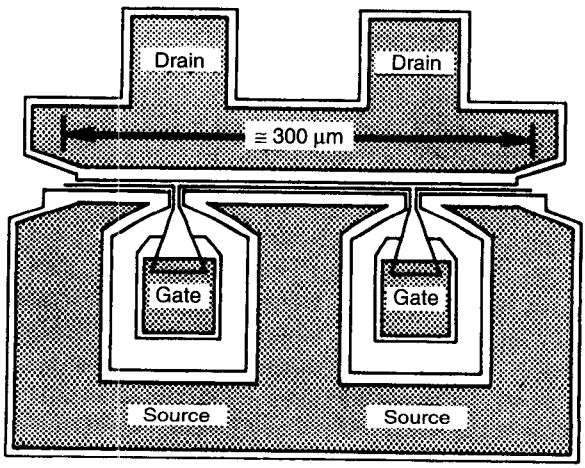
The selection of a FET for a particular application is straightforward but sometimes tedious. One simply simulates the performance of the available FETs for the particular application and then selects the optimal FET based on the relevant trade-offs, which may include

- cost,
- reliability, and
- electrical performance.

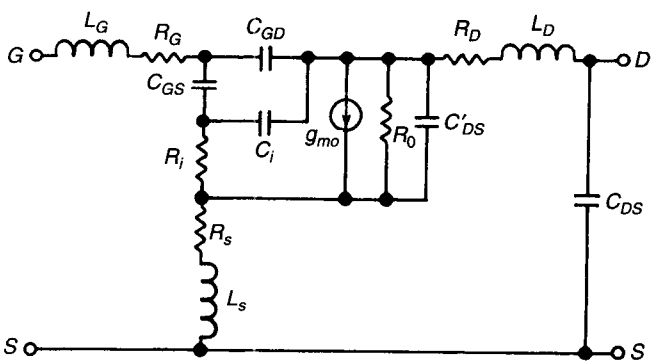
Table 10.2 lists a few of the typical low-noise FETs available commercially as of late 2001. The most important dimensions of a FET are the FET gate

length and total gate periphery. Typical gate lengths are 0.3–1.0 μm. The FET gate is usually made up of several fingers. A typical FET might have four fingers, each 75 μm wide, making a total gate periphery of 300 μm, as shown in Fig. 10.10.

The electrical parameters available in vendor catalogs include S parameters at two bias conditions (generally at 10 mA and $\frac{1}{2}I_{DSS}$), optimum reflection coefficient for minimum noise figure, equivalent noise resistance R_n, minimum noise figure, associated gain, and maximum available gain at various frequencies. Electrical parameters for a 300-μm-gatewidth FET are shown in Tables 10.3 and 10.4. The maximum frequency f_T (at which the FET has current gain unity) is a useful figure of merit for a FET because it defines an approximate



(a)



Component	Value
R_G	2.0 Ω
R_i	1.0 Ω
R_s	2.5 Ω
R_0	300 Ω
R_D	4.0 Ω
L_G	0.02 nH
L_S	0.05 nH
L_D	0.02 nH
C_{GS}	0.4 pF
C_{GD}	0.02 pF
C_i	0.001 pF
C'_{DS}	0.02 pF
C_{DS}	0.03 pF
g_{mo}	35 mS

(b)

Figure 10.10 (*a*) Four-finger MESFET configuration. (*b*) Equivalent circuit.

Table 10.3 S Parameters of FET

Frequency (GHz)	$\lvert S_{11}\rvert$	$\angle S_{11}$	$\lvert S_{12}\rvert$	$\angle S_{12}$	$\lvert S_{21}\rvert$	$\angle S_{21}$	$\lvert S_{22}\rvert$	$\angle S_{22}$	K	MSG
2.0	0.98	−30	0.02	75	2.63	160	0.73	−7	0.18	21.2
4.0	0.93	−56	0.04	62	2.36	142	0.71	−13	0.35	17.7
6.0	0.87	−78	0.05	53	2.05	128	0.69	−19	0.53	16.1
8.0	0.83	−96	0.06	46	1.77	116	0.68	−24	0.72	14.7
10.0	0.79	−111	0.07	42	1.54	106	0.67	−28	0.91	13.4
12.0	0.77	−122	0.07	40	1.36	98	0.67	−32	1.10	11.1
14.0	0.75	−132	0.07	40	1.20	92	0.66	−37	1.29	9.2
16.0	0.74	−140	0.07	41	1.08	86	0.66	−41	1.48	7.9
18.0	0.73	−148	0.07	43	0.98	81	0.66	−45	1.67	6.8
20.0	0.72	−154	0.07	46	0.90	76	0.66	−49	1.83	5.9

Note: $V_D = 3$ V, $I_{DS} = 10$ mA.

maximum frequency ($f_T = g_m/2\pi C_{GS}$). The value of transconductance g_m is generally available or can be obtained from S parameters along with C_{GS}. In addition to transconductance and C_{GS} at microwave frequencies, a FET exhibits many parasitic elements that reduce electrical performance and must be accounted for in the design. An equivalent lumped-element model for a FET as shown in Fig. 10.10 is very handy and can be obtained by computer optimization to replicate the measured S parameters.

Figure 10.11*a* shows the minimum noise figure for various kinds of microwave and millimeter-wave transistors, whereas Fig. 10.11*b* depicts state-of-the-art power performance for single-chip MMIC amplifiers at microwave and millimeter-wave frequencies.

10.4.2 Narrow-Band Low-Noise Design

The design of a single-stage narrow-band low-noise amplifier can be carried out step by step as follows:

Table 10.4 Characteristics of Low-Noise FET

Frequency (GHz)	Minimum NF (dB)	G_A (dB)	Γ_{opt}	R_n	Z_S^* (Ω)	Z_L^* (Ω)
4	0.73	14.8	$0.66\angle 44$	13.1	$58.9 + j95.1$	$113.8 + j115.9$
8	1.44	11.1	$0.54\angle 86$	12.7	$29.0 + j44.6$	$86.0 + j114.7$
12	2.11	8.7	$0.55\angle 118$	13.1	$19.1 + j26.8$	$63.9 + j107.6$
16	2.75	6.8	$0.60\angle 140$	12.6	$14.2 + j17.1$	$47.7 + j99.0$
20	3.33	5.4	$0.64\angle 155$	13.6	$11.5 + j10.7$	$37.3 + j89.7$

Note: $V_D = 3$ V, $I_{DS} = 10$ mA.

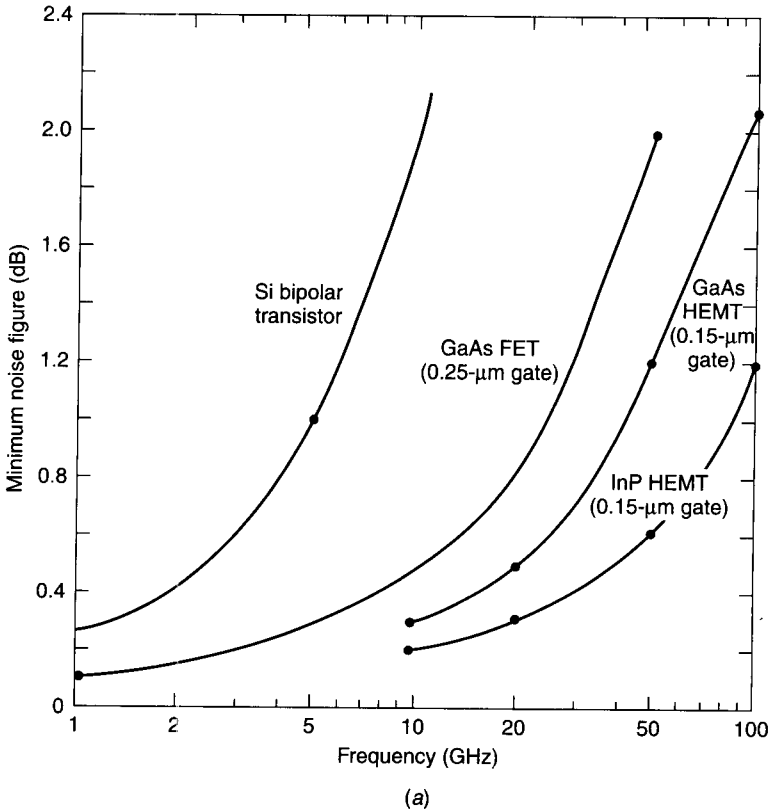
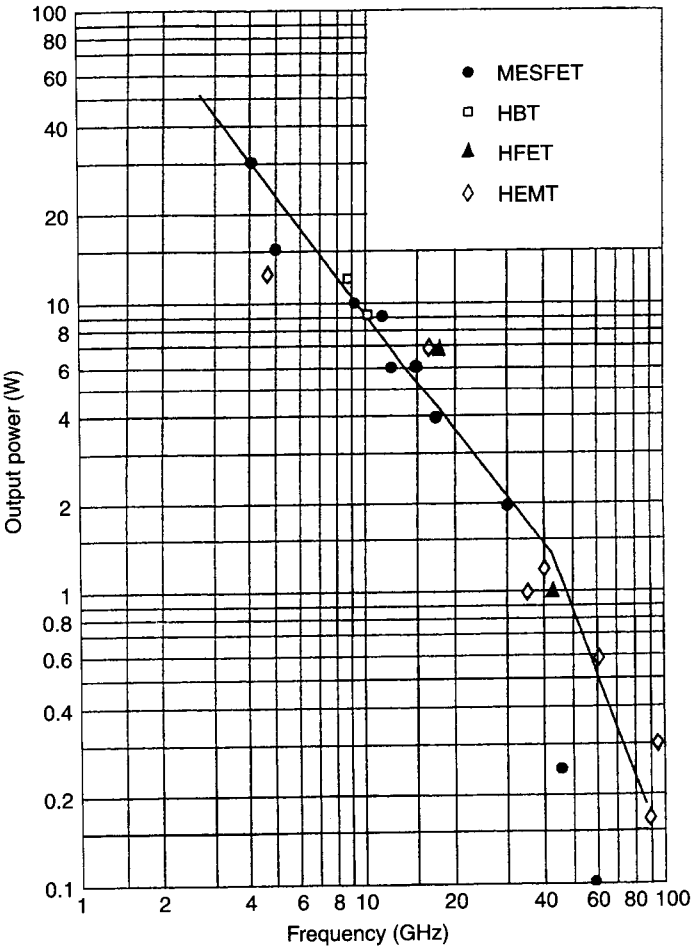


Figure 10.11 (*a*) Minimum noise figure versus frequency for various microwave devices. (*b*) Performance status of single-chip power MMIC amplifiers using MESFET, HFET, HEMT, and HBT technologies.

1. Select a GaAs FET with a noise figure lower and a gain higher than the design value.
2. Calculate its stability factor K.
3. If $K > 1$, select suitable input and output matching networks that include biasing circuitry and complete their design.
4. If $K < 1$, plot the regions of instability on the reflection planes and select matching networks that avoid the unstable regions.
5. Calculate the amplifier performance using analytical methods or CAD tools. Check stability of the amplifier in the band as well as outside the band.
6. Work out the realization of the amplifier.

Now we describe an example to illustrate the procedure. Design a low-noise single-stage amplifier using a microstrip on 0.25-mm-thick alumina substrate ($\epsilon_r = 9.9$) with the following specifications:



(*b*)

Figure 10.11 *(Continued)*

Frequency	12 GHz,
Bandwidth	5%,
Gain (min)	8 dB,
Noise figure (max)	2.5 dB,
Output VSWR (max)	1.2:1.

The FET listed in Tables 10.3 and 10.4 satisfies the requirements and $K > 1$. A simple amplifier configuration consisting of two matching elements at the input and two at the output, as shown in Fig. 10.12, can be selected. At 12 GHz, the input matching circuit must transform 50 Ω to 19.1 + j26.8 Ω, and

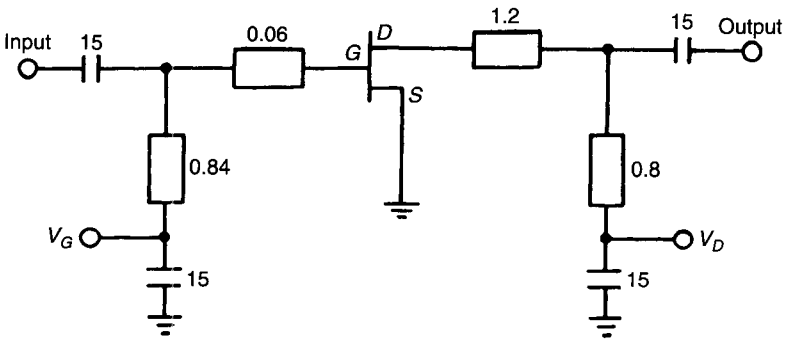


Figure 10.12 Complete amplifier schematic. Microstrip lines width is 0.11 mm. All dimensions are in millimeters. Capacitances are in picofarads.

the output matching circuit must transform 50 Ω to $63.9 + j107.6$ Ω. These matching circuit elements can be determined using either a Smith chart or the analytical methods described in Chapter 4. Here 70 Ω as the characteristic impedance of microstrip lines has been used. Physical dimensions for matching elements on alumina substrate ($\epsilon_r = 9.9$, $h = 0.25$ mm) are given in Fig. 10.12. In order to neglect the effect of DC blocking and bypass capacitors, their values are chosen so that their reactance values are less than 1 Ω at the lowest operating frequency. The calculated performance of this amplifier is shown in Fig. 10.13.

In multistage low-noise amplifiers, the input port is matched for optimum noise figure, and the interstages and output port are matched for maximum and flat gain. In this case, usually all FETs except in the first stage are biased approximately at half I_{DSS}.

10.4.3 Maximum-Gain Amplifier Design

The design procedure for a maximum-gain amplifier is the same as described for the low-noise one, except that all the stages are matched for maximum gain. The following example illustrates the design of a maximum-gain amplifier.

Example. Design a maximum-gain amplifier at 14 GHz with a gain of 8.5 dB and VSWR better than 1.5.

The FET listed in Table 10.3 satisfies the requirements, and $K = 1.29$. Equations (10.21) and (10.22) can be used to compute the simultaneously conjugately matched source and load impedances for maximum gain as follows:

$$C_1 = 0.451\underline{/42^\circ} \qquad C_2 = 0.325\underline{/134^\circ}$$

$$B_1 = 0.9119 \qquad B_2 = 0.6634$$

$$\text{MG} = 9.2 \text{ dB}$$

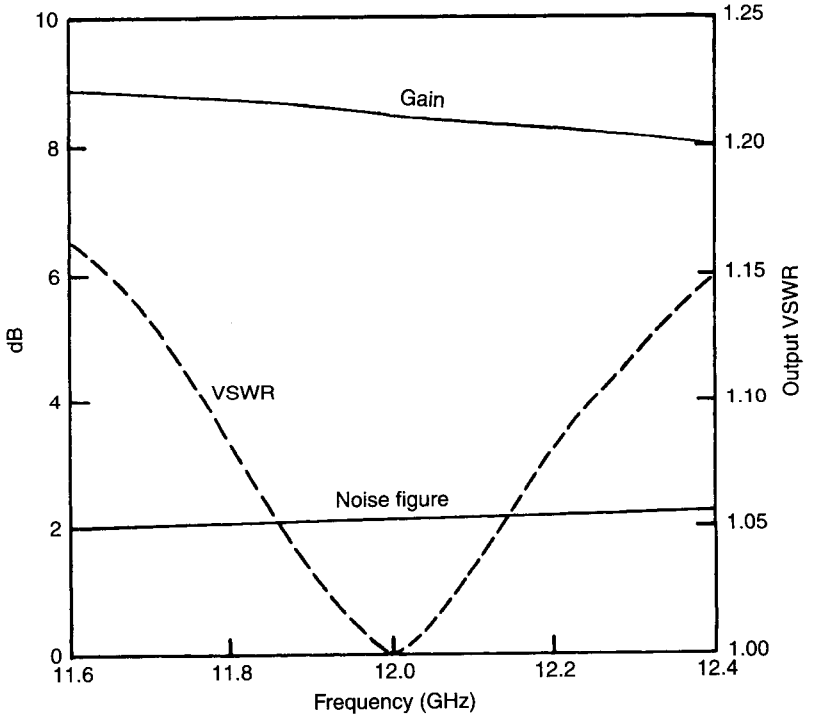


Figure 10.13 Calculated response of low-noise amplifier.

$$\Gamma_{SM} = 0.8624/\underline{136°} \qquad \Gamma_{LM} = 0.8148/\underline{46°}$$

$$Z_{SM} = 4.2 + j19.1\ \Omega \qquad Z_{LM} = 31.6 + j110.2\ \Omega$$

$$Y_{SM} = 0.0111 - j0.0499\ \text{S} \qquad Y_{LM} = 0.0024 - j0.0084\ \text{S}$$

A simple amplifier configuration consisting of two matching elements at the input and two at the output, as shown in Fig. 10.14, can be selected. At 14 GHz, the input matching circuit must transform 50 Ω to $4.2 + j19.1$ Ω, and the output matching circuit must transform 50 Ω to $31.6 + j111.2$ Ω. The line lengths for matching circuit elements using 50-Ω lines were determined using a Smith chart. The physical dimensions for microstrip sections on 0.25-mm-thick alumina substrate ($t = 6$ μm and $\epsilon_r = 9.9$) are shown in Fig. 10.14.

10.4.4 Broadband Amplifiers

The broadband amplifier design is carried out by considering the power gain rolloff of the FET with frequency (usually 6 dB/octave), the gain–bandwidth limitations of the input and output of the FET, and the overall amplifier stability versus frequency. The design of a broadband, common-source amplifier

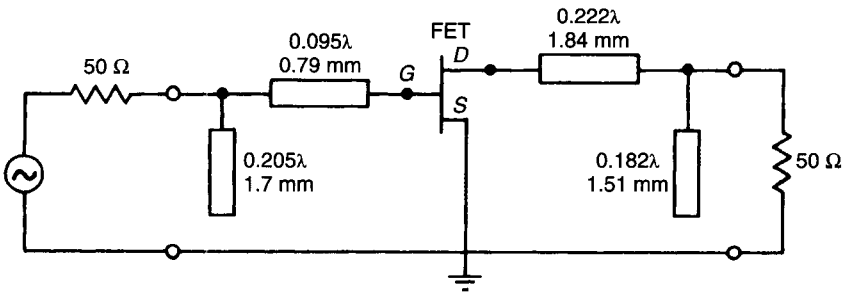


Figure 10.14 Amplifier schematic. Microstrip width is 0.24 mm and $\lambda = 8.28$ mm at 14 GHz.

requires a compromise between several competing requirements. The input impedance to the gate of a GaAs FET can be approximated conceptually as a series RC circuit composed of the gate resistance, the source resistance, the intrinsic resistance, and the gate capacitance. The Q of the circuit is $1/\omega RC$. For the purpose of broadband matching to the gate, the lower the Q, the easier it will be to design and realize conjugate matching circuits. There are fundamental limits to how well matched an RC circuit can be over a given bandwidth, but practical considerations are more important. If the matching circuit becomes too elaborate, it cannot be built accurately, and circuit losses mount rapidly with circuit complexity. The requirements of low FET device noise are in opposition to reduction of the input Q.

There are various techniques used to realize broadband amplifiers. Five of them are shown in Fig. 10.15. In balanced configurations [23] a single-ended reflective match amplifier (as described in the previous section) is designed, and a matched pair is used to realize a balanced configuration using two broadband Lange couplers. The single-ended amplifier is usually mismatched for flat gain, low noise figure, and good stability. The reflections from the amplifier are terminated in 50 Ω, which usually guarantees stability. If one stage fails, the overall gain drops about 6 dB, which may provide useful fault tolerance for some applications. In feedback amplifiers [24], a series RL feedback is used between the drain and gate of the FET. This configuration improves input match, output match, and stability by lowering gain at lower frequencies. The active matching method uses common-gate FET configuration at the input and common-drain FET configuration at the output of the main FET for matching to 50 Ω. This configuration is useful for monolithic amplifiers working up to 10 GHz [25]. The resistive matching technique uses resistors as a part of matching networks and is very similar to reflective matching methods. The distributed amplifier configuration (also commonly referred to as a traveling-wave amplifier) provides the unique capability of adding device transconductance without adding device parasitic capacitance. In this configuration, FETs with series inductors behave like an effective low-pass transmission line. By terminating these lines with resistive loads, the unwanted signals are dissipated while the desired signals are added in phase at the output of the amplifier. This results

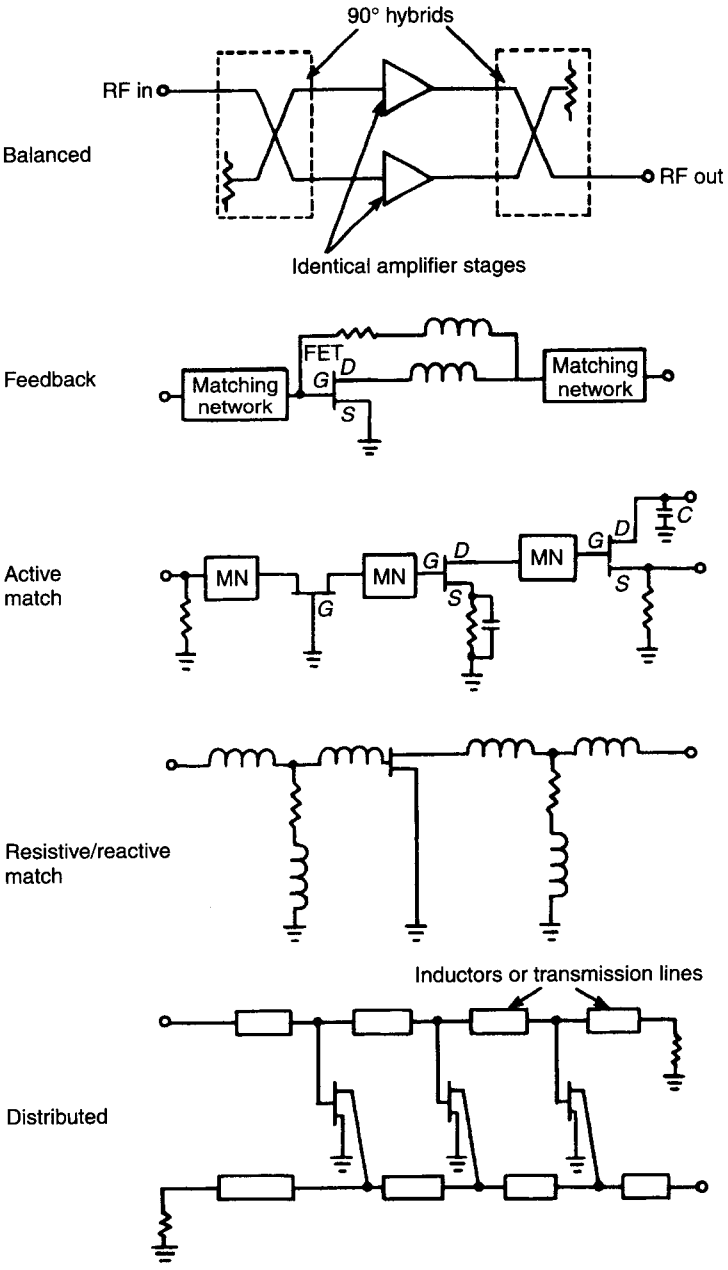


Figure 10.15 Broadband amplifier configurations.

in an excellent gain-bandwidth product with flat gain and low VSWR. This technique has been successfully used in monolithic amplifiers [26, 27] as well as in hybrid amplifiers [28].

A qualitative comparison of techniques for broadbanding of amplifiers previously described is given in Table 10.5, and a quantitative comparison has been described by Niclas [29]. Broadband amplifiers can be designed using the impedance matching techniques described in Chapter 4 or CAD tools. The CAD tools play a very important role in designing broadband amplifiers.

The rapid increase of data traffic requires large-capacity optical communication systems. Such systems demand data rates as high as 40 Gb/s. The most speed limited component in such systems is a preamplifier in the receiver that requires a low noise figure, flat-gain response, and ultralarge bandwidth (DC to 40 GHz). Various amplifier topologies including traveling wave and device technologies including MESFET, PHEMT, and HBT are being pursued to develop high-speed preamplifiers in hybrid or monolithic form.

A six-cell monolithic distributed amplifier was designed using M/A-COM's multifunction self-aligned gale (MSAG) process [21] for 20-Gb/s optical communication applications. The schematic of the amplifier is shown in Fig. 10.16. The 1000-pF, 150-nF, and 0.01-μF capacitors were connected externally to extend the frequency range to 500 kHz. Figure 10.17 shows the measured gain and input and output return losses. The amplifier has typically 10 dB gain and a maximum VSWR 2:1 in the 500 kHz–20 GHz band. The gain flatness was within 2 dB.

10.5 POWER AMPLIFIERS

Requirements for power amplifiers vary drastically from one application to another. Basic requirements for such amplifiers are high gain, higher linearity, high power-added efficiency (PAE), high reliability, small size, and low cost. Usually communication applications require linear operation, while for radar applications high PAE is of prime importance. Personal communication systems working in the 800 MHz–2.5 GHz range use different digital modulation and access schemes. They require high efficiency and linear power amplifiers for hand-held as well as for base-station applications.

For power amplifiers, the input signal level is high, and consequently the output current is either in the cutoff or saturation region during a portion of the input signal cycle. This leads to the classification of power amplifiers into three basic modes of operation: Class A, class B, and class C. These classes are implemented by suitably biasing the active devices. When the output current flows for a full period of the input voltage cycle, the amplifier is operated in the class A mode. If the output current flows for a half period of the input voltage cycle, the amplifier is designated a class B amplifier and is biased at cutoff. If the output current flows for less than a half period of the input voltage cycle, the amplifier is called a class C amplifier and is biased below cutoff. The class

Table 10.5 Comparison of Techniques Used for Broadbanding of Amplifiers

Balanced Circuit	Feedback	Active Matching	Resistive Matching	Distributed Approach
Good up to a 2-octave bandwidth	Good up to many-octave bandwidth	Good up to many-octave bandwidth	Good up to many-octave bandwidth	Good up to many-octave bandwidth
Requires matched FET pairs and quadrature couplers	Requires large-g_m FETs and uses negative as well as positive feedback	Requires a common-gate FET at the input and a common-drain FET at the output	Requires large-g_m FETs	Requires many small FETs
Size is relatively large	Size is relatively small	Size is moderate	Moderate size	Size is moderate
Provides very good impedance match	Provides good impedance match	Provides good impedance match	Provides poor impedance match	Provides good impedance match
Has low noise figure	Has high noise figure	Has moderate noise figure	Has high noise figure	Has moderate noise figure
Effect of fabrication tolerance is small	Effect of fabrication tolerance is moderate	Effect of fabrication tolerance is moderate	Effect of fabrication tolerance is small	Effect of fabrication tolerance is moderate
Cascading of two or more gain modules is easy	Modular approach is easy	Modular approach is easy	Modular approach is not easy	Modular approach is easy

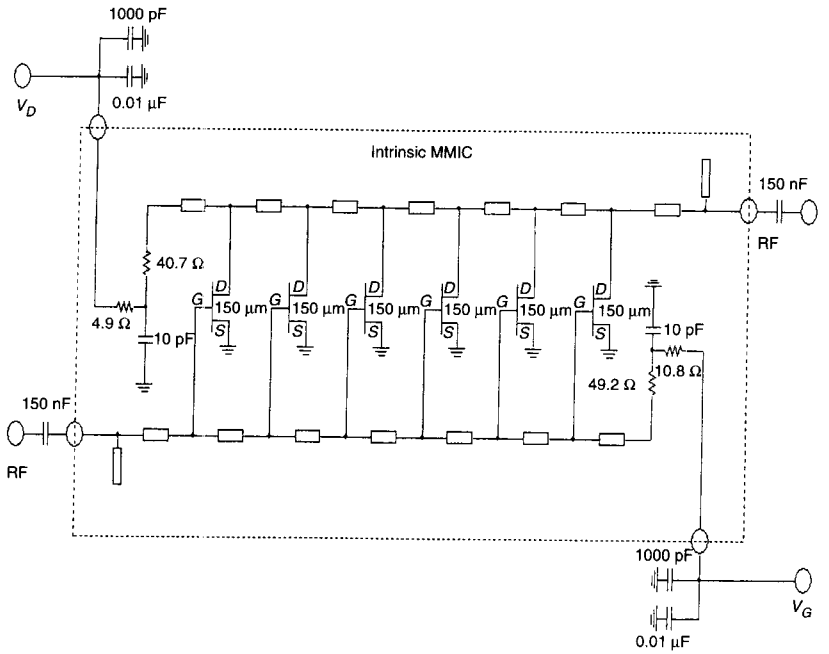


Figure 10.16 Schematic of distributed amplifier.

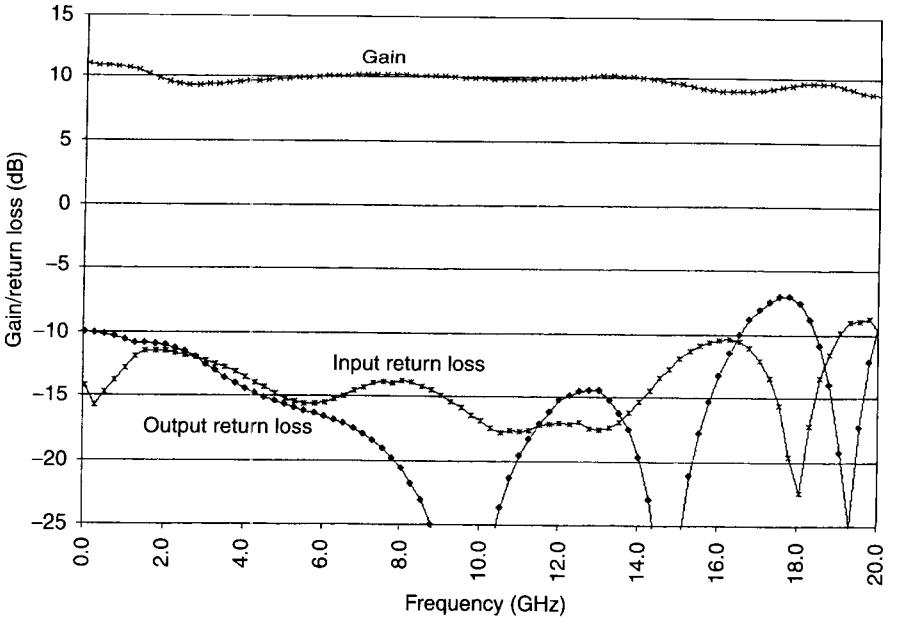


Figure 10.17 Calculated performance of 2–19-GHz distributed amplifier.

A amplifier has the best linearity among these three types, whereas the class C amplifier has the best efficiency.

In the basic modes of operation, the harmonics at the device output are resistively terminated in the load. There are two other popular classes of power amplifiers in which harmonics are reactively terminated to reduce the power dissipation. These are known as classes E and F. In class E, the active device works as a switch and has a high-Q tuned circuit at the output of the device to provide a designed reactive load at the fundamental frequency and open termination at the second and third harmonics. In the class F amplifier configuration, the reduced power dissipation is achieved by employing some sort of impedance matching technique (such as resonant circuits) to terminate harmonic frequencies in desired loads (short circuits for even harmonics and open circuits for odd harmonics). Again, the active device is basically operated as a switch and the theoretical efficiency approaches 100%.

As discussed above, RF power amplifiers can be designed to operate under a variety of PAE and linearity conditions with the various amplifier types identified with a class designation. In practice, all amplifier classes operate at reduced efficiencies due to inherent parasitic losses and nonideal operating conditions. Thus, the selection of power amplifier technology is very important to meet system requirements in terms of output power, DC power consumption, linearity conditions (modulation schemes), frequency range, size, weight, and cost. In any class, there is a trade-off between the efficiency and linearity; higher efficiency means poorer linearity.

Power-added efficiency is defined as

$$\text{PAE} = \frac{\text{output signal power} - \text{input signal power}}{\text{DC power}} = \frac{P_{\text{out}} - P_{\text{in}}}{P_{\text{DC}}}$$

$$= \frac{P_{\text{out}}}{P_{\text{DC}}}\left(1 - \frac{1}{G}\right) = \eta_d\left(1 - \frac{1}{G}\right) \tag{10.31}$$

where η_d is the drain efficiency. Power-added efficiency is a measure of how efficiently the DC bias power is converted into output RF power. Thus, the PAE is an important power amplifier parameter for several reasons. If the amplifier is battery powered, the battery life will be approximately proportional to the PAE. Equally important, the power not converted to microwave energy is dissipated as heat, which must be removed to preserve the reliability of the amplifier. For high-efficiency amplifiers, single-stage gain on the order of 10 dB or higher is required.

Designing a power amplifier for systems that will have many closely spaced signals is very challenging. Intermodulation products typically fall on top of adjacent signals, potentially wrecking system performance. The simplest and most widely used approach to countering intermodulation is to run the power amplifier at power levels well below $P_{1\,\text{dB}}$. There are many amplifiers in the field today running about 10 dB below $P_{1\,\text{dB}}$. This works, but at the cost of

having a much larger amplifier than one might expect, running at low PAE. Backing down 10 dB might reduce a typical power amplifier from 50 to 5% PAE.

Power amplifiers in communication satellites often counter intermodulation by breaking the frequency band into narrow subbands allowing the filtering of the intermodulation products. The design trade-off of introducing much more hardware complexity for improved PAE is often chosen because prime power and cooling are major concerns for satellite operation. Cost constraints for terrestrial systems usually do not support channelization schemes.

Power amplifiers can be designed for improved intermodulation performance (typically 3–5 dB) by using either nonlinear modeling or load-pull data to select optimum input match and output load impedance for the design of output power level for best intermodulation performance. An improvement of 3 dB may mean cutting prime power requirements by 50%.

There is a great deal of activity underway in support of the telecommunication industry to develop alternative solutions to obtain high efficiency with excellent intermodulation performance. High linearity in power amplifiers can also be achieved either by using highly linear devices such as pulse/spike doped [30, 31] or by using predistortion techniques [32–49] at the circuit level or cancellation technique such as feedforward [50–52]. The intermodulation-distortion characteristics of GaAs FETs and their relation to doping profile have been studied, and it was found that pulsed doped FETs have 5–20 dB better P_{3rd} or IP3 performance compared to conventional FETs. Predistortion techniques are more practical at the circuit level whereas cancellation techniques are more complicated.

The nonlinearity in an amplifier results in nonlinear behavior in amplitude (AM–AM) and phase (AM–PM) responses. Thus the distortion introduced in an amplifier can be explained in terms of AM–AM and AM–PM characteristics and strongly depends upon the class of operation. In predistortion techniques, the amplitude and phase variation with input power are compensated for by adding extra elements in the circuit at the input of the power amplifier. Large number of linearization techniques have been used to linearize a nonlinear amplifier. These predistortion techniques include using dual-gate FET [32], series feedback [33], IF feedback [34], active feedback [35, 36], harmonic feedback [37–39], series [40, 41] and shunt diode [42], passive FET [43], cascode [44], push-pull [45], second-order intermodulation component feedforwarding [46], current dumping [47], branch FET [48], and Doherty techniques [49]. Several of these techniques have shown drastic improvement in IP2/IP3 and ACPR. In most methods, accurate knowledge of the amplitude and phase of the input signal is required along with accurate nonlinear device models predicting AM–AM and AM–PM characteristics. A predistortion technique is simple in concept but not so simple to implement in a stable reproducible way.

Distortion cancellation techniques are used at the system level and are more complex. An overview of these techniques can be found in the literature [50–52], feedforward being the most popular. These techniques are very com-

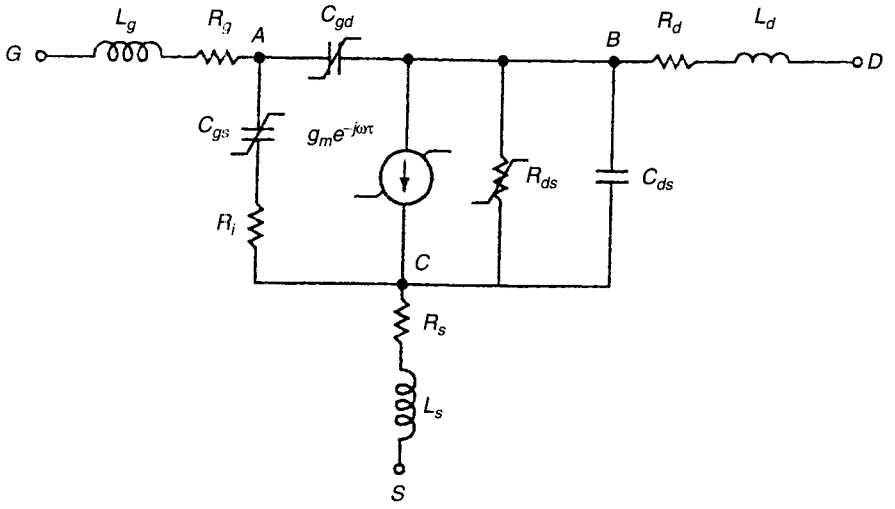
plicated for handset applications but are commonly used for base-station applications. The basic idea in a feedforward design is that a relatively small comparator amplifier is used to inject a signal at the output to cancel differences between the output signal and the input. Because it is not within the scope of this book to go into such detail, readers are referred to the above-mentioned references.

10.5.1 Device Models: Linear and Nonlinear

The application of modern CAD tools offers an improved approach to reducing the design time for power amplifiers. As the sophistication and accuracy of these tools improve, significant reductions in design cycle time can be realized and "first-pass" design success can be achieved for ICs. Accurate models for FETs, HFETs, HEMTs, and HBTs are an essential part of these tools. The most commonly used method of developing models for active devices is by measuring their DC characteristics and S parameters. This modeling approach gives quick and accurate results, although they are generally limited to just the devices measured. Equivalent-circuit (EC) model parameter values are extracted by computer correlation to the measured DC and S parameter data. The accuracy of the measurement-based models depends on the accuracy of the measurement systems, the calibration techniques, and the calibration standards. On-wafer measurements using high-frequency probes provide accurate, quick, nondestructive, and repeatable results up to millimeter-wave frequencies. Measurement-based models fall in two groups: linear (active devices for linear operation) and nonlinear (active devices for nonlinear operation).

Linear Model. In current-measurement-based linear modeling, devices are electrically characterized by measuring DC and RF parameters at the operating bias conditions. A lumped-element equivalent-circuit model that describes the frequency-dependent electrical characteristics is chosen and the model parameter values are extracted to replicate the measured S parameter. The model parameter extraction is generally based on statistical data with average and standard deviation values, which will help in centering designs for high yield.

An EC model for a MESFET is shown in Fig. 10.18*a* and its parameter values are shown in Fig. 10.18*b*. The parameters C_{gs}, C_{gd}, g_m, and R_{ds} are a strong function of device bias conditions. At given bias conditions, this model describes basic linear operation of a FET and the model reproduces the small-signal RF terminal characteristics of the device with good accuracy. The model is widely used to extrapolate S parameters to frequencies for which experimental data are not available and can be scaled to different sizes of the same device. The main disadvantages of the equivalent-circuit model are difficulty in scaling to different physical structures, frequency independence of circuit elements, no time dependence feature, and the inherent limitation to linear circuits.

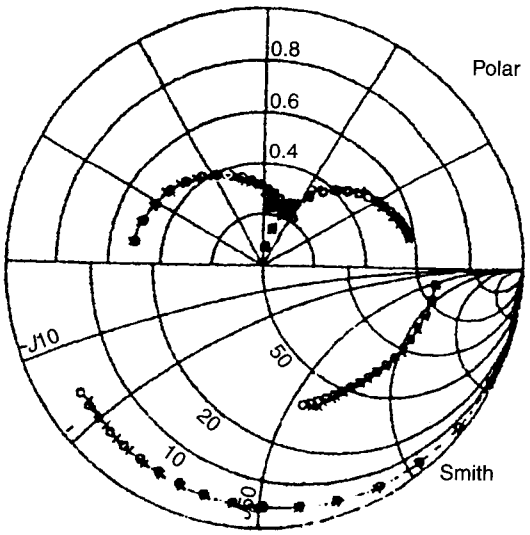


(*a*)

V_{ds} = 2.5 V Bias at 50% of I_{dss}

Measured X Computed O

S_{11}	
S_{12}	Radius: 0.20
S_{21}	Radius: 5.00
S_{22}	
1.000	18.00 GHz
E_{11}	2.20%
E_{12}	3.54%
E_{21}	10.18%
E_{22}	2.10%
F_t	18.25 GHz
$g_{m/C}$	15.57 GHz
$g_{m,\,extr}$	31.28 mS
g_m	34.80 mS
C_{gs}	332.02 fF
C_{gd}	24.49 fF
C_{ds}	91.55 fF
T_t	2.00 ps
R_g	0.20 Ω
R_{gs}	42.21 kΩ
R_{in}	1.13 Ω
R_s	3.24 Ω
R_{ds}	226.17 Ω
R_d	5.44 Ω
L_g	72.3 pH
L_s	0.8 pH
L_d	50.0 pH



(*b*)

Figure 10.18 Equivalent-circuit model and typical small-signal model values for 300-μm power FET biased at V_{ds} = 2.5 V, I_{ds} = 50% I_{dss}; R_{gs} is across C_{gs}, not shown in (*a*).

Nonlinear Model. Many nonlinear equivalent-circuit models have been reported in the literature [18, 53–57]. All of these models have the same basic configuration as shown in Fig. 10.18*a* plus the drain current generator. The EC model parameters and common-source DC or pulsed *I–V* curves that are in qualitative agreement with experimental data are obtained. Significant differences occur, however, in the quantitative behavior of the models, both in comparison to each other and in comparison with experimental data. The differences in the various models are the expressions used to characterize the drain current generator and gate–source and gate–drain capacitances. Some of the most commonly referred models are Curtice, Curtice–Ettenberg, Stratz, Materka, and TriQuint Own Model (TOM).

The measurement-based models need to include terminal voltage-dependent equivalent-circuit elements such as I_{ds} (g_m, R_{ds}), C_{gs}, and C_{gd}, as shown in Fig. 10.18*a*, to predict accurately the nonlinear behavior of the active device. Commonly used representation of the nonlinear MESFET model is given by

$$I_{ds} = (A_0 + A_1 V_1 + A_2 V_1^2 + A_3 V_1^3) \tanh(\alpha V_{ds}) \tag{10.32}$$

where

$$V_1 = V_{gs}[1 + \beta(V_{ds0} - V_{ds})] \tag{10.33}$$

and

$$C_{gs} = C_{gs0} f(V_{gs}, V_{gd}) \tag{10.34a}$$

$$C_{gd} = C_{gd0} g(V_{gs}, V_{gd}) \tag{10.34b}$$

Here, V_{gs}, V_{gd}, and V_{ds} are the terminal voltages between the gate–source, gate–drain, and drain–source of the device, respectively. The source of the FET is normally grounded. The A_i (where $i = 0, 1, 2, 3$) coefficients and constants α, β, and V_{ds0} are evaluated using measured DC or pulsed *I–V* data. The quantities C_{gs0} and C_{gd0} are extracted from the measured *S* parameters at the operating DC bias conditions while f and g, which are functions of both V_{gs} and V_{gd}, are determined from measured *S* parameters over a large range of DC bias conditions to cover the full range of device operation. Inclusion of the tanh function to describe drain current is the generally accepted technique used to extend operation to the saturation region.

Basically there are three steps in the development of nonlinear equivalent-circuit models:

- Extract coefficients for I_{ds} to match with measured *I–V* data. Important data are near the knee of the curves and break down near pinchoff.
- Measure *S* parameters, extract small-signal model values, and derive coefficients for gate–source and gate–drain capacitances to describe dependence of the model on gate and drain voltages.

- Validate the model by comparing measured and simulated data with 50 Ω input and output for $P_{1\,dB}$ compression point and power levels for other harmonics. Simulations are generally carried out using harmonic balance analysis.

The main advantage of the equivalent-circuit models is the ease with which they can be integrated into RF circuit simulators. For linear operation (i.e., small-signal) the interface is direct since the entire device and circuit model are simulated in the frequency domain. For nonlinear applications the device models are formulated in the time domain and are interfaced with the frequency-domain linear circuit simulators by means of the harmonic balance method [58, 59]. The RF performance obtained from these simulators can be satisfactory to good for a well-defined circuit, especially for mildly nonlinear applications such as class A power amplifiers not operating in hard saturation. The large-signal equivalent-circuit models generally do not scale well with varying operational conditions such as frequency or bias. As the circuit becomes increasingly nonlinear, simulator performance degrades.

The main disadvantage of the equivalent-circuit model is inherent inaccuracy resulting from simplifications in the model formulation, such as neglect of domain capacitance and the interdependence of the nonlinear elements. In the actual device all of the nonlinear elements are interdependent. For example, in a MESFET it is not possible to change the device transconductance without also changing elements such as the gate–source capacitance. Perhaps the most significant limitation of the equivalent-circuit models, however, is the need to experimentally characterize the devices that are to be used. The devices must be designed, fabricated, and characterized before the CAD models can be defined. A simple change in any design parameter (such as gate width or channel impurity concentration) requires a complete recharacterization since scaling techniques are difficult to apply. This limits the designer's flexibility in obtaining optimum performance of ICs where tailoring the device design for special applications would be desirable.

10.5.2 Load-Line Modelling

The electrical device parameters available in vendor catalogues or foundry libraries include S parameters or equivalent-circuit models at recommended bias conditions, nonlinear model parameters, optimum input and output impedances for maximum output power or PAE, and associated gain at nominal frequencies. Data for best linearity design are not available. Small-signal S parameters/EC models and nonlinear models are required to calculate the stability and electrical performance of an amplifier. Source-pull and load-pull data are essential to accurately characterize power devices for optimum amplifier design. Power amplifier design based on load line [20, 60] is briefly described below.

The power characteristics of a GaAs MESFET amplifier are elementary

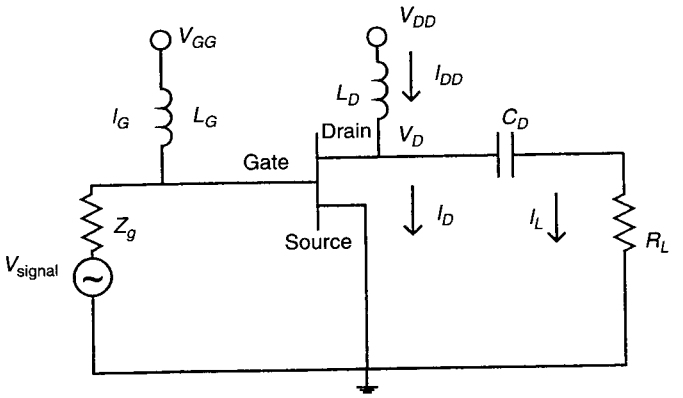


Figure 10.19 Schematic for common-source amplifier.

consequences of voltage and current clipping, not esoteric nonlinear device behavior. Power amplifier design based upon voltage and current waveform limits is called load-line design. Load-line design has a couple of key advantages as compared to nonlinear simulation. First, since load-line simulation is much faster than nonlinear simulation, it is better suited to the initial design phase of topology selection and first optimization. For the same reason, it is useful for circuit thermal simulation and manufacturing tolerance analysis. Second, although load-line design is approximate, it is never far off. If load-line simulation shows a problem, there is a problem. Load-line design is applicable to power design using any device that has reasonably sharp, nondestructive current-limiting operation at a fixed bias voltage.

Consider the common-source amplifier shown in Fig. 10.19. First, we will treat the FET as an ideal voltage-controlled current source that can sink a maximum drain current I_{peak} when the gate voltage is high (i.e., channel is completely open) and pinches off to zero drain current when the gate bias is low (pinch-off condition). Further, we will assume that the signal voltage to the gate is sufficiently large that the FET drain current is effectively a square wave between zero current and $I_{\max}$ at the input frequency ω. As we will see shortly, $I_{\max}$ may be less than I_{peak} depending on drain voltage behavior. The gate and drain bias inductors (L_G, L_D) are assumed to be ideal chokes ($\omega L_G \gg Z_g$, $\omega L_D \gg R_L$) and the drain capacitor C_D is assumed to be an ideal blocking capacitor ($1/\omega C_D \ll R_L$).

Looking at the current paths for the drain bias, one can see that

$$I_{DD} = I_D + I_L \tag{10.35}$$

Since I_{DD} is purely a DC term (since $\omega L_D \gg RL$) and since I_L has no DC component (because of C_D), I_{DD} is the average of the drain current, which for a square-wave drain current is just $\frac{1}{2} I_{\max}$.

Similarly,

$$V_D = V_{DD} + I_L R_L \tag{10.36}$$

Combining (10.35) and (10.36), one gets

$$V_D = V_{DD} + (\tfrac{1}{2} I_{\max} - I_D) R_L \tag{10.37}$$

Since we have assumed that I_D is a square wave varying from zero to $I_{\max}$ at frequency ω, V_D is likewise a square wave in voltage but 180° out of phase such that V_D is minimum when I_D is at $I_{\max}$. Since the FET will sink current only if $V_D \geq 0$, we come to the following condition, which determines $I_{\max}$:

$$(V_D)_{\min} = V_{DD} + (\tfrac{1}{2} I_{\max} - I_{\max}) R_L \geq 0 \tag{10.38}$$

or

$$I_{\max} \leq \frac{2V_{DD}}{R_L} \tag{10.39}$$

Finally, we have the FET device drain current limit,

$$I_{\max} \leq I_{\text{peak}} \tag{10.40}$$

These limits will come into play depending on the value of R_L, as shown in Fig. 10.20. If $R_L < 2V_{DD}/I_{\text{peak}}$, the drain current will range from zero to I_{peak} while the drain voltage will vary around $V_{DD} \pm (\tfrac{1}{2} I_{\text{peak}}) R_L$. This situation is described as current clipping because if one used a FET with higher I_{peak}, the output signal would be larger but varying V_{DD} slightly would not affect the output.

If $R_L > 2V_{DD}/I_{\text{peak}}$, the drain voltage swings from zero to $2V_{DD}$ while the drain current varies from zero to $2V_{DD}/R_L$. This situation is described as volt-

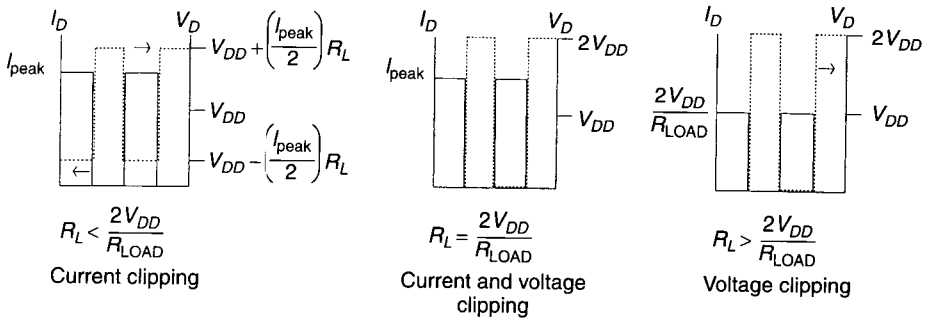


Figure 10.20 FET drain voltage and current waveforms for three ranges of R_L.

age clipping because if V_{DD} is raised, the output signal will increase while higher FET I_{peak} would not affect the output signal.

When $R_L = 2V_{DD}/I_{peak}$, one obtains maximum current and voltage swings of zero to I_{peak} and $2V_{DD}$ to zero, respectively. Now the amplifier is simultaneously voltage and current clipping.

The power taken from the drain power supply, P_{DC}, is given by

$$P_{DC} = V_{DD}\left(\tfrac{1}{2}I_{max}\right) \tag{10.41}$$

The total power delivered to the load is given by

$$P_{LT} = \left(\tfrac{1}{2}I_{max}\right)^2 R_L \tag{10.42}$$

but the power delivered to the load at the fundamental frequency, P_L, is what we usually care about, where $P_L < P_{LT}$ because the square-wave output contains all odd harmonics. Doing the Fourier integral, one obtains

$$P_L = \frac{2}{\pi^2} I_{max}^2 R_L \tag{10.43}$$

Thus about 81% of the output power is at the fundamental.

The drain efficiency, η_D, is the percentage of the DC power converted into RF power and delivered to the load at the fundamental:

$$\eta_D = \frac{P_L}{P_{DC}} = \frac{4}{\pi^2}\left(\frac{I_{max}}{V_{DD}}\right) R_L \tag{10.44}$$

The output power to the load and the drain efficiency are plotted versus R_L in Fig. 10.21. As can be seen, P_L is maximized at $R_L = 2V_{DD}/I_{peak}$ while η_D is maximized so long as $R_L \geq 2V_{DD}/I_{peak}$.

It is interesting to note that we have calculated output power and drain efficiency with no knowledge of the FET gate bias conditions. We could do this because under hard saturation the FET self-biases, approaching the square-wave approximation irrespective of small-signal biasing.

If we combine (10.35) and (10.36) to focus upon the drain current, we obtain

$$V_D = V_{DD} + (I_{DD} - I_D)R_L \tag{10.45}$$

This is the load-line equation. By overlaying the drain current and voltage relationship from (10.35) with FET drain I–V characteristics, we obtain curves similar to those shown in Fig. 10.22. In this case, we have used measured pulsed drain I–V data from the M/A-COM MSAG 5A, 625-μm power FET and shown a power optimal load line assuming $V_{DD} = 10$ V.

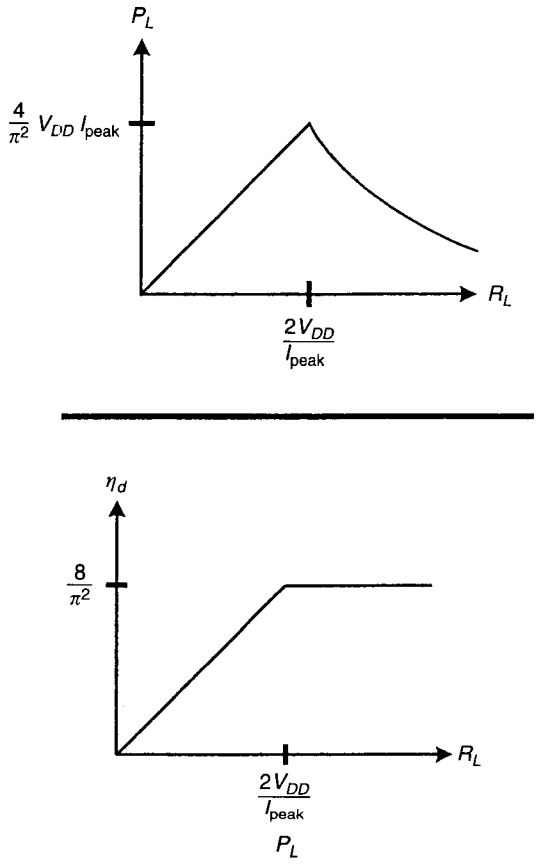


Figure 10.21 P_L and η_D as function of R_L.

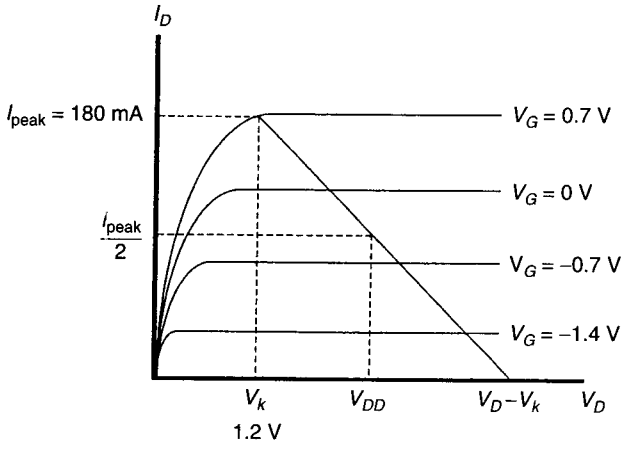


Figure 10.22 M/A-Com MSAG 5A, 625-μm power FET pulse drain I–V curves plotted with power optimal load line for $V_{DD} = 10$ V.

Note that the FET exhibits a knee voltage ($V_k \sim 1.2$ V) below which the drain current falls off sharply with drain voltage. This requires modification of our earlier equations to account for V_k:

$$I_{\max} \leq \frac{2(V_{DD} - V_k)}{R_L} \tag{10.46}$$

This reduces the output voltage swing maximum from V_{DD} to $V_{DD} - V_k$, which reduces the output power and drain efficiency for $R_L \geq 2V_{DD}/I_{\text{peak}}$ by a factor of $(V_{DD} - V_k)/V_{DD}$.

It is important to use pulsed I–V versus DC I–V measurements to determine I_{peak} and V_k. Relatively slow traps cause I_{peak} to be less (sometimes substantially) at microwave frequencies than at DC. Typically using pulses of 10 μs or less is adequate to reveal I–V changes with frequency at least up through the *Ku* band. The MSAG 625-μm FET has a peak DC drain current of about 280 mA, versus the pulsed drain peak shown of about 180 mA.

The power and drain efficiency calculations for the MSAG 625-μm FET at the peak power load line are as follows:

$$P_L = \frac{4}{\pi^2} I_{\text{peak}}(V_{DD} - V_k) = 0.64 \text{ W} \tag{10.47}$$

$$\eta_D = \frac{8}{\pi^2}\left(\frac{V_{DD} - V_k}{V_{DD}}\right) = 71.3\% \tag{10.48}$$

Measurements of the MSAG FET power and drain efficiency at 14 GHz and about 1.5 dB gain compression are $P_L = 0.56$ W and $\eta_D = 70\%$, respectively. Despite drastic assumptions, the agreement between the simple load line and measurements in general agree, provided device parasitic reactances at the operating frequency do not have significant effect on the device calculations. The above described simple theory also does not take harmonic contributions into account.

10.5.3 Power Amplifier Design

The design of power amplifiers for a particular application and frequency range is straightforward but sometimes tedious in the sense that they have to meet the physical, electrical, and thermal characteristics and cost requirements. Amplifier performance requirements in terms of, for example, frequency band, gain, power output, PAE, linearity, and input and output VSWR are determined by the FET sizes, the circuit design topology, matching networks, the number of gain stages, the aspect ratio for the FETs between the stages, design methodology, fabrication technology, and packaging. More often they involve tradeoffs in terms of size, electrical performance, reliability, and cost.

The design of narrow-band and wide-band amplifiers using power GaAs FETs requires several device and circuit considerations as given below:

1. First and foremost, one must select the proper power device types and their sizes or power levels to meet design requirements (available power supply, power output and frequency range). Silicon bipolar transistors can deliver more RF power than GaAs FETs in the L and S band. About 20–30% margin should be included from the device P_{out} to amplifier P_{out}.
2. Power transistors with higher breakdown voltage are desirable. Use transistors that are close to industry standards. Transistors on thin substrates with via holes have low series inductance and better heat dissipation.
3. Operate the amplifier circuit within its safest operating bias. Do not exceed the maximum breakdown voltage and current ratings.
4. Junction-to-case thermal resistance should be as low as possible for better performance and reliability.
5. Load-pull measurements are essential to accurately characterize power devices for optimum amplifier design.
6. Internally matched transistors help in reducing the effect of package parasitics. They offer higher efficiency and larger bandwidth.
7. Design the input matching network for maximum power transfer, while designing the output matching network for maximum power out. Matching circuits should offer minimum gain outside the desired frequencies.
8. Use lumped elements or lumped-distributed circuit elements for matching low impedance to 50 Ω in order to realize a compact circuit. Also use low-loss circuit elements at the output since the efficiency is reduced more by a given amount of loss in the output than in the input.
9. Use low loss and 85–90% efficient power-combining techniques for high-power modules.
10. For broadband amplifiers use low-Q matching networks.

Next we discuss how FET size scaling works to achieve higher power levels. In order to transfer maximum power, low-loss matching networks must be used to match source impedance to the input of the device and to match the output impedance of the device to the load impedance. The GaAs FET's input and output impedances decrease as the power level of the device goes up (as the gate width becomes larger). The input impedance is the most difficult to match. Most of the time, a combination of lumped and distributed elements is used in realizing impedance matching networks.

An approximate large-signal simplified equivalent circuit of a GaAs power FET is shown in Fig. 10.23. Various circuit elements can be approximated for an X-band 1-μm-gate-length FET biased at 10 V as follows:

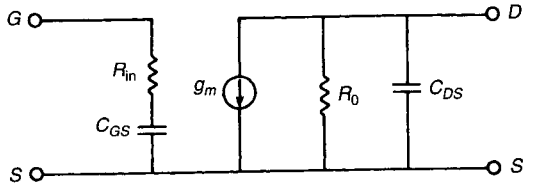


Figure 10.23 Simplified large-signal equivalent circuit for FET.

$$R_{\text{in}} = R_G + R_i + R_S \simeq \frac{2}{W} \quad (\Omega) \tag{10.49a}$$

$$C_{GS} \simeq W \quad (\text{pF}) \tag{10.49b}$$

$$g_m \simeq 80W \quad (\text{mS}) \tag{10.49c}$$

$$R_0 \simeq \frac{70}{W} \quad (\Omega) \tag{10.49d}$$

$$C_{DS} \simeq 0.12W \quad (\text{pF}) \tag{10.49e}$$

$$P_{1\,\text{dB}} \simeq 0.5W \quad (\text{W}) \tag{10.49f}$$

where W is the total gate width of the power FET in millimeters. For example, a device delivering about 2 W power out at 10 GHz must be 4 mm and has an input impedance about $0.5-j4\ \Omega$ and output impedance about $13.7-j0.41\ \Omega$.

Two parameters that are very useful for comparing different power amplifier configurations and classes are the power output and power-added efficiency. Here class A has been chosen for maximum linear power output. The circuit topologies are similar to those of small-signal amplifiers. The design of a narrow-band power amplifier is described step by step in the following.

C-Band Power Amplifier Design Example. Design a power amplifier at 5.5 GHz with 1.7 W power output, 25% power-added efficiency, and 6 dB gain at the 1-dB compression point. The power FET to be used has the following characteristics measured at 5.5 GHz and $P_{\text{in}} = 26$ dBm, $V_{DS} = 9$ V, and $I_{DSS} = 1000$ mA:

$$\text{Gain} = 7\ \text{dB} \qquad P_{\text{out}} = 2\ \text{W} \qquad \eta_{\text{add}} = 35\%$$

$$S_{11} = 0.89\angle -157^\circ \qquad S_{21} = 1.5\angle 84^\circ \qquad S_{12} = 0.049\angle 54^\circ$$

$$S_{22} = 0.31\angle -145^\circ$$

$$\Gamma_S = 0.73\angle 150^\circ \qquad \Gamma_L = 0.32\angle 160^\circ$$

From (10.18),

$$K = 1.267$$

Therefore the transistor is unconditionally stable at 5.5 GHz. Substituting Γ_S and Γ_L into (10.1) yields $G = 8.1$ dB. Thus the device has enough gain to design a 6-dB-power gain-amplifier when circuit losses are incorporated.

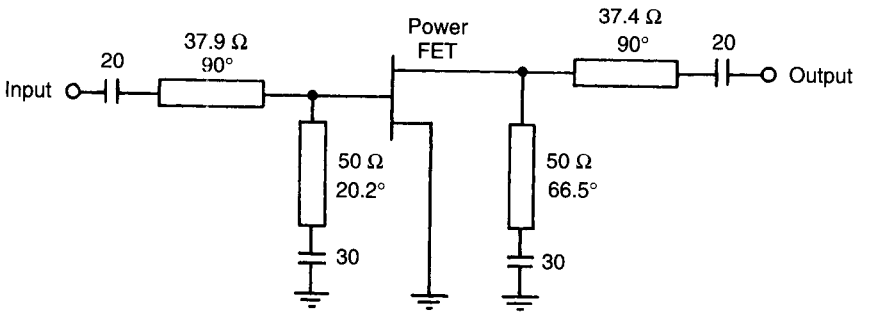


Figure 10.24 Schematic of 1.7-W-power amplifier. Electrical line lengths are at 5.5 GHz and capacitances in picofarads.

From $\Gamma_S = 0.73\underline{/150°}$, $Z_S = 8.35 + j13.05\ \Omega$, and $Y_S = 0.0348 - j0.0544$ S. A short-circuited stub has an input admittance of

$$Y = \frac{-jY_0}{\tan \beta l} = -j0.0544$$

when $Y_0 = 0.02$ and $\beta l = 20.2°$. Thus the characteristic impedance of the stub is 50 Ω. To match the parallel conductance of 0.0348 S to the source admittance of 0.02 S, a quarter-wave transformer of characteristic impedance $Z_0 = \sqrt{1/(0.02 \times 0.0348)} = 37.9\ \Omega$ is used. The complete input matching network is shown in Fig. 10.24.

For the output matching network, $\Gamma_L = 0.32\underline{/160°}$, $Z_L = 26.34 + j6.42$, and $Y_L = 0.0358 - j0.00874$.

A short-circuited stub has an input admittance $Y = -jY_0/\tan \beta l = -j0.00874$ when $Y_0 = 0.02$ and $\beta l = 66.5°$. In this case also the characteristic impedance of the stub is 50 Ω. To match the parallel conductance of 0.0358 S to the load admittance of 0.02 S, a quarter-wave transformer of characteristic impedance $Z_0 = \sqrt{1/(0.02 \times 0.0358)} = 37.4\ \Omega$ is used. The complete output matching network is shown in Fig. 10.24.

Wireless HBT Power Amplifier Design Example.* The example provided here is for a GSM power amplifier (PA), whose typical performance specifications are given below:

Frequency range	880–915 MHz
Power gain	30 dB
Output power	3 W
PAE	50%
Input VSWR	2:1
Supply voltage	3.5 V
Control current	5 mA @ 2.7 V

*This example was provided by Dr. Thomas Winslow of M/A-COM, Roanoke, VA.

To design such an amplifier, we selected 3-μm-emitter-width HBT-based monolithic technology. The HBT technology has gained acceptance as a cost-effective alternative to MESFET-based power amplifiers and has several advantages over MESFET power amplifiers. An HBT power amplifier operates from a single positive DC power supply resulting in reduced overall amplifier design complexity. The HBT power amplifiers are capable of very high power densities, which reduces the overall chip size and cost compared to MESFET power amplifiers. However, the thermal design of HBT power amplifiers is much more critical than with MESFET amplifiers. Great care must be taken to prevent thermal runaway in HBT power amplifier designs. The process of designing an HBT power amplifier can be divided into two important areas: electrical design and thermal design.

Electrical Design. To keep the device operational current density at ~0.1–0.2 mA/μm^2, the load line and class of operation of the output stage must be properly designed. Typically, the first stage is biased class A, 0.1 mA/μm^2 current density, but the last two stages are biased deep class AB or class B, resulting in a quiescent current density of approximately 0.008 mA/μm^2. Utilizing thick emitter metallization and proper via placement for minimum unit cell thermal resistance, the output stage was designed to operate at ~0.14 mA/μm^2 current density under RF power. The output stage is sized at 11,880 μm^2, which requires six parallel arms with 11 cells per arm, as seen in Fig. 10.25. Each cell consists of two fingers having a total area of $2 \times 3 \times 30 = 180$ μm^2. To meet the target gain specification, the power amplifier requires three stages. The device size ratio of the last two stages is 6.6:1 and the device size ratio of the first two stages is 5:1. The fundamental load impedance of the output stage as represented by a parallel *RC* combination is approximately 2 Ω and −26 pF. The fundamental load impedance of the second stage is approximately 13 Ω and −3.9 pF. The first-stage load impedance is easier to match than the last stage and is approximately 65 Ω with less than −1 pF capacitance.

The output match is formed using the output wirebonds and package inductance in series with a 80-mil-long, 50-Ω transmission-line segment followed by two shunt 10-pF, high-*Q* adjacent capacitors in parallel forming 20 pF. A secondary output capacitance of 3 pF is located 600 mils beyond a 20-pF capacitor by using another 50-Ω transmission-line section. The transmission lines can be formed on standard 10-mil-thick FR4 dielectric using 1-oz copper traces.

The collector bias can be fed into the output stage through a quarter-wave line, which can also act as a second-harmonic short or a low DC resistance 18-nH choke. The load impedance for the second stage is realized using the input capacitance of the third stage in parallel with a shunt inductance formed by series–parallel wirebonds directly off the collector of the second stage followed by a large-valued shunt RF bypassing capacitor. The collector bias of the second stage can be fed through this low-impedance point. The shunt inductive bias path is in parallel with the capacitive input of the last stage,

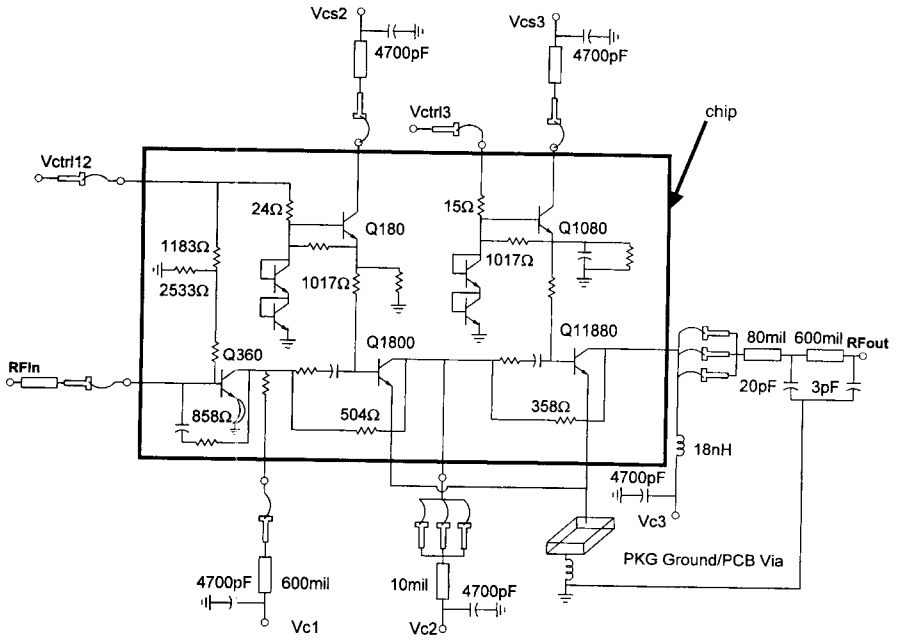


Figure 10.25 Schematic of GSM power amplifier for handset applications. Package grounding effects must be considered to maximize performance and stability.

forming a parallel resonance slightly offset from the center frequency of the band. The second- to third-stage interstage match is highly frequency tunable and is very sensitive but allows for a simple and very compact layout, requiring no on-chip inductors. The interstage match between the first and second stages of the amplifier is more forgiving and can be formed by using either a shunt transmission line or shunt inductance connected to the collector of the first stage. Bias for the first stage is supplied through bypassing on this line, as seen in the schematic in Fig. 10.25. All three stage devices have parallel feedback resistors for stabilization. The current design also includes a control circuit on the chip for stable operation of the PA over 0 to 50°C temperature.

Thermal Design Considerations. Two important thermally related issues must be addressed when designing an HBT PA. The first issue is ballasting. When large numbers of small unit HBT cells (e.g., 2 fingers by 3 μm × 30 μm = 180 μm^2) are joined together in parallel to form an output stage necessary for generating the desired target output power, small variations between unit devices can cause current hogging during operation. Gallium arsenide HBTs have a positive temperature coefficient, which means that as a HBT cell draws more current as it heats up; it further increases the operational junction temperature. Current hogging is a phenomenon due to the slight variation between unit HBT cells causing a few to draw more current than their neighbors. Ulti-

mately, a few cells will tend to draw excessive amounts of current (hogging) and reach a junction temperature that can cause serious performance degradation with serious lifetime degradation or even catastrophic junction failure. The simplest method to prevent this is by either base or emitter ballasting. Ballasting involves the placement of a small resistor in the emitter of each unit cell or in the base bias path of each unit cell to prevent current hogging. When a "hot" HBT cell within any array starts to draw slightly more current than the surrounding cells, the voltage across the ballasting resistance will increase, which will reduce the voltage applied across the base–emitter junction of the current hogging cell. This breaks the positive thermal feedback within the unit cell and will allow even distribution of heat across the output array of an HBT PA as it draws current under operation.

Another important aspect of HBT PA thermal design is the maximum operating current density. Since HBT power amplifiers are not as current limited as MESFET PAs, they can be designed to operate at very high current densities (and thus junction temperature) by a proper selection of load line and bias. Care must be taken to design the HBT PA to operate at a current density that will keep the junction temperature of the unit cells below 150°C. This will depend on process and layout specific factors such as emitter metallization thickness, wafer thickness, via location, cell pitch, and emitter dimensions. Excessive junction temperature can cause degradation in operational lifetime, degradation in RF performance, or even destructive failure.

Backside vias are useful for decreasing the effective thermal resistance of the last stage of a PA because they help to transfer heat from the emitters to the backside of the chip. The thermal resistance of the HBT PA is approximately 18°C/W. The layout symmetry of the output stage can help to improve stability, thermal uniformity, and power gain of the overall amplifier. See Fig. 10.26 for a layout of the example amplifier. The typical measured performance of this HBT PA design is shown in Fig. 10.27. Peak PAE is 54% at an output power of 35.5 dBm and 30 dB gain.

10.5.4 Design of Internally Matched Power FET Amplifier

Since the total gate width is very large in the case of high-power GaAs FETs, the impedance of GaAs FET chips becomes so low that input and output impedances are affected by parasitic capacitance and inductance of a package. It is difficult to match the amplifier circuits out of a package, especially of high frequencies. One of the most practical methods of designing broadband power FET amplifiers is to use internal matching [61–63] within a microwave package to deal with the low input impedance of the device.

Lumped elements and/or distributed elements for matching networks can be used. For broadband and power levels greater than 5 W, lumped elements are generally preferred for input matching of power FETs. Lumped inductors are realized by using bonding wires, and capacitors are of the metal–insulator–metal type that use high-dielectric-constant ceramics. Capacitors must have

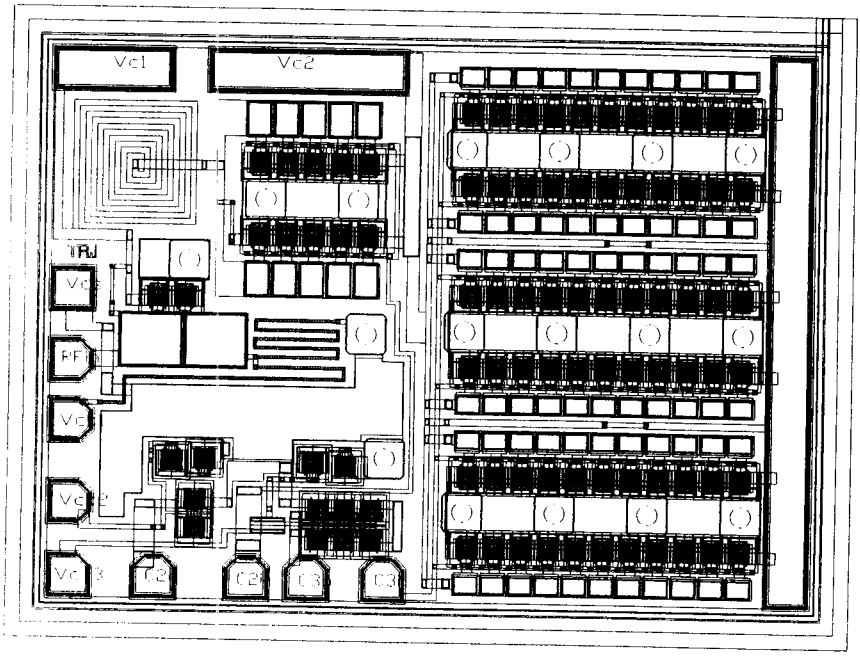


Figure 10.26 GSM power amplifier chip layout.

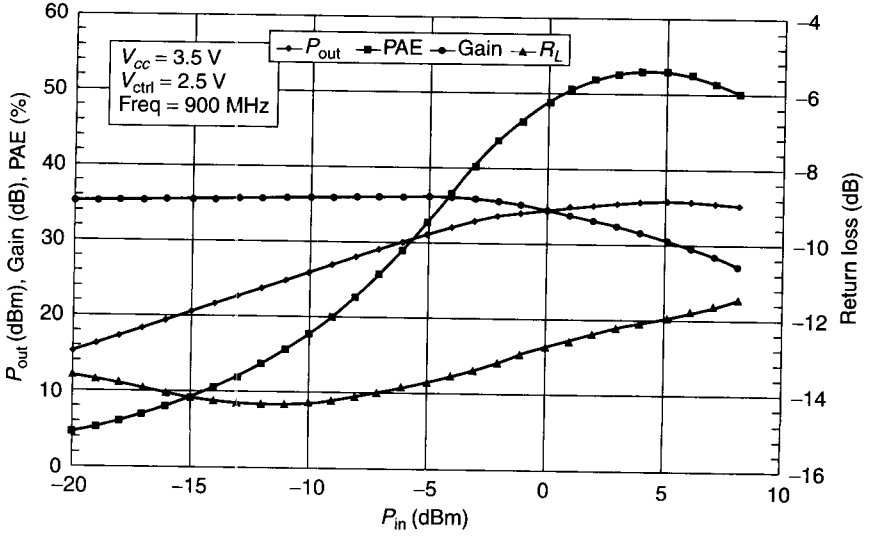


Figure 10.27 Example measured performance of GSM power amplifier design.

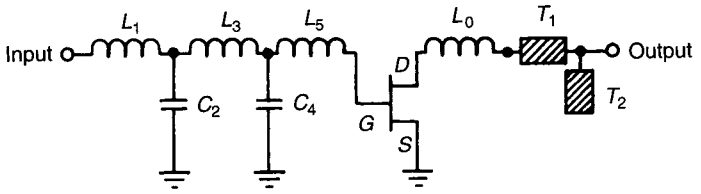


Figure 10.28 Schematic of internally matched power FET.

small parasitic inductance and resistance, sufficient thermal and mechanical strength, a small temperature coefficient, 40 V or higher breakdown voltage, and low cost. Since the output impedance is much higher than the input impedance, the output matching networks are realized using both lumped and distributed circuit elements. Microstrip lines on ceramic substrates are generally used for distributed circuit elements.

The schematic of an internally matched power FET is shown in Fig. 10.28. A 2.2-GHz, 140-W internally matched heterojunction FET used in a push–pull amplifier for cellular base-station applications is depicted in Fig. 10.29. This particular device has a total gate periphery of 332 mm, which gives a power output density of about 0.42 W/mm. The FET has two input and output leads for push–pull combining using external baluns. The design and fabrication of this circuit have been described elsewhere [63].

10.5.5 Power-Combining Techniques

Although there are fundamental limitations to the power that can be generated from a single device, the achievable power levels can be significantly increased

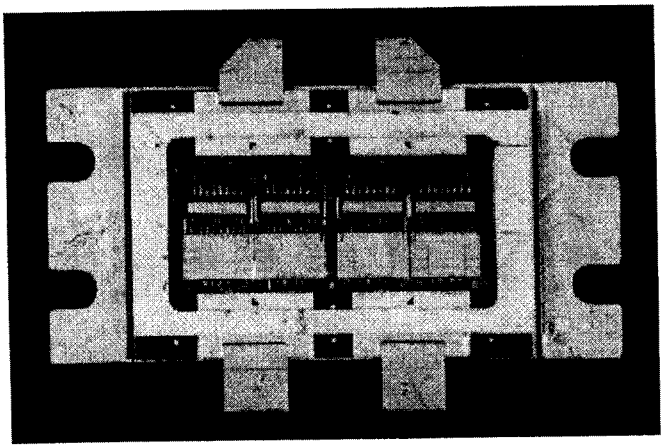


Figure 10.29 Photograph of a 2.2-GHz, 140-W internally matched FET. (After Takenaka et al. [63]. Reprinted with permission of IEEE.)

by combining a number of devices operating coherently or by accumulating the power from a number of discrete devices. This may be done in one of two ways: by combining power either at the device level or at the circuit level. Most combining techniques can provide "graceful degradation" in the case of failure of one or more devices in the combiner.

Device-Level Combining. Device-level combining is accomplished by clustering the devices in a region whose extent is small compared with a wavelength and is generally limited to the number of devices that can be combined efficiently. Higher power is obtained by bonding several devices onto a heat sink or a common carrier and connecting them to input and output matching circuitry, as shown in Fig. 10.30. The entire circuit may be hermetically sealed and used as a single device. Monolithic power amplifiers have shown great promise as an alternative to chip combining. These amplifiers have built-in power FETs and matching networks. The advantages of monolithic power amplifiers are small size, light weight, low cost, and totally matched circuits. Power levels of 25 and 14 W at 6 and 10.5 GHz, respectively, have been demonstrated using this technique.

Circuit-Level Power Combining. There are two categories of circuit level power combiners: (1) those that combine the output of N devices in a single step (known as N-way combiners) and (2) tree or chain combining structures. The desirable characteristics of a combining structure are minimum loss in the matching elements, minimum loss in the combining networks, minimum amplitude and phase imbalance, a good combiner input VSWR, even distribution of dissipated heat in the devices, and efficient removal of dissipated heat.

N-Way Combiners. N-way combining structures are the simpler of the two, and they avoid the use of several combining stages, thus making it possible to achieve high-efficiency combining. The structures can be either cavity or nonresonant combining structures. In cavity structures, cylindrical or rectangular cavities are used to combine the power output from a number of devices. These combiner–dividers have low loss ($\simeq$0.2 dB) and a combining efficiency of 85–90%. Note that, in general, combining and dividing circuits are identical. An

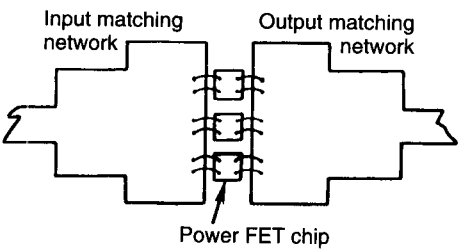
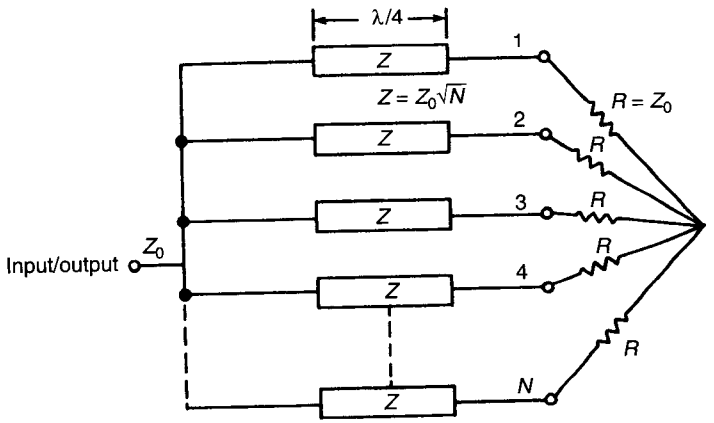


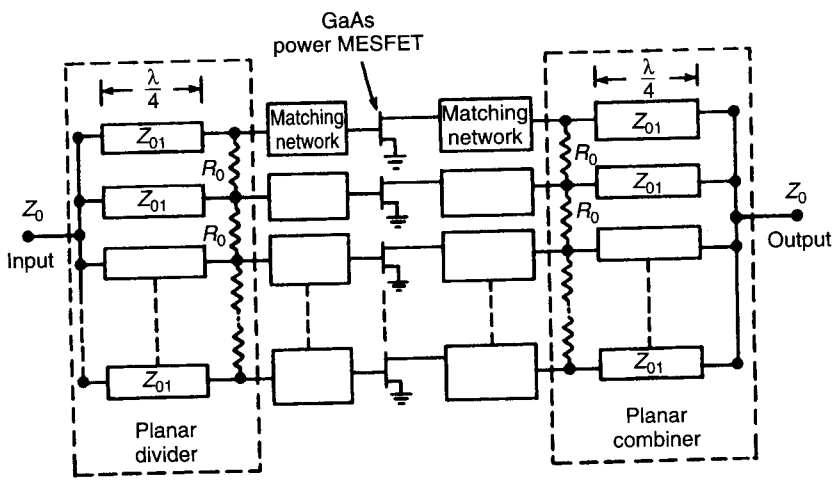
Figure 10.30 Chip combining of power FETs.

80-W FET amplifier using an eight-way combiner–divider working from 5.9 to 6.4 GHz has been developed [64].

Many nonresonant N-way combining techniques are also available. Essentially three types of N-way combiners are used for combining large numbers of amplifiers: Wilkinson, radial, and planar. The N-way Wilkinson divider [65] shown in Fig. 10.31*a* has the advantage of low loss, moderate bandwidth, and good amplitude and phase balance. However, its major disadvantage for power applications is the "floating starpoint" isolation resistors. These resistors require a nonplanar crossover configuration, which limits the power-handling



(*a*)



(*b*)

Figure 10.31 (*a*) Wilkinson N-way divider–combiner. (*b*) Modified N-way divider–combiner. Isolation resistors supress possible oscillations.

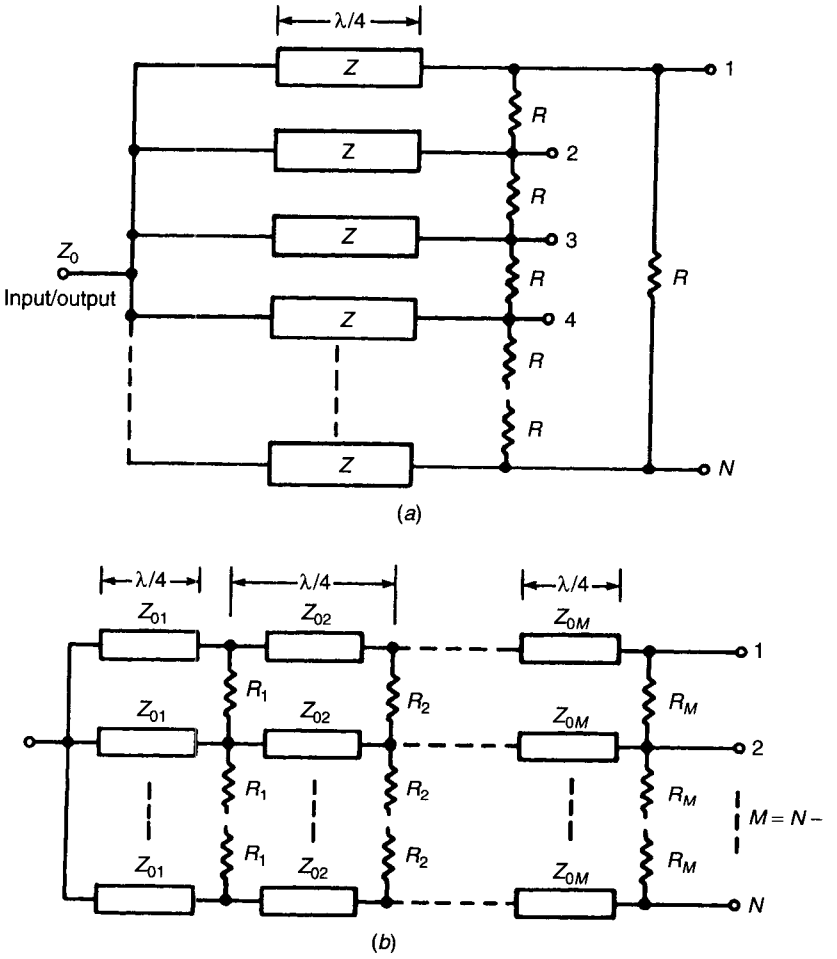
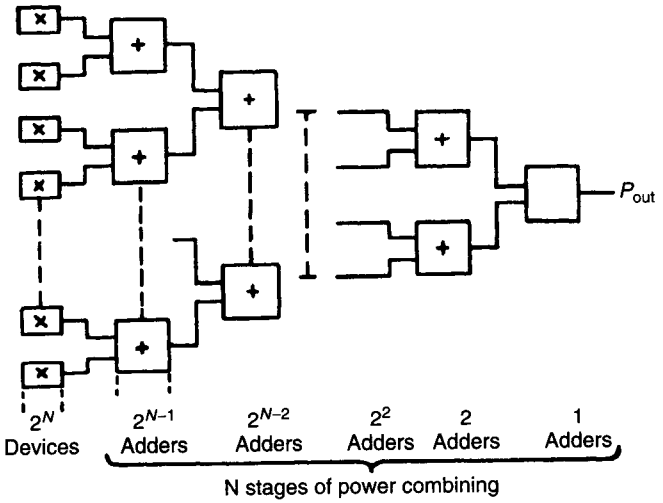
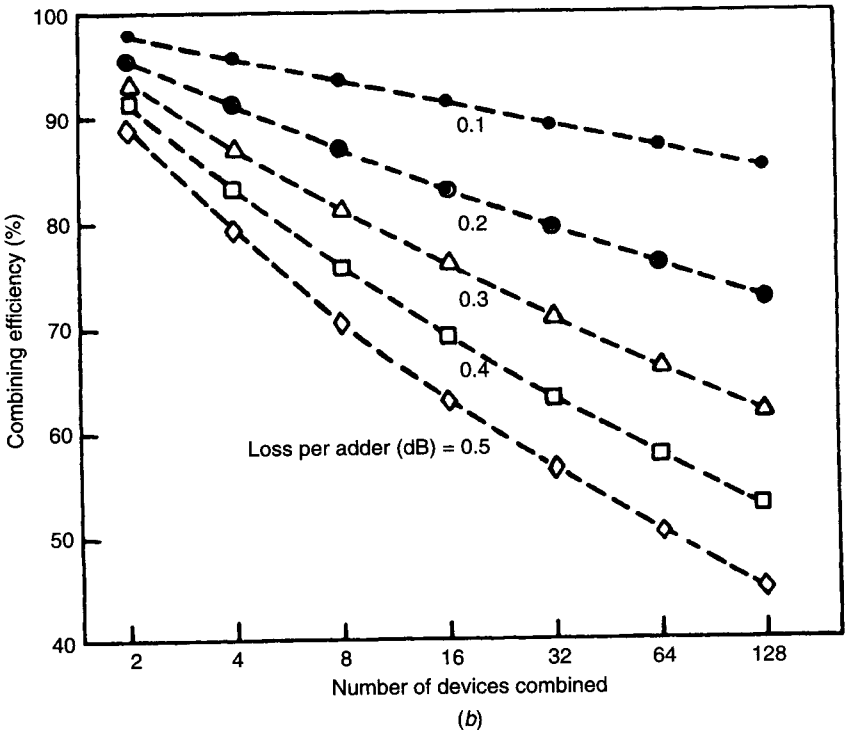


Figure 10.32 *N*-way divider–combiner: (*a*) radial line; (*b*) planar.

capability of the combiner. Fortunately, a simplified version as shown in Fig. 10.31*b* can also be used [66]. This particular arrangement has a combining efficiency on the order of 90% and shows much promise for chip combining as well as for MMIC applications. Radial-line and planar *N*-way combiners are shown in Fig. 10.32. The radial-line [67] combiner has low loss, inherent phase symmetry, and good isolation. Its main disadvantage is that it requires a three-dimensional structure. On the other hand, the planar *N*-way combiner–divider [68] requires $(N-1) \times N$ quarter-wave sections for maximum isolation and thus is very large. A compact structure using a tapered microstrip line has also been described [69]. The inherent redundancy in the *N*-way combiner makes it possible to obtain a graceful degradation characteristic. In an *N*-amplifier



(*a*)



(*b*)

Figure 10.33 (*a*) Corporate combining structure. (*b*) Combining efficiency for corporate combining structure.

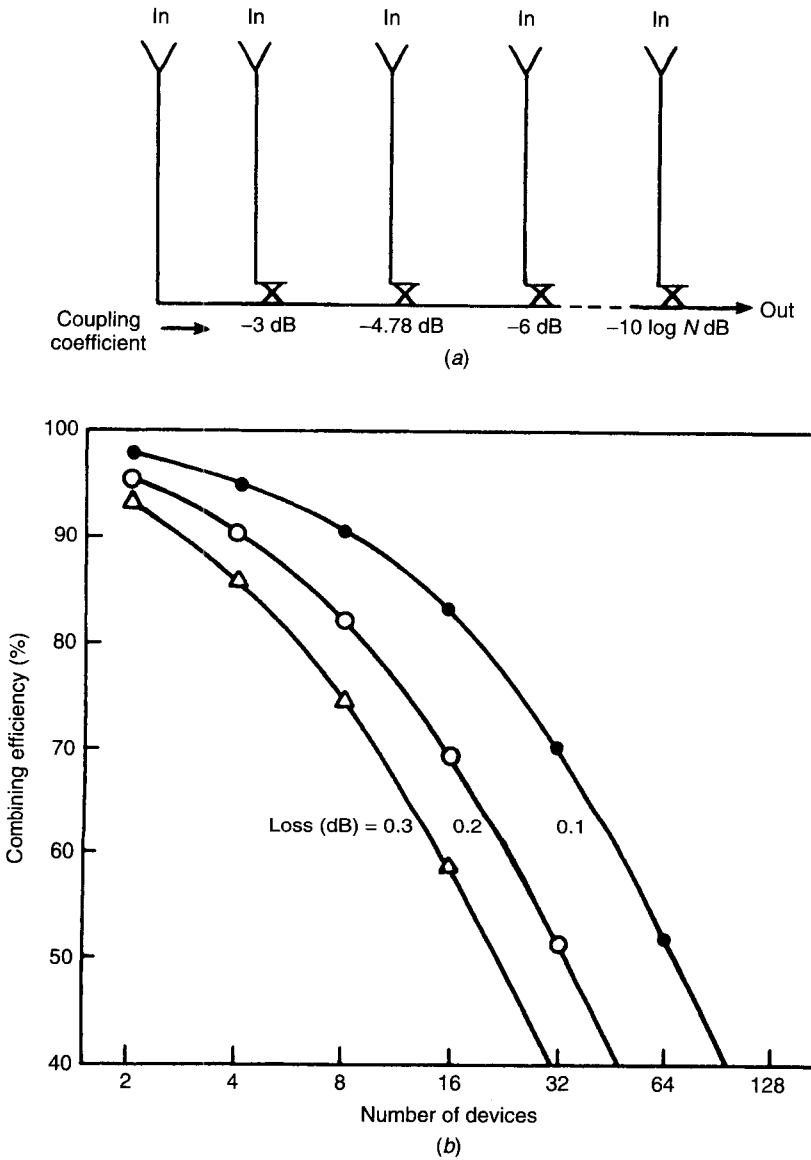
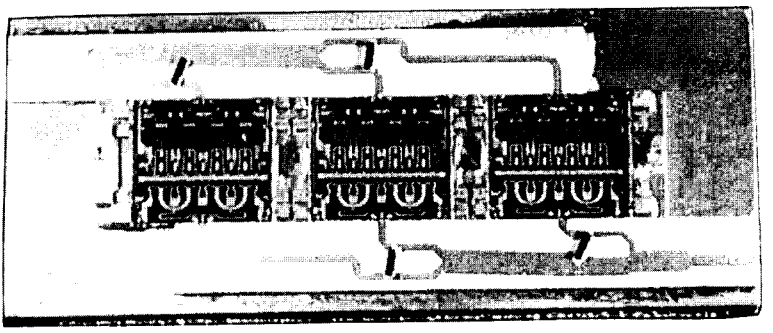


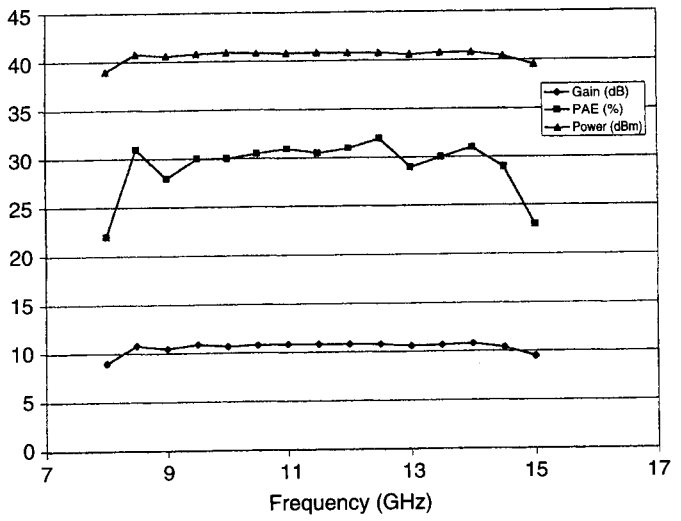
Figure 10.34 (*a*) Serial combining structure. (*b*) Combining efficiency for chain combining structure. Loss in decibels refers to the loss in each power path in each stage's coupler.

combiner with F failed amplifiers, the power output, relative to maximum output power, is given by $P_o/P_{\max} = (N - F)/N^2$.

Corporate Structures. A corporate structure (or tree) for combining power from two-way adders or combiners is shown in Fig. 10.33*a*. The loss in adders



(*a*)



(*b*)

Figure 10.35 (*a*) Photograph of 12-W HPA using three 5-W MMIC amplifiers. (*b*) Measured performance of 12-W HPA ($V_{DS} = 10$ V).

limits the combining efficiency. Figure 10.33*b* illustrates [70] the combining efficiency of the corporate structure versus the number of devices for various loss-per-adder values. The number of devices combined in this way is binary. Examples of two-way adders are directional couplers, hybrids, and the two-way Wilkinson combiner and are described in Chapter 5. Among the two-way adders, the Lange coupler is usually preferred because of its good isolation and wide-band properties. However, cascading these to obtain high-order combining becomes impractical beyond a four-way combiner due to the relatively high coupler loss.

Serial/Traveling-Wave Structure. A serial or chain combiner is shown in Fig. 10.34*a*. Here each successive stage of an N-way combiner adds $1/N$ of the

power delivered to the output. The number of the stage determines the required coupling coefficients for that stage, as indicated in the figure. One advantage of the chain structure is that another stage can be added by simply connecting the new source to the line after the Nth stage through a coupler with $10 \log(N + 1)$ coupling coefficient. The roles of input and output ports are changed for the divider structure. Losses in the couplers reduce the combining efficiency and bandwidth attainable with this approach. Combining efficiency for the chain structure for each path is shown [70] in Fig. 10.34*b*. The four-way structure on the microstrip can be used to realize combining efficiency on the order of 90% over an octave or greater bandwidth. The combiner is compact in size and has good input and output VSWR.

Example. Design a 12-W power amplifier at 12 GHz with 40% bandwidth and 30% power-added efficiency using monolithic IC amplifier chips having 50 Ω source and load impedances. The performance of these chips is measured as $P_{\text{out}} = 5$ W, gain = 12 dB, and $\eta_{\text{add}} = 35\%$.

In order to obtain 12 W power output we need three such chips and two low-loss three-way divider/combiner circuits. The three-way divider/combiner used is of the traveling-wave type similar to that described by Willems et al. [71] and fabricated on a 15-mil-thick alumina substrate ($\epsilon_r = 9.9$). Figure 10.35*a* shows the assembly of a 12-W high power amplifier (HPA) using three 5-W MMIC chips, and Fig. 10.35*b* shows the typical measured performance. The output power, gain, and PAE values are about 12 W, 11 dB, and 30%, respectively. The measured input return loss was better than 12 dB.

A comparison of various circuit-level power-combining techniques appears in Table 10.6.

Table 10.6 Comparison of Circuit-Level Power-Combining Techniques

Combining Technique	Advantages	Disadvantages
N-way W/G cavity	Low loss, high efficiency	Nonplanar, complex assembly narrow band
N-way Wilkinson	Low loss, moderate bandwidth, good isolation, high efficiency	Nonplanar, low power
N-way radial line	Low loss, good isolation	Nonplanar, complex assembly
N-way planar	Large bandwidth, good isolation, moderate loss	Large size, low efficiency
Corporate structure	Good isolation, large bandwidth	Impractical beyond four-way combining due to low efficiency
Serial/chain structure	More flexible, octave or greater bandwidth, good efficiency, good isolation, good input and output match	High resolution, fabrication required complex design, impractical beyond five-way combining due to fabrication limitations

REFERENCES

1. Harrison, A. E., *Klystron Tubes*, McGraw-Hill, New York, 1947.
2. Liao, S. Y., *Microwave Devices and Circuits*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
3. Gandhi, O. P., *Microwave Engineering and Applications*, Pergamon, New York, 1981.
4. Pierce, J. R., *Travelling Wave Tubes*, Van Nostrand, Princeton, NJ, 1950.
5. Blackwell, L. A., and K. L. Kotzebue, *Semiconductor-Diode Parametric Amplifiers*, Prentice-Hall, Englewood Cliffs, NJ, 1964.
6. Penfield, P., and R. P. Rafuse, *Varactor Applications*, MIT Press, Cambridge, MA, 1962.
7. Chang, K. K. N., *Parametric and Tunnel Diodes*, Prentice-Hall, Englewood Cliffs, NJ, 1964.
8. Watson, H. A., *Microwave Semiconductor Devices and Their Circuit Applications*, McGraw-Hill, New York, 1969.
9. Sterzer, F., "Tunnel Diode Devices," in L. Young (Ed.), *Advanced in Microwaves*, Vol. 2, Academic, New York, 1967, pp. 1–41.
10. Sterzer, F., "Transferred Electron (Gunn) Amplifiers and Oscillators for Microwave Applications," *Proc. IEEE*, Vol. 59, Aug. 1971, pp. 1155–1163.
11. Eastman, L. F., *Gallium Arsenide Microwave Bulk and Transit-Time Devices*, Artech House, Dedham, MA, 1973.
12. Haddad, G., *Avalanche Transit-Time Devices*, Artech House, Dedham, MA, 1973.
13. Ha, T. T., *Solid-State Microwave Amplifier Design*, Wiley, New York, 1981.
14. Gonzalez, G., *Microwave Transistor Amplifiers Analysis and Design*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
15. Fukui, H. (Ed.), *Low-Noise Microwave Transistors and Amplifiers*, IEEE Press, New York, 1981.
16. Pengelly, R. S., *Microwave Field-Effect Transistors—Theory, Design and Applications*, Wiley, New York, 1982.
17. Vendelin, G. D., A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley, New York, 1990.
18. Walker, J. L. B. (Ed.), *High-Power GaAs FET Amplifiers*, Artech House, Norwood, MA, 1993.
19. Chang, K., *Microwave Solid-State Circuits and Applications*, Wiley, New York, 1994.
20. Cripps, S. C., *RF Power Amplifiers for Wireless Communications*, Artech House, Norwood, MA, 1999.
21. Chang, K., I. Bahl, and V. Nair, *RF and Microwave Circuit Design*, Wiley, New York, 2002.
22. Slovick, M., "Measuring ACPR in CDMA Amplifiers," *Microwave J.*, Vol. 41, Dec. 1998, pp. 74–80.
23. Kurokawa, K., "Design Theory of Balanced Transistor Amplifiers," *Bell Syst. Tech. J.*, Vol. 44, Oct. 1965, pp. 1675–1698.
24. Niclas, K. B., et al., "The Matched Feedback Amplifier: Ultrawide-Band Micro-

wave Amplification with GaAs MESFETs," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, Apr. 1980, pp. 285–294.

25. Peterson, W. C., et al., "A Monolithic GaAs 0.1 to 10 GHz Amplifier," *IEEE Int. Microwave Symp. Dig.*, 1981, pp. 354–355.
26. Ayasli, Y., et al., "A Monolithic GaAs 1–13 GHz Traveling-Wave Amplifier," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, July 1982, pp. 976–981.
27. Ayasli, Y., et al., "2–20 GHz GaAs Traveling-Wave Amplifier," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-32, Jan. 1984, pp. 71–77.
28. Giamand, P., "A Complete Small Size 20 to 30 GHz Hybrid Distributed Amplifier Using a Novel Design Technique," *IEEE Int. Microwave Symp. Dig.*, 1986, pp. 343–346.
29. Niclas, K. B., "Reflective Match, Lossy Match, Feedback and Distributed Amplifiers: A Comparison of Multi-Octave Performance Characteristics," *IEEE Int. Microwave Symp. Dig.*, 1984, pp. 215–217.
30. Ikalainen, P. K., L. C. Witkowski, and K. R. Varian, "Low-Noise, Low DC Power Linear Amplifiers," *IEEE Microwave Millimeter-Wave Monolithic Circuits Symp. Dig.*, 1993, pp. 159–162.
31. Pedro, J. C., "Evaluation of MESFET Nonlinear Intermodulation Distortion Reduction by Channel-Doping Control," *IEEE Trans. Microwave Theory Tech.*, Vol. 45, Nov. 1997, pp. 1989–1997.
32. Kumar, M., J. C. Whartenby, and H. J. Wolkstein, "Predistortion Linearizer Using GaAs Dual-Gate MESFET for TWTA and SSPA Used in Satellite Transponders," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-33, Dec. 1985, pp. 1479–1488.
33. Nakayama, M., et al., "A Novel Amplitude and Phase Linearing Technique for Microwave Power Amplifiers," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1995, pp. 1951–1954.
34. Voyce, K. G., and J. H. McCandless, "Power Amplifier Linearization Using IF Feedback," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1989, pp. 863–866.
35. Pedro, J. C., and J. Perez, "An MMIC Linearized Amplifier Using Active Feedback," *IEEE Microwave Millimeter-Wave Monolithic Circuits Symp. Dig.*, 1993, pp. 113–116.
36. Nishikawa, K., and T. Tokumitsu, "An MMIC Low-Distortion Variable-Gain Amplifier Using Active Feedback," *IEEE Trans. Microwave Theory Tech.*, Vol. 43, Dec. 1995, pp. 2812–2816.
37. Moazzam, M. R., and C. S. Aitchison, "A Low Third Order Intermodulation Amplifier with Harmonic Feedback Circuitry," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1996, pp. 827–830.
38. Ali, F., M. R. Moazzam, and C. Aitchison, "IMD Elimination and ACPR Improvement for an 800 MHz HBT MMIC Power Amplifier," *IEEE RFIC Symp. Dig.*, 1998, pp. 69–71.
39. Jing, D., et al., "New Linearization Method Using Interstage Second Harmonic Enhancement," *IEEE Microwave Guided Wave Lett.*, Vol. 8, Nov. 1998, pp. 402–404.
40. Yamauchi, K., et al., "A Novel Series Diode Linearizer for Mobile Radio Power Amplifiers," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1996, pp. 831–834.
41. Haskins, C., T. Winslow, and S. Raman, "FET Diode Linearizer Optimization for Amplifier Predistortion in Digital Radios," *IEEE Microwave Guided Wave Lett.*, Vol. 10, Jan. 2000, pp. 21–23.

42. Yamauchi, K., et al., "A Microwave Miniaturized Linearizer Using a Parallel Diode with a Bias Feed Resistance," *IEEE Trans. Microwave Theory Tech.*, Vol. 45, Dec. 1997, pp. 2431–2434.

43. Katz, A., S. Moochalla, and J. Klatskin, "Passive FET MMIC Linearizers for *C*, *X* and *Ku*-Band Satellite Applications," *IEEE Microwave Millimeter-Wave Monolithic Circuits Symp. Dig.*, 1993, pp. 155–158.

44. Hayashi, H., M. Nakatsugawa, and M. Muraguchi, "Quasi-Linear Amplification Using Self Phase Distortion Compensation Technique," *IEEE Trans. Microwave Theory Tech.*, Vol. 43, Nov. 1995, pp. 2557–2564.

45. Meharry, D. E., "Multi-Octave Transformer Coupled Differential Amplifier for High Dynamic Range," *IEEE GaAs IC Symp. Dig.*, 1998, pp. 225–228.

46. Yang, Y., and B. Kim, "A New Linear Amplifier Using Low-Frequency Second-Order Intermodulation Component Feedforwarding," *IEEE Microwave Guided Lett.*, Vol. 9, Oct. 1999, pp. 419–421.

47. Sandhiya, P., et al., "Current Dumping: A Novel Linearization Technique for RF Power Transistors," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1999, pp. 815–818.

48. Kim, M. G., et al., "An FET-Level Linearization Method Using a Predistortion Branch FET," *IEEE Microwave Guided Wave Lett.*, Vol. 9, June 1999, pp. 233–235.

49. Campbell, C. F., "A Fully Integrated *Ku*-Band Doherty Amplifier MMIC," *IEEE Microwave Guided Wave Lett.*, Vol. 9, Mar. 1999, pp. 114–116.

50. Pothecary, N., *Feedforward Linear Power Amplifiers*, Artech House, Norwood, MA, 1999.

51. Kenington, P. B., "Methods Linearize RF Transmitters and Power Amps, Part 1," *Microwaves RF*, Vol. 37, Dec. 1998, pp. 102–116.

52. Kenington, P. B., "Methods Linearize RF Transmitters and Power Amps, Part 2," *Microwaves RF*, Vol. 38, Jan. 1999, pp. 79–89.

53. Curtice, W. R., "A MESFET Model for Use in the Design of GaAs Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, May 1980, pp. 448–456.

54. Materka, A., and T. Kacprzak, "Computer Calculations of Large-Signal GaAs FET Amplifier Characteristics," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-33, Feb. 1985, pp. 129–135.

55. Curtice, W. R., and M. Ettenberg, "A Nonlinear GaAs FET Model for Use in the Design of Output Circuits for Power Ampligiers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-33, Dec. 1985, pp. 1383–1394.

56. Statz, H., et al., "GaAs FET Device and Circuits Simulation in SPICE," *IEEE Trans. Electron Dev.*, Vol. ED-34, Feb. 1987, pp. 160–169.

57. Trew, R. J., "MESFET Models for Microwave CAD Applications," *Int. J. Microwave Millimeter-Wave Computer-Aided Eng.*, Vol. 1, Apr. 1991, pp. 143–158.

58. Nakhla, M. S., and J. Vlach, "A Piecewise Harmonic-Balance Technique for Determination of Periodic Response of Nonlinear Systems," *IEEE Trans. Circuits Syst.*, Vol. CAS-23, Feb. 1976, pp. 85–91.

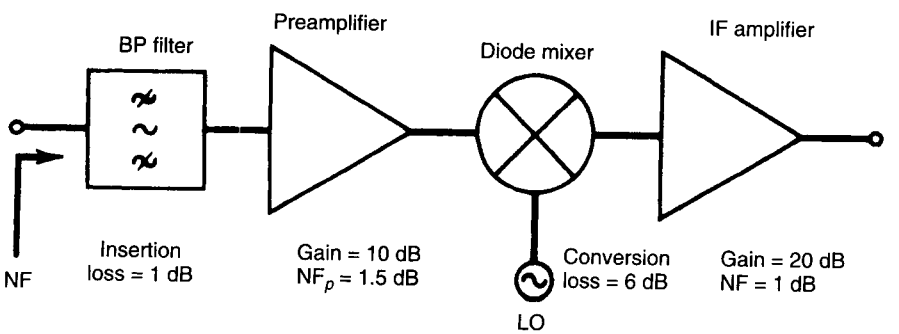
59. Rizzoli, V., and A. Neri, "State of the Art and Present Trends in Nonlinear Microwave CAD Techniques," *IEEE Trans. Microwave Theory Tech.*, Vol. 36, Feb. 1988, pp. 343–365.

60. Griffin, E. L., "Application of Loadline Simulation to Microwave High Power Amplifiers," *IEEE Microwave Mag.*, Vol. 1, June 2000, pp. 58–66.

61. Honjo, K., et al. "15-Watt Internally Matched GaAs FETs and 20-watt Amplifier Operating at 6 GHz," *IEEE Int. Microwave Symp. Dig.*, 1979, pp. 289–291.
62. Honjo, K., Y. Takayama, and A. Higashisaka, "Broad-Band Internal Matching of Microwave Power GaAs MESFETs," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, Jan. 1979, pp. 3–8.
63. Takenaka, I., et al., "Low Distortion High Power GaAs Pseudomorphic Hetero-junction FETs for *L*/*S*-Band Digital Cellular Base Stations," *IEEE MTT-S Int. Microwave Symp. Dig.*, 2000, pp. 1711–1715.
64. Okubo, N., et al., "A 6-GHz 80-W GaAs FET Amplifier with TM-mode Cavity Power Combiner," *IEEE Int. Microwave Symp. Dig.*, 1983, pp. 276–278.
65. Wilkinson, E., "An *N*-Way Hybrid Power Divider," *IRE Trans. Microwave Theory Tech.*, Vol. MTT-8, Jan. 1960, pp. 116–118.
66. Saleh, A. A. M., "Planar Electrically Symmetric *n*-Way Hybrid Power Dividers/Combiners," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, June 1980, pp. 555–563.
67. Schellenberg, J. M., and M. Cohn, "A Wideband Radial Power Combiner for FET Amplifiers," *IEEE Int. Solid State Circuits Conf. Dig.*, Feb. 1978, pp. 164–165.
68. Nagai, N., E. Maekawa, and K. Ono, "New *n*-way Hybrid Power Dividers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-25, Dec. 1977, pp. 1008–1012.
69. Yau, W., and J. M. Schellenberg, "An *n*-Way Broadband Planar Power Combiner/Divider," *Microwave J.*, Vol. 29, Nov. 1986, pp. 147–151.
70. Russel, K. J., "Microwave Power Combining Techniques," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, May 1979, pp. 472–478.
71. Willems, D., I. Bahl, and B. Kruger, "A Quasi-Microstrip Travelling-Wave Power Divider/Combiner for Use in High-Density Packages," *IEEE Microwave Guided Wave Lett.*, Vol. 3, May 1993, pp. 148–149.

PROBLEMS

10.1 Determine the noise figure of the receiver whose block diagram is shown below. Compare its performance with another receiver where the diode mixer is replaced by a dual-gate FET mixer having gain of 7 dB and noise figure of 10 dB.



10.2 Lumped-element values of a GaAs FET are given in Fig. 10.10. Determine the stability of the device and identify the element values that give rise to instability.

10.3 The scattering parameters of a GaAs FET are given in Table 10.3, which shows that the FET is potentially unstable at 6 GHz. Draw stability circles at 6 GHz and show how one can stabilize the device.

10.4 The scattering parameters for two different transistors at 8, 10, and 12 GHz are given below. Calculate the stability factor K in each case and draw input and output stability circles on the Smith chart. Show which transistor is suitable for a 10-GHz amplifier.

Frequency (GHz)	S_{11}	S_{21}	S_{12}	S_{22}
		Device 1		
8	$0.847\angle -89°$	$2.43\angle 121°$	$0.029\angle 60°$	$0.744\angle -19°$
10	$0.810\angle -104°$	$2.15\angle 112°$	$0.030\angle 61°$	$0.734\angle -23°$
12	$0.782\angle -117°$	$1.91\angle 103°$	$0.030\angle 67°$	$0.726\angle -27°$
		Device 2		
8	$0.820\angle -89°$	$2.21\angle 118°$	$0.028\angle 66°$	$0.644\angle -24°$
10	$0.778\angle -104°$	$1.95\angle 108°$	$0.030\angle 70°$	$0.635\angle -29°$
12	$0.746\angle -117°$	$1.73\angle 99°$	$0.032\angle 78°$	$0.628\angle -34°$

10.5 If the RF output voltage v_o of a FET amplifier is represented by

$$v_o = C_1 v_i + C_2 v_i^2 + C_3 v_i^3$$

where the input voltage $v_i = A \cos \omega_1 t$, show that the gain of the amplifier, G, in decibels is given by

$$G = 20 \log(C_1 + 0.75 C_3 A^2)$$

10.6 In Problem 10.5, if $C_1 = 10$, $C_2 = 0$, and $C_3 = -1$, determine the 1-dB compression point for the amplifier having input and output impedances of 50 Ω.

10.7 Derive an expression for the noise figure for n identical amplifiers, with gain G and noise figure F, cascaded in series. If $G = 5$ and $F = 2$, what is the maximum number of amplifiers that can be cascaded to obtain a noise figure less than 2.3?

10.8 A three-stage low-noise amplifier has a noise figure of 3 dB and gain of 30 dB over 1 GHz bandwidth at room temperature. Calculate the DR and DR_f when the 1-dB compression point for the amplifier is 15 dBm.

10.9 The scattering parameters of a GaAs FET at half I_{DSS} are given in Problem 10.4 (device 2). Design a 12-dB-gain amplifier at 10 GHz with input and output VSWR better than 1.5:1 when matched to 50-Ω source and load over a 10% bandwidth.

10.10 The common-source scattering parameters of a GaAs FET at 8 GHz are

$$S_{11} = 0.847/\underline{-89^\circ} \qquad S_{21} = 2.43/\underline{121^\circ}$$
$$S_{12} = 0.03/\underline{60^\circ} \qquad S_{22} = 0.744/\underline{-19^\circ}$$

Determine the common-gate and common-drain S parameters and the stability factor in all the three cases.

10.11 Design a low-noise amplifier at 4 GHz using a GaAs FET whose S parameters are given in Table 10.3 and other design parameters are given in Table 10.4. Desirable values for the noise figure and associated gain are 1.2 and 12 dB, respectively.

10.12 Design a two-stage amplifier with 20 ± 1 dB gain from 8–12 GHz. The source and load impedances are 50 Ω and the minimum acceptable VSWR at output is 1.5.

10.13 Design a broadband, 10-dB flat-gain MESFET amplifier using hybrid circuit matching networks in the input and output for the 6–8-GHz range. The S parameters are given in Table 10.3. The substrate to be used is RT/duroid ($\epsilon_r = 2.22$ and $h = 0.63$ mm). The effects of dispersion, strip thickness, and discontinuities may be ignored.

10.14 Derive an expression for the noise figure F of a balanced amplifier in terms of the individual amplifier noise figures F_1, F_2 and gains G_1, G_2 assuming the 3-dB 90° divider/combiner is ideal.

10.15 Using Eqs. (10.49a)–(10.49f), design a 2-W-power amplifier at 10 GHz assuming $V_{DSB} = 20$ V, $V_{DS} = 9$ V, and $I_{DS} = 0.7$ A. The amplifier operates in the class A mode.

10.16 Assuming that all individual amplifiers in an N-way amplifier are perfectly matched and have the same transmission phase and 100% efficiency of the divider–combiner, prove that the P_{out} of this amplifier is

given by

$$P_{\text{out}} = \frac{1}{N}\left(\sum_{n=1}^{N}\sqrt{P_n}\right)^2$$

where P_n is the output power of the nth individual amplifier.

10.17 Calculate the efficiency of combining and power output of a four-way chain divider–combiner realized using couplers having 0.1 dB insertion loss. The amplifier has 3 W power output, and the coupling factors are −3, −4.78, and 6 dB.

11

DETECTORS AND MIXERS

Robert G. Harrison

11.1 INTRODUCTION

Detectors and mixers use the nonlinear i–v characteristics of semiconductor devices to perform frequency extraction and translation operations on an input signal. Traditionally, two-terminal semiconductor diodes have been employed; today three-terminal active devices are sometimes used too. *Detection* is a form of rectification that extracts the AM component of a modulated RF signal: In principle, detection of an unmodulated RF signal produces a DC output signal. *Mixing* involves the multiplication of an input signal at frequency f_1 by another signal at f_2 to produce a translated output signal either at the sum frequency $f_1 + f_2$ or at the difference frequency $|f_1 - f_2|$.

11.1.1 Basics of Detection and Mixing

Detection. Detection resembles low-frequency power rectification but involves more subtleties. The function of a detector is to produce a DC or low-frequency "video" voltage that follows the amplitude modulation of an input RF signal. Most microwave detectors utilize a metal–semiconductor junction such as the Schottky barrier or Mott diode. Other two-terminal devices used as detectors include the tunnel and back diodes, the resonant tunneling diode (RTD), thermal devices,* and cryogenic devices such as the superconductor–

* Detectors such as thermocouples convert RF input energy into heat. Diode detectors are preferred for applications such as swept measurements because they have a faster response to varying power levels than do the thermal types; they also tend to have higher impedance levels and are inherently more efficient RF-to-DC converters than thermal devices.

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

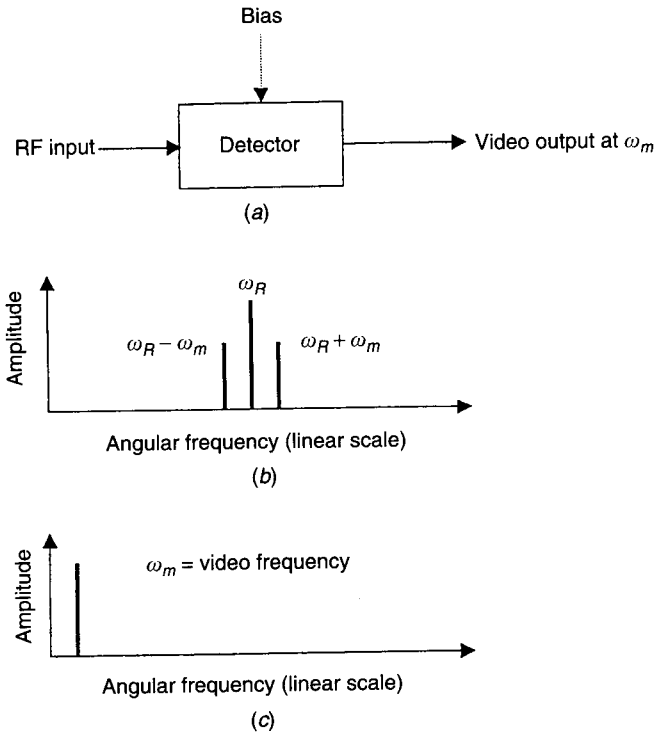


Figure 11.1 (*a*) Diagram of detecting function. (*b*) Spectrum of input RF signal ω_R with AM sidebands due to modulating signal ω_m. (*c*) Idealized output spectrum, consisting only of the so-called *video* signal.

insulator–superconductor (SIS) tunneling junction. For inputs up to about 20 μW conventional diodes and thermal detectors exhibit square-law characteristics and therefore provide a true rms measurement of all types of input signals (continuous wave (CW), pulsed, or modulated).

In Fig. 11.1*a* an input signal consisting of an RF carrier at ω_R and AM modulation sidebands at $\omega_R \pm \omega_m$ (see Fig. 11.1*b*) enters an idealized "detector block" that may have a DC bias input. The detector output, shown in Fig. 11.1*c*, is ideally a "video" signal at ω_m only. In reality the output contains undesired frequencies also; see Section 11.2.3.

The current flow in a forward-biased Schottky barrier is due to the injection of majority carriers from the semiconductor into the metal. This means that long reverse recovery times typical of *p–n* junctions, due to minority-carrier charge storage (diffusion) capacitance, are eliminated. Generally, Schottky diodes satisfy the basic detector requirements of minimum capacitance and low series resistance. However, for applications such as radio astronomy, where extreme sensitivity and low noise are of overriding importance, cryogenic detectors are preferred. Small-signal operation depends on the *slope* and *curvature* of the *i–v* characteristic near the bias point. On the other hand, large-signal

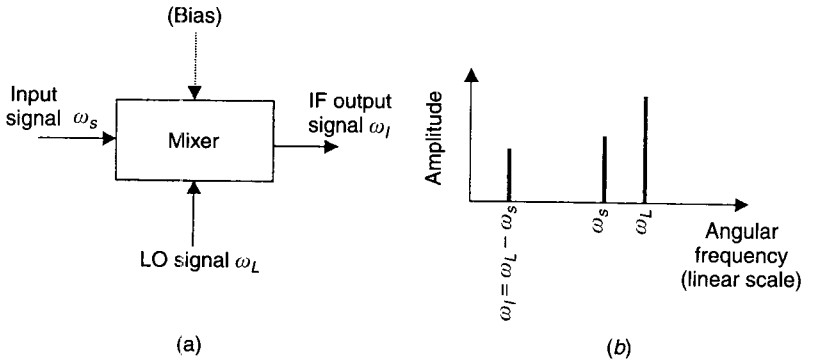


Figure 11.2 (*a*) Diagram of mixing function for a down converter. (*b*) Idealized output spectrum when $\omega_s < \omega_L$.

operation depends on the slope of the i–v characteristic far from the bias point; here the diode functions more like a *switch* than a nonlinear resistor. In large-signal detection the diode conducts over a portion of the input cycle and the output current (or voltage) follows the peaks of the input signal waveform.

Mixing. A resistive mixer converts a low-power *input signal* ω_s from one frequency to another by combining it with a higher power *local oscillator* (LO) signal ω_L in a device that has a resistive but nonlinear i–v transfer function.* Figure 11.2*a* illustrates the mixing process for a *down-converting* mixer, as in a receiver. An RF signal ω_s is mixed with a higher frequency ω_L (close to ω_s) to produce a low-frequency *intermediate frequency* (IF) $\omega_I = \omega_L - \omega_s$. Figure 11.2*b* shows the idealized output spectrum when $\omega_s < \omega_L$.

For an *up-converting* mixer, as in a transmitter, ω_s would be a low-frequency input signal that is mixed with a high-frequency ω_L to produce a high-frequency output $\omega_I = \omega_s + \omega_L$. In reality, the output spectrum of a mixer contains numerous other undesired sum and difference frequencies of ω_s and ω_L and their harmonics.

Mixers can be made using most of the nonlinear resistive devices that are used in detectors. In addition, *active mixers* with reduced conversion loss, even gain, can be made using the resistive nonlinearities inherent in three-terminal active devices such as GaAs FETs, HEMTs, and HBTs. Active detectors have also been described [1].

11.1.2 Applications of Detectors and Mixers

Mixer (superheterodyne) type receivers can achieve sensitivities of −100 dBm or more, but detector ("video-crystal") receivers can reach only about −50 dBm. Some applications of microwave detectors and mixers are listed below.

*Mixing action can also be obtained in nonlinear *reactive* devices, resulting in efficient (but typically narrow-band) *parametric* mixers. Certain parametric mixers can have gain [2].

Applications of Microwave Detectors

- Automatic gain control (AGC)
- Automatic leveling control (ALC) loops for microwave signal sources
- Conventional AM demodulation, as in microwave receivers
- Detector logarithmic video amplifier (DLVA) receivers [3]
- Logarithmic oscilloscopes, typically used to display the attenuation of a microwave filter versus frequency
- Monitoring microwave power levels and modulation envelope waveforms in transmitter installations
- Null detection in measurement systems
- Power detectors in six-port network analyzers
- Ultra-wide-band receivers
- Virtual batteries.*

Applications of Microwave Mixers

- Control systems
- Electronic countermeasure (ECM) systems
- Phase detectors in phase-locked loops
- Radio and radio astronomy
- Radars and radar astronomy
- Receivers and transmitters
- Regenerative frequency dividers (see Chapter 13)
- Spectrum analyzers
- Vehicle proximity detection.

11.1.3 Nonlinear Resistive Devices for Detection and Mixing

Important devices for detection and mixing are *pn*-junction and Schottky barrier diodes. Also used are the Mott diode, the tunnel diode, the back diode, the RTD, and the planar doped barrier (PDB) diode.

PN-Junction Diodes and Schottky Barrier Diodes. These are discussed in Chapter 8. For both, the relevant characteristic for detectors and mixers is the i–v curve of Fig. 11.3*a*, represented here by the exponential-diode equation

$$i = I_s\left[\exp\left(\frac{q_e v}{nkT}\right) - 1\right] = I_s(e^{\Lambda v} - 1) \tag{11.1}$$

where the right-hand form is a convenient shorthand and

* In this application, part of the received RF carrier is used to power a small MMIC receiver [4].

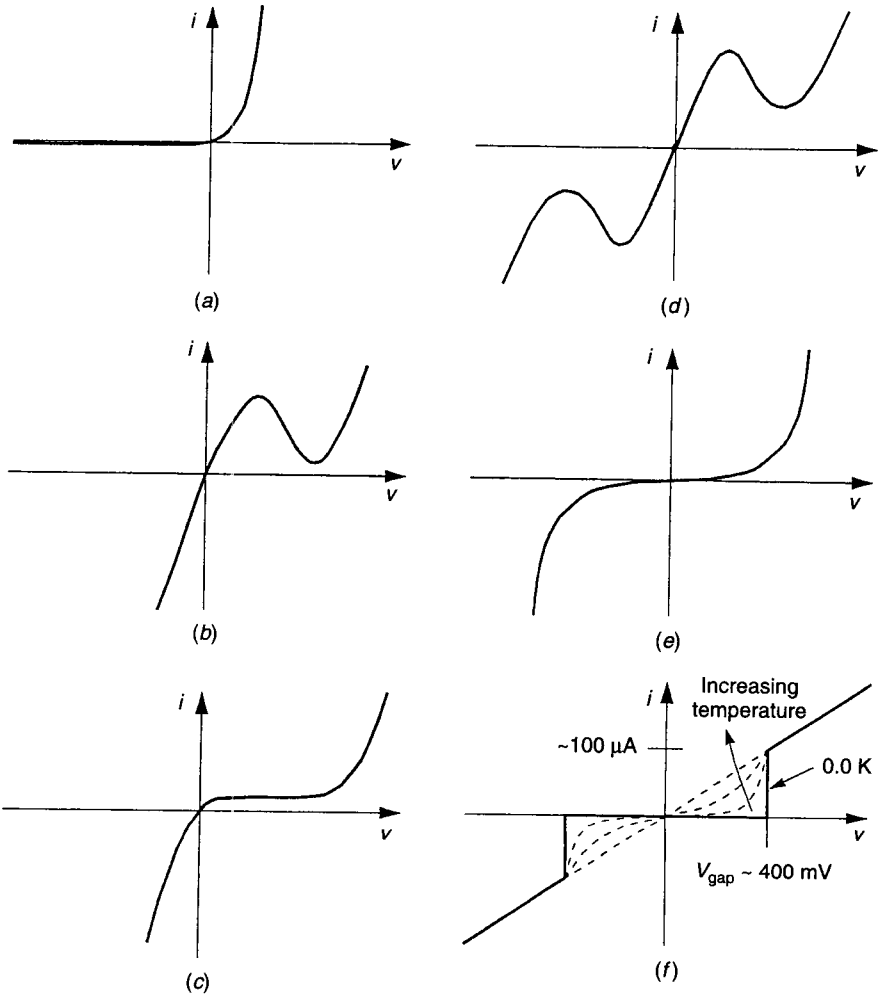


Figure 11.3 Static current–voltage (i–v) characteristics of some devices used in detectors and mixers: (*a*) *pn*-junction, Schottky, and Mott diodes; (*b*) tunnel diode; (*c*) back diode; (*d*) resonant tunneling diode; (*e*) planar doped barrier diode; (*f*) superconductor–insulator–superconductor tunneling junction.

I_s = diode saturation current

q_e = electron charge magnitude

k = Boltzmann's constant (1.38066×10^{-23} J/K)

n = diode ideality factor

$\Lambda = q_e/(nkT)$

T = temperature (K)

Ideally n is unity but is larger in real Schottky diodes due to high doping concentrations, oxide at the metal–semiconductor interface, and edge effects (in small diodes). A good Schottky has $n \sim 1.08$. At low temperatures I_s is T dependent, so that (11.1) is inaccurate. For temperatures below ~50 K, n itself becomes T dependent [5].

The Mott Diode. The Mott diode [6] has a junction capacitance C_j almost independent of bias voltage. In mixers, the advantage is that the minimum conversion loss and minimum noise figure occur under nearly the same circuit conditions, whereas with a Schottky, the conditions differ because of C_j variations with bias, tuning, and LO level. A futher advantage is that the Mott diode avoids parametric frequency conversion via nonlinear capacitive reactance, which can worsen the conversion loss and increase noise. Because of the absence of high-field regions, Mott diodes can provide reduced mixer noise temperature [7], comparable to that of SIS quasi-particle tunneling junction mixers. The Mott diode can be modeled like a Schottky but with a fixed junction capacitance.

Tunnel Diodes, Back Diodes, and Resonant Tunneling Diodes (RTD). Tunnel diodes, back diodes, and resonant tunneling diodes have all been used as detectors and mixers, and can have negative-differential-resistance (NDR) regions.

Passive detectors must absorb RF power to operate. However, a negative-resistance device such as a tunnel diode or RTD can be biased at the edge of its NDR region (see Fig. 11.3*b*), such that the negative resistance cancels the loss resistance, leaving a pure nonlinear resistance (at small signal levels). This increases the sensitivity [8].

The back diode is a form of tunnel diode designed to have strong i–v nonlinearity near zero bias and negligible negative resistance (see Fig. 11.3*c*).

The RTD is a quantum-well device [9–11] in which quantum-mechanical resonance causes tunneling-current maxima, and hence NDR regions, at specified voltages (see Fig. 11.3*d*).

Planar Doped Barrier Diode (PDB). The PDB diode is an all-GaAs device made by molecular beam epitaxy (MBE) [12–15]. High-resistivity Si devices have also been made [16]. Being a majority-carrier device, its frequency capabilities are like those of a Schottky, but it has the advantage that the degree of symmetry/asymmetry and the barrier height can be independently controlled. In its symmetrical form (Fig. 11.3*e*), it is useful for subharmonically pumped mixers (Section 11.3.7). The PDB diode consists of a sequence of $n^+/i/p^+/i/n^+$ layers, in which a thin ($x \sim 2$–10 nm) depleted p^+-doped acceptor sheet is positioned within a wide (~500 nm) undoped i region delimited by n^+ contact layers. Typical p^+ layer sheet doping density is $N_a \sim 10^{12}$ cm^{-2}; volume doping density of the n^+ layers is $N_d \sim 10^{18}$ cm^{-3} [13]. By adjusting the doping profiles and the lengths of the i layers, one can control the barrier height, the capaci-

tance per unit area, and the degree of asymmetry. A symmetrical PDB diode with series resistance R_s can be modeled as

$$v = \frac{2}{\Lambda} \operatorname{arcsinh}\left(\frac{i}{MI_s}\right) + iR_s$$

where $M \sim 2$ for a Si PDB [16].

Superconductor–Barrier–Superconductor (SIS) Quasi-Particle Tunneling Junction. This provides the most extreme i–v nonlinearity known; see Fig. 11.3*f*. It consists of a symmetrical sandwich of two superconducting electrodes separated by a thin insulating layer. For appreciable tunneling current the insulator must be thin enough for the quantum-mechanical wavefunctions associated with the two sides to overlap in the insulator. In a microwave version the associated capacitance can be absorbed into the embedding network. Alternatively, the capacitance can be tuned out by external elements since the series resistance R_s is *zero* in SIS tunnel junctions. The best devices use a Nb/AlO_x–Al/Nb sandwich structure. Since they have to be operated at ~1 K, their use is restricted to applications requiring extreme sensitivity and low noise, like millimeter-wave radio astronomy. For a quantum-mechanical discussion, see Tucker [17].

11.1.4 Noise in Detector and Mixer Devices

This section considers noise in devices conforming to (11.1). Thermal, shot, and flicker noise components influence detectors and mixers from microwave to submillimeter frequencies and should be considered in modeling and optimization.

Thermal Noise (Johnson or Nyquist Noise). Thermal noise is due to the random motion of charge carriers in conductive materials above 0 K. The thermal noise power in a measurement bandwidth B is

$$P_{\text{thermal}} = 4kTB \tag{11.2}$$

and is independent of frequency, current, or resistance. For a diode with series resistance R_s the mean-square thermal noise voltage is

$$\langle v_{\text{thermal}}^2 \rangle = 4kTBR_s \tag{11.3}$$

This is the microwave equivalent of the Rayleigh–Jones approximation of Planck's black-body radiation law [18, 19]:

$$\langle v_{\text{thermal}}^2 \rangle = \frac{4hfBR_s}{\exp(hf/(kT)) - 1} \tag{11.4}$$

where h is Planck's constant (6.6262×10^{-34} J · s) and f is the frequency. Approximation (11.3) is valid for $f < 100$ GHz and $T > 100$ K. An alternative way of expressing (11.3) is the mean-square noise current

$$\langle i^2_{\text{thermal}} \rangle = \frac{4kTB}{R_s} \tag{11.5}$$

Shot Noise. Shot noise is a manifestation of the fact that the physical current is a summation of the random flow of individual electrons. Assuming that the transit time is much smaller than the reciprocal of the measurement frequency, the shot noise diode current is [19]

$$\langle i^2_{\text{shot}} \rangle = 2q_e(I_{\text{dc}} + 2I_s)B \tag{11.6}$$

where I_{dc} is the forward-bias current. By differentiating (11.1), the incremental resistance at this bias is

$$R_j(I_{\text{dc}}) = \frac{1}{\Lambda(I_{\text{dc}} + I_s)} = \frac{nkT}{q_e(I_{\text{dc}} + I_s)} \tag{11.7}$$

From this and (11.6) the shot noise power is

$$P_{\text{shot}} = \langle i^2_{\text{shot}} \rangle R_j = \langle v^2_{\text{shot}} \rangle / R_j = 2nkTB\left(1 + \frac{I_s}{I_{\text{dc}} + I_s}\right) \tag{11.8}$$

This shows that at zero bias, a diode with $n \sim 1$ has nearly equal shot noise and thermal noise powers per unit bandwidth [20].

Flicker Noise. Flicker noise ($1/f$ noise), is due to imperfections at the metal/semiconductor interface. Flicker noise power depends on both measurement bandwidth and bias current and is approximately proportional to $1/f$. In a bandwidth B between limits f_H and f_L the mean square flicker noise voltage can be written (Problem 11.1)

$$\langle v^2_f \rangle = 2nkTf_nR_j\left(1 + \frac{I_s}{I_s + I_{\text{dc}}}\right) \ln\left(\frac{f_H}{f_L}\right) \tag{11.9}$$

where f_n is the "noise corner frequency" at which the mean square shot and flicker noise voltages are equal. A way of avoiding the "$1/f$" problem and at the same time greatly improving receiver sensitivity is to use a mixer in front of the detector, and to set the mixer difference frequency $|f_{\text{R}} - f_{\text{L}}| > f_n$.

Summing the thermal, shot, and flicker components from (11.3), (11.8), and (11.9), the *total mean-square noise voltage* is

$$\langle v_\Sigma^2 \rangle = 4kTB\left\{R_s + \frac{n}{2}R_j\left(1 + \frac{I_s}{I_s + I_{dc}}\right)\left(1 + \frac{f_n}{B}\ln\left[\frac{f_H}{f_L}\right]\right)\right\} \quad (11.10)$$

This result simplifies when $I_s \ll I_{dc}$.

The "*white noise*" *temperature ratio* t_w measures the noisiness of a biased diode compared with that of of a pure resistor [21]. It is defined by

$$t_w = \frac{\text{total noise power available from a diode, in bandwidth } B}{\text{thermal noise power available from a resistor at } T_0, \text{ in bandwidth } B}, \quad (11.11)$$

where T_0 is a reference temperature, conventionally 290 K. "White" implies frequency independence. It can be shown [21] that for a diode obeying (11.1),

$$t_w = \frac{1}{(R_j + R_s)}\frac{T}{T_0}\left[R_s + \frac{1}{2}n\left(1 + \frac{I_s}{I_{dc} + I_s}\right)R_j\right] \quad (11.12)$$

When $R_j \gg R_s$, (11.12) reduces to the familiar

$$t_w \cong \frac{1}{2}n\frac{T}{T_0}\left(1 + \frac{I_s}{I_{dc} + I_s}\right) \quad (11.13)$$

For forward bias, t_w is usually <1. However, t_w does not include the flicker noise component that is important for detectors. Assuming that $I_{dc} \gg I_s$, $R_j \gg R_s$, and $n \sim 1.0$, a modified form of (11.12) that does include the flicker noise is [20] the *frequency-dependent noise ratio*

$$t_n(f) \cong t_w + \frac{1}{2}\frac{f_n}{f} \quad (11.14)$$

where the last term represents the flicker noise component.

11.2 DETECTORS

This section introduces a basic detector circuit and reviews several detector performance measures. It concludes with three ways of analyzing Schottky and *pn*-junction detectors:

(a) A *second-order* diode model demonstrates demodulation of an AM carrier.

(b) A *linearized* diode model can be employed to derive performance measures theoretically.

(c) A *nonlinear* exponential-diode model is used to predict the large-signal and temperature performance.

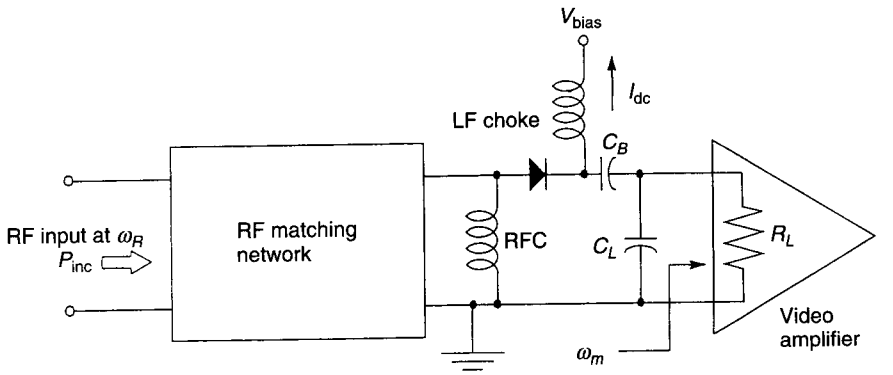


Figure 11.4 Simplified detector circuit.

The small-signal approaches (a, b) provide spectral and frequency response data; the large-signal analysis (c) predicts detector performance over a wide dynamic range.

11.2.1 Principle of Operation

Figure 11.4 shows a simplified detector circuit. Incident RF power P_{inc} at ω_R is applied via a matching network such that P_{inc} is absorbed in the resistive part $R_v = R_j + R_s$ of the diode impedance. Bypass capacitor C_L short circuits the output for RF and prevents P_{inc} getting into the video load resistor R_L (where it would be wasted). An external bias I_{dc} can be applied to the diode; the RF choke (RFC) provides a DC return, while the large blocking capacitor C_B, which has negligible reactance at the video frequency ω_m, isolates I_{dc} from the video amplifier.

11.2.2 Detector Sensitivity Measures

The sensitivity of a low-level detector depends on the rectification efficiency, output impedance, and noise properties of the diode; the input impedance, bandwidth, and noise properties of the video amplifier following the diode; and the RF matching network. Rectification efficiency is quoted either as voltage sensitivity or as current sensitivity. The following measures relate primarily to *pn*-junction and Schottky detectors.

Voltage Sensitivity. The voltage sensitivity γ_v (also denoted β_γ, γ, or K) is defined as the ratio of the *increase* in the open-circuit video output voltage ΔV_v^{oc} to the incident microwave power P_{inc}:

$$\gamma_v = \frac{\Delta V_v^{oc}}{P_{inc}} \qquad (\text{mV/mW}) \tag{11.15}$$

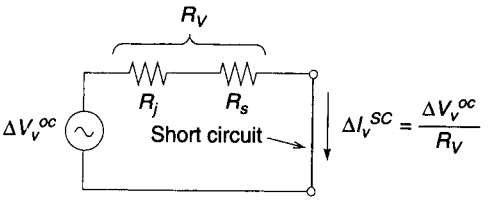


Figure 11.5 Detector equivalent circuit at video frequency ω_m used to obtain a relation between γ_v and β_I.

Sometimes γ_v is defined under loaded conditions [22]. "Open circuit" implies a load $R_L \gg 50\ \Omega$. To ensure detector operation in the square-law range, γ_v is usually measured at P_{inc} in the range between -20 dBm and the noise floor, over which the video output power is proportional to the square of the incident input power:

$$P_{\text{out}} = \frac{(\Delta V_v^{\text{oc}})^2}{R_v} = \frac{(\gamma_v P_{\text{inc}})^2}{R_v} \tag{11.16}$$

where $R_v = R_j + R_s$ is the so-called video resistance.* Depending on the detector type, γ_v is typically 300–20,000 mV/mW. A theoretical expression for γ_v will be obtained in Section 11.2.3.

Current Sensitivity. The current sensitivity β_I (or current responsivity S_I) is defined as the ratio between the *increase* in the short-circuit video output current ΔI_v^{sc} and P_{inc}:

$$\beta_I = \frac{\Delta I_v^{\text{sc}}}{P_{\text{inc}}} \qquad (\text{mA/mW}) \tag{11.17}$$

Referring to Fig. 11.5, the relationship between β_I and the voltage sensitivity γ_v is

$$\beta_I P_{\text{inc}} = \frac{\gamma_v P_{\text{inc}}}{R_v} \tag{11.18}$$

or

$$\beta_I = \frac{\gamma_v}{R_v} \tag{11.19}$$

*The video resistance R_v, the real part of the small-signal diode impedance Z_v, depends on I_{dc}.

assuming matched input and output ports [23]. The quantity β_I varies with bias; a theoretical expression will be obtained in Section 11.2.3.

Minimum Detectable Signal and the Figure of Merit M. From (11.15) and (11.19) the open-circuit video "signal" output voltage due to an incident noise power P_{inc} is

$$\Delta V_v^{\text{oc}} = \gamma_v P_{\text{inc}} = \beta_I R_v P_{\text{inc}} \tag{11.20}$$

If the thermal (Johnson) noise of the diode dominates, then the video noise is

$$V_n = \sqrt{4kTBR_v} \tag{11.21}$$

where B is the video bandwidth. The minimum detectable power, also called *minimum detectable signal* (MDS), is found by equating V_n and ΔV_v^{oc}, so that the signal-to-noise ratio is 1.0, with the result*

$$\text{MDS} = P_{\text{min}} = \frac{\sqrt{4kTB}}{\beta_I \sqrt{R_v}} = \frac{\sqrt{4kTB}}{M} \tag{11.22}$$

where [24]

$$M = \beta_I \sqrt{R_v} = \frac{\gamma_v}{\sqrt{R_v}} \qquad (W^{-1/2}) \tag{11.23}$$

is the detector *figure of merit* [20] and is generator and bandwidth independent. However, it does not include shot noise or $1/f$ noise due to non-zero bias currents, so M is of limited use for semiconductor detectors.†

Noise Equivalent Power. Noise equivalent power (NEP) is defined as the input power P_{in} needed to produce an output signal-to-noise ratio of 1.0 for a bandwidth of 1 Hz. This measure is appropriate when the video output frequency is low and when low-noise video amplifiers are used. According to [25],

$$\text{NEP} = \frac{2nkT}{q_e}\left[1 + \left(\frac{f_{\text{in}}}{f_c}\right)^2\right]\sqrt{\left(\frac{4kT}{R_j}t_w\right)\left(1 + \frac{R_s}{R_j}\right)\left(1 + \frac{f_n}{f}\right)} \qquad (\text{W/Hz}^{1/2}) \tag{11.24}$$

* Minimum detectable power was originally [21] defined as the smallest amplitude signal that could be observed as an oscilloscope trace when its position along the trace was unknown. A highly subjective measure, it corresponds to a signal-to-noise (S/N) ratio of ~1, agreeing with (11.23).

† A modified form of M that includes the effect of I_{dc} is discussed elsewhere [24].

where f_{in} = microwave input frequency
f = video output frequency
f_n = noise corner frequency
f_c = diode cutoff frequency (q.v.), $=1/[2\pi R_s C_j(0)]$

If $t_w \sim 1.0$, $R_s \ll R_j$, $f_{\text{in}} \ll f_c$, and $f \gg f_n$, then

$$\text{NEP} \sim \frac{2nkT}{q_e}\sqrt{\frac{4kT}{R_j}} \tag{11.25}$$

As will be shown in Section 11.2.3, the maximum value of the current sensitivity is

$$\beta_I^{\max} = \frac{q_e}{2nkT} \tag{11.26}$$

Consequently, NEP can also be written as

$$\text{NEP} = \frac{1}{\beta_I^{\max}}\sqrt{\frac{4kT}{R_j}} \tag{11.27}$$

A typical value of NEP is 10^{-11} W/Hz$^{1/2}$ (equivalently -80 dBm/Hz$^{1/2}$).

Tangential Signal Sensitivity (TSS). Despite its subjective nature, TSS remains a common sensitivity rating for detector diodes [26–28]. It is defined as the signal input power (in dBm) needed to shift the video output noise* level by an amount equal to its peak magnitude V_{pk}. At the TSS condition [24] (see Fig. 11.6),

$$V_{\text{pk}} = \Delta V_v^{\text{oc}} - V_{\text{pk}} \tag{11.28}$$

By assuming an approximately sinusoidal noise voltage waveform, it can be shown (Problem 11.2) that the TSS power level is

$$\text{TSS} \approx \frac{2\sqrt{2}\sqrt{4kTBR_v}}{\gamma_v} = 2\sqrt{2}(\text{MDS}) \qquad \text{(W)} \tag{11.29}$$

According to the traditional procedure [20], an observer measures TSS by setting the level of an RF input pulse so that the amplitude of the video output pulse is apparently equal to the average peak-to-peak noise, as viewed on an oscilloscope (see Fig. 11.6). When the lower noise peaks on the output pulse

*Due to the diode itself and to the amplifier.

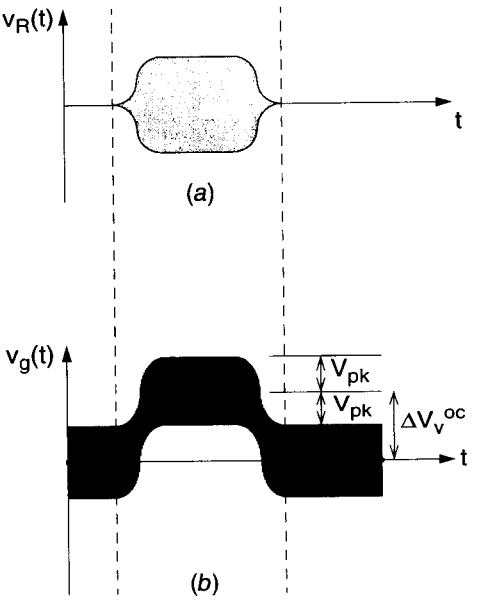


Figure 11.6 (*a*) RF input pulse. (*b*) Video output pulse, as seen at receiver output, for TSS input power level. Note that the value of TSS should be accompanied by a statement of the video bandwidth.

are colinear with the upper noise peaks on either side, P_{inc} equals TSS. Typically, TSS is in the −50- to −40-dBm range (for a 3-dB video bandwidth of 2 MHz). Even in wide-band systems, detection is so inefficient that the incoming (antenna) noise need not normally be considered. Virtually all the noise comes from the detector diode and the video amplifier.

Detector Sensitivity in Practice. In practice the sensitivity of a low-level detector depends on numerous factors, including the rectification efficiency, quoted as either voltage sensitivity or current sensitivity; the output impedance and noise properties of the diode; the input impedance, bandwidth, and noise properties of the video amplifier that follows the diode; and the RF matching network.

11.2.3 Small-Signal Detector Theory

If a Schottky or *pn*-junction detector is operated in its square-law region where the incident RF voltage v_{in} is much less than nkT/q_e (~26 mV at 300 K), a small-signal analysis is possible. If v_{in} is $\sim nkT/q_e$ or larger, then a large-signal analysis (Section 11.2.4) is needed.

Demodulation of Amplitude-Modulated RF Carrier. Detection of an AM RF carrier can be explained using a *second-order* diode model. This is derived

from (11.1) by writing the junction voltage v as the sum of a steady bias V_0 and a small-signal voltage Δv,

$$v = V_0 + \Delta v \tag{11.30}$$

and retaining the first three terms of a Taylor expansion of (11.1):

$$i(\Delta v) = i|_{V_0} + \Delta v \frac{di}{dv}\bigg|_{V_0} + \frac{1}{2!}\Delta v^2 \frac{d^2 i}{dv^2}\bigg|_{V_0} + \frac{1}{3!}\Delta v^3 \frac{d^3 i}{dv^3}\bigg|_{V_0} + \cdots \tag{11.31}$$

Here, the constant term

$$i|_{V_0} = i(V_0) = I_s(e^{\Lambda V_0} - 1) = I_{\text{dc}} \tag{11.32}$$

is the bias current. The first derivative is

$$\frac{di}{dv}\bigg|_{V_0} = \Lambda I_s e^{\Lambda V_0} = \Lambda(I_{\text{dc}} + I_s) = \frac{1}{R_j} = G_j \tag{11.33}$$

which defines the dynamic resistance R_j (conductance G_j) of the junction, and the second derivative is

$$\frac{d^2 i}{dv^2}\bigg|_{V_0} = \Lambda^2 I_s e^{\Lambda V_0} = \Lambda^2(I_{\text{dc}} + I_s) = \Lambda G_j \tag{11.34}$$

Writing the total diode current as $i = I_{\text{dc}} + \Delta i$ and putting (11.32), (11.33), and (11.34) into (11.31), the second-order small-signal diode model at the bias point (V_0, I_{dc}) is

$$\Delta i = (\Delta v)G_j + \frac{1}{2!}\Lambda(\Delta v)^2 G_j \tag{11.35}$$

The diode junction voltage is now made an amplitude-modulated RF carrier:

$$\Delta v = V[1 + m\cos(\omega_m t)]\cos(\omega t) \tag{11.36}$$

where V = peak carrier amplitude
ω_m = modulation frequency
ω = carrier frequency
m = modulation index, lying in the range $0 \le m \le 1$

Inserting (11.36) into (11.35), the frequencies and relative amplitudes of Table 11.1 are found (Problem 11.3). The first-order terms are linear in diode voltage Δv, but the second-order terms are proportional to $(\Delta v)^2$. Figure 11.7 shows

Table 11.1 Components of Output Spectrum of Detector

Frequency	Relative Amplitude	Relative Amplitude when $m = \frac{1}{2}$, $\Lambda V = 2$	Order of Term
0, detected signal	$1 + \frac{1}{2}m^2$	1.125	Second
ω_m, modulation	$2m$	1.0	Second
$2\omega_m$	$\frac{1}{2}m^2$	0.125	Second
$\omega \pm \omega_m$	$\frac{2m}{\Lambda V}$	0.5	First
ω, carrier	$\frac{4}{\Lambda V}$	2.0	First
$2(\omega \pm \omega_m)$	$\frac{1}{4}m^2$	0.0625	Second
$2\omega \pm \omega_m$	m	0.5	Second
2ω	$1 + \frac{1}{2}m^2$	1.125	Second

the resulting input and output spectra for $m = 0.5$ and $\Lambda V = 2.0$. In practice, the desired output at frequency ω_m is separated from the unwanted components by a low-pass filter (C_L in Fig. 11.4); a blocking capacitor (C_B in Fig. 11.4) removes the DC term.

Theoretical Sensitivity Measures. To calculate the bias dependence and *frequency response* of the detector voltage and current sensitivities γ_v and β_I, one must use a linearized small-signal detector model in which the diode parameters at the bias point are given by $R_j(I_{dc})$, as in (11.7), and the depletion capacitance

$$C_j(V_0) = \frac{C_j(0)}{(1 - V_0/V_{bi})^\gamma} \tag{11.37}$$

where $C_j(0)$ is the zero-bias value, V_{bi} is the built-in voltage, $\gamma \simeq 0.5$, and V_0 is the value of v at bias I_{dc}.

Theoretical Voltage Sensitivity. Under matched conditions, when the modulation frequency is much less than the carrier frequency and I_{dc} comes from an ideal current source, the open-circuit voltage sensitivity is [22]

$$\gamma_v^{oc} = \frac{\Delta V_v^{oc}}{P_{in}} = \frac{\Lambda}{2} \frac{R_j}{1 + \rho(1 + v^2)} \qquad \text{(V/W)} \tag{11.38}$$

where $\rho = R_s/R_j$ and $v = \omega C_j R_j$. The frequency dependence comes from the sharing of input power between R_j and R_s. The voltage sensitivity can also be defined under loaded conditions:

$$\gamma_v^{loaded} = \frac{\gamma_v^{oc}\sigma}{1 + \sigma + \rho} \qquad \text{(V/W)} \tag{11.39}$$

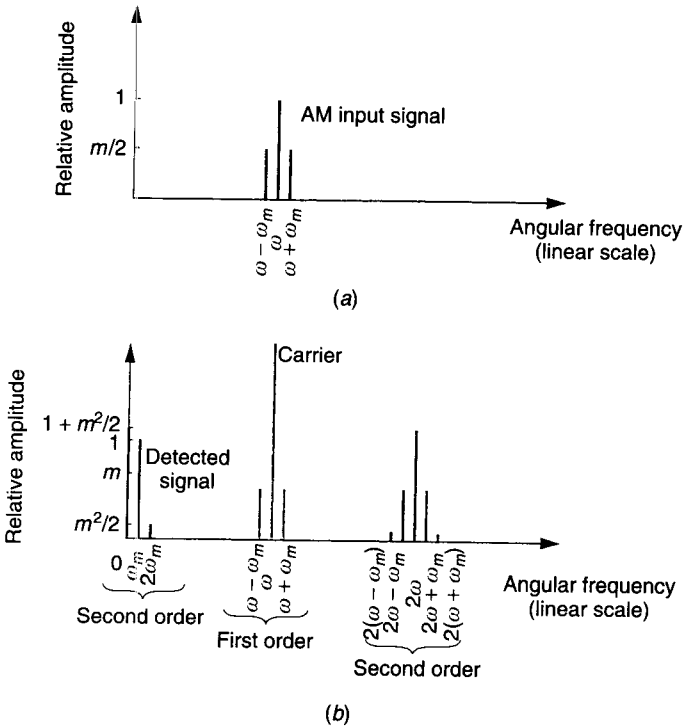


Figure 11.7 Demodulation of AM signal: (*a*) input spectrum; (*b*) output spectrum assuming square-law detector with $m = 0.5$, $\Lambda V = 2.0$.

where $\sigma = R_L/R_j$. Since both R_j and C_j depend on I_{dc}, γ_v^{oc} is bias dependent:

$$\gamma_v^{oc} = \frac{1}{2I_s}\left\{(1+\zeta) + k(1+\zeta)^2 + \frac{1}{k}\left(\frac{f}{f_c(0)}\right)^2\left[1 - \frac{1}{\Lambda V_{bi}}\ln(1+\zeta)\right]^{-2\gamma}\right\}^{-1} \tag{11.40}$$

where $\zeta = I_{dc}/I_s$

$k = \Lambda R_s I_s$

$f_c(0)$ = small-signal diode cutoff frequency, $= [2\pi R_s C_j(0)]^{-1}$.

Figure 11.8 shows the loaded voltage sensitivity as a function of bias and frequency calculated from (11.38) and (11.40). This demonstrates that the bias current can often be optimized for maximum loaded sensitivity. In the unloaded case, there is no such optimum current.

Theoretical Current Sensitivity. From (11.19), $\beta_I = \gamma_v/R_v$, so using (11.38) the current sensitivity is

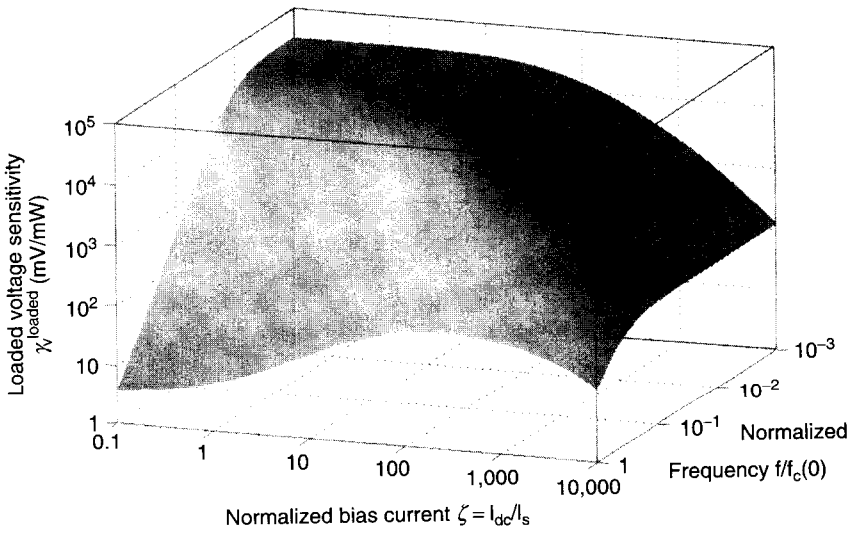


Figure 11.8 Calculated loaded voltage sensitivity γ_V^{loaded}, according to (11.39) and (11.40), as function of frequency f and bias current I_{dc}. Diode parameters [22] are $\gamma = 0.5$, $I_s = 0.2$ μA, $\Lambda = 22\ \text{V}^{-1}$, $R_s = 24\ \Omega$, $V_{bi} = 0.45$ V, and $C_j(0) = 23$ fF. Video load $R_L = 3$ kΩ, and cutoff frequency $f_c(0) = 288.3$ GHz.

$$\beta_I = \frac{\Lambda}{2} \frac{1}{(1+\rho)[1+\rho(1+\nu^2)]}. \tag{11.41}$$

The current sensitivity is maximum when $\rho \to 0$, that is, when $R_s \ll R_j$:

$$\beta_I^{\max} = \frac{\Lambda}{2} = \frac{q_e}{2nkT} \tag{11.42}$$

which is almost independent of any semiconductor properties, except at cryogenic temperatures, and is equal to 19.3 mA/mW at 300 K if $n = 1.0$.

11.2.4 Large-Signal Detector Theory

Detector characteristics change as the incident power level increases. An accurate large-signal analysis using the Ritz–Galérkin (RG) method [29] predicts detector performance over a large dynamic range and shows that, contrary to general belief, slopes can be much steeper than square law. At high levels the results agree with a piecewise-linear analysis.

Ritz–Galérkin Analysis. The RG analysis assumes an ideal exponential diode [see (11.1)] in series with R_s. The incident RF power P_{inc}, distinct from the absorbed power P_{abs} is the power impinging on the input terminals,

regardless of the matching or nonlinearity. Absorbed power P_{abs} is that portion of P_{inc} actually absorbed by the diode. The rest of P_{inc} is reflected. Only under "matched" conditions is P_{abs} equal to P_{inc}. But matching is a linear concept and is meaningless for large-signal operation, so the objective here is to relate P_{inc} to the change ΔV_0 in the video voltage, providing a solution corresponding to what is provided in manufacturer's data sheets. The analysis [29] results in the solution

$$\mathbb{I}_0(X) = \left(1 + \zeta + \frac{Y}{a}\right) \exp[(1+b)Y + k\zeta] \tag{11.43}$$

where $X = \Lambda V_g$ is the normalized RF input voltage
$Y = \Lambda V_0$ is the normalized video output voltage
$a = \Lambda R_L I_s$ is a load parameter
$b = (R_g + R_s)/R_L$ is a resistance ratio
$k = \Lambda R_s I_s$
$\zeta = I_{dc}/I_s$ is a bias current parameter

and $\mathbb{I}_0(X)$ is the zero-order modified Bessel function of the first kind and argument X. A relation between P_{inc} and ΔV_0 can be obtained from (11.43) by noting that P_{inc} is the power that would be absorbed by a conjugately matched linear load:

$$P_{inc} = \frac{V_g^2}{8R_g}. \tag{11.44}$$

From (11.43) and (11.44), the nonlinear algebraic relationship between P_{inc} and V_0 is

$$\mathbb{I}_0(\Lambda\sqrt{8R_g P_{inc}}) = \left(1 + \frac{I_{dc}}{I_s} + \frac{V_0}{R_L I_s}\right) \times \exp\left[\left(1 + \frac{R_g + R_s}{R_L}\right)\Lambda V_0 + \Lambda R_s I_{dc}\right]$$

(nonlinear transfer function of detector) (11.45)

This includes the effect of the bias current I_{dc}. The actual detector response $V_0(P_{inc})$ can be found from (11.45) by calculating the *input* P_{inc} as a function of the *output* V_0. In the absence of input power, the static output $V_0(0)$ is found by solving* (11.45) when $P_{inc} = 0$, that is, when $\mathbb{I}_0(0) = 1.0$. Then

$$\Delta V_0 = V_0(P_{inc}) - V_0(0) \tag{11.46}$$

*The $\mathbb{I}_0(x)$ function can be inverted using the bisection method, an efficient alternative to the Newton root-finding method when there is a single root.

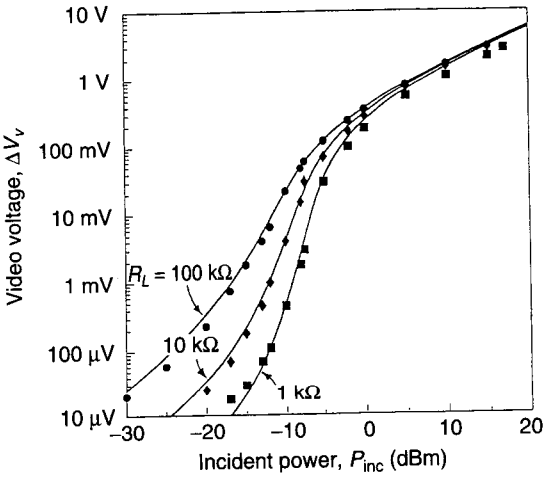


Figure 11.9 Comparison between theoretical and measured (1.8-GHz) detector transfer functions for three different video load resistances. Diode and circuit parameters are $I_s = 1.7 \times 10^{-9}$ A, $n = 1.02$, and $R_s = 10\ \Omega$. Circuit parameters are $R_g = 50\ \Omega$, $I_{dc} = 0$, and $T = 300$ K. Measured data: (●) $R_L = 100$ kΩ; (♦) $R_L = 10$ kΩ; (■) $R_L = 1$ kΩ. (After Harrison and Le Polozec [29]. Reprinted with permission of IEEE.)

The curves in Fig. 11.9 were obtained in this way, using typical diode and circuit parameters.* As the load R_L increases, (11.45) predicts a gradual change from "square law" to "linear" behavior: This could not be predicted by a linear analysis. As R_L is reduced, there is a range of power levels where the slope $\partial(\Delta V_0)/\partial(P_{inc})$ is *much steeper* than square law. In the case illustrated, the maximum slope corresponds to a *tenth-power* law. There is good agreement with measurement over a dynamic range of at least 45 dB.

Voltage Sensitivity. The open-circuit voltage sensitivity $\gamma_v^{oc} = \Delta V_0^{oc}/P_{inc}$ has the power dependence shown by the $R_L = \infty$ curve in Fig. 11.10. The sensitivity γ_v^{oc} is constant at low incident power levels and gradually decreases at higher power levels. For finite R_L, the *loaded* voltage sensitivity γ_v shows a peak near the value of P_{inc} corresponding to the maximum slope of the transfer characteristic in Fig. 11.9. Measured results show similar behavior [30].

Piecewise-Linear (PWL) Analysis. For high signal levels, the circuit can be analyzed by a PWL method. Here the diode is represented as a resistor R_s in series with an ideal switch that is open for reverse bias ($v \leq 0$) and closed for forward bias ($i > 0$). The following transcendental equation is found for $\alpha = V_0/V_g$ (Problem 11.5):

*To get useful results, *accurate* values for the diode parameters are needed. For a suitable method see Waugh [4].

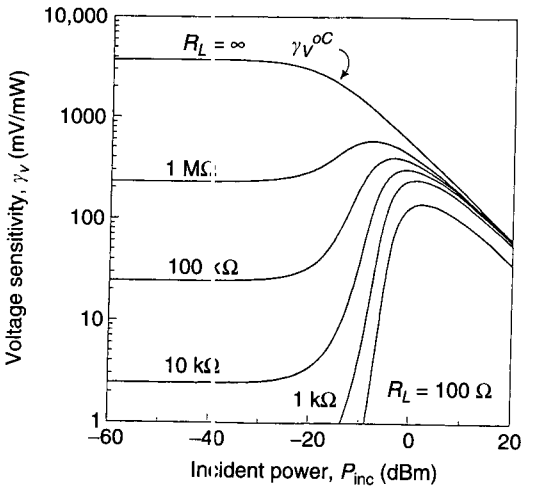


Figure 11.10 Calculated dependence of voltage sensitivity γ_v on P_{inc} and video load R_L. Diode and circuit parameters as in Fig. 11.9. (After Harrison and Le Polozec [29]. Reprinted with permission of IEEE.)

$$\sqrt{1-\alpha^2} = \alpha(\pi b + \arccos \alpha) \tag{11.47}$$

where $b = (R_g + R_s)/R_L$, as before. Using (11.44), the solution for "linear" detector operation is

$$V_0 = \alpha\sqrt{8R_g P_{inc}} \quad \text{(high-level transfer function of detector)} \tag{11.48}$$

where α is the fixed value given by solving (11.47). Note that (11.48) predicts true linear operation of the detector circuit. The two methods agree at high incident power levels (for $R_L = \infty$), but the PWL method gives incorrect results at low power levels.

11.2.5 Detector Design Considerations

Detector design is a compromise between several competing requirements, including operating frequency, bandwidth, sensitivity, dynamic range, and temperature extremes.

Diode Selection. Equation (11.40) shows that for high sensitivity γ_v, the diode cutoff frequency

$$f_c(0) = \frac{1}{2\pi R_s C_j(0)} \tag{11.49}$$

should be large relative to the operating frequency f. Typically, a value

$f_c(0) > 10f$ would be used, corresponding to

$$C_j(0) \sim \frac{1}{20\pi R_s f} \tag{11.50}$$

However, too small a value of $C_j(0)$ would result in a high-Q diode hard to match over a broad bandwidth. For broadband applications, diodes with $C_j(0)$ values two to three times larger are often chosen. Additionally, (11.38) and (11.39) indicate that for highest sensitivity one requires $\rho \ll 1$ and $\sigma \gg 1$, that is, $R_s \ll R_j$ and $R_L \gg R_j$, where R_j is the bias-dependent "junction resistance" given by (11.7). The actual bias current I_{dc} may be selected to maximize γ_v^{loaded}; see Fig. 11.8.

A further consideration is the diode parasitic series inductance L_p. In some designs, the diode is chosen (or designed) such that L_p resonates with $C_j(0)$.

Design Example: Broadband Detector. The following steps are taken to design a planar 6–12-GHz detector [24].

1. *Diode.* In accordance with the above guidelines, we choose

   $$f_c(0) > 120 \text{ GHz}$$

   since $f_{\max} = 12$ GHz. Then, assuming a typical value of 15 Ω for R_s, (11.50) gives

   $$C_j(0) \sim 0.09 \text{ pF}$$

   However, this is a broadband application, so a larger value, $C_j(0) = 0.2$ pF, is selected.
2. *Diode Impedance.* Because octave bandwidth is required, the diode is shunted with a low-Q impedance provided by a quarter-wave open-circuited radial stub. This results in the frequency-dependent locus Z_1 in Fig. 11.11. A small bias ($I_{dc} = 50$ μA) increases the real part of the diode impedance, thereby improving the matching; see locus Z_2.
3. *Matching Circuit.* Adopting the approach of [31] to match Z_2 to 50 Ω, the following components are needed:
   - A series line to rotate Z_2 until the resulting admittance Y_3 has equal susceptances at the band edges. In this example a line with $Z_0 = 85$ Ω and electrical length 44° at 9 GHz is used.
   - A short-circuited stub to reduce the band-edge susceptances to near zero. The required stub impedance is

   $$Z_{st} \approx \left[Y_{st} \tan\left(\frac{\pi}{2}\frac{f_1}{f_2}\right)\right]^{-1} \tag{11.51}$$

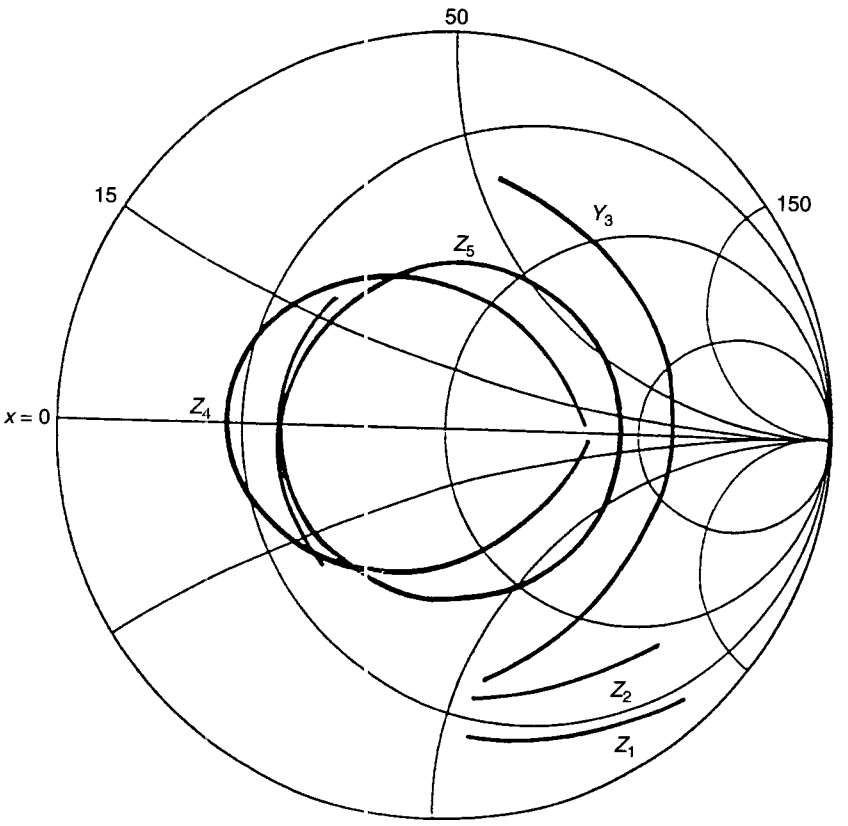


Figure 11.11 Smith chart impedance and admittance trajectories for designing a 6–12-GHz detector. Diode parameters: $C_j(0) = 0.2$ pF, $R_s = 8\ \Omega$, $L_p = 0.09$ nH, $C_p = 0.02$ pF, $I_s = 10^{-7}$ A, $I_{dc} = 50\ \mu$A, $R_j = 518\ \Omega$. (After Irvine, [24]. Reprinted with permission of John Wiley & Sons.)

where f_1 = low-frequency band edge
f_2 = high-frequency band edge
Y_{st} = susceptance at f_1

In this example $Z_{st} = 27\ \Omega$ and the resulting impedance is shown as Z_4.

- A series line to equalize the VSWR over the band. A section with characteristic impedance 43 Ω and electrical length 90° (at 9 GHz) results in the Z_5 impedance trajectory in Fig. 11.11. The reflection coefficient magnitude is $|\Gamma| \approx 0.44$ (VSWR ≈ 2.6) over the band.

The final circuit schematic is given in Fig. 11.12*a*, and a microstrip implementation is shown in Fig. 11.12*b*. Table 11.2 lists the voltage sensitivity γ_v and the value γ_v' corrected for mismatch loss.

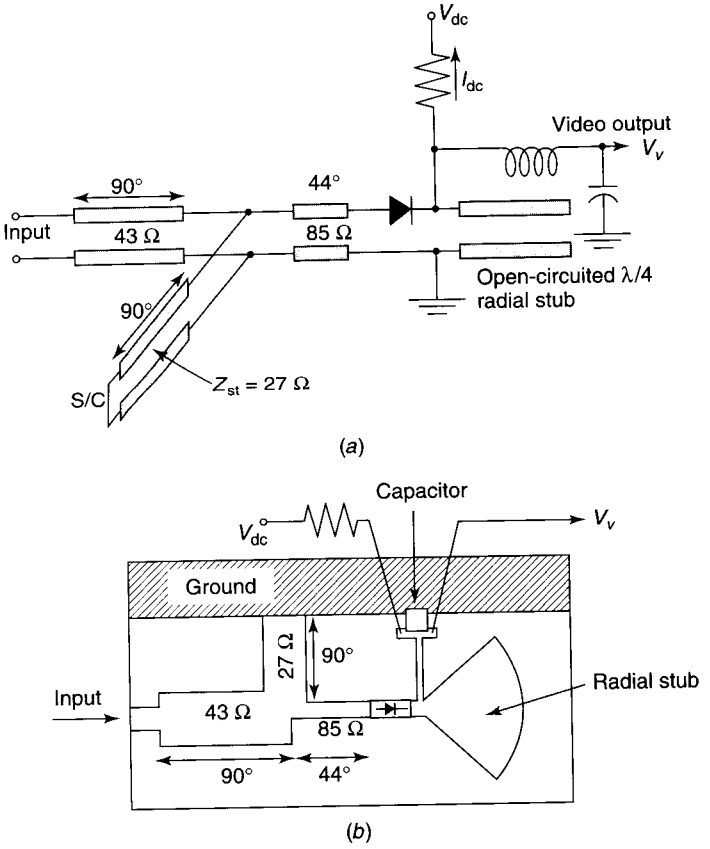


Figure 11.12 A 6–12-GHz detector: (*a*) detector schematic; (*b*) microstrip implementation. Note that an RF choke is not needed in this design because the short-circuited stub provides a return for the bias current I_{dc}. All electrical lengths are quoted for 9 GHz. (After Irvine [24]. Adapted with permission of John Wiley & Sons.)

Table 11.2 Microstrip Detector Performance

Frequency (GHz)	Voltage Sensitivity, γ_v (mV/μW)	Reflection Coefficient Magnitude, $\|\Gamma\|$	Sensitivity (corrected for mismatch loss) γ_v' (mV/μW)
6.0	6.6	0.44	5.3
8.0	5.2	0.44	4.2
10.0	4.3	0.44	3.5
12.0	3.5	0.44	2.8

The required linewidths for a substrate with a particular permittivity ϵ_r and thickness can be obtained from Chapter 2. Computer optimization of the circuit, including discontinuity models, will further improve the performance.

More information on designing Schottky diode detector circuits is given in the literature [31–35], while a PDB detector is described by Fraser [36]. A preamplified monolithic Schottky detector is described by Wang et al. [37]: At 94 GHz the nonamplified detector chip has a voltage sensitivity $\gamma_v = 7.2$ V/mW and a tangential signal sensitivity TSS = −45 dBm. With a two-stage HEMT preamplifier these figures become $\gamma_v = 300$ V/mW and TSS = −62 dBm, also measured at 94 GHz.

11.3 MIXERS

Mixers use a resistive nonlinearity to produce an output spectrum containing the sum and difference frequencies of two input signals.

In receivers, down-converting mixers are used, in which the signal frequency ω_s and the local oscillator (LO) frequency ω_L are both higher than the intermediate-frequency (IF) output signal ω_I; see Fig. 11.13*a*. In this case ω_s is an RF signal, denoted here by ω_R. It is mixed with another RF signal ω_L called the *local oscillator* to produce a low-frequency difference signal $\omega_I = \omega_R - \omega_L$ called the *intermediate frequency* and a higher one $\omega_\Sigma = \omega_R + \omega_L$ called the *sum*. The desired IF is selected by a low-pass filter.

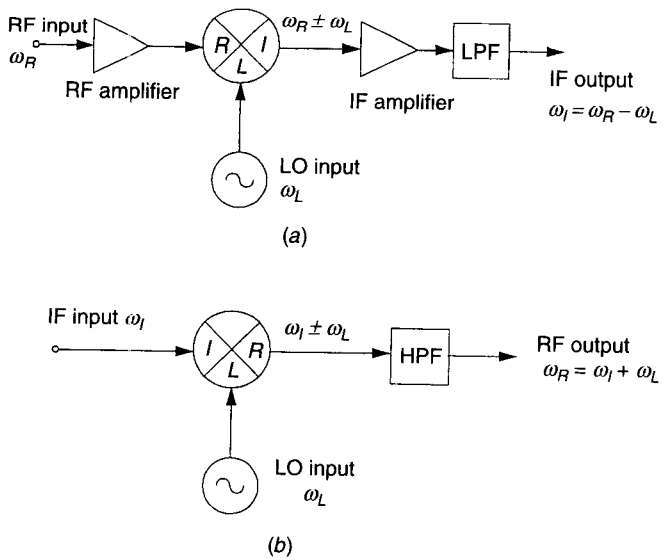


Figure 11.13 (*a*) Mixer used as down converter in receiver. (*b*) Mixer used as up converter in transmitter.

In a transmitter (see Fig. 11.13*b*), the input signal ω_s is at the IF frequency ω_I and an up-converting mixer produces an RF output at the sum frequency $\omega_R = \omega_I + \omega_L$. The unwanted difference $\omega_I - \omega_L$ is filtered out.

11.3.1 Mixer Spectral Products

The i–v characteristic of an arbitrary single-valued nonlinear resistor (mixer diode) can be represented by the general polynomial

$$i = a_0 + a_1 v + a_2 v^2 + a_3 v^3 + \cdots + a_N v^N \tag{11.52}$$

where a_0 accounts for any DC bias current. This expression includes the ideal exponential diode, for which

$$i = I_s \left(\sum_{N=0}^{\infty} \frac{1}{N!} (\Lambda v)^N - 1 \right) \tag{11.53}$$

Let the RF input signal be

$$v_R(t) = V_R \sin(\omega_R t) \tag{11.54}$$

and the LO signal be

$$v_L(t) = V_L \sin(\omega_L t) \tag{11.55}$$

Then there will be an infinite series of mixing products given by

$$\begin{aligned} i = {} & a_0 + a_1[V_R \sin(\omega_R t) + V_L \sin(\omega_L t)] + a_2[V_R \sin(\omega_R t) + V_L \sin(\omega_L t)]^2 \\ & + \cdots + a_N[V_R \sin(\omega_R t) + V_L \sin(\omega_L t)]^N \end{aligned} \tag{11.56}$$

Thus the output contains a DC term (usually ignored)

$$I_{\text{dc}} = a_0 + \tfrac{1}{2} a_2 (V_R^2 + V_L^2) + \cdots$$

as well as the original signals. The primary mixing products $\omega_L \pm \omega_R$ come from the *second-order* term:

$$\begin{aligned} & a_2 \{ \tfrac{1}{2} V_R^2 [1 - \cos(2\omega_R t)] + V_R V_L [\cos(\omega_R - \omega_L)t + \cos(\omega_R + \omega_L)t] \\ & \quad + \tfrac{1}{2} V_L^2 (1 - \cos(2\omega_L t)) \} \end{aligned} \tag{11.57}$$

and their amplitudes are proportional to coefficient a_2. This term also generates the second-order harmonics $2\omega_R$ and $2\omega_L$. The third- and higher order terms

generate higher order harmonics and mixing products of the form

$$\omega_{M,N} = M\omega_L \pm N\omega_R \tag{11.58}$$

where M, N are positive integers.

The desired mixer output is normally the IF, denoted here by ω_I. In general, ω_I can be the lower sideband $\omega_L - \omega_R$ or the upper sideband $\omega_L + \omega_R$. The lower sideband is usually the one of interest and the term "mixer" generally implies this. The two possible situations are as follows:

First Case: $\omega_R < \omega_L$. In this case Fig. 11.14*a* shows, in addition to the desired ω_I output, two other frequencies of importance: the *image* frequency $2\omega_L - \omega_R$ and the upper sideband, or *sum*, frequency $\omega_R + \omega_L$.

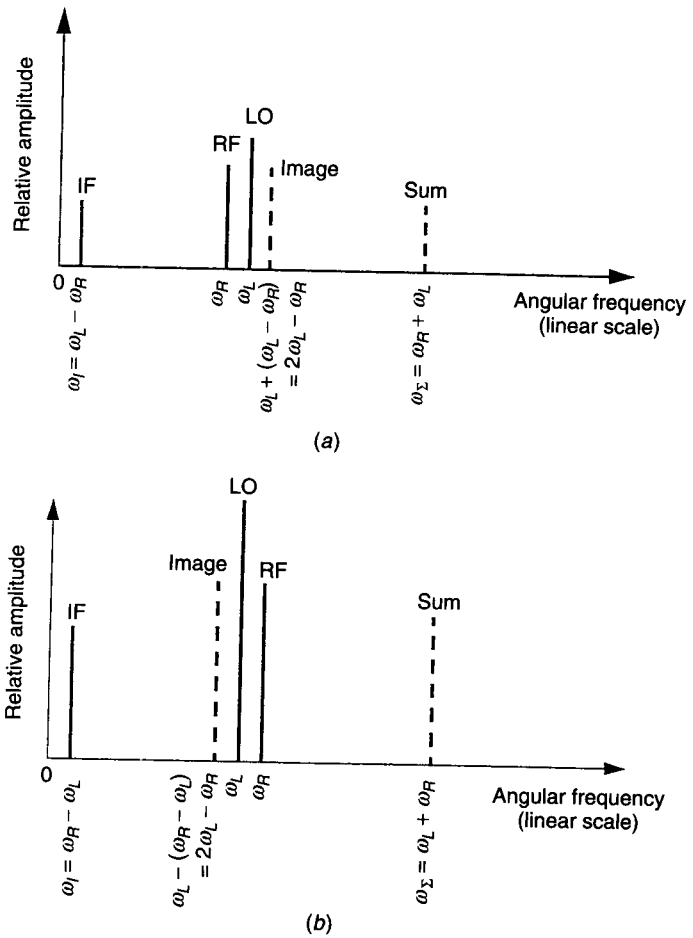


Figure 11.14 Mixer spectra when (*a*) $\omega_R < \omega_L$ and (*b*) $\omega_R > \omega_L$. Generally, ω_I is small compared with either ω_R or ω_L.

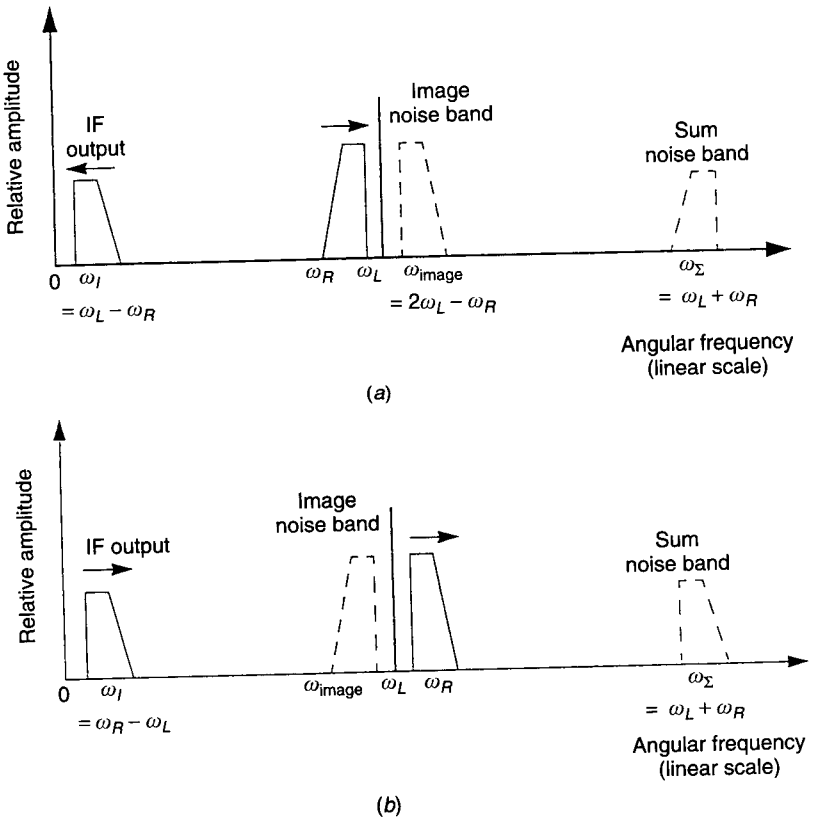


Figure 11.15 Mixer spectra for finite-bandwidth RF signal. (*a*) When $\omega_R < \omega_L$; increasing RF frequency results in *decreasing* IF frequency. (*b*) When $\omega_R > \omega_L$; increasing RF frequency results in *increasing* IF frequency.

Second Case: $\omega_R > \omega_L$. Here the IF, image, and sum frequencies are as shown in Fig. 11.14*b*.

In either case the objective is to produce ω_I with the least amount of conversion loss (L_c), noise, and nonlinear distortion (which implies the generation of spurious frequencies). When the RF signal has a finite bandwidth, the corresponding spectra are as in Figs. 11.15*a*,*b*.

11.3.2 Image-Enhanced Mixers

A very broadband down-converting mixer produces the sum (ω_Σ) and IF (ω_I) frequencies with equal efficiencies; for example, see Fig. 11.14*b*. This results in the conversion of *equal amounts* of RF power to the ω_Σ and ω_I frequencies. One of the objectives of mixer design is to match the desired ω_I power to the

input of the IF amplifier. The unwanted frequencies ω_Σ and ω_{image} would then be reflected back toward the RF input. However, if suitable filters are included in the microwave realization, the ω_Σ and the ω_{image} sidebands can be reflected back to the diode for *another mixing operation*. This is a good idea because these sidebands contain useful power, and their reuse results in reduced conversion loss L_c. This technique is the basis of the so-called image-enhanced and image-and-sum-enhanced mixers.

To see how the enhancement works, consider the result of mixing the LO and image signals. With appropriate phasing,

$$\omega_L - \omega_{\text{image}} = \omega_L - (2\omega_L - \omega_R) = \omega_R - \omega_L = \omega_I$$

as desired. This operation is called *image enhancement* or *image recovery*. For *sum enhancement*, the mixing process of interest is

$$\omega_\Sigma - 2\omega_L = \omega_R + \omega_L - 2\omega_L = \omega_R - \omega_L = \omega_I$$

again resulting in the desired IF frequency.

To obtain enhancement in a practical mixer, the circuit is so arranged that the various contributions to the IF signal are all *in phase* with the principal frequency transformation $\omega_R - \omega_L$; see Section 11.3.7. The result is an increase in the net output power P_{out} and a decrease in the conversion loss L_c.

11.3.3 Conversion Loss and Conversion Gain

The sensitivity of a microwave receiver using a mixer is limited in that only part of the available RF input is actually converted into IF power. The conversion loss L_c of a mixer is defined as

$$L_c = 10 \log_{10}\left(\frac{P_R^{\text{avail}}}{P_I}\right) \quad \text{(dB)} \tag{11.59}$$

where P_R^{avail} is the available RF input power and P_I is the IF output power. This definition implies that P_R is much less than the LO power P_L. It gives L_c as positive decibels for lossy mixers and is a measure of the efficiency of conversion of RF energy into IF energy.

Mixer conversion loss (or gain) depends on the nature of the nonlinear elements used in the mixer as well as their circuit embedding. Conversion loss occurs when the elements are positive nonlinear resistors; conversion gain is possible when the v–i characteristic includes negative-resistance regions, as in RTDs, or when *active* three-terminal nonlinear devices such as MESFETs and HEMTs are used. Here we discuss conversion loss, with the understanding that the results apply also to conversion gain.

In a *single-sideband* (SSB) system only one sideband is converted to IF, so that half (3 dB) of the conversion loss is inevitable. The remaining losses can be

minimized by using sufficient LO drive to make the mixer diodes approximate ideal switches. An ideal switch has zero impedance in the "on" state and infinite impedance in the "off" state.

In a *double-sideband* (DSB) system, on the other hand, an IF output is generated for RF signals both above and below the LO frequency; the RF inputs are "mutual images" [38]. Such mixers are often used in SSB receivers and as radiometers in radio astronomy applications.

Although some parametric, superconducting, and active mixers can provide conversion gain rather than conversion loss, there will always be significant loss components due to (i) input and output mismatches, (ii) device and circuit losses, and (iii) parasitics associated with the nonlinear device. All of these losses can be functions of power, frequency, and bias.

The first component of L_c, *mismatch loss*, depends on the matching at the RF and IF ports. Mismatch at the input reduces the available RF power at the diode while output mismatch reduces the IF power. The overall mismatch loss can be expressed as (Problem 11.7)

$$L_1 = 10\left[\log_{10}\frac{(1+S_R)^2}{4S_R} + \log_{10}\frac{(1+S_I)^2}{4S_I}\right] \quad \text{(dB)} \qquad (11.60)$$

where S_R and S_I are the VSWRs at the RF and IF ports, respectively.

The second component of L_c is *conversion loss* within the diode itself. This depends primarily on the i–v characteristic and also on the circuit embedding: the conditions at the RF and IF ports. There are several ways of treating this component, here denoted L_2, some of which will be discussed below. It can be shown [38, 39] that L_2 approaches a minimum value of

$$L_2^{\min}(\text{DSB}) = 3 \text{ dB}$$

for non-enhanced broadband mixers, even using ideal (perfect-switch) diodes. Thus in the most efficient possible DSB mixers the unused mixing products would be reactively terminated so that real power could flow at the RF, image frequency, and IF only. The lost energy would be equally divided between conversion to the image frequency and reflection loss at the RF port (3:1 VSWR mismatch) [38]. In a broadband scenario where matched loads are presented to the unused mixing products, it can be shown [24, 38] that the minimum is given as

$$L_2^{\min}(\text{DSB}) = 20\log_{10}(\pi/2) = 3.92 \text{ dB}$$

This will be discussed in Section 11.3.4. The above values represent absolute minima for the L_2 of nonenhanced mixers. For *image-enhanced* mixers, in which the image frequency is provided with a reactive (lossless) termination so that the frequency transformation described in Section 11.3.2 is achieved, a

minimum conversion loss of $L_2 = 0$ dB is theoretically obtainable if the image is either open or short circuited [39].

The third component of L_c is *RF signal loss due to device parasitics*. Here, parasitics are features that detract from an ideal nonlinear resistor. For a diode, the important ones are the junction capacitance $C_j(v)$ and the series resistance R_s. The diode "parasitic loss" is the ratio of the available RF input power to the power actually delivered to the diode junction:

$$L_3 = 10 \log_{10}\left(\frac{P_R}{P_{\text{jct}}}\right) \quad \text{(dB)} \tag{11.61}$$

If the LO power is small enough for the time-averaged junction resistance $\langle R_j \rangle$ and depletion capacitance $\langle C_j \rangle$ to be meaningful concepts, then L_3 can be expressed as [40–42]

$$L_3 = 10 \log_{10}[1 + \rho(1 + v^2)] \quad \text{(dB)} \tag{11.62}$$

where $\rho = R_s/\langle R_j \rangle$ and $v = \omega \langle C_j \rangle \langle R_j \rangle$ are both drive-level dependent.

The overall mixer insertion loss is the decibel sum of the mismatch, device, and parasitic losses:

$$L_c = L_1 + L_2 + L_3 \quad \text{(dB)} \tag{11.63}$$

The measured L_c of SSB mixers is usually slightly better than for DSB mixers [40].

11.3.4 Ideal Mixers

Saleh [38] investigated resistive mixers in which the diode is a nonlinear resistor whose effective resistance $\partial v/\partial i$ is modulated (pumped) by a large-amplitude LO signal. The diode is then represented as a linear time-varying resistor* $r(t)$ or conductance $g(t)$. Such representations remain valid when an RF signal is simultaneously applied, provided that the RF power is small compared with the LO power. He classified mixer circuits into Z, Y, H, and G types, according to the type of transfer matrix relating the terminal voltages and currents. Figure 11.16 shows three equivalent ideal Z mixers. Since the function of the LO signal is only to create the $r(t)$ waveform, it need not be shown explicitly. "Ideal" bandpass filters allow currents to flow only at the specified frequencies. Ideal Y mixers are effectively duals of ideal Z mixers.

Conversion loss L_c is minimized when $r(t)$ or $g(t)$ is a *square wave* [39]. In the limiting case of alternation between infinite and zero resistance, the mixer diode acts as an ideal switch. This can be approached in reality when the LO

*A lowercase r is used to emphasize that it is a *time-varying* resistance.

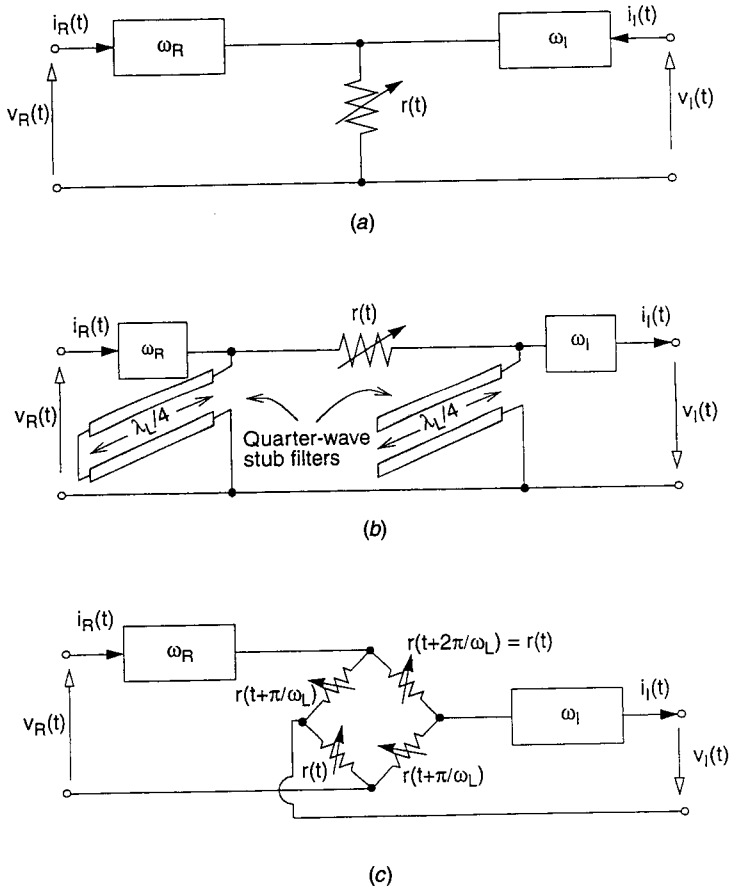


Figure 11.16 Three equivalent linear-time-varying-resistance Z mixers: (*a*) shunt Z mixer; (*b*) series Z mixer; (*c*) balanced Z mixer. The boxes represent ideal bandpass filters at the indicated frequencies. In each case the time-varying resistors are controlled by the LO at frequency ω_L (not shown). (From Saleh [39]. Adapted with permission of MIT Press.)

frequency is low enough and the LO power large enough for the IF output level to become nearly independent of the LO level. Under these conditions the diode is being switched fully on and fully off.

Irvine [24] considered a version of the Z mixer of Fig. 11.16*b* where $r(t)$ is replaced by an ideal switch. As seen in Fig. 11.17, when the duty ratio p approaches zero, the minimum possible L_c also approaches 0 dB regardless of the ratio $R_{\max}/R_{\min}$, and the waveform approaches a train of impulse functions, an extreme idealization. Figure 11.17 shows that $L_c^{\min}$ is actually a strong function of p [43].

For any Z or Y mixer, including ring configurations, one can investigate ultimate theoretical limits on L_c and on intermodulation (IM) suppression by

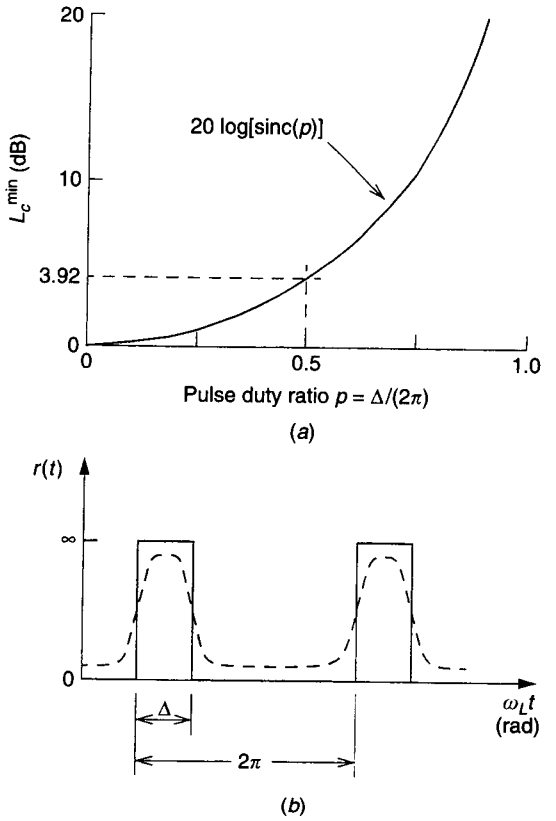


Figure 11.17 (*a*) Minimum possible conversion loss of ideal-switch mixer. (*b*) Time-varying resistance waveform for ideal-switch mixer (—) and more realistic resistance waveform (– – – –).

replacing the $r(t)$'s or $g(t)$'s by ideal switches. For example, if $p = 0.5$, the mixer output is found by multiplying $v_R(t)$ by the Fourier expansion of a square-wave function at the LO frequency:

$$\begin{aligned} v_{\text{out}}(t) &= V_R \sin(\omega_R t)\frac{4}{\pi}\sum_{N=\text{odd}}\frac{1}{N}\sin(N\omega_L t) \\ &= \frac{4}{\pi}V_R\{\tfrac{1}{2}[\cos(\omega_L-\omega_R)t-\cos(\omega_L+\omega_R)t] \\ &\quad +\tfrac{1}{6}[\cos(3\omega_L-\omega_R)t-\cos(3\omega_L+\omega_R)t]+\cdots\} \end{aligned} \tag{11.64}$$

For the $\omega_I = \omega_L - \omega_R$ and $\omega_\Sigma = \omega_L + \omega_R$ products the conversion loss is

$$L_C^{\min} = 10\log_{10}\left(\frac{P_R^{\text{avail}}}{P_I^{\text{avail}}}\right) = 20\log_{10}\left[\frac{V_R}{(2/\pi)V_R}\right] = 20\log_{10}(\pi/2) \equiv 3.92 \text{ dB} \tag{11.65}$$

For an *arbitrary* duty ratio p one finds that [24]

$$L_c^{\min} = \frac{\sin(\pi p)}{\pi p} \equiv 20 \log_{10}[\text{sinc}(p)] \qquad \text{(dB)} \tag{11.66}$$

as plotted in Fig. 11.17*a*. The corresponding time-varying resistance waveform for the ideal-switch mixer is shown in Fig. 11.17*b*.

Using an S parameter approach, Kelly [39] studied the more general case of DSB ideal-switch mixers. A DSB mixer is one in which the IF can be generated for ω_R either above or below ω_L; such mixers are basic components of SSB receivers and radiometers. Beyond specifying reactive idler terminations, he made no assumptions about the nature of these terminations or about p. He showed that the ultimate $L_c^{\min}$ for a DSB mixer is 3.0 dB, the lost power being shared equally between frequency conversion to the image frequency and reflection at the RF port. He also considered the *broadband* DSB mixer, where the spectra of ω_R and the idlers actually overlap. In this case it is impossible to terminate all the idlers reactively. Instead, matched loads are provided at the idler frequencies. He showed that in this case the conversion–gain scattering parameter $S_{21} = (2/\pi)\sin(\pi p)$ is maximized by setting $p = 0.5$, so that

$$L_c^{\min} = 20 \log_{10}(S_{21}) = 3.92 \text{ dB} \tag{11.67}$$

This is a *general* ultimate limit for any resistive (conductive) mixer where all the idler frequencies are match terminated.

11.3.5 Exponential-Diode Mixers

When high LO power is unavailable, exponential-diode mixer models are more suitable than ideal-switch models. Schottky i–v characteristics can be exponential over a seven-decade range of currents, making possible accurate closed-form calculations [43].

Exponential-diode mixer theory was investigated by Saleh [38]. Figure 11.18 shows an example of a Y mixer using an exponential diode. As before, only the

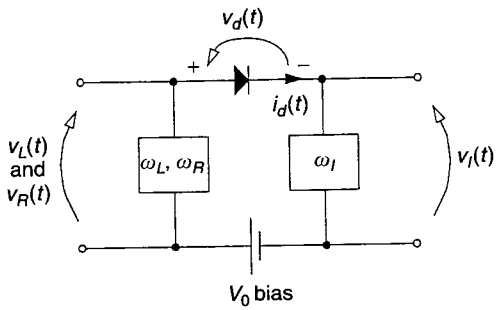


Figure 11.18 A Y mixer using an exponential diode. The boxes denote ideal notch filters that present zero impedance at the indicated frequencies but open circuits at all other frequencies.

LO and IF ports are shown. The time-varying conductance $g(t)$ is obtained by pumping the diode with the LO. If $V_L \gg V_R$, the RF signal sees the diode approximately as a linear, periodically time-varying resistor. In practice, some nonlinear distortion always exists. We wish to find the $g(t)$ waveform. Because of the ideal filters, the diode voltage v_d consists of a DC term and a fundamental at ω_L only (since $V_L \gg V_R$):

$$v_d(t) = V_0 + v_L(t) = V_0 + V_L \cos(\omega_L t) \tag{11.68}$$

Representing the diode by (11.1) and making use of the integral

$$\frac{1}{2\pi}\int_0^{2\pi} e^{x\cos(\theta)} \cos(N\theta)\, d\theta = \mathbb{I}_N(x) \tag{11.69}$$

where $\mathbb{I}_N(x)$ is the modified Bessel function of order N and argument x, it can be shown (Problem 11.8) that if $\Lambda V_L \gtrsim 20$, the diode current has a DC component

$$I_{\text{dc}} \cong I_s \frac{e^{\Lambda(V_0+V_L)}}{\sqrt{2\pi\Lambda V_L}} \tag{11.70}$$

and an LO component

$$I_L \cong 2I_{\text{dc}} \tag{11.71}$$

The time-varying conductance waveform is found to be

$$\begin{aligned} g(t) = \frac{di_d}{dv_d} &= \Lambda I_s e^{\Lambda\{V_0 + V_L\cos(\omega_L t)\}} \\ &= g_0 + 2g_1\cos(\omega_L t) + 2g_2\cos(2\omega_L t) + \cdots \end{aligned} \tag{11.72}$$

where

$$g_N = \Lambda I_s e^{\Lambda V_0} \mathbb{I}_N(\Lambda V_L) \qquad N = 0, 1, 2, \ldots \tag{11.73}$$

Figure 11.19 shows conductance waveforms for $\Lambda V_L = 30$. All Fourier components $g(t)$ are in phase because of the purely resistive nonlinearity.

The Saleh theory [39] gives different conversion loss minima, depending on whether the image frequency $\omega_{\text{image}} = 2\omega_L - \omega_R$ is short circuited or open circuited.

1. *Short-Circuited Image.* In this case, if $\Lambda V_L \gg 1$ (it usually is), then the minimum L_c is approximately

$$L_c^{\min} \cong 1 + \frac{2}{\sqrt{\Lambda V_L}} \tag{11.74}$$

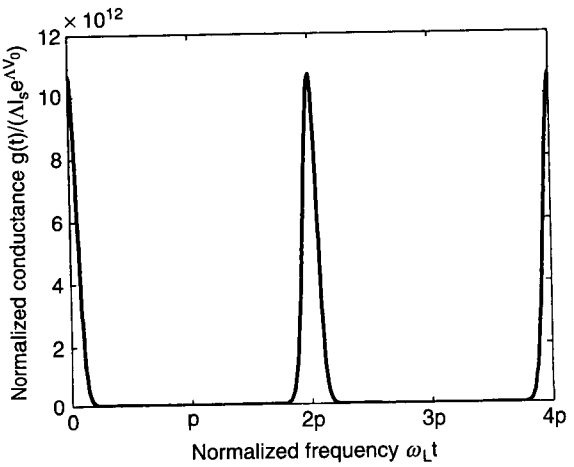


Figure 11.19 Time-varying conductance waveform for ideal-exponential diode mixer when $\Lambda V_0 = 0$, $\Lambda V_L = 30$. The waveform resembles a train of impulse functions.

2. *Open-Circuited Image.* Here, again assuming $\Lambda V_L \gg 1$,

$$L_c^{\min} \cong 1 + \sqrt{\frac{2}{\Lambda V_L}} \tag{11.75}$$

3. *Broadband Condition.* If ω_R and ω_{image} are provided with equal matched terminations, then the mixer operates under a broadband condition. For $\Lambda V_L \gg 1$

$$L_c^{\min} \cong 2\left(1 + \frac{\sqrt{2}}{\Lambda V_L}\right) \tag{11.76}$$

The nonapproximated results for these three conditions are plotted in Fig. 11.20. The parameter L_c always decreases with increasing V_L, so that $g(t)$ approaches a train of impulse functions. For high LO drive levels $L_c^{\min}$ tends toward 0 dB for both the short-circuited and open-circuited image conditions but towards 3 dB for the broadband matched termination condition.

Note that according to (11.70) a significant amount of RF input power gets converted to DC, as it does in a detector. This second-order effect is *generally missing from classical mixer analyses.*

11.3.6 Mixer Performance Measures

Intermodulation in Mixers. Intermodulation (IM) is familiar as a sensitive linearity test for microwave amplifiers. It can also be used to test linearity in mixers. In both amplifiers and mixers IM is usually a consequence of saturation

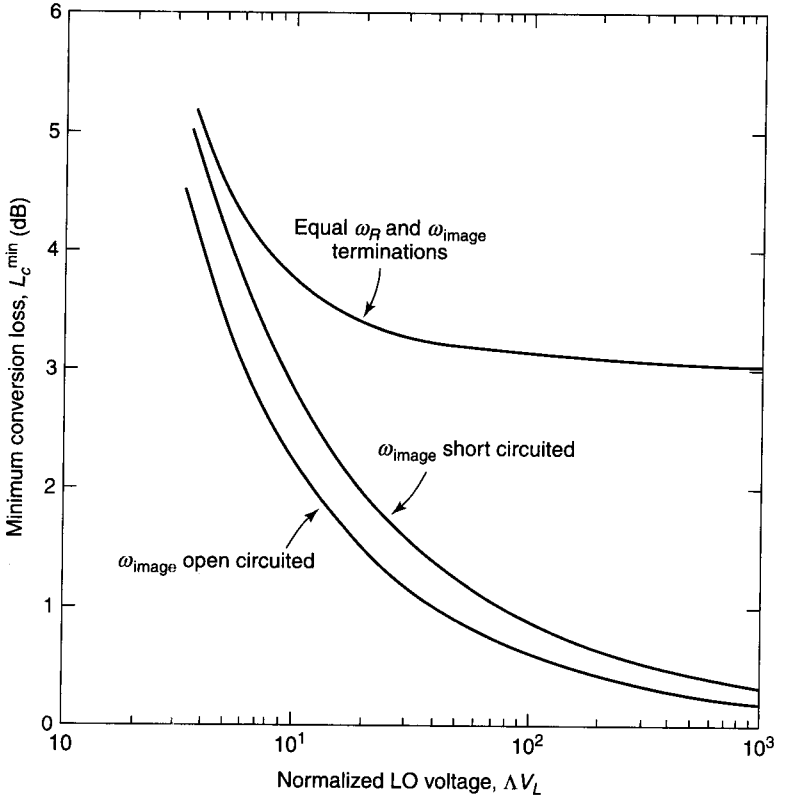


Figure 11.20 Minimum possible conversion loss L_c^{min} Y mixer using single exponential diode for three different termination conditions. (After Saleh [39]. Reprinted with permission of MIT Press.)

in the input–output transfer characteristic, or more generally some variation of gain (or attenuation) and/or phase with input power level.

Regarding a mixer as a nonlinear device with inputs at ω_R and ω_L, *all* its outputs, apart from harmonics and possible subharmonics, are really IM products. For small-signal RF inputs $P_R \ll P_L$, the frequency translation process from ω_R to the desired mixer output $\omega_L - \omega_R$ or $\omega_R - \omega_L$ is a linear operation in the sense that the IF output P_I is proportional to P_R. Under this condition, mixers can be analyzed by treating the LO signal as the sole forcing function and then investigating how the resulting time-varying phenomena interact with a small RF input signal. This results in the classical "small-signal" or "linear" theory of mixers, in which IM effects need not be considered. But if P_R approaches P_L, then in common with other nonlinear devices, the mixer transfer function will exhibit saturation. This causes intermodulation distortion (IMD) products that corrupt the IF output spectrum. Small-signal assumptions then collapse.

Unwanted mixing products like $M\omega_L - N\omega_R$ or $M\omega_R - N\omega_L$ (M, N integers) are called "single-tone" or "harmonic" IMD products because they occur with only one RF input present, but they are really the same kind of thing as the IMD products generated by a saturating amplifier under two-tone excitation. They are sometimes also referred to as "idlers" (not to be confused with the idlers used in parametric frequency converters). If $N = M$, the IMD products are harmonics of ω_I, since

$$N\omega_I = N\omega_L - N\omega_R$$

These products can cause trouble if the IF bandwidth is wide and have been investigated by Orloff [44] for a single-exponential-diode mixer.

If there are two closely spaced RF inputs, then high-order two-tone IMD products can appear. This corresponds to a nonlinear network producing IMD products when driven by *three* simultaneous input signals: two (or more) small ones and one big one. Thus if $\omega_L > (\omega_{R1}, \omega_{R2})$, as in Fig. 11.21, then the following IMDs can appear at the IF output in addition to the wanted signals $\omega_L - \omega_{R2}$ and $\omega_L - \omega_{R1}$:

$$\omega_L - [N\omega_{R1} - (N-1)\omega_{R2}] \quad \text{(upper IM sidebands)}$$
$$\omega_L - [N\omega_{R2} - (N-1)\omega_{R1}] \quad \text{(lower IM sidebands)}$$

with $N = 2, 3, 4, \ldots$. For simplicity, Fig. 11.21 omits the ordinary two-tone IM frequencies caused by cross-products between f_{R1} and f_{R2} (such spectral lines

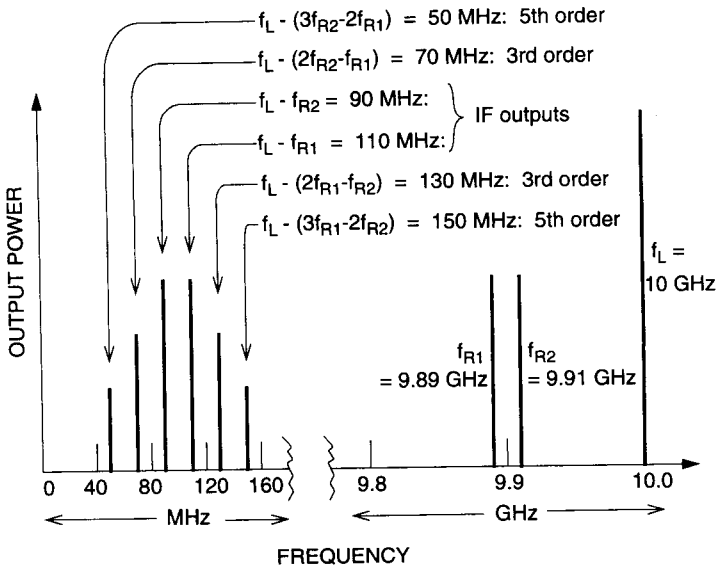


Figure 11.21 Example of two-tone IMD spectrum in mixer with $f_L > (f_{R1}, f_{R2})$. Desired outputs $f_L - f_{R1}$ and $f_L - f_{R2}$ are shown, together with unwanted third- and fifth-order IMD products.

would be interleaved with the ones shown). The most significant two-tone IMDs are the third-order ones $\omega_L - (2\omega_{R1} - \omega_{R2})$ and $\omega_L - (2\omega_{R2} - \omega_{R1})$ because they appear right beside the desired IF outputs and can easily occur within the IF bandwidth.

The details of a spectrum such as Fig. 11.21 depend on the precise shape of the nonlinear transfer function of the mixer as well as the frequencies and absolute power levels of the various input signals. A brief discussion is given by Kollberg [45]. The quantitative prediction of the magnitudes of the various spectral lines is much more difficult than calculating their frequencies, particularly when there are more than two RF input frequencies.

Dynamic Range and Linearity. The input-power/output-power diagram for a single-input mixer (Fig. 11.22) resembles that of a saturating amplifier. The spurious-free dynamic output power range extends from the threshold thermal noise level up to the level at which IMD products begin to exceed the threshold level. Normally the third-order single-tone IMDs are the first spurious frequencies to appear as the RF input power P_R is increased. The level at which these products appear depends on the LO level, the diode characteristics, and the type and design of the mixer. Figure 11.22 also shows the 1-dB compression point, which is the input P_R level for which L_c increases by 1 dB. It also shows the theoretical third-order intercept point, the value of P_R for which the linearly extrapolated characteristics of the primary response and of the third-order IM response intersect. Normally the mixer is operated with P_R below the 1-dB

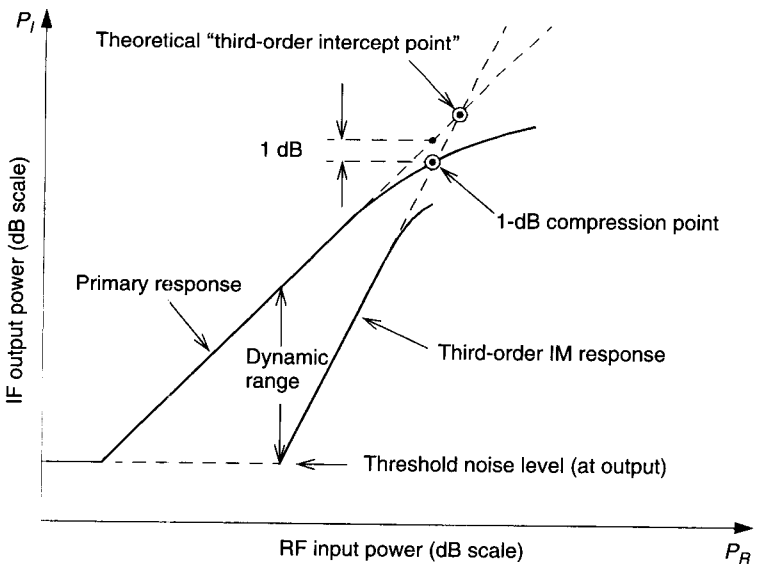


Figure 11.22 The dynamic range of a mixer is typically limited by the appearance of third-order IM products. The level at which this occurs depends on the mixer type, the diode characterixtics, and the LO level. Operation is usually restricted to the region below the 1-dB compression point.

compression point. The greater the value of the third-order intercept point, the better the suppression of the third-order IMD. Typical values are 9–11 dB above the 1-dB compression point.

If the mixer diodes are represented as switches at high power levels, the compression (departure from linearity) of the response can be attributed to the diode conduction angle becoming dependent on the RF power as well as on the LO level [46].

Harmonic Suppression. When a mixer is pumped by the LO, harmonics of ω_L are generated by the resistive nonlinearity. According to the Page–Pantell inequality (see Chapter 13), the *power* in each harmonic is, at most, proportional to $1/N^2$, where N is the order of the harmonic. However, there is no guarantee that the harmonic *currents* will be proportional to $1/N^2$ [47].

In subharmonically pumped mixers (Section 11.3.7), a specific harmonic of ω_L, usually the second, is required for proper mixer operation. Otherwise an attempt is made to suppress harmonic outputs by exploiting mixer symmetry or, as a last resort, by filtering.

In receiver applications, it is important to prevent harmonics of ω_L escaping from the input (RF) port, whence they could be radiated into the environment via the antenna.

A further consideration is that noise at or near the harmonics of ω_L can be down converted to the IF band, thereby degrading the mixer noise figure. This can be minimized either by employing a harmonic-enhanced design or by intentional harmonic filtering using the inevitable device parasitics.

Isolation. Isolation between mixer ports is the attenuation experienced by a signal fed in at one port and measured at another. High isolation means low leakage between ports. In single-ended mixers (Section 11.3.7) isolation depends on filters; whereas in the various forms of balanced mixers isolation is achieved through circuit symmetry. The important isolation parameters are the following:

1. LO-to-RF isolation: the attenuation of the LO drive level P_L, as measured at the RF port, with the IF port terminated in $Z_0 = 50\ \Omega$.
2. LO-to-IF isolation: the attenuation of the LO drive level as measured at the IF port, the RF port being terminated in $Z_0 = 50\ \Omega$.

Normally, only the LO-to-X isolations are specified since $P_L \gg P_R$.

Isolation measurements are usually done with P_L set so that the diodes are in their normal operating region. The RF-to-LO isolation can be measured by injecting a large RF signal and terminating the LO port in 50 Ω.

RF Matching. The RF input VSWR expresses how well a mixer circuit is matched to the RF source at a particular combination of ω_R and P_L. This is important for minimization of overall mixer conversion loss. In mixers where

the RF and LO inputs share a common port, good matching reduces the possibility of LO power being radiated from the antenna.

IF Matching. The IF impedance Z_{IF} is the effective impedance of the pumped mixer diodes and associated circuitry, considered a Thévenin source. It is a function of both P_L and the RF impedance. If DC bias is applied to the diode(s), the effective conduction angle will be altered, thereby changing Z_{IF}. Good IF matching is a further requirement for minimum conversion loss. For least overall receiver noise figure (Section 11.3.8), the IF amplifier should be specified to have its minimum noise figure for a source impedance equal to Z_{IF}. Poor IF matching can degrade the effective noise figure of the following IF amplifier [48].

11.3.7 Types of Mixer Circuits

Mixers are typically used in receiver front ends. For a low system noise figure, a microwave GaAs FET or HEMT low-noise amplifier (LNA) may be placed ahead of the mixer. At millimeter and sub-millimeter wavelengths where LNAs are not available, it can be advantageous to cool mixers to cryogenic temperatures. However, when maximum dynamic range is needed, it is best to apply the RF directly to the mixer. The mixer then establishes the noise figure, sensitivity, bandwidth, and dynamic range of the entire receiver. An upper limit to the dynamic range is fixed by the maximum RF power P_R for which (a) the IF output power P_I remains below the 1-dB compression point (Section 11.3.6) and (b) undesired signals within the IF bandwidth (such as intermodulation products) are held below the tangential signal sensitivity (TSS) noise level.

Single-Ended Mixers. Single-ended mixers are typically based on the single-diode Z mixer. An example, corresponding to Fig. 11.16*b* is given in Fig. 11.23. Here a directional loop filter provides isolation between the LO and RF ports.* Short-circuited and open-circuited $\lambda_L/4$ stubs provide reactive terminations to the idler frequencies $\omega_N = N\omega_L + \omega_I$. The RF bandpass filter rejects out-of-band input signals that might otherwise mix with ω_L to produce spurious signals at the IF port. It also prevents residual LO power from escaping from the RF port. Further practical design considerations for single-ended mixers are discussed elsewhere [48].

Image-and-Sum Enhanced Microstrip Single-Ended Mixer. Figure 11.24 depicts a planar fixed-LO, 12-GHz image-and-sum enhanced single-ended mixer. The directional loop filter [49, 50] is sharply selective at ω_L; it provides some LO-to-RF isolation and reduces the effect of AM noise coming from the LO source on the overall mixer noise figure. The low-pass sum-frequency filter

*RF-to-LO isolation is often important, for example to prevent LO power leaking into the environment via the receiver antenna.

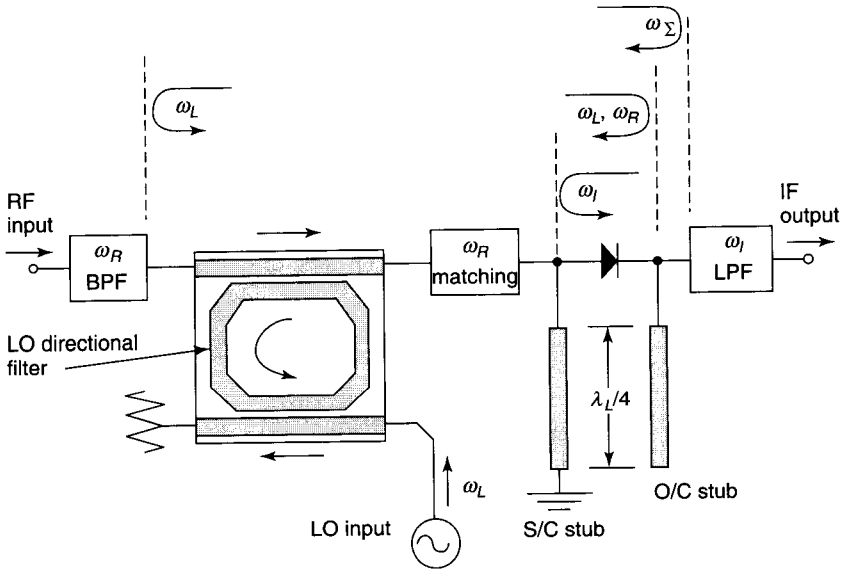


Figure 11.23 Practical realization of series Z mixer. Compare with Fig 11.16*b*. The "hooks" denote locations where the indicated frequencies are reflected for image and sum enhancement; see Fig. 11.24. (From [24]. Adapted with permission of John Wiley & Sons.)

simultaneously prevents the ω_Σ signal from propagating toward the RF input port and reflects it back to the diode for reuse, as indicated in Fig. 11.23. Image enhancement (Section 11.3.2) is obtained by properly positioning the image-reject filter. For wide-band image enhancement, these filters should have their reflection planes as close to the diode as possible. The circuit includes a return path for the DC component of the diode current, an IF bypass network (to prevent ω_I getting out of the RF port), separate tuning $\lambda/4$ stubs for the RF and LO frequencies, a sum-filter, and a lumped-element low-pass filter for the IF output.

In such a mixer the input to the diode has to be well matched to minimize reflection of ω_L back to the RF port: the LO signal could be radiated from the antenna unless an additional ferrite isolator is used. Figure 11.25 summarizes the performance of this mixer using a GaAs beam-lead Schottky diode. The results show L_c as a function of f_R with the LO fixed at 11.5 GHz. Minimum L_c of 3.8 dB is obtained for $f_R = 12.175$ GHz. Also shown is the L_c at the image frequency, reaching a peak of 24 dB at $f_{\text{image}} = 2f_L - f_R = 2 \times 11.5 - 12.175 = 10.825$ GHz.

Single-Balanced Mixers. Single-ended mixers have several shortcomings:

(a) Special filtering is needed to prevent LO power escaping from the RF port.

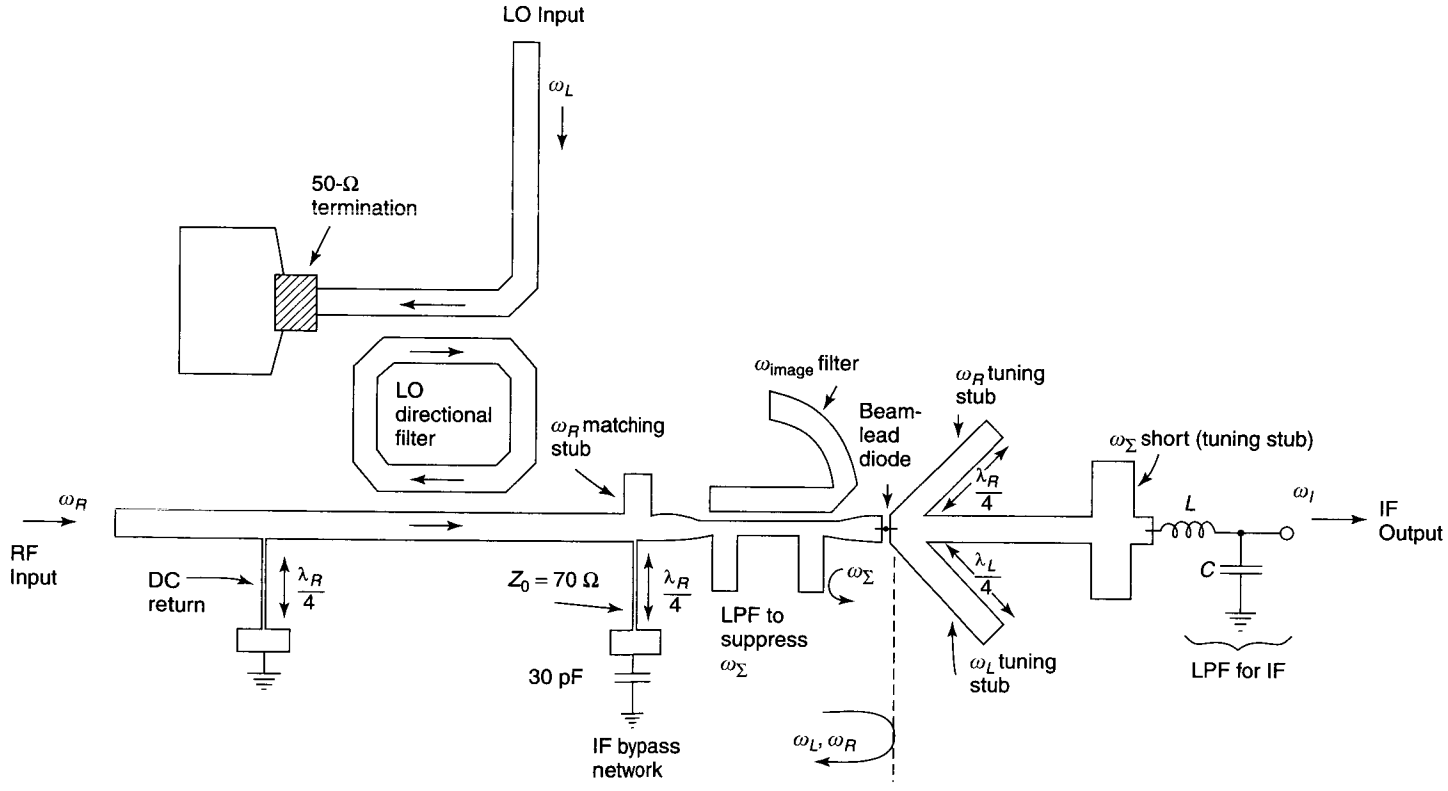


Figure 11.24 Example of practical 12-GHz image-and-sum enhanced single-ended mixer realized in microstrip. (Courtesy of E. J. Denlinger, David Sarnoff Research Center.)

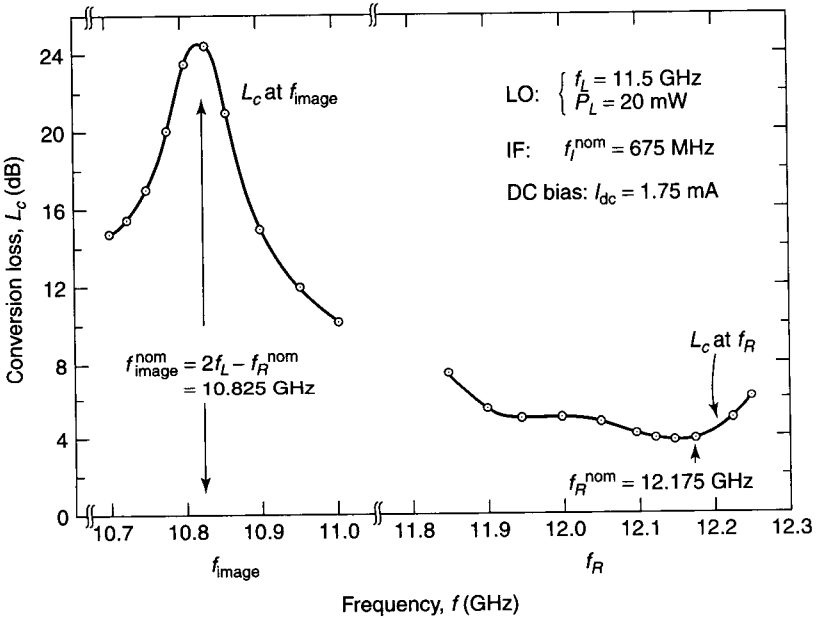


Figure 11.25 Measured performance of 12-GHz microstrip Z mixer of Fig. 11.24. Using a GaAs beam-lead diode, the minimum L_c is ~3.8 dB. The noise figure NF is ~5.5 dB when an IF amplifier with a noise figure of 1.6 dB is employed. (Courtesy of E. J. Denlinger, David Sarnoff Research Center.)

(b) Both noise figure and conversion loss will be degraded if a directional coupler or directional loop filter is used to bring in the LO signal.

(c) If the bandwidth of the AM noise associated with the LO signal overlaps the RF bandwidth, the mixer noise figure will again be degraded.

A solution is to combine two single-ended mixers in a balanced configuration using a 90° or 180° hybrid.

In Fig. 11.26*a* a prototype 180° hybrid single-balanced mixer (SBM) uses a "wire" balun* transformer to combine two single-ended mixers in parallel and 180° out of phase. This configuration suppresses even harmonics of one of the input signals, usually the LO. The degree of the suppression depends on circuit balance. The balun provides LO-to-RF isolation. The LO signal is applied to the two diodes 180° out of phase, whereas the RF signal appears in phase, as indicated by the broken and solid arrows. Any AM noise and/or spurious signals entering from the LO port will appear 180° out of phase at the two diodes; the resulting IF noise contributions at the output therefore cancel. Mixers using

* *Balun* means "balance-to-unbalance."

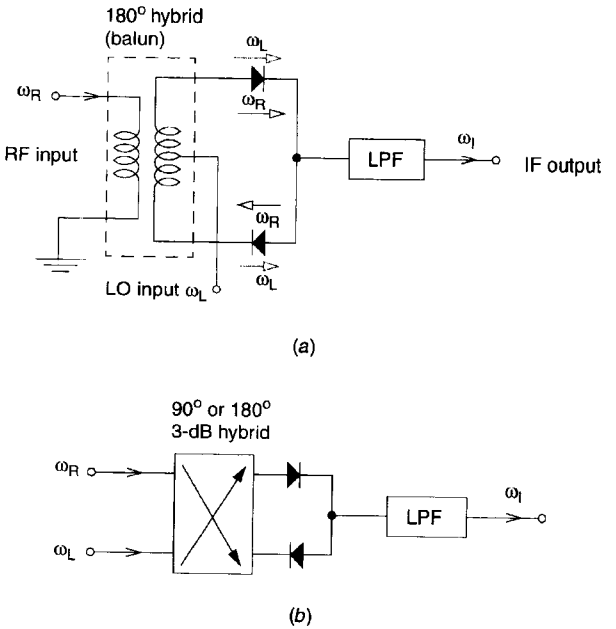


Figure 11.26 (*a*) Prototype SBM using 180°-hybrid (balun) transformer; (*b*) SBM using 90° hybrid.

this type of wire hybrid are limited to frequencies below ~2 GHz because of parasitics, but if realized in MMIC form, much higher frequencies are possible.

Figure 11.26*b* indicates a basis for microwave SBMs; here the wire balun is replaced by a 90° or 180° 3-dB hybrid: such hybrids can be realized in numerous different ways. Both 90° and 180° SBMs have the following advantages over single-ended mixers:

- reduced spurious signals,
- cancellation of the DC component at the IF output, and
- RF and LO inputs conveniently separated.

180° Single-Balanced Mixers. One type of 180° hybrid is the "classical" $6\lambda/4$ rat-race coupler [51] (see Fig. 11.27) consisting of three $\lambda/4$ arms and a $3\lambda/4$ section that provides an additional 180° phase shift in the fourth arm. A disadvantage is that the required $\lambda/2$ path difference is exact at only one frequency, resulting in a narrow bandwidth of operation. The bandwidth can be increased dramatically by using instead a $4\lambda/4$ rat-race hybrid, in which the 180° phase shift is achieved by a frequency-insensitive phase inverter. An early example [52] used a two-wire line for the ring; a simple Möbius twist produced the phase inversion. More recently, an octave-bandwidth microstrip rat-race hybrid used a pair of $\lambda/4$ coupled lines in the fourth arm [51].

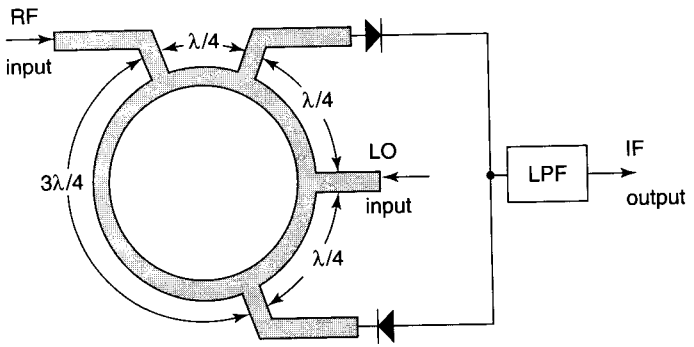


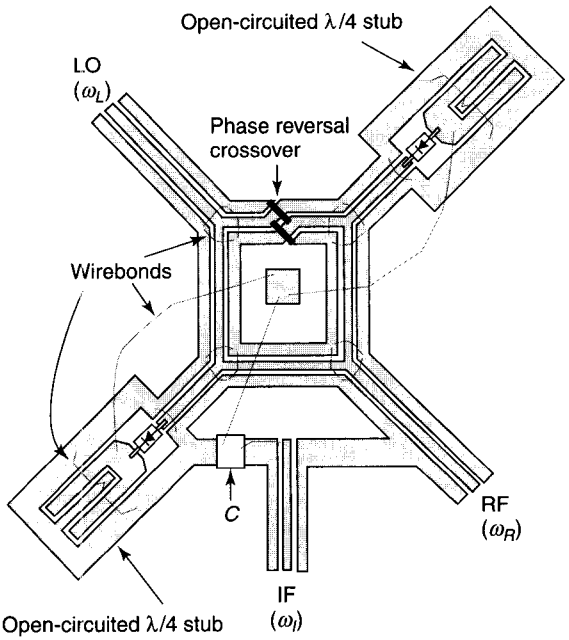
Figure 11.27 Conventional narrow-band SBM using $6\lambda/4$ rat-race hybrid ring. Wavelength λ refers to center frequency.

Wide-Band SBM Design. In the wide-band SBM design [53] shown in Fig. 11.28*a*, the rat race is fabricated using finite-ground coplanar waveguide (FCPW) [54] and the phase inversion is done using a crossover in the middle of the fourth arm. Whereas the conventional rat race uses a ring of impedance $\sqrt{2}$ times the port impedance (i.e., 70.7 Ω in a 50-Ω system), resulting in a Butterworth response, in this design a 55-Ω ring produces a second-order Chebyshev response with 12-dB return loss. Figure 11.28*b* demonstrates that an SBM using this type of 180° hybrid can operate over a bandwidth of 7.5 to 46 GHz (~4 octaves) with a conversion loss $L_c \sim 6$–8 dB.

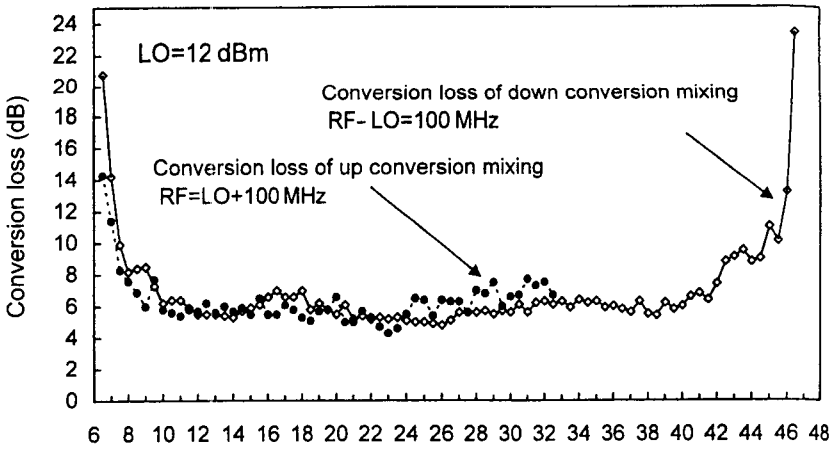
90° Single-Balanced Mixers. An alternative to the 180° hybrid is a 90° hybrid, such as the 3-dB branch-line coupler; see Figs. 11.29*a,b*. The SBMs designed with 90° hybrids have properties different from those using 180° hybrids. In general, there is suppression of the harmonics and IM products of both the RF and LO signals. If the two 90° output ports are equally terminated, all reflected power will go to the fourth, normally isolated, port. Therefore good diode balance will give a low VSWR (typically less than 1.5:1) at both the RF and LO ports. The LO-to-RF isolation now depends on the diode impedance matching and consequently tends to be poor unless the diodes are an integrated pair. Since multisection 3-dB branch-line couplers (see Chapter 5) can be designed for wide bandwidths, this topology can also used for wide-band SBM designs, up to an octave.

Double-Balanced Mixers. Single-ended and single-balanced mixers have the following problems:

- The LO-to-RF isolation is often inadequate.
- A large LO power is required.
- They are susceptible to AM noise on the LO.
- The bandwidth is narrow.



(a)



(b)

Figure 11.28 (*a*) A FCPW wide-band implementation of single-balanced mixer shown in Fig. 11.27. The long wirebonds are part of the IF circuit and do not affect the RF/LO behavior. (*b*) Multioctave down- and up-conversion losses L_c versus frequency. (After Chang et al. [53]. Reprinted with permission of IEEE.)

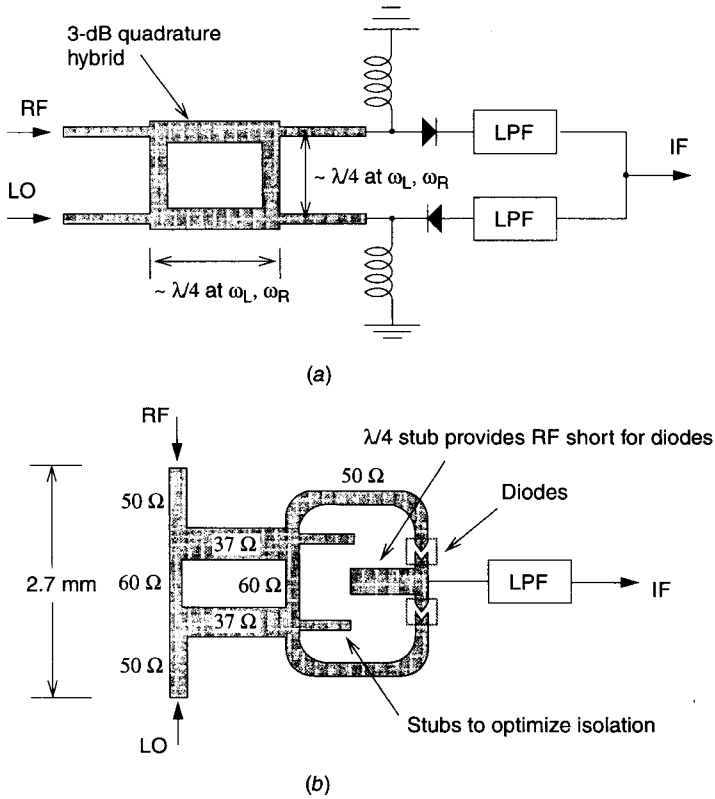


Figure 11.29 (*a*) Single-balanced mixer using 90° branch-line hybrid coupler. The usual impedance ratio between the main and side arms of the coupler is $\sqrt{2}$:1. (*b*) Example of monolithic implementation of (*a*). Here the impedance ratio of the hybrid has been increased to 1.6:1 to improve the isolation and bandwidth at the expense of the input matching (a VSWR increase from 1.1 to 1.2). In this example f_R = 30.5–31.8 GHz (at 0.1 mW), f_L = 29 GHz (at 30 mW), and f_I = 1.5–2.8 GHz. The conversion loss is less than 6.5 dB over the band. (After Chu et al. [55]. Reprinted with permission of IEEE.)

These problems can be solved by putting two SBMs together using either 180° or 90° hybrids (baluns). For example, one could take two SBMs as in Fig. 11.26*a* and connect them in parallel and 180° out of phase, as indicated in Fig. 11.30*a*. Figure 11.30*b* shows a realization in which diodes D_1 and D_2 "belong" to the upper SBM and diodes D_3 and D_4 to the lower SBM. This topology is equivalent to the classical prototype double-balanced mixer (DBM) in Fig. 11.31. Such a structure incorporates the four-diode ring of Fig. 11.16*c*, which ideally phase–cancels not only reflections at the LO port but also noise at both RF and IF ports. Isolation between all three ports is achieved by circuit balance. Theoretically, this isolation is complete if all four diodes are perfectly balanced and the circuit is exactly symmetrical. In practice, excellent balance can be achieved by integrating all four diodes, for example as a beam-lead quad chip. Using such a ring, the mutual isolation between mixer ports can

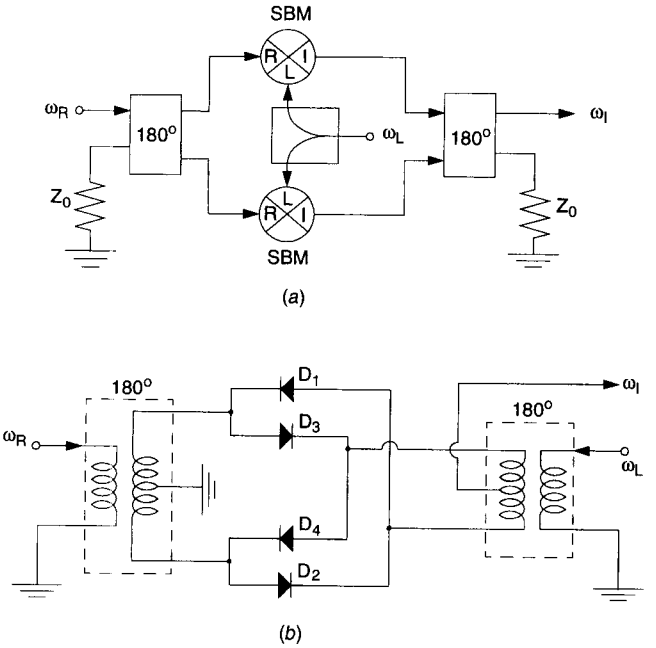


Figure 11.30 (*a*) Schematic showing how a DBM can be made by combining two SBMs in parallel and 180° out of phase. (*b*) Prototype classical DBM corresponding to (*a*).

exceed 30 dB and is fairly independent of frequency. The symmetry also suppresses the even-order harmonics of both ω_R and ω_L. This in turn reduces the level of the IM products (see Section 11.3.6): At the IF port all the IMs are theoretically balanced out, except for those involving *odd-order* harmonics of ω_R and ω_L [46]. Another consideration is that the four-diode bridge increases

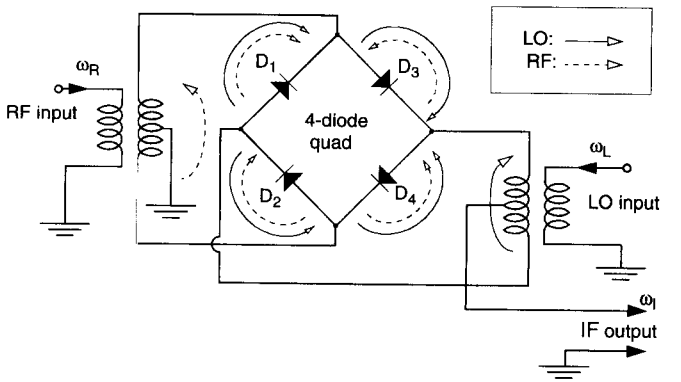


Figure 11.31 Prototype DBM. The topology is the same as in (*b*); the diode designations correspond.

the compression point and reduces the IM levels below those of an SBM because there is less RF voltage across each diode for a given RF input power level. Further increases in compression point and reduction in IM level can be provided by multiple-diode configurations, for example a 12-diode ring using 3 diodes per branch.

If, as is usual, the LO amplitude is much larger than the RF amplitude, opposite pairs of diodes get turned on and off alternately by the LO voltage. The LO power P_L required tends to be higher than for an SBM because in a DBM the LO source sees two diodes in series. A typical value is $P_L = +10$ dBm, and a typical 3-dB intercept point is +18 dBm.

The "wire-type" 180° baluns in Fig. 11.31 limit operation to ~3 GHz. If they are realized as airbridge transformers in MMIC technology, much higher operational frequencies are possible.

Planar DBM Design. The operation of a planar DBM can be illustrated by referring to the microstrip/slotline implementation of Fig. 11.32 [56]. Comparing this with the "prototype" DBM of Figs. 11.30 and 11.31, one sees that the circuits are functionally equivalent but differ in topology. Thus the 180° RF hybrid configuration using both sides of the substrate is different from the wire-type 180° hybrid in that there are no crossing transmission lines. Rather it resembles the classical waveguide magic tee. As in Fig. 11.30*a*, there are two SBMs, but because of the topology of the 180° hybrid, the LO and IF connections to them are made in a different way. The RF is applied via a microstrip on one side of the substrate and is magnetically coupled to the symmetrical $3\lambda/4$-slotline arms on the other side using a well-known $\lambda/4$-microstrip-to-slot-line transition [57]. The signal splits symmetrically so that equal RF fields appear across the two slotlines. The two arms are then brought together, making a pair of coupled slotlines (equivalent to a section of coplanar waveguide). The LO input is applied via a $\lambda/4$-microstrip-to-CPW transition so that equal LO electric fields appear across each slot of the pair. Finally, the two slotlines are separated to provide the two output ports of the magic tee. Thus the complete magic tee uses the orthogonal even and odd modes of coupled slotlines in such a way that at the output the RF signal appears in the odd mode and the LO appears in the even mode. The directions of the RF and LO electric fields are represented by broken and solid arrows, respectively: They are in phase in one arm but 180° out of phase in the other. Comparison with Fig. 11.31 shows that the similarly numbered diodes in the two figures see equivalent combinations of these fields.

The two SBMs are each similar to a design described by Ogawa et al. [58]. Because of the different topology, the IF signals from the two SBMs appear in phase, so that a simple 0° 3-dB coupler can be used to combine them.

Finally, unwanted frequencies appearing at the IF port are removed by a five-section microstrip low-pass filter.

Analysis of this DBM using ideal-exponential-diode theory [56] shows that when the LO signal is

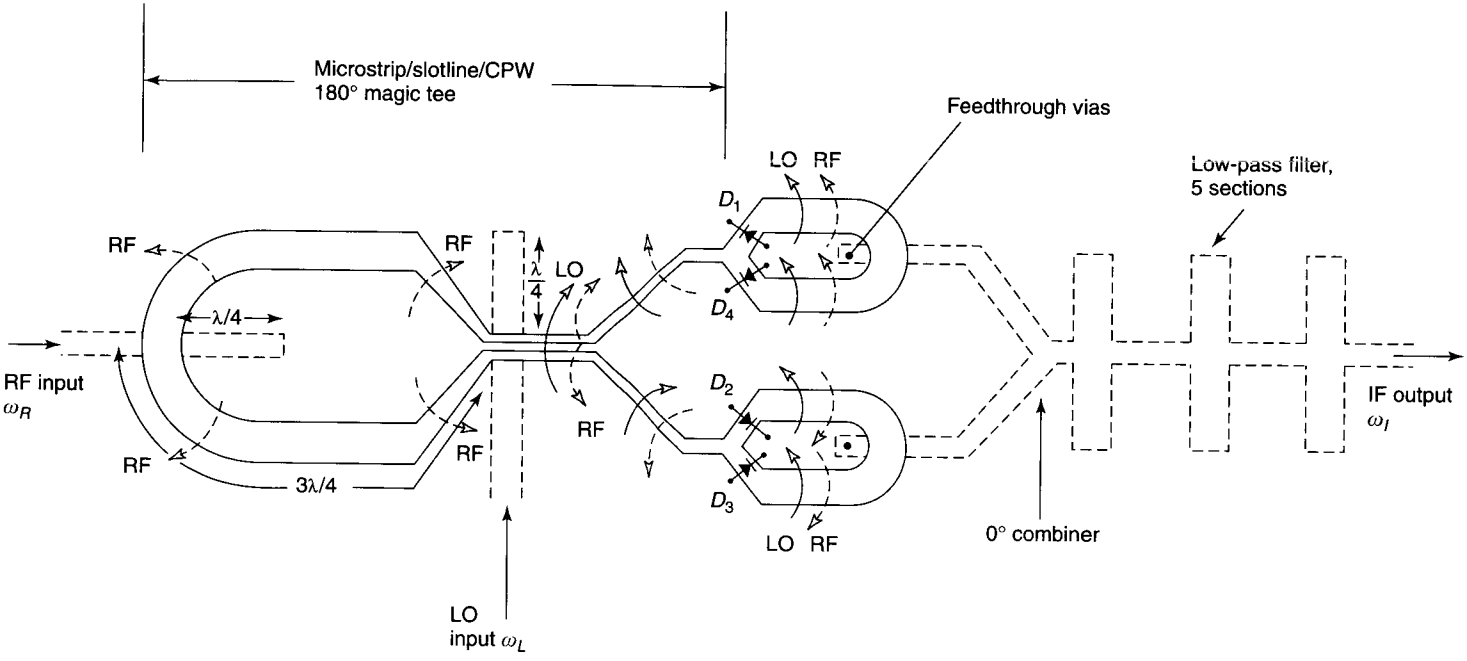


Figure 11.32 Planar-substrate DBM topology designed for $f_R = 18$–21 GHz. The solid lines indicate slotlines and coupled slotlines [equivalent to coplanar waveguides (CPWs)]; broken lines indicate microstrip on the reverse side of the substrate. (After Ogawa et al. [56]. Reprinted with permission of IEEE.)

$$v_L(t) = V_L \cos(\omega_L t)$$

the IF current at the microstrip output port is

$$\begin{aligned} i_I(t) = {} & 4I_s X_R \mathbb{I}_1(X_L) \cos(\omega_L - \omega_R)t \quad \text{(desired output at } \omega_I\text{)} \\ & + 4I_s X_R \mathbb{I}_1(X_L) \cos(\omega_L + \omega_R)t \quad \text{(terms of the form } N\omega_L \pm \omega_R\text{,} \\ & + 4I_s X_R \mathbb{I}_3(X_L) \cos(3\omega_L - \omega_R)t \quad \text{where } N \text{ is a nonzero odd integer)} \\ & + \cdots \end{aligned}$$

where I_s is the diode saturation current, $X_R = \Lambda V_R$, $X_L = \Lambda V_L$, $\Lambda = q_e/(nkT)$, and $\mathbb{I}_N(x)$ is the modified Bessel function of order N and argument x. This demonstrates that the output from the IF port includes the wanted frequency ω_I, the sum frequency $\omega_I + \omega_R$, and harmonic sidebands but *not* the image $2\omega_L - \omega_R$.

In a similar way it can be shown that the currents appearing at the RF and LO ports are

$$\begin{aligned} i_R(t) = {} & 4I_s X_R \mathbb{I}_0(X_L) \cos(\omega_R t) \quad \text{(fundamental RF signal)} \\ & + 4I_s X_R \mathbb{I}_0(X_L) \cos(2\omega_L - \omega_R)t \quad \text{(terms of the form } M\omega_L \pm \omega_R\text{,} \\ & + 4I_s X_R \mathbb{I}_0(X_L) \cos(4\omega_L - \omega_R)t \quad \text{where } M \text{ is an even integer)} \\ & + \cdots \\ i_L(t) = {} & 4I_s X_L [\mathbb{I}_0(X_L) + \mathbb{I}_2(X_L)] \cos(\omega_L t) \\ & + 4I_s X_L [\mathbb{I}_2(X_L) + \mathbb{I}_4(X_L)] \cos(3\omega_L t) \\ & + 4I_s X_L [\mathbb{I}_4(X_L) + \mathbb{I}_6(X_L)] \cos(5\omega_L t) \\ & + \cdots \end{aligned}$$

Thus only the fundamental and odd harmonics of ω_L appear at the LO port.

An implementation of this DBM [56] achieved an isolation between the three ports better than 20 dB in the 18–21-GHz range and a conversion loss $L_c < 5$ dB over the RF range $19 < f_R < 20$ GHz for an LO power $P_L = 13.7$ dBm.

Double Double-Balanced Mixers. Double double-balanced mixers (DDBMs) are the only type that can have overlapping RF and IF bandwidths and still maintain isolation between the RF, LO, and IF ports. Most DDBMs need separate RF, LO, and IF balun transformers and double-ring diodes [59]. Such nonplanar circuits are unsuitable for conventional MMIC processes.

A slotline/CPW uniplanar DDBM was described by Eisenberg et al. [60], and a design using a lumped-element hybrid analogous to the 180° rat race was

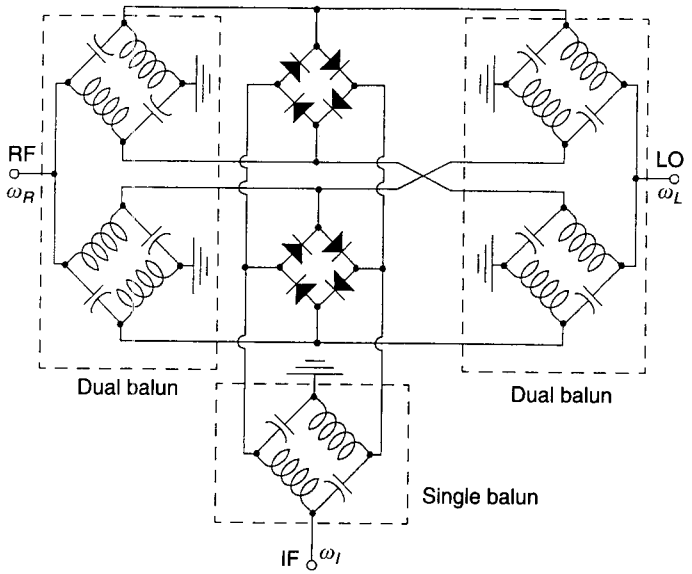
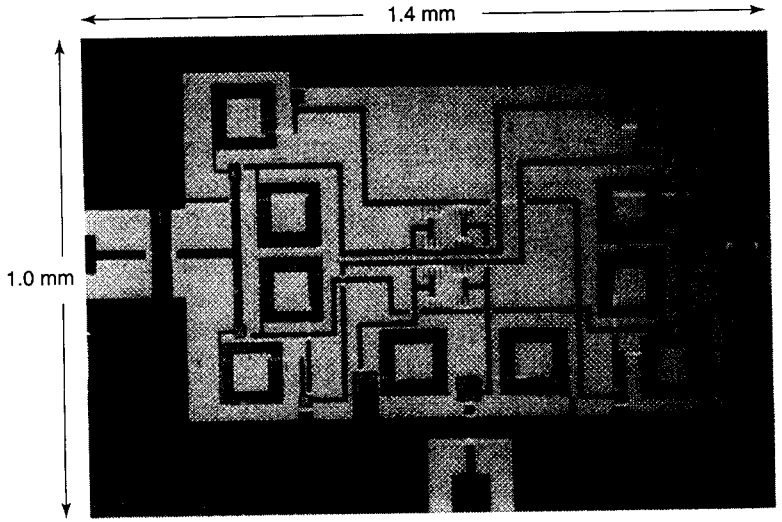


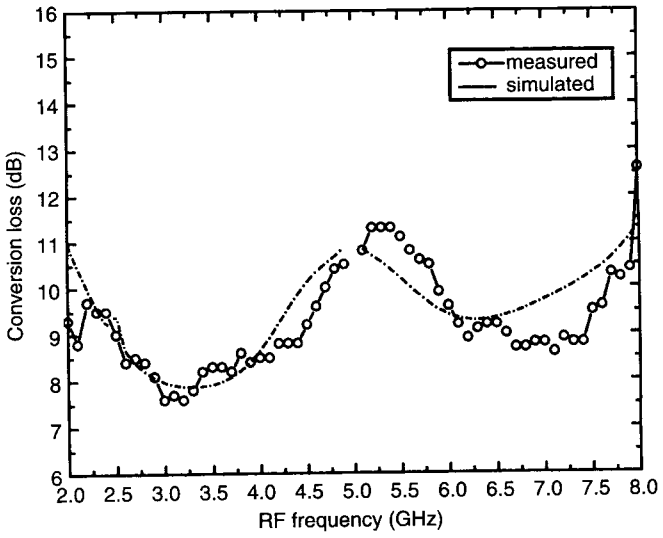
Figure 11.33 Circuit diagram of monolithic DDBM. (After Chiou and Lin [62]. Adapted with permission of IEEE.)

given by Ton et al. [61]; both are complex and large. The lumped-element DDBM in Fig. 11.33 is suitable for use up to ~5 GHz and has minimum size. It consists of two lumped-element dual baluns for the RF and LO inputs, a single balun for the IF, and two ring diode quads. The dual baluns each consists of two out-of-phase power splitters in parallel, each composed of one high-pass filter (HPF) and one low-pass filter (LPF) in parallel. Such a mixer has been implemented [62] (see Fig. 11.34*a*) using *n*-implanted diodes with $R_s = 32\ \Omega$, $C_j(0) \sim 58$ fF. The LO and RF dual baluns use 3-nH inductors and 0.27-pF capacitors for 5 GHz center frequency; the LO balun uses 4-nH inductors and 1.2-pF capacitors for a center frequency of 2 GHz. Even though this is an all-lumped-element design, it operates up to 5 GHz because of its small dimensions (1 × 1.4 mm chip size). Figure 11.34*b* gives the measured conversion loss for a fixed LO frequency $f_L = 5$ GHz: it is seen that the RF and IF bandwidths do indeed overlap, since the IF frequency extends up to 3 GHz when the RF frequency goes down to 2 GHz.

Subharmonically Pumped Mixers. When a mixer is to be used with high RF frequencies, it may be advantageous to derive the LO from a source at a submultiple of the desired LO frequency. The nonlinear mixing element itself is then designed to also perform the necessary multiplication of the source frequency up to the LO frequency. This is useful at millimeter and submillimeter frequencies where the cost of fundamental-frequency power is high. Such sub-



(*a*)



(*b*)

Figure 11.34 (*a*) Lumped-element MMIC DDBM. (*b*) Conversion loss of DDBM with fixed LO frequency of 5 GHz and LO power of 17 dBm o: measured, --- simulated. (After Chiou and Lin [62]. Adapted with permission of IEEE.)

harmonically pumped (SHP) mixers typically operate with the LO at around half the signal frequency. They were first reported in 1975 simultaneously by Cohn et al. [63] and Schneider and Snell [64]. Confusingly, these mixers are called both subharmonic and harmonic mixers; here the former term is preferred.

Advantages of SHP Mixers. In addition to the above points, SHP mixers offer certain advantages:

1. Isolation between the LO and IF ports, typically >30 dB, is at least 10 dB better than other mixer circuits and is a consequence of the (ideally perfect) isolation between the effective pump frequency (usually $2\omega_L$) and ω_I.
2. Because conversion of ω_L to undesired frequencies is suppressed, L_c is low. (These frequencies include even-order harmonics of ω_L and mixing products involving odd-order products of ω_L or ω_R).
3. Degradation of the overall system by LO noise is reduced [63, 65].
4. Because there is a large frequency separation between the external LO frequency and ω_R, there is no need for the hybrid couplers and/or diplexers used to couple the ω_L and ω_R signals to the diodes in ordinary mixers [66].

SHP Mixer Circuits. A single-ended SHP mixer circuit is shown in Fig. 11.35. For operation with the pump at half the desired LO frequency, a nonlinear resistance with an odd-order i–v characteristic is required. This is usually obtained with an antisymmetric diode pair. In this configuration the two diodes protect each other against burnout due to excessive inverse voltage; furthermore, the total diode current has no DC component, obviating the need for a DC return path. A planar doped barrier (PDB) diode (Fig. 11.3*e*) can also be used.

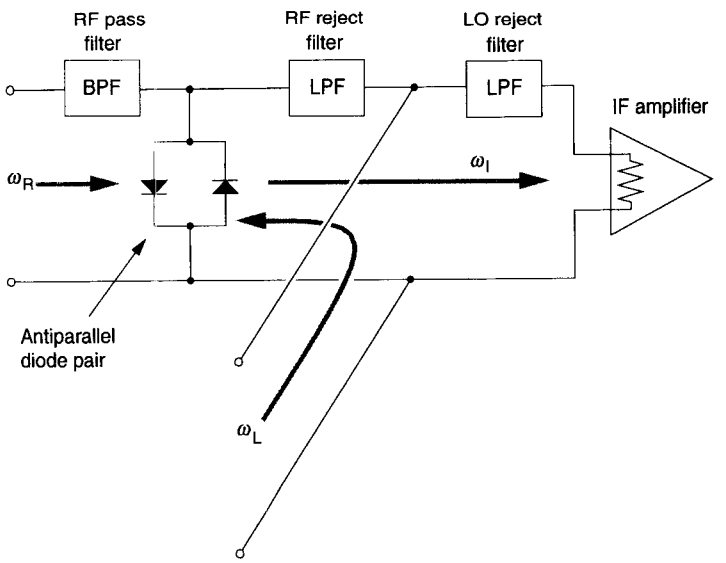


Figure 11.35 Conceptual SHP mixer circuit. Other variations are possible, depending on the bandwidths of the RF, LC, and IF signals and the transmission medium.

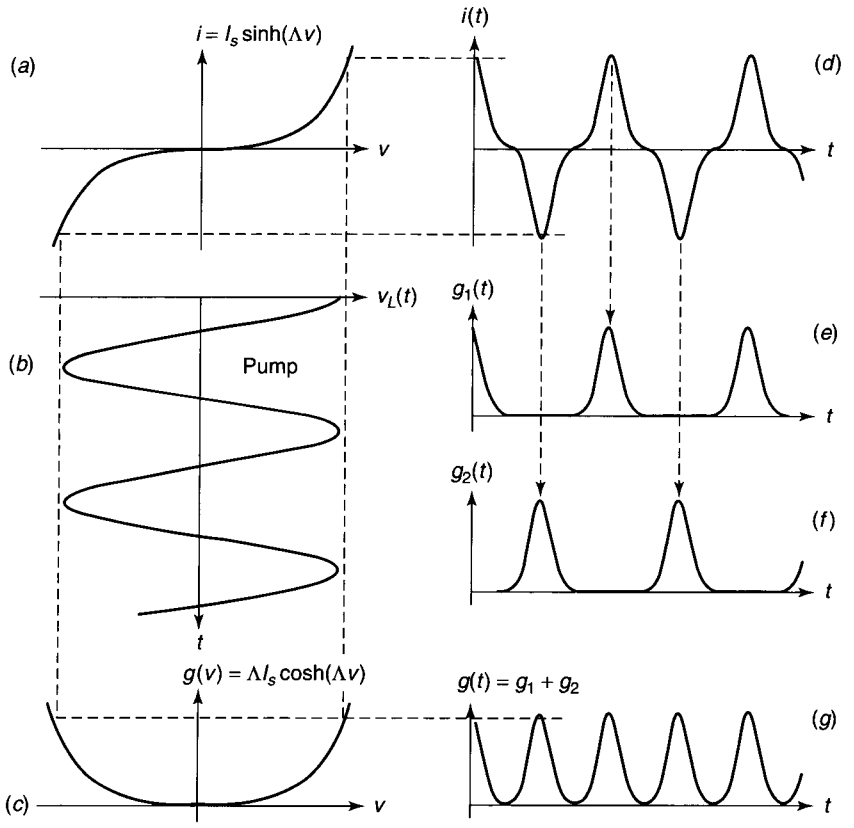


Figure 11.36 Transfer functions and waveforms of second-subharmonic mixer: (*a*) nonlinear *v–i* characteristic of antisymmetric diode pair; (*b*) LO pump waveform $v_L(t)$; (*c*) conductance characteristic $g(v)$; (*d*) current $i(t)$ flowing in diode pair; (*e*) time-varying conductance $g_1(t)$ of diode 1; (*f*) time-varying conductance $g_2(t)$ of diode 2; (*g*) total time-varying conductance $g(t)$ of diode pair.

The objective of the circuit in Fig. 11.35 is to achieve a time-varying conductance $g(t)$ that goes through two cycles for one LO cycle. Figure 11.36 shows how this happens. The *i–v* characteristic is a sinh function of the applied voltage v, so its derivative $\partial i/\partial v = g(v)$ is a cosh function of v. Consequently, one diode has peaks of conductance $g_1(t)$ at positive $i(t)$ excursions; the other has peaks of $g_2(t)$ at negative $i(t)$ excursions. Since the diodes conduct on alternate LO cycles, the frequency-doubled total conductance function is $g(t) = g_1(t) + g_2(t)$.

Analysis of SHP Mixers. Subharmonically pumped mixers have been analyzed assuming identical diodes [63, 64, 66, 67] and unbalanced diodes [68]. Kerr [67] did a detailed investigation of noise and loss in both SHP and balanced mixers. An image rejection SHP was analyzed by Weiner et al. [69].

Assuming that diode capacitances C_j and resistances R_s are negligible and the voltage across the diode pair $v_L(t) = V_L \cos(\omega t)$, then it can be shown [64] that only odd-order harmonic currents flow in the diode pair. The conductance is

$$g(t) = 2\Lambda I_s \cosh[\Lambda V_L \cos(\omega_L t)] \tag{11.77}$$

so that

$$g(t) = 2g_0 + 4g_1 \cos(2\omega t) + 4g_2 \cos(4\omega t) + \cdots \tag{11.78}$$

where

$$g_N = \Lambda I_s \mathbb{I}_N(\Lambda V_L) \qquad N = 0, 1, 2, \ldots$$

This is the $g(t)$ waveform shown in Fig. 11.36*g*. The conversion loss L_c of the SHP mixer can be found from this using the Saleh theory [39]. Comparing (11.78) with the corresponding equation (11.72) for the single-diode Y mixer, one sees that the expressions for the conductance waveforms are similar, except that the constituent frequencies are now all multiplied by 2, and that the coefficients of the time-varying terms are twice as large. Consequently, the expressions for the minimum possible L_c are the same as for the single-diode Y mixer under fundamental pumping. Bearing in mind that the image frequency is

$$\omega_{\text{image}} = 2\omega_L - \omega_I = 2\omega_L - (\omega_R - 2\omega_L) = 4\omega_L - \omega_R$$

the expressions for L_c in the short-circuited image, open-circuited image, and broadband image conditions are exactly the same as in (11.74)–(11.76), respectively, for the single-diode mixer. This means that the L_c curves of Fig. 11.20 apply to the SHP mixer also.

The L_c of an SHP mixer can be reduced by image enhancement. On the other hand, L_c increases drastically with both diode-pair asymmetry and network unbalance. For example, it was shown [68] that a $C_j(0)$ imbalance of only 0.14 fF led to an 8-dB increase of L_c in a particular SHP mixer that had $f_R = 103$ GHz, $f_L = 50$ GHz, and $f_I = 3$ GHz.

Further design considerations include the possibility of circulating currents being excited in the loop of a two-diode pair [63] and parasitic loop inductance causing multiple conduction in the diodes [67, 68]. Such concerns are absent when antisymmetric PDB diodes are used [70].

Total Diode Current. The total current in the diode pair is given by

$$i(t) = 2I_s \sinh[\Lambda V_R \cos(\omega_R t) + \Lambda V_L \cos(\omega_L t)] \tag{11.79}$$

which is valid for arbitrarily large values of V_R and V_L [whereas the linearized expression $i(t) = g(t)v(t)$ (see [63]) is not]. Expansion of (11.79) shows that the total diode current $i(t)$ involves only frequencies $n\omega_R + m\omega_L$, where

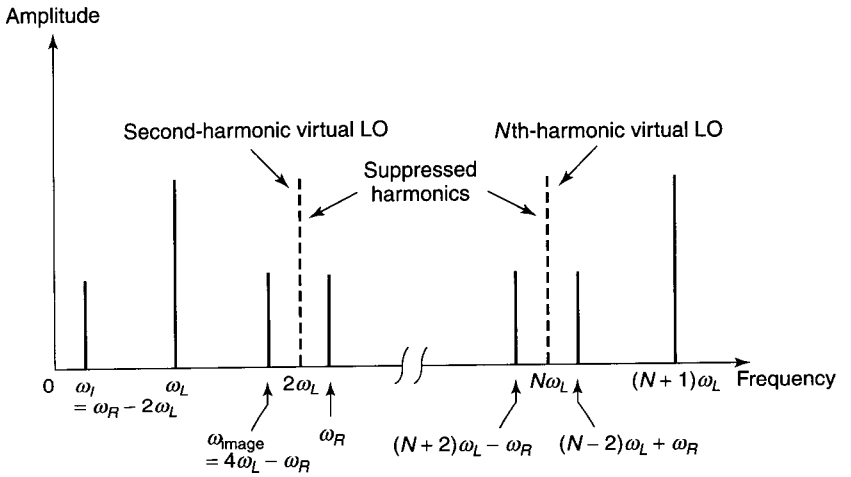


Figure 11.37 Simplified spectrum showing operation of SHP when second harmonic of external LO is used and $\omega_R > 2\omega_L$. The symmetry of the antiparallel diode pair suppresses the even-order harmonics of the LO frequency ω_L. The suppressed harmonics are indicated as the broken lines. The general case for mixing on the Nth harmonic is also shown.

$n + m =$ odd integer. Among these frequencies, $|\omega_R - 2\omega_L|$ is the desired IF frequency ω_I, while $|\omega_R - 4\omega_L|$ is the image frequency. The terms $2\omega_R + \omega_L$ and $4\omega_R + \omega_L$ are canceled in a single-balanced SHP mixer.

Figure 11.37 shows a typical SHP mixer spectrum* for $N = 2$. It also shows the general case for $N > 2$. There can be RF input signals at either $N\omega_L + \omega_I$ or $N\omega_L - \omega_I$ (the SSB case) or at both (the DSB case); they are down converted to ω_I by the pump frequency $N\omega_L$. The multiplied-up LO frequency $(2\omega_L)$ and all even-order harmonics of ω_L are suppressed, as well as the fundamental and other odd-harmonic mixing products [63].

SHP Mixer Designs. Microwave SHP mixers have been fabricated in stripline [64] and slotline [63]. A millimeter-wave suspended-stripline version operated at $f_R = 230$ GHz with $f_L = 115$ GHz [71]. Other planar designs include an *image rejection* SHP mixer [69] and a quasi-optical configuration with an antiparallel diode pair at the center of a bowtie antenna [72]. The symmetric PDB diode (Section 11.1.3) has been used in coplanar [73] and finline [70] SHP mixers for f_R up to ~125 GHz. A monolithic SHP mixer for $f_R \sim 175$ to 182 GHz was designed using InP HEMT diodes [74].

11.3.8 Mixer Noise

Mixer noise mechanisms are similar to those of detectors but have the complication that noise spectra are transferred to different frequency ranges via the

*S. Bafia, private communication.

mixing process. In addition, one must consider *intermodulation, harmonic suppression* (Section 11.3.6), and *noise factor* (*noise figure*) implications on the overall mixer design.

Mixer Noise Temperature Ratio. The noise temperature ratio of a mixer* is defined as

$$t_m \triangleq \frac{\text{available IF output noise power in bandwidth } B}{kT_0B} \tag{11.80}$$

when the mixer and all terminations are at $T_0 = 290$ K (by convention).

Noise Factor and Noise Figure. Because of the difference between the noise factors of linear and nonlinear circuits, the noise factors (noise figures) of receivers and mixers will be discussed separately.

Receiver Noise Factor (Noise Figure). The overall *noise factor F of a receiver* is a measure of the noise added by the receiver circuits, excluding $1/f$ (flicker) noise. The noise figure NF is F expressed in decibels: NF $= 10 \log_{10}(F)$ decibels. By definition [75, 76],

$$F \triangleq \frac{N_o}{N_iG} = \frac{N_o}{N_i} \cdot \frac{S_i}{S_o} = \frac{S_i/N_i}{S_o/N_o} \tag{11.81}$$

where G = receiver gain (can be <1), $=S_o/S_i$

S_i = available signal input power,

S_o = available signal output power,

N_i = available input noise power in bandwidth B, $=kT_0B$

N_o = available output noise power in bandwidth B, $=kT_1B$

T_0 = standard noise temperature = 290 K (this makes kT_0/q_e equal to 25.0 mV),

T_1 = output noise temperature of receiver

Since $S_o = GS_i$ and the input noise power $N_i = kT_0B$, the amplified input noise is kT_0BG. One can *represent* the extra noise power added by the receiver by kT_eBG, where T_e is a fictitious equivalent input noise temperature. Then the total output noise is

$$N_o = k(T_0 + T_e)BG$$

so that the receiver's noise factor can be written as

*Not to be confused with the diode white noise temperature ratio t_w (Section 11.1.4).

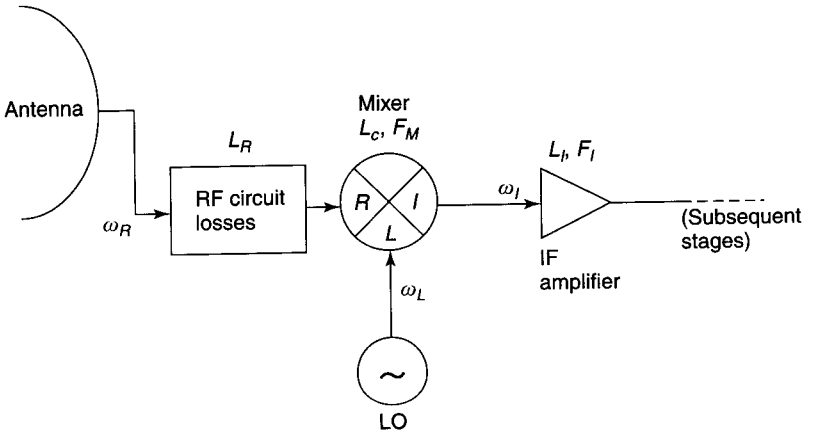


Figure 11.38 General configuration of microwave receiver: L_R, L_C, and L_I are losses associated with RF circuit, mixer, and IF amplifier, respectively; F_M and F_I are noise factors of mixer and IF amplifer.

$$F = 1 + \frac{T_e}{T_0} \tag{11.82}$$

The *noise temperature* T_e is often more useful for calculations than is the noise factor F. The noise factor of a receiver (Fig. 11.38) can be found from that of a cascade [48] and is

$$F_T = L_R + (F_M - 1)L_R + (F_I - 1)L_R L_C + \cdots \tag{11.83}$$

where L_R is the RF loss (assumed noiseless), L_C is the mixer loss, F_I is the IF amplifier noise factor, and higher order terms are negligible if the IF gain is large. The term F_M is the mixer noise factor, discussed next.

Mixer Noise Factor. There have been misconceptions about the noise factor of a passive mixer [77, 67]. The following definition [77] agrees with the experimental facts:

$$F_M = \frac{P_{\text{IF}}^{\text{total}}(B)}{P_{\text{IF}}^{\text{term}}(B)} \tag{11.84}$$

where $P_{\text{IF}}^{\text{total}}(B)$ is the available total IF noise output power in bandwidth B, and $P_{\text{IF}}^{\text{term}}(B)$ is the available IF noise output power in bandwidth B *due to the signal terminations only*.

Like the noise factor F of a *linear* system, F_M measures how much noise is added to an input signal by a mixer. Since a mixer is a *nonlinear* multiport with multiple frequency conversions, the definition of F_M has to be consistent with

the type of mixer under evaluation. Below, F_M is evaluated for three different circumstances.

1. Noise Factor of an IRM. The image-reject mixer (IRM) is also known as a SSB mixer [67] and an image-filtered mixer [48]. Assuming that *no power is dissipated at the image frequency* (no resistive termination at that frequency), that there are reactive (noiseless) terminations at all sidebands except ω_R and ω_I, that the external RF termination is at reference temperature T_0 (290 K), and that the mixer diode(s) generate noise power kT_DB, where $T_D > T_0$ is the (fictitious) diode noise temperature, it can be shown [Problem 11.9(a)] that the IRM noise factor is

$$F_M^{\text{IRM}} = \left(\frac{T_D}{T_0}\right)(L_r - 1) + 1 \tag{11.85}$$

where L_r is the RF-to-IF conversion loss. For Schottkys the ratio T_D/T_0 is the noise temperature ratio t_w (Section 11.1.4). By (11.80) the noise temperature ratio of an IRM is found to be [Problem 11.9(b)]

$$t_m^{\text{IRM}} = t_w(1 - L_r^{-1}) + L_r^{-1} \tag{11.86}$$

agreeing with Kerr [67]. Equations (11.85) and (11.86) show that for Schottky mixers

$$F_M^{\text{IRM}} = L_r t_m^{\text{IRM}} \tag{11.87}$$

a relationship often quoted for mixers in general but seldom justified.

2. Double-Sideband Noise Factor. Here it is assumed that there are input signals at both RF and image ports, that there are (noisy) terminations at temperature T_0 at those ports, and that the RF-to-IF and image-to-IF conversion losses L_r and L_i are different (results simplify when $L_r = L_i$). The DSB noise factor and noise temperature ratio are [Problem 11.9(c)]

$$F_M^{\text{DSB}} = \frac{T_D}{T_0}\left(\left[\frac{1}{L_r} + \frac{1}{L_i}\right]^{-1} - 1\right) + 1 \quad \text{(DSB noise factor)} \tag{11.88}$$

$$t_m^{\text{DSB}} = \frac{T_D}{T_0}\left(1 - \left[\frac{1}{L_r} + \frac{1}{L_i}\right]\right) + \left(\frac{1}{L_r} + \frac{1}{L_i}\right)$$

$$\text{(DSB noise temperature ratio)} \tag{11.89}$$

agreeing with Kerr [67]. Comparing (11.88) and (11.89), one can still write

$$F_M^{\text{DSB}} = L_c t_m^{\text{DSB}} \tag{11.90}$$

provided that $L_c = (1/L_r - 1/L_i)^{-1}$.

3. Single-Sideband Noise Factor. Again the RF and image terminations at T_0 ambient both contribute to the total output noise, resulting in [Problem 11.9(d)]

$$F_M^{\text{SSB}} = \frac{T_D}{T_0}\left(L_r - \left[1 + \frac{L_r}{L_i}\right]\right) + \left(1 + \frac{L_r}{L_i}\right) \quad \text{(SSB noise factor*)} \quad (11.91)$$

$$t_m^{\text{SSB}} = \frac{T_D}{T_0}\left(1 - \left[\frac{1}{L_r} + \frac{1}{L_i}\right]\right) + \left(\frac{1}{L_r} + \frac{1}{L_i}\right)$$

(SSB noise temperature ratio) (11.92)

It is still possible to write

$$F_M^{\text{SSB}} = L_r t_m^{\text{SSB}} \quad (11.93)$$

Relation between F_M^{SSB} and F_M^{DSB}. If $L_r = L_i = L_c$, then (11.88) and (11.91) give the well-known result

$$F_M^{\text{SSB}} = 2F_M^{\text{DSB}} \quad (11.94)$$

Simplification of Receiver Noise Factor. With the above interpretations, the mixer noise factor is

$$F_M = L_c t_m \quad (11.95)$$

Putting this into (11.83),

$$F_T = L_R + (L_c t_m - 1)L_R + (F_I - 1)L_R L_c$$

Simplifying, the overall noise factor of the receiver in Fig. 11.38 is [75]

$$F_T = L_R L_c (F_I + t_m - 1) \quad (11.96)$$

REFERENCES

1. Ratni, M., B. Huyart, E. Bergeault, and L. Jallet, "RF Power Detector Using a Silicon MOSFET," *IEEE MTT-S Int. Symp. Dig.*, Jun. 1998, pp. 1139–1142.
2. Penfield, P., and R. P. Rafuse, *Varactor Applications*, MIT Press, Cambridge, MA, 1962.

*In the IEEE definition of F_M^{SSB} the noise contribution from the image termination is *ignored* by postulating that the termination is at the nonphysical temperature 0.0 K; see Maas [48].

3. Tsui, J. B., *Microwave Receivers with EW Applications*, Wiley-Interscience, New York, 1986, pp. 16–18.
4. Waugh, R. W., "Designing the Virtual Battery," in *Proc. Third Annual Wireless Symposium*, Santa Clara CA, Feb. 1995, pp. 160–168. See also Agilent Application Note No. 1088, www.semiconductor.agilent.com, "Designing the Virtual Battery," Nov. 1999.
5. Zirath, H. H. G., S. M. Nilsen, H. Hjelmgren, L. P. Ramberg, and E. Kollberg, "Temperature Variable Noise and Electrical Characteristics of Au-GaAs Schottky Barrier Millimeter-Wave Mixer Diodes," *IEEE Trans. Microwave Theory Tech.*, Vol. 36, No. 11, Nov. 1988, pp. 1469–1475.
6. Mott, N. F., "Note on the Contact between a Metal and an Insulator or Semiconductor," *Proc. Cambridge Phil. Soc.*, Vol. 34, 1939, pp. 568–572.
7. Keen, N. J., W. M. Kelly, and G. T. Wrixon, "Pumped Schottky Diodes with Noise Temperatures Less Than 100 K at 115 GHz," *Electron. Lett.*, Vol. 15, Oct. 1979, pp. 689–690.
8. Montgomery, M. D., "The Tunnel Diode as a Highly Sensitive Microwave Detector," *Proc. IRE*, Vol. 49, Apr. 1961, pp. 826–827.
9. Tsu, R., and L. Esaki, "Tunneling in a Finite Superlattice," *Appl. Phys. Lett.*, Vol. 22, No. 11, Jun. 1973, pp. 562–564.
10. Chang, L. L., L. Esaki, and R. Tsu, "Resonant Tunneling in Semiconductor Double Barriers," *Appl. Phys. Lett.*, Vol. 24, No. 12, Jun. 1974, pp. 593–595.
11. Capasso, F., S. Sen, F. Beltram, and A. Y. Cho, "Resonant Tunneling Devices and Their Applications," in R. K. Watts (Ed.), *Submicron Integrated Circuits*, Wiley, New York, 1989.
12. Malik, R. J., R. T. AuCoin, R. L. Ross, K. Board, C. E. C. Wood, and L. F. Eastman, "Planar Doped Barriers in GaAs by Molecular Beam Epitaxy," *Electron Lett.*, Vol. 16, No. 22, 1980, pp. 836–838.
13. Dixon, S., and R. J. Malik, "Subharmonic Planar Doped Barrier Mixer Conversion Loss Characteristics," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-31, No. 2, Feb. 1983, pp. 155–158.
14. Dale, I., A. Condie, S. Neylon, and M. J. Kearney, "Planar Doped Mixer and Detector Diodes as Alternatives to Schottky Diodes for Both Microwave and Millimeter Wave Applications," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. 1, Long Beach, CA, June 13–15, 1989, pp. 467–470.
15. Lee, T.-H., J. R. East, C.-Y. Chi, G. M. Rebeiz, R. J. Dengler, I. Mehdi, P. H. Siegel, and G. I. Haddad, "The Fabrication and Performance of Planar Doped Barrier Diodes as 200 GHz Subharmonically Pumped Mixers," *IEEE Trans. Microwave Theory Tech.*, Vol. 42, No. 4, Apr. 1994, pp. 742–749.
16. Güttich, U., K. M. Strohm, and F. Schäffler, "D-band Subharmonic Mixer with Silicon Planar Doped Barrier Diodes," *IEEE Trans. Microwave Theory Tech.*, Vol. 39, No. 2, Feb. 1991, pp. 366–368.
17. Tucker, J. R., and M. J. Feldman, "Quantum Detection at Millimeter Wavelengths," *Rev. Modern Phys.*, Vol. 57, No. 4, Oct. 1985, pp. 1055–1113.
18. Pozar, D. M., *Microwave Engineering*, Addison-Wesley, Reading, MA, 1990. Chap. 11, p. 584.

19. van der Ziel, A., *Noise: Sources, Characterization, Measurement*, Prentice-Hall, Englewood Cliffs, NJ, 1970.
20. Uhlir, A., "Characterization of Crystal Diodes for Low-Level Microwave Detection," *Microwave J.*, Vol. 6, No. 7, Jul. 1963, pp. 59–67.
21. Torrey, H. C., and C. A. Whitmer, *Crystal Rectifiers*, MIT Radiation Laboratory Series, Vol. 15, McGraw-Hill, New York, 1948.
22. Warner, F. L., "Detection of Millimeter and Submillimeter Waves," in F. A. Benson (Ed.), *Millimeter and Submillimeter Waves*, Iliffe, London, 1969, pp. 464–468.
23. Bhartia, P., and I. J. Bahl, *Millimeter Wave Engineering and Applications*, Wiley, New York, 1984, pp. 421–432.
24. Irvine, J., "Detectors and Mixers," in P. Bhartia and I. J. Bahl (Eds.), *Microwave Solid State Circuit Design*, Wiley, New York, 1988, Chap. 11.
25. Anand, Y., and W. J. Moroney, "Microwave Mixer and Detector Diodes," *Proc. IEEE*, Vol. 59, No. 8, Aug. 1971, pp. 1182–1190.
26. Tsui, J., "Tangential Sensitivity of EW Receivers," *Microwave J.*, Vol. 24, No. 10, Oct. 1981, pp. 99–102.
27. Hughes, R. S., "Practical Approach Makes TSS Measurement Simple," *Microwaves*, Vol. 21, No. 1, Jan. 1982, pp. 69–72.
28. Hughes, R. S., "Determining Maximum Sensitivity and Optimum Maximum Gain for Detector-Video Amplifiers with RF Preamplification," *Microwave J.*, Vol. 28, No. 11, Nov. 1985, pp. 165–178.
29. Harrison, R. G., and X. Le Polozec, "Non-Squarelaw Behavior of Diode Detectors Analyzed by the Ritz–Galerkin Method," *IEEE Trans. Microwave Theory Tech.*, Vol. 42, No. 5, May 1994, pp. 840–846.
30. Barnard, D. L., H. Black, and J. A. Thalmann, "Automatic Calibration for Easy and Accurate Power Measurements," *Hewlett-Packard J.*, Vol. 43, No. 2, Apr. 1993, pp. 95–100.
31. Hewlett-Packard, "Impedance Matching Techniques for Mixers and Detectors," Hewlett-Packard Application Note No. 963, Hewlett-Packard, San Jose, CA.
32. Hewlett-Packard, "Dynamic Range Extension of Schottky Detectors," Hewlett-Packard Application Note No. 956-5, Hewlett-Packard, San Jose, CA.
33. Agilent Technologies, "The Zero Bias Schottky Diode Detector at Temperature Extremes—Problems and Solutions," Agilent Technologies Application Note No. 1090, www.semiconductor.agilent.com.
34. Chang, C. C., D. L. Lynch, M. D. Sohigian, G. F. Anderson, T. Schaffner, and G. I. Roberts, "A Zero-Bias GaAs Millimeter Wave Integrated Detector Circuit," *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1982, pp. 206–208.
35. Shillady, R. W., "High Dynamic Range Video Detector," *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1986, pp. 301–304.
36. Fraser, A., "A Planar Doped Barrier Detector for General-Purpose Applications," *Microwave J.*, Vol. 30, No. 5, May 1987, pp. 309–315.
37. Wang, H., W. Lam, T. N. Ton, D. C. W. Lo, K. L. Tan, G. S. Dow, B. Allen, and J. Berenz, "A Monolithic W-Band Preamplifier Diode Detector," *MTT-S Int. Microwave Symp. Dig.*, Vol. 1, 1993, pp. 365–368.

38. Kelly, A. J., "Fundamental Limits on Conversion Loss of Double Sideband Resistive Mixers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-25, No. 11, Nov. 1977, pp. 867–869.
39. Saleh, A. A. M., "Theory of Resistive Mixers," Ph.D. Thesis, Massachusetts Institute of Technology, 1970. Also published as *MIT Research Monograph No. 64*, 1971.
40. Elder, H. E., and V. J. Glinski, "Detector and Mixer Diodes and Circuits," in H. A. Watson (Ed.), *Microwave Semiconductor Devices and Their Circuit Applications*, McGraw-Hill, New York, 1969, Chap. 12.
41. Kerr, A. R., "Low-Noise Room-Temperature and Cryogenic Mixers for 80–120 GHz," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-23, No. 10, Oct. 1975, pp. 781–787.
42. Vernon, F. H., M. F. Millea, M. F. Bottjer, A. H. Silver, R. J. Pedersen, and M. McColl, "The Super-Schottky Diode," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-25, No. 4, Apr. 1977, pp. 286–294.
43. Barber, M. R., "Noise Figure and Conversion Loss of the Schottky Barrier Mixer Diode," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-15, No. 11, Nov. 1967, pp. 629–635.
44. Orloff, L. M., "Intermodulation Analysis of a Crystal Mixer," *Proc. IEEE*, Vol. 52, No. 2, Feb. 1964, pp. 173–179.
45. Kollberg, E., "Mixers and Detectors," in K. Chang (Ed.), *Handbook of Microwave and Optical Components*, Vol. 2, Wiley, New York, 1990, Chap. 2.
46. Henderson, B., "Mixers in Microwave Systems," *Microwave Syst. News*, Oct. 1989, pp. 64–74; Nov. 1989, pp. 71–75.
47. Egami, S., "Nonlinear, Linear Analysis and Computer-Aided Design of Resistive Mixers," *IEEE Trans Microwave Theory Tech.*, Vol. MTT-22, No. 3, Mar. 1974, pp. 270–275.
48. Maas, S., *Microwave Mixers*, Artech House, Dedham, MA, 1986, pp. 154–158.
49. Matthaei, G., L. Young, and E. M. T. Jones, *Microwave Filters, Impedance Matching Networks, and Coupling Structures*, McGraw-Hill, New York, 1964, pp. 843–847.
50. Weir, W. B., and D. K. Adams, "Wideband Multiplexers Using Directional Filters," *Microwaves*, Vol. 8, No. 5, May 1969, pp. 44–50.
51. Maas, S., *Nonlinear Microwave Circuits*, Artech House, Dedham, MA, 1988, p. 220.
52. Tyminski, W. V., and A. E. Hylas, "A Wideband Hybrid Ring for UHF," *Proc. IRE*, Jan. 1953, pp. 81–87.
53. Chang, C.-Y., C.-C. Yang, and D. C. Niu, "A Multioctave Bandwidth Rat-Race Singly Balanced Mixer," *IEEE Microwave Guided Wave Lett.*, Vol. 9, No. 1, Jan. 1999, pp. 37–39.
54. Brauchler, F., S. Robertson, J. East, and L. P. B. Katehi, "W-band Finite Ground Coplanar (FGC) Line Circuit Elements," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. III, June 1996, pp. 1845–1848.
55. Chu, A., W. E. Courtney, and R. W. Sudbury, "A 31-GHz Monolithic GaAs Mixer/Preamplifier Circuit for Receiver Applications," *IEEE Trans. Electron Devices*, Vol. ED-28, No. 2, Feb. 1981, pp. 149–154.

56. Ogawa, H., M. Aikawa, and K. Morita, "K-Band Integrated Double-Balanced Mixer," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, No. 3, Mar. 1980, pp. 180–195.

57. Mariani, E. A., C. P. Heintzman, J. P. Agrios, and S. B. Cohn, "Slot Line Characteristics," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-17, No. 12, Dec. 1969, pp. 1091–1096.

58. Ogawa, H., K. Yamamoto, and N. Imai, "A 26-GHz High-Performance MIC Transmitter-Receiver for Digital Radio Subscriber Systems," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-32, No. 12, Dec. 1984, pp. 1551–1555.

59. Maas, S. A., *Microwave Mixers*, 2nd ed., Artech House, Norwood, MA, 1992, p. 344.

60. Eisenberg, J. A., J. S. Panelli, and W. Ou, "Slotline and Coplanar Waveguide Team to Realize a Novel MMIC Double Balanced Mixer," *Microwave J.*, Sept. 1992, pp. 123–131.

61. Ton, T. N., G. S. Dow, T. H. Chen, M. Lacon, T. S. Lin, and S. Bui, "An X-Band Monolithic Double Double-Balanced Mixer for High Dynamic Receiver Application," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. 1, May 8–10, 1990, pp. 197–200.

62. Chiou, H.-K., and H.-H. Lin, "A Miniature MMIC Double Doubly Balanced Mixer Using Lumped Dual Balun for High Dynamic Receiver Application," *IEEE Microwave Guided Wave Lett.*, Vol. 7, No. 8, Aug. 1997, pp. 227–229.

63. Cohn, M., J. E. Degenford, and B. A. Newman, "Harmonic Mixing with an Antiparallel Diode Pair," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-23, No. 8, Aug. 1975, pp. 667–673.

64. Schneider, M. V., and W. W. Snell, "Harmonically Pumped Stripline Down-Converter," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-23, No. 3, Mar. 1975, pp. 271–275.

65. Henry, P. S., B. S. Glance, and M. V. Schneider, "Local Oscillator Noise Cancellation in the Subharmonically Pumped Downconverter," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-24, May 1976, pp. 254–257.

66. McMaster, T. F., M. V. Schneider, and W. W. Snell, "Millimeter Wave Receivers with Subharmonic Pump," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-24, Dec. 1976, pp. 948–952.

67. Kerr, A. R., "Noise and Loss in Balanced and Subharmonic Pumped Mixers: Parts I & II," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, Dec. 1979, pp. 938–950.

68. Hicks, R. G., and P. J. Khan, "Numerical Analysis of Subharmonic Mixers Using Accurate and Approximate Models," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-30, Dec. 1982, pp. 2113–2119.

69. Weiner, D., J. Griffin, and L. McCarty, "The Image-Rejection Harmonic Mixer," *IEEE MTT-S Int. Microwave Symp. Dig.*, Dallas, June 15–17, 1982, pp. 36–38.

70. Güttich, U., K. M. Strohm, and F. Schäffler, "D-Band Subharmonic Mixer with Silicon Planar Doped Barrier Diodes," *IEEE Trans. Microwave Theory Tech.*, Vol. 39, No. 2, Feb. 1991, pp. 366–368.

71. Künzi, K. F., and H. Berger, "The Development of Subharmonically Pumped Mixers," *AGARD Conf. Proc. Millimeter and Submillimeter Wave Propagation and Circuits*, Sept. 1978, pp. 18-1–18-5.

72. Stephan, K. D., and T. Itoh, "A Planar Quasi-Optical Subharmonically Pumped Mixer Characterized by Isotropic Conversion Loss," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-32, No. 1, Jan. 1984, pp. 97–102.
73. Dixon, S., and R. J. Malik, "Subharmonic Planar Doped Barrier Mixer Conversion Loss Characteristics," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-31, No. 2, Feb. 1983, pp. 155–158.
74. Kok, Y.-L., H. Wang, M. Barsky, R. Lai, M. Sholley, and B. Allen, "A 180-GHz Monolithic Sub-Harmonic InP-Based HEMT Diode Mixer," *IEEE Microwave Guided Wave Lett.*, Vol. 9, No. 12, Dec. 1999, pp. 529–531.
75. Friis, H. T., "Noise Figures of Radio Receivers," *Proc. IRE*, Vol. 32, 1944, pp. 419–429.
76. Haus, H. A., R. Adler, R. S. Engelbrecht, M. T. Lebenbaum, S. W. Harrison, and W. W. Mumford, "Description of Noise Performance of Amplifiers and Receiving Systems," *Proc. IEEE*, Vol. 51, Mar. 1963, pp. 436–442.
77. Held, D. N., and A. R. Kerr, "Conversion Loss and Noise of Microwave and Millimeter-Wave Mixers: Part I—Theory," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-26, No. 2, Feb. 1978, pp. 49–55.

PROBLEMS

11.1 *Flicker Noise.* Assume that the mean-square flicker noise voltage of a diode can be written

$$\langle v_{\text{flick}}^2 \rangle = \int_{f_L}^{f_H} \frac{A}{f}\, df$$

where f_L and f_H are the lower and upper limits of the measurement bandwidth and A is a constant. By considering a very small measurement bandwidth $\delta f \ll f_L$, evaluate A in terms of the noise corner frequency f_n at which the mean-square flicker and shot noise voltages are equal. Hence derive Eq. (11.10).

11.2 *Tangential Signal Sensitivity.* Referring to Fig. 11.6, at the TSS condition

$$V_{\text{pk}} = \Delta V_v^{\text{oc}} - V_{\text{pk}}$$

Assuming that the noise voltage waveform can be viewed as a collection of sinusoids, show that the TSS power level can be written as

$$\text{TSS} \approx \frac{2\sqrt{2}V_n}{\gamma_v} \qquad \text{(W)}$$

Hence derive Eq. (11.29), including the relation between TSS and MDS.

11.3 *Output Spectrum of a Detector.* Verify the expressions in Table 11.1 for the relative amplitudes of the output components of the spectrum of a detector.

11.4 *Detector Voltage Sensitivity.* Assuming a well-matched RF input, find the voltage sensitivity of a detector that uses a Schottky diode with parameters $I_s = 10^{-7}$ A, $n = 1.06$, $C_j(0) = 0.25$ pF, $R_s = 15\ \Omega$, and $R_L = 1$ M Ω when (a) $f = 2$ GHz, $T_0 = 23°C$, $I_{dc} = 0$; (b) $f = 2$ GHz, $T_0 = 23°C$, $I_{dc} = 40\ \mu A$; (c) $f = 10$ GHz, $T_0 = 23°C$, $I_{dc} = 0$; and (d) $f = 10$ GHz, $T = 100°C$, $I_{dc} = 0$.

11.5 *PWL Analysis of a High-Level Detector.* Let a high-level detector diode be represented as a resistor R_s in series with an ideal switch S that is open for reverse bias ($v \leq 0$) and closed for forward bias ($i > 0$). Assume that the video output voltage v_{out} has a constant value V_0 for a given incident RF power P_{inc}. Given that the Thévenin equivalent RF source is a voltage $V_g \cos(\omega t)$ in series with a resistance R_g and that the bias current source is omitted:

(a) Sketch (i) the detector equivalent circuit and (ii) the v_g–i_d transfer function and $v_g(t)$ and $i_d(t)$ waveforms.

(b) Obtain an expression for the time-averaged diode current.

(c) From this, assuming a load R_L, derive Eq. (11.47), which gives the ratio $\alpha = V_0/V_g$.

(d) Hence derive relation (11.48) between V_0 and P_{inc}. Compare the result with Fig. 11.9 when $R_L = \infty$.

11.6 *Detector Impedance.* Find the small-signal impedance of a Schottky detector diode with parameters $I_s = 10^{-8}$ A, $n = 1.05$, $C_j(0) = 0.25$ pF, $R_s = 10\ \Omega$, parasitics $L_p = 1$ nH, and $C_p = 0.25$ pF, and ambient temperature $T_0 = 20°C$ when (a) $f = 2$ GHz, $I_{dc} = 0$; (b) $f = 6$ GHz, $I_{dc} = 0$; and (c) $f = 2$ GHz, $I_{dc} = 50\ \mu A$.

11.7 *Mixer Mismatch Loss.* Referring to Section 11.3.3, derive expression (11.60) for the overall mismatch loss of a mixer:

$$L_1 = 10\left[\log_{10}\frac{(1+S_R)^2}{4S_R} + \log_{10}\frac{(1+S_I)^2}{4S_I}\right] \quad \text{(dB)}$$

where S_R and S_I are the VSWRs at the RF and IF ports, respectively.

11.8 *Exponential-Diode Mixer.* Referring to Section 11.3.5 and assuming a negligibly small RF voltage ($V_R \ll V_L$), show that the DC component of the mixer diode current is

$$I_{dc} \cong \frac{I_s e^{\Lambda(V_0+V_L)}}{\sqrt{2\pi\Lambda V_L}}$$

if $\Lambda V_L \gtrsim 20$. *Hints*:

$$\text{(i)} \qquad e^{x\cos\theta} = \mathbb{I}_0(x) + 2\sum_{k=1}^{\infty} \mathbb{I}_k(x)\cos(k\theta) \quad \text{(Sonine's expansion)}$$

where $\mathbb{I}_k(x)$ is a modified Bessel function of the first kind, order k and argument x.

$$\text{(ii)} \qquad \lim_{x\to\infty} \mathbb{I}_N(x) = \frac{e^x}{\sqrt{2\pi x}} \quad \text{(independently of } N\text{)}$$

11.9 *Image-Reject Mixer Noise.*

(a) Derive expression (11.85) for the noise factor of an ideal image-reject mixer (IRM):

$$F_M^{\text{IRM}} = \left(\frac{T_D}{T_0}\right)(L_r - 1) + 1$$

where T_D is the diode noise temperature, T_0 is the ambient temperature, and L_r is RF-to-IF conversion loss.

(b) Derive expression (11.86) for the noise temperature ratio of an ideal IRM:

$$t_m^{\text{IRM}} = t_w(1 - L_r^{-1}) + L_r^{-1}$$

where t_w is the noise temperature ratio of a Schottky diode.

(c) Derive expression (11.88) for the double-sideband noise factor of an IRM:

$$F_M^{\text{DSB}} = \left(\frac{T_D}{T_0}\right)\left[\left(\frac{1}{L_r} + \frac{1}{L_i}\right)^{-1} - 1\right] + 1$$

(d) Show that the single-sideband noise factor of an IRM is

$$F_M^{\text{SSB}} = \left(\frac{T_D}{T_0}\right)\left[L_r - \left(1 + \frac{L_r}{L_i}\right)\right] + \left(1 + \frac{L_r}{L_i}\right)$$

12

MICROWAVE CONTROL CIRCUITS

K. C. Gupta

Microwave switches, phase shifters, and attenuators are three important "control" circuits used extensively for controlling the signal flow at microwave frequencies. A single phased array radar system, for example, may use thousands of these circuits for precise electronic control of the radiated beam. Traditionally, two types of active devices used commonly in control circuits are *pin* diodes [1] and GaAs MESFETs [2]. Recently, MEMS devices [3] have been developed for RF switches and variable capacitors that are appropriate and offer challenges to *pin* diodes and MESFETs for use in microwave control components. Applications of MEMS in RF circuits are discussed in Chapter 14. The current chapter describes microwave control circuits using semiconductor devices. However, many of these circuit configurations and design approaches are applicable for MEMS circuits also. Before looking at circuit design methods, we briefly look at the equivalent circuit representations for *pin* diodes and GaAs MESFETs.

12.1 *pin* DIODE AND MESFET MODELING FOR CONTROL CIRCUITS

12.1.1 *pin* Diodes

The physics of *pin* diodes has been described in Chapter 8. The equivalent circuit of a packaged *pin* diode may be written as shown in Fig. 12.1. In forward

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

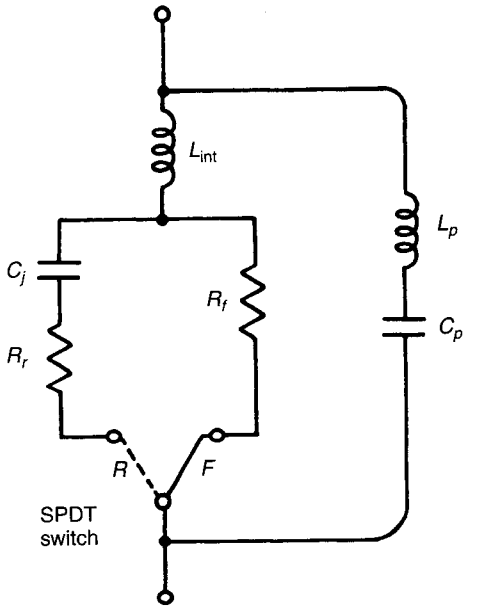


Figure 12.1 Equivalent circuit of packaged *pin* diode.

bias, the single-pole double-throw (SPDT) switch is in position F, and in reverse bias it is put in position R. Inductance L_p and capacitance C_p are contributed by the package. Typical values for two commercially available diodes are given in Table 12.1 (included in Section 12.2.2).

12.1.2 GaAs MESFETs

Two different modes of operation of GaAs MESFETs are used for the design of control circuits at microwave frequencies. These are known as (1) active and (2) passive modes. In the active mode, single-gate or dual-gate MESFETs are used as three-terminal active devices. Equivalent circuits for the active mode are identical to those used for amplifier designs and have been discussed in Chapter 7 (on solid state active devices) and Chapter 10 (on amplifiers).

In the passive mode of operation, MESFETs are used as passive two-terminal devices, with the gate terminal acting as a port for the control signal only. The RF connections are made to the drain and the source terminals only, and the gate terminal looks into an open circuit (high impedance) for the RF signal. The RF impedance between the drain and the source terminals depends upon the DC control voltage at the gate terminal. For switching applications, low-impedance and high-impedance states are obtained by making the gate voltage equal to zero and by using a gate voltage greater (numerically) than the pinchoff voltage, respectively.

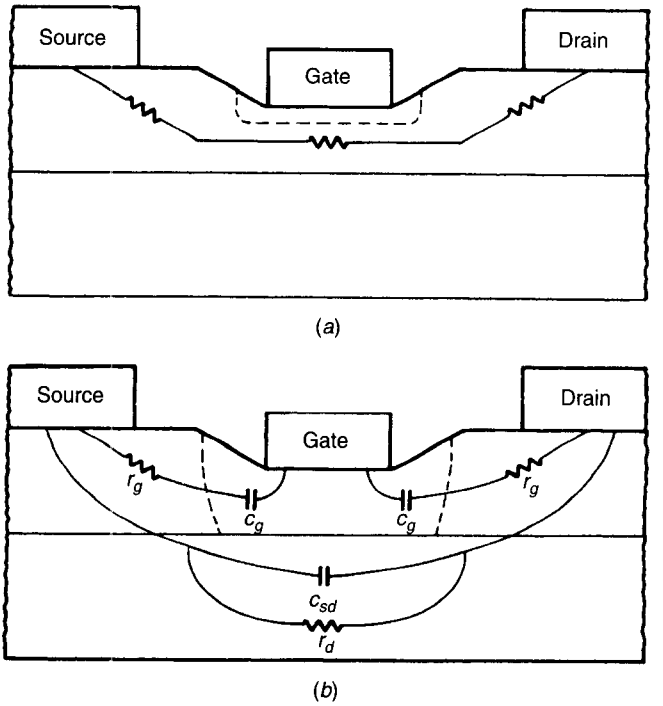


Figure 12.2 Device configurations and equivalent circuits for MESFETs in passive mode. (*a*) Low-impedance state. (*b*) High-impedance state.

For equivalent circuits for these two states, refer to Fig. 12.2. In the low-impedance state, the channel presents a resistive path to the RF current between the drain and the source. The value of this resistance R_{on} can be estimated by dividing the total current path in four different types of regions as discussed by Ayasli [2]. Typical value of R_{on} for a 1×1000 μm^2 gate MESFET suitable for *X*-band operation is about 2.5 Ω at around 10 GHz. Attempts at reducing the value of R_{on} using self-aligned gate technology have been reported [4].

For the high-impedance state of MESFETs operating in the passive mode, an equivalent circuit may be derived by referring to Fig. 12.2*b*. As the channel is now pinched off, the capacitance of the depletion layer (represented by C_g) appears in series between the source and drain terminals. Also in this case, the capacitance C_{sd} (and the leakage resistance r_d) between the source and the drain terminals needs to be incorporated in the equivalent circuit. These components are present in the low-impedance state also, but need not be included in the equivalent circuit because of the small value of R_{on} appearing in parallel. A typical high-impedance-state equivalent circuit is shown in Fig. 12.3. For a 1×1000 μm^2 gate MESFET operating at 10 GHz, typical values of the various elements in this equivalent circuit are $C_{sd} \simeq 0.25$ pF, $r_d \simeq 3$ kΩ, and

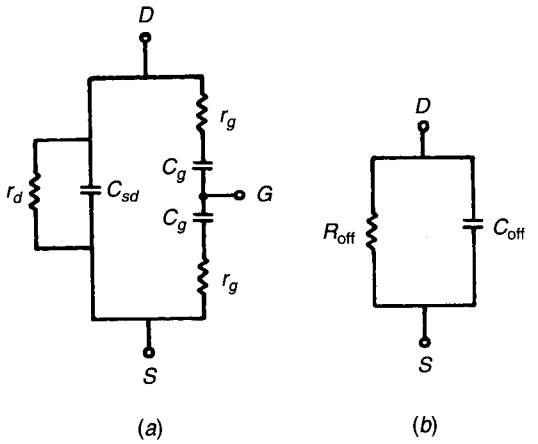


Figure 12.3 (*a*) High-impedance-state equivalent circuit of MESFET in passive mode. (*b*) Simplified circuit for high-impedance state.

$C_g \simeq 0.2$ pF. As r_g is much smaller than the reactance of C_g [$r_g \ll 1/(\omega C_g)$], the equivalent circuit of Fig. 12.3*a* can be replaced by a parallel combination of R_{off} and C_{off} (shown in Fig. 12.3*b*), where

$$C_{\text{off}} = C_{sd} + \frac{C_g}{2} \tag{12.1}$$

$$R_{\text{off}} = \frac{2r_d}{2 + r_d\omega^2 C_g^2 r_g} \tag{12.2}$$

where ω is the operating frequency in radians/second. Typical values for C_{off} and R_{off} are (for 1×1000 μm^2 gate devices at 10 GHz) about 0.2 pF and 2 kΩ, respectively. It may be noted that R_{off} [as indicated by (12.2)] is now a function of the operating frequency ω.

12.2 DESIGN OF SWITCHES

The design of microwave switching circuits using *pin* diodes or GaAs MESFETs in passive mode is discussed in this section.

12.2.1 Basic Configurations

There are two basic configurations that may be used for a simple single-pole single-throw (SPST) switch designed to control the flow of microwave signals along a transmission line. These are shown in Figs. 12.4 and 12.5. In the first case, the switching device (SD) is mounted in series with one of the conductors

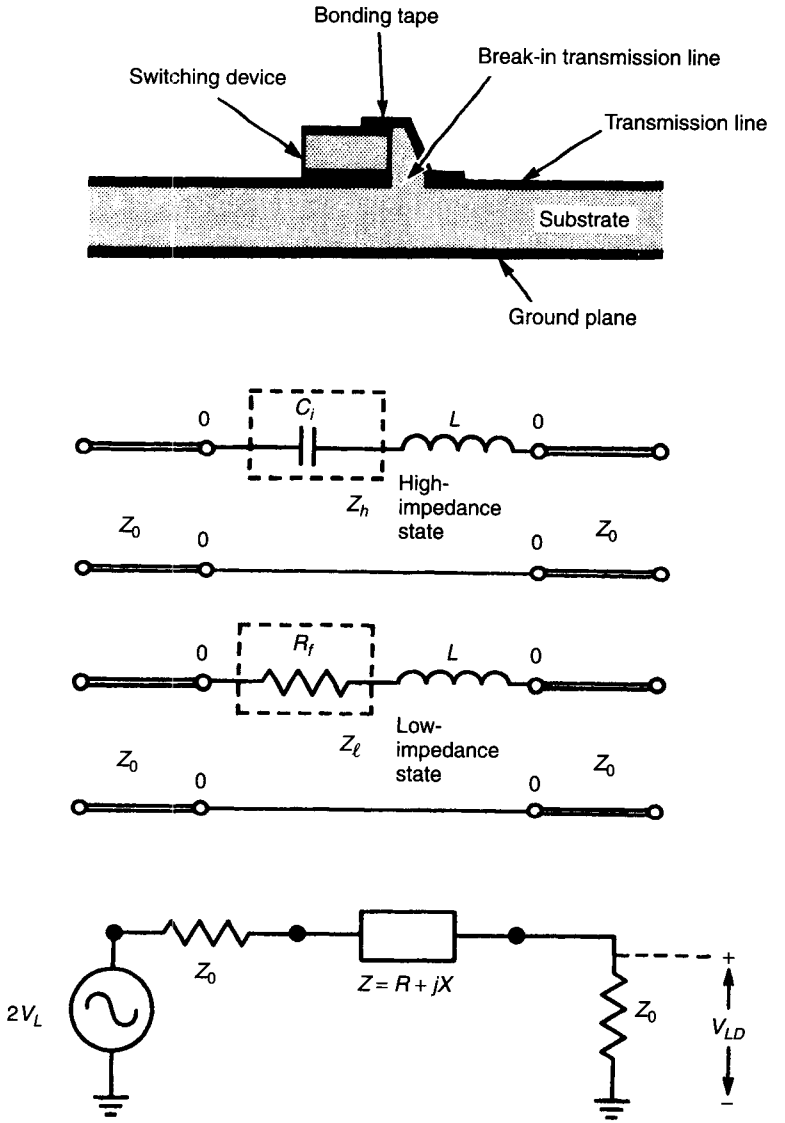


Figure 12.4 Series-mounted switching device in microstrip circuit and simplified equivalent circuits for high- and low-impedance states. Term Z is impedance of switching device SD.

of the transmission line (strip conductor of the microstrip line). In the second case (Fig. 12.5), the switching device is mounted in shunt across the two conductors of a transmission line. Inductances L in Figs. 12.4 and 12.5 are inductances of bonding ribbons, and C_i and R_f constitute the simplified equivalent circuits for the high-impedance and low-impedance states respectively of the device chip.

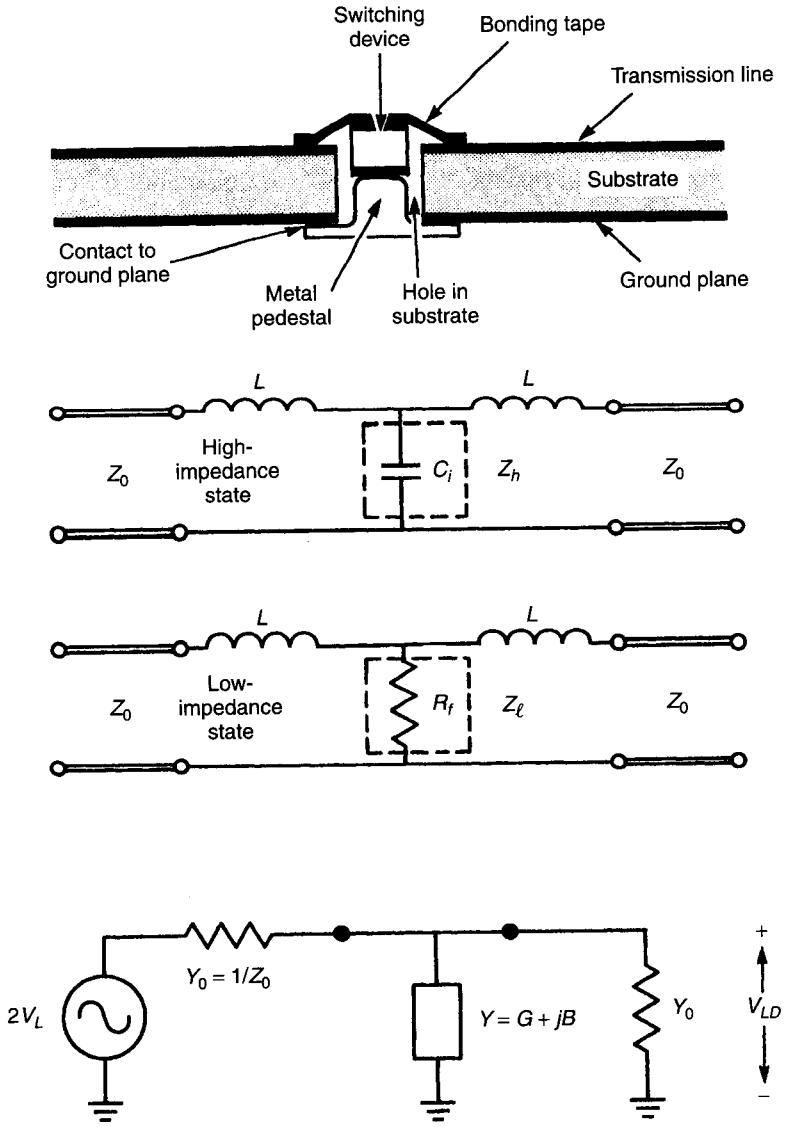


Figure 12.5 Shunt-mounted switching device in microstrip circuit and simplified equivalent circuits for high- and low-impedance states. Y is admittance of the switching device SD.

The two configurations are complementary in the sense that for the series configuration, the low-impedance state of the device allows the signal to propagate; whereas in the shunt-mounted configuration, the signal is delivered to the load when the device is in the high-impedance state. In the OFF states (when no signal is delivered to the output load) of both of these types of switches, the microwave power incident on the switch is mostly reflected back.

However, small fractions of power get (1) dissipated in device resistances or circuit losses, and (2) get transmitted to the load (accounting for a finite isolation or nonzero transmission) because of device/circuit imperfections.

12.2.2 Insertion Loss and Isolation

Because of a finite nonzero impedance of switching devices in the low-impedance state and a definite noninfinite impedance in the high-impedance state, the switching circuits are not perfect. The performance of a practical switch can be expressed by specifying its insertion loss and isolation.

Insertion loss is defined as the ratio of the power delivered to the load in the ON state of the ideal switch to the actual power delivered by the practical switch (in the ON state). It is usually expressed in decibels. The insertion loss may be calculated by considering the equivalent circuits shown in Figs. 12.4 and 12.5. If V_L denotes the actual voltage across the load in the ideal switch, the insertion loss (IL) may be written as

$$\text{IL} = \left|\frac{V_L}{V_{LD}}\right|^2 \tag{12.3}$$

where V_{LD} is the voltage across the load in the practical switch. For the series configuration (Fig. 12.4) simple circuit analysis yields:

$$V_{LD} = \frac{2V_L}{2 + Z/Z_0} \tag{12.4}$$

where Z $(=R + jX)$ is the impedance of the switching device. The insertion loss is given by

$$\begin{aligned}\text{IL} &= \left|\frac{2 + Z/Z_0}{2}\right|^2 \\ &= 1 + \frac{R}{Z_0} + \frac{1}{4}\left(\frac{R}{Z_0}\right)^2 + \frac{1}{4}\left(\frac{X}{Z_0}\right)^2\end{aligned} \tag{12.5}$$

where R and X are the resistance and the reactance of the switching device in the low-impedance state.

For the shunt configuration (Fig. 12.5), the voltage across the load may be written as

$$V_{LD} = \frac{2V_L Y_0}{2Y_0 + Y} \tag{12.6}$$

and the insertion loss becomes

Table 12.1 Equivalent-Circuit Parameters for Two Commercially Available *pin* Diodes

Parameters	MA47892-109	MA47899-030
C_j	1 pF	0.1 pF
R_f	0.4 Ω	1 Ω
R_r	0.5 Ω	4 Ω
L_{int}	0.3 nH	0.3 nH
C_p	0.08 pF	0.18 pF
τ	5 μs	0.5 μs
f_{cs} $[=1/(2\pi C_j\sqrt{R_r R_f})]$	350 GHz	800 GHz

$$\text{IL} = \left|\frac{2Y_0 + Y}{2Y_0}\right|^2 = \left|1 + \frac{G + jB}{2Y_0}\right|^2$$

$$= 1 + \frac{G}{Y_0} + \frac{1}{4}\left(\frac{G}{Y_0}\right)^2 + \frac{1}{4}\left(\frac{B}{Y_0}\right)^2 \tag{12.7}$$

where $Y_0 = 1/Z_0$ and G and B are the real and imaginary parts of the admittance Y of the switching device in the high-impedance state. It may be noted that the similarity of (12.5) and (12.7) originates from the dual nature of series and shunt configurations.

Isolation is a measure of the performance of the switch when it is in the OFF state. Isolation is defined as the ratio of the power delivered to the load for an ideal switch in the ON state to the actual power delivered to the load when the switch is in the OFF state. For the series configuration, the OFF state exists when the device is in the high-impedance state. The isolation in this case is also given by (12.5) with R and X replaced by the corresponding values in the high-impedance state. Similarly, the isolation for the shunt configuration is given by (12.7) when we use the low-impedance-state values for G and B.

Example. As an example, let us calculate insertion loss and isolation for two switches using *pin* diodes with parameters listed in Table 12.1. Package capacitance is ignored and the switches are considered to operate at 3.18 GHz with $Z_0 = 50$ Ω. For MA-47892 we have $R = 0.4$ Ω and $X = \omega L = 2\pi \times 3.18 \times 10^9 \times 0.3 \times 10^{-9} = 6$ Ω in forward bias. The reverse-bias impedance is $0.5 - j44$ Ω.

Series Switch Configuration

$$\text{IL} = 1 + \frac{0.4}{50} + 0.25\left(\frac{0.4}{50}\right)^2 + 0.25\left(\frac{6}{50}\right)^2 = 0.05 \text{ dB} \tag{12.8}$$

$$\text{Isolation} = 1 + \frac{0.5}{50} + 0.25\left(\frac{0.5}{50}\right)^2 + 0.25\left(\frac{44}{50}\right)^2 = 0.805 \text{ dB} \tag{12.9}$$

For this design, the value of isolation is unacceptably low because the reverse-bias impedance is comparable to the 50 Ω characteristic impedance of the transmission line. However, when we use MA-47899 with a smaller reverse-bias capacitance, we have, in forward bias $Z_f = R_f + jX_f = 1 + j6\ \Omega$, and in reverse bias

$$Z_r = R_r + j\left(\omega L_{\text{int}} - \frac{1}{\omega C_j}\right) = 4 - j494\ \Omega$$

These yield

$$\text{IL} = 1 + \frac{1}{50} + 0.25\left(\frac{1}{50}\right)^2 + 0.25\left(\frac{6}{50}\right)^2 = 1.0237 = 0.1\ \text{dB} \tag{12.10}$$

$$\text{Isolation} = 1 + \left(\frac{4}{50}\right) + 0.25\left(\frac{4}{50}\right)^2 + 0.25\left(\frac{494}{50}\right)^2 = 14.06\ \text{dB} \tag{12.11}$$

This performance is certainly more reasonable.

Shunt Switch Configuration. When MA-47892 is used in the shunt configuration, IL and isolation are given by (12.7). We have

$$Y_f = \frac{1}{Z_f} = \frac{1}{0.4 + j6} = 0.011 - j0.166\ \text{S} \tag{12.12}$$

which yields

$$\begin{aligned}\text{Isolation} &= 1 + \frac{0.011}{0.02} + 0.25\left(\frac{0.011}{0.02}\right)^2 + 0.25\left(\frac{0.166}{0.02}\right)^2 \\ &= 18.85 = 12.75\ \text{dB}\end{aligned} \tag{12.13}$$

In reverse bias,

$$Y_r = \frac{1}{Z_r} = \frac{1}{0.5 - j494} = 0.00026 + j0.0272\ \text{S} \tag{12.14}$$

and the insertion loss is calculated to be

$$\begin{aligned}\text{IL} &= 1 + \frac{0.00026}{0.02} + \frac{1}{4}\left(\frac{0.00026}{0.02}\right)^2 + \frac{1}{4}\left(\frac{0.0272}{0.02}\right)^2 \\ &= 1.336 = 1.26\ \text{dB}\end{aligned} \tag{12.15}$$

Corresponding calculations for MA-47899 in shunt configuration yield an isolation of 12.84 dB and an insertion loss of 0.029 dB.

12.2.3 Compensation of Device Reactances

An examination of (12.5) and (12.7), for IL and isolation of series and shunt switches, indicates that the performance of switching circuits is limited by the device reactance X or the susceptance B. Compensation of device reactances therefore provides a mechanism for improving the switch performance.

Let us consider the insertion loss of a shunt-mounted switch. The switching device is in the high-impedance state. In this state, both *pin* diodes and GaAs MESFETs (passive mode) and also MEMS devices discussed in Chapter 14 may be represented by a parallel combination of a high resistance R and a small capacitance C. The total admittance of the high-impedance state can be reduced by connecting an inductive susceptance of an equal magnitude in parallel with the capacitance. This can be achieved either by mounting a lumped inductor or by incorporating a shorted (shorter than quarter-wave) stub. These arrangements are shown in Fig. 12.6. At the design frequency, the device capacitance C_i and externally added inductance L_c form a parallel resonant circuit, and the switching device presents only a resistive impedance R_i across the transmission line of characteristic impedance Z_0. Thus at the resonance of L_c and C_i, B in (12.7) becomes zero and the insertion loss of the switch may be written as

$$\text{IL} = 1 + \frac{G}{Y_0} + \frac{1}{4}\left(\frac{G}{Y_0}\right)^2 \tag{12.16}$$

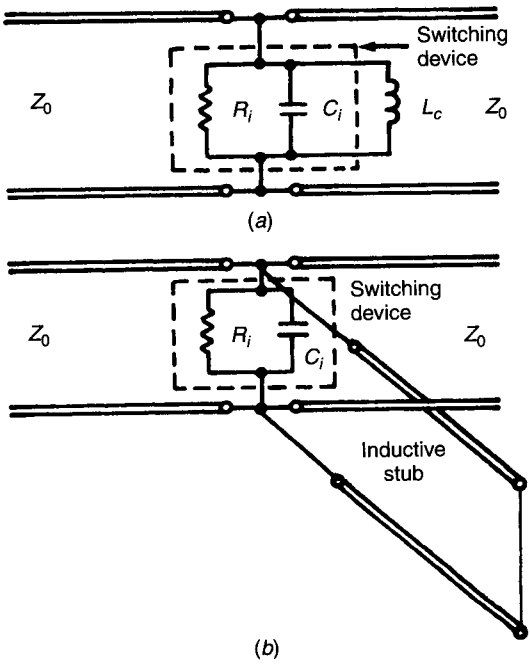


Figure 12.6 Compensation of capacitance of switching device in high-impedance state by using (*a*) a lumped inductance and (*b*) an inductive stub.

If this inductive compensation is introduced in the example of the *pin* diode switch considered earlier, we find that the insertion loss of the shunt switch using MA-47892 is reduced to

$$\text{IL} = 1 + \frac{0.00026}{0.02} + \frac{1}{4}\left(\frac{0.00026}{0.02}\right)^2 = 1.01304 = 0.0563 \text{ dB} \qquad (12.17)$$

from the earlier value of 1.26 dB calculated in (12.15).

Let us consider another example of a MESFET switch with device parameters at 10 GHz being

$$R_i = R_{\text{off}} = 3 \text{ k}\Omega \qquad C_i = C_{\text{off}} = C_{sd} + \frac{C_g}{2} = 0.25 \text{ pF}$$

Without any compensating inductance, the insertion loss for a shunt-mounted switch is found to be

$$\begin{aligned}\text{IL} &= 1 + \frac{0.00033}{0.02} + \frac{1}{4}\left(\frac{0.00033}{0.02}\right)^2 + \frac{1}{4}\left(\frac{2\pi 10^{10}(0.25 \times 10^{-12})}{0.02}\right)^2 \\ &= 1 + 0.01667 + 0.00007 + 0.15421 = 1.17095 = 0.69 \text{ dB} \qquad (12.18)\end{aligned}$$

When the capacitance is compensated by a suitable value of L_c, the insertion loss reduces to

$$\text{IL} = 1 + \frac{0.00033}{0.02} + \frac{1}{4}\left(\frac{0.00033}{0.02}\right)^2 = 1.01674 = 0.0721 \text{ dB} \qquad (12.19)$$

Introduction of a compensating inductance L_c in MESFET switching circuits is easier because there is no DC potential difference between the drain and the source terminals. Resonated GaAs FET devices are used extensively in microwave switching circuits [5]. However, when a compensating inductance is to be introduced in a *pin* diode or MEMS switching circuit, we need a DC blocking capacitor in series with L_c (lumped or stub) so that the reverse bias on the *pin* diode or the MEMS device is not disturbed.

Isolation of shunt-mounted switches can also be improved by compensating the inductive reactance of the switching device in the low-impedance state by adding a series capacitance of suitable value.

12.2.4 Single-Pole Double-Throw (SPDT) Switches

Single-pole double-throw switches require a minimum of two switching devices. Two basic configurations for SPDT switches, namely series mounted and shunt mounted, are shown in Fig. 12.7. In the series configuration, the input signal is routed to output 1 when the switching device SD1 is in the low-impedance state and device SD2 is in the high-impedance state. In the shunt configuration

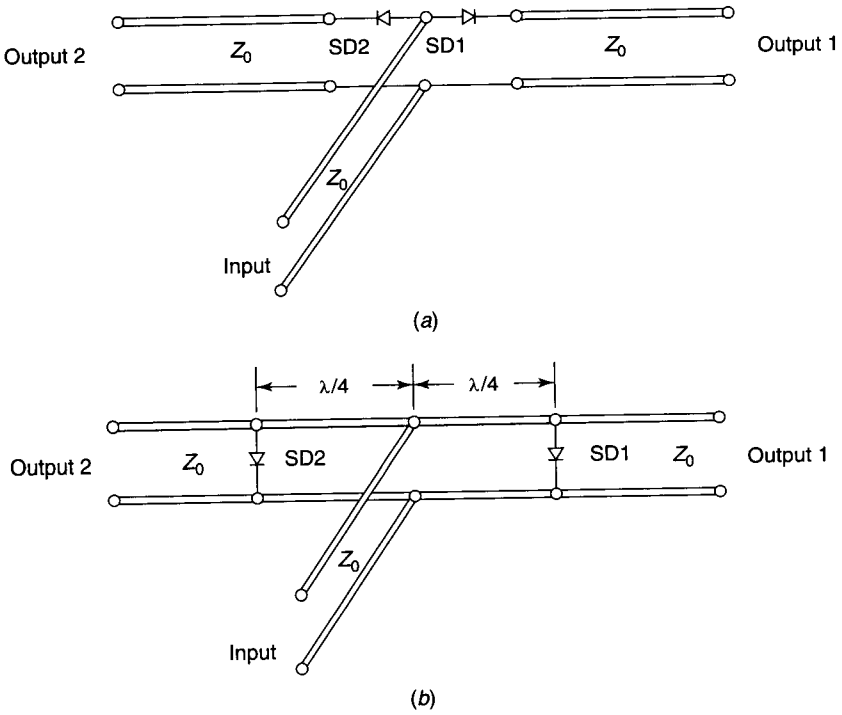


Figure 12.7 (*a*) SPDT switch using series-mounted devices. (*b*) SPDT switch using shunt-mounted devices.

shown in Fig. 12.7*b*, the signal is routed to output 1 when device SD1 is in the high-impedance state and device SD2 is in the low-impedance state. Thus, in either configuration, at any time, one of the devices is in the low-impedance state while the other one is in the high-impedance state.

The bandwidth of the SPDT in shunt configuration is limited because of the $\frac{1}{4}\lambda$ line lengths required between the transmission-line junction and the locations of the two switching devices. Figure 12.8 shows the performance of two SPDT configurations when the switching devices are MA-47899 *pin* diode chips. The design is centered around 3 GHz frequency. For a shunt-mounted switch, the variation of insertion loss with frequency limits the operating bandwidth. An example of a shunt-mounted SPDT switch using MESFETs is shown in Fig. 12.9. The design frequency in this case is 10 GHz, and the improvement in performance by adding two compensating inductances is also shown in the figure.

The design concept of SPDT switches may also be extended to single-pole multiple-throw switches. An SP3T switch [6] will use a minimum of three switching devices. Microstrip construction of a typical SP3T switch is shown in Fig. 12.10. In this configuration there are six devices, two associated with each

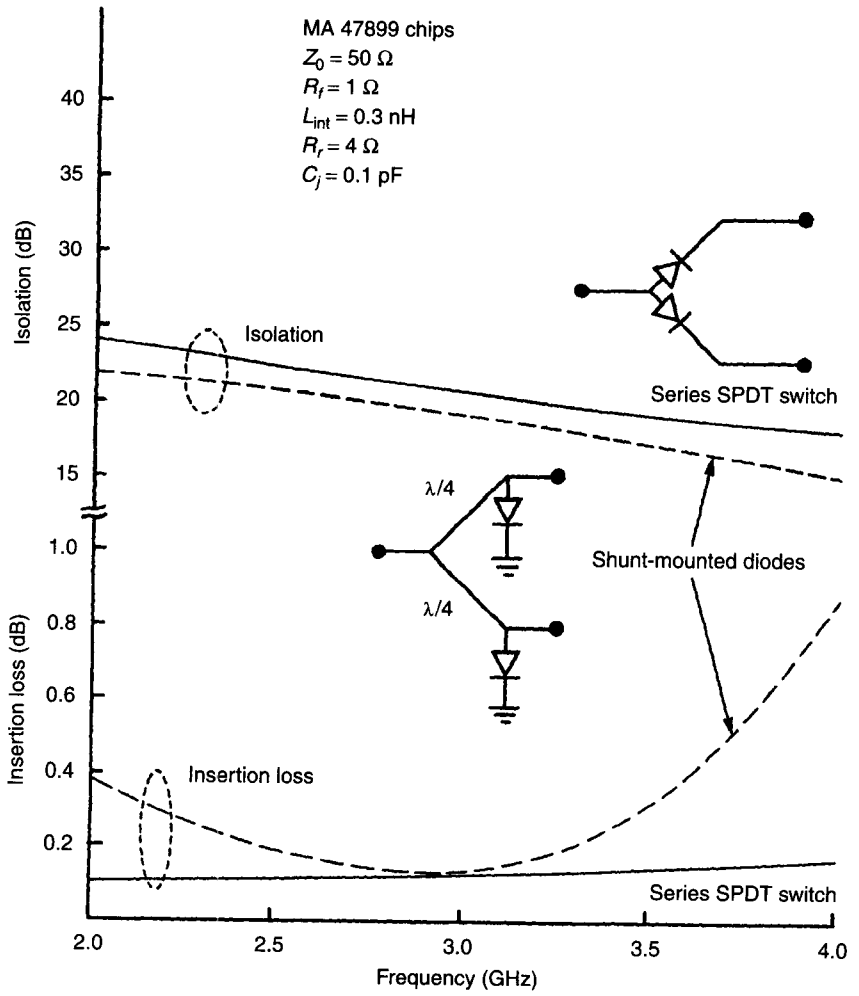


Figure 12.8 Typical insertion loss and isolation performance of SPDT switches using two *pin* diodes.

of the output ports. Using a combination of series- and shunt-mounted diodes improves the switch's performance. The DC blocking capacitors and RF chokes are needed for DC biasing.

12.2.5 Series–Shunt Switching Configurations

In Section 12.2.1, we looked at the characteristics of two basic switching configurations using series-mounted and shunt-mounted switching devices. One could realize a better switching performance when both series- and shunt-

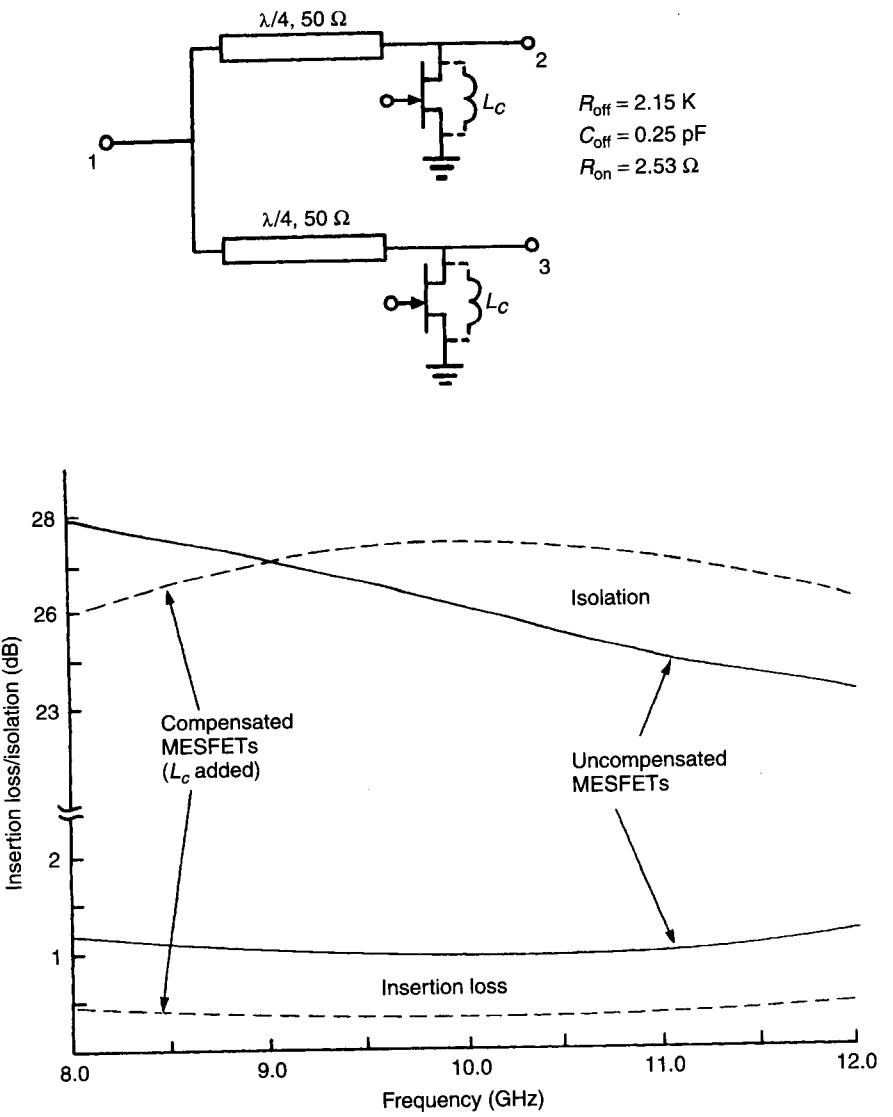


Figure 12.9 Performance of SPDT switch using two shunt-mounted MESFETs.

mounted devices are included in a single circuit (as shown for the SP3T switch of Fig. 12.10).

The simplest series–shunt switching configuration is shown in Fig. 12.11*a*. This switch is ON when the series device is in the low-impedance state and the shunt device is in the high-impedance state. In the OFF state of the switch, the series device is in the high-impedance state and the shunt device is in the low-impedance state. This switching circuit may be analyzed in terms of the equiv-

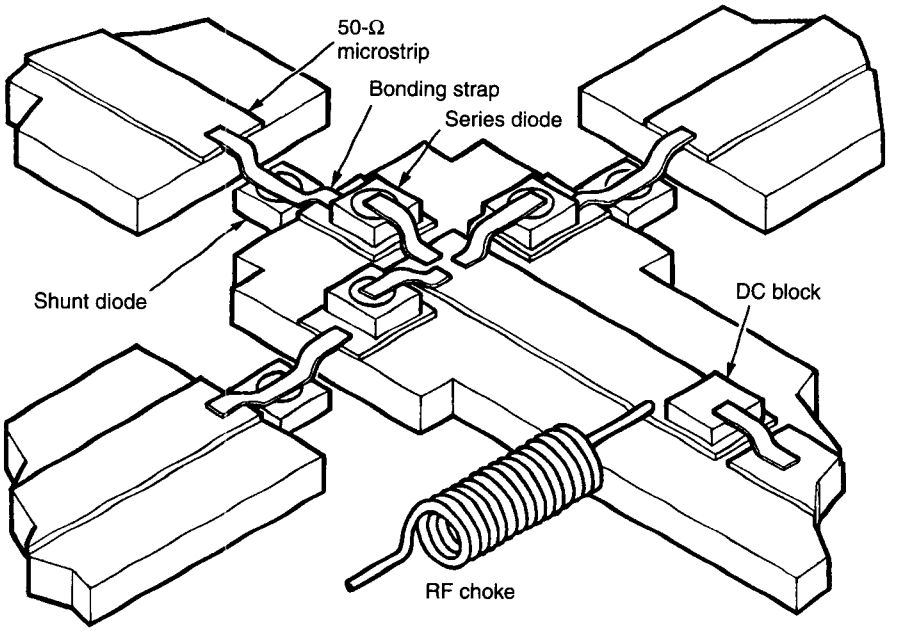


Figure 12.10 Microstrip construction of SP3T switch using three series-mounted and three shunt-mounted devices. (After Chorney [6]. Reprinted with permission of Horizon House Publications.)

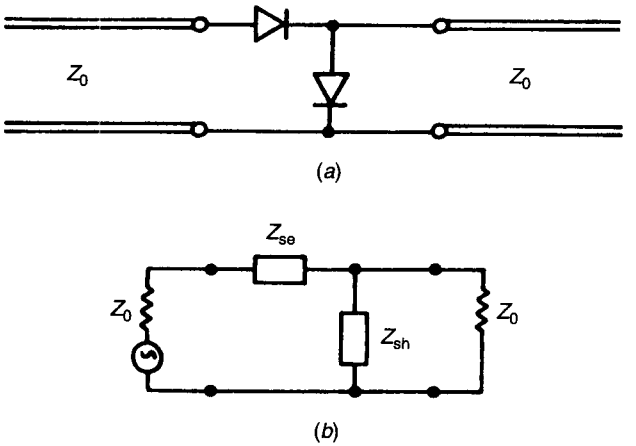


Figure 12.11 (*a*) Simplest series–shunt switching configuration. (*b*) Equivalent circuit for the series–shunt configuration shown.

alent circuit shown in Fig. 12.11*b*. As shown in this figure, Z_{se} is the impedance of the series-mounted device and Z_{sh} is the impedance of the shunt-mounted device. For the ON state the device impedance Z_{se} is denoted by the low-impedance Z_l, and the device impedance Z_{sh} is denoted by the high-impedance Z_h. From simple circuit analysis, the insertion loss may be written as

$$\text{IL} = \left| \frac{1}{2} + \frac{(Z_0 + Z_h)(Z_0 + Z_l)}{2Z_0 Z_h} \right|^2 \tag{12.20}$$

Similarly,

$$\text{Isolation} = \left| \frac{1}{2} + \frac{(Z_0 + Z_l)(Z_0 + Z_h)}{2Z_0 Z_l} \right|^2 \tag{12.21}$$

Expressions (12.5) and (12.7) for series and shunt switches may be obtained as limiting cases of (12.20) for Z_{sh} $(=Z_h) \to \infty$ and for Z_{se} $(=Z_l) \to 0$, respectively. It may be noted that if nonidentical devices are used in the series and the shunt locations, values of Z_h and Z_l in (12.20) could be different from those in (12.21). As an example, let us calculate the performance of a series–shunt switch (using MA-47899 *pin* diodes in chip form and operating at 6.37 GHz) and compare the results with those for simple series and shunt switches using the same devices. This comparison is depicted in Table 12.2. We note that isolation obtained with a series–shunt configuration is much better (more than twice in decibels) than that for either the series or shunt switch. The insertion loss for the series–shunt configuration is worse than that for a shunt switch but better than that for a series switch.

At first glance it looks surprising that a switch using two (lossy) *pin* diodes can have a smaller insertion loss than that with a single diode. However, a detailed analysis shows that use of a series–shunt switch reduces the reflection loss (compared to that for a series switch) and thereby improves the insertion loss.

Wide-Band Series–Shunt Switch Configurations. Use of multiple switching devices in a series–shunt configuration can lead to ultra-wide-band

Table 12.2 Comparison of Three Switching Configurations (with MA-47899 Chip *pin* Diodes at 6.37 GHz)

Configuration	Insertion Loss (dB)	Isolation (dB)
Series–shunt (2 devices)	0.108	20.17
Series switch (2 device)	0.147	8.29
Shunt switch (1 device)	0.063	7.52

switches. The basic concept involves the use of a ladder network structure that behaves like a low-pass filter when series devices are in the low-impedance state (inductive) and shunt devices are in the high-impedance state (capacitive). When the bias levels on the series- and shunt-switching devices are interchanged, the network behaves as a high-pass filter providing a high insertion loss below the cutoff frequency. If this cutoff frequency of the high-pass filter configuration is arranged to lie slightly above the cutoff frequency of the low-pass configuration, the network behaves as a switch with the bandwidth of the low-pass filter configuration.

Let us consider an example of a three-device T-structure switch shown in Fig. 12.12*a*. In the ON state of the switch, the series devices are in the low-impedance state and shunt devices are in the high-impedance state. Wire-bonded devices exhibit a series inductance in the low-impedance state leading to an equivalent circuit of the switch as shown in Fig. 12.12*b*. If the effect of device resistances is ignored, the circuit reduces to the low-pass filter configuration shown in Fig. 12.12*c*. The cutoff frequency and the impedance level of this low-pass filter configuration are obtained by comparing with a three-element maximally flat low-pass design. For a prototype filter of this type, we have $L = L_s = 1$ H and $C = C_j = 2$ F (corresponding to $Z_0 = 1\ \Omega$ and cutoff frequency $\omega_c = 1$ rad/s).

Transforming this prototype to have a cutoff frequency ω_c and an input–output impedance level of $Z_0\ \Omega$, we have

$$L = \frac{Z_0}{\omega_c}\ \text{H} \tag{12.22}$$

$$C = \frac{2}{Z_0\omega_c}\ \text{F} \tag{12.23}$$

Thus for $L = L_s$ and $C = C_j$, the cutoff frequency ω_c and the impedance level Z_0 are given by

$$\omega_c = \sqrt{\frac{2}{L_s C_j}} \tag{12.24}$$

$$Z_0 = \sqrt{\frac{2L_s}{C_j}} \tag{12.25}$$

Taking typical values of L_s and C_j to be 0.2 nH and 0.05 pF (typical of chip or beam-lead *pin* diodes), we get a 3-dB cutoff frequency $f_c\ (=\omega_c/2\pi) = 71.18$ GHz and an impedance level $Z_0 = 89.4\ \Omega$. It may be noted that adjustments of ω_c and Z_0 are possible only by changing L_s and C_j. For *pin* diodes, the capacitance C_j depends on the junction area. Unless the designer has an in-house device fabrication facility, only a very limited choice of C_j can be obtained from commercially available devices. On the other hand, L_s is primarily con-

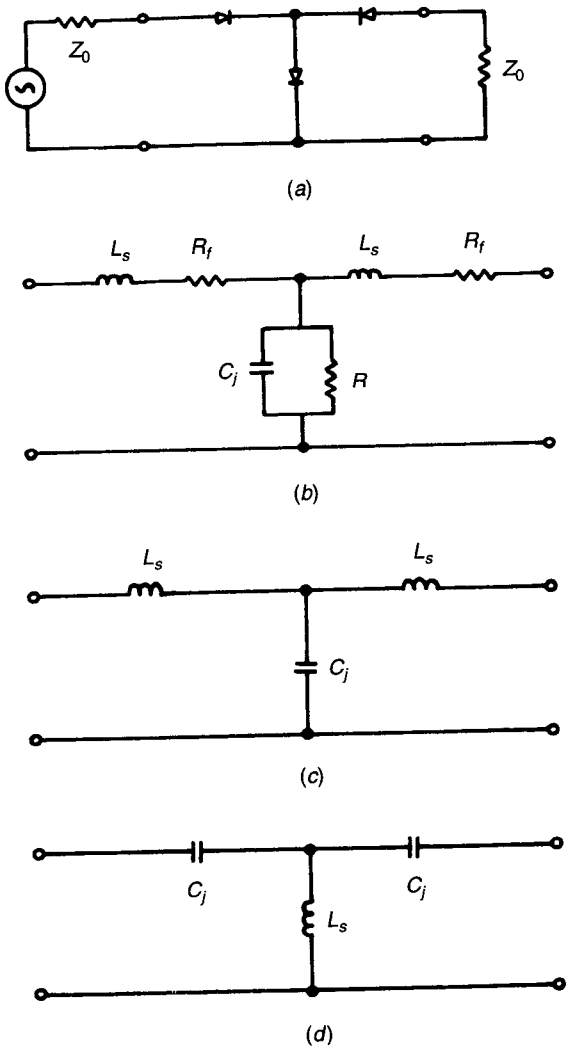


Figure 12.12 (*a*) Series–shunt switch using three switching devices. (*b*) Equivalent circuit with series devices in low-impedance state and shunt device in high-impedance state (ON state for switch). (*c*) Simplified equivalent circuit (ignoring resistances) for ON state. (*d*) Simplified equivalent circuit for switch in OFF state.

tributed by the bonding wire inductance and can be adjusted (to some extent) by changing the length of the bonding wire.

When the series devices (Fig. 12.12*a*) are in the high-impedance state and the shunt device is in the low-impedance state, the switch is in the OFF state. The equivalent circuit for this state is shown in Fig. 12.12*d* and corresponds to a high-pass *LC* filter circuit. Low-frequency cutoff and the impedance level of

this high-pass filter may be obtained by comparing it with a prototype ($\omega_c = 1$ rad/s and $Z_0 = 1\ \Omega$) three-element high-pass filter. For a prototype filter circuit $C = C_j = 1$ F and $L = L_s = 0.5$ H. Transforming the cutoff frequency to ω_c and the impedance level to $Z_0\ \Omega$, we get

$$C = \frac{1}{\omega_c Z_0} \text{ F} \tag{12.26}$$

$$L = \frac{0.5 Z_0}{\omega_c} \text{ H} \tag{12.27}$$

In terms of L_s and C_j, we obtain

$$\omega_c = \sqrt{\frac{0.5}{L_s C_j}} \tag{12.28}$$

$$Z_0 = \sqrt{\frac{2L_s}{C_j}} \tag{12.29}$$

Comparing the result with the corresponding results for the ON state, we find that whereas Z_0 has the same value as for the low-pass filter (12.25), the value of the cutoff frequency for the high-pass filter (OFF state) is only half that of the low-pass filter corresponding value (12.24) to the ON state. Since the low-frequency cutoff in the OFF state implies only 3-dB isolation, the operating range of the switch is still smaller. This operating range can be increased by choosing the shunt-mounted device such that the product $L_s C_j$ for the shunt device is one-fourth of that for the series-mounted device, but the ratio L_s/C_j is equal in two cases. This selection will make values of the two cutoff frequences equal.

Considering an example with three identical switching devices (with $L_s = 0.2$ nH, $C_j = 0.1$ pF, and $R_f = 0.5\ \Omega$), the insertion loss and isolation of the series–shunt switch are plotted in Fig. 12.13 as a function of frequency up to 30 GHz. We note that the insertion loss is less than 0.1 dB throughout this range, but the performance is limited by the value of isolation, which decreases monotonically. An isolation better than 20 dB is obtained for frequencies up to 12.4 GHz.

Five-Device Series–Shunt Configuration. Performance obtained from the series–shunt switch configuration discussed previously can be improved by increasing the number of switching devices from three to five as shown in Fig. 12.14 (inset). When five identical devices similar to that for the three-device switch are used, the insertion loss and isolation obtained are shown in Fig. 12.14. It may be noted that the isolation is improved for frequencies below 19 GHz and is better than 20 dB for frequencies up to 16.3 GHz. In this frequency range, the insertion loss is only marginally worse ($\simeq$0.1 dB in place of $\simeq$0.08 dB for the three-device case when $R_f = 0.5\ \Omega$).

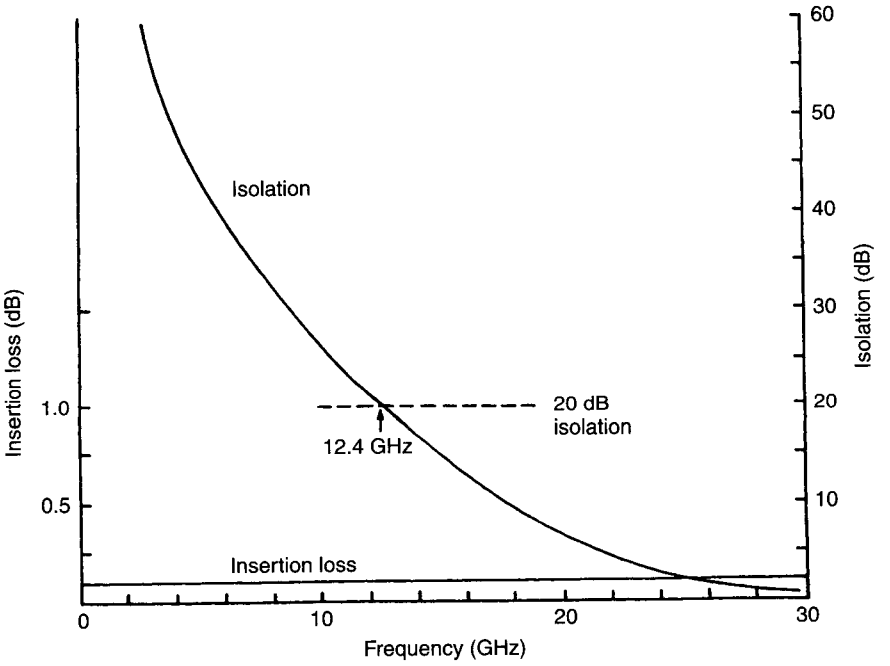


Figure 12.13 Insertion loss and isolation performance of three-device series–shunt switch circuit.

It may be noted that when five identical devices are used, the ON and the OFF state equivalent circuits do not correspond exactly to the five-element low-pass and high-pass filter designs. For realizing these filter configurations, as described in Chapter 6, three different types of switching devices (with different L_s and C_j values) are required. At times, this selection becomes difficult because of the limitations on parameter values of available devices.

12.2.6 Switching Speed Considerations

The time required to change the state of a switch from OFF to ON (or vice versa) becomes an important consideration in several switching circuits, such as those used for modulation or for ultra-high-speed digital circuits. Some of the switching speed considerations are discussed briefly in this section.

Definitions. Different terms, like turn-on delay, turn-on switching time, turn-off delay, and turn-off switching time, are used for microwave switches. These may be explained with the help of Fig. 12.15*a* and *b*. Figure 12.15*a* shows an experimental setup for observing switching speed on a cathode-ray oscilloscope (CRO). In an alternative experimental configuration, the crystal detector and the dual-trace oscilloscope are replaced by a sampling oscilloscope used for

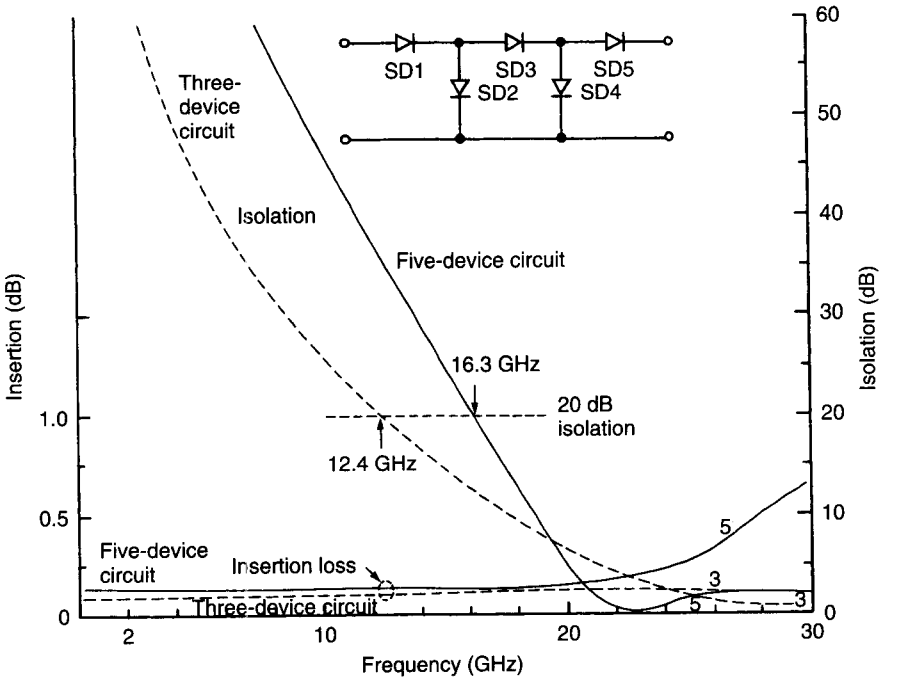


Figure 12.14 Insertion loss and isolation performance of five-device series–shunt switching configuration and its comparison with performance of a three-device switch (shown in dotted curves).

direct observation of microwave waveforms. In either case, the two waveforms significant for measurement of switching speed are the input from the pulse generator to the switch driver and the envelope of (or the detected) microwave signal. By comparing these waveforms on the calibrated time axis (Fig. 12.15*b*), we can define various switching speed terms. *Turn-on delay* is the time interval between the instants when the input signal reaches 90% of its peak value and when the detected RF envelope reaches 10% of its peak value. *Turn-on switching speed* is defined as the time interval between the instants when the envelope of the RF output rises from 10% of its peak value to 90% of its peak value. *Turn-off delay* and *turn-off switching* are defined in a similar manner and are shown in Fig. 12.15*b*. Ten percent and 90% points on the detected RF envelope correspond to 10- and 0.5-dB points when a square law detector is used to measure the output power, but correspond to 20- and 1-dB points, respectively, when a sampling oscilloscope is used. Thus for interpreting switching speed results, it is necessary to know the experimental setup used.

Speed Limitations Imposed by Switching Devices. When *pin* diodes are used for microwave switching, the main factor limiting the switching speed is the time required to remove the charge from the intrinsic region when the diode

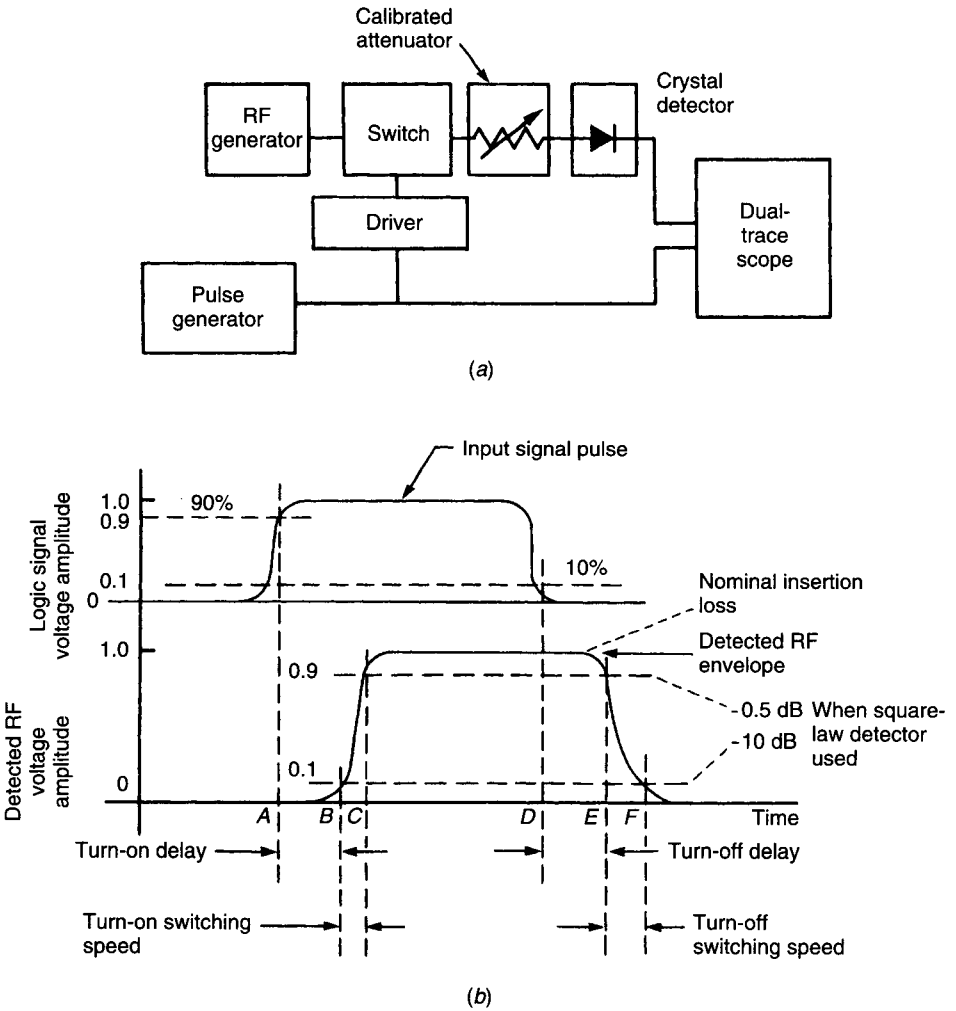


Figure 12.15 (*a*) Experimental setup for measuring speed of switching circuits. (*b*) Illustration of various delay terms used for microwave switches.

bias is switched from forward to reverse. This charge removal time depends on the width (thickness) of the intrinsic layer. A decrease in the width W of the intrinsic layer makes the charge removal faster but at the same time reduces the reverse breakdown voltage and hence the power-handling capability of the *pin* diode switches [7]. Since the power-handling capability is proportional to the square of the breakdown voltage, it is also proportional to the square of width W. On the other hand, switching time τ is proportional to width W. These simple (although crude) arguments point out a trade-off between switching time τ and the square root of the power-handling capability P. Garver [7] has shown

that approximately

$$\tau = \frac{\sqrt{P}}{25} \text{ ns} \tag{12.30}$$

when P is the power-handling capability in watts. Practical values of τ are about 25 times those given by relation (12.30).

Enhancement in switching speed of *pin* diodes can be obtained by using GaAs *pin* diodes [8] in place of the Si *pin* diodes commonly used. Electron mobility in GaAs is about four times that in silicon. This leads to faster switching times as well as lower drive current requirements for GaAs *pin* diodes. Switching speeds of MEMS switches (Chapter 14) are much smaller than semiconductor devices.

Switching Speed Limitation Imposed by Biasing Network. In addition to the switching speed limitation imposed by the semiconductor devices used in switching circuits, switching speed may also be affected by circuit considerations. Considering the DC bias terminal(s) as a separate port, a single-pole single-throw switching circuit may be viewed as a three-port network [9], as shown in Fig. 12.16. A low-pass filter arrangement is needed at the bias port to ensure that the RF signal does not leak away via the bias port. High-pass filters are needed at the RF input and output ports to ensure that the DC bias (or switching pulse) does not interfere with the other parts of the circuit. In the simplest form, these high-pass filters may simply be DC blocking capacitors at either end. These filters increase the rise time of the switching pulse and thereby reduce the switching speed. Rise time τ of a pulse (10–90% levels) passing through a filter is related to the 3-dB bandwidth (BW) of the filter, as

$$\tau = \frac{0.44}{\text{BW}} \tag{12.31}$$

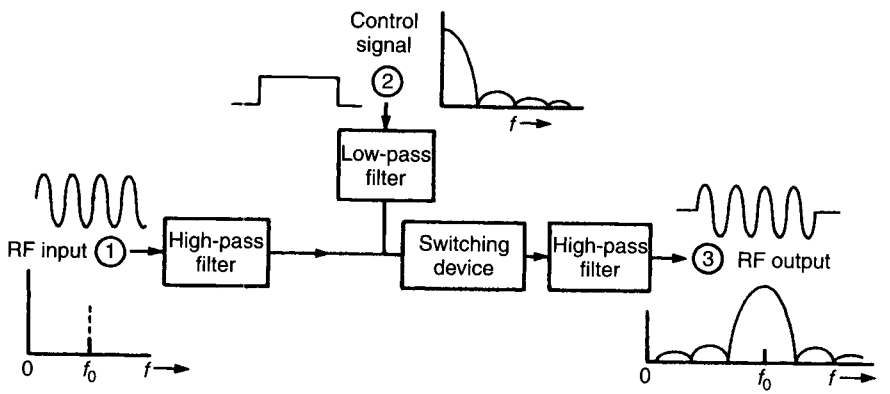


Figure 12.16 Three-port network representation of switching circuits.

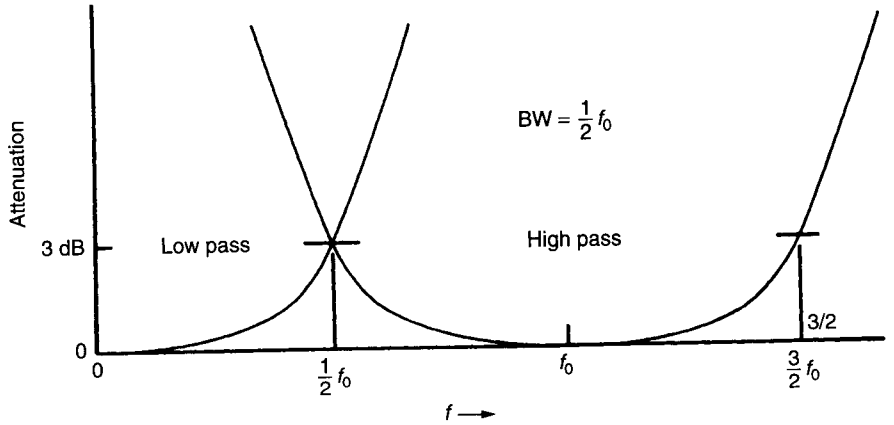


Figure 12.17 Optimum filter characteristics for fast switching considerations.

Thus, in order to decrease the rise time, the 3-dB bandwidth of the filters should be as large as possible. Since we have a low-pass filter and a high-pass filter in the path of the switching pulse, optimum filter characteristics are as shown in Fig. 12.17. With this arrangement, the switching time contributed by filters is given [9],

$$\tau = \sqrt{\tau_1^2 + \tau_2^2} = \sqrt{2}\tau_F = \frac{1.24}{f_0} \tag{12.32}$$

where τ_1 and τ_2 are rise times for the two filters (we assume $\tau_1 = \tau_2 = \tau_F$) and f_0 is the operating frequency. Thus at an operating frequency of 1 GHz, the minimum switching delay contributed by biasing filters is about 1.24 ns.

Switching speed limitation caused by the biasing network becomes a limiting factor for MESFET switches because the MESFET device itself is inherently much faster than *pin* diodes.

A detailed overview of switches including circuit types, theory, power handling, dynamic range, and biasing schemes can be found in Chang et al. [10].

12.3 DESIGN OF PHASE SHIFTERS

12.3.1 General

A phase shifter is a two-port network with the provision that the phase difference between the output and the input signals may be controlled by a control signal (DC bias). Various phase shifter designs may be classified as shown in Fig. 12.18. Phase shifters are called *digital* when the differential phase shift can

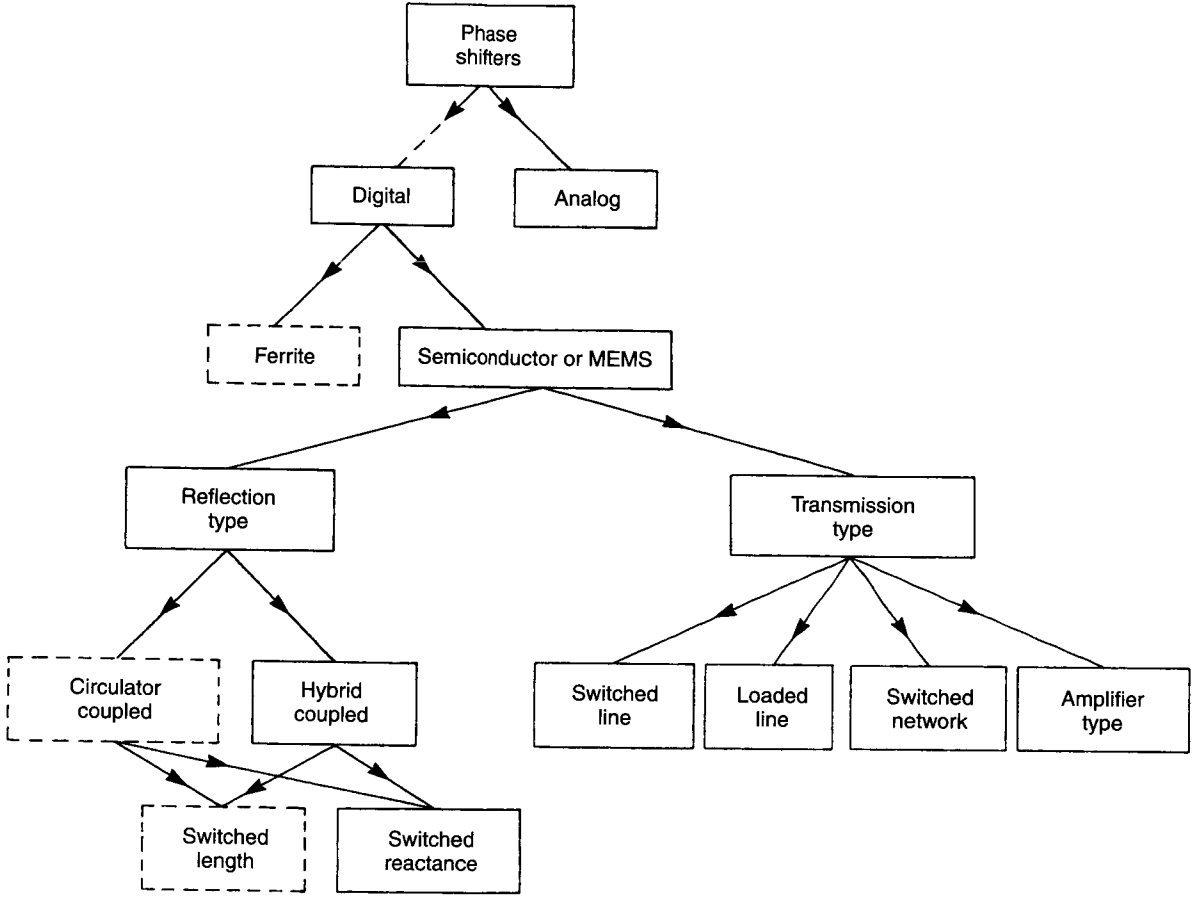


Figure 12.18 Various types of microwave phase shifters.

be changed by only a few predetermined discrete values, such as 180°, 90°, 45°, 22.5°, and 11.25°. On the other hand, in *analog* phase shifters, the differential phase shift can be varied in a continuous manner by a corresponding continuous variation of the control signal. Digital phase shifters find extensive applications in phased-array antenna systems. Phase control of the signals fed to the various elements of the array allows the direction of the radiated beam to be scanned electronically. Digital phase shifters are more compatible with the computer control of beam scanning in phased-array antenna systems.

There are two distinct methods for designing digital phase shifters at microwave frequencies. One is to use the properties of ferrimagnetic materials for obtaining switchable phase shift. Ferrite phase shifters had gone through rapid developments during the 1960s, and a fairly good description of ferrite phase shifter design is available in Skolnik's *Radar Handbook* [11]. The other major design for digital phase shifters uses semiconductor or MEMS devices. These phase shifters are, in general, more compact, have lower switching times, and require lower drive power when compared to ferrite phase shifters. Here, we discuss semiconductor phase shifters only.

As shown in Fig. 12.18, phase shifters using semiconductor devices can be either of the reflection type or the transmission type. In *reflection-type* shifters, the basic design unit is a one-port network, and it is the phase shift of the reflected signal that is changed by the control signal. These basic one-port phase shifters can be converted into useful two-port components either by using a circulator or a hybrid. Because of the ease of integration, the hybrid-coupled reflection-type phase shifters are more common. As far as the design of the basic one-port reflection phase shifter bit is concerned, one may use a switched-length type or a switched-reactance type of design. Other designs for semiconductor phase shifters may be called transmission type. Here there are three subgroups: switched-line type of phase shifters, loaded-line phase shifters, and switched-network type of phase shifters. A special class of "transmission" phase shifters, called amplifier type, use GaAs MESFETs in an active amplifier mode. This type of phase shifter is nonreciprocal and cannot be designed by using *pin* diodes or MEMS devices. Design methods for the various types of phase shifters are discussed in this section.

12.3.2 Switched-Line Phase Shifters

Design of switched-line type of phase shifters is conceptually the simplest. The basic configuration for a single-bit phase shifter is shown in Fig. 12.19. Two SPDT switches are used to route the signal via one of the two alternative transmission path lengths l_1 or l_2. When the signal passes through the longer path, it goes through an additional phase delay given by

$$\Delta\phi = \beta(l_2 - l_1) = \frac{2\pi f}{v_p}(l_2 - l_1) \tag{12.33}$$

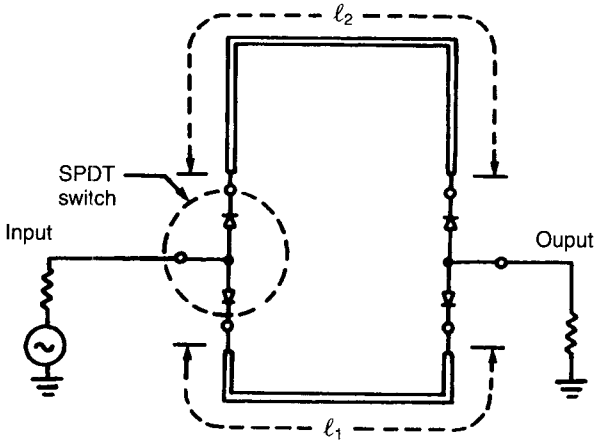


Figure 12.19 Single-bit switched-line phase shifter.

where v_p is the phase velocity. The insertion loss of a phase shifter of this type is equal to that of the two SPDT switches (plus any line losses that may be present).

An interesting characteristic of this type of phase shifter, as can be seen from (12.33), is that the differential phase shift $\Delta\phi$ is directly proportional to frequency. Because of this feature, switched-line phase shifters are also called switchable time delay networks. The time delay τ_d is given by

$$\tau_d = \frac{l_2 - l_1}{v_p} \tag{12.34}$$

One of the common problems in the design of the switched-line phase shifters is caused by off-path resonances (Fig. 12.20). This may be illustrated by a specific example. Consider the design of a 45° phase bit for which l_1 and l_2 are chosen to be 160° and 205°, respectively, at the center frequency of 1.5 GHz. Further, the *pin* diodes used for SPDT switches can be represented by an $R_f = 1\ \Omega$ in forward bias and a $C_j = 0.2$ pF in the reverse bias. The equivalent

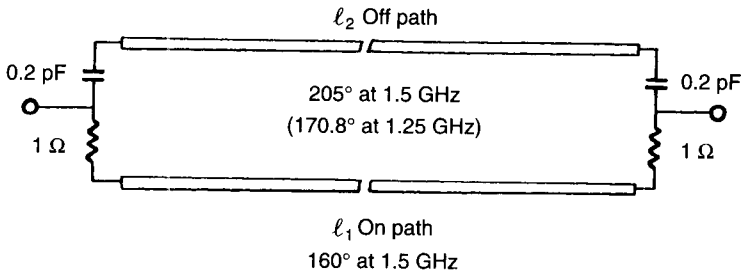


Figure 12.20 Equivalent circuit for switched-line phase shifter example.

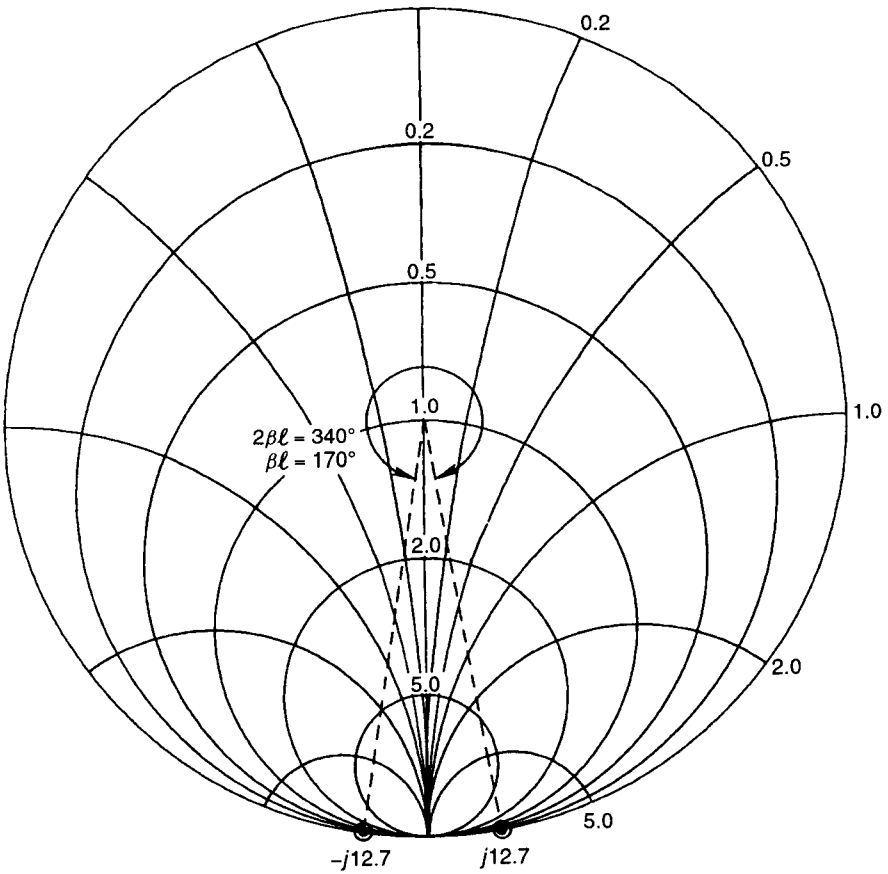


Figure 12.21 Smith chart illustration of insertion loss resonance in switched-line phase shifters.

circuit, for l_1 in the on path, is shown in Fig. 12.20. The design works quite well at the center frequency, but if the frequency is varied to 1.25 GHz, the length l_2 becomes 170.8° and, when associated with reactances of 0.2-pF capacitors at each end, shows a resonance. This resonance phenomenon is illustrated in Fig. 12.21. The 0.2-pF capacitors have a reactance of about $-j636\ \Omega$ each and add up to an effective line length of about 4.5° at each end. This causes a resonance that is exhibited by a peak in the insertion loss curve as well as by a rapid phase shift variation with frequency around 1.25 GHz. Frequency variation of phase shift and insertion loss are shown in Fig. 12.22. When the signal is routed through the path l_2, the insertion loss resonance of the length l_1 is observed near 1.58 GHz.

The insertion loss resonances may be avoided by a suitable choice of lengths l_1 and l_2. For example, in the present case, l_1, l_2 may be 50° and 95°, respectively. Another method of avoiding these resonances is by terminating the

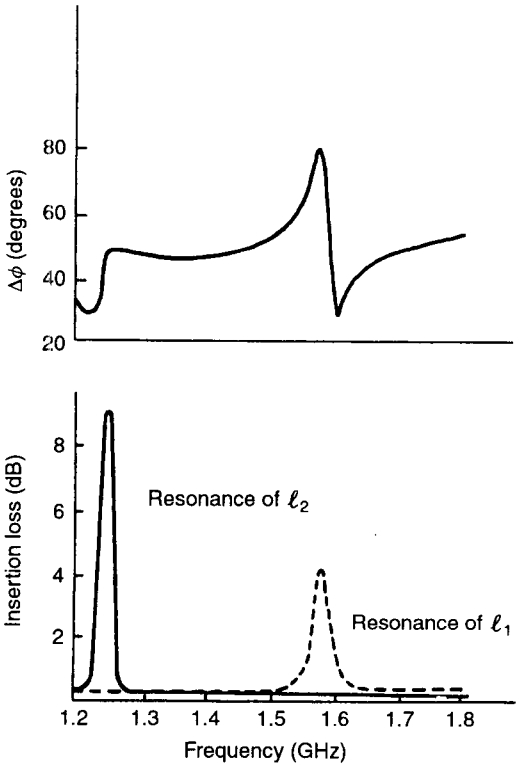


Figure 12.22 Insertion loss resonances in switched-line phase shifter circuit shown in Figs. 12.19 and 12.20. (After White [1]. Reprinted with permission of Van Nostrand.)

off path by matched loads. This does, however, require additional switching devices.

12.3.3 Loaded-Line Phase Shifters

A very common design for 45° and 22.5° phase bits is known as the loaded-line phase shifter. The mechanism of phase shift in this circuit is based on the loading of a uniform transmission line by a small reactance as shown in Fig. 12.23. It can be shown that the transmitted wave undergoes a phase shift $\Delta\phi$ that depends upon the normalized susceptance $b = B/Y$. The reflection caused by b is given by

$$\Gamma = \frac{1-(1+jb)}{1+(1+jb)} = \frac{-jb}{2+jb} \tag{12.35}$$

The voltage associated with the transmitted wave is $V_T = V_I + V_R$, where V_I and V_R are voltages of the incoming wave and the reflected wave, respectively.

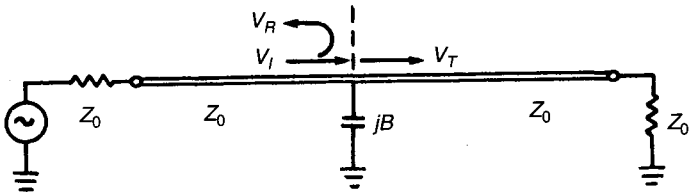


Figure 12.23 Circuit for illustrating basic mechanism of loaded-line phase shifter.

The transmission coefficient T can therefore be written as

$$T = \frac{V_T}{V_I} = \frac{V_I + V_R}{V_I} = 1 + \Gamma = \frac{2}{2 + jb} \tag{12.36}$$

$$V_T = TV_I = V_I \frac{2}{2 + jb} = V_I \left(\frac{4}{4 + b^2}\right)^{1/2} \exp\left[-j \tan^{-1}\left(\tfrac{1}{2}b\right)\right] \tag{12.37}$$

The phase difference introduced (phase of V_I – phase of V_T) may therefore be written as

$$\Delta\phi = \tan^{-1}\left(\tfrac{1}{2}b\right) \tag{12.38}$$

If the normalized susceptance $b = 0.2$ (i.e., capacitive), $\Delta\phi = \tan^{-1}(0.1) = 0.1$ rad $= 5.7°$. If, on the other hand, we load the line with $b = -0.2$ (a shunt inductor), the phase delay $\Delta\phi$ is negative, that is, $\Delta\phi = \tan^{-1}(-0.1) \simeq -5.7°$. That is, the transmitted wave advances in phase as compared with the transmitted wave for $b = 0$. Another important performance parameter here is the insertion loss given by the magnitude of the ratio of V_T to V_I, which may be written as

$$\frac{V_T}{V_I} = \left(\frac{4}{4 + b^2}\right)^{1/2} \tag{12.39}$$

As $|b|$ is equal in two states (only its sign is changed), insertion loss has the same value in two states. For $b = 0.2$, $V_T/V_I = 0.995$, which corresponds to a loss of 0.04 dB. Although, in the present case, the value of insertion loss is not significant, the presence of this loss is an unfavorable feature of the circuit shown in Fig. 12.23.

This drawback of reflection from the susceptance producing the phase shift can be overcome by using two identical susceptances separated by a quarter-wave line length. Such an arrangement is shown in Fig. 12.24. If we consider partial reflections caused by the two susceptances, these reflected waves are almost equal in magnitude and out of phase (180° phase difference) when looking into the input terminals. Thus these undesirable reflections cancel each

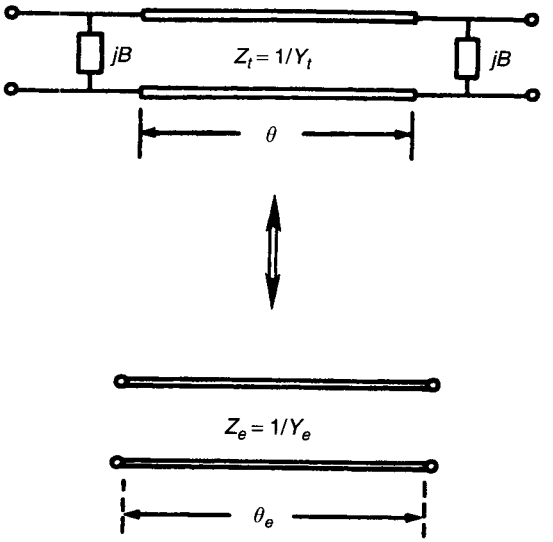


Figure 12.24 Circuit configuration for loaded-line phase shifter and equivalent representation.

other. A quantitative assessment of the situation may be obtained by writing a uniform transmission line equivalent circuit as shown in Fig. 12.24. Equivalent Y_e and θ_e may be obtained by comparing $ABCD$ matrices of the two networks. An $ABCD$ matrix of the transmission line of length θ, shunt-loaded by susceptances at either end, may be obtained by multiplying $ABCD$ matrices of the shunt susceptance, of the line section, and of the second susceptance at the other end. This yields

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ jB & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & jZ_t \sin\theta \\ jY_t \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jB & 1 \end{bmatrix}$$

$$= \begin{bmatrix} (\cos\theta - BZ_t \sin\theta), & j(Z_t \sin\theta) \\ j(2B\cos\theta + Y_t \sin\theta - B^2 Z_t \sin\theta), & (\cos\theta - BZ_t \sin\theta) \end{bmatrix} \quad (12.40)$$

Note that B on right side of this equation denotes susceptance of the switching devices (not to be confused with the B term of the $ABCD$ matrix). $ABCD$ of the equivalent transmission line may be written as

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cos\theta_e & jZ_e \sin\theta_e \\ jY_e \sin\theta_e & \cos\theta_e \end{bmatrix} \quad (12.41)$$

We can find the length of the equivalent uniform line by comparing the A terms in (12.40) and (12.41), as

$$\cos\theta_e = \cos\theta - BZ_t \sin\theta \quad (12.42)$$

The equivalent admittance Y_e is obtained by comparing the ratios C/B, as

$$Y_e = Y_t[1 - (BZ_t)^2 + 2BZ_t \cot \theta]^{1/2} \tag{12.43}$$

When $\theta = 90°$, relations (12.42) and (12.43) may be written as

$$\cos \theta_e = -BZ_t \tag{12.44}$$

or

$$\theta_e = \tfrac{1}{2}\pi + BZ_t + \tfrac{1}{6}(BZ_t)^3 \tag{12.45}$$

$$Y_e = Y_t[1 - (BZ_t)^2]^{1/2} \tag{12.46}$$

Thus we note that $\theta_e > \theta$ for capacitive susceptance and $\theta_e < \theta$ for inductive susceptances. Also $Y_e < Y_t$ and its magnitude is independent of the sign of B. Thus, if only the sign of B is changed (and its magnitude kept unchanged), the circuit can remain matched in both states of the phase shifter. For example, if $b_1 = 0.2$ and $b_2 = -0.2$, we have $\theta_{e1} = 101.54°$ and $\theta_{e2} = 78.46°$, resulting in a differential phase shift $\Delta\phi = \theta_{e2} - \theta_{e1} = 23.08° \simeq 0.4$ rad. It may be noted that if $(b_1 - b_2)$ is small, $\Delta\phi$ is approximately $(b_1 - b_2)$ radians. Also in the present case, $Y_e/Y_t = 0.98$. That is, if the desired impedance level is 50 Ω, Z_t should be $0.98 \times 50 = 49$ Ω. If Z_t is retained as 50 Ω, it will result in an input VSWR of 1.02.

Various Configurations for Loaded-Line Phase Shifters [12]. Various designs for loaded-line phase shifters differ to the extent that the susceptances B_1 and B_2 for two states of the phase shifter are realized by different circuit configurations. We will discuss two types of circuits used for this purpose. The basic design equations are obtained by combining (12.42) and (12.43) with (12.38) and may be written as

$$Y_T = Y_0 \sec(\tfrac{1}{2}\Delta\phi) \sin \theta \tag{12.47}$$

$$B_{1,2} = Y_0[\sec(\tfrac{1}{2}\Delta\phi) \cos \theta \pm \tan(\tfrac{1}{2}\Delta\phi)] \tag{12.48}$$

1. *Main-Line Mounted Circuit.* In this case, the switching devices are mounted directly across the main line. The idea is to use the high-impedance state capacitance C_j and the low-impedance state inductance L_s directly for B_1 and B_2, respectively. However, it becomes necessary to add an external inductance L_e in series with the device as shown in Fig. 12.25. Susceptances B_1 and B_2 are now given by

$$B_1 = \frac{\omega C_j}{1 - \omega^2 L C_j} \tag{12.49a}$$

$$B_2 = \frac{-1}{\omega L} \tag{12.49b}$$

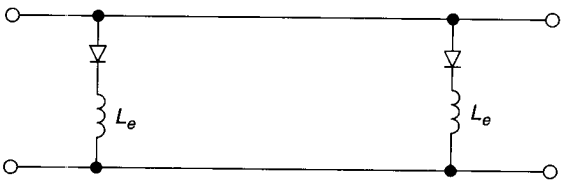


Figure 12.25 Main-line mounted-type loaded-line phase shifter circuit.

where $L = L_s + L_e$. Using (12.48) and (12.49), we can derive the following relationship between Y_0 and θ for $\Delta\phi$ and C_j:

$$Y_0 = \omega C_j \frac{\sin \Delta\phi}{\sin^2(\Delta\phi/2) - \cos^2 \theta} \tag{12.50}$$

Usually one would like to select θ so as to obtain the maximum bandwidth from the phase shifter circuit, which occurs at $\theta = 90°$ [12]. For this case

$$Y_0 = 2\omega C_j \cot(\tfrac{1}{2}\Delta\phi) \tag{12.51}$$

This puts restrictions on the input–output line impedance unless C_j can be chosen arbitrarily, which becomes difficult because of the limited availability of switching device capacitances. Because of these difficulties, this design configuration has not become very popular.

2. *Stub-Mounted-Type Circuit.* In the circuit arrangement shown in Fig. 12.26, switching devices are mounted at the ends of two shunt-connected stubs separated by a line length θ. This allows greater design flexibility. Stub impedance Z_b and length θ_b are determined from device reactances X_f and X_r (in

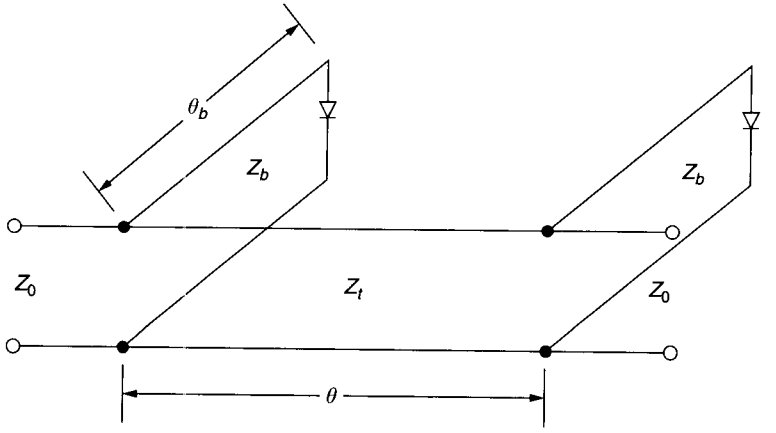


Figure 12.26 A stub-mounted-type loaded-line phase shifter circuit.

low- and high-impedance states, respectively) as follows:

$$Z_b = \left(\frac{X_f - X_r - X_f X_r(B_1 - B_2)}{B_1 - B_2 - B_1 B_2(X_f - X_r)}\right)^{1/2} \tag{12.52}$$

$$\tan \theta_b = \frac{Z_b(1 - X_f B_1)}{X_f - B_1 Z_b^2} \tag{12.53}$$

where Z_t and θ may be selected for wide bandwidth. It has been reported [12] that the best bandwidths are obtained for $\theta \sim 90°$. These design methods make use of device reactances only and assume the device resistances to be negligible.

12.3.4 Reflection-Type Phase Shifters

Another important class of phase shifters is constituted by reflection-type phase shifters. The basic concept is illustrated in Fig. 12.27*a*. A subnetwork with

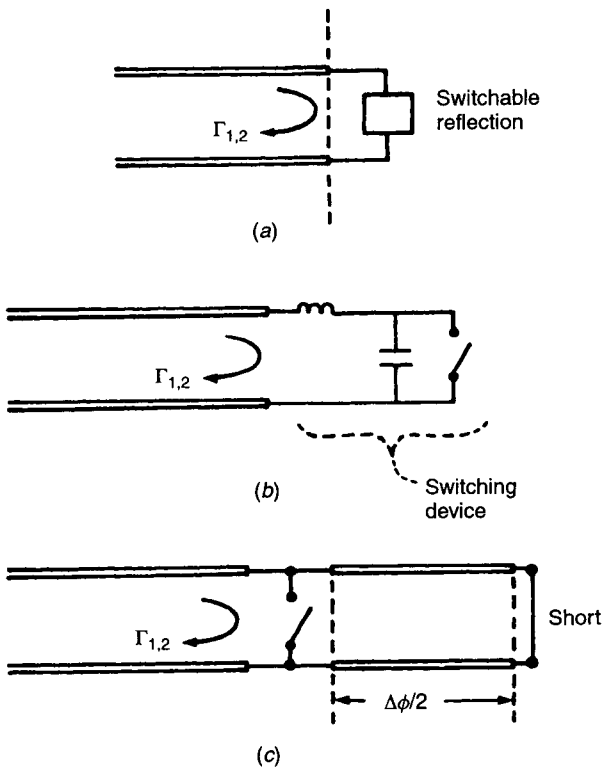


Figure 12.27 (*a*) Basic concept of reflection-type phase shifters. (*b*) Switchable-reactance-type reflection phase shifter. (*c*) Switchable-length-type reflection phase shifter.

switchable reflection coefficient terminates a uniform transmission line. When the reflection coefficient is switched from $\Gamma_1 = |\Gamma_1| \angle \phi_1$ to $\Gamma_2 = |\Gamma_2| \angle \phi_2$, the reflected signal undergoes a differential phase shift $\Delta\phi = \phi_1 - \phi_2$. In any state, the ratio of the reflected power to the incident power is given by $|\Gamma|^2$. Ideally $|\Gamma|$ should be unity so that there is no loss associated with the phase shifting operation.

Subnetworks providing the switchable reflection coefficients may be of two different types. In the first group, the reactance terminating the line is changed (say from inductive to capacitive) as shown in Fig. 12.27*b*. Phase shifters using these subnetworks are called switchable reactance types and constitute the more commonly used variety of reflection phase shifters. In the second group of these phase shifter circuits, an additional line length is added at the reflection plane by using an SPST switch. The concept is similar to that used in switched-line phase shifters. Here, as shown in Fig. 12.27*c*, the signal travels a longer path when the switch is open. The differential phase shift, in this case, is twice the electrical length of the shorted line switched in by the SPST switch.

Transformation of a Reflection Phase Bit into a Two-Port Network. Most of the systems using phase shifters require the phase shifting circuits to be two-port transmission networks. A reflection phase bit can be converted into a two-port network either by using a circulator or by a 90° hybrid. These two arrangements are shown in Fig. 12.28*a* and *b*, respectively. It may be noted that hybrid-coupled arrangements require two identical phase bits (and hence twice the number of devices). However, this circuit is preferred to the circulator-coupled phase shifters because of the following two features: (1) 90° hybrids are a lot more integrable in MICs and MMICs than the circulators, and (2) the use of two switching devices in hybrid-coupled phase shifters increases the power-handling capability by a factor of 2, which is needed in several systems.

Design of Reflection-Type Phase Shifters. A hybrid-coupled reflection-type phase shifter consists of basically three elements: a hybrid, two transforming networks, and two switching devices. Design of the hybrid can be considered separately from the rest of the phase shifter circuit and will not be discussed here. One may refer to Chapter 5 or [13] for designs of hybrids in strip-line configuration. The remaining circuit consists of two identical networks connected to the two coupled ports of the hybrid. These networks may be called *phase shift networks*. An important part of the phase shift network is the switching device itself. The remaining network may be called the *transforming network*, as it transforms the device impedances in two states to appropriate values at the hybrid ports, so that the desired phase shift characteristics are obtained.

Various designs for hybrid-coupled reflection phase shifters use different configurations for the two transforming networks. Let us look in more detail at the specific functions these transforming networks are intended to perform.

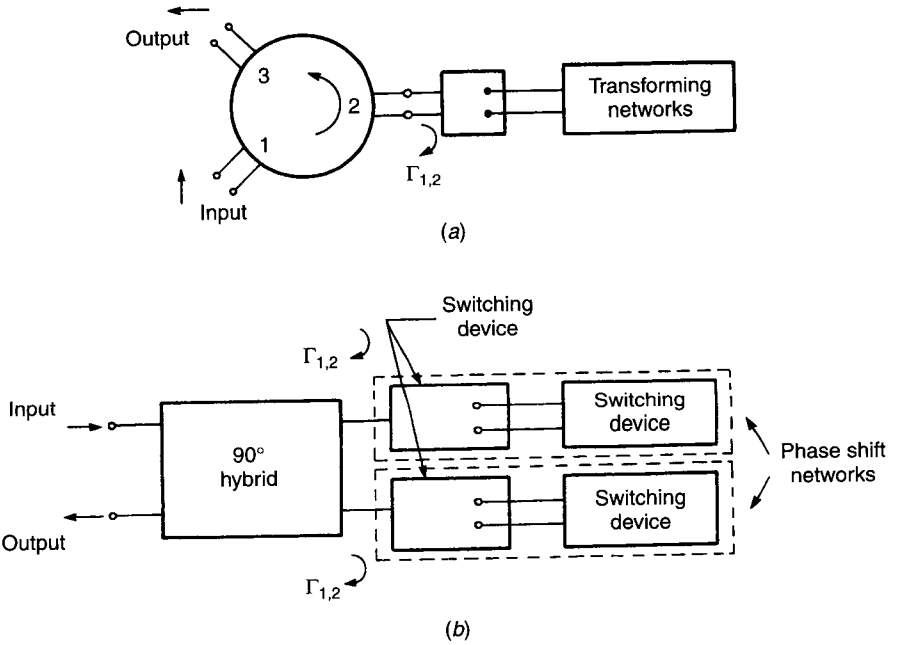


Figure 12.28 (*a*) Circulator-coupled reflection-type phase shifter. (*b*) Hybrid-coupled reflection-type phase shifter.

Design of Transforming Networks. In order to obtain a phase shift $\Delta\phi$, the reflection coefficient values (as seen at two coupled ports of the hybrid) should differ in phase by $\Delta\phi$. Ideally their magnitudes should be unity (no loss). Thus for a 180° bit, values of the reflection coefficients in two states would lie on the outer circle ($|\Gamma| = 1$) of the Smith chart and their locations would be diametrically opposite.

Although an infinite number of such combinations are possible, it can be shown that for a reflection phase shifter using a two-branch branch-line hybrid, the maximum phase shifter bandwidth is obtained [14] when the two values are located at points corresponding to normalized reactances equal to $+j$ and $-j$ on the Smith chart (i.e., when located symmetrically with respect to the $X = 0$ axis). Similarly for other values of $\Delta\phi$, Γ_1 and Γ_2 in two states should also be located symmetrically with respect to the $X = 0$ axis. These locations for $\Delta\phi = 180°$, 90°, and 45° are shown in Fig. 12.29. Note that for all values, other than $\Delta\phi = 180°$, there are two alternative sets of values for Γ_1 and Γ_2. Strictly speaking, for the positions shown in the right half of the Smith chart, the angle of Γ_f minus the angle of Γ_r would be 360° minus the corresponding values for points shown in the left half of the chart.

Based on the preceding discussion, we can divide the function of the transforming network into two parts.

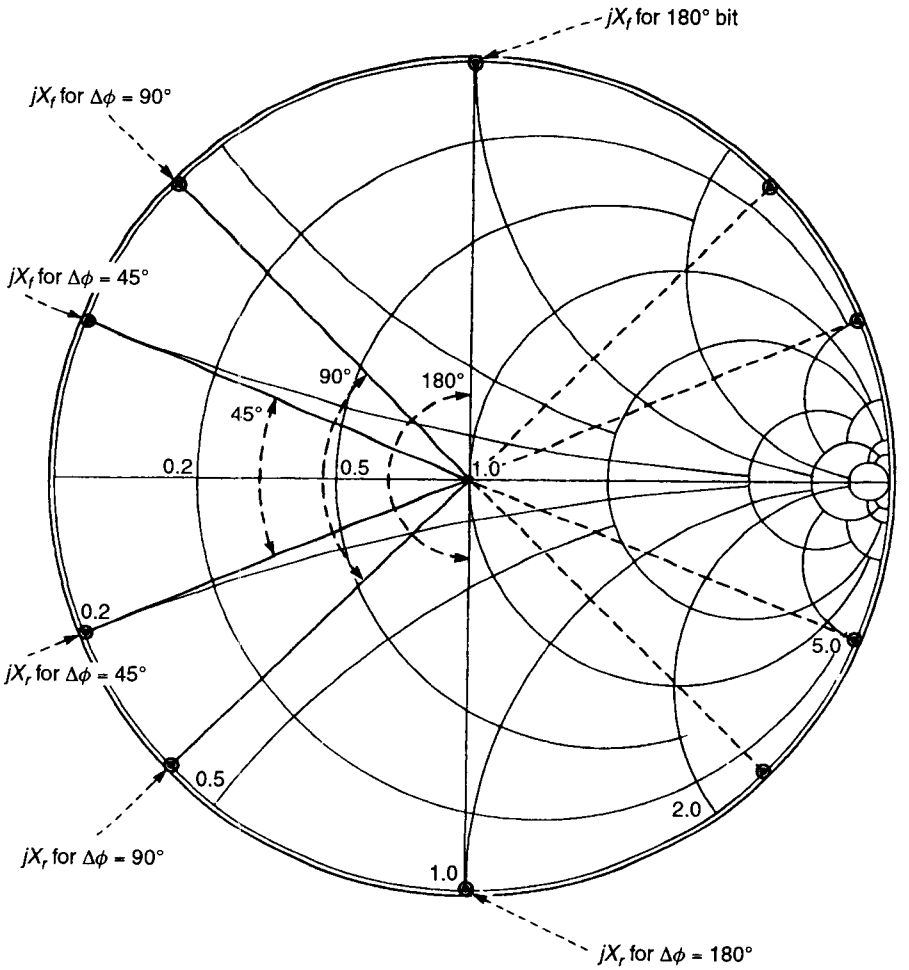


Figure 12.29 Desired location of reflection coefficients for two states of reflection-type phase shifter using two branch-line hybrids.

1. It arranges Γ_f and Γ_r so that the desired phase shift $\Delta\phi$ is obtained.
2. It locates Γ_f and Γ_r symmetrically about the $X = 0$ axis on the Smith chart so that the condition for maximum bandwidth is satisfied.

It may be noted that the second part of the function just mentioned is implemented by connecting a suitable line length between the transforming networks designed for part 1 and the ports of the hybrid. This line length will rotate Γ_f and Γ_r points around the Smith chart without disturbing the phase relationship between them. For a given set of device impedances in two bias

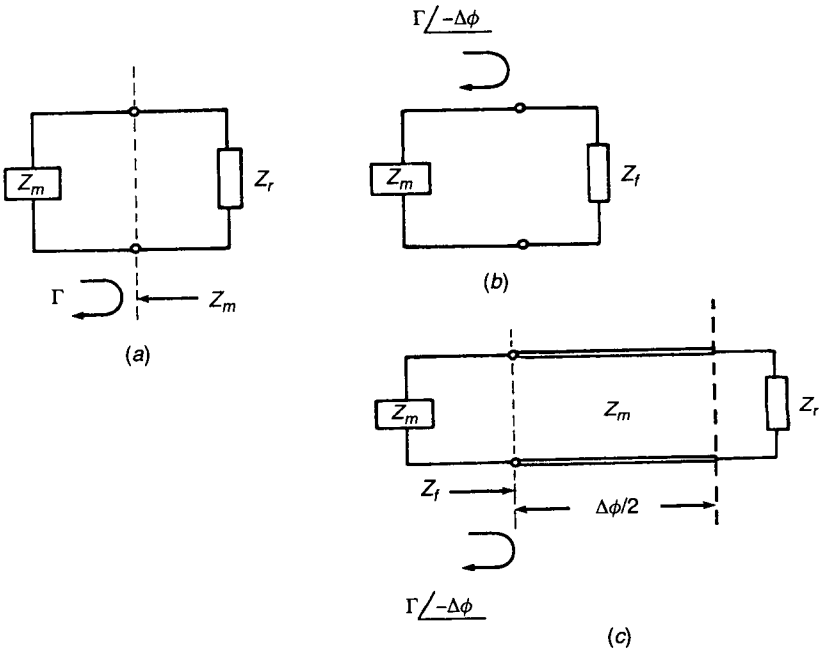


Figure 12.30 Equivalent networks for calculating Z_m (Thévenin impedance of transforming network).

states (Z_f and Z_r), a desired phase shift $\Delta\phi$ can be obtained by arranging for a suitable impedance (say Z_m) to be seen when looking from the device terminals toward the rest of the network. Values of Z_m may be evaluated by considering the representation shown in Fig. 12.30.

Let a network with the equivalent Thévenin impedance Z_m terminated by an impedance Z_r (of the high-impedance state of the switching device) produce a reflection coefficient Γ, as shown in Fig. 12.30*a*. When the same network is terminated with an impedance Z_f (of the low-impedance state of the device), we want a reflection coefficient of $\Gamma/-\Delta\phi$ (Fig. 12.30*b*). A reflection coefficient of $\Gamma/-\Delta\phi$ will also be produced when a transmission-line section of length $\Delta\phi/2$ and impedance Z_m is connected between the network and Z_r, as shown in Fig. 12.30*c*. Thus networks shown in Fig. 12.30*b* and *c* should be equivalent. For this equivalence to hold good, the following relationship should be satisfied:

$$Z_f = Z_m \frac{Z_r + jZ_m \tan(\Delta\phi/2)}{Z_m + jZ_r \tan(\Delta\phi/2)} \tag{12.54}$$

For low-loss switching devices, Z_f and Z_r may be approximated by jX_f and jX_r, respectively. In this case, (12.54) may be rewritten to express Z_m explicitly

in terms of X_f, X_r, and $\Delta\phi$. That is, given device reactances, the desired phase shift $\Delta\phi$ may be obtained by choosing Z_m given by

$$Z_m = \frac{X_f - X_r}{2\tan(\Delta\phi/2)} \pm \sqrt{\left(\frac{X_f - X_r}{2\tan(\Delta\phi/2)}\right)^2 - X_f X_r} \tag{12.55}$$

For a 180° bit, Z_m becomes

$$Z_m = \sqrt{-X_f X_r} \tag{12.56}$$

and for a 90° bit, we should have

$$Z_m = \frac{X_f - X_r}{2} \pm \sqrt{\left(\frac{X_f - X_r}{2}\right)^2 - X_f X_r} \tag{12.57}$$

It should be remembered that the reactance X_r has a negative value (it is capacitive). This design approach holds good even when the switching device is lossy. However, for lossy devices Eqs. (12.55)–(12.57) get modified and Z_m may be complex.

Phase Shifter Using a $\frac{1}{4}\lambda$ Transforming Network. The basic configuration is shown in Fig. 12.31. The design is simple. A quarter-wave line of impedance Z_t is used to transform the hybrid impedance Z_0 to a value Z_m, which the device should look into in order to provide the desired phase shift $\Delta\phi$. So

$$Z_t = \sqrt{Z_0 Z_m} \tag{12.58}$$

where Z_m is given by (12.55), (12.56), or (12.57). The line length θ_t is selected so that the impedances Z_b (in two bias states) are located symmetrically with respect to the $X = 0$ axis on the Smith chart. For a 180° phase bit, Z_b in the low-impedance state should be $-jZ_0$ (since a 90° line length has already been added). This yields

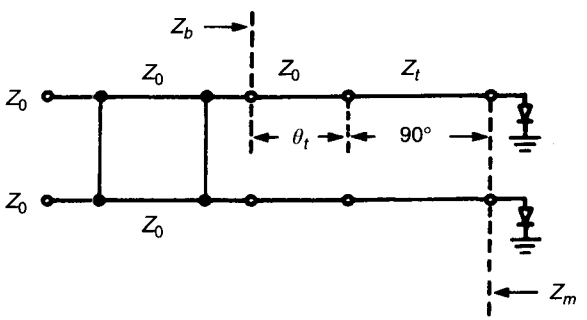


Figure 12.31 Reflection phase shifter using $\frac{1}{4}\lambda$ transforming network.

$$-jZ_0 = Z_b = jZ_0 \frac{-Z_t^2/X_f + Z_0 \tan\theta_t}{Z_0 + (Z_t^2/X_f)\tan\theta_t} \tag{12.59}$$

which gives

$$\theta_t|_{\Delta\phi=180^\circ} = \tan^{-1}\left(\frac{Z_t^2 - Z_0X_f}{Z_t^2 + Z_0X_f}\right) \tag{12.60}$$

Similarly, for a 90° bit, we should have

$$Z_b = -j2.4142Z_0$$

which yields

$$\theta_t|_{\Delta\phi=90^\circ} = \tan^{-1}\left(\frac{Z_t^2 - 2.4142X_fZ_0}{Z_0X_f + 2.4142Z_t^2}\right) \tag{12.61}$$

Phase Shifter Using Impedance-Transforming Hybrid. Rather than using a $\frac{1}{4}\lambda$ transformer to change the impedance level from Z_0 to Z_m, this impedance transformation can be integrated within the hybrid design itself. Such a hybrid is called an impedance-transforming hybrid and has been described in [15].

12.3.5 Switched-Network Phase Shifters

This class of phase shifters may be considered to be a conceptual evolution from switched-line phase shifters. A basic block diagram of this type of phase shifter is shown in Fig. 12.32. When the input signal, originally passing through network 1, is switched to pass through network 2, we get a differential phase shift $(\phi_2 - \phi_1)$. The switched-line phase shifter is a special case of a switched-network phase shifter with the two networks being sections of transmission

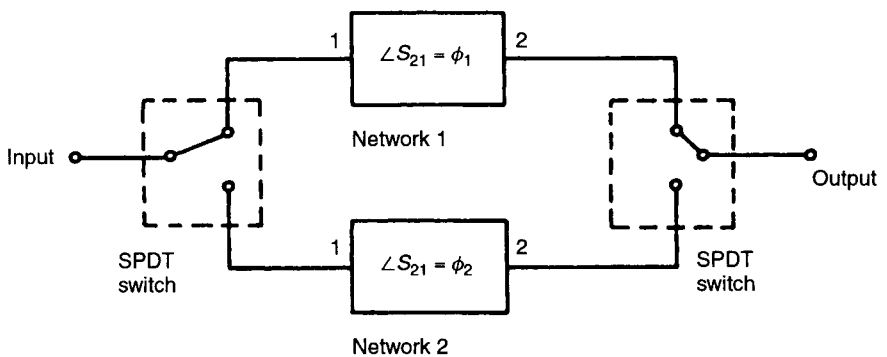


Figure 12.32 Basic block diagram of switched-network phase shifter.

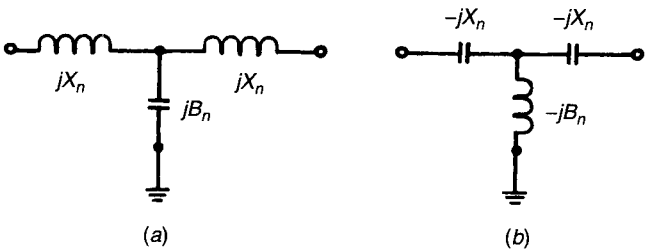


Figure 12.33 (*a*) Low-pass and (*b*) high-pass networks used in switched-network phase shifters.

lines of different lengths. The main advantage of generalizing a switched-line phase shifter into a switched-network configuration is that one can design the variations of ϕ_1 and ϕ_2 with frequency appropriately and obtain a wider bandwidth or a desired frequency response of the phase shifter.

The most commonly used networks in switched-network phase shifters are the low-pass and high-pass filter configurations shown in Fig. 12.33.

The normalized $ABCD$ matrix of the network in Fig. 12.33*a* may be written as

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}_n = \begin{bmatrix} A & B/Z_0 \\ CZ_0 & D \end{bmatrix} = \begin{bmatrix} 1 & jX_n \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jB_n & 1 \end{bmatrix} \begin{bmatrix} 1 & jX_n \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 - B_nX_n & j(2X_n - B_nX_n^2) \\ jB_n & 1 - B_nX_n \end{bmatrix} \tag{12.62}$$

where X_n and B_n are the reactance and susceptance shown in Fig. 12.33 normalized with respect to transmission-line impedance Z_0 and admittance Y_0, respectively. The transmission coefficient S_{21} is given, in terms of the normalized $ABCD$ matrix, by

$$S_{21} = \frac{2}{A + B + C + D} = \frac{2}{2(1 - B_nX_n) + j(B_n + 2X_n - B_nX_n^2)} \tag{12.63}$$

The transmission phase ϕ is given by

$$\phi = \tan^{-1}\left(-\frac{B_n + 2X_n - B_nX_n^2}{2(1 - B_nX_n)}\right) \tag{12.64}$$

When both B_n and X_n change signs (as shown in Fig. 12.33*b*), the phase ϕ retains the same magnitude but changes sign. Amplitude of S_{21} does not change. Thus the phase shift $\Delta\phi$ caused by switching between low-pass and high-pass networks is given by

$$\Delta\phi = 2\tan^{-1}\left(-\frac{B_n + 2X_n - B_nX_n^2}{2(1 - B_nX_n)}\right) \tag{12.65}$$

For the phase shifter to be matched we need

$$|S_{11}| = 0 \tag{12.66}$$

Since we are considering a lossless case,

$$|S_{11}| = \sqrt{1 - |S_{21}|^2} \tag{12.67}$$

This leads to the following relationship between B_n and X_n:

$$B_n = \frac{2X_n}{X_n^2 + 1} \tag{12.68}$$

Thus the phase shift $\Delta\phi$ can be expressed in terms of X_n alone as

$$\Delta\phi = 2\tan^{-1}\left(\frac{2X_n}{X_n^2 - 1}\right) \tag{12.69}$$

which yields X_n in terms of $\Delta\phi$ as

$$X_n = \tan\left(\tfrac{1}{4}\Delta\phi\right) \tag{12.70}$$

Substituting X_n in (12.68) yields

$$B_n = \sin\left(\tfrac{1}{2}\Delta\phi\right) \tag{12.71}$$

The π-section filter shown in Fig. 12.34, may also be used in place of the T configuration shown in Fig. 12.33. For this case

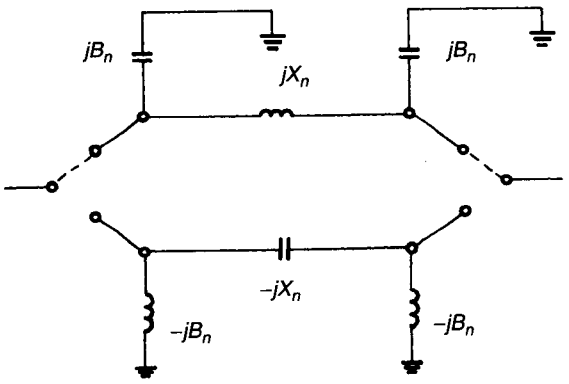


Figure 12.34 Switched-network phase shifter using π network.

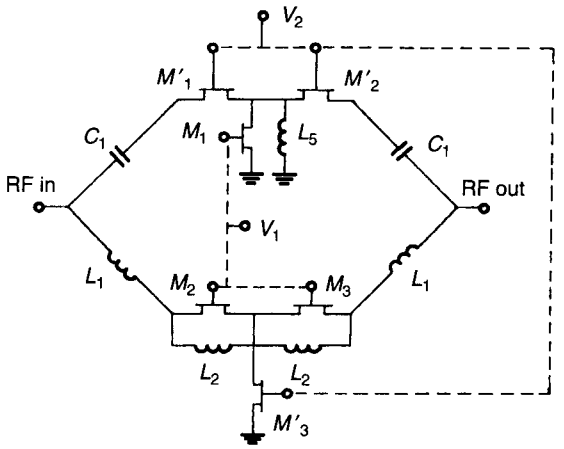


Figure 12.35 Switched-network phase shifter using six MESFETs. (After Ayasli et al. [16]. Reprinted with permission of IEEE.)

$$B_n = \tan\left(\tfrac{1}{4}\Delta\phi\right) \tag{12.72}$$

$$X_n = \sin\left(\tfrac{1}{2}\Delta\phi\right) \tag{12.73}$$

Bandwidth of these phase shifters may be discussed by considering variations of ϕ_1 and ϕ_2 with frequency. For a T-configuration low-pass filter, phase delay ϕ_1 increases with frequency; while for a high-pass filter, phase advance ϕ_2 decreases with frequency. As ϕ_1 is phase delay and ϕ_2 is phase advance, they have opposite signs and $\Delta\phi = \phi_1 - \phi_2 = |\phi_1| + |\phi_2|$. Thus the two variations tend to compensate for each other and $\Delta\phi$ stays relatively constant.

A different implementation of a switched-network phase shifter using six MESFETs has been reported [16]. The configuration is shown in Fig. 12.35. Six MESFETs are used in two T configurations. The MESFETs are used in the passive mode with control voltage on gate terminals used for switching the device from the high-impedance to the low-impedance state and vice versa. Series elements of one T and shunt element of the other T are controlled by a single gate control voltage V_1. Thus three MESFETs M_1, M_2, and M_3 are always in the same state (high impedance or low impedance) and the other three MESFETs M'_1, M'_2, and M'_3 controlled by V_2 are in the opposite state. External inductances and capacitances L_1 and C_1 are included to provide reactances required for low-pass and high-pass filters.

This six-MESFET design has been used for construction of a monolithic phase shifter in the 2–8 GHz frequency range [16].

12.3.6 Amplifier-Type Phase Shifters

Various phase shifter designs discussed in Sections 12.3.2–12.3.5 may be realized either by using *pin* diodes or MESFETs in the passive mode or RF MEMS

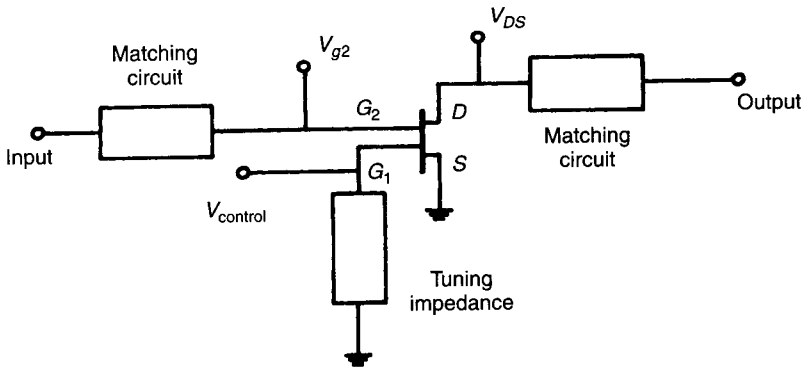


Figure 12.36 Circuit configuration for tuned-gate dual-gate MESFET phase shifter.

devices (Chapter 14). Use of MESFETs in the active amplifier mode makes several novel phase shifter designs possible. Most of the interesting active phase shifter designs have been realized using dual-gate MESFETs. Available active phase shifter designs may be grouped into three classes: (1) tuned gate dual-gate MESFET phase shifter, (2) active phase shifter using switchable SPDT amplifiers, and (3) active phase shifter using vector modulator circuits.

Active Phase Shifters Using Tuned-Gate Dual-Gate MESFET. This circuit was the first reported [17] active phase shifter circuit but has not been very popular since. In this design, the second gate (nearer to the drain terminal) of the dual-gate MESFET is used as the signal input gate, and the first gate (nearer to the source terminal) is used as the control gate. A schematic of this type of phase shifter is shown in Fig. 12.36. As in the case of the MESFET amplifier design, matching circuits are needed both at input and output diodes. A tuning reactance (mostly inductive) is connected between the first gate G_1 and the ground. The control voltage for controlling the phase shift between the output and input signals is applied to gate 1. Phase control is obtained because of interaction between changing device parameters (such as gate 1 to source capacitance C_{g1s}) and the externally connected tuning impedance (which could be a series inductance). A detailed analysis of the interaction is not available. Reported results point out that a phase shift of up to 100° can be obtained at 12 GHz. A continuous variation of phase shift (up to 70°) has been reported with a 3 dB gain in the 11.9–12.2 GHz frequency range [17]. Narrow bandwidth is one of the limitations of this type of phase shifter.

Active Phase Shifters Using Switchable SPDT Amplifiers. A block diagram of this type of active phase shifter is shown in Fig. 12.37. The input signal is switched between two identical amplifiers. At the output of one of the amplifiers, an additional line length is introduced to provide the required phase shift $\Delta\phi$. The two signal paths are combined in a power combiner circuit. A Wilkinson-type power combiner introduces a loss of 3 dB (as only one input

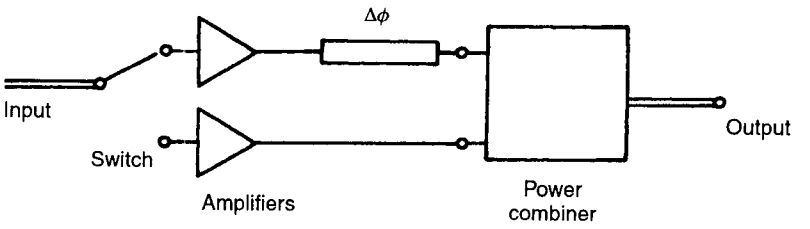


Figure 12.37 Block diagram of phase shifter using SPDT amplifiers.

signal is present at any time), but in view of the gain available in amplifiers, this 3 dB loss is not considered serious. A convenient feature of this design is that the phase shifting part of this circuit is independent of the rest of the design, which remains invariant for different sizes of phase bits. A detailed circuit example of such a phase shifter is available in [18].

Active Phase Shifter Using Vector Modulator Circuits. A vector modulator is a circuit that is capable of independently varying the amplitude and the phase of an input signal by desired amounts. The concept is shown schematically in Fig. 12.38*a*. In principle, ϕ could be anywhere between 0° and 360°, and A could be any reasonable factor less than or greater than 1. Two different

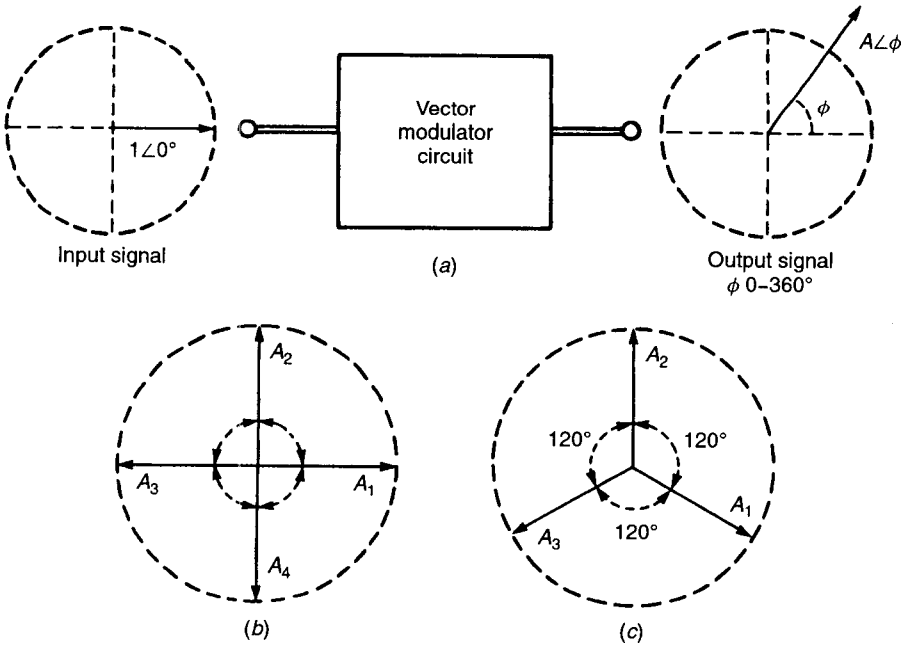


Figure 12.38 (*a*) Concept of a vector modulator circuit. (*b*) Four vectors spaced 90° apart, used for realization of vector modulator circuits. (*c*) Three vectors (120° apart) used for realization of vector modulator circuits.

schemes have been proposed for implementing a vector modulator circuit at microwave frequencies. One of these uses four vectors spaced 90° apart and pointing in four different directions as shown in Fig. 12.38*b*. Amplitudes of the four component vectors A_1, A_2, A_3, and A_4 are controlled independently by four different controlled-gain amplifiers. Of course, only two of these four components are nonzero at any time. Thus the vector sum could lie in any of the four quadrants and the amplitude could be adjusted by controlling the gains of two amplifiers active at that time. It may be noted that four component vectors are not needed and three vectors spaced 120°, as shown in Fig. 12.38*c*, could serve the same purpose. This provides an alternative implementation for vector modulator circuits, and its realization has been described in [19]. Clearly, a vector modulator has more versatility than that required for a digital phase shifter. In fact, it is this versatility that makes the circuit very attractive for monolithic phase shifter applications.

Phase Shifter Using Segmented-Gate MESFETs. We note that in phase shifter circuits based on the vector modulator concept, the phase shift depends upon the ratio of two gains. In order to design a digital phase shifter, we need a repeatable digital variation of the gain ratio. A monolithic device introduced [20] for this purpose is known as a segmented dual-gate MESFET. The total gate width of a dual-gate MESFET is divided into several sections. A signal–gate connection is common to all of these sections, but the control gate is broken into several segments, one corresponding to each of the sections. External connections are brought out from each one of the second gate segments, so that various sections of the MESFET may be switched ON or OFF selectively. A typical schematic diagram of such a device [20] is shown in Fig. 12.39. Here, the control gate (total width 750 μm) is divided into four segments of different widths. Segments *A*, *B*, *C*, and *D* have widths of 50, 100, 200, and 400 μm, respectively. Each one of these segments may be switched ON or OFF

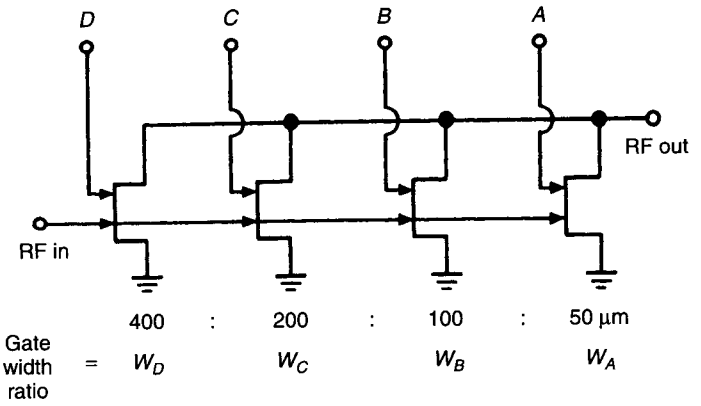


Figure 12.39 Circuit representation of a segmented dual-gate device. (After Hwang et al. [20]. Reprinted with permission of IEEE.)

independently. Since the gain of the MESFET is, to a reasonable approximation, linearly proportional to the gate width, the gain can be varied by switching ON a particular segment or a combination of segments. If we consider the gain provided by the 50-μm segment as one unit, the gain of the device shown in Fig. 12.39 can be varied in steps of these units from a single unit (when only the 50-μm segment is on) through to 15 units (when all four segments are turned on). This corresponds to a gain variation over a range of 23.52 dB. In the fabrication process of GaAs MESFET devices, the gate width is a parameter that is least sensitive to processing variables. So even if the gain of individual segments may vary from device to device, the ratios of gains exhibit an excellent repeatability. Also, these ratios are relatively independent of temperature variations.

Variable-gain amplifiers may be used for the design of phase shifters. A schematic block diagram of such a phase shifter is shown in Fig. 12.40. The input signal is divided into two equal parts by using a 90° hybrid. Thus for an input of $1\angle 0°$, the two signals at the output of block A are $0.707\angle 0°$ and $0.707\angle -90°$, respectively. These two signals are fed to variable-gain amplifiers. If the phase shift required from the phase shifter is ϕ, the gain of the amplifier in the upper signal path is set proportional to $\cos\phi$, while that of the amplifier in the lower signal path is adjusted proportionally to $\sin\phi$. The two signals coming out of these two amplifiers may be represented by $0.707K\cos\phi\angle -\theta°$ and $0.707K\sin\phi\angle -\theta - 90°$, where K is the proportionality factor for the amplifier gains and θ is the phase common to the two channels introduced by the amplifiers and the connecting lines. The two signals are added in an in-phase power combiner circuit. The output of the combiner is $0.5K[\cos(\theta+\phi) - j\sin(\theta+\phi)]$, that is, a phase shift of $\theta+\phi$ compared to the signal at the input. When the amplifier in the lower path is switched OFF (making $\phi = 0$), the output signal lags the input by a phase angle θ. This can be considered to be the reference state. Then, the differential phase shift ϕ is obtained when gains are set proportional to $\cos\phi$ and $\sin\phi$.

A digital phase shifter based on the concept shown in Fig. 12.40 can be obtained by using segmented dual-gate devices for $\cos\phi$ and $\sin\phi$ amplifiers. One of the possible arrangements is to use a device with three segments with widths in the ratio of 8:4:1. A fairly good approximation to $\cos\phi$ and $\sin\phi$ can

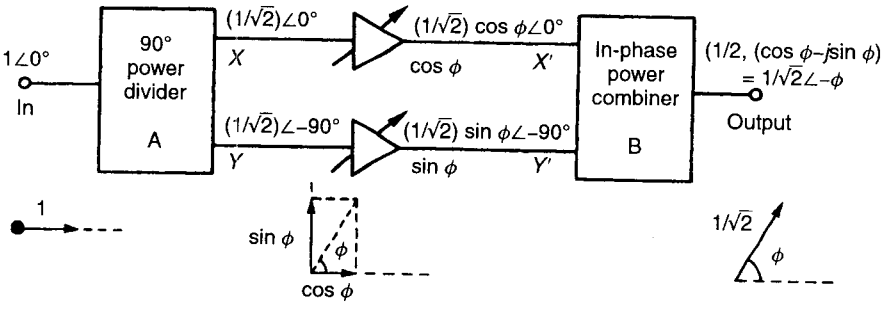


Figure 12.40 Phase shifter using two variable-gain amplifiers.

Table 12.3 The $\frac{1}{13}$ Approximation for Generating Sine and Cosine Functions with Segmented Gate Device

θ degrees	$\sin\theta$	1/13 Approximation	W_8	W_4	W_1
0	0	0/13	0	0	0
22.5	0.383	5/13 (0.384)	0	1	1
45.0	0.707	9/13 (0.692)	1	0	1
67.5	0.924	12/13 (0.923)	1	1	0
90.0	1.000	13/13	1	1	1

Note: 0 = off, 1 = on.

Source: After Hwang et al. [20]. Reprinted with permission of IEEE.

now be obtained for $\phi = 0$, 22.5°, 45°, 67.5°, and 90°. Table 12.3 shows this so-called $\frac{1}{13}$ approximation obtained by switching the three segments ON or OFF. In this table, the 0's in the last three columns show that corresponding segments are OFF, and the 1's indicate the corresponding segments are switched ON.

For differential phase shifts greater than 90°, $\cos\phi$ becomes negative. This requires a phase inversion in the upper signal path (Fig. 12.40). In this case, a possible solution is to use a 180° hybrid combiner in place of the in-phase power combiner shown in the figure. An alternative solution for obtaining ϕ greater than 90° is to cascade two stages with individual phase shifts less than 90°.

12.4 DESIGN OF LIMITER CIRCUITS

Apart from the switches and phase shifters discussed in previous sections, a *limiter* is an important control component used at microwave frequencies. An ideal limiter has no attenuation when low power is incident upon it but has an attenuation that increases with increasing power (above a threshold level) to maintain the output power constant. Input–output characteristics for ideal limiters (solid line) and practical limiters (dotted lines) are shown in Fig. 12.41. The most common application of microwave limiters is to prevent transmitter power in a radar from reaching the receiver directly and burning its sensitive input stage. Limiters are also used to protect the receivers from other nearby radar transmitters. Other applications of limiters are for reducing the amplitude modulation of swept-frequency oscillators and for reducing amplitude modulation in phase-detection systems.

12.4.1 Various Phenomena Used for Limiting

Limiters can be compared with SPST switches in the ON state when the incident power is below the limiting threshold value and to switches in the OFF

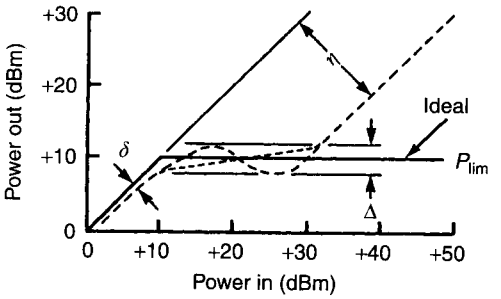


Figure 12.41 Characteristics of ideal and practical limiters. (After Garver [9]. Reprinted with permission of Artech House.)

state when the input exceeds the threshold. Consequently, practical limiter circuits are characterized by a finite insertion loss δ and a finite isolation η as shown in Fig. 12.41. Sometimes, another parameter, vacillation Δ about the limit power level P_{lim}, is also specified.

There are three phenomena, exhibited by microwave semiconductor diodes, that may be used for limiting. Let us discuss these very briefly.

1. *Rectification.* This is the most commonly used limiting technique at lower frequencies. A basic circuit (*a*) using two rectifying diodes, the *I–V* relationship (*b*), and the clipped sine wave (*c*) (obtained when the input is above the limiting threshold) are shown in Fig. 12.42. Point contact and Schottky barrier

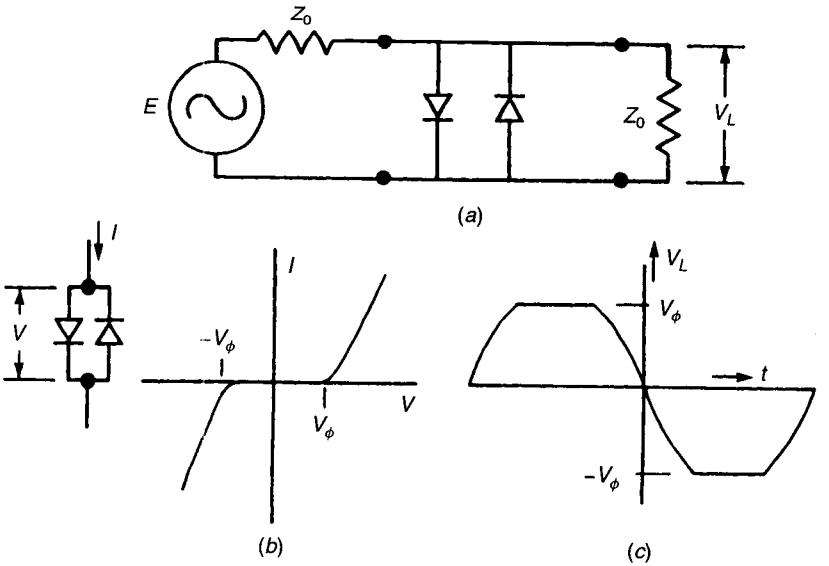


Figure 12.42 (*a*) Basic limiter circuit using two rectifying diodes. (*b*) The resulting *I–V* characteristic. (*c*) Output voltage waveform.

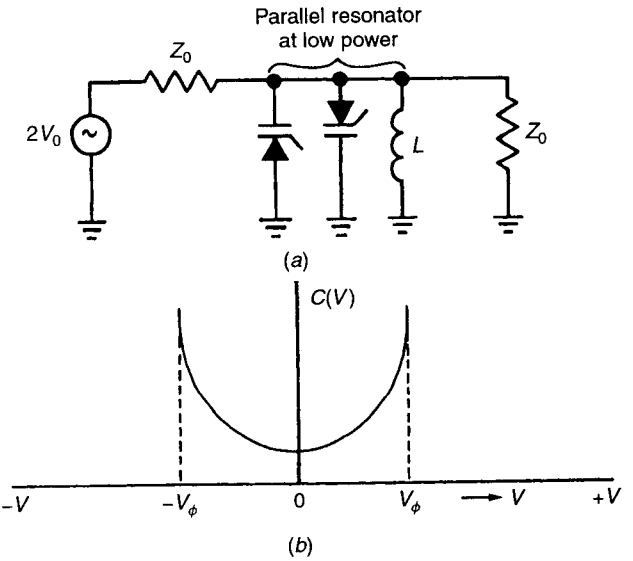


Figure 12.43 (*a*) Two varactor diodes set parallel with opposite polarity and (*b*) the resulting C–V characteristic for $|V| < |V_\phi|$.

diodes (not *pin* diodes) can be used for rectification at microwave frequencies. The main problem in using this method is its very low power-handling capability. The rectifying diodes have a very thin depletion layer (in order to have a sufficiently rapid turn-on time at microwave frequencies) and have a small junction area (in order to keep the device capacitance small). The resulting small volume of the device cannot protect it from burning out when very high power microwave signals are present.

2. *Capacitance Variation with Voltage.* Varactor diodes have a junction capacitance that is voltage dependent and that responds rapidly enough to change the characteristics at microwave frequency. If two diodes are set parallel in opposite polarity (as shown in Fig. 12.43*a*), the resulting C–V characteristic is shown in Fig. 12.43*b*. When these diodes are used in a parallel resonant circuit, we obtain a reflection coefficient that varies as a function of incident power. Also, some rectification occurs at high powers and provides additional limiting. These limiters also suffer from low power-handling capability.

3. *RF Conductivity Modulation.* This phenomenon is exhibited by *pin* diodes when biased at zero voltage and subjected to high-power RF current signals. When a *pin* diode is forward biased, the free-carrier concentration in the *i* region is not perfectly uniform but, because of the limited lifetime of the carriers, has a shape shown in Fig. 12.44*a*. Now suppose that the DC bias is removed and replaced by a short-circuit path (DC return) as shown in Fig. 12.44*b*. It has been shown by Leenov [21] that, when a large microwave current has been established, the distribution of the resulting carrier concentration will

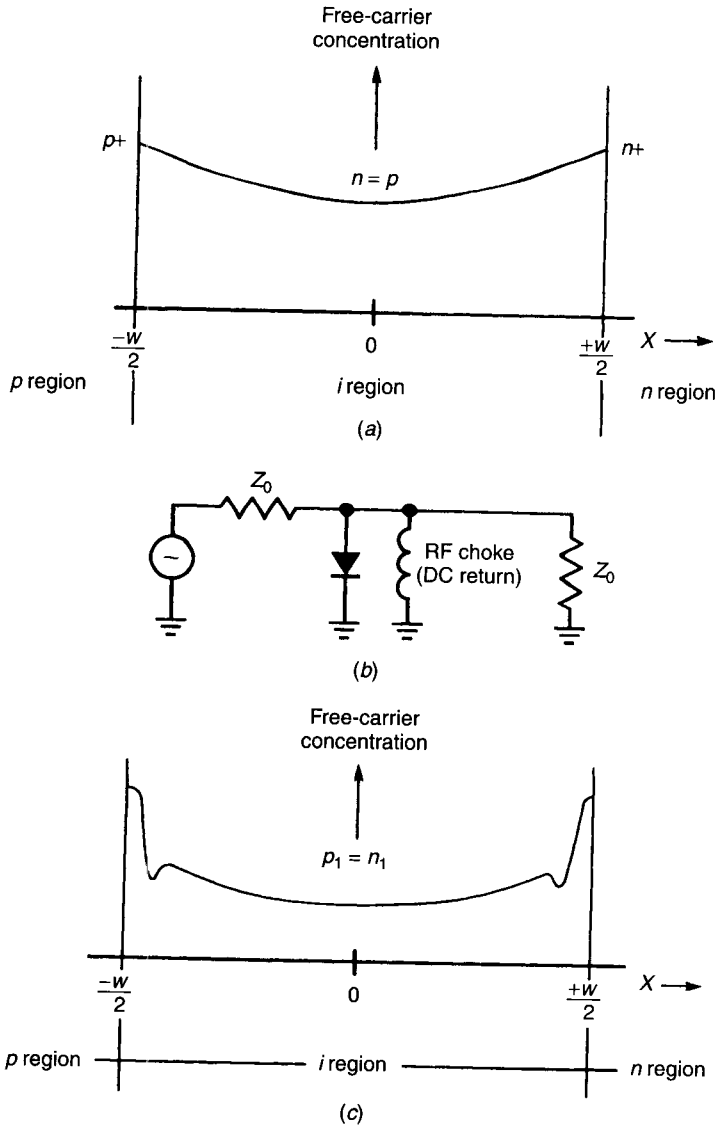


Figure 12.44 (*a*) The *i*-layer carrier concentration in a forward-biased *pin* diode. (*b*) Limiter circuit using a zero-biased *pin* diode with a DC return path. (*c*) Carrier concentration in *i*-layer *pin* diode in (*b*) for high-level RF excitation. (After White [1]. Reprinted with permission of Van Nostrand.)

be as shown in Fig. 12.44*c*. The large microwave current causes some partial carrier injection near the p^+ and n^+ boundaries during the forward-going half-cycles of the RF signal. Not all of this injected charge is withdrawn when the RF voltage reverses. The result is a trickling of electrons and holes into the *i* region, and these get distributed as shown in Fig. 12.44*c*. This steady state

distribution is attained after a few RF cycles and remains static during the remainder of the RF pulse, serving to conductivity modulate the i region. During the transient period in which the charge is built up in the i region, the *pin* limiter provides relatively little limiting to the high-power RF signal. The RF signal passes through the *pin* limiter with little attenuation during this spike leakage period, after which a relatively high attenuation (limiting) is achieved and the small fraction of the power that passes through is called the flat leakage. This *pin* diode limiting phenomenon is illustrated in Fig. 12.45. In order to distinguish clearly from rectification-type limiting, it may be noted that (for the limiter circuit shown in Fig. 12.45), the *pin* diode presents substantially the same conductivity to both forward- and reverse-going halves of the RF cycle.

12.4.2 *pin* Diode Limiters

It has been shown by Leenov [1, 21] that the resistance of the i region of a *pin* diode activated through a microwave current is given by

$$R_i = \frac{W}{\sqrt{D_{ap}/2\pi f}} \frac{1}{e/kT} \frac{1}{I_{rf}} \tag{12.74}$$

when the i layer width W is much less than the carrier diffusion length L ($=43\sqrt{\tau}$ μm, where τ is the average carrier lifetime in microseconds). In the preceding relation, D_{ap} is the diffusion coefficient, W is in centimeters, and I_{rf} is in amperes rms. For Si diodes at room temperature relation (12.74) may be simplified as

$$R_i = \frac{W\sqrt{f}}{20I_{rf}} \tag{12.75}$$

where R_i is in ohms, W is in microns, f is in gigahertz, and I_{rf} is in amperes rms.

It may be noted that the action of a high-power microwave signal is much less effective in producing conductivity modulation of the i layer than is a direct current. For example, at 1 GHz the RF current required to reduce R_i to a value of 1 Ω [for a diode with $W = 50$ μm (2 mil)] is $I_{rf} = 50\sqrt{1}/(20 \times 1) = 2.5$ A (rms), whereas the corresponding requirement for direct current will be only about 50 mA.

After the incident high-power microwave pulse ceases, the concentration of holes and electrons in the i region does not disappear immediately but decays exponentially with a time constant equal to the average carrier lifetime. During this recovery period the limiter has a relatively high insertion loss. The *recovery time* is defined as the time required for the limiter's insertion loss to return to within 3 dB of its low-level insertion loss following the cessation of the high-power pulse. In a typical radar system this period corresponds to echos from nearby targets, and a higher insertion loss in the receiver path may be tolerated.

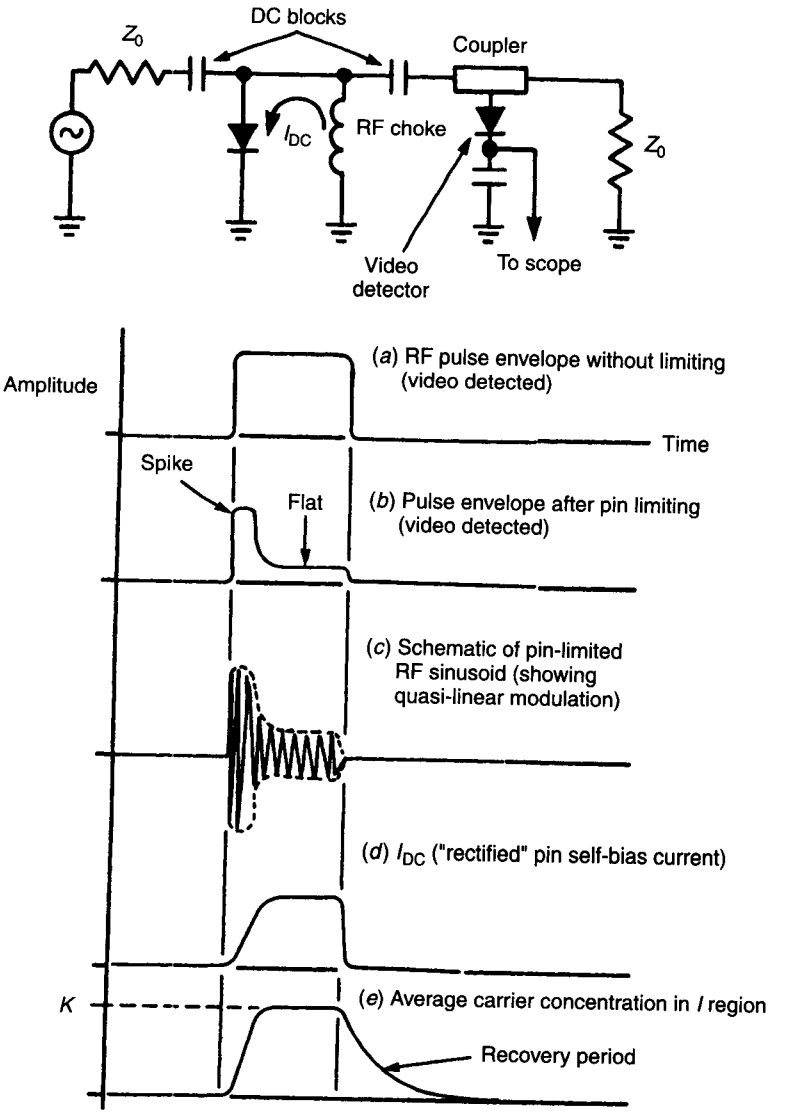


Figure 12.45 Various waveforms in *pin* diode limiter. (After White [1]. Reprinted with permission of Van Nostrand.)

The recovery time is related to the diode minority carrier lifetime τ. For a given τ, the recovery time is linearly proportional to the peak RF power applied up to some peak power level above which recovery time increases rapidly because of thermal heating of the diode. Measurement of the recovery time thus allows a nondestructive monitoring of the power-handling capacity of a *pin* diode limiter circuit.

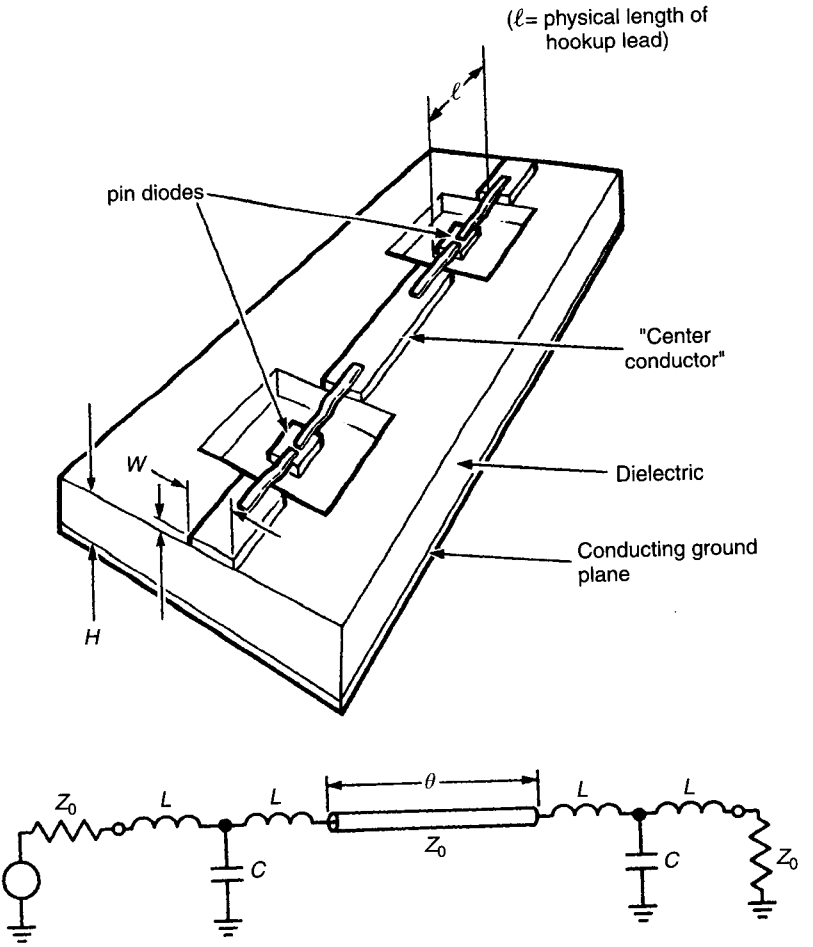


Figure 12.46 Limiter in microstrip configuration and its equivalent circuit. (After White [1]. Reprinted with permission of Van Nostrand.)

12.4.3 Limiters in Microstrip Configuration

A practical implementation of a limiter circuit using two shunt-mounted diodes spaced along a microstrip line is shown in Fig. 12.46. The diodes are mounted on the ground plane (after punching out holes in the substrate dielectric material). Flexible wire (or rectangular ribbon) leads provide mechanical stress relief in the connection of the strip conductor to the diode. In addition, by proper selection of lead lengths and sizes, series inductances are realized that can be used to form a matched T filter and thereby tune out the reflection that would be otherwise introduced by the diode capacitance. An equivalent circuit (with reversed-bias *pin* diodes represented by capacitances C) is also shown in Fig. 12.46. The circuit configuration shown in Fig. 12.46 is also well suited for *pin* diode switches discussed earlier.

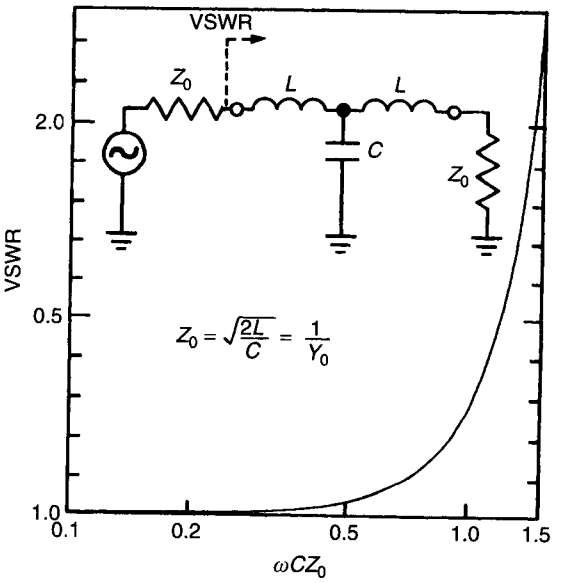


Figure 12.47 VSWR for matched-T diode configuration. (After White [1]. Reprinted with permission of Van Nostrand.)

If the total inductance, $2L$, of the two bonding straps satisfies

$$Z_0 = \sqrt{\frac{2L}{C}} \tag{12.76}$$

the equivalent T circuit will resemble a length of a transmission line and its transmission match will be frequency independent in the range for which $\omega L \ll Z_0$ and $\omega C \ll (1/Z_0)$. The VSWR resulting from a single T network of this type is shown in Fig. 12.47. If we consider a diode with $C = 0.32$ pF, mounted in a 50-Ω system, half of the incident power is reflected at 20 GHz (where $\omega CZ_0 = 2$), corresponding to a VSWR value of 5.8. The VSWR value is 1.28 at 10 GHz, which is a more realistic upper frequency for a limiter in this configuration. An estimate for strap inductance L is obtained by treating it as a microstrip transmission line with air dielectric and an average height above the ground plane given by means of diode chip thickness and microstrip dielectric substrate thickness used. If Z_0 is the characteristic impedance of this microstrip line, the inductance per unit length is given by (Z_0/v_p) where v_p is the phase velocity for the air medium.

12.5 DESIGN OF VARIABLE ATTENUATORS

Voltage-controlled variable attenuators are important control elements and are widely used for automatic gain control circuits. They are indispensable

for temperature compensation of gain variation in broadband amplifiers. Both *pin* diodes and GaAs MESFET devices are used for design of variable attenuators. Apart from the use of MESFETs in the passive mode (described in Section 12.5.2), active MESFET amplifier circuits may also be used for variable-attenuation circuits. Dual-gate MESFET amplifiers with controlled voltage applied to the second gate are ideal for this purpose.

In this section, we discuss variable attenuators using *pin* diode and passive-mode MESFETs only.

12.5.1 *pin* Diode Attenuators

The fact that the resistance of the intrinsic layer of *pin* diodes is a strong function of the DC bias current is employed for designing current-controlled variable attenuators at microwave frequencies.

The resistance of the intrinsic layer under forward bias is given by

$$R_i = \frac{W^2}{2\mu_{ap}\tau I_0} \tag{12.77}$$

where W is the width of the i layer, μ_{ap} is the ambipolar mobility (610 cm^2/V-s in silicon), τ is the lifetime of carriers, and I_0 is the DC bias current. The derivation of (12.77) is based on several simplifying assumptions. For practical diodes [22] variation with I_0 is more like $I_0^{-0.87}$ than $I_0^{-1.0}$. It is often advisable to make an experimental measurement of the variation of the R_f (which includes R_j and contact resistance) with bias current.

An important characteristic of a variable attenuator is that its input impedance should remain constant, so that the attenuator remains matched over its operating range. One way of realizing this is the π network shown in Fig. 12.48. This circuit consists of three *pin* diodes for R_1, R_2, R_3, respectively. For matching, the impedance of the network to the right of AA' (say Z_A) in parallel with R_2 should equal Z_0, that is,

$$\frac{1}{Z_0} = \frac{1}{R_2} + \frac{1}{Z_A} \tag{12.78}$$

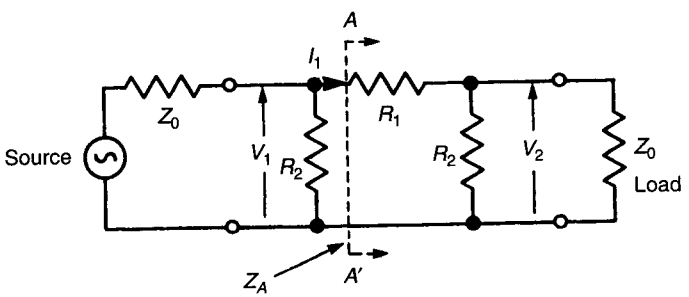


Figure 12.48 Resitive π network for *pin* attenuators.

where $Z_A = R_1 + R_2Z_0/(R_2 + Z_0)$. Since we have $V_1 = I_1Z_A$ and $V_2 = I_1R_2Z_0/(R_2 + Z_0)$, the attenuation ratio V_1/V_2 may be written as

$$\frac{V_1}{V_2} = \frac{Z_A}{R_2Z_0/(R_2 + Z_0)} \tag{12.79}$$

Substituting for Z_A we have

$$K = \frac{V_1}{V_2} = \frac{R_2 + Z_0}{R_2 - Z_0} \tag{12.80}$$

Resistance values R_2 and R_1 can be expressed in terms of the characteristic impedance Z_0 and the voltage ratio K as

$$R_2 = \frac{Z_0(K + 1)}{K - 1} \tag{12.81}$$

$$R_1 = \tfrac{1}{2}Z_0(K - 1/K) \tag{12.82}$$

The values of bias currents required to produce the necessary resistance values for R_1 and R_2 are obtained by using (12.77) or by experimental measurements.

12.5.2 MESFET Attenuators

For attenuators employing MESFETs in the passive mode, the basic device mechanism used is the change in the low field resistance of a zero-biased FET controlled by the gate voltage. Three MESFETs may be connected in T or π configuration, as shown in Fig. 12.49*a* and *b*, respectively. The MESFETs may

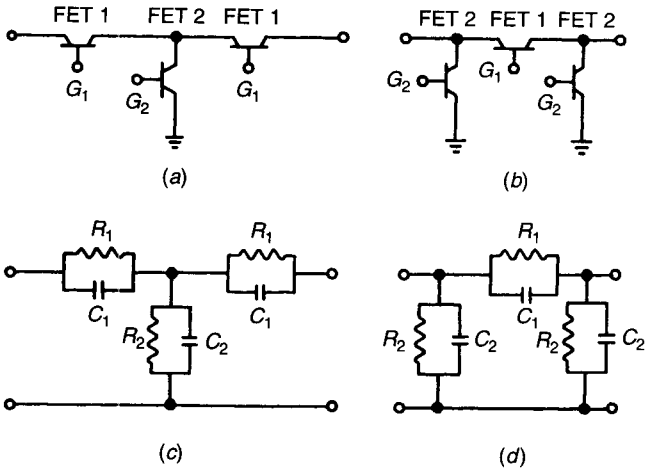


Figure 12.49 T [(*a*) and (*c*)] and π [(*b*) and (*d*)] attenuator networks using MESFETs and equivalent network models.

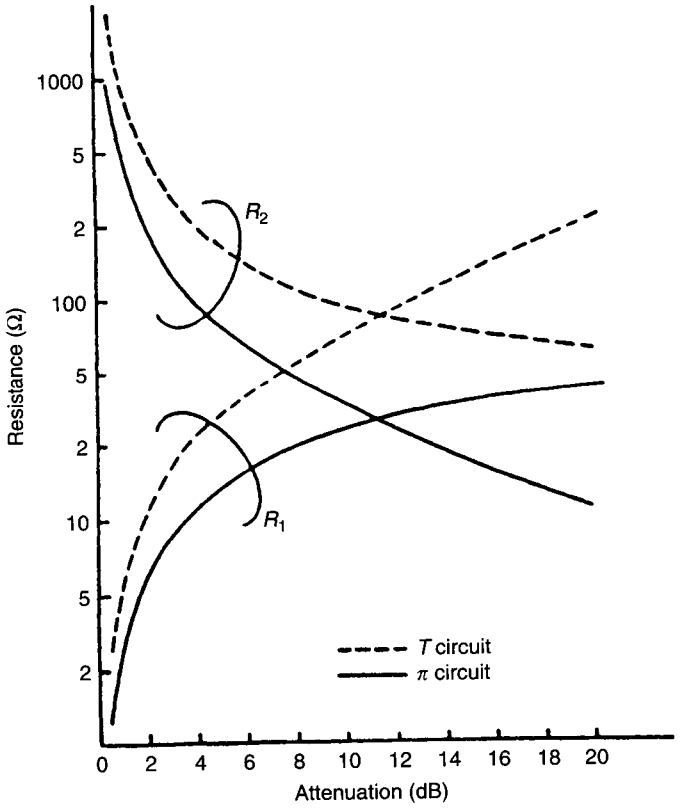


Figure 12.50 Values of resistance R_1 and R_2 needed for attenuator design. (After Tajima et al. [23]. Reprinted with permission of IEEE.)

be modeled by a parallel combination of R and C, as shown in Fig. 12.49*c* and *d*, where the values of R_1 and R_2 vary as a function of the gate voltages. The value of R varies from zero-bias value (R_{on}) to a large value R_{off} when the gate voltage reaches the pinchoff voltage. The values of the capacitances C_1 and C_2 may be considered to be fairly constant with gate voltages.

At lower microwave frequencies the effect of capacitors can be neglected. In such a case, values of R_1 and R_2 needed for various values of attenuation (for $Z_0 = 50\ \Omega$) are shown in Fig. 12.50 [23]. For the same R_1, the π configurations have a lower loss than the T configurations. The minimum value of R_1 ($=R_{on}$) can be reduced by increasing the number of gate fingers (although C_1 will also increase). Larger values of C_1 limit the dynamic range of attenuation at higher frequencies. From the point of view of dynamic range, the T configuration is better than the π configuration.

An attenuator with T-circuit topology (Fig. 12.51) has been reported in [23]. Gate widths of 600 and 200 μm were chosen for the series and shunt

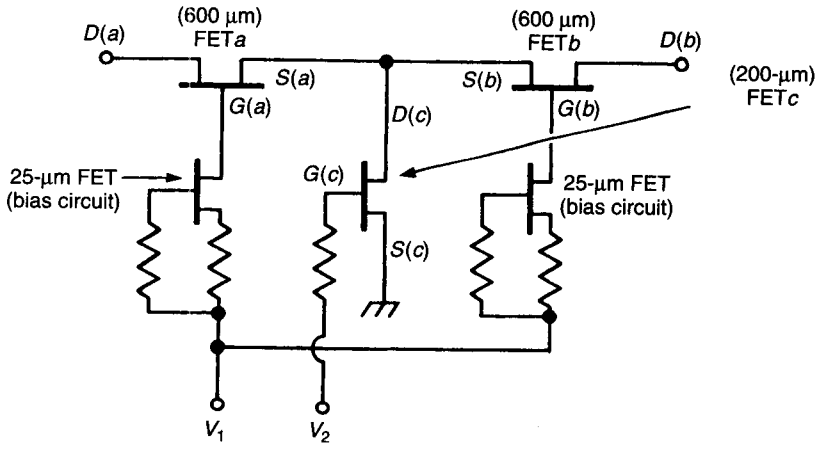


Figure 12.51 Schematic of a MESFET attenuator. (From Tajima et al. [23]. Reprinted with permission of IEEE.)

MESFETs, respectively. Input and output microstrips were connected to drain contacts of FETs *a* and *b*, while source contacts were connected by a common air bridge that served as a connection to the drain contact of FET *c*. The source of FET *c* is grounded through a via hole. Isolation between the RF circuit and DC control circuit was achieved by thin-film resistors and FET resistors that have 25-μm-wide gates.

Experimental results [23] show this attenuator to have a 12-dB dynamic range of attenuation over the 2–18 GHz frequency band, and 17 dB over the 2–12 GHz frequency range. Return loss was better than 10 dB, and minimum insertion loss about 2 dB. The performance did not degrade up to 600-mV input signal.

Acknowledgements. Discussions and assistance provided by Dr. Huantong Zhang in revision of this chapter are gratefully acknowledged.

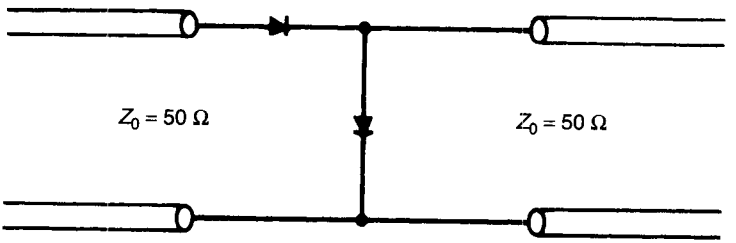
REFERENCES

1. White, J. F., *Microwave Semiconductor Engineering*, Van Nostrand, Princeton, NJ, 1982, Chap. 2.
2. Ayasli, Y., "Microwave Switching with GaAs FET's," *Microwave J.*, Vol. 25, Nov. 1982, pp. 61–74.
3. Yao, J. J., "RF MEMS From a Device Perspective," *J. Micromech. Microengg.*, Vol. 10, 2000, pp. R9–R38.
4. Sokolov, V., J. J. Geddes, A. Contolatis, P. E. Bauhahn, and C. Chao, "A *Ku*-Band GaAs Monolithic Phase Shfter," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-31, 1983, pp. 1077-1082.

5. Pengelly, R. S., *Microwave Field-Effect Transistor-Theory, Design and Applications*, Research Series Press (Wiley), New York, 1982, Chap. 9.
6. Chorney, P., "Multi-Octave, Multi-Throw, *p-i-n* Diode Switches," *Microwave J.*, Vol. 17, Sept. 1974, pp. 39–42, 48.
7. Garver, R. V., "*PIN* Diode Switches: Speed vs. Power," *Microwave J.*, Vol. 21, Feb. 1978, p. 53.
8. Barratt, C., A. Christou, N. Jansen, R. E. Neidert, M. L. Ruess, Jr., and C. W. Young, "New GaAs *p-i-n* Diodes with Lower Dissipation Loss, Faster Switching Speed at Lower Drive Power," *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1983, pp. 507–509.
9. Garver, R. V., *Microwave Diode Control Devices*, Artech House, Dedham, MA, 1976, pp. 63–65.
10. Chang, K., I. Bahl, and V. Nair, *RF and Microwave Circuit and Component Design for Wireless Systems*, Wiley, New York, 2002, Chap. 7.
11. Stark, L., R. W. Burns, and W. P. Clark, "Phase Shifters for Arrays," in M. I. Skolnik (Ed.), *Radar Handbook*, McGraw-Hill, New York, 1970, pp. 12.1–12.65.
12. Bahl, I. J., and K. C. Gupta, "Design of Loaded-Line *p-i-n* Diode Phase Shifter Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, Mar. 1980, pp. 219–224.
13. Mongia, R., I. Bahl, and P. Bhartia, *RF and Microwave Coupled-Line Circuits*, Artech House, Norwood, MA, 1999.
14. Trivedi, D. K., "Design, Performance and Optimization of Reflection-Type *pin* Diode Phase Shifters," Ph.D. Thesis, Indian Institute of Technology, Kanpur, 1981.
15. Eking, R. B., "On the Design of Impedance Transforming Directional Couplers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-19, Apr. 1971, pp. 415–416.
16. Ayasli, Y., S. W. Miller, R. L. Mozzi, and L. K. Hanes, "Wideband S-C Band Monolithic Phase Shifter," *IEEE Monolithic Microwave Millimeter Wave Circuit Symp. Dig.*, San Francisco, May 1984, pp. 11–13.
17. Tsironis, C., and P. Harrop, "Dual-Gate MESFET Phase Shifter with Gain at 12 GHz," *Electron. Lett.*, Vol. 16, July 1980, pp. 553–554.
18. Vorhaus, J. L., R. A. Pucel, and Y. Tajima, "Monolithic Dual-Gate FET Digital Phase Shifters," *IEEE Trans., Microwave Theory Tech.*, Vol. MTT-30, 1982, pp. 991–992.
19. Gazit, Y., and H. C. Johnson, "A Continuously Variable *Ku*-Band Phase/Amplitude Control Module," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1981, pp. 436–438.
20. Hwang, Y., Y. Chen, and R. Naster, "A Microwave Phase and Gain Controller with Segmented Dual Gate MESFETs in GaAs MMIC," *IEEE Microwave Millimeter Wave Circuits Symp. Dig.*, San Francisco, June 1984, pp. 1–5.
21. Leenov, D., "The Silicon *p-i-n* Diode as a Microwave Radar Protector at Megawatt Level," *IEEE Trans. Electron Dev.*, Vol. ED-11, 1964, pp. 53–61.
22. Hewlett-Packard, "An Attenuator Design Using *p-i-n* Diodes," Application Note No. 912, Hewlett Packard Inc, Palo Alto (CA).
23. Tajima, Y., et al., "GaAs Monolithic Wideband (2–18 GHz) Variable Attenuators," *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1982, pp. 479–481.

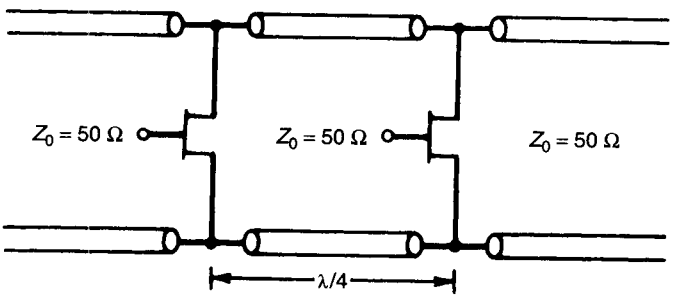
PROBLEMS

12.1 An SPST switch employs two *pin* diodes in the series–shunt configuration shown below. Calculate the insertion loss and isolation at 6 GHz when two *pin* diodes have $C_j = 0.1$ pF, $R_f = 1\ \Omega$, and $L_{\text{int}} = 0.3$ nH. Package capacitances may be ignored.



Compare the results with the switch performance when only one diode is used in (a) series configuration and (b) shunt configuration.

12.2 An SPST switch uses two GaAs MESFETs (in passive mode) mounted in shunt across a transmission line ($Z_0 = 50\ \Omega$) and spaced a quarter wavelength apart, as shown in the figure below. The switch is designed for a center frequency of 10 GHz where the MESFET equivalent circuit parameters are $R_{\text{on}} = 2.7\ \Omega$, $R_{\text{off}} = 3\ \text{k}\Omega$, and $C_{\text{off}} = 0.25$ pF. Compare the performance with conventional shunt-mounted switch designs (Fig. 12.5) when (a) only one MESFET is used and (b) two MESFETs are used at the same location (in parallel).



12.3 Evaluate and plot the response (insertion loss and isolation) of the switching circuit in Problem 12.2 as a function of frequency (from 7.5 to 12.5 GHz). Assume the MESFET equivalent circuit to be valid over this frequency range. How is this performance modified when the high-impedance state capacitances of the MESFET are compensated by appropriate shunt inductances?

12.4 A *pin* diode chip is mounted in shunt across a 50-Ω microstrip line with two bonding tapes as shown in Fig. 12.5. The length of the bonding tapes and hence their inductance L can be adjusted. Find the value of the inductances needed to optimize the insertion loss at 6 GHz. In reverse bias the *pin* diode chip may be represented by a capacitance $C_j = 0.1$ pF.

12.5 An SP3T switch is designed using *pin* diode chips ($C_j = 0.1$ pF, $R_f = 1\ \Omega$, $R_r = 4\ \Omega$, and $L_{\text{int}} = 0.3$ nH). The center design frequency is 3 GHz. Calculate and plot switch performance over the frequency range 2–4 GHz for (a) series configuration and (b) shunt configuration.

12.6 An SP3T switch is designed using three GaAs MESFETs (passive mode) in series configuration. The MESFETs may be modeled by $R_{\text{on}} = 2.7\ \Omega$, $R_{\text{off}} = 3\ \text{k}\Omega$, and $C_{\text{off}} = 0.25$ pF. Calculate the insertion loss and isolation at an operating frequency of 10 GHz when all the transmission lines have a characteristic impedance of 50 Ω each.

12.7 Calculate how the performance of the SP3T switch in Problem 12.6 is modified when:

(a) MESFET chips are used in a hybrid MIC circuit; and, in addition to the MESFET parameters specified earlier, we have a series inductance of 0.2 nH because of bonding wires.

(b) MESFET capacitance in the high-impedance state is compensated by a parallel inductance of suitable value at 10 GHz. (Bonding wire inductances are not present in this case.)

12.8 A 180° single-bit switched-line phase shifter is designed using two SPDT switches employing two MESFETs each in series configurations. The MESFETs may be modeled by $R_{\text{on}} = 2.7\ \Omega$, $R_{\text{off}} = 3\ \text{k}\Omega$, and $C_{\text{off}} = 0.25$ pF. The design frequency is 10 GHz and the electrical length of the smaller of the switched lines is taken to be 30°. Plot the differential phase shift and insertion loss (for both the states) as a function of frequency from 7.5 to 12.5 GHz. Do you observe any off-path resonances? Compare the phase shift response with the performance you would expect if the SPDT switches were ideal. (Any available microwave circuit analysis program may be used for this problem.)

12.9 A main-line-mounted loaded-line type of phase shifter is designed for operation at 10 GHz using *pin* diode chips. These diodes can be modeled by $R_f = 1\ \Omega$ and $L_{\text{int}} = 0.3$ nH in forward bias and by a series combination of $L_{\text{int}} = 0.3$ nH, $R_r = 4\ \Omega$, and $C_j = 0.1$ pF in reverse bias. The two diodes are spaced 90° apart and the characteristic impedance of the lines at external ports is Z_0. Find the value of extra inductance that needs to be added and the line impedance Z_0 needed to obtain a differential phase shift of 22.5°. (R_f and R_r may be ignored for this design.)

12.10 Use any available microwave circuit analysis program to plot the frequency response (VSWR, $\Delta\phi$, and insertion loss) of the phase shifter designed in Problem 12.9 over a frequency range of 8.0–12.0 GHz. Include the diode resistances in two bias states and recompute the phase shifter performance in the above-mentioned frequency range in order to observe the effect of diode resistances.

12.11 A loaded-line phase shifter is designed using two stub-mounted MESFETs (in passive mode). The MESFETs may be represented by $R_{on} = 2.7\ \Omega$ in the low-impedance state and $C_{off} = 0.25$ pF in the high-impedance state. Find the stub length and impedance for a 22.5° phase shifter bit to operate at 10 GHz. The spacing between stubs is 90° and the lines at two ports have $Z_0 = 50\ \Omega$. R_{on} can be ignored.

12.12 Use any available microwave circuit analysis program to plot the frequency response (VSWR, $\Delta\phi$, and insertion loss) of the phase shifter designed in Problem 12.11 over a frequency range of 8.0–12.0 GHz. Repeat these computations by including R_{on} in the low-impedance state and compare the results.

12.13 A reflection-type phase shifter is designed using *pin* diode chips with $R_f = 0.4\ \Omega$, $R_r = 0.5\ \Omega$, $C_j = 1$ pF, and $L_{int} = 0.3$ nH. The design frequency is 3.18 GHz. Calculate the value of Z_m (the equivalent impedance the *pin* diodes should look into) (a) when the desired phase shift $\Delta\phi$ is 180° and (b) when $\Delta\phi$ is 90°. (Effect of diodes resistances may be ignored.)

12.14 Equation (12.55) for Z_m is valid for lossless devices only. Derive a corresponding expression for the case when the device resistances cannot be neglected; that is, in the low-impedance state $Z_f = R_f + jX_f$, and in the high-impedance state $Z_r = R_r + jX_r$.

12.15 Consider the reflection phase shifter design of Problem 12.13. Repeat the calculations for Z_m taking the effects of diode resistances into account. (Use the expression Z_m derived in Problem 12.14.)

12.16 When reflection-type phase shifters are designed in monolithic configuration, bonding wire inductance is not present. These monolithic devices can be represented by $Z_l = R_l$ in the low-impedance state and by $Z_h = R_h + jX_h$ in the high-impedance state. The expression for Z_m derived in Problem 12.14 can be used in this case also. Find the values of Z_m for $\Delta\phi = 180°$ and $\Delta\phi = 90°$ when $Z_l = 0.4\ \Omega$ and $Z_h = 0.5 - j50\ \Omega$.

12.17 Design a 180° reflection phase shifter using a quarter-wave transforming section, a two-branch hybrid, and *pin* diode chips with $C_j = 1$ pF and $L_{int} = 0.3$ nH. The center frequency for design is 3.0 GHz. Diode resistances may be ignored. Simulate the circuit using any available

microwave circuit analysis program and compute $\Delta\phi$, VSWR, and insertion loss (in two states) for frequencies ranging from 2.5 to 3.5 GHz (in 0.1-GHz steps). The effect of discontinuity reactances may be ignored. Take the substrate parameters as $\epsilon_r = 2.5$, $h = 0.8$ mm, and $t = 18$ μm.

12.18 Consider the case when the diodes used in the design of Problem 12.17 are lossy with $R_f = 0.4$ and $R_r = 0.5$ Ω. How will the performance of the phase shifter ($\Delta\phi$, VSWR, and insertion loss over 2.5–3.5 GHz) be modified when the design remains unchanged? Use any available microwave circuit analysis program for these computations.

12.19 Design a 180° reflection phase shifter using a two-branch impedance transforming hybrid and *pin* diodes with $C_j = 1$ pF and $L_{\text{int}} = 0.3$ nH. The design frequency is 3.18 GHz. The effect of diode resistances may be neglected. Simulate this circuit using any available microwave circuit analysis program (microstrip configuration with $\epsilon_r = 2.5$, $h = 0.8$ mm, and $t = 18$ μm) and compare its performance with the design of Problem 12.17.

13

FREQUENCY MULTIPLIERS AND DIVIDERS

Robert G. Harrison

13.1 INTRODUCTION

This chapter considers circuits that perform frequency multiplication and division, functions that are unlike the analog operations of frequency addition and subtraction that are performed by mixers.

13.1.1 Basics of Frequency Multiplication and Division

Several different phenomena can be used to realize frequency multiplication or frequency division. Frequency multiplication is easily achieved, but conditions for frequency division are more stringent.

The simplest frequency-multiplying device is the nonlinear resistor (varistor), which has a nonlinear static i–v relation. At RF frequencies pn-junction diodes can be used, but at microwave frequencies low-parasitic Schottky diodes (Chapter 8) or specialized devices such as Mott diodes [1, 2] are required. Microwave three-terminal devices including GaAs FETs, HBTs, HEMTs, and SiGe bipolar transistors can be operated as "active varistors." When a two-terminal (passive) positive nonlinear resistor multiplies a frequency by an integer N, the maximum attainable efficiency is $1/N^2$ [3], but with an active three-terminal varistor, multiplication with gain is possible [4].

It is impossible [5] to obtain frequency division from a passive nonlinear resistor. However, frequency division *is* possible using two-terminal virtual-negative-resistance devices, such as Gunn diodes: The natural oscillation fre-

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

quency f_0, determined by the transit time of traveling space-charge domains, can be entrained by an injected signal near a harmonic of f_0. Active nonlinear-resistance devices can be used in regenerative (Miller) frequency dividers [6, 7] in which an input signal is applied to the LO port of a mixer, while the IF output is amplified and fed back to the RF port. With appropriate filtering in the feedback path, frequency division can be obtained.

Two-terminal nonlinear-*reactance* devices can be used to realize both parametric frequency multipliers and dividers. Manley and Rowe proved [8] that two-terminal nonlinear-reactance multipliers and dividers have a maximum possible efficiency of 100%. Under certain conditions parametric frequency converters can act as chaos generators [9]; this may be highly undesirable in practical applications. The usual nonlinear-reactance device is the varactor. Classical varactors are *pn*-junction or Schottky diodes designed to exploit the asymmetrical voltage-dependent depletion layer capacitance, a high-Q nonlinear reactance under reverse bias. A recent variant is the heterostructure-barrier varactor* (HBV) [10, 11]. The symmetrical HBV has maximum nonlinearity at zero bias and can operate as an efficient frequency tripler. In all varactors, the underlying mechanism is a nonlinear charge–voltage relationship.

In the step recovery diode (SRD) the significant nonlinearity is the charge storage associated with the diffusion capacitance. The SRDs generate sharp pulses in the time domain, translating into a rich spectrum of harmonics in the frequency domain.

Injection-locked oscillator (ILO) multipliers and dividers work by synchronizing (entraining) an oscillator with an injected signal near a subharmonic or harmonic of the natural frequency.

Finally, both static and dynamic digital circuits can be designed to operate as wide-band microwave frequency dividers.

Noise Augmentation in Frequency Multipliers. The FM (and phase) noise of an input signal applied to an Nth-order multiplier will be increased by a factor of at least N^2. This is equivalent to an increase of $10 \log_{10}(N^2) = 20 \log_{10}(N)$ decibels, amounting to 6 dB if $N = 2$.

Noise Reduction in Frequency Dividers. Similarly, the FM (and phase) noise of a frequency-divided signal can be reduced by a factor of at best $1/N^2$, or equivalently by $-20 \log_{10}(N)$ decibels.

13.1.2 Applications

This section reviews applications of frequency multipliers and dividers.

Applications of Frequency Multipliers. Frequency multipliers are used in microwave and millimeter-wave transmitter and receiver chains and to pro-

* Previously known as the *quantum barrier varactor* or the *single-barrier varactor*.

vide LO sources at frequencies where fundamental oscillators are unavailable. Extension to millimeter-wave frequencies is important for communications, electronic warfare, radar, satellites, atmospheric physics, and spectroscopic remote sensing. Sources at submillimeter frequencies are needed for plasma diagnostics, spectroscopy, and for radio-astronomy mixers.

At low microwave frequencies oscillators based on active three-terminal devices are commonly used. At intermediate frequencies (20–60 GHz), when two-terminal Gunn diode or IMPATT oscillators are too hot and/or noisy, or where small size and low cost are important, as in phased-array radars, monolithic multipliers have the advantage. At millimeter and submillimeter wavelengths (30–100 GHz), passive multipliers based on either resistive nonlinearities (varistors) or reactive nonlinearities (varactors) are the best choices: Varistors provide bandwidth; varactors offer efficiency and output power. Schottky diodes can be used in either role and have generated milliwatts of power at frequencies up to ~1 THz [12]. Previously the only millimeter-wave devices were unreliable whisker-contacted Schottkys in waveguide mounts. With the microchannel structure [13] reliable planar multipliers are feasible at ever-increasing frequencies.

New nonlinear-reactance devices such as the HBV are being used in both lumped (fixed-tuned) and traveling-wave (wide-band) multiplier structures.

Applications of Frequency Dividers. Frequency dividers are used as prescalers, for frequency translation* and bandwidth compression, in communications systems, in frequency-modulating phase-locked loops, and for phase noise reduction.

Prescalers. Divide-by-N circuits can be used as front-end prescalers in microwave frequency counters and digital frequency discriminators. The maximum input frequency is set by the capability of the first divider stage.

Frequency Translation. Since frequency dividers divide both input *bandwidth* and *frequency*, they can translate high-frequency (typically FM) analog signals to lower frequencies for digital processing. The fractional bandwidth $\Delta f/f$ is invariant under frequency division, whereas down conversion using a mixer results in an *increase* in $\Delta f/f$; see Fig. 13.1. Digital frequency memories represent an application where wide-band frequency division is required.

Communications Systems. In communications systems frequency dividers can improve frequency stability and reduce noise. Conventionally, frequency stability is obtained by locking a voltage-controlled oscillator (VCO) to a low-frequency reference oscillator using a phase-locked loop (Fig. 13.2). A frequency multiplier then provides a microwave output at f_o. With an external

* Parametric frequency dividers have also been used in low-cost no-battery theft prevention systems [14].

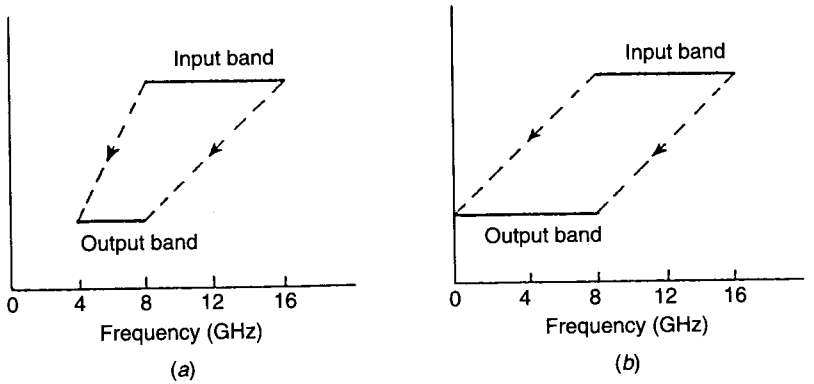


Figure 13.1 (*a*) Example of frequency translation by division: fractional bandwidth is $\Delta f/f$ is 66% (one octave) at both input and output. (*b*) Frequency translation by (mixer) down conversion: $\Delta f/f$ is 66% at the input, but at the output it is 100% (1.5 octaves).

reference f_r, at say 100 MHz, a sample of the VCO output at f_1, nominally also 100 MHz, is applied to the second input of the phase comparator. If f_r and f_1 are the same, the comparator output is a DC voltage dependent on the phase difference between the two signals. If f_r and f_1 are different, the output will be an AC signal at the difference frequency. In either event, upon loop lockup, the frequency f_1 will be nearly equal to f_r. For an output at $f_o = 10$ GHz, a multiplier with $N = 100$ would be required. This is unsatisfactory since the output will contain sidebands at 10.1 and 9.9 GHz due to heterodyne mixing in the

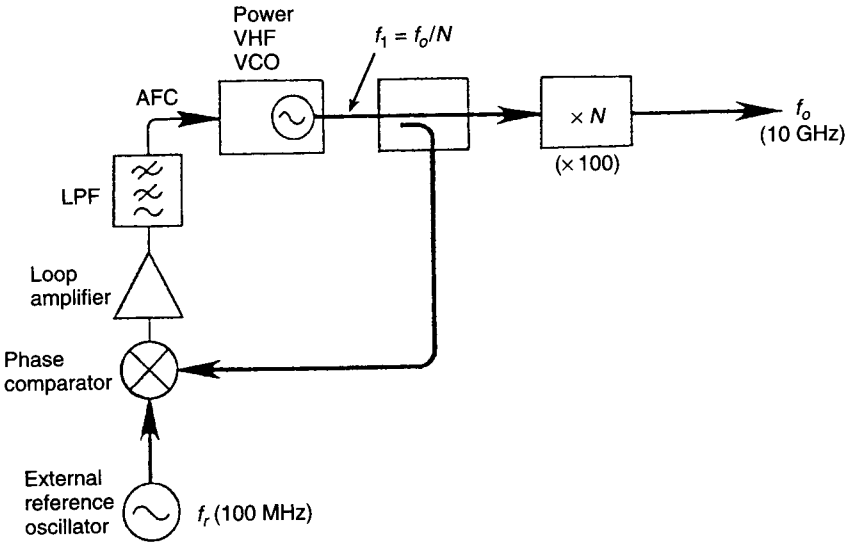


Figure 13.2 Conventional phase-locked loop (PLL) using $\times N$ multiplier to obtain microwave output.

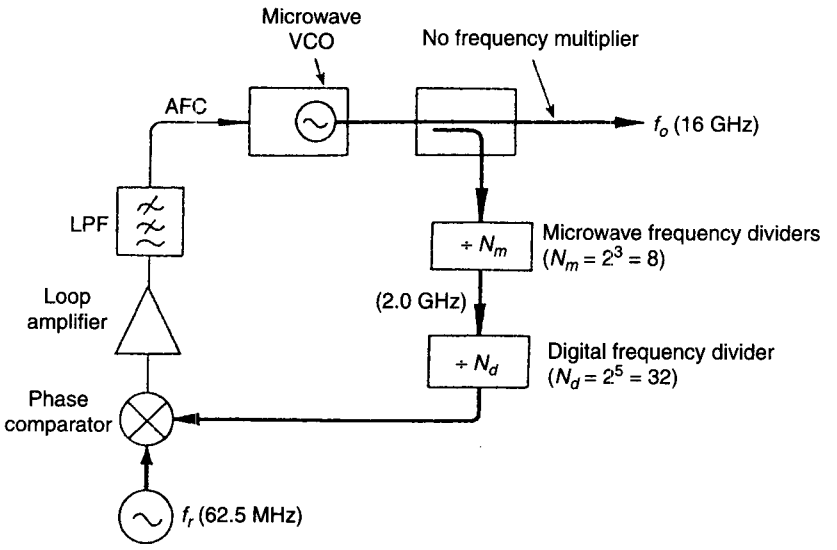


Figure 13.3 Phase-locked loop with direct reference-oscillator control of microwave VCO.

multiplier. These sidebands would be difficult to filter out. Moreover the power in these sidebands would be enhanced by a factor 20 $\log_{10}N$ due to the multiplication process itself. For $N = 100$, the enhancement would be an unacceptable 40 dB. In addition, the multiplier will increase any FM noise on the signal f_1 (which tracks that on f_r) by the factor N. Finally, the noise figure of the multiplier would contribute to the overall noise figure of the system.

A better approach is to use a *microwave* VCO and to *frequency divide* a sample of the output at f_o for phase comparison with an LF reference f_r; see Fig. 13.3. Here the reference f_r controls the microwave VCO directly. A sample of the output f_o, for example, at 16 GHz, is frequency divided by $N_m = 2^K$, using a sequence of K microwave frequency halvers. In this example, $K = 3$ so $N_m = 8$ and the input to the digital divider is only 2 GHz. Further digital division then reduces the frequency by a factor N_d to permit phase comparison with the reference f_r. Here $N_d = 32$, giving a total division ratio $N_m N_d = 256$, so that a reference frequency $f_r = 62.5$ MHz would be chosen. The advantage is that the FM noise of the output signal f_o is that of the reference oscillator. There is no degradation due to frequency multipliers, since none are used.

Frequency-Modulated PLL. In a conventional PLL, direct FM of the VCO is possible; see Fig. 13.4*a*. However, the modulating frequency f_m must be greater than about 1.5 times the 3-dB loop bandwidth [15, 16]; otherwise the PLL will treat the modulation as an error signal and cancel it out. At the same time, the maximum deviation Δf must not exceed f_m, otherwise the loop could lock to a modulation sideband. Thus the modulation index must be less than 1.0, limiting the utility of the conventional system. However, if frequency modulation is

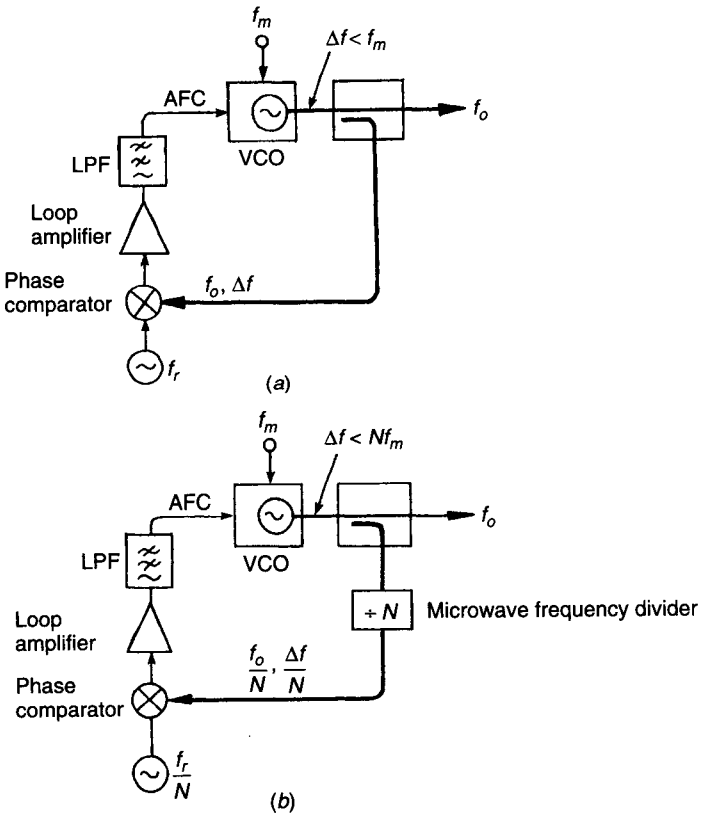


Figure 13.4 Frequency modulation applied to (*a*) conventional phase-locked loop (PLL), and (*b*) PLL with microwave frequency divider.

applied to a PLL with a frequency divider by N in the loop (see Fig. 13.4*b*), the deviation Δf and the output frequency f_o will both be divided by N. Then the reference frequency can also be divided by N, so that the phase comparison can be done at f_o/N rather than at f_o. Such a system operates correctly when $\Delta f/(Nf_m) < 1.0$, that is, the effective modulation index of the system has been increased by a factor N.

Frequency Counters. A conventional way of measuring the frequency of signals consisting of RF pulses or with large FM components is to use heterodyne or transfer oscillator techniques. These depend on mixing the unknown signal with the Nth harmonic of a local oscillator. In the heterodyne case, a low enough output frequency can be measured using a digital counter. In the transfer oscillator case, the LO frequency is tuned until the frequency difference between the input and the Nth harmonic of the LO is zero. A digital counter then measures the LO frequency. Knowing this and the harmonic number N,

the input frequency can be calculated. Heterodyne oscillators have limited ability to measure the frequency of pulsed signals. Transfer oscillators can measure the fundamental frequency of pulsed signals but are unable to handle signals with large FM components. An alternative is to *frequency divide* the input signal by a factor M, count the output digitally, and multiply the result by M. This approach avoids the disadvantages of heterodyne and transfer oscillator techniques but requires frequency dividers with good RF pulse response and FM capability, such as the parametric frequency halvers of Section 13.3.2.

Phase Noise Reduction. In principle the phase noise of a microwave source can be reduced by a factor $20 \log_{10} N$ by following it by a divide-by-N stage. In practice there will be some phase noise degradation at very small offsets from the reference frequency. The phase noise incurred in translating a reference f_r down to a lower frequency f_r/N can be compared for (a) down conversion by mixing and (b) frequency division. Let the reference oscillator instantaneous voltage be

$$v_r(t) = V_r \sin[\omega_r + \psi \sin(\omega_n t)] \tag{13.1}$$

where V_r = peak reference oscillator voltage
$\omega_r = 2\pi f_r$ = reference oscillator angular frequency
ψ = peak phase noise deviation
ω_n = angular frequency of the phase noise that frequency modulates ω_r

and all the phase noise is represented by $\psi \sin(\omega_n t)$. In the mixer case, assume that the LO signal is

$$v_L(t) = V_L \sin(\omega_L t) \tag{13.2}$$

and that the output signal of interest is given by the product

$$v_o(t) = v_r(t) v_L(t) \tag{13.3}$$

Then it can be shown (Problem 13.1) that case (a) has a worst-case timing error of $N\psi/\omega_r$, whereas case (b) provides a peak timing error of only ψ/ω_r, that is, N times less.

Application Requiring Simultaneous Use of Multipliers and Dividers. Demodulation of suppressed-carrier PSK signals requires carrier reconstitution from the PSK signals themselves. One way of doing this for a biphase signal is shown in Fig. 13.5. A sample of the PSK signal (*a*) is frequency doubled to eliminate the phase coding. To remove sidebands due to the phase switching, the signal is narrow-band filtered by a tracking PLL to produce a sinusoidal signal (*b*). Next, it is frequency divided by 2 to produce a *recovered carrier* (*c*) at the

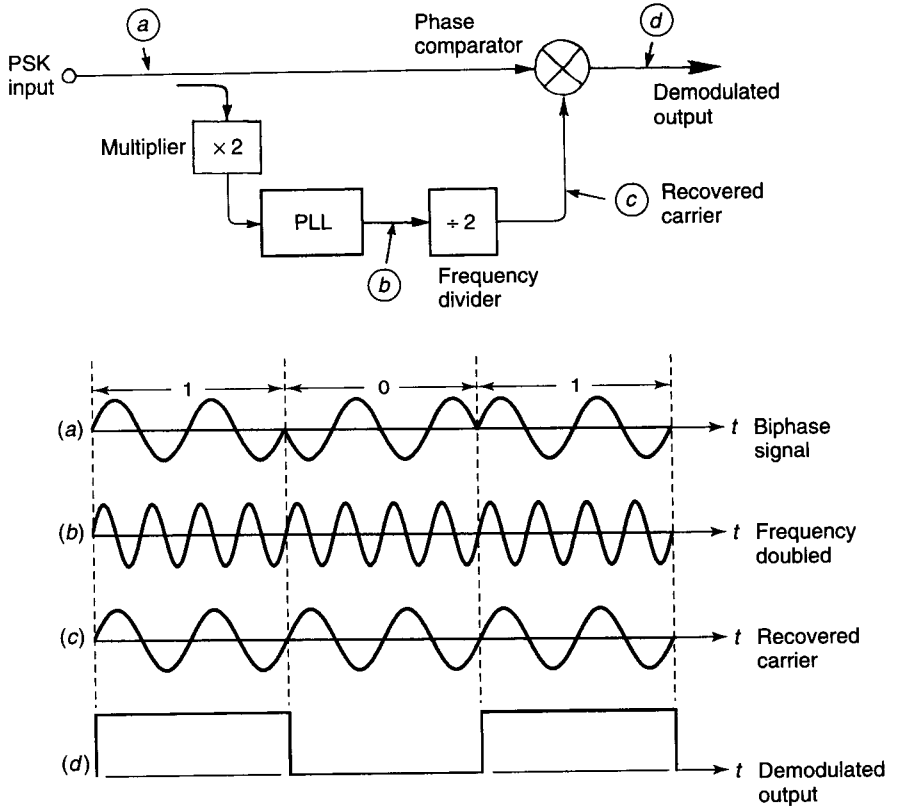


Figure 13.5 Phase-shift keying (PSK) carrier recovery scheme.

appropriate frequency. Finally the input PSK signal and the recovered carrier are fed to the phase comparator. This produces an output DC level (d) that depends on whether the input phases are equal or opposite, that is, it produces a demodulated output. Similar schemes are possible for 4-, 8-, 16-phase, and so on, PSK signals. This means that by using appropriate frequency division and multiplication techniques, microwave PSK signals can be demodulated directly, whereas in conventional systems, demodulation is done at an IF frequency in the 10- to 100-MHz range.

13.2 FREQUENCY MULTIPLICATION

Frequency multiplication involves nonlinearity. Depending on the circuit embedding, the nonlinear resistive or reactive characteristics of passive semiconductor devices can be used to achieve frequency multiplication through

harmonic generation. Harmonic generation in nonlinear resistive devices is accompanied by loss, but multiplication in nonlinear reactive devices can in principle be lossless. Real diodes have both resistive nonlinearities (conduction current flows in ohmic regions) and reactive nonlinearities (displacement current flows in depletion regions). An indicator of the mode of operation is the device Q: resistive multipliers tend to have low Q and low efficiency with wide bandwidth; reactive multipliers display high Q and good efficiency with narrow bandwidth. When active devices are used as frequency multipliers, both types of nonlinearity may be involved.

13.2.1 Types of Multipliers

Frequency multiplication can be achieved in at least six different ways:

1. *Nonlinear-resistance multipliers* exploit harmonic generation due to static nonlinear v–i (voltage–current) diode characteristics.
2. *Nonlinear-reactance multipliers* depend on parametric multiplication using the nonlinear q–v (charge–voltage) relationship of varactor diodes.
3. *Step-recovery diodes* (*SRDs*) can provide high-order multiplication and comb-spectrum generation as well as pulse sharpening.
4. *Nonlinear transmission lines* (*NLTLs*): Transmission lines periodically loaded with either asymmetrical or symmetrical varactors can be used as untuned multipliers. They can provide both efficiency and wide bandwidth [17].
5. *Active multipliers* use simultaneous harmonic generation and gain in active devices such as GaAS FETs, HBTs, HEMTs, and SiGe bipolars. The primary multiplication mechanism is the nonlinearity of the transconductance transfer function.
6. *Injection-locked oscillator* (*ILO*) *multipliers* involve synchronizing an oscillator to a submultiple of its natural frequency. Locking gain can be very high (~50 dB or more), but the operational bandwidth is correspondingly small (the fractional bandwidth is $\sim 1/Q$).

Item 3 has been dealt with in Chapter 8; the other items (except 4) are discussed below.

13.2.2 Nonlinear-Resistance Multipliers

Frequency multipliers based on monotonic-positive nonlinear resistors are useful when wide bandwidth and/or high power is needed. They can be used at higher frequencies (up to the infrared) and over wider bandwidths than varactor multipliers, which have frequency limitations due to the saturation velocity of the neutral space-charge layer [18]. Additionally, resistive multipliers are very stable and are not prone to parametric oscillations. The penalty

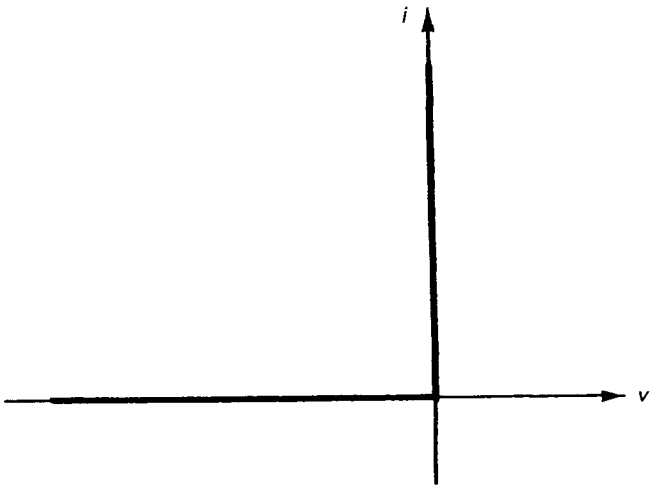


Figure 13.6 Ideal lossless rectifier characteristic.

for using resistive multipliers is a lower conversion efficiency than for reactance (varactor) multipliers (Section 13.2.3) or active multipliers (Section 13.2.4).

Fundamental limitations of resistive multipliers are reviewed next; this will be followed by a discussion of doublers and triplers.

The Page–Pantell Inequality. For a positive nonlinear resistor (varistor) such that the voltage v is a single-valued function of the current i, and $\partial i/\partial v \geq 0$, all v, the power P_N produced at the Nth harmonic of an input at frequency f_{in} is related to the power P_1 absorbed by the resistor at f_{in} by the Page–Pantell inequality [19–22]:

$$\frac{P_N}{P_1} \leq \frac{1}{N^2} \tag{13.4}$$

This applies to *all* resistive nonlinearities conforming to the above specification. This includes the ideal rectifier of Fig. 13.6, which is frequently (but erroneously) assumed to be lossless and to provide 100% multiplier efficiency. It also applies to the ideal exponential diode (see Chapter 8) defined by

$$i = I_s(e^{\Lambda v} - 1) \tag{13.5}$$

where I_s is the saturation current and $\Lambda = q_e/(nkT)$ is the inverse of the thermal voltage. Equation (13.4) shows that no matter how ingenious the circuit designer, resistive* doublers, triplers, and quadruplers can *never* have efficiencies exceeding 25, 11.1, and 6.25%, respectively.

* Excluding negative-resistance devices such as resonant tunneling diodes.

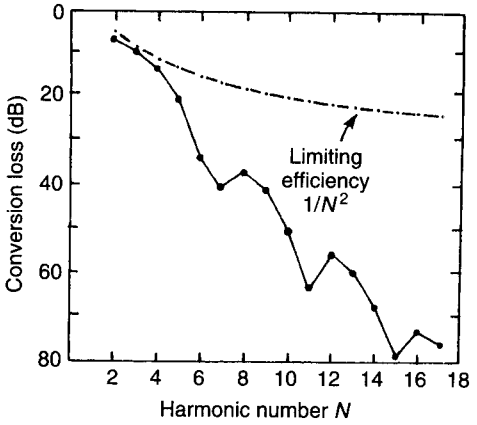


Figure 13.7 Multiplier efficiency vs. harmonic number, exponential law analysis. The dots are connected for clarity: $nkT/q_e = 25$ mV, $I_s = 1$ μA, $R_s = 100$ Ω. (After Benson and Winder [23]. Reprinted with permission of IEE.)

A harmonic generator using the exponential diode of (13.5) with a series resistance R_s has been analyzed [23]; the result is shown in Fig. 13.7. For $N > 4$ the conversion loss increases with N much faster than is predicted by (13.4).

Resistive Doublers. A prototype resistive doubler is the "balanced full-wave rectifier" of Fig. 13.8*a*. In microwave implementations the transformer is

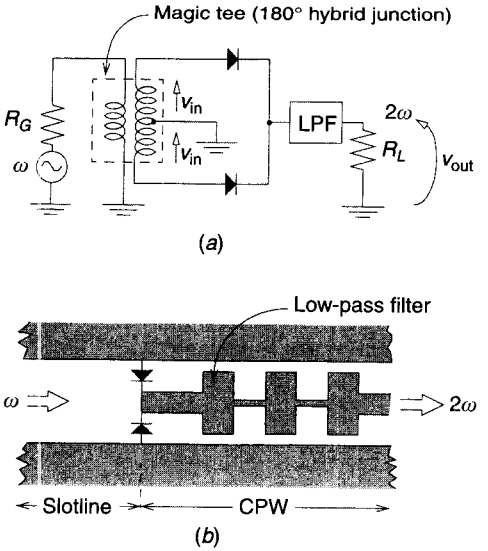


Figure 13.8 (*a*) Full-wave balanced doubler. Low-pass filter eliminates unwanted harmonics $4\omega, 6\omega, 8\omega, \ldots$. (*b*) Slotline–coplanar waveguide implementation of (*a*).

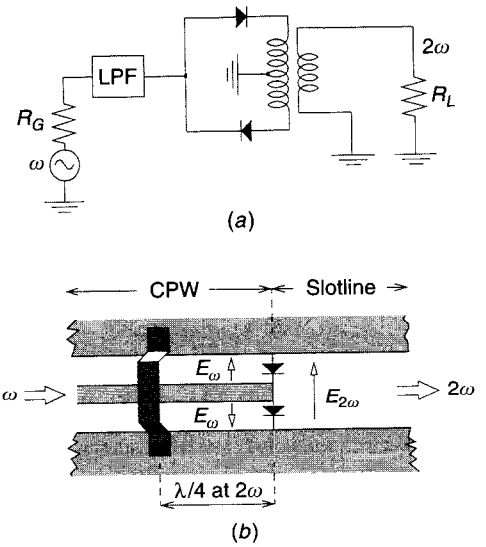


Figure 13.9 (*a*) Balanced resistive frequency doubler with series diodes. (*b*) Coplanar waveguide (CPW)–slotline implementation of (*a*). Open-head arrows indicate *E*-field directions at ω and 2ω. Airbridge provides backshort at 2ω.

replaced by a magic-tee, such as a slotline coplanar-waveguide junction (Fig. 13.8*b*). The symmetry provides isolation between the input and output circuits.

A variant [24] employs series diodes; see Fig. 13.9*a*. Here the low-pass filter (LPF) prevents generated harmonics from being propagated back toward the source. A realization is shown in Fig. 13.9*b*. An airbridge placed a quarter-wavelength from the diodes acts as a backshort at the second harmonic. It also minimizes the fourth harmonic, at which it is a half-wavelength away from the diodes.

Exponential Diode Doubler Theory. The analysis of the balanced resistive doubler is similar in the parallel-diode and series-diode cases. Here, the parallel-diode case is investigated. The magic-tee (180° hybrid junction) is represented by a transformer, as in Fig. 13.8*a*. It is assumed that the transformer is ideal, with each output voltage equal to the input voltage v_{in}, that the Schottky diodes have identical characteristics described by (13.5), and that the diode parasitics C_j and R_j, which do not contribute to resistive doubling, are negligible.

Using normalizations $x = \Lambda v_{\text{in}}$, $y = \Lambda v_{\text{out}}$, $a = \Lambda R_L I_s$, the doubler transfer function is

$$y = 2a(e^{-y} \cosh x - 1) \tag{13.6}$$

The *even* function $\cosh(x)$ is the origin of even-order harmonics. Assuming that

$v_{in} \cong V_{in} \cos(\omega t)$ and that $V_{in} < nkT/q_e$, the output is approximately (Problem 13.2)

$$\begin{aligned} y \cong\ & 2a[\mathbb{I}_0(X) - 1] && \text{(DC-rectified term)} \\ & - 4a\mathbb{I}_1(X) \cos(2\omega t) && \text{(frequency-doubled term)} \\ & + 4a\mathbb{I}_2(X) \cos(4\omega t) && \text{(frequency-quadrupled term)} \end{aligned} \tag{13.7}$$

where $X = \Lambda V_{in}$, and $\mathbb{I}_n(X)$ is the nth order modified Bessel function of the first kind and argument X. Thus the doubler generates a DC output as well as even harmonics. An LPF selects the second harmonic. The appearance of multifrequency diode voltages precludes a large-signal analysis in terms of modified Bessel functions, as was done for detectors in Chapter 11. Nevertheless, insight can be obtained by plotting the inverse of the voltage transfer function (13.6); see Fig. 13.10. Both positive and negative peaks of v_{in} cause positive peaks of v_{out}, leading to doubling action. Realistic values of x indicate an almost linear relationship between x and y once the diodes are turned on, so that one can use a piecewise linear (PWL) approximation in which each diode is modeled as a switch in series with a fixed voltage V, to calculate the doubler input and output powers and hence efficiency.

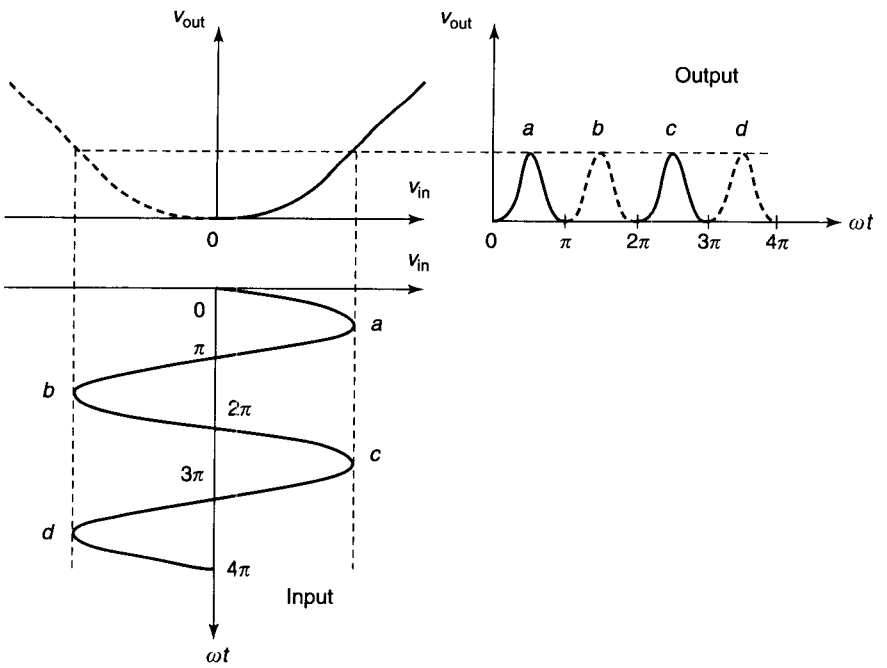


Figure 13.10 Mechanism of balanced resistive doubler.

Piecewise-Linear Doubler Theory. The topology of Fig. 13.8*a* is analyzed, and parasitics are neglected. The diode model is

$$\begin{aligned} v &= V \qquad i \geq 0 \\ i &= 0 \qquad v < V \end{aligned} \tag{13.8}$$

where V is a constant voltage ~1 V. Again assuming that R_s is negligible, that $R_G \ll R_L$, and that $v_{\text{in}} \cong V_{\text{in}} \cos(\omega t)$, the input power is (Problem 13.3)

$$P_{\text{in}} = \frac{V_{\text{in}}^2}{\pi R_L}[\arccos(b) - b\sqrt{1-b^2}] \tag{13.9}$$

where $b = V/V_{\text{in}}$, and the conversion efficiency is

$$\eta_c = \frac{P_{2\omega}}{P_{\text{in}}} = \frac{8}{9\pi} \frac{(1-b^2)^3}{\arccos(b) - b\sqrt{1-b^2}} \tag{13.10}$$

Figure 13.11*a* shows the conversion loss $L_c = 10 \log_{10}(\eta_c)$ versus P_{in}. The normalization power $P_{\text{norm}} = V^2/(2R_L)$ is taken to be +15 dBm. The minimum L_c is 6.62 dB, compared with the theoretical minimum possible 6.02 dB. The PWL analysis agrees reasonably well with reality. In the real doubler, after reaching a minimum, L_c increases with input power. This is due to diode-current saturation, not modeled in the PWL analysis. The wideband frequency response is given in Fig. 13.11*b*.

Another implementation [24] used the CPW–slotline circuit of Fig. 13.9*b* to make a low-spurious (<−40 dBc) wideband (20–40 GHz) doubler for network analyzer and synthesized sweeper applications.

Resistive Triplers. To estimate the ultimate performance of resistive triplers, the theoretical efficiency of a Schottky barrier diode tripler to 900 GHz was investigated [25]; see Fig. 13.12. To suppress reactive nonlinearity, the depletion capacitance (Chapter 8) was assumed voltage invariant, as in a Mott diode [1]. The maximum efficiency was ~7%, compared with 11.1% for an ideal tripler.

For generalized *odd-order* harmonic generation, the optimum nonlinear resistor, which gives an efficiency of $1/N^2$, has the v–i characteristic of Fig. 13.13*a*: It is an ideal limiter [22]. This conclusion is relevant to the operation of active multipliers (Section 13.2.4). In practice, such a characteristic is approximated using an antiparallel diode pair or a planar-doped barrier (PDB) diode; see Chapter 11. Using a PWL method, Clay [22] found that the Z tripler* of

*Analogously to "Z mixers" (Chapter 11), in a "Z multiplier" the nonlinear element resides in the shunt branch common to the input and output circuits. In a "Y multiplier" however, the nonlinear device is placed in a series branch between the input and output circuits, not a good location for heat-sinking purposes.

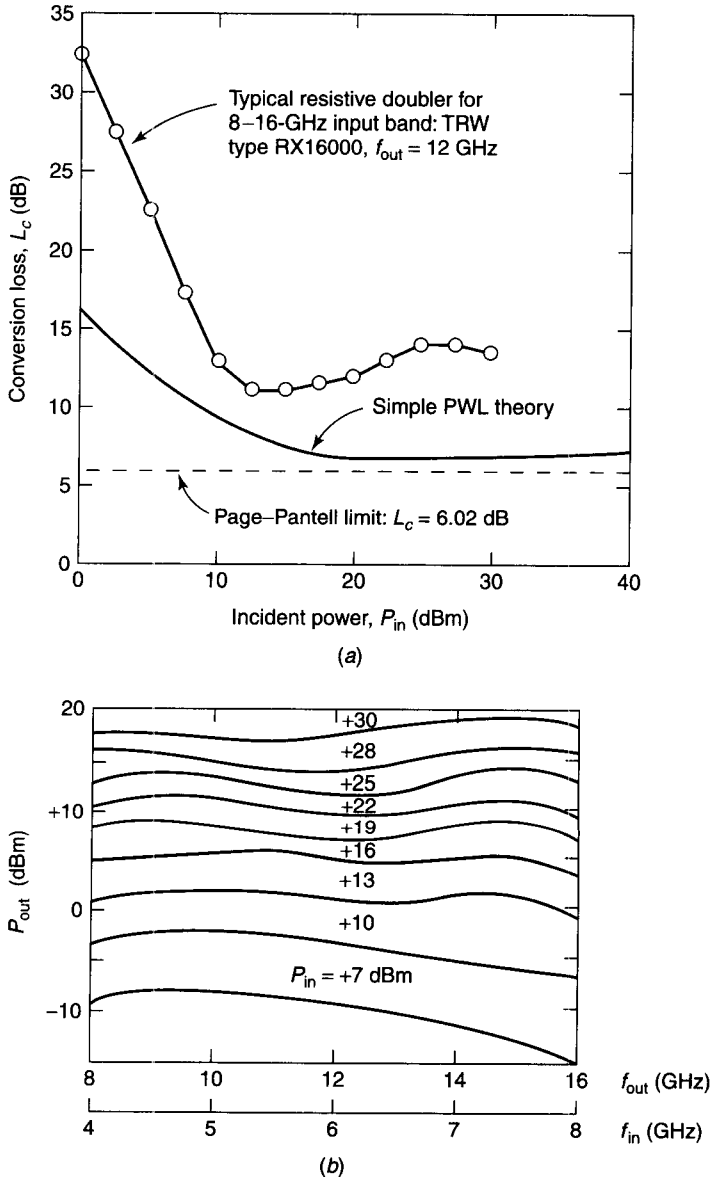


Figure 13.11 (*a*) Theoretical conversion loss L_c of balanced resistive doubler (—) predicted by simple PWL theory, compared with performance of typical commercial unit (–o–), and absolute minimum L_c for a doubler (- - -). (*b*) Typical wideband frequency response of Schottky diode resistive doubler (type RX 16000) for range of values of P_{in}. (Courtesy of TRW Microwave Inc.).

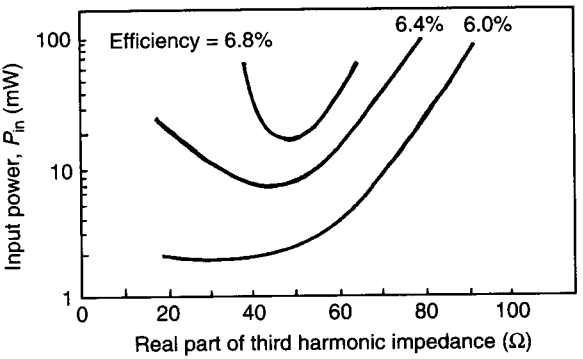


Figure 13.12 Contours of constant efficiency for tripler to 900 GHz, calculated as function of input power P_{in} and real part of third-harmonic embedding impedance. Bias current I_{dc} is chosen at each point to optimize efficiency. Assumed Schottky diode parameters are $R_s = 10\ \Omega$, $I_s = 10^{-6}$ A, $C_j = 1.0$ fF, $V_{bi} = 1.0$ V, $\gamma = 0$, and $n = 1.2$. (After Benson and Frerking [25]. Reprinted with permission of IEEE.)

Fig. 13.13*a* has an efficiency agreeing with the Page–Pantell limit. Rauschenbach et al. [26] analyzed a reflective tripler utilizing an antiparallel pair of ideal exponential diodes and found a peak efficiency ~6.7%, consistent with Fig. 13.12.

The antiparallel diode configuration of Fig. 13.13*b* was used in a broadband tripler [27] for the 25–36.67-GHz input band (75–110 GHz output). In this application a nine-element Chebyshev LPF matched the diodes over the input band; the 45-GHz filter cutoff frequency controlled harmonic generation.

13.2.3 Nonlinear-Reactance Multipliers

Radio-frequency receivers need local oscillator (LO) sources. Three-terminal active-device oscillators are available at low frequencies, and Gunn oscillators up to ~150 GHz. Harmonic multipliers are the main sources at higher frequencies. There are several types of nonlinear-reactance (parametric) multipliers:

- Early parametric multipliers used *pn*-junction varactors and had limited frequency capability due to diffusion capacitance effects.
- Exploitation of the diffusion capacitance led to the development of SRDs suitable for spectral-comb generation and high-order multiplication (without idler circuits) up to the low-GHz range. Step recovery diode multipliers are discussed in Chapter 8.
- Schottky varactors are majority-carrier devices free from diffusion-capacitance limitations. They are useful in narrow-band low-noise multipliers, at frequencies into the terahertz range, and at powers up to ~100 mW for a single device.* A problem (shared by *pn*-junction varactors) is that high-order ($N > 2$) multipliers tend to require idler circuits (causing

*The quoted frequencies and powers have not been obtained simultaneously.

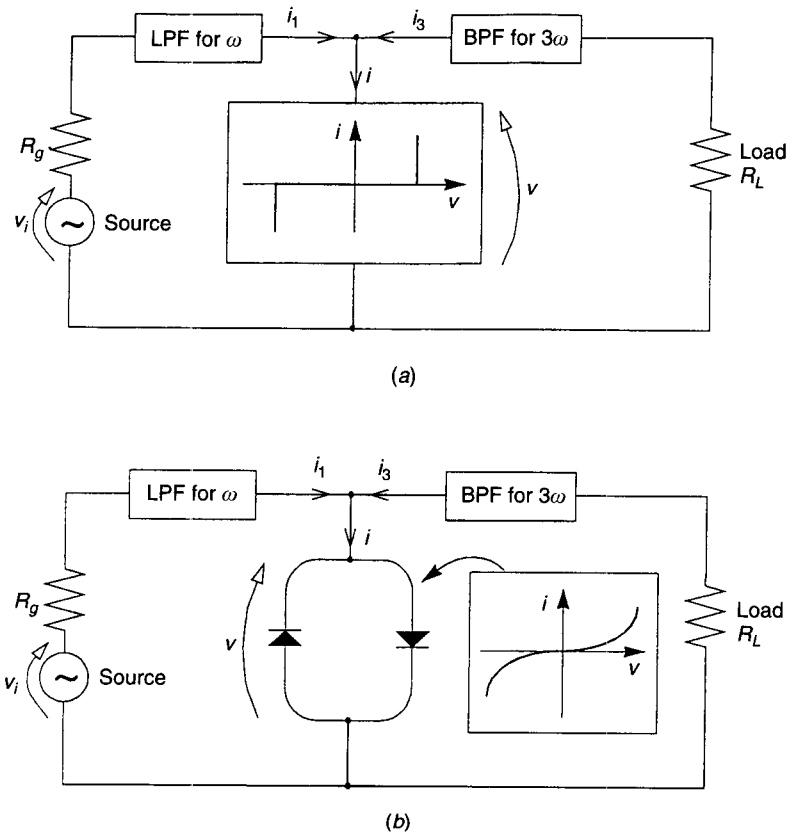


Figure 13.13 (*a*) Idealized resistive tripler circuit. (*b*) Practical approximation to (*a*) using antiparallel diode pair (or PDB diode) for which $i = 2I_s \sinh(\Lambda v)$.

extra loss). For certain devices and configurations idlers are essential. Low-loss Schottky varactors at cryogenic temperatures are the devices of choice for low-noise frequency multipliers, for example as LO sources for radio astronomy.

- Idlers are unnecessary in triplers using symmetrical varactors. The most promising such device is the HBV [10, 28]; others such as the back-to-back barrier-i–n^+ varactor [29] and the dual-anode high-electron-mobility varactor [30] have been investigated. Whereas a fundamental source at 60 GHz for wireless communications might be expensive, a 20-GHz Gunn oscillator followed by an efficient HBV tripler could provide an inexpensive solution. Symmetrical varactors can also be used in higher–odd-order multipliers such as quintuplers.
- When *efficient* wide-band multipliers are needed, nonlinear-transmission-line (NLTL) multipliers employing transmission lines periodically loaded with varactors are an attractive alternative [17].

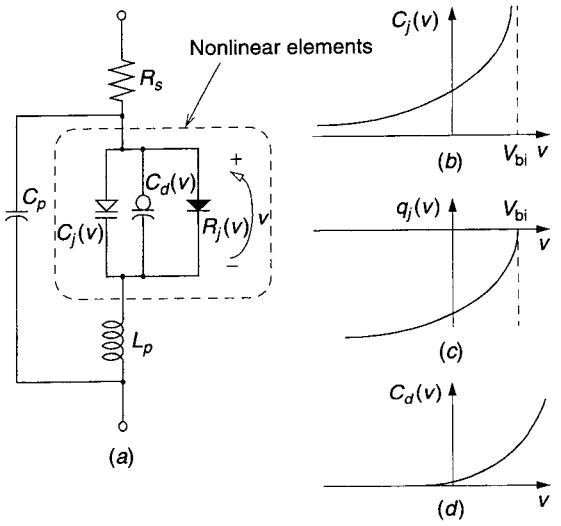


Figure 13.14 (*a*) Simplified nonlinear equivalent-circuit model applicable to *pn*-junction, Schottky barrier, and step recovery varactors: C_p and L_p are parasitics; breakdown effects not included. (*b*) Depletion capacitance characteristic $C_j(v)$. (*c*) Corresponding depletion charge characteristic $q_j(v)$ (parabolic when $\gamma = \frac{1}{2}$). (*d*) Exponential charge storage characteristic $q_d(v)$, modeled by $C_d(v)$.

Manley–Rowe Equations for Parametric Multipliers. The Manley–Rowe equations [8] show that for parametric nonlinear-reactance multipliers, the power P_N generated at the Nth harmonic of the input frequency f_{in} is related to the power P_1 absorbed at f_{in} by $P_N/P_1 \leq 1$, that is, the maximum possible efficiency is 100%.

Classical Varactor Multipliers. The simplified nonlinear equivalent circuit of Fig. 13.14*a* can model *pn* junction and Schottky barrier varactors, as well as SRDs, and is similar to the diode models used in simulators such as Spice. For analysis, the large-signal behaviors of the depletion capacitance $C_j(v)$ and diffusion capacitance $C_d(v)$ are appropriately modeled by the junction charge

$$q_j(v) = C_j(0)V_{bi}(\gamma - 1)^{-1}\left(1 - \frac{v}{V_{bi}}\right)^{1-\gamma} \tag{13.11}$$

and the diffusion charge

$$q_d(v) = \tau_L I_s[\exp(\Lambda v) - 1] \tag{13.12}$$

respectively, where $C_j(0)$ = depletion capacitance at zero bias
V_{bi} = built-in voltage

γ = capacitance-law exponent
τ_L = average lifetime of injected minority carriers
I_s = diode reverse saturation current
$\Lambda = q_e/(nkT)$
q_e = electron charge magnitude
n = diode ideality factor
k = Boltzmann constant (1.38×10^{-23} J/K)
T = temperature in kelvins

An alternative expression $q_d = \tau_L i$ can be found from (13.5) and (13.12). The nonlinear components of the equivalent-circuit model of Fig. 13.14*a* are $R_j(v)$, representing the varactor conduction current defined by (13.5); $C_j(v)$, representing the depletion capacitance $\partial q_j/\partial v$, storing charge $q_j(v)$ [see (13.11)]; and $C_d(v)$, representing the diffusion charge (absent in Schottky varactors) [see (13.12)]. The linear components are R_s, the series resistance of the diode, L_p the parasitic inductance, and C_p the parasitic capacitance. Classical varactor multiplier theory [5, 31, 32] is based on three assumptions:

1. The varactor is "fully pumped," implying full drive from reverse breakdown to zero volts but *not into reverse conduction or forward bias.*
2. The varactor obeys the "abrupt-junction" charge model* with $\gamma = \frac{1}{2}$, yielding the useful *parabolic* voltage–charge characteristic:

$$v(q_j) = V_{\text{bi}}\left[1 - \left(\frac{q_j}{q_j(0)}\right)^2\right] \tag{13.13}$$

   where $q_j(0) = -2C_j(0)V_{\text{bi}}$ is the depletion charge at zero bias.
3. The only parasitic of importance is R_s.

Assumption 1 is appropriate for low-noise frequency multipliers where shot noise due to conduction currents must be avoided, but not for SRDs where forward conduction is an essential feature of the operation.

Under *fully pumped* conditions one can define a large-signal cutoff frequency

$$\mathbf{F_c} \triangleq \frac{1}{2\pi R_s}\left\{\frac{1}{C_j(V_{\text{BR}})} - \frac{1}{C_j(0)}\right\} \tag{13.14}$$

where V_{BR} is the reverse breakdown voltage. Here, $\mathbf{F_c}$ is *not* the same as the small-signal cutoff frequency f_c quoted in data sheets:

$$f_c(V_{\text{dc}}) = \frac{1}{2\pi R_s C_j(V_{\text{dc}})} \tag{13.15}$$

* This model is valid only for $q_j < 0$.

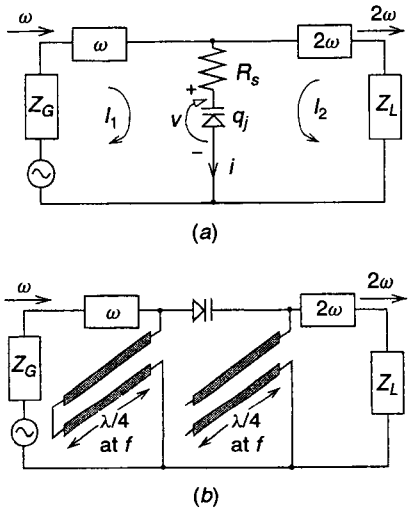


Figure 13.15 (*a*) Classical varactor shunt-mode doubler in which nonlinear reactive element is located in branch common to input and output circuits. The ω and 2ω filters are assumed to be ideal and to pass currents at indicated frequencies only. This topology is analogous to shunt Z-mixer topology of Fig. 11.16*a*. (*b*) Classical varactor series-mode doubler similar to series Z mixer in Fig. 11.16*b*. Note that matching is not included in these diagrams.

where V_{dc} is the bias voltage (typically specified as 0 or -6 V). Using (13.11), (13.14), and (13.15), the large-signal cutoff frequency can be expressed in terms of known quantities as

$$\mathbf{F_c} \triangleq f_c(0)\left[\left(1+\frac{V_B}{V_{bi}}\right)^{\gamma}-1\right] \tag{13.16}$$

For a comprehensive treatment of fully pumped classical depletion-varactor multipliers see Penfield and Rafuse [5], which includes in the varactor model only C_j and R_s and assumes a varactor embedding constraining the varactor current, and hence the time-varying charge, to the fundamental frequency f_{in}, the Nth harmonic Nf_{in}, and necessary idler frequencies when $N > 2$. Figure 13.15*a* shows an idealized shunt doubler ($N = 2$). Ideal filters restrict the diode current to only two frequency components:

$$i = I_1 \cos(\omega t) + I_2 \cos(2\omega t + \theta) \tag{13.17}$$

Tang [31] found closed-form solutions for fully pumped doublers with $\gamma = \frac{1}{2}$. This theory was extended [32] to include not only skin effect, which increases R_s at higher frequencies, but also *partial* varactor pumping, necessary in high-spectral-purity applications where the noisy breakdown region must be avoided. Numerical simulations [33] can predict the performance of depletion multipliers with arbitrary capacitance laws and drive levels.

Figure 13.15*b* shows a practical shunt doubler. Here the shorted stub ($\lambda/4$ at the input frequency f) is transparent at f, but shorts currents at $2f$. The open stub ($\lambda/2$ at $2f$) provides a short for currents at f.

Figure 13.16*a* compares the maximum possible efficiencies η_c of fully pumped "$\gamma = \frac{1}{2}$" doublers for the maximum-efficiency/maximum-output-power case [5] with those of the minimum-dissipation/fixed-input-power case investigated by Tang [31], both as functions of $f_{in}/\mathbf{F_c}$. Tang's expression is

$$\eta_c = \frac{1 - k\sqrt{2}(f_{in}/\mathbf{F_c})}{1 + 2k\sqrt{2}(f_{in}/\mathbf{F_c})} \quad \text{where } k = \sqrt{5\sqrt{5} + 11} \simeq 4.71 \tag{13.18}$$

Also shown in Fig. 13.16*a* is the low-frequency approximation [33, 34]

$$\eta_c = \exp(-2\alpha f_{in}/\mathbf{F_c}) \tag{13.19}$$

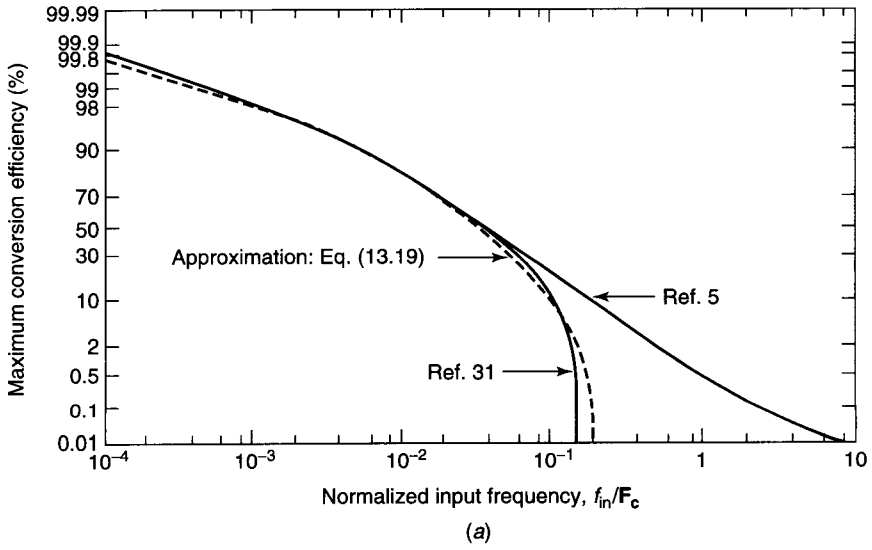
The coefficient α is $\simeq 11.1$ for a doubler with $\gamma = \frac{1}{2}$. The curve for (13.18) resembles the result of Archer's simulations [35], using the Siegel and Kerr modification of Gwarek's technique [36, 37]. Such curves demonstrate the importance of choosing a varactor with $\mathbf{F_c}$ much greater than the maximum input frequency. Figure 13.16*b* gives the optimum input and load resistances [5], with which Archer's results agree. Further design curves for the input, output, and dissipated powers, together with optimum source and load terminations and bias voltages are given in [5] for multipliers of orders $N = 2, 3, 4, 5, 6$, and 8.

Unlike nonlinear-resistance multipliers using varistors or three-terminal active devices, or those employing SRDs, efficient depletion-varactor Z multipliers require input and output tuning, as well as tuned idler resonators (for $N > 2$), and are consequently narrow-band circuits. High efficiency also requires the input, output, and idler circuits to be coupled only via the nonlinear reactance, and input and output circuits to be conjugately matched to the dynamic (bias dependent) varactor impedances at the respective input and output frequencies. A further requirement for efficiency is that, ideally, varactor currents flow only at the input, output, and idler frequencies [34].

For treatments of AM-to-AM and AM-to-PM conversion, additive noise, and hysteresis phenomena in "$\gamma = \frac{1}{2}$" multipliers, see Bava et al. [38, 39].

Designing Classical Varactor Frequency Multipliers. The following considerations affect the performance of classical varactor parametric multipliers [40]:

1. At low input powers, multiplier performance is mainly determined by the DC bias point and the embedding impedances presented to the diode(s) at the input frequency and the harmonic output frequency. Good impedance matching should be maintained over the desired operating band. The output power and conversion efficiency can be optimized by tweaking the bias point. Simulations tend to overestimate the output power and efficiency.



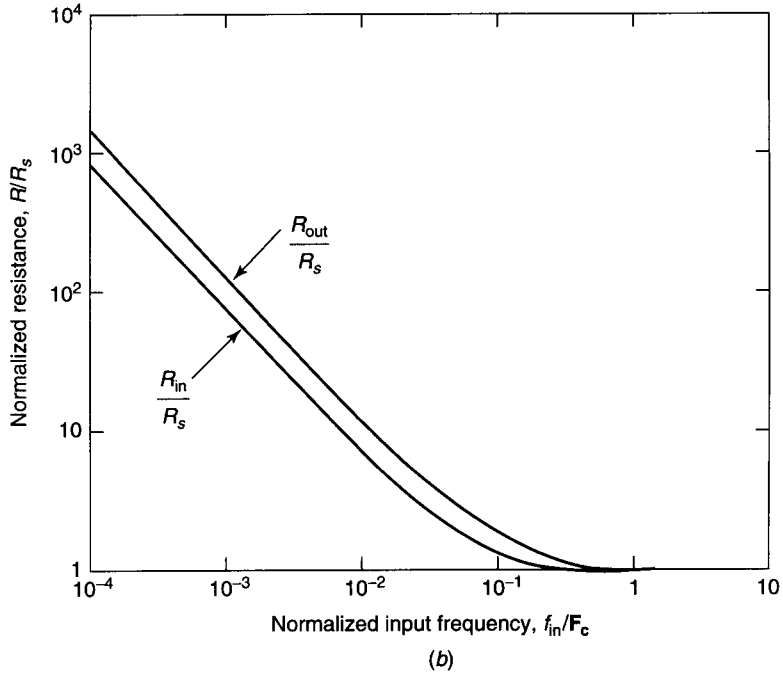


Figure 13.16 (*a*) Maximum efficiency of classical fully pumped depletion varactor doublers with $\gamma = \frac{1}{2}$, as function of input frequency f_{in}. (*b*) Input and load resistances R_{in} and R_{out} for classical fully pumped depletion varactor doublers with $\gamma = \frac{1}{2}$, for maximum efficiency. These resistances, which are the real parts of the diode impedances at the two frequencies, need to be matched to the source and load impedances. (After Penfield and Rafuse [5]. Reprinted with permission of MIT Press.)

2. At maximum power levels, the operation is also affected by the diode breakdown voltage V_{BR}, series resistance R_s, current saturation mechanisms [18], and impact ionization [40], all of which can affect conversion efficiency η_c. These problems are usually alleviated by using stacked varactors [41, 42]. Since the power capability of varactor multipliers is proportional to V_{BR}, connecting n diodes in series increases the overall breakdown voltage to nV_{BR} and the output power by a factor of n^2. To maintain overall values of capacitance and series resistance, the active area per diode has to be increased by a factor n.
3. The best efficiency is obtained in the transition region between low and high power levels.
4. Low-input-frequency multipliers should be optimized for maximum output power since input power is cheap.
5. High-input-frequency multipliers should be designed for maximum efficiency since input power is expensive. This requires optimization of both the devices and circuits.

Microwave Varactor Multipliers. Multipliers for microwave outputs (1–30 GHz) are seldom specified in new designs because sources such as Gunn oscillators are widely available in this range. When such multipliers *are* used, high power operation mandates a shunt-mode configuration [43, 44] for good heat sinking; see Fig. 13.15*a*. At low power levels heat sinking is not a problem and series multipliers (Fig. 13.15*b*) can be used. An example is a quadrupler to 8.5 GHz designed [45] according to the Burckhardt theory [33], including a second-harmonic idler, with efficiency of ~38%. The series mode is advantageous for "$N > 2$" multipliers since the $q_j(v)$ nonlinearity is a *square-root* function (generating harmonics above the second) so that idlers are less essential than in the shunt mode where the $v(q_j)$ function is *parabolic* [in which case a varactor with $\gamma = \frac{1}{2}$ generates no harmonics above the second, and therefore requires idler(s) if $N > 2$]. A further advantage of the series mode in planar designs is that a via is not needed to ground the varactor.

Millimeter-Wave Varactor Multipliers are very much a current concern since communications systems are now migrating to higher frequencies where congestion is less. Until the late 1980s millimeter-wave multipliers used whisker-contacted multiple-anode Schottky varactors in crossed-waveguide mounts [46]. High-Q waveguide circuits can incorporate adjustable tuners and backshorts but require expensive machining, are hard to adjust, and are unsuitable for mass production. These problems are exacerbated as frequencies increase, and MMIC designs become essential. Developments in planar Schottky diodes and circuits now make true MMIC millimeter-wave designs feasible, albeit with lower efficiency than in waveguide.

Narrow-Band High-Efficiency Example. A millimeter-wave classical varactor doubler, with both diode and circuit designed from first principles, is described

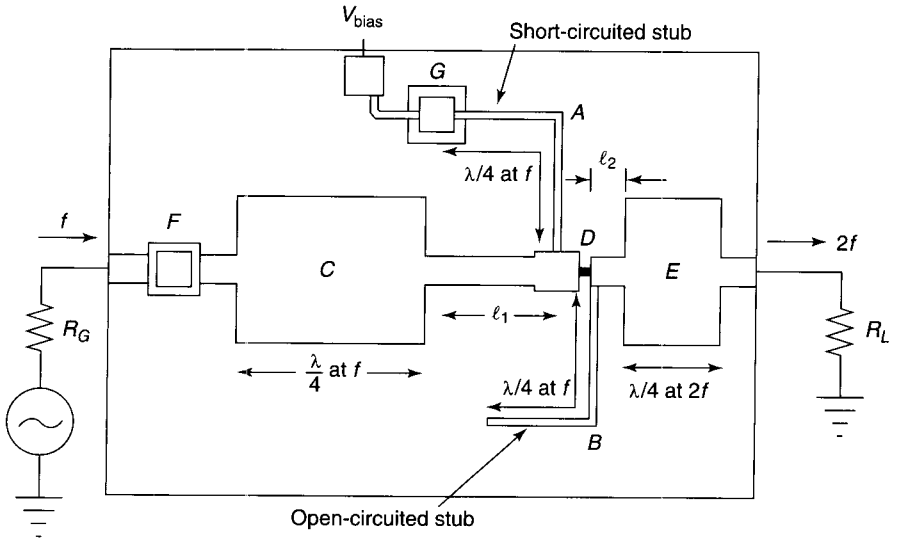
in [47]. To get a high cutoff frequency, the Schottky varactor has a small, thin, heavily doped epilayer. This in turn sets the breakdown voltage V_{BR} and the maximum current capability, and hence the maximum power handling level. For a low R_s the thickness of the n^+ layer is made greater than its skin depth. The doubler is to have a 94-GHz output with an input power of 350 mW (25.4 dBm) at 47 GHz. The conversion efficiency η_c and output power P_{out} are to be maximized. Measured parameters of the optimized diode are: series resistance $R_s = 0.82\ \Omega$, junction capacitance $C_j(0) = 0.17$ pF, breakdown voltage $V_{BR} = 19$ V, and cutoff frequency $f_c(0) = 1140$ GHz. The circuit of Fig. 13.17*a* is a series doubler (see Fig. 13.15*b*) with input and output matching networks. To maximize η_c and P_{out}, the effective diode capacitances are resonated with inductive impedances at the input and output frequencies f and $2f$, respectively. The 50-Ω lines of lengths l_1 and l_2 are used as the inductors. Then $\lambda/4$ impedance transformers transform the external 50-Ω impedances to the optimum diode resistances R_{in} and R_{out}; see Fig. 13.16*b*. It is required that the input and output circuits be isolated from each other, except via the diode, simultaneously with input and output tuning. In Fig. 13.17*a*:

- The short-circuited input stub A is $\lambda/4$ at frequency $f = 47$ GHz, but is $\lambda/2$ at $2f = 94$ GHz, thereby creating a reflective short for currents at $2f$, without affecting currents at f.
- The open-circuited output stub B is also $\lambda/4$ at f, behaving as a reflective short for currents at the input frequency, but as a $\lambda/2$ section at $2f$, so that it does not peturb the output signal.
- A 50-Ω section of length l_1 resonates C_j at f. Input matching is completed by the matching transformer C, ($\lambda/4$ at f) of impedance $\sqrt{R_G R_{in}}$.
- A short 50-Ω section of length ℓ_2 resonates C_j at $2f$, while the $\lambda/4$ (at $2f$) matching transformer E, of impedance $\sqrt{R_{out} R_L}$, provides the output match.
- The remaining components are the DC blocking capacitor F and the RF-grounding capacitor G.

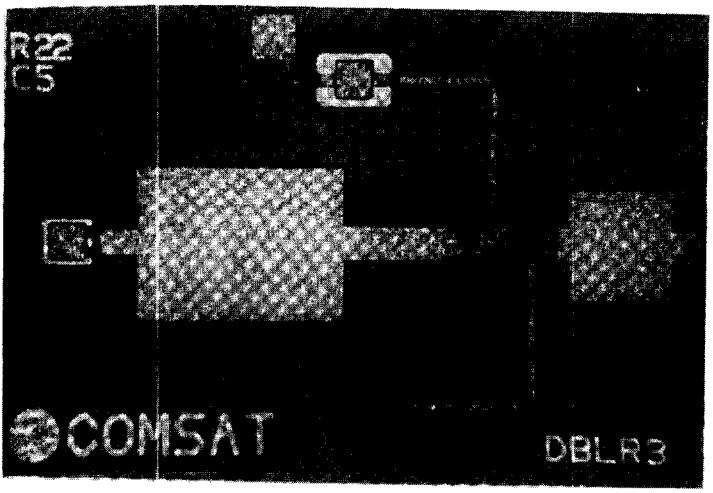
As shown in Fig. 13.18, under the originally specified conditions, this narrow-bandwidth 94-GHz doubler produces a peak η_c of 19.7% at 18.1 dBm output, and a $\eta_c = 25\%$ at 17.4 dBm output.

Wide-Bandwidth Lower Efficiency Example. A doubler for the 70–80-GHz output band was designed in finite-ground coplanar waveguide FCPW technology* by Papapolymerou et al. [48]. The airbridge Schottky diodes (Fig. 13.19), which were fabricated during the processing of the 500-μm-thick vapor-

*The advantage of FCPW [51] over CPW technology is that vias are not needed to suppress moding. The coupled-slotline mode is suppressed by (a) symmetry and (b) airbridges at discontinuities. The ground strip must be narrow enough to avoid moding, but wide enough to reduce loss. FCPW lines are equivalent to lossy TEM lines.



(*a*)



(*b*)

Figure 13.17 A 94-GHz MMIC single-diode frequency doubler. (*a*) Identification of circuit components on vapor-phase-epitaxial (VPE) substrate. (*b*) Micrograph of doubler. The 90-μm-thick GaAs VPE chip is 1.0 × 1.5 mm. (After Chen et al. [47]. Reprinted with permission of IEEE.)

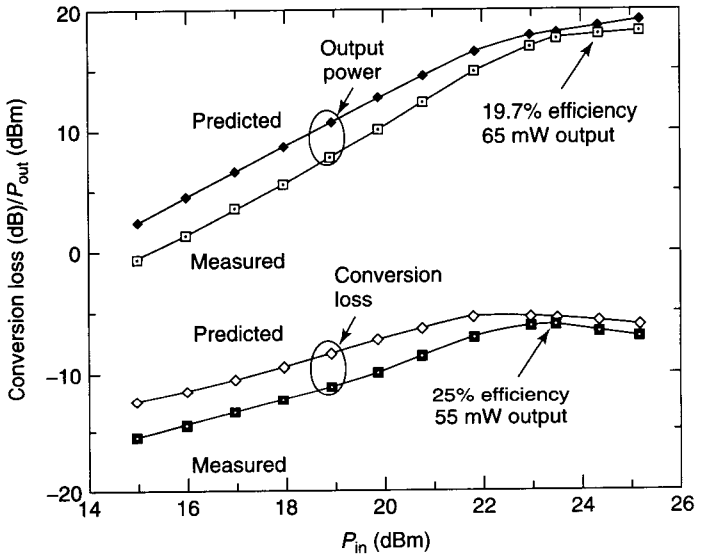


Figure 13.18 Performance of 94-GHz MMIC doubler at $V_{bias} = -7.0$ V. (After Chen et al. [47]. Reprinted with permission of IEEE.)

phase epitaxy (VPE) GaAs substrate, have $R_s = 2.2\ \Omega$, $C_j(0) = 88$ fF, $V_{BR} = 11.8$ V, $f_c(0) = 822$ GHz and impedance $51.5 - j106\ \Omega$ (at 40 GHz, $V_{bias} = -3.5$ V). The shunt-mode circuit is similar to Fig. 13.15*a*. In the equivalent circuit, Fig. 13.20*a*, A and H are 50-Ω sections. Open-circuited stub B is a second-harmonic trap, $\sim\lambda/4$ long at $2f$. Lines C and E are 50-Ω matching

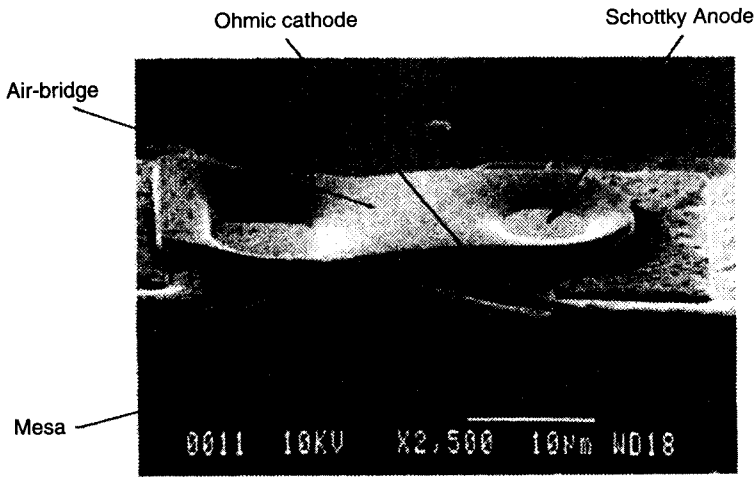


Figure 13.19 Scanning electron microscopy (SEM) photograph of airbridge Schottky barrier diode. (After Papapolymerou et al. [48]. Reprinted with permission of IEEE.)

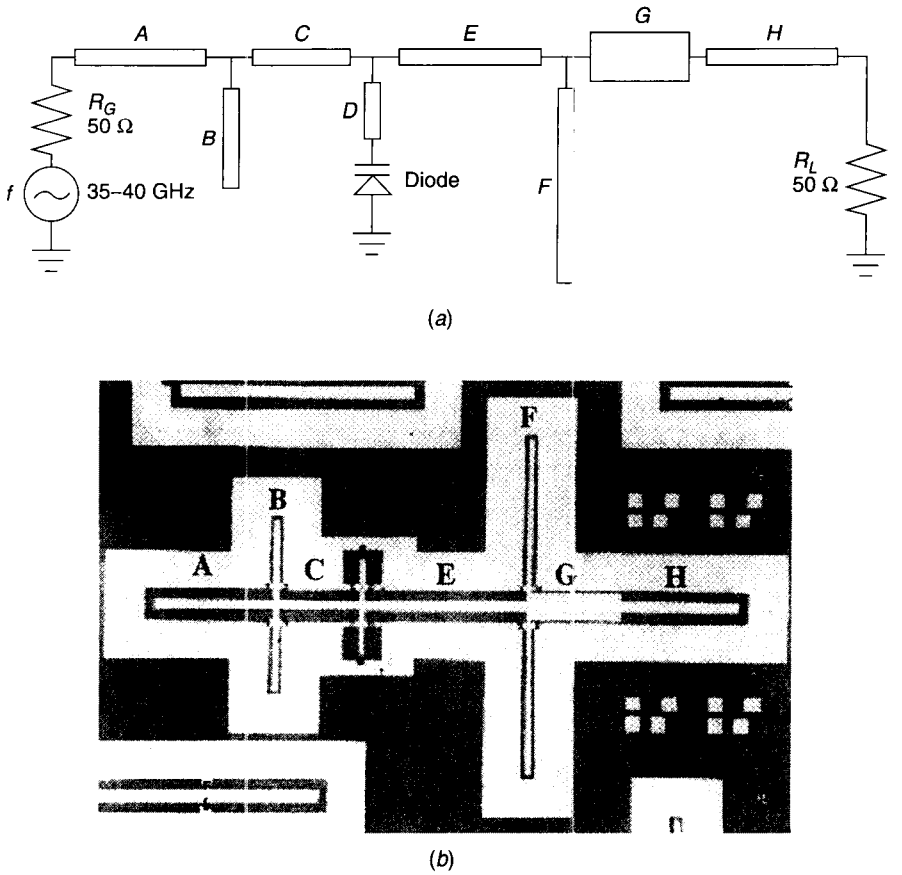


Figure 13.20 (*a*) Equivalent circuit of shunt-mode MMIC doubler for 70- to 80-GHz output band. (*b*) Micrograph of circuit, implemented in FCPW technology. Airbridges are used at the stub corners to suppress the coupled-slotline mode. Size of 500-μm-thick GaAs substrate is ~2 × 3 mm. (After Papapolymerou et al. [48]. Reprinted with permission of IEEE.)

sections; the average diode capacitance $\langle C_j \rangle$ is canceled out by the high-impedance inductive line D. Open-circuited stub F on the output side is a fundamental trap, $\sim\lambda/4$ long at f. Low-impedance line G is another matching section. In the actual circuit, Fig. 13.20*b*, the diodes and stubs are replicated to maintain FCPW symmetry. The design was refined using East's multiple-reflection program [49]. Figure 13.21 shows the wideband performance: for a 20-dBm input, η_c remains above 10% ($L_c < 10$ dB) over the 70- to 80-GHz output band. At the 80-GHz output frequency, η_c increases to 18% at an input of 27 dBm.

Step Recovery Diode Multipliers. During the initial development of *pn*-junction parametric multipliers in the 1960s, it was found that unexpectedly

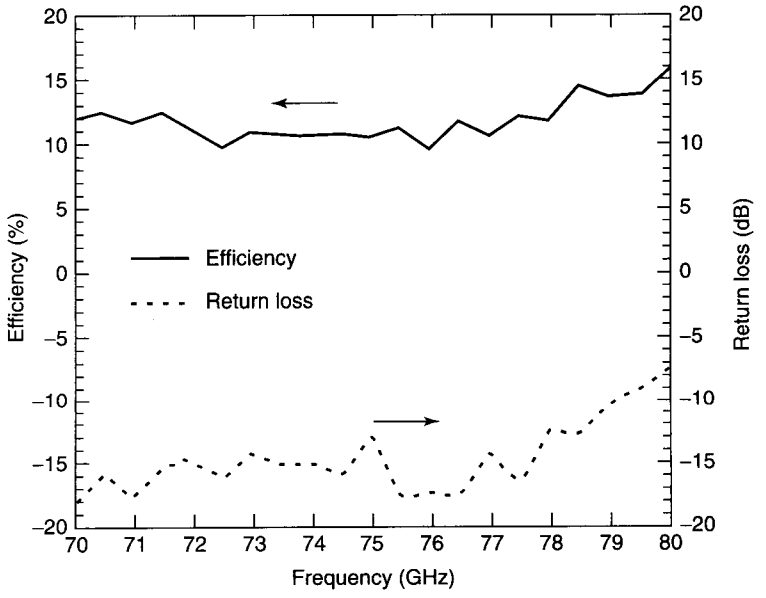


Figure 13.21 (*a*) Efficiency η_c and return loss of FCPW doubler vs. output frequency. Input level is +20 dBm. (*b*) Efficiency and output power vs. input power, at 80-GHz output, and $V_{bias} = -3.5$ V. (After Papapolymerou et al. [48]. Reprinted with permission of IEEE.)

high power handling [50] and high η_c could be achieved by driving the varactor into forward bias for part of the cycle. This could not be explained by the existing "depletion-capacitance" theories [5]; moreover the optimum circuit and maximum accumulated junction charge were quite different from predictions. These effects were found to be due to minority-carrier charge storage (low-frequency manifestation: *diffusion capacitance*) in *pn*-junction varactors. The result was the emergence of a separate class of varactor device: the SRD. If the diode is alternately forward and reverse biased, such as by a sinusoidal voltage, then under appropriate conditions there will be (a) storage of injected minority-carrier diffusion charge q_d during the forward-bias regime and (b) withdrawal of q_d during the reverse-bias phase, leading to a reverse conduction current that ceases "instantly" as soon as all of the stored q_d has been removed, so that the device suddenly approximates an open circuit. The resulting picosecond-scale step recovery transient generates fast time-domain pulses or, equivalently, a comblike spectrum of high-order harmonics. Hence the appellation SRD for varactors designed to exploit this behavior. Other names include *diffusion varactor, charge storage varactor*, and *snapback diode*. In SRDs the diffusion capacitance is dominant, and the depletion capacitance is nearly constant under reverse bias. With the SRD, it is possible to make high-order harmonic generators, up to the 20th harmonic or more, with outputs up to ~50 GHz. The mechanism of such multipliers (see Chapter 8) is quite different from that of depletion-varactor multipliers. Charge storage is enhanced by maximizing the average minority carrier lifetime τ_L. In many applications idler circuits are unnecessary, simplifying circuit design and improving bandwidth (at the expense of efficiency). The SRD is very similar to the *pin* diode; the differences are in the application and dimensioning of the device.

Symmetrical Varactor Multipliers. Early multiplier designs were restricted to the nonlinear $C_j(v)$ characteristics of *pn*-junction and Schottky barrier diodes. Now molecular-beam epitaxy (MBE) and metallorganic chemical vapor deposition (MOCVD) allow the precise fabrication of structures such as the heterostructure-barrier varactor* (HBV) [52], and $C_j(v)$ characteristics can be optimized for specific applications. A symmetric-$C_j(v)$ varactor requires no bias and generates only odd harmonics. Fewer idlers are needed than for classical multipliers, none for tripling. The HBV offers more design options than classical varactors since its semiconductor layer compositions, doping profiles, barrier thickness, barrier composition, number of barriers, device geometry, and area can all be varied [53]. Other symmetrical devices such as the *high-electron-mobility varactor* [54, 55], the *barrier-intrinsic-n^+* varactor [56, 57], and the *back-to-back barrier-n-n^+* varactor [58] have also been described.

Heterostructure Barrier Varactor. The HBV consists of a large-bandgap semiconductor sandwiched between symmetrical smaller-bandgap depletion

* Originally known as the *quantum barrier varactor* [52] and later as the *single-barrier varactor*.

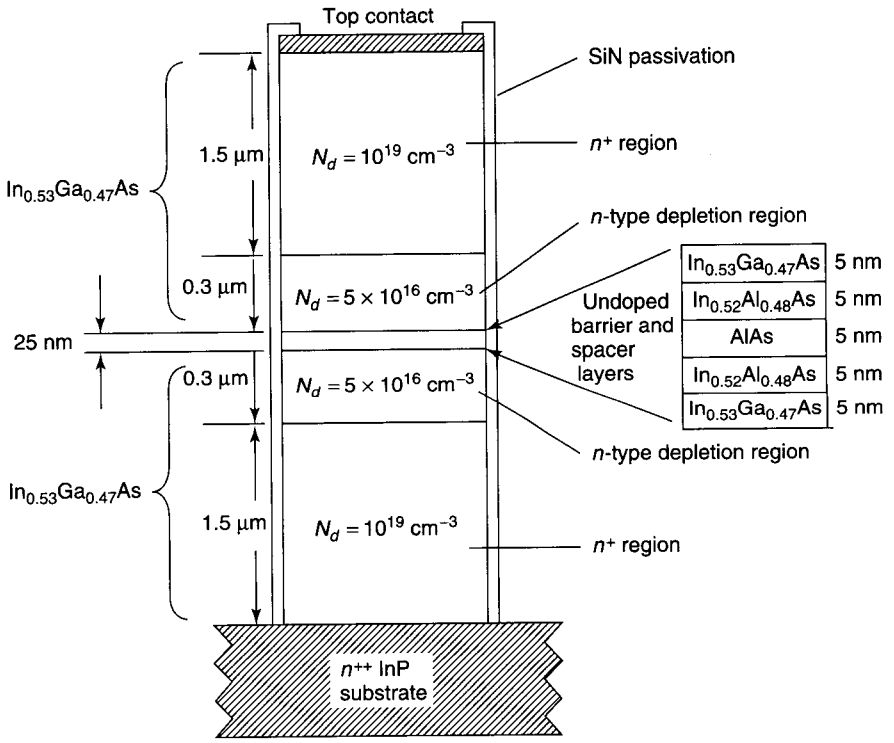


Figure 13.22 Structure of typical single-barrier heterostructure barrier varactor [63] (not to scale). This device is fabricated on a highly doped InP substrate. A practical device requires a passivating layer, for example, SiN, to protect the exposed layers from atmospheric contamination. The vertical scale is greatly exaggerated.

regions. The voltage-dependent thickness of the depletion layers within these regions is the origin of the symmetric $C_j(v)$ characteristic. To fabricate an HBV, a sequence of dissimilar "semiconductor alloys" is deposited on a substrate, usually by means of MBE. Figure 13.22 shows an example grown on an n^{++} InP substrate. The essential features are (a) the central blocking *barrier*, here a 5-nm undoped AlAs layer, to inhibit conduction current; (b) two *n*-type *depletion regions*; (c) two n^+ regions, highly doped to minimize R_s; (d) a top metal contact; and (e) undoped *spacer* layers, here $In_{0.53}Ga_{0.47}As$ and $In_{0.52}Al_{0.48}As$.

Symmetrical Varactor Model. In the model of Fig. 13.23*a* $C_j(v)$, denoted by the "symmetrical-varactor" symbol, stores charge q_j. As seen in Fig. 13.23*b*, the geometrical symmetry of the device leads to electrical symmetry of the $C_j(v)$ characteristic. The maximum capacitance $C_j(0)$ and the maximum nonlinearity both occur at zero bias, a feature leading to circuit simplification. The $q_j(v)$ characteristic, Fig. 13.23*c*, is antisymmetric. The undesired conduction

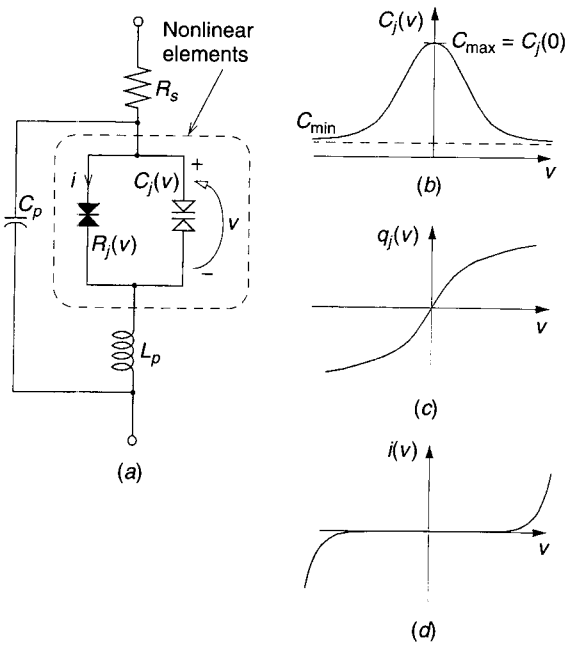


Figure 13.23 (*a*) Simplified nonlinear equivalent-circuit model for symmetrical varactors: C_p and L_p are parasitic elements; breakdown effects not included. Compare with Fig. 13.14*a*. (*b*) Symmetric depletion capacitance characteristic $C_j(v) = \partial q_j/\partial v$ (neglects C_p). (*c*) Corresponding antisymmetric $q_j(v)$ characteristic (empirically a cubic). (*d*) Antisymmetric $i(v)$ characteristic representing undesired leakage current, modeled by $R_j(v)$.

current* $i(v)$ (Fig. 13.23*d*), due to indirect transitions [28], is represented by the "nonlinear resistor" $R_j(v)$. R_s accounts for skin-effect resistance, contact resistance, and the resistance of the undepleted epilayers; it is weakly voltage dependent. The parameters C_p and L_p are device parasitics. The optimum composition of the device layers is the subject of continuing research and is critical for (a) minimum leakage current $i(v)$, (b) minimum R_s, and (c) the specific C_{max}/C_{min} ratio needed to optimize multiplier efficiency.

Symmetrical-Varactor Multiplier Circuit. Figure 13.24 is a simplified model of a symmetrical shunt-mode tripler. The classical varactor theories do *not* work for symmetrical varactors like HBVs.

The antisymmetric $q_j(v)$ characteristic of Fig. 13.23*c* can be approximated† by the empirical "cubic" charge model‡ [59]:

*Early HBVs had excessive leakage (conduction) current, but high-Q HBVs are now possible [60].

† For a physical explanation of HBV characteristics, see Jones et al. [53].

‡ For barrier–n–n^+ or barrier–intrinsic–n^+ varactors, a raised-cosine function would be more appropriate [58].

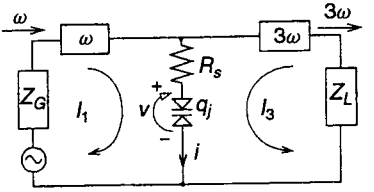


Figure 13.24 Symmetrical varactor shunt-mode tripler. The nonlinear reactive element is located in the branch common to the input and output circuits. The ω and 3ω filters are assumed to be ideal. Compare with Fig. 13.15*a*.

$$v(q_j) = V_0\left[\frac{q_j}{q_0} + \beta\left(\frac{q_j}{q_0}\right)^3\right] \tag{13.20}$$

where v is the internal voltage across the HBV (excluding R_s), q_j is the depletion region charge, and V_0, q_0, and β are constants. For curve-fitting measured data, (13.20) can be differentiated to yield an "inverse" expression for $C_j(v)$:

$$\frac{v}{V_0} = \pm\left[\frac{1}{3\beta}\left(\frac{C_j(0)}{C_j(v)} - 1\right)\right]^{1/2} \pm \beta\left[\frac{1}{3\beta}\left(\frac{C_j(0)}{C_j(v)} - 1\right)\right]^{3/2} \tag{13.21}$$

For a given peak capacitance $C_j(0)$, the parameters V_0 and β determine the shape of the $C_j(v)$ characteristic. The *measured* values of $C_{\max}$ and $C_{\min}$ will of course depend also on the parasitic capacitance C_p.

Closed-Form Large-Signal Analysis of Symmetric-Varactor Triplers. Similarly to the closed-form analysis of the classical abrupt-junction doubler [31], a closed-form analysis of the symmetric shunt-mode tripler of Fig. 13.24 is possible [28, 59]. This circuit has advantages compared with the classical tripler:

- No idler at 2ω is needed.
- The $C_j(v)$ nonlinearity is greatest at zero volts, leading to good efficiency at low input power levels. This is important at millimeter-wave frequencies where available input power is small.
- No bias is needed. This is a plus because wide-band bias circuits can be difficult to design at millimeter-wave frequencies.

The analysis uses the *cubic-charge model* (13.20) for the varactor, augmented with loss resistance R_s. In Fig. 13.24 the varactor is placed between ideal bandpass filters that pass currents at ω and 3ω only, so that the total varactor current is

$$i = I_1\cos(\omega t) + I_3\cos(3\omega t + \theta) \tag{13.22}$$

and the depletion charge is

$$q_j = \frac{I_1}{\omega}\sin(\omega t) + \frac{I_3}{3\omega}\sin(3\omega t + \theta) \tag{13.23}$$

It is assumed (a) that the input power at ω is insufficient to drive the varactor voltage v to the breakdown value V_{BR} and (b) that there is perfect q_j–v antisymmetry, so that the average charge is zero. Substituting (13.23) into (13.20), the voltage across the lossless $C_j(v)$ is found to include only the fundamental and the third, fifth, seventh, and ninth harmonics:

$$v = V_1 \cos(\omega t - \zeta_1) + V_3 \cos(3\omega - \zeta_2) + \cdots + V_9 \cos(9\omega - \zeta_2) \quad (13.24)$$

where V_n and ζ_n are amplitudes and phases to be determined. Imposing the circuit conditions and the constraints (a) that fixed power P_{in} is supplied to the nonlinear $C_j(v)$ (excluding R_s), and (b) that dissipation is minimized with respect to I_1, one finds [Problem 13.4(a)] the optimization condition

$$I_1 = (12)^{1/8}\left(\frac{P_{in}}{M \sin\theta}\right)^{1/4} = \sqrt{3} I_3 \quad (13.25)$$

where $M = \beta V_0/[4(\omega q_0)^3]$. The tripler conversion efficiency is

$$\eta_c = \frac{\text{power output}}{\text{power input}} = \frac{P_{in} - 0.5 I_3^2 R_s}{P_{in} + 0.5 I_1^2 R_s} \quad (13.26)$$

Substituting (13.25) into (13.26), the optimized efficiency is

$$\eta_c = \frac{\sqrt{6\sqrt{3}}(P_{in} M \sin\theta)^{1/2} - R_s}{\sqrt{6\sqrt{3}}(P_{in} M \sin\theta)^{1/2} + 3R_s} \quad (13.27)$$

which is maximized for a phase $\theta = 90°$. This is also the minimum-dissipation condition. The input and output impedances are found to be [Problem 13.4(b)]

$$Z_{in} = \left(\frac{V_1}{I_1}\cos\zeta_1 + R_s\right) + j\left(\frac{V_1}{I_1}\sin\zeta_1\right) \quad (13.28)$$

$$Z_L = \left(-\frac{V_3}{I_3}\cos(\zeta_3 - \theta) - R_s\right) + j\left(-\frac{V_3}{I_3}\sin(\zeta_3 - \theta)\right) \quad (13.29)$$

Interestingly, the real parts of Z_{in} and Z_L are related by $R_{in}/R_L = 1/(3\eta_c)$.

A typical design starts with a fixed input power P_{in}, the input frequency ω, and the diode parameters V_0 and β. From (13.25), I_1 and I_3 can be determined. Setting $\theta = 90°$, one can solve for the unknown amplitudes V_1, V_3 and phases ζ_1, ζ_3. The conversion efficiency follows from (13.27), and the input and load impedances from (13.28) and (13.29), respectively.

Example. Figure 13.25 plots the maximum efficiency η_c^{max} versus frequency for a symmetric-HBV tripler with $C_j(0) = 300$ fF, $V_0 = 0.455$ V, $\beta = 3.2$ (so that $C_{min} = 50$ fF at 2 V), and $R_s = 10\ \Omega$. It is compared with η_c^{max} for a classical

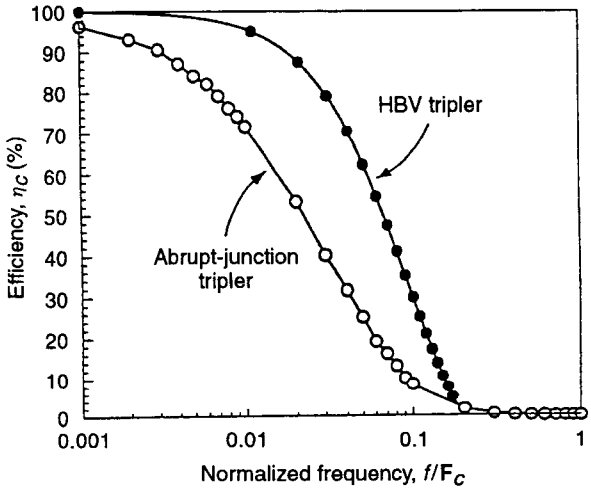


Figure 13.25 Maximum conversion efficiency of heterostructure–barrier–varactor tripler (with HBV parameters $V_0 = 0.455$ V, $\beta = 3.2$) compared with that of classical tripler (with abrupt-junction varactor parameter $\gamma = \frac{1}{2}$) using short-circuited idler [59]. The frequency axis is normalized to the varactor large-signal cutoff frequency $\mathbf{F_c}$.

tripler calculated according to [31]; see (13.18), assuming a "$\gamma = \frac{1}{2}$" varactor with $C_{\min} = 50$ fF and $R_s = 10\ \Omega$, so that its large-signal cutoff frequency is $\mathbf{F_c} \sim 1/(2\pi R_s C_{\min}) = 319$ GHz, and assuming a short-circuited idler at the second harmonic. The enhanced efficiency of the HBV tripler is evident.

Comparison with Harmonic Balance (HB) Simulation. The above simple tripler model gives results in remarkable agreement with more exact HB simulations [28] using a form of the Kerr multiple-reflection algorithm [61, 62]. Table 13.1 compares (a) data for a tripler from 10 to 30 GHz calculated from a curve-fitted cubic charge model ($V_0 = 1.02$ V, $\beta = 0.32$, $R_s = 1\ \Omega$) with (b) data derived from a HB simulation using measured $C_j(v)$ and $i(v)$ data for the same device.

Higher Power HBV Multipliers. Higher-power-HBV multipliers can be designed by replacing the single-barrier HBV with a multiple-barrier device

Table 13.1 Comparison of Cubic-Charge Model with Harmonic-Balance Simulation

Method	P_{in}	P_{out}	η_c	R_{in}	X_{in}	R_L	X_L
Harmonic balance	10.2 mW	7.4 mW	72.62%	4.67 Ω	−52.69 Ω	11.23 Ω	24.67 Ω
Cubic charge model	10.2 mW	7.3 mW	71.66%	4.70 Ω	−47.19 Ω	10.13 Ω	21.90 Ω

[63, 64]. For a stack of N barriers of a given cross-sectional area, the power handling can be increased by a factor N^2, and the capacitance decreased by a factor N. Alternatively, for a specified capacitance, the area can be increased, simplifying the fabrication process.

Idlers in Symmetrical-Varactor Multipliers. The outcome of placing a parabolic $v(q)$ device [see (13.13)] in a shunt-varactor multiplier is frequency *doubling*, no other harmonics being generated. With such a $v(q)$, idlers are essential to obtain harmonic numbers $N > 2$. The success of the shunt topology (including good heat sinking) led to misconceptions about idlers, such as "you always need idlers to multiply by any factor greater than 2." However, in the series-varactor Y-multiplier topology the parabolic $v(q)$ is replaced by a square-root law $q(v)$, producing a power series with terms of all integer orders, so (inefficient) multiplication can be done *without* idlers. Of course, the fact that SRDs could easily produce high-order harmonics, at least $\times 10$, was attributed to their being somehow "a different kind of device." Yet for many years "dual-mode" varactors, which combine the properties of both, have been widely available. A symmetrical varactor with a cubic $v(q)$ characteristic, placed in a Z multiplier naturally performs *frequency tripling* without an idler, but idlers are required for $\times 5, \times 7, \ldots$ multipliers.

Stability of Varactor Multipliers. Criteria for designing periodically driven varactor circuits with guaranteed stability are given in [65].

Example of a Practical HBV Tripler. Figure 13.26 shows a tripler from ~13 GHz to ~36 GHz realized in microstrip on a 0.127-mm Al_2O_3 substrate [28, 66]. The design is such that the input and output ports are mutually isolated from one other. Components of the circuit are:

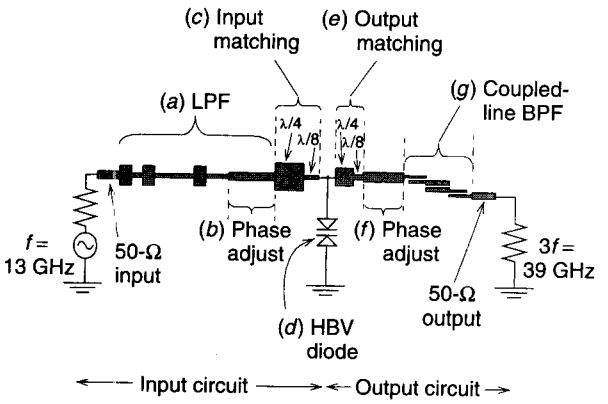


Figure 13.26 Symmetrical HBV shunt-mode tripler design realized in microstrip (Courtesy of K. Krishnamurthi [28]). (Not to scale).

(a) An input low-pass filter (LPE) to pass f_{in} but reject $3f_{in}$.
(b) A 50-Ω phase-adjust section.
(c) An input matching circuit which avoids frequency-sensitive stubs. A $\lambda/8$ line resonates the HBV to provide a real impedance (~2 Ω) that is transformed by a low-impedance $\lambda/4$ line to 50 Ω for connection to the LPF.
(d) The triple-stack symmetric HBV, which has $C_j(0) = 0.6$ pF, $R_s = 0.3$ Ω, and $V_{BR} = 14$ V. The device characteristics are found from S parameter measurements (see Krishnamurthi [28]). An HB simulation is performed to estimate the impedance of the HBV under full drive. The diode is mounted on a heat sink and the top of the mesa is connected to the input and output circuits.
(e) The output matching circuit resembles the input matching circuit.
(f) A second 50-Ω phase-adjust section.
(g) A 3-section coupled-line BPF, each section being $\lambda/4$ at the center frequency $3f_{in}$.

The design procedure is as follows:

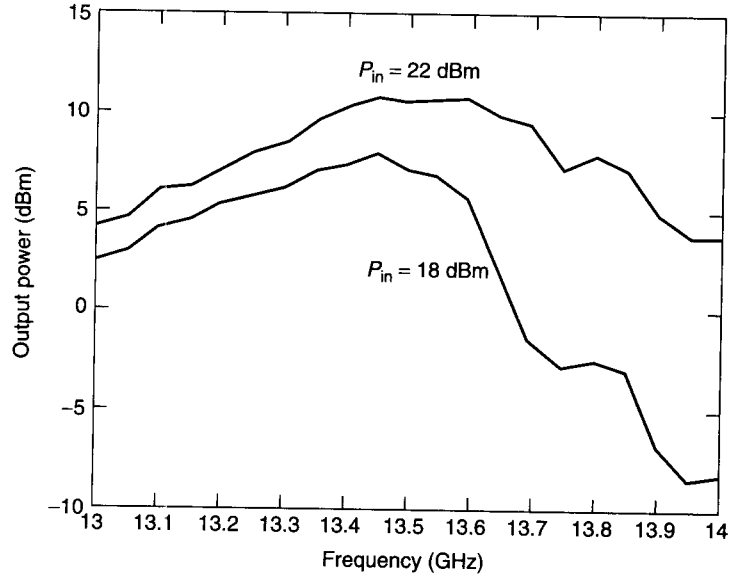
1. Determine $Z_{in}(f_{in})$ and $Z_{out}(3f_{in})$ at the HBV using either the Siegel and Kerr method [67] or by means of the "cubic" charge model (13.20).
2. Design input and output circuits to match $Z_{in}(f_{in})$ and $Z_{out}(3f_{in})$ to a 50-Ω external environment.
3. Design the input LPF and the output BPF.
4. Determine the lengths of the 50-Ω phase-adjust sections such that the input filter presents an open circuit to the HBV input at $3f_{in}$, while the output filter provides an open circuit at the HBV output plane at f_{in}.

Figure 13.27*a* shows the measured frequency response for $P_{in} = 18$ and 22 dBm, while Fig. 13.27*b* depicts the P_{in}–P_{out} transfer characteristic for $f_{in} = 13.45$ GHz. The best efficiency in this realization was $\eta_c = 8.9\%$, although 20% was obtained using the same device in a waveguide circuit [63].

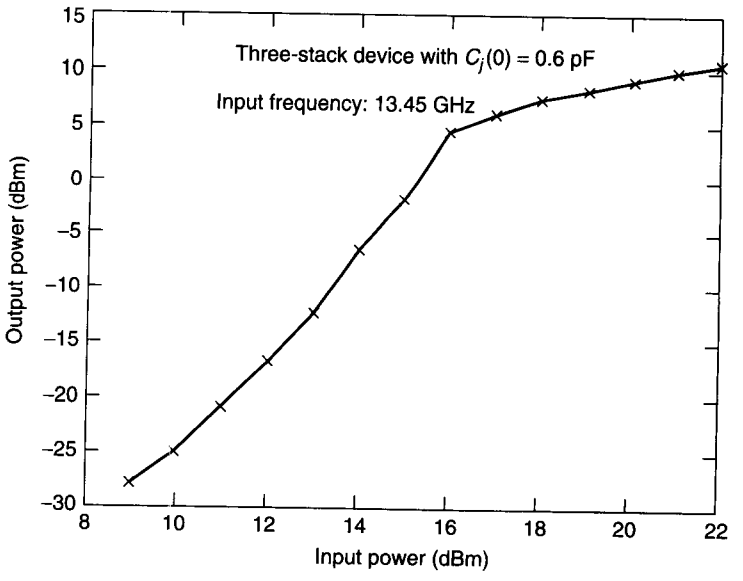
More recently, Mélique et al. have obtained $P_{out} = 5$ mW at 5% efficiency for tripling to 216 GHZ using an InGaAs/InAlAs/AlAs barrier scheme on an InP substrate [68], and $P_{out} = 9.5$ mW at 12.3% efficiency for tripling to 250 GHZ using an HBV with reduced R_s due to a more efficient pseudomorphic step-like InGaAs/InAlAs/AlAs blocking barrier [69].

13.2.4 Active Frequency Multipliers

Compared with diode multipliers, active frequency multipliers have reduced conversion loss (sometimes gain [70–72]), inherent input–output isolation, MMIC process compatibility (since active devices are "planar" whereas var-



(*a*)



(*b*)

Figure 13.27 (*a*) Frequency response of HBV tripler for P_{in} = 18 and 22 dBm. (*b*) P_{in}–P_{out} transfer characteristic for tripling to 40.35 GHz [66].

actors are "vertical"), better bandwidth (in untuned designs), but lower frequency capability. They are used in automotive collision avoidance radar [73], digital microwave radio, microwave and millimeter-wave receiver and transmitter chains, instrumentation, communications networks, and phase-locked loops and tracking systems. In systems needing a low-phase-noise LO, such as digital microwave radio, a solution is to multiply up from a "clean" VCO at say ~3 GHz [74].

Evolution. Active multipliers have been made using a variety of three-terminal devices including MESFETs, HEMTs, and HBTs. Before ~1990 most active multipliers used GaAs MESFETs in single-device or balanced designs executed in microstrip or CPW technology; see [70–72, 75–79]. Since then, the tendency has been to employ pseudomorphic HEMTs (pHEMTs) with gate lengths of the order of ~0.15 μm in CPW and FCPW (finite-ground CPW) technologies. Single-device designs include doublers [79–80], triplers [73, 81–83], and quadruplers [73], while balanced and multiple-device designs include doublers [78–79, 84–85] and triplers [86]. Active multipliers based on HBTs have been addressed in [74] and [83].

Active Multiplier Theory. Active frequency multipliers exploit nonlinear characteristics inherent in certain three-terminal devices. Unlike varactor multipliers, which employ the nonlinear reactance associated with voltage-dependent depletion layers, the important nonlinearity in active multipliers is the transconductance, which can be modeled as a voltage-controlled current source. For a description of the formerly important MESFET active multipliers, see Harrison [75]. The following focuses on HEMT multipliers.

The HEMT. Figure 13.28 shows the layout of a pseudomorphic* high-electron-mobility transistor (pHEMT). The short dimension of the narrow stripe is the gate "length" L_g, while its long dimension is the gate "width" W_g. Figure 13.29 is an oblique view of a small region (broken rectangle) including part of the

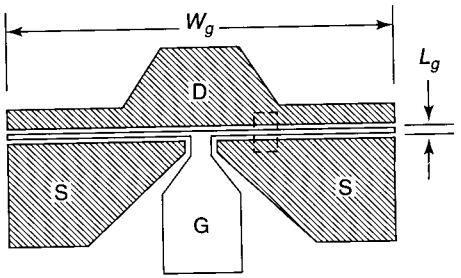


Figure 13.28 Top view of microwave HEMT. The gate (pad G) is the narrow stripe of width W_g and length L_g midway between the source (two pads S) and drain (D). This location of the gate pad minimizes the gate resistance R_g. The outlined rectangle is the region shown in Fig. 13.29.

*The key feature of the pHEMT is *an electrically active interface* between strained but lattice-matched (pseudomorphic) layers of AlGaAs on a GaAs substrate. Other layer systems such as InGaAs/GaAs and InGaAs/InAlAs/InP are also used.

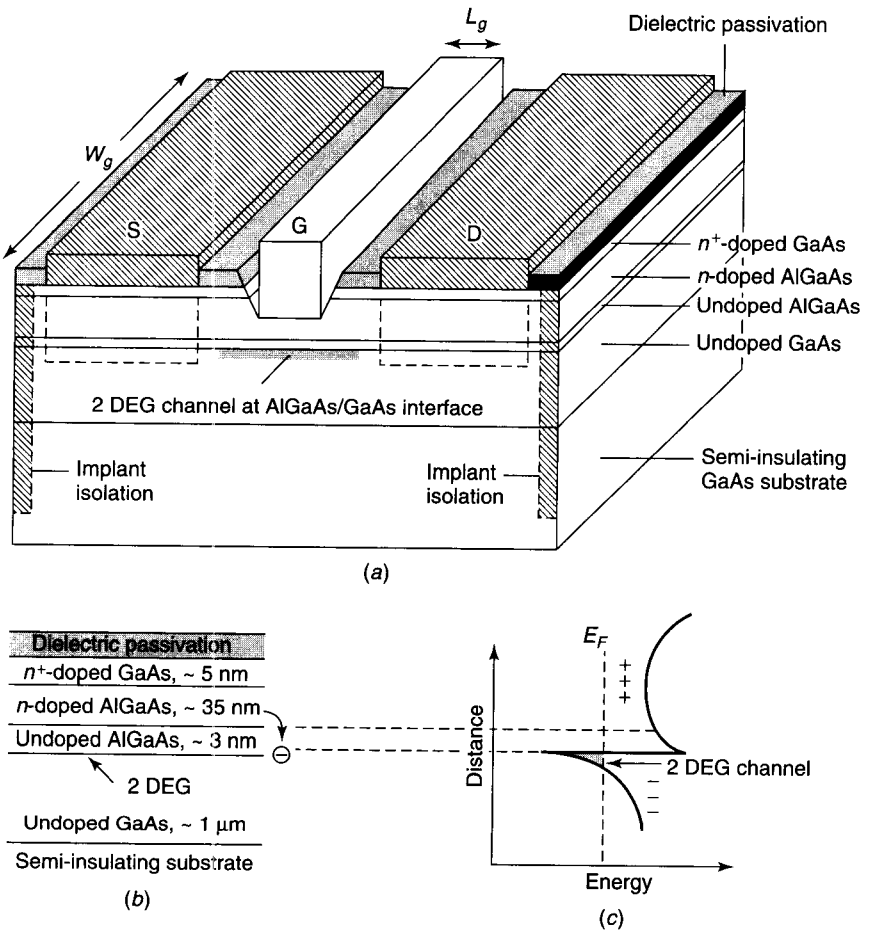


Figure 13.29 (*a*) Cross section of part of conventional GaAs/AlGaAs microwave HEMT structure. (*b*) Enlargement of MBE layers. (*c*) Energy band diagram (unbiased device). (Note: none of (*a*), (*b*), or (*c*) are to scale).

gate stripe. Like the HBV, the HEMT is a heterojunction device. Here the heterojunction between undoped AlGaAs and undoped GaAs separates electrons provided by donors in the doped AlGaAs region. This places a large electron sheet density ($\sim 10^{12}$ cm^{-2}) in a thin layer ($\sim$8 nm), referred to as a two-dimensional electron gas (2DEG), confined in the undoped GaAs layer close ($\sim$20 nm) to the gate. These confined electrons can move rapidly (hence "high electron mobility"). In the off condition, where the gate-source voltage v_{GS} is less than the pinchoff voltage V_p, no drain current i_D flows. In the on state when $v_{GS} > V_p$, the built-in voltage of the heterojunction propels the depletion region back into the AlGaAs layer, the 2DEG is no longer depleted, and the device turns on.

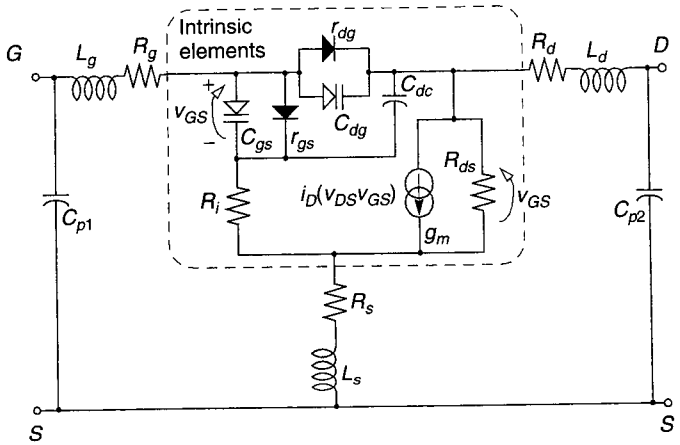


Figure 13.30 Generic nonlinear equivalent circuit applicable to HEMT or to MESFET device. The nonlinear elements are C_{gs}, C_{dg}, r_{gs}, r_{dg}, and g_m.

Numerous HEMT models have been proposed, and detailed expressions for the current–voltage characteristics have been derived from the device physics; see, for example, the charge control model* of De and Meindl [87].

The generic model of Fig. 13.30 is valid for HEMTs and MESFETs operating in their active regions. For application in active multipliers, the main difference between an HEMT and an MESFET is in the transconductance (g_m) characteristic, represented by the nonlinear relation between the drain current i_D and the terminal voltages v_{GS} and v_{DS}. In the HEMT, g_m peaking occurs as v_{GS} is increased [88]. For design purposes a fast-executing simulation algorithm based on one of several empirical models is useful. Angelov [84] models the g_m nonlinearity as a $\tanh(x)$ function whose $\text{sech}^2(x)$ derivative has a "bell" shape like that of the measured $g_m(v_{GS})$ function. This model applies to both conventional AlGaAS/GaAs and pseudomorphic AlGaAs/InGaAs/GaAs HEMTs; its parameters can be extracted from experimental $i_D(v_{GS}, v_{DS})$ and $g_m(v_{GS})$ data. The important nonlinearity for active multipliers is the $i_D(v_{GS}, v_{DS})$ transconductance characteristic, represented in [84] by

$$i_D = I_{\text{pk}}[1 + \tanh(\psi)][1 + \lambda v_{DS}] \tanh(\alpha v_{DS}) \tag{13.30}$$

where I_{pk} is the "intrinsic" i_D (excluding current in R_{ds}) at which g_m is maximum. The quantity ψ represents a power series in the voltage difference $(v_{GS} - V_{\text{pk}})$:

$$\psi = P_1(v_{GS} - V_{\text{pk}}) + P_2(v_{GS} - V_{\text{pk}})^2 + P_3(v_{GS} - V_{\text{pk}})^3 + \cdots \tag{13.31}$$

* *Charge control* refers to the dependence of the channel charge density on the gate bias voltage.

where V_{pk} is the gate voltage v_{GS} for maximum g_m, and α and λ are adjustable parameters. In some HEMTs V_{pk} depends weakly on v_{DS}: this can be modeled as

$$V_{pk} = V_{pk}(0) + \gamma v_{DS} \tag{13.32}$$

where γ is a constant.

The nonlinear capacitances $C_{gs}(v_{GS})$ and $C_{dg}(v_{DG})$ of the equivalent circuit are of second order importance in this application. They can be modeled like Schottky capacitances [89]:

$$C_{gs}(v_{GS}) = C_{gs}(0)\left(1 - \frac{v_{GS}}{V_{bi}}\right)^{-m} \tag{13.33}$$

where $C_{gs}(0)$ and m are adjustable parameters. Alternatively a $\tanh(x)$ model for $C_{gs}(v_{DS}, v_{GS})$ is possible [84].

Active HEMT Multiplier Design. Figure 13.31 shows an idealized $i_D(v_{GS})$ characteristic, neglecting the influence of v_{DS}, with the gate biased to the pinchoff voltage V_p for class B operation. An incoming sinusoidal waveform $v_{GS}(t)$ is transformed into an nonsinusoidal drain current $i_D(t)$ waveform with significant harmonic content. For frequency multiplication by an integer (say 2 or 3) there is a trade-off between the class of operation (determined by the gate bias V_{GS}), the conversion loss of the multiplier (determined by the effective gain of the active device in its circuit embedding), and the circuit complexity. A simple way to optimize the HEMT conduction angle θ and bias is to represent the device as voltage-controlled switch with fixed g_m in the *on* condition. The resulting $i_D(t)$ waveform is a sequence of cosine tips, expressible as a sum of Fourier components:

$$i_D = I_{dc} + I_1 \cos(\omega t) + I_2 \cos(2\omega t) + \cdots + I_N \cos(N\omega t) \tag{13.34}$$

where $I_{dc} = (2I_{max}/\pi)(t_o/T)$, $T = 1/f$ is the period, t_o/T is the *on* aperture, and

$$I_N = \frac{4I_{max}}{\pi}\left(\frac{t_o}{T}\right)\left|\frac{\cos(N\pi t_o/T)}{1 - (2Nt_o/T)^2}\right| \tag{13.35}$$

Figure 13.32 shows the harmonic currents as functions of θ $(=2\pi t_o/T)$. For more detailed analyses of generic active multipliers see [88].

In a practical multiplier there will be numerous device–circuit interactions, so accurate simulations are needed to determine the optimum circuit configuration and biasing. The interactions can make the $i_D(t)$ waveform unlike that in Fig. 13.31. In the following example the $i_D(t)$ waveform does resemble Fig. 13.31, but in the subsequent one it is very different.

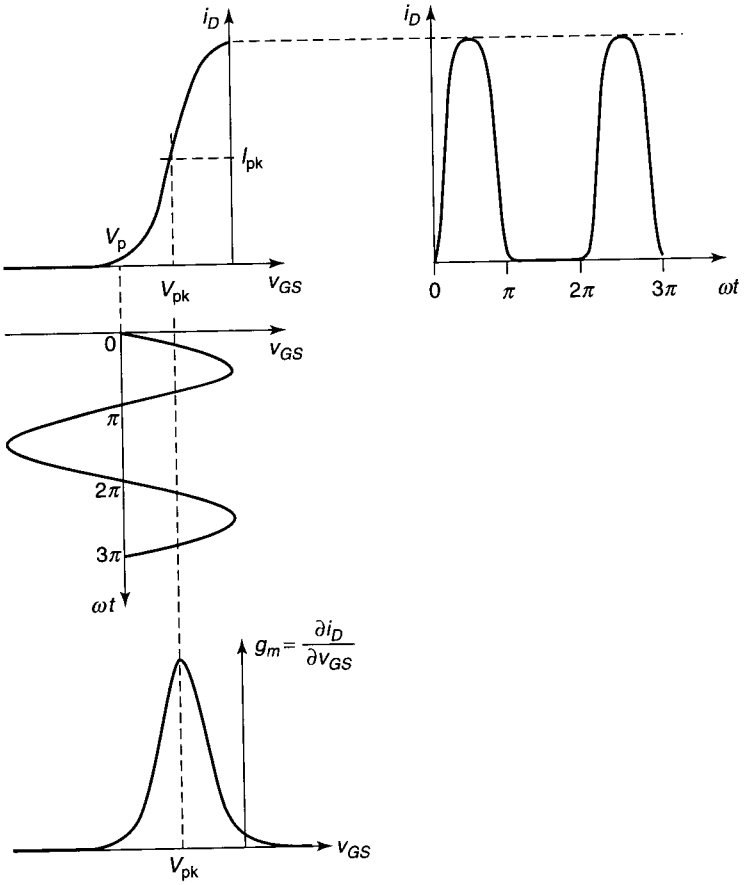


Figure 13.31 Idealized drain current waveform $i_D(t)$ in response to sinusiodal $v_{GS}(t)$ for HEMT device biased at $V_{GS} \sim V_p$. This is a class B bias condition.

Practical Active Frequency Multipliers

(a) A *balanced active doubler* from ~20 to ~40 GHz is shown in Fig. 13.33*a*. Balanced designs suppress feedthrough of the fundamental and odd-order harmonics. Here [84] the empirical model in Fig. 13.30 and Eqs. (13.30)–(13.33) are used to design a circuit consisting of a 180° 3-dB coupler feeding equal-amplitude out-of-phase signals to the gates of two HEMTs, each biased for i_D clipping. The drain output lines are directly connected, so that the fundamental and odd-order harmonic currents from the two devices have opposite phases and therefore cancel, but the second-harmonic currents are in phase and so add. The HEMT model was implemented in Hewlett-Packard's MDS software [90] using equivalent circuit parameters derived from DC and *S* parameter measurements up to 62 GHz. Table 13.2 summarizes extracted pHEMT

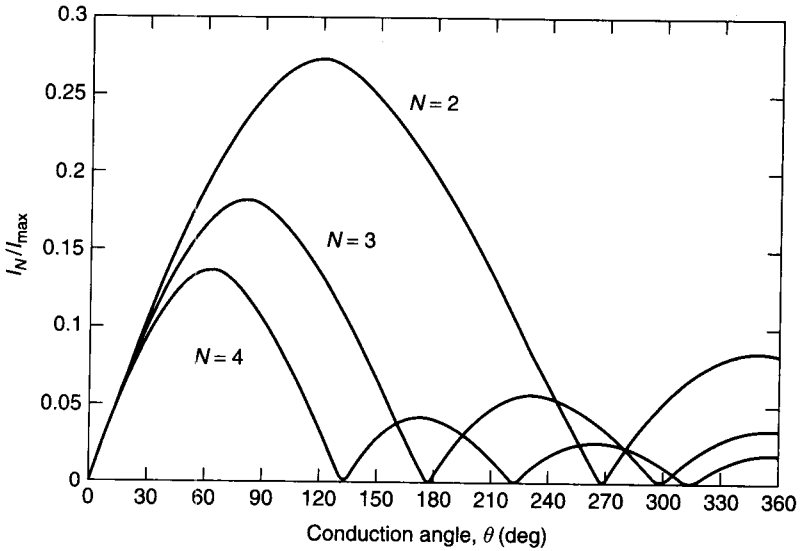


Figure 13.32 Normalized harmonic current components I_N as functions of conduction angle θ. (After Boudiaf et al. [81]. Reprinted with permission of IEEE.)

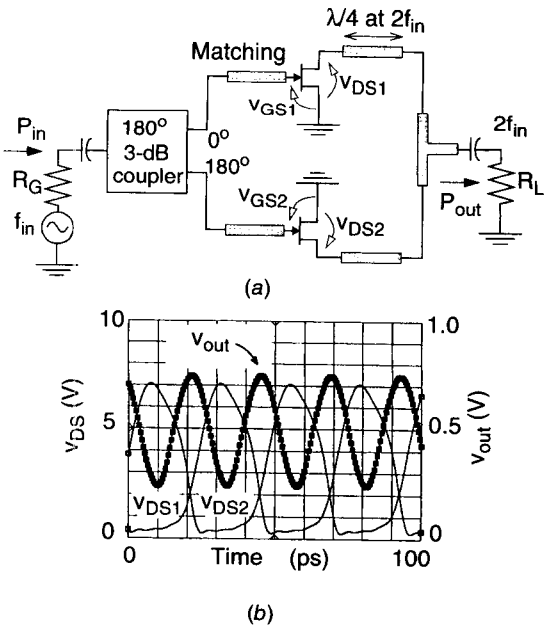


Figure 13.33 (*a*) Block diagram of balanced pHEMT frequency doubler. Biasing is not shown. (*b*) Simulated voltage waveforms $v_{DS1}(t)$, $v_{DS2}(t)$, and $v_{out}(t)$ of doubler. (After Angelov et al. [84]. Reprinted with permission of IEEE.)

Table 13.2 Extracted pHEMT Model Parameters

$R_g = 8\ \Omega$	$R_{gs} = 5\ \Omega$	$R_s = 5\ \Omega$	$R_d = 6\ \Omega$	$R_{ds} = 1\ \text{k}\Omega$	$C_{ds} = 14\ \text{fF}$	$C_{gs} = 40\ \text{fF}$
$C_{gd} = 13\ \text{fF}$	$I_{\text{pk}} = 15\ \text{mA}$	$P_1 = 2.6$	$P_3 = 7.0$	$V_{\text{pk}} = 0.2\ \text{V}$	$\alpha = 3.0$	$\lambda = 0.015$

Source: Angelov et al. [84].

model parameters. HB simulations using 8 to 10 harmonics show the optimum output line length to be $\lambda/4$ at the output frequency $2f$. This decouples the transistors at f but matches them to the output at $2f$. Figure 13.33*b* shows the voltage waveforms recovered by inverse Fourier transformation. The $v_{DS1}(t)$ and $v_{DS2}(t)$ waveforms are interleaved and sum to produce a sinusoidal output $v_{\text{out}}(t)$. The analysis shows that optimum conversion efficiency, output power, and spectral purity are obtained with DC bias such that $I_D = I_{\text{pk}}$. Although the exact bias condition (class A, B, ...) is not critical, it is important to avoid destructively high instantaneous values of $v_{DS}(t)$.

The discrete-pHEMT doubler, realized on a 0.127-mm Al_2O_3 substrate, is shown in Fig. 13.34. The 180° 3-dB coupler is a rat-race hybrid ring; all circuit components are optimized for doubler bandwidth and efficiency. The δ-doped pHEMT devices have $f_{\max} \sim 180$ GHz. Figure 13.35*a* gives the P_{in}–P_{out} curve for $f_{\text{in}} = 21$ GHz, while 13.35*b* shows the frequency response for $P_{\text{in}} = 6$ dBm, $V_{GS} = 0.5$ V, $V_{DS} = 3$ V. Overall bandwidth is limited by the 3-dB coupler.

(b) A *single-ended active tripler* from ~12.6 to ~38 GHz [81] has the block diagram of Fig. 13.36*a*. Figure 13.36*b* is a photograph of the tripler chip. The

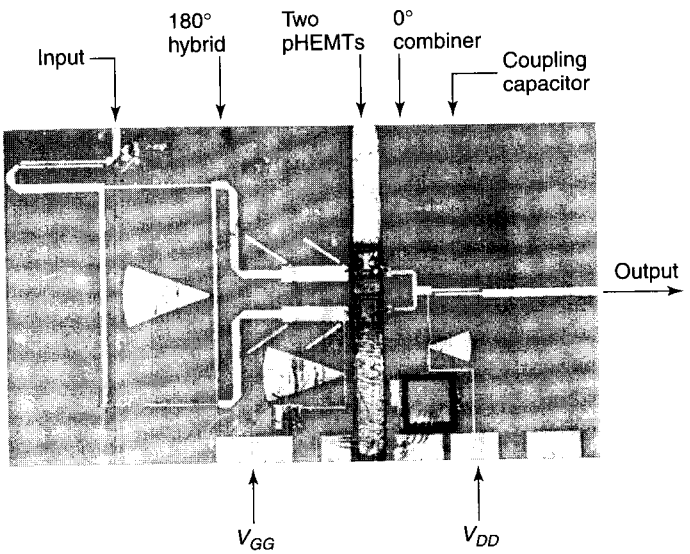


Figure 13.34 The pHEMT doubler circuit fabricated on a 5-mil Al_2O_3 substrate. (After Angelov et al. [84]. Reprinted with permission of IEEE.)

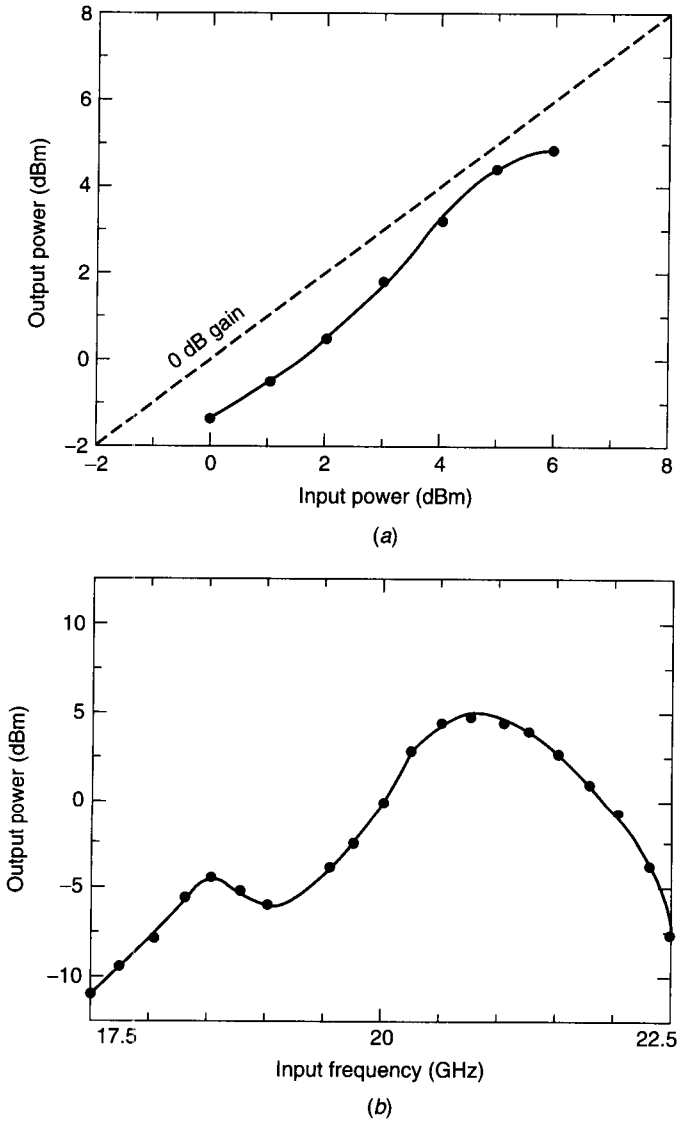
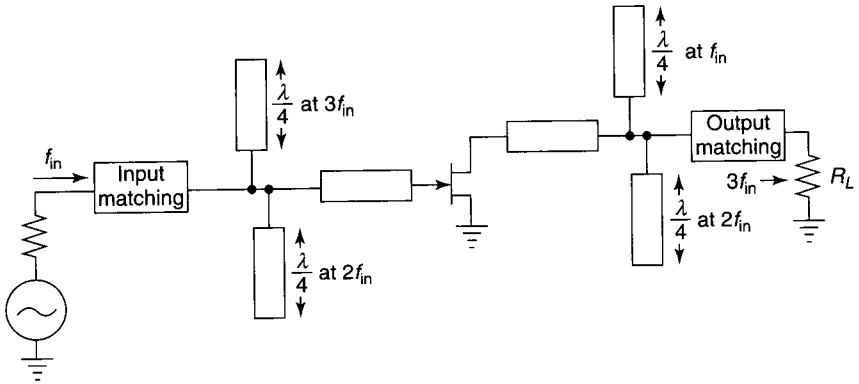
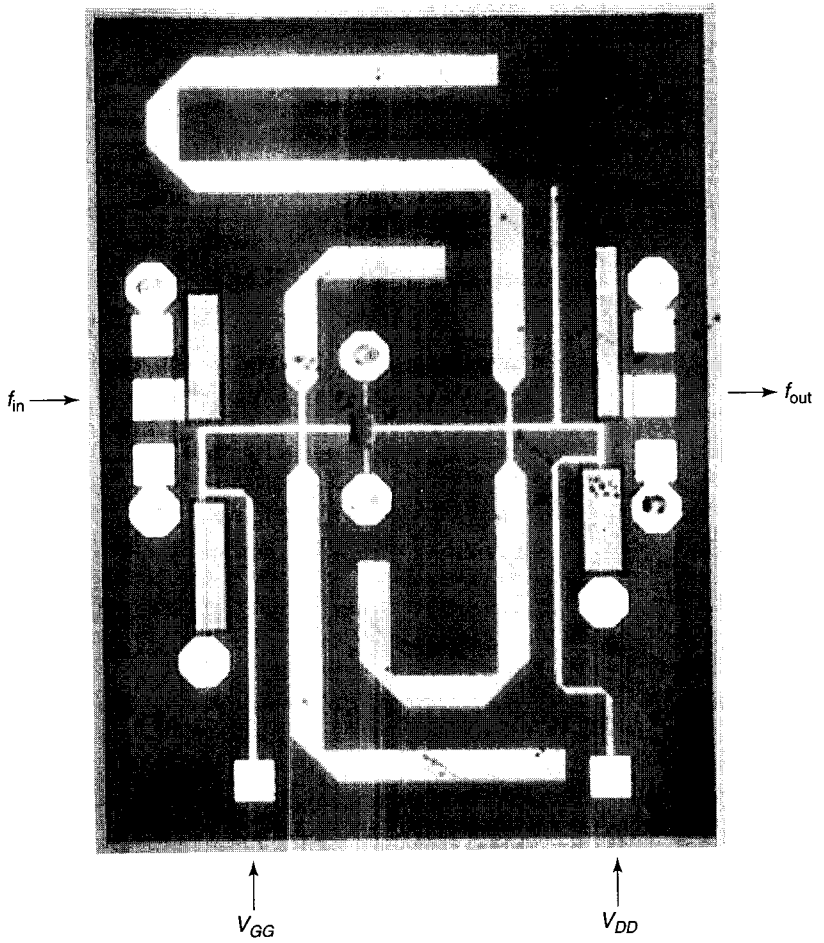


Figure 13.35 Measured results for pHEMT balanced doubler, biased at $V_{DS} = 3$ V, $V_{GS} = +0.5$ V: (*a*) output power P_{out} vs. input power P_{in} for $f_{in} = 21$ GHz; (*b*) frequency response for $P_{in} = 6$ dBm. (After Angelov et al. [84]. Reprinted with permission of IEEE.)

GaAs pHEMT used here has $L_g = 0.2$ μm and $f_T = 62$ GHz. Simulations show that biasing with $V_{gs} \sim V_p$ for class B provides the best overall performance. More precisely, $V_{gs} = -0.7$ V gives the optimum efficiency $\eta_c = 11\%$, but $V_{gs} = -0.6$ V gives the maximum $P_{out} = 3.8$ dBm. Figure 13.37*a* depicts the optimized class B dynamic loadline for $V_{DS} = 3$ V, while Fig. 13.37*b* shows the



(*a*)



(*b*)

Figure 13.36 (*a*) Block diagram of single-ended pHEMT tripler. (*b*) Photograph of 1.5 × 2-mm tripler chip layout. Substrate thickness is 100 μm. (After Boudiaf et al. [81]. Reprinted with permission of IEEE.)

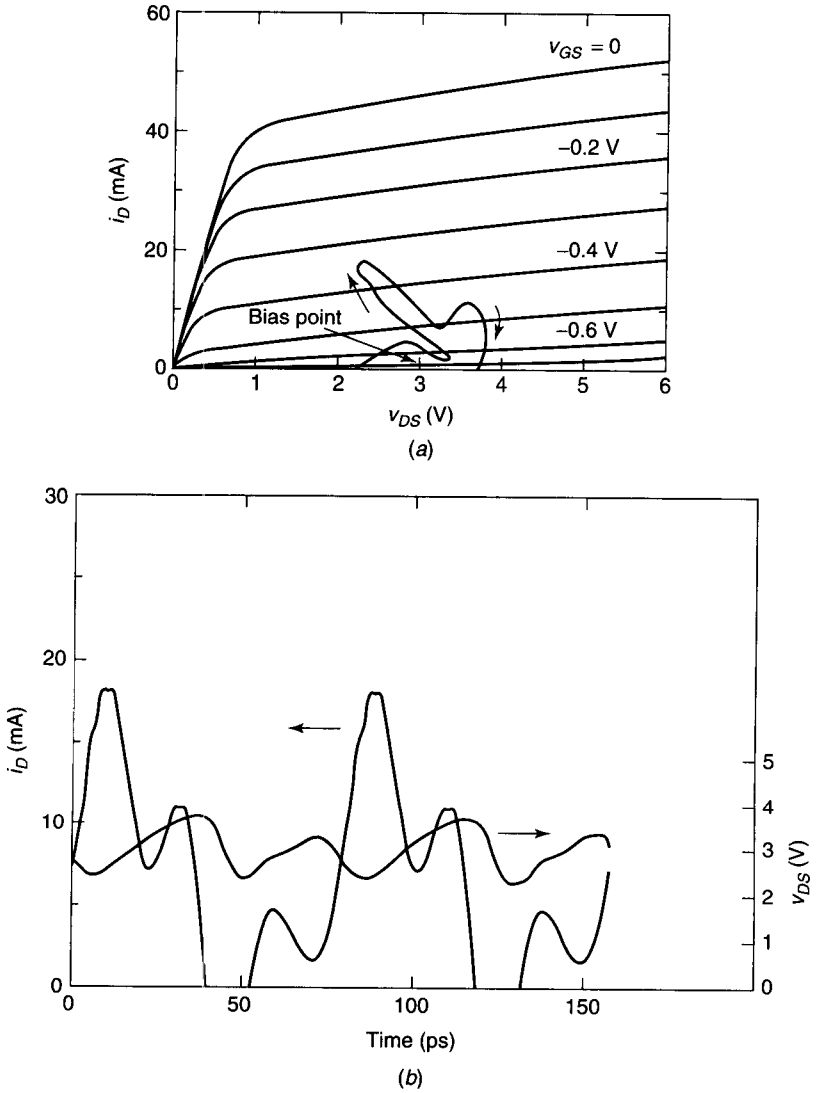


Figure 13.37 Simulated pHEMT tripler results: (*a*) optimized class-B dynamic load line for $V_{DS} = 3$ V; (*b*) simulated $i_D(t)$ and $v_{DS}(t)$ waveforms, showing strong third-harmonic component. (After Boudiaf et al. [81]. Reprinted with permission of IEEE.)

corresponding $i_D(t)$ and $v_{DS}(t)$ waveforms, both having strong third-harmonic components. Figure 13.38*a* gives the measured gain and fundamental and second-harmonic rejection levels, all as functions of P_{in}. Conversion efficiency is seen Fig. 13.38*b*. The measured phase noise degradation is 9 ± 1 dB, which is consistent with the theoretical value of $20 \log(3) = 9.54$ dB for a resistive tripler.

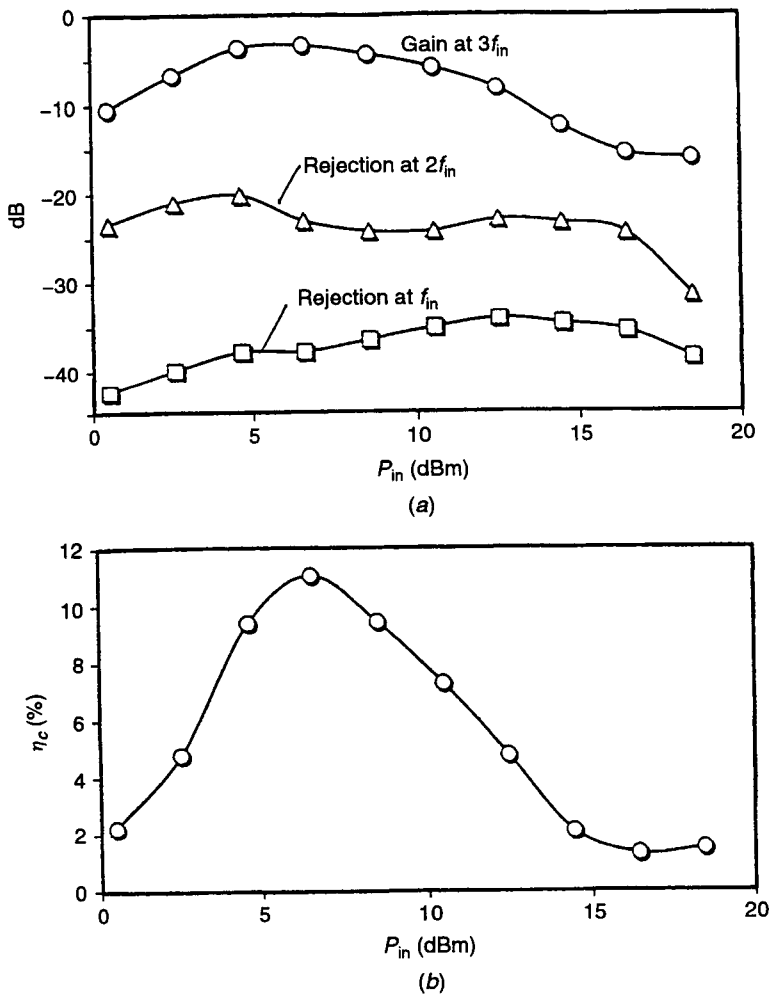


Figure 13.38 Measured results for pHEMT tripler at $f_{out} = 38.64$ GHz: (*a*) gain (○), rejection at $2f_{in}$ (△) and at f_{in} (□), all versus input power P_{in}; (*b*) conversion efficiency η_c versus P_{in}. (After Boudiaf et al. [81]. Reprinted with permission of IEEE.)

13.3 FREQUENCY DIVISION

13.3.1 Types of Frequency Dividers

Frequency division can be achieved in numerous ways, including (i) parametric subharmonic generation using pumped varactors, (ii) exploitation of charge storage phenomena in SRDs, (iii) mixer-with-feedback (regenerative) frequency dividers, (iv) transit-time phenomena in Gunn diodes, (v) digital frequency dividers, (vi) injection-locked oscillators (ILOs), and (vii) phase-locked loops

incorporating multipliers in the feedback path. Of these, method (i) can provide octave-bandwidth (67%) dividers [91] with low phase noise but require high input power and have conversion loss ~12 dB. Under high drive conditions parametric dividers (like other parametric frequency converters) can exhibit unwanted instabilities and even chaos. The designer should consult the literature on this subject [92, 93]. The related SRD method (ii) has received scant attention [94]. Regenerative (Miller) dividers using method (iii) have ~15% bandwidths and are widely used in contemporary communications systems; they are reviewed here. Method (iv), exploiting the transit-time delay of charge domains in transferred-electron devices, was expected to be important [95]. However, they had phase noise problems and bandwidths of only ~8%. They are unsuited for MMICs because of excessive power dissipation (a high electric field is required over the device length). Method (v), resulting from the improving speed of digital dividers, is widely used and provides fractional bandwidths up to ~100%; penalties are power consumption and poor phase noise performance. The ILO method (vi), related to method (iii), provides high conversion gain (~40 dB) but narrow bandwidth (~3%). If the harmonic locking signal is removed, ILOs free run and can also exhibit chaotic behavior [96]. Method (vii), a conventional PLL application, is outside the scope of this chapter.

In the following, methods (i), (iii), and (vi) are reviewed.

13.3.2 Parametric Frequency Dividers

Varactor Parametric Divider Theory. The Manley–Rowe equations [8] predict that for ideal parametric frequency dividers (PFDs), the power $P_{(1/N)}$ generated at the Nth subharmonic of f_{in} is related to the input power P_1 by

$$\frac{P_{(1/N)}}{P_1} \leq 1 \tag{13.36}$$

that is, the maximum possible efficiency is 100%.

Filter-based PFDs using the topology of Fig. 13.15*a* have been investigated using HB simulations [97]. Here we analyze a more practical balanced topology. Figure 13.39*a* shows the schematic of a balanced two-varactor PFD; for a nonlinear model of the abrupt-junction or Schottky varactors see Fig. 13.14. For these devices, exponent γ is $\sim\frac{1}{2}$, so that the "parabolic" charge model of (13.13) applies. The even/odd modality of the circuit lends itself to closed-form analysis [98], resulting in a pair of coupled differential equations. The first describes the even-mode excitation:

$$(\xi_s + \xi_p)\frac{du}{d\tau} + 1 - \tfrac{1}{4}(u^2 + z^2) = x(\tau) + X_0 \tag{13.37}$$

and the second, describing the odd-mode resonance, is [assuming $\omega_0 C_j(0) R_L \gg 1$]

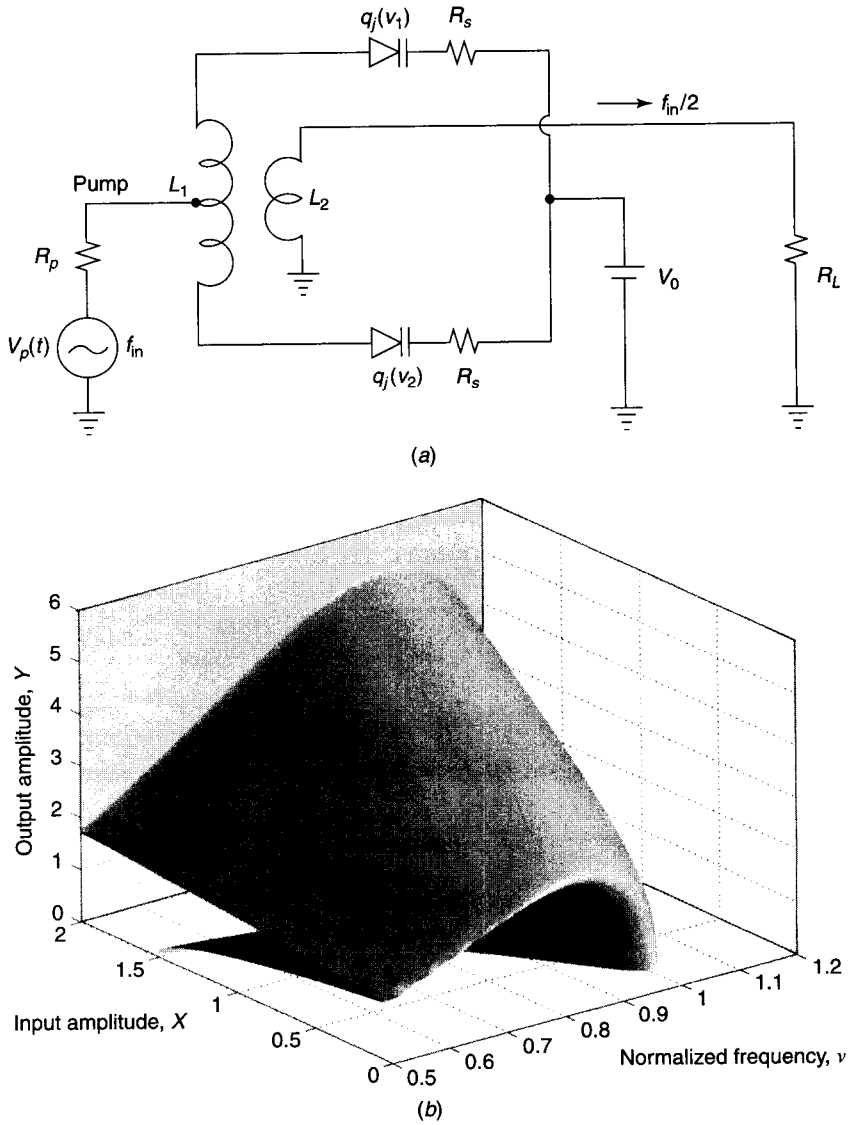


Figure 13.39 (*a*) Equivalent circuit of balanced parametric frequency divider by 2 (frequency halver). (*b*) Theoretical response surface according to Eq. (13.39), assuming that $\xi_p = 0.1$, $\xi_s = 0.1$, $X_0 = -0.1$.

$$\frac{d^2z}{d\tau^2} + \xi_s z = \tfrac{1}{2}uz \tag{13.38}$$

where $u = [q_j(v_1) + q_j(v_2)]/V_{bi}$ is the normalized sum of the varactor charges
$z = [q_j(v_1) - q_j(v_2)]/V_{bi}$ is the normalized difference between the varactor charges

$\tau = \omega_0 t$ is the normalized time
$\omega_0 = [L_1 C_j(0)/2]^{-1/2}$ is the small-signal resonant frequency at zero bias
$\xi_s = \omega_0/[2\pi f_c(0)] = \omega_0 C_j(0) R_s$ is the ratio between ω_0 and the varactor cutoff frequency
$\xi_p = \omega_0 C_j(0) R_p$ is a pump resistance parameter
$x = X\cos(2\nu\tau)$ is the normalized pump voltage v_p at frequency 2ω
$\nu = \omega/\omega_0$ is the normalized output frequency (detuning)
$X = V_p/V_{bi}$ is the pump amplitude
$X_0 = (V_{dc}/V_{bi})$ is the normalized bias voltage
V_{bi} is the varactor built-in voltage

Eliminating u between (13.37) and (13.38), and assuming that $z(\tau) \approx Z\cos(\nu\tau + \psi)$, one can use "algebraic" harmonic balance (HB) to solve (Problem 13.6) for the charge amplitude Z and phase ψ. The output voltage v_{out} is related to z by $d^2z/d\tau^2 = y$, where $y = v_{out}/V_{bi}$. Assuming that $y(\tau) \approx Y\cos(\nu\tau + \theta)$, the steady state solution is

$$\left[\frac{3}{2}(\nu^4 - 1 + X_0) + \xi_s\left(\frac{9}{2}\xi_s + 8\xi_p\right)\nu^2 + \frac{5Y^2}{64\nu^4}\right]^2 + [2\xi_s\nu^3]^2 = X^2 \quad (13.39)$$

Figure 13.39*b* shows the resulting halving response as a function of frequency and amplitude when $\xi_s = 0.1$, $\xi_p = 0.1$, $X_0 = -0.1$. For normalized frequencies $\nu \lesssim 1$ there is a region of hysteresis with respect to both ν and X. The output power at the divided frequency results from a virtual negative resistance created at frequency ω when the varactors are pumped at frequency 2ω. The position of the "nose" of the response curve moves toward higher frequencies as the bias X_0 is made more negative. The balanced design has inherent input–output isolation, minimizing the need for RF filters, thereby improving transient response.

What the algebraic analysis does *not* predict is that the PFD, like any nonlinear reactance circuit of second order or higher, can become chaotic under high drive, which may preclude its use in certain applications. A method of investigating this using HB simulators is given in [93].

Practical PFD Circuit. A practical circuit is shown in Fig. 13.40*a*. The symmetrical configuration consists of a substrate of thickness h with an input microstrip of width w on one side and a collinear slotline of width s on the other. The nonlinear capacitances of the varactors, together with the overlapping microstrip/slotline section of length ℓ, form a parametric subharmonic resonator. An input signal at f entering the microstrip excites the varactors in phase (even mode). Because of the nonlinear coupling mechanism between this mode and the subharmonic resonance (odd mode), energy is transferred from $2f$ to f, causing subharmonic currents to flow in the path indicated by the

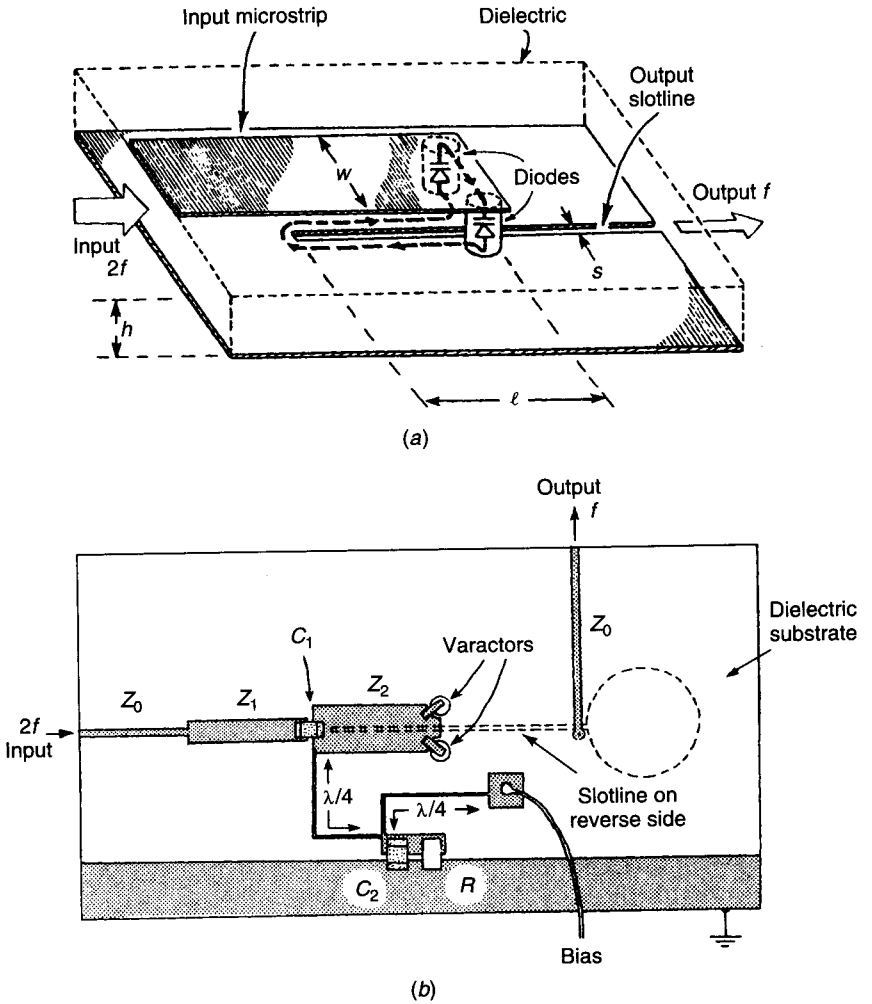


Figure 13.40 Microwave parametric frequency halver. (*a*) Basic symmetric subharmonic resonator consisting of varactor-loaded microstrip section overlapping slotline. (*b*) Practical implementation for 2–4-GHz input band. The varactors are connected to the end of the microstrip and contact the ground plane on the reverse side through vias. C_1 is a blocking capacitor; the bias circuit consists of capacitor C_2, resistor R, and two high-impedance quarter-wave sections.

heavy broken line. Since the E field in the microstrip is approximately orthogonal to the E field across the slot, undesired input–output coupling is minimized. A practical implementation for the 2–4-GHz input band is given in Fig. 13.40*b*. It uses a two-section input matching transformer and a slotline/microstrip output transition. Octave-bandwidth performance is seen in Fig. 13.41. The measured hysteresis region is much smaller than predicted because the analysis omits parasitics and loss mechanisms.

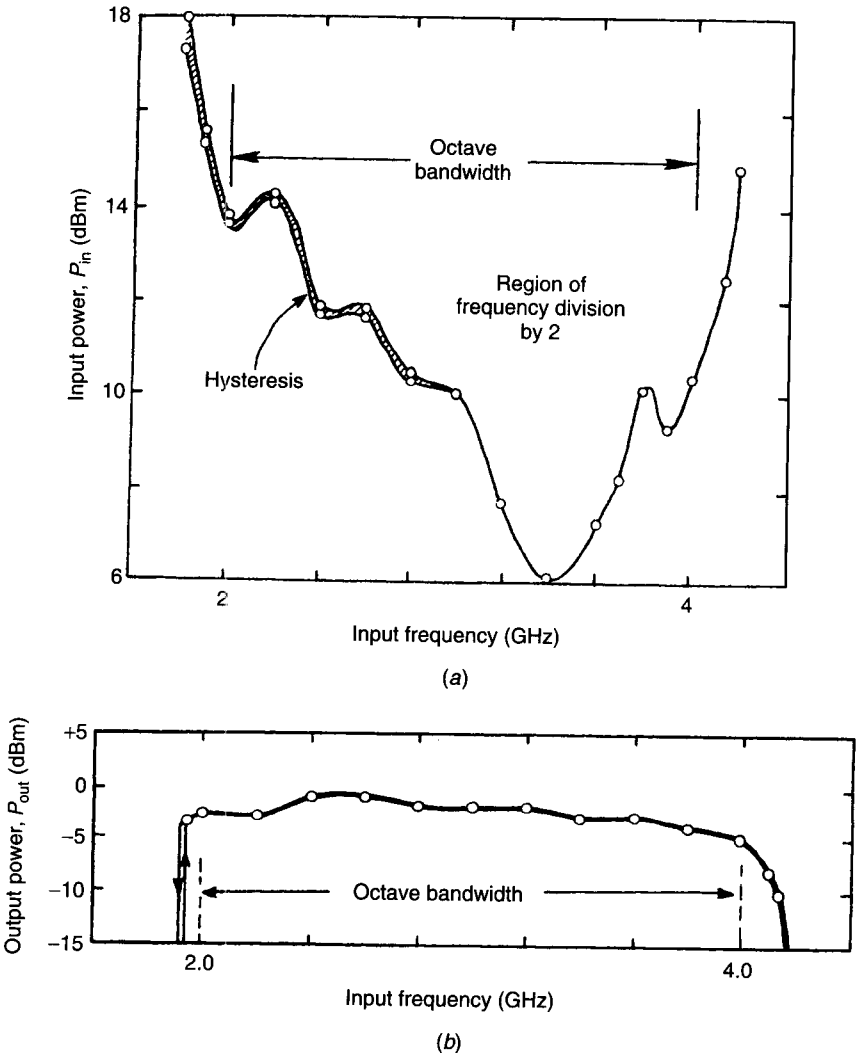


Figure 13.41 Measured performance of octave-bandwidth balanced parametric frequency halver of Fig. 13.40: (*a*) region of frequency division by 2; (*b*) power output P_{out} at divided-by-2 frequency for $P_{in} = +15$ dBm. Bias is $V_{dc} = 0.87$ V.

13.3.3 Regenerative (Mixer-with-Feedback) Frequency Dividers

A useful frequency division approach is the regenerative "mixer with feedback," or Miller divider [99], which requires overall loop gain and therefore active devices.

The Concept. Figure 13.42*a* shows a scheme for frequency division by an integer N. There is a *wide-band mixer* and a *low-pass filter* in the forward path

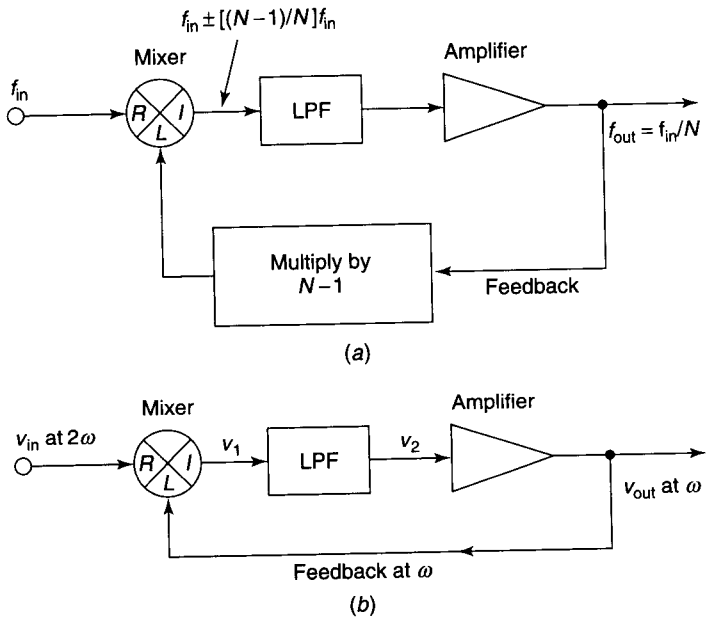


Figure 13.42 (*a*) Generalized scheme for regenerative frequency division by *N*. For division by 2, the multiplier is not required. (*b*) Generic circuit for analysis of divide-by-2 RFD.

and a *frequency multiplier* in the feedback path. Three conditions must be satisfied: (1) a finite-amplitude noise signal must be present in the loop to provide an initial condition for regeneration, (2) the small-signal loop gain must exceed unity, and (3) for operation as a true regenerative frequency divider (RFD) with zero output for zero input, the loop gain should be less than unity when the input is removed [100]. If condition 3 is not met, the circuit will behave as an ILO frequency divider; see Section 13.3.4. For division by N, an input at frequency f_{in} is mixed with a signal at $(N-1)f_{in}/N$ to give the dominant sidebands

$$f_{in} \pm \frac{(N-1)f_{in}}{N} = \frac{(2N-1)f_{in}}{N} \quad \text{or} \quad \frac{f_{in}}{N} \tag{13.40}$$

The $(N-1)f_{in}/N$ signal is obtained from the lower sideband f_{in}/N by means of a times-$(N-1)$ multiplier. A low-pass filter rejects the upper sideband, and an amplifier overcomes mixer and multiplier losses. If the multiplier is omitted, then $N = 2$, and the circuit is a frequency halver. For stability with FM inputs, the loop delay due to dispersive elements such as filters must be much less than the FM modulation period. For CW inputs, the loop group delay limits the bandwidth, which is bounded by asynchronous and submultiple-division modes [101], and possible regions of chaotic behavior.

Theory. The starting condition for the RFD of Fig. 13.42*b* will be found from Miller's original theory [99]. Then the steady state will be investigated using a combination of the Miller theory and a stability analysis due to Derksen et al. [102].

The exact behavior of an RFD depends on the characteristics of all the loop components. We assume that the amplifier is linear and that the mixer is a wideband four-diode-ring double-balanced mixer (DBM) (Chapter 11), with output current given by the even-symmetry function

$$i_I = I_s \sinh(\Lambda v_R) \sinh(\Lambda v_L) \tag{13.41}$$

where i_I is the current at the IF port and $v_R(t)$ and $v_L(t)$ are the voltages at the RF and LO ports.

Starting Condition. In Fig. 13.42*b*, let the 2ω input signal be

$$v_{\text{in}} = V_{\text{in}} \cos(2\omega t) \tag{13.42}$$

and assume that the feedback signal at the halved frequency ω has a phase angle θ:

$$v_{\text{out}}(t) = V_{\text{out}} \cos(\omega t + \theta) \tag{13.43}$$

For small signals just exceeding noise, and assuming that the amplifier presents a resistive impedance R_{in} to the mixer IF port, the mixer function* (13.41) can be replaced by the product [99]

$$v_1(t) \cong 2Kv_{\text{in}}(t)v_{\text{out}}(t) \tag{13.44}$$

where $K = \frac{1}{2} R_{\text{in}} I_s \Lambda^2$. From (13.42), (13.43), and (13.44)

$$v_1 = KV_{\text{in}}V_{\text{out}}[\cos(3\omega t + \theta) + \cos(\omega t - \theta)] \tag{13.45}$$

If the 3ω component is removed by the LPF, if all loop losses are included in the gain A_v, and if the total phase shift (at ω) around the loop from the mixer IF port back to its LO port is ϕ_1, then the output is

$$v_{\text{out}}(t) = A_v KV_{\text{in}}V_{\text{out}} \cos(\omega t - \theta + \phi_1) \tag{13.46}$$

Equating (13.43) and (13.46), the starting condition is $V_{\text{in}} \geq 1/(KA_v)$, or for a diode DBM

$$V_{\text{in}} \geq \frac{2}{RI_s\Lambda^2 A_v} \tag{13.47}$$

* Using the approximation $\sinh(x) \cong x$ for small x.

Steady State. For the steady state solution, assume that the DBM acts as ideal switch. Then the large-signal feedback modulates the incoming signal by a square-wave function, and the steady state amplitude is [99]

$$V_{\text{out}} = \frac{4}{3\pi} A_v V_{\text{in}} K_1 \frac{1}{\sqrt{1 + 3\cos^2(\phi)}} \tag{13.48}$$

where K_1 is a constant, and the phase of the feedback signal is

$$\theta = \tfrac{1}{2} \arctan[\tfrac{1}{2} \tan(\phi)] \tag{13.49}$$

Stability Analysis. In Fig. 13.42*b*, let the filter output be $v_2(t) = V_2 \cos[\omega t + \mu(\theta)]$ where $\mu(\theta)$ is a *phase function* characterizing the mixer plus LPF. According to Derksen et al. [102] the criterion for RFD stability is

$$\left|\frac{\partial \mu}{\partial \theta}\right| < 1.0 \tag{13.50}$$

The result of the calculation is that the RFD is stable when

$$\tfrac{1}{2}(N-1)\pi + \theta_c < \theta < \tfrac{1}{2}N\pi - \theta_c \qquad N = 1, 2, 3, \ldots \tag{13.51}$$

where $\theta_c = \frac{1}{4} \arccos \frac{1}{3} = 0.308$ rad $= 17.6°$ is the critical value of θ. The resulting RFD amplitude-phase response is shown* in Fig. 13.43. The LPF passband selects only one of the possible sequence of responses. Using (13.49) the corresponding critical value of the loop phase ϕ is $\phi_c = 0.955$ rad $= 54.7°$, and the stable ranges of ϕ are given by

$$(M-1)\pi + \phi_c < \phi < M\pi - \phi_c \qquad M = 1, 2, 3, \ldots \tag{13.52}$$

The centers of these ranges are at $\phi_M = \frac{1}{2}(2M-1)\pi$ radians. In Fig. 13.43 the first two ranges are centered at $\pi/2$ and $3\pi/2$. The *frequency* response of the RFD can be found by assuming flat loop gain and constant group delay T_0 (ϕ increasing linearly with ω). Then the stable ranges are given by

$$\frac{(M-1)\pi + \phi_c}{T_0} < \omega < \frac{M\pi - \phi_c}{T_0} \qquad \text{(rad/s)} \tag{13.53}$$

Each stable range has the same *absolute* bandwidth

$$\Delta f = (\pi - 2\phi_c)/(2\pi T_0) \qquad \text{(Hz)} \tag{13.54}$$

*In the case illustrated, the amplifier has a *compressive* transfer function $A_v = A_0[(1 - V_{\text{out}})/V_{\text{sat}}]$ where A_0 and V_{sat} are constants.

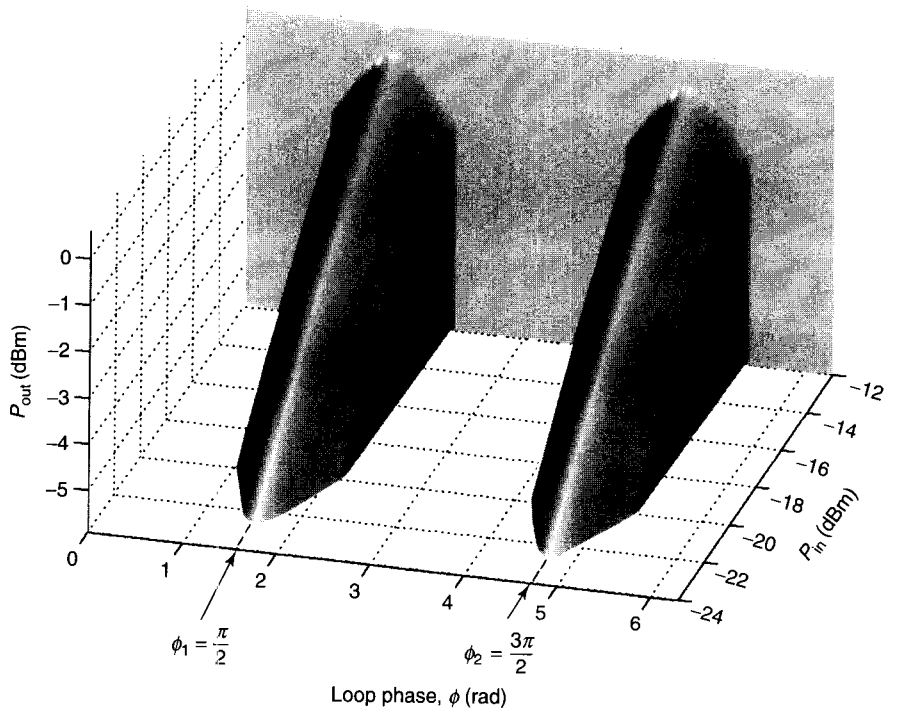


Figure 13.43 Regenerative frequency divider: regions of stable frequency division by 2, vs. loop phase ϕ, according to simple algebraic theory. The first two regions are centered at $\phi_1 = \frac{1}{2}\pi$ and $\phi_2 = \frac{3}{2}\pi$.

but an M-dependent *fractional* bandwidth

$$\frac{\Delta f}{f} = \frac{2(\pi - 2\phi_c)}{(2M-1)\pi} \tag{13.55}$$

From (13.54), the loop delay T_0 should be minimized for maximum absolute bandwidth, but (13.64) indicates that M should be minimized for maximum fractional bandwidth.

Comparison with Measurement. The measured response of a discrete-component regenerative frequency divider by 2 (Fig. 13.44) is similar to the algebraic predictions, including the sudden cessation of regeneration at the band edges. The fractional bandwidth is ~5.9%, a consequence of the long delay in the feedback path.

Practical MMIC Regenerative Frequency Dividers. For acceptable bandwidth, the electrical length of the feedback path must be minimized; this is

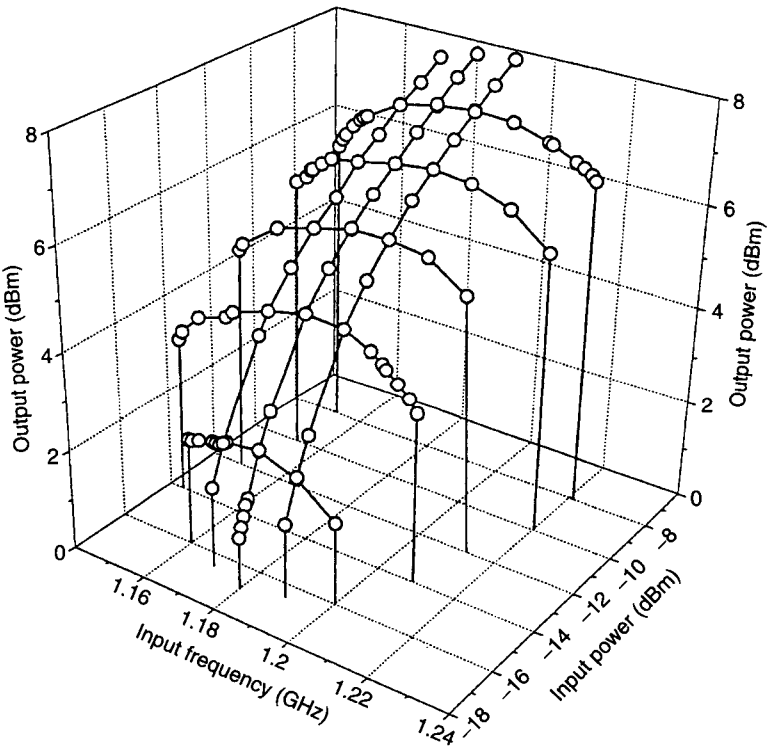


Figure 13.44 Measured amplitude and frequency response of discrete-component regenerative frequency divide-by-2 circuit. Compare with theoretical response of Fig. 13.43.

favored by MMIC implementations. In some designs a single active device performs both mixing and amplifying functions; see Table 13.3.

The RFDs using single-gate GaAs FETs all have conversion loss; those with dual-gate FETs provide conversion gain. Figure 13.45*a* shows a typical schematic: The feedback loop consists of the FET, the capacitor C_3 and the airbridge inductor L_2. Figure 13.45*b* is a micrograph of the GaAs MMIC chip. Figure 13.46 shows that for $f_{in} \sim 9.7$ GHz the maximum $\Delta f/f$ is 8.2%. The frequency response for a fixed input power is given in Fig. 13.47. The response surface resembles that of Fig. 13.44 even though the physical circuits are very different: the regions of operation are the same shape, the frequency responses are bounded by abrupt transitions from on to off, and the amplitude transfer functions are the same shape. Dissimilarities (due to the small gain of an FET compared with that of a discrete amplifier) include the much larger threshold power requirement of the integrated RFD, and the ~8 dB conversion loss of the integrated divider compared with the large conversion gain of the discrete-component example.

In other designs, multiple active devices perform the mixing and amplifica-

Table 13.3 Single-Active-Device RFDs

Active Device	Technology	Division Ratio	Input Frequency (GHz)	Fractional Bandwidth $\Delta f/f$ (%)	Gain (dB)	Reference (in date order)
GaAs FET	Microstrip	2:1	~16	6	−4	[100]
Schottky diode mixer + GaAs FET	GaAs chips, microstrip	2:1	~14	7	−10	[103]
GaAs FET	GaAS MMIC	2:1	~9.7	8.2	−8	[104]
Two-gate GaAs FET	Microstrip	2:1	~7	17–20	+4	[105]
Two-gate GaAs FET	GaAs MMIC	2:1	~15	4	+2.5	[106]

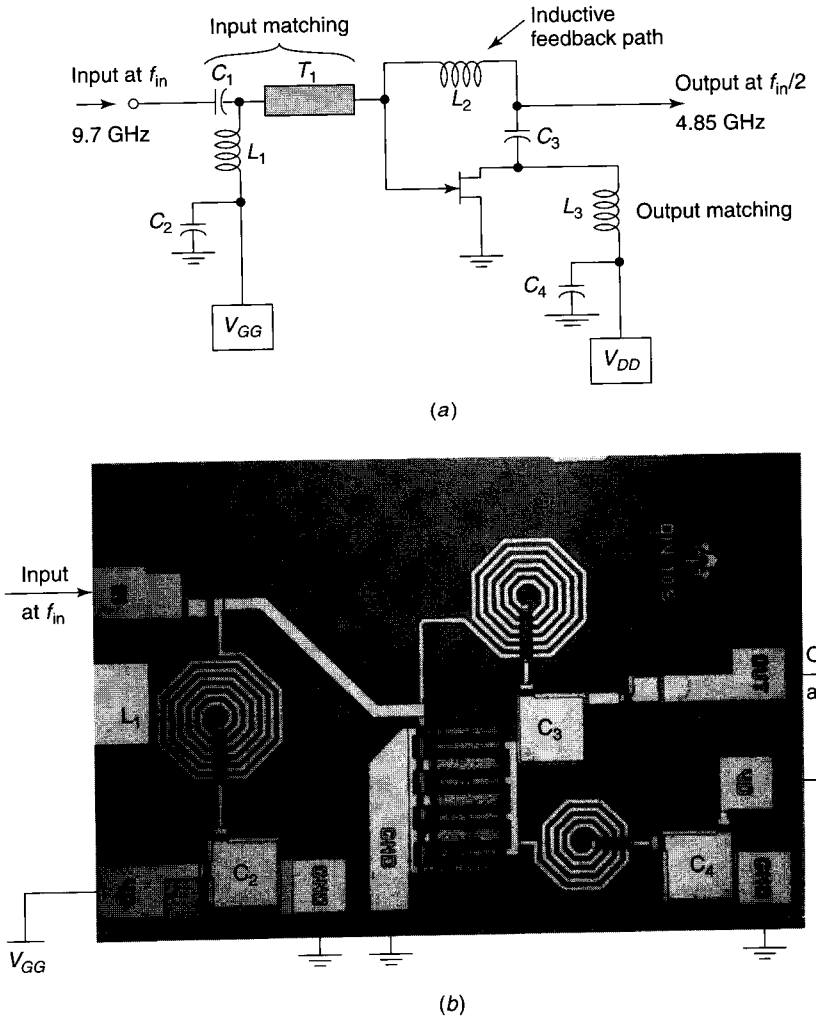
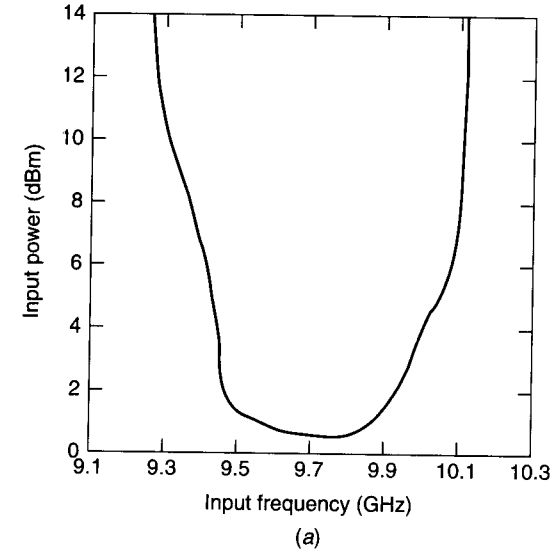


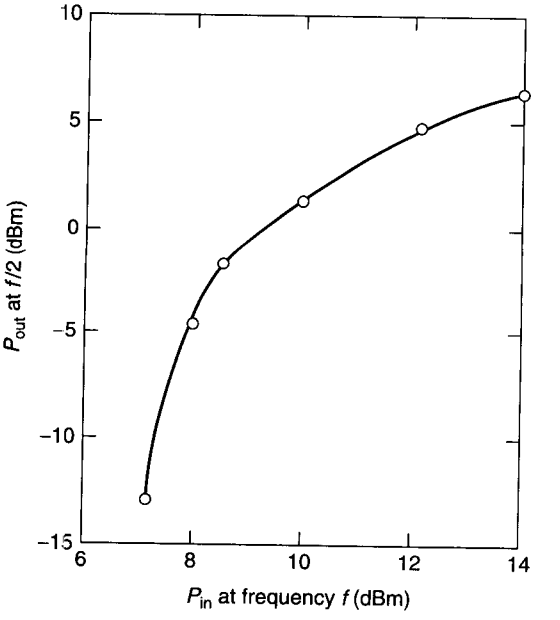
Figure 13.45 (*a*) Schematic of MMIC GaAs FET regenerative frequency halver. (*b*) Micrograph of MMIC RFD. Chip size is 1 × 0.65 mm. (After Stubbs et al. [104]. Adapted with permission of IEEE.)

tion functions; see Table 13.4. Some of these [107–109] use quasi-digital circuit techniques to create multiple-active-device RFDs with bandwidths up to 100%. They can operate at frequencies above what is possible from a conventional "digital" flip-flop divider in the same technology [107].

RFDs should be designed not to oscillate in the absence of an input signal; this can be ensured by reducing the positive feedback until oscillation cannot



(*a*)



(*b*)

Figure 13.46 (*a*) Region of frequency division of GaAs MMIC regenerative halver. Minimum input power is $P_{in} = 0.5$ dBm; fractional bandwidth is 8.2% at $P_{in} = +12$ dBm. (*b*) Amplitude transfer characteristic of halver at input frequency $f = 9.5$ GHz. (After Stubbs et al. [104]. Adapted with permission of IEEE.)

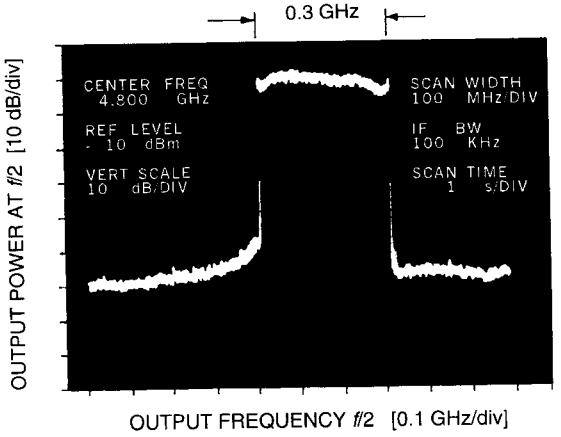


Figure 13.47 Frequency response of GaAs MMIC regenerative frequency halver. P_{in} is constant at +4 dBm and $f/2$ spectral line is displayed while f is swept over 0.6 GHz bandwidth. Note the flat top and the steep edges of response. (After Stubbs et al. [104]. Adapted with permission of IEEE.)

occur [100, 104]. If oscillation does take place, then the circuit may behave as an injection-locked oscillator (ILO) divider; see the next section.

13.3.4 Injection-Locked Oscillator Frequency Dividers

Injection-locked oscillator frequency dividers (ILOFDs) are similar to RFDs except that they act as free-running oscillators in the absence of an input signal. Where this is not a disadvantage ILOFDs can provide large conversion gain at the expense of narrow operational bandwith, which is approximately inversely proportional to the oscillator Q. The ILOFDs are possible because the frequency of an oscillator can be synchronized by a small-amplitude injected signal near a harmonic of the natural frequency f_0.

The ILOFDs were described as early as 1932 [115]. The Adler–Kurokawa theory [116, 117] treats injection locking in microwave solid state oscillators but not frequency division. Tunnel diode ILOFDs at ~7 GHz could provide frequency division by 9 with conversion gain ~38 dB (118); those based on IMPATT and Gunn diode oscillators [119, 120], division up to 6 with conversion gain ~20 dB. More recent work uses a GaAs MESFET oscillator as an ILOFD [121].

Theory. For a simplified theory of injection-locked oscillator (ILO) frequency division, consider an ILO described by a differential equation of the form

$$\frac{d^2y}{d\tau^2} + \frac{1}{Q}\frac{d}{d\tau}[n(y)] + y = vX\cos(v\tau) \tag{13.56}$$

Table 13.4 Multiple-Active-Device RFDS

Devices	Technology	Division Ratio	Input Frequency (GHz)	$\Delta f/f$ (%)	Gain (dB)	Reference (in date order)
1 Dual-gate, 2 single-gate GaAs FETs	GaAs MMIC	4:1	~9	23	−5	[110]
2 Si Darlington pairs	Si MMIC	2:1	~13.5	20	~+10	[111]
22 bipolar transistors, 2 diodes	Si MMIC	2:1	1.5–5	100	N/A	[107]
17 bipolar transistors, 2 diodes	Si MMIC	2:1 and 8:1	2–7.3	87	N/A	[108]
21 bipolar transistors	Si MMIC	16:1	28	80	N/A	[109]
2 HEMTs	Microstrip	2:1	48	3–5	−10 to −13	[112]
2 GaAS MESFETs	GaAS MMIC	2:1	3.7–4.3	15	+9	[113]
GaAs FETs	Low-power GaAs MMIC	6:1	9.7–10.7	9.8	−5	[114]

where $y = v/V_0$ is the normalized oscillator voltage
$X = V/V_0$ is the normalized injected voltage
$\tau = \omega_0 t$ is the normalized time
$\nu = \omega/\omega_0$ is the normalized frequency
ω_0 is the free-running frequency
Q is the oscillator quality factor
$\boldsymbol{n}(y)$ is an N-shaped zero-memory nonlinear (ZMNL) function

This equation was studied by Hayashi et al. [122], who assumed an asymmetrical $\boldsymbol{n}(y) = -y + ay^2 + by^3$ and used "algebraic" harmonic balance to find solutions for division by 2 and 3. Minakova et al. [123] assumed a symmetrical $\boldsymbol{n}(y) = -y + by^3 + dy^5$ and found solutions for division by 2 and 4.

For a high-Q oscillator, the waveform $y(\tau)$ is close to sinusoidal, and the normalized frequency ν is 1.0. When $X > 0$ and ν is near an integer N, the oscillator can lock to $(1/N)$th of the input frequency, realizing the divide-by-N

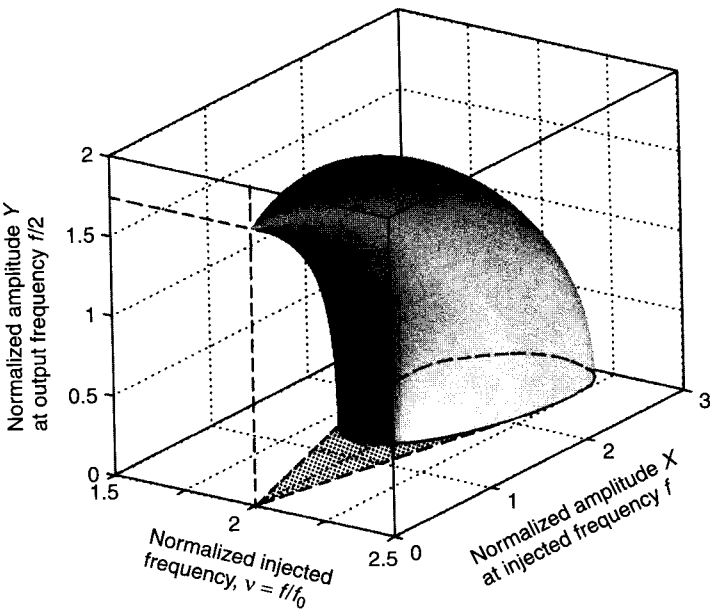


Figure 13.48 Theoretical $\frac{1}{2}$-frequency response surface of idealized ILO frequency divider. Parameters: $a = \frac{2}{3}$, $b = \frac{4}{9}$, $Q = 6.7$. The upper part of the surface corresponds to the stable solution, the lower part to the unstable solution. There is also a periodic response at the injected frequency (not shown). The maximum fractional bandwidth is $\Delta f/f \sim 10\%$. (After Hayashi et al. [122]. Adapted with permission of IEEE.)

function. An approximate solution is assumed:

$$y \cong \frac{X}{1-v^2}\cos(v\tau) + Y\cos\left(\frac{v\tau}{N}+\theta\right) \tag{13.57}$$

Halver Case. When $N = 2$, the amplitude Y of the divided-by-2 oscillation is given by [122]

$$\frac{3}{4}bY^2 = 1 - \frac{3}{2}b\left[\frac{vX}{1-v^2}\right]^2 \pm \sqrt{a^2\left[\frac{vX}{1-v^2}\right]^2 - Q^2\left[\frac{2}{v}-\frac{v}{2}\right]^2} \tag{13.58}$$

Figure 13.48 is a three-dimensional plot of this solution in (v, X, Y) space. The projection of this surface upon the (v, X) plane (shaded) shows combinations of injected frequency and power that can lead to stable frequency division by 2. For values of X below a certain threshold, this surface does not intersect the (v, X) plane. This means that conditionally stable frequency halving in this range is only possible if the initial conditions place the amplitude Y *inside* the appropriate part of the response surface.

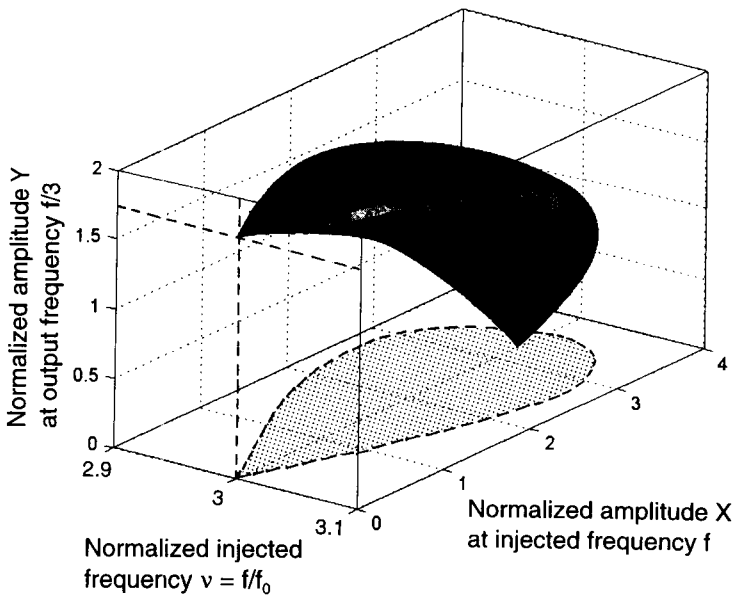


Figure 13.49 Theoretical $\frac{1}{3}$-frequency response surface of idealized ILO frequency divider. Parameters: $a = \frac{2}{3}$, $b = \frac{4}{9}$, $Q = 6.7$. The upper part of the surface corresponds to the stable solution, the lower part to the unstable solution. There is also a periodic response at the injected frequency. The maximum fractional bandwidth is $\Delta f/f \sim 5\%$. (After Hayashi et al. [122]. Adapted with permission of IEEE.)

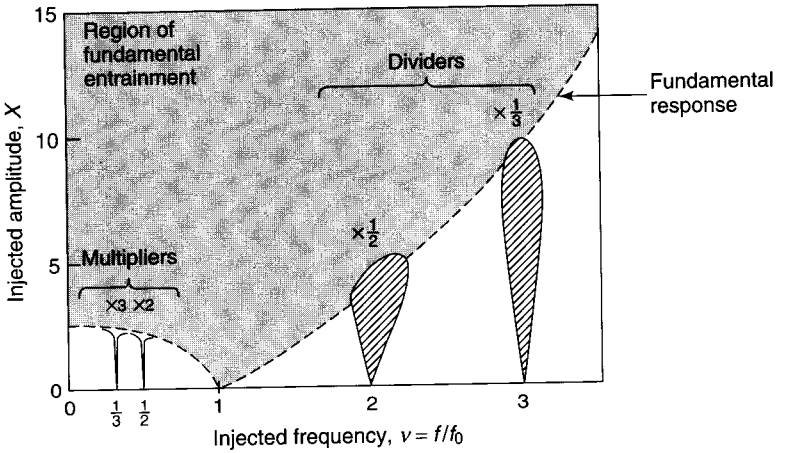


Figure 13.50 Regions of synchronization for idealized Hayashi ILO. (After Hayashi et al. [122]. Adapted with permission of IEEE.)

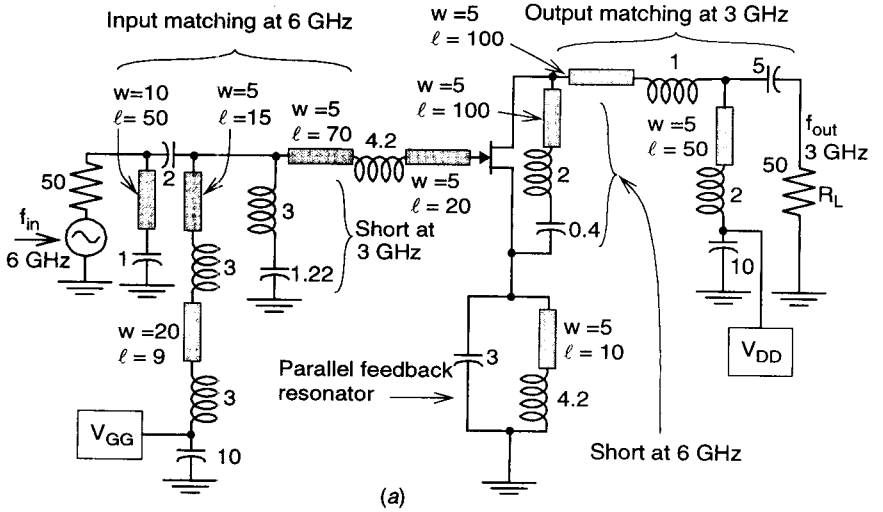
Thirder Case. In a similar fashion, setting $N = 3$, the amplitude solution found for the divide-by-3 case is

$$\frac{3}{4}bY^2 = \left(B + \frac{3}{8}bA^2\right) \pm \sqrt{\left(B + \frac{3}{8}bA^2\right)^2 - B^2 - Q^2\left(\frac{3}{v} - \frac{v}{3}\right)^2} \qquad (13.59)$$

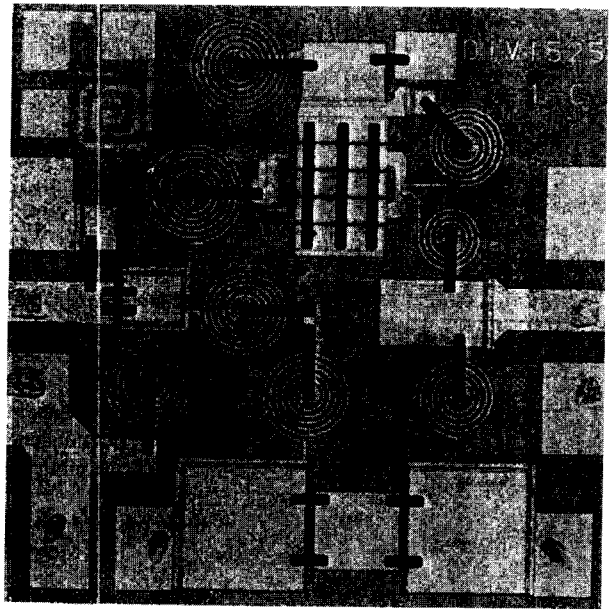
where $A = vX/(1 - v^2)$ and $B = 1 - (\frac{3}{2})bA^2$. The parameter a does *not* appear in this solution. This is to be expected since a is the coefficient of y^2 in the $\boldsymbol{n}(y)$ function. The response surface for the injection-locked oscillator thirder is completely disconnected from the (v, X) plane; see Fig. 13.49.

Regions of Synchronization. The regions of synchronization in the "injected-signal" (v, X) plane are shown in Fig. 13.50. At $v \sim \frac{1}{3}$ and $\frac{1}{2}$ there are regions of frequency *multiplication* by 2 and 3, respectively; at $v \sim 2$ and 3 are the regions of frequency *division* by 2 and 3. Above the broken-line boundary the oscillation is periodic at the injected frequency; below it the oscillation is quasi-periodic.

Practical ILOFD Frequency Halver. A divider of this type has been investigated in detail by Quéré et al. [121, 124, 125]. Figure 13.51*a* is the equivalent circuit of the 6- to 3-GHz GaAs MMIC divider shown in Fig. 13.51*b*. The behavior of this divider was investigated by means of a full harmonic-balance simulation and found to agree with meaurement. The results are summarized in Figure 13.52. The region of halving in (*a*) corresponds to the "$\times\frac{1}{2}$" region in Fig. 13.50, while (*c*) shows cross sections of the response surface at three values of injected power. The similarities between the simple algebraic theory and the actual behavior of the MMIC are apparent.

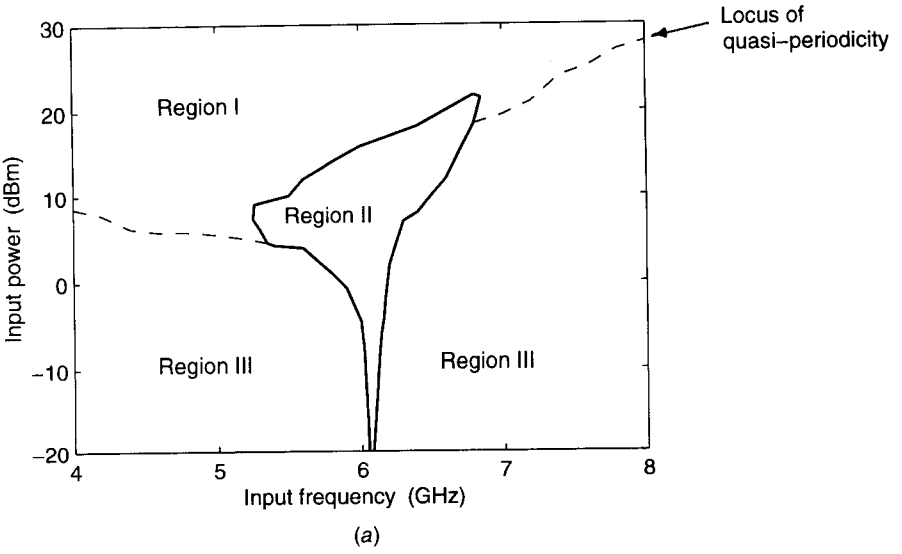


(*a*)

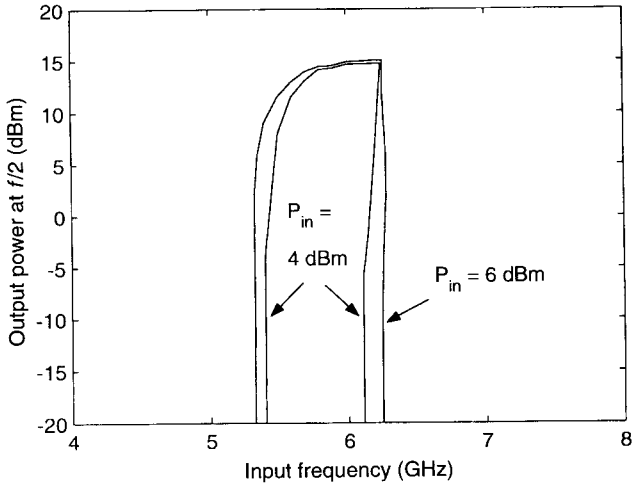


(*b*)

Figure 13.51 (*a*) Equivalent circuit of a monolithic injection-locked-oscillator divider by 2 using GaAS MESFET in low-Q 3-GHz oscillator circuit. Capacitances in pF, inductances in nH, resistances in Ω, and lengths in μm. Transistor has 0.5 × 150-μm gate fingers. (*b*) Photograph of monolithic GaAs chip. (After Quéré et al. [121]. Reprinted with permission of IEEE.)



(*a*)



(*b*)

Figure 13.52 (*a*) Region of division by 2 for practical ILOFD using low-Q oscillator with natural frequency $f_0 \sim 3$ GHz. In region I there is fundamental entrainment, as in Fig. 13.50. In region II division by 2 takes place (compare with $\times\frac{1}{2}$ in Fig. 13.50); in region III there is a quasiperiodic oscillation. (*b*) Amplitude response curves at two different injected power levels. Fundamental response at injected frequency not shown. (After Suarez et al. [124] and Morales et al. [125]. Reprinted with permission of IEEE.)

REFERENCES

1. Mott, N. F., "Note on the Contact between a Metal and an Insulator or Semiconductor," *Proc. Cambridge Phil. Soc.*, Vol. 34, 1939, pp. 568–572.
2. Maas, S. A., *Microwave Mixers*, Artech House, Dedham, MA, 1986, pp. 35–38.
3. Page, C. H., "Harmonic Generation with Ideal Rectifiers," *Proc. IRE*, Vol. 46, No. 10, Oct. 1958, pp. 1738–1740.
4. Harrison, R. G. "A Nonlinear Theory of Class C Transistor Amplifiers and Frequency Multipliers," *IEEE J. Solid-State Circuits*, Vol. SC-2, No. 3, Sept. 1967, pp. 93–102.
5. Penfield, P., and R. P. Rafuse, *Varactor Applications*, MIT Press, Cambridge, MA, 1962.
6. Miller, R. L., "Fractional-Frequency Generators Utilizing Regenerative Modulation," *Proc. IRE*, Vol. 27, No. 7, Jul. 1939, pp. 446–457.
7. Derksen, R. H., V. Lück, and H.-M. Rein, "Stability Ranges of Regenerative Frequency Dividers Employing Double Balanced Mixers in Large-Signal Operation," *IEEE Trans. Microwave Theory Tech.*, Vol. 39, No. 10, Oct. 1991, pp. 1759–1761.
8. Manley, J. M., and H. E. Rowe, "Some General Properties of Nonlinear Elements—Part I. General Energy Relations," *Proc. IRE*, Vol. 44, No. 7, Jul. 1956, pp. 904–913.
9. Hunt, E. R., and G. Johnson, "Keeping Chaos at Bay," *IEEE Spectrum*, Nov. 1993, pp. 32–36.
10. Kollberg, E. L., and A. Rydberg, "Quantum-Barrier-Varactor Diode for High-Efficiency Millimeter-Wave Multipliers," *Electron. Lett.*, Vol. 25, 1989, pp. 1696–1697.
11. Krishnamurthi, K., S. M. Nielsen, and R. G. Harrison, "GaAs Single-Barrier Varactors for Millimeter-Wave Triplers: Guidelines for Enhanced Performance," *IEEE Trans. Microwave Theory Tech.*, Vol. 42, No. 12, Dec. 1994, pp. 2512–2516.
12. Räisänen, A. V., "Frequency Multiplier for Millimeter and Submillimeter Wavelengths," *Proc. IEEE*, Vol. 80, No. 11, Nov. 1992, pp. 1842–1852.
13. Bishop, W. L., K. McKinney, R. J. Mattauch, T. W. Crowe, and G. Green, "A Novel Whiskerless Schottky Diode for Millimeter and Submillimeter Wave Application," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. II, Las Vegas, NV, June 1987, pp. 607–610.
14. Sloan, G. R., "The Modeling, Analysis, and Design of Filter-Based Parametric Frequency Dividers," *IEEE Trans. Microwave Theory Tech.*, Vol. 41, No. 2, Feb. 1993, pp. 224–228.
15. Payne, J. B., "Recent Advances in Solid-State Phase Locked Microwave Signal Sources: Part I," *Microwave Syst. News*, Feb./Mar. 1976, pp. 79–87.
16. Payne, J. B., "Recent Advances in Solid-State Phase Locked Microwave Signal Sources: Part II," *Microwave Syst. News*, Apr./May 1976, pp. 118–129.
17. Rodwell, M. J. W., S. T. Allen, R. Y. Yu, M. G. Case, M. Reddy, E. Carman, J. Pusl, M. Kamegawa, Y. Konishi, and R. Pullela, "Active and Nonlinear Wave Propagation Devices in Ultrafast Electronics and Optoelectronics," *Proc. IEEE*, Vol. 82, No. 7, July 1994, pp. 1037–1058.

18. Kollberg, E. L., T. J. Tolmunen, M. A. Frerking, and J. R. East, "Current Saturation in Submillimeter Wave Varactors," *IEEE Trans. Microwave Theory Tech.*, Vol. 40, No. 5, May 1992, pp. 831–838.
19. Page, C. H., "Frequency Conversion with Positive Nonlinear Resistance," *J. Res. Natl. Bur. Stand.*, Vol. 56, Apr. 1956, pp. 179–182.
20. Page, C. H., "Harmonic Generation with Ideal Rectifiers," *Proc. IRE*, Vol. 46, No. 10, Oct. 1958, pp. 1738–1740.
21. Pantell, R. H., "General Power Relationships for Positive and Negative Nonlinear Resistive Elements," *Proc. IRE*, Vol. 46, No. 12, Dec. 1958, pp. 1910–1913.
22. Clay, R., *Nonlinear Networks and Systems*, Wiley-Interscience, New York, 1971.
23. Benson, F. A., and F. Winder, "Nonlinear-Resistance Harmonic Generators," *Electron. Lett.*, Vol. 3, No. 12, Dec. 1967, pp. 534–535.
24. Zellers, J. R. "A 0.01- to 40-GHz Switched Frequency Doubler," *Hewlett-Packard J.*, Vol. 42, No. 2, Apr. 1991, pp. 31–33.
25. Benson, K., and M. A. Frerking, "Theoretical Efficiency for Triplers Using Non-ideal Varistor Diodes at Submillimeter Wavelengths," *IEEE Trans. Microwave Theory Tech.*, Dec. 1985, pp. 1367–1374.
26. Rauschenbach, K., C. A. Lee, and M. V. Schneider, "High-Frequency Efficient Reflection Multiplier," *IEEE Trans. Microwave Theory Tech.*, Vol. 39, No. 3, Mar. 1991, pp. 575–579.
27. Sayed, M. M., and G. F. Anderson, "50-to-110-GHz High-Performance Millimeter-Wave Source Modules," *Hewlett-Packard J.*, Vol. 42, No. 2, Apr. 1991, pp. 50–64.
28. Krishnamurthi, K., "Heterostructure Varactors on InP and GaAs for Millimeter-Wave Frequency Triplers," Ph.D. Dissertation, Carleton University, Ottawa, Canada, Jan. 1995.
29. Hwu, R. J., C. F. Jou, W. W. Lam, U. Lieneweg, D. C. Streit, N. C. Luhmann, Jr., J. Maserjian, and D. B. Rutledge. "Watt-Level Millimeter-Wave Monolithic Diode-Grid Frequency Multipliers," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. I, May 1988, pp. 533–536.
30. Peatman, W. C. B., T. W. Crowe, and M. Shur, "A Novel Schottky/2-DEG Diode for Millimeter- and Submillimeter-Wave Multiplier Applications," *IEEE Electron Dev. Lett.*, Vol. 13, No. 1, Jan. 1992, pp. 11–13.
31. Tang, C. C. H., "An Exact Analysis of Varactor Frequency Multipliers," *IEEE Trans. Microwave Theory Tech.*, Vol. 14, No. 4, Apr. 1966, pp. 210–212.
32. Sard, E. W., "Closed-Form Analysis of the Abrupt-Junction Varactor Doubler," *IEEE Trans. Microwave Theory Tech.*, Vol. 27, No. 6, June 1979, pp. 604–611.
33. Burckhardt, C. B., "Analysis of Varactor Frequency Multipliers for Arbitrary Capacitance Variation and Drive Level," *Bell Syst. Tech. J.*, Vol. 44, Apr. 1965, pp. 675–692.
34. Rafuse, R. P., "Recent Developments in Parametric Multipliers," *Proc. National Electronics Conference*, Vol. 19, Oct. 1963, pp. 461–470.
35. Archer, J. W., and R. A. Batchelor, "Multipliers and Parametric Devices," in K. Chang (Ed.), *Handbook of Microwave and Optical Components, Vol. 2: Microwave Solid State Components*, Wiley-Interscience, New York, 1990, pp. 142–191.

36. Gwarek, W. K., "Nonlinear Analysis of Microwave Mixers," M.S. Thesis, Massachusetts Instiute of Technology, Cambridge, MA, Sept. 1974.

37. Siegel, P. H., and A. R. Kerr, "Computer Analysis of Microwave and Millimeter Wave Mixers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, No. 3, Mar. 1980, pp. 275–276.

38. Bava, E., G. P. Bava, A. Godone, and G. Rietto, "Analysis of Varactor Frequency Multipliers: Nonlinear Behaviour and Hysteresis Phenomena," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, No. 2, Feb. 1979, pp. 141–147.

39. Bava, E., G. P. Bava, A. Godone, and G. Rietto, "Transfer Functions of Amplitude and Phase Fluctuations and Additive Noise in Varactor Doublers," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, No. 8, Aug. 1979, pp. 753–757.

40. de la Fuente, J. G., V. Krozier, and F. M. Martos, "Modeling and Design Aspects of Millimeter-Wave Schottky Varactor Frequency Multipliers," *IEEE Microwave Guided Wave Lett.*, Vol. 8, No. 11, Nov. 1998, pp. 387–398.

41. Chu, A., W. E. Courtney, L. J. Mahoney, R. W. McLelland, and H. A. Atwater, "GaAs Monolithic Frequency Doublers with Series Connected Varactor Diodes," *IEEE Microwave Millimeter-Wave Monolithic Circuits Symp. Dig.*, San Francisco, CA, May 1984, pp. 74–77.

42. Staecker, P. W., M. E. Hines, F. Occhiutti, and J. F. Cushman, "Multi-Watt Power Generation at Millimeter-Wave Frequencies Using Epitaxially-Stacked Varactor Diodes," *IEEE Int. Microwave Symp. Dig.*, Vol. II, June 1987, pp. 917–920.

43. Swan, C. B., "Design and Evaluation of a Microwave Varactor Tripler," *IEEE Int. Solid-State Circuits Conf. Dig.*, Feb. 1965, pp. 106–107.

44. Staecker, P. W., and R. W. Chick, "10 to 40 GHz Doubler-Doubler Chain for Satcom Applications," *Microwave J.*, Vol. 28, No. 12, Dec. 1985, pp. 87–108.

45. Horton, J. B., "A Thin-Film X-Band Varactor Quadrupler," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-15, No. 12, Dec. 1967, pp. 752–754.

46. Faber, M. T., and J. W. Archer, "A High-Efficiency Frequency Doubler for 100 GHz," *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1985, pp. 363–365.

47. Chen, S.-W., T. C. Ho, K. Pande, and P. D. Rice, "Rigorous Analysis and Design of a High-Performance 94 GHz MMIC Doubler," *IEEE Trans. Microwave Theory Tech.*, Vol. 1, No. 12, Dec. 1993, pp. 2317–2322.

48. Papapolymerou, J., F. Brauchler, J. East, and L. P. B. Katehi, "*W*-Band Finite Ground Coplanar Monolithic Multipliers," *IEEE Trans. Microwave Theory Tech.*, Vol. 47, No. 5, May 1999, pp. 614–619.

49. East, J., E. Kollberg, and M. Frerking, "Performance Limitations of Varactor Multipliers," *Proc. 4th Int. Symp. Space Terahertz Technol.*, Mar. 1993, pp. 312–325.

50. Luettgenau, G., J. Williams, and H. Miyahira, "A Practical Approach to the Design of Parametric Frequency Multipliers," TRW Bulletin No. 92-562, Pacific Semiconductors Inc., Lawndale CA, 1962.

51. Brauchler, F., "Finite Ground Coplanar Passive and Active Circuits," Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 1996.

52. Kollberg, E. L., and A. Rydberg, "Quantum-Barrier-Varactor Diode for High-Efficiency Millimeter-Wave Multipliers," *Electron. Lett.*, Vol. 25, 1989, pp. 1696–1697.

53. Jones, J. R., S. H. Jones, G. B. Tait, and M. F. Zybura, "Heterostructure Barrier Simulation Using an Integrated Hydrodynamic Device/Harmonic-Balance Circuit Analysis Technique," *IEEE Microwave Guided Wave Lett.*, Vol. 4, No. 12, Dec. 1994, pp. 411–413. Correction: Vol. 5, No. 2, Feb. 1995, p. 62.

54. Peatman, W. C. B., T. W. Crowe, and M. Shur, "A Novel Schottky/2-DEG Diode for Millimeter- and Submillimeter-Wave Multiplier Applications," *IEEE Electron Dev. Lett.*, Vol. 13, No. 1, Jan. 1992, pp. 11–13.

55. Koh, P. J., W. C. B. Peatman, and T. W. Crowe, "Millimeter Wave Tripler Evaluation of a Metal/2-DEG Schottky Diode Varactor," *IEEE Microwave Guided Wave Lett.*, Vol. 5, No. 2, Mar. 1995, pp. 73–75.

56. Lieneweg, U., B. R. Hancock, and J. Maserjian, "Barrier-Intrinsic-N^+ (BIN) Diodes for Near-Millimeter Wave Generation," *Dig. 12th Int. Conf. Infrared and Millimeter Waves*, 1987.

57. Hwu, R. J., "A mm-Wave High Power Monolithic Back-to-Back Varactor Diode Frequency Tripler," *Microwave J.*, Vol. 36, No. 5, May 1993, pp. 240–253.

58. Lieneweg, U., T. J. Tolmunen, M. A. Frerking, and J. Maserjian, "Modeling of Planar Varactor Frequency Multiplier Devices with Blocking Barriers," *IEEE Trans. Mcirowave Theory Tech.*, Vol. 40, No. 5, May 1992, pp. 839–845.

59. Krishnamurthi, K., and R. G. Harrison, "Analysis of Symmetric-Varactor Frequency Triplers," *IEEE Int. Microwave Symp. Dig.*, Atlanta, GA, June 1993, pp. 649–652.

60. Krishnamurthi, K., S. M. Nielsen, and R. G. Harrison, "GaAs Single-Barrier Varactors for Millimeter-Wave Triplers: Guidelines for Enhanced Performance," *IEEE Trans. Microwave Theory Tech.*, Vol. 42, No. 12, Dec. 1994, pp. 2512–2516.

61. Kerr, A. R., "A Technique for Determining the Oscillator Waveforms in a Mixer," *IEEE Trans. Microwave Theory Tech.*, Oct. 1975, pp. 828–831.

62. Rydberg, A., H. Grönqvist, and E. Kollberg, "Millimeter- and Submillimeter-Wave Multipliers Using Quantum-Barrier-Varactor (QBV) Diodes," *IEEE Electron Dev. Lett.*, Vol. 11, No. 9, Sept. 1990, pp. 373–375.

63. Krishnamurthi, K., and R. G. Harrison, "Millimeter-Wave Frequency Tripling Using Stacked Heterostructure-Barrier Varactors on InP," *IEE Proc. Microwaves Antennas Propagation*, Vol. 143, No. 4, Aug. 1996, pp. 272–276.

64. Mélique, X., A. Maestrini, R. Farré, P. Mounaix, M. Favreau, O. Vanbésien, J.-M. Goutoule, F. Mollot, G. Beaudin, T. Nähri, and D. Lippens, "Fabrication and Performance of InP-Based Heterostructure Barrier Varactors in a 250-GHz Tripler," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 6, June 2000, pp. 1000–1006.

65. Sandberg, I. W., and G. J. J. van Zyl, "Designing Periodically Driven Varactor Circuits with Guaranteed Stability," *IEEE Microwave Guided Wave Lett.*, Vol. 9, No. 6, June 1999, pp. 230–232.

66. Krishnamurthi, K., E. Boch, and R. G. Harrison, "A Ka-Band Planar Tripler Based on a Stacked Symmetric InP Heterostructure-Barrier Varactor," *IEEE Int. Microwave Symp. Dig.*, Vol. II, May 1995, pp. 549–552.

67. Siegel, P. H., A. R. Kerr, and W. Hwang, "Topics in the Optimization of Millimeter-Wave Mixers," NASA Technical Paper 2287, 1984.

68. Mélique, X., C. Mann, P. Mounaix, J. Thornton, O. Vanbésien, F. Mollot, and D. Lippens, "5-mW and 5% Efficiency 216-GHz InP-Based Heterostructure Barrier Varactor Tripler," *Microwave Guided Wave Lett.*, Nov. 1998, pp. 384–386.

69. Mélique, X., A. Maestrini, E. Lheurette, P. Mounaix, M. Favreau, O. Vanbésien, J. M. Goutoule, G. Beaudin, T. Nähri, and D. Lippens, "12% Efficiency and 9.5 dBm Output Power from InP-based Heterostructure Barrier Varactor Triplers at 250 GHz," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. I, June 1999, pp. 123–126.

70. Gilmore, R., "Design of a Novel FET Frequency Doubler Using a Harmonic Balance Algorithm," *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1986, pp. 585–588.

71. Camargo, E., and F. S. Correra, "A High Gain GaAs MESFET Frequency Quadrupler," *IEEE MTT-S Int. Microwave Symp. Dig.*, Las Vegas, June 1987, pp. 177–180.

72. Hirota, T., and H. Ogawa, "Uniplanar Monolithic Frequency Doublers," *IEEE Trans. Microwave Theory Tech.*, Vol. 37, No. 8, Aug. 1989, pp. 1249–1254.

73. Campos-Roca, Y., L. Verweyen, M. Neumann, M. Fernández-Barciela, M. C. Currás-Francos, E. Sánchez-Sánchez, A. Hülsmann, and M. Schlechtweg, "Coplanar pHEMT MMIC Frequency Multipliers for 76-GHz Automotive Radar," *Microwave Guided Wave Lett.*, Vol. 9, No. 6, June 1999, pp. 242–244.

74. Zhang, X., and Y.-H. Yun, "A DC to X-Band Frequency Doubler Using GaAs HBT MMIC," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1997, pp. 1213–1216.

75. Harrison, R. G., "Frequency Multipliers and Dividers," in I. Bahl and P. Bhartia (Eds.), *Microwave Solid State Circuit Design*, Wiley, New York, 1988, Chap. 13.

76. Rauscher, C., "High Frequency Doubler Operation of GaAs FET Field Effect Transistors," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-31, No. 6, June 1983, pp. 462–473.

77. Guo, C., E. Ngoya, R. Quere, M. Camiade, and J. Obregon, "Optimal CAD of MESFET Frequency Multipliers with and without Feedback," *IEEE MTT-S Int. Microwave Symp. Dig.*, May 1988, pp. 1115–1118.

78. Hiraoka, T., T. Tokumitsu, and M. Akaike, "A Miniaturized Broad-Band MMIC Frequency Doubler," *IEEE Trans. Microwave Theory Tech.*, Vol. 38, No. 12, Dec. 1990, pp. 1932–1936.

79. Tuko, M. A., and I. Wolff, "Novel 36 GHz GaAs Frequency Doublers Using (M)MIC Coplanar Technology," *IEEE Int. Microwave Symp. Dig.*, Vol. III, June 1992, pp. 1167–1170.

80. Chow, P. D., Y. Hwang, D. Garske, J. Velebir, and H. C. Yen, "A 44 GHz Doubler/Amplifier Chain," *IEEE Int. Microwave Symp. Dig.*, Vol. I, May 1990, pp. 603–606.

81. Boudiaf, A., D. Bachelet, and C. Rumelhard, "A High-Efficiency and Low-Phase-Noise 38-GHz pHEMT MMIC Tripler," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 12, Dec. 2000, pp. 2546–2553.

82. Zhang, G., R. D. Pollard, and C. M. Snowden, "A Novel Technique for HEMT Tripler Design," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. II, June 1996, pp. 663–666.

83. Thibaud, B., D. Barataud, M. Campovecchio, J. M. Nebus, S. Tranchant, P. Quentin, and D. Floriot, "CAD Oriented Design Methods of Frequency Multipliers: Application to a Millimeterwave MMIC PHEMT Tripler and a Microwave HBT Doubler," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. IV, June 1999, pp. 1747–1750.

84. Angelov, I., H. Zirath, N. Rorsman, and H. Grönqvist, "A Balanced Millimeter Wave Doubler Based on Pseudomorphic HEMTs," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. I, June 1992, pp. 353–356.

85. Yanev, A. S., B. N. Todorov, and V. Z. Ranev, "A Broad-Band Balanced HEMT Frequency Doubler in Uniplanar Technology," *IEEE Trans. Microwave Theory Tech.*, Vol. 46, No. 12, Dec. 1998, pp. 2032–2035.

86. Fudem, H., and E. C. Niehenke, "Novel Millimeter Wave Active MMIC Triplers," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. II, June 1998, pp. 387–390.

87. De, V. K., and J. D. Meindl, "Analytical Drain Current Models for AlGaAs/GaAs MODFET's Including Subthreshold Conduction," *IEEE J. Solid-State Circuits*, Vol. 28, No. 5, May 1993, pp. 596–604.

88. O'Ciardha, E., S. U. Lidholm, and B. Lyons, "Generic-Device Frequency Multiplier Analysis—a Unified Approach," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 7, July 2000, pp. 1134–1141.

89. Wang, G.-W., I. Ichitsubo, W. H. Ku, Y.-K. Chen, and L. F. Eastman, "Large-Signal Time-Domain Simulation of HEMT Mixers," *IEEE Trans. Microwave Theory Tech.*, Vol. 36, No. 4, Apr. 1988, pp. 756–758.

90. Microwave Design System (MDS), Hewlett-Packard Inc. (Now replaced by Advanced Design System, Agilent Corp.)

91. Kalivas, G. A., and R. G. Harrison, "A New Slotline-Microstrip Frequency Halver," *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1985, pp. 683–686.

92. Rizzoli, V., and A. Lipparini, "General Stability Analysis of Periodic Steady State Regimes in Nonlinear Microwave Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. 33, No. 1, Jan. 1985, pp. 30–37.

93. Collantes, J.-M., and A. Suárez, "Period-Doubling Analysis and Chaos Detection Using Commercial Harmonic Balance Simulators," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 4, Apr. 2000, pp. 574–581.

94. Ryan, W. D., and H. B. Williams, "The Carrier-Storage Frequency Divider: A Steady-State Analysis," *IEEE Trans. Circuit Theory*, Vol. 11, Sept. 1964, pp. 396–403.

95. Huang, C. I., J. Tsui, R. T. Kemerley, and G. L. McCoy, "Dynamic Microwave Frequency Division Charactristics of Coplanar Transferrred-Electron Devices in a Reistive Circuit," *IEEE Trans. Microwave Theory Tech.*, Vol. 24, No. 1, Jan. 1976, pp. 61–63.

96. Suárez, A., and J. Morales, "Chaos Prediction in an MMIC Frequency Divider in Millimetric Band," *IEEE Microwave Guided Wave Lett.*, Vol. 8, No. 1, Jan. 1998, pp. 21–23.

97. Sloan, G. R., "The Modeling, Analysis, and Design of Filter-Based Parametric Frequency Dividers," *IEEE Trans. Microwave Theory Tech.*, Vol. 41, No. 2, Feb. 1993, pp. 224–228.

98. Harrison, R. G., "Theory of the Varactor Frequency Halver," *IEEE MTT-S Int. Microwave Symp. Dig.*, May 1983, pp. 203–205.

99. Miller, R. L., "Fractional-Frequency Generators Utilizing Regenerative Modulation," *Proc. IRE*, Vol. 27, No. 7, July 1939, pp. 446–457.

100. Rauscher, C., "Regenerative Frequency Division with a GaAs FET," *IEEE Trans. Microwave Theory Tech.*, Vol. 32, No. 11, Nov. 1984, pp. 1461–1468.

101. Ahamed, S. V., J. C. Irvine, and H. Seidel, "Study and Fabrication of a Frequency Divider-Multiplier Scheme for High-Efficiency Microwave Power," *IEEE Trans. Commun.*, Vol. COM-24, No. 2, Feb. 1976, pp. 243–249.

102. Derksen, R. H., V. Lück, and H.-M. Rein, "Stability Ranges of Regenerative Frequency Dividers Employing Double Balanced Mixers in Large-Signal Operation," *IEEE Trans. Microwave Theory Tech.*, Vol. 39, No. 10, Oct. 1991, pp. 1759–1761.

103. Ohira, T., K. Araki, T. Tanaka, and H. Kato, "14 GHz Band GaAs Monolithic Analogue Frequency Divider," *Electron. Lett.*, Vol. 21, No. 22, Oct. 1985, pp. 1057–1058.

104. Stubbs, M. G., and S. P. Stapleton, "A Single-Stage Monolithic Regenerative 1/2 Analog Frequency Divider," *IEEE Gallium Arsenide Integrated Circuit Symp. Dig.*, Grenelefe, FL, Oct. 1986, pp. 199–201.

105. Kaminsky, D., P. Goussu, R. Funck, and A. G. Bert, "A Dual-Gate GaAs FET Analog Frequency Divider," *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1983, pp. 352–354.

106. Kanzawa, K., M. Hagio, M. Kazumura, and G. Kano, "A 15 GHz Single-Stage GaAs Dual-Gate FET Monolithic Analog Frequency Divider with Reduced Input Threshold Power," *IEEE Trans. Microwave Theory Tech.*, Vol. 36, No. 12, Dec. 1988, pp. 1908–1912.

107. Derksen, R. H., H.-M. Rein, and K. Wörner, "Monolithic Integration of a 5.3 GHz Regenerative Frequency Divider Using a Standard Bipolar Technology," *Electron. Lett.*, Vol. 21, No. 22, Oct. 1985, pp. 1037–1039.

108. Derksen, R. H., and H.-M. Rein, "7.3 GHz Dynamic Frequency Dividers Monolithically Integrated in a Standard Bipolar Technology," *IEEE Trans. Microwave Theory Tech.*, Vol. 36, No. 3, Mar. 1988, pp. 537–541.

109. Kurisu, M., G. Uemura, M. Ochuchi, H. Takemura, T. Morikawa, and T. Tashiro, "A Si Bipolar 28-GHz Dynamic Frequency Divider," *IEEE J. Solid-State Circuits*, Vol. 27, No. 12, Dec. 1992, pp. 1799–1804.

110. Madihian, M., and K. Honjo, "11 GHz Band Monolithic VCO with 1/4 Analog Frequency Divider," *IEEE Gallium Arsenide Integrated Circuit Symposium Dig.*, Monterey CA, Nov. 1985, pp. 133–136.

111. Kipnis, I., "Silicon Bipolar MMIC for Frequency Conversion Applications up to 20 GHz," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. II, June 1987, pp. 855–858.

112. Angelov, I., H. Zirath, N. Rorsman, and C. Karlsson, "48/24 GHz and 20/10 GHz Regenerative Frequency Dividers," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. II, June 1996, pp. 971–974.

113. Branger, B., E. LaPorte, J. C. Nallatamby, M. Prigent, and L. LaPierre, "Comparison between Simulations and Measurements of Large Signal and Nonlinear

Noise Behaviors of MMIC Analog Frequency Divider by Two," *IEEE Microwave Guided Wave Lett.*, Vol. 7, No. 5, May 1997, pp. 127–129.

114. Ohira, T., and Y. Suzuki, "Novel Frequency Division Technique for Very Low Power GaAs Monolithic Microwave Prescalers," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. I, June 1999, pp. 193–196.
115. Mandelstam, L., and N. Papalexi, "Über Rezonanzerscheinungen bei Frequenzteilung," *Zeit. Phys.*, Vol. 73, 1932, pp. 223–248.
116. Adler, R., "A Study of Locking Phenomena in Oscillators," *Proc. IRE*, Vol. 34, Jun. 1946, pp. 351–357. Reprinted in *Proc. IEEE*, Vol. 61, No. 10, Oct. 1973, pp. 1380–1385.
117. Kurokawa, K., "Injection Locking of Microwave Solid State Oscillators," *Proc. IEEE*, Vol. 61, No. 10, Oct. 1973, pp. 1386–1410.
118. Mashock, J. A., and T. K. Ishii, "An Interesting Pulling Effect in a Tunnel Diode Oscillator," *IEEE Trans. Microwave Theory Tech.*, Vol. 17, No. 6, June 1969, pp. 347–348.
119. Tager, A. S., and A. M. Tsebiyev, "Synchronization of a Microwave Avalanche-Transit Time Diode Oscillator by an External Harmonic Signal," *Radio Eng. Electron. Phys.* (former USSR), Vol. 14, No. 7, 1969, pp. 1110–1116.
120. Huang, C. I., J. Tsui, R. T. Kemerley, and G. L. McCoy, "Dynamic Microwave Frequency Division Characteristics of Coplanar Transferred-Electron Devices in a Resistive Circuit," *IEEE Trans. Microwave Theory Tech.*, Vol. 24, No. 1, Jan. 1976, pp. 61–63.
121. Quéré, R., E. Ngoya, M. Camiade, A. Suarez, M. Hessane, and J. Obregon, "Large Signal Design of Broadband Monolithic Microwave Frequency Dividers and Phase-Locked Oscillators," *IEEE Trans. Microwave Theory Tech.*, Vol. 41, No. 11, Dec. 1993, pp. 1928–1938.
122. Hayashi, C., H. Shibayama, and Y. Nishikawa, "Frequency Entrainment in a Self-Oscillatory System with External Force," *IRE Trans. Circuit Theory*, Vol. CT-7, No. 4, Dec. 1960, pp. 413–422.
123. Minakova, I. I., and A. G. Fedoseyev, "Parametric Synchronization of a Symmetrical Tunnel-Diode Oscillator," *Radio Eng. Electron. Phys.* (former USSR), Vol. 18, No. 2, Feb. 1973, pp. 220–225.
124. Suárez, A., J. C. Sarkassian, R. Sommet, E. Ngoya, and R. Quéré, "Stability Analysis of Analog Frequency Dividers in the Quasi-Periodic Regime," *IEEE Microwave Guided Wave Lett.*, Vol. 4, No. 5, May 1994, pp. 138–140.
125. Morales, J., A. Suárez, and R. Quéré, "Accurate Determination of Frequency Dividers Operating Bands," *IEEE Microwave Guided Wave Lett.*, Vol. 6, No. 1, Jan. 1996, pp. 46–48.

PROBLEMS

13.1 *Phase Noise Reduction.* Referring to Eqs. (13.1)–(13.3), show

(a) that when a reference signal at frequency ω_r is translated down to ω_r/N using a mixer, the worst-case timing error is $N\psi/\omega_r$, but

(b) that when the translation is done using frequency division, the peak timing error is only ψ/ω_r.

13.2 *Resistive Doubler: Exponential Diode.*

(a) Verify that the voltage transfer function of the resistive full-wave balanced doubler of Fig. 13.8*a* is $y = 2a(e^{-y}\cosh x - 1)$, where $x = \Lambda v_{\text{in}}$, $y = \Lambda v_{\text{out}}$, $a = \Lambda R_L I_s$. For this approximate calculation, neglect R_G and the diode resistance R_s.

(b) Assuming that $v_{\text{in}} = V_{\text{in}}\cos(\omega t)$, and that $V_{\text{in}} < nKT/q_e$, show that the output voltage v_{out} is approximately $2a[\mathbb{I}_0(\Lambda V_0) - 1] - 4a\mathbb{I}_1(\Lambda V_0)\cos(2\omega t) + 4a\mathbb{I}_2(\Lambda V_0)\cos(4\omega t)$, where $\mathbb{I}_k(x)$ is a modified Bessel function of the first kind, order k, and argument x.

(c) Sketch the transfer function, showing the graphical relationship between the $v_{\text{in}}(t)$ and $v_{\text{out}}(t)$ waveforms.

$$\textit{Hint:}\qquad \cosh(x\sin\theta) = \mathbb{I}_0(x) + 2\sum_{k=1}^{\infty}(-1)^k\mathbb{I}_{2k}(x)\cos(2k\theta)$$

13.3 *Resistive Doubler: PWL Diode.*

(a) Using the circuit topology of the previous problem, but assuming the ideal-diode model (13.8), that R_s is negligible, that $R_G \ll R_L$, and that $v_{\text{in}} \cong V_{\text{in}}\cos(\omega t)$, show that the average input power to the doubler is

$$P_{\text{in}} = \frac{V_{\text{in}}^2}{\pi R_L}(\arccos(b) - b\sqrt{1-b^2})$$

where $b = V/V_{\text{in}}$.

(b) Show that the average output power at the second harmonic is

$$P_{2\omega} = \frac{8V_{\text{in}}^2(1-b^2)^3}{9\pi^2 R_L}$$

where $b = V/V_{\text{in}}$, and hence obtain (13.10).

13.4 *Symmetric-Varactor Tripler.*

(a) For the shunt-mode tripler of Fig. 13.24, assume a symmetric heterostructure barrier varactor (HBV) conforming to the cubic-charge model (13.20), with series resistance R_s. Assuming (i) that the HBV is placed between ideal band-pass filters that pass currents at ω and 3ω only, (ii) that the HBV is not driven into breakdown, (iii) that a fixed P_{in} is supplied to the nonlinear $C_j(v)$ (excluding R_s), and (iv) that dissipation is minimized with respect to I_1, derive the optimization condition (13.25).

(b) Derive the input and output impedances as given by (13.28) and (13.29).

13.5 *Varactor Devices and Circuits.*

(a) Why are idlers used in certain parametric frequency conversion circuits but not in others?

(b) What are the physical origins of the the $C_j(v)$ characteristics of classical Schottky varactors and of heterostructure barrier varactors?

(c) For what types of frequency multipliers are HBVs preferable to classical varactors? Explain.

13.6 *Classical Varactor Frequency Halver.*

(a) Using the notation of Section 13.3.2 and Fig. 13.39*a*, derive the coupled differential equations (DEs) (13.37) and (13.38) that model the even-mode excitation and odd-mode resonance of a balanced parametric frequency halver. Assume that the diodes have $\gamma = \frac{1}{2}$.

(b) By eliminating the sum-charge variable u between (13.37) and (13.38) obtain a single nonlinear DE in the difference charge variable z.

(c) Assuming that $z(\tau) \approx Z \cos(\nu\tau + \psi)$, that is, ignoring odd harmonics, solve for Z and ψ. Finally, assuming the normalized output voltage to be approximately $y(\tau) \approx Y \cos(\nu\tau + \theta)$, obtain (13.39).

(d) Explain the limitations of this analysis.

13.7 *Frequency Dividers.* Compare the relative advantages and disadvantages (including bandwidth, maximum operational frequency, phase noise, insertion loss/gain, power dissipation, generation of spurious frequencies, and suitability for monolithic integration) of the following types of frequency dividers:

(a) varactor parametric dividers,

(b) regenerative frequency dividers,

(c) digital frequency dividers (static and dynamic),

(d) transferred-electron-device (Gunn) dividers, and

(e) injection-locked-oscillator dividers.

14

RF MEMS DEVICES AND CIRCUIT APPLICATIONS

Ramesh Ramadoss and K. C. Gupta

14.1 INTRODUCTION

A microelectromechanical system (MEMS) [1–3] integrates electrical and mechanical functions in a single component using microfabrication or micromachining technology similar to semiconductor integrated circuit processing. MEMS technology enables the integration of microelectronics with the actuation, sensing, and control capabilities of microsensors and microactuators. Typical sizes for MEMS components range from micrometers to millimeters. The characteristics of MEMS devices are low cost, high functionality, small size, and low weight compared to semiconductor devices. Cost and weight reduction is the result of utilizing semiconductor batch-processing techniques.

MEMS has proved to be a revolutionary technology in many application arenas, including accelerometers, pressure sensors, micro-optics, inkjet nozzles, optical scanners, and fluid pumps. MEMS has potential applications in many fields, including telecommunications, information storage, robotics, the aerospace industry, medicine, optics, and materials science.

MEMS devices employed in radio-frequency applications are termed RF MEMS. These are a new class of devices and components with low insertion loss, high isolation, high Q, small size, and low power consumption that enable new system capabilities. Applications of MEMS in RF technology can be broadly classified into two categories: active devices and passive components. In active devices, mechanical movement of microstructures (fabricated by micromachining processes) is used for functioning of the device (e.g., RF MEMS

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

switch, RF MEMS variable capacitor). In passive RF MEMS components, micromachining processes are used for fabrication of low-loss transmission lines and RF components (e.g., micromachined transmission lines, resonators, antennas).

This chapter deals with RF MEMS devices and their circuit applications. Section 14.2 describes fabrication processes and assembly techniques used in RF MEMS. This is followed by a section on mechanical structures and electrostatic actuators. RF MEMS devices and circuit applications are discussed in Sections 14.4 and 14.5, respectively.

14.2 RF MEMS FABRICATION AND ASSEMBLY

14.2.1 Fabrication Processes

The MEMS fabrication process involves microelectronics fabrication processes enhanced with specialized techniques and is generally termed "micromachining." Micromachining is the key feature of the MEMS technology. The most commonly used micromachining techniques for RF MEMS are surface micromachining, bulk micromachining, and LIGA* techniques. Surface micromachining is extensively used for fabrication of RF MEMS devices and components.

Surface micromachining [1–3] involves deposition, patterning, and etching of thin films on the surface of a substrate (such as silicon, glass, alumina, or metal). The sequence of steps involved in the fabrication of a cantilever structure by surface micromachining is shown in Fig. 14.1*a*. The concept of a sacrificial layer shown in step 2 in Fig. 14.1*a* is the basis for fabrication of freestanding microstructures by surface micromachining. A sacrificial (soluble or removable) layer is deposited on a silicon substrate (with an Si_3N_4 isolation layer) by chemical vapor deposition (CVD). The purpose of the sacrificial layer is to provide temporary support for the structural layer during the subsequent fabrication steps. The commonly used sacrificial layers include metals (e.g., Au, Al), ceramics (SiO_2 and Si_3N_4), plastics (e.g., photoresist, polymethyl methacrylate) and polyimides. Openings are etched entirely through the sacrificial layer to provide anchoring points for the structural layer. A thin film of structural material (typically polysilicon) is deposited and etched to define a cantilever structure. After etching the sacrificial layer, the patterned structure is separated from the substrate (except at the anchoring point) by the thickness of the removed sacrificial layer to form a freestanding cantilever structure. The process can be repeated to form a stack of structural layers of different thicknesses with sacrificial layers sandwiched between them. This multilayer surface micromachining process can be used to create complex MEMS structures. A

*LIGA is a German acronym for lithography (*LI*thographie), electroforming (*G*alvanoforming), and molding (*A*bformung) [2, 3].

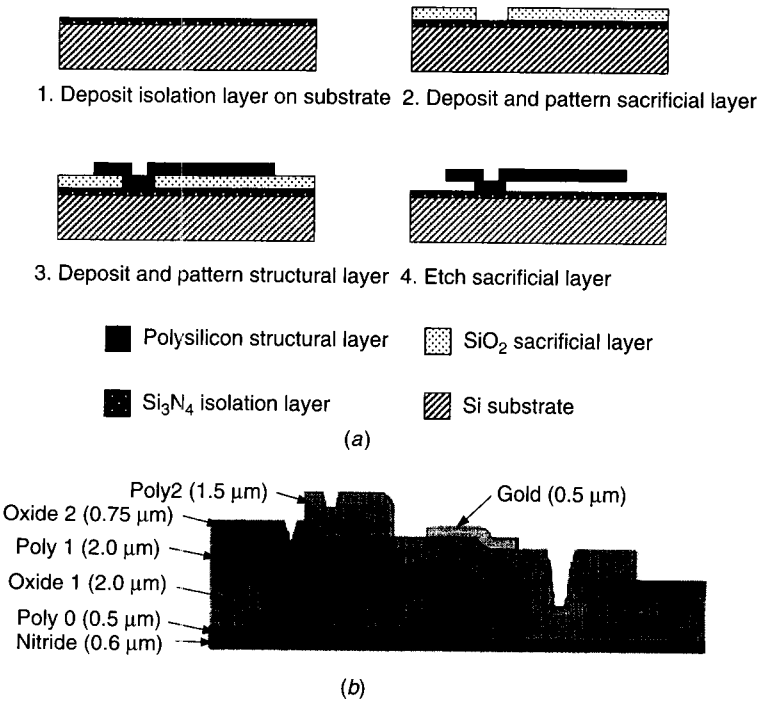


Figure 14.1 Surface micromachining process: (*a*) sequence of steps (1–4) in fabrication of a cantilever structure; (*b*) various layers in MEMS structures fabricated by MUMPs process.

three-layer polysilicon surface micromachining process, known as Multi-User MEMS Processes or MUMPs* [4], is useful for general-purpose design and fabrication of MEMS devices. The MUMPs fabrication process (see Fig. 14.1*b*) allows for a nitride layer, three polysilicon layers (poly 0, poly 1, and poly 2), two oxide layers (oxide 1 and oxide 2), and a gold layer; all deposited on a silicon wafer (100-mm diameter *n*-type ⟨100⟩ with 1–2 Ω-cm resistivity).

Bulk micromachining [1–3] is a process for making three-dimensional microstructures from a starting substrate (such as silicon, quartz, SiC, GaAs, InP, Ge, and glass) by selectively removing unwanted portions of the substrate. The most commonly used substrate material is silicon. In this process, microstructures are created by photolithography and etching techniques. Etching is the key process in bulk micromachining. Available etching methods fall into three categories in terms of the state of the etchant: wet, vapor, and plasma. Bulk micromachining based on wet chemical etching is widely used for fabrication of RF MEMS components. Wet bulk micromachining uses aqueous etchants in conjunction with etch masks and etch stops to shape the silicon

*MUMPs is a registered trademark of Cronos Integrated Microsystems (a JDS Uniphase company), Research Triangle Park, NC.

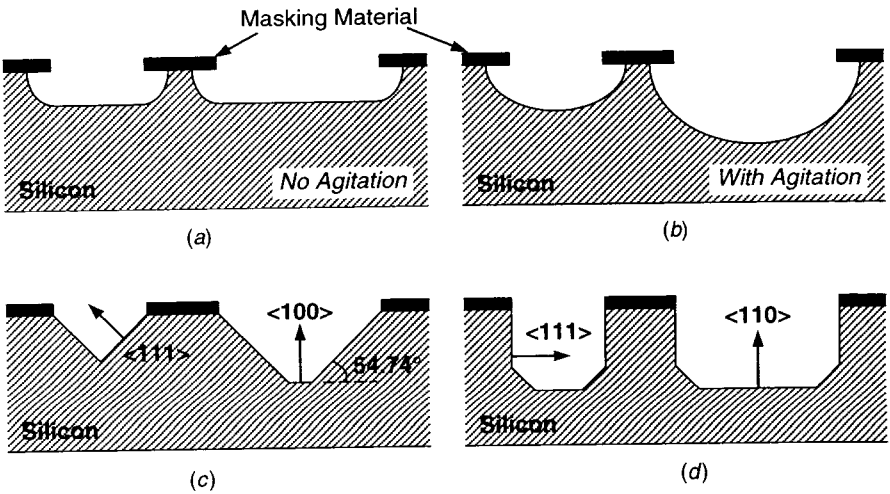


Figure 14.2 Profiles obtained by wet chemical etching processes in bulk micromachining: (*a*) isotropic etching; (*b*) isotropic etching with agitation; (*c*) anisotropic etching of ⟨100⟩ surface; (*d*) anisotropic etching of ⟨111⟩ surface.

substrate. Wet etching relies on the oxidation of silicon compounds that can be removed from the substrate. Wet etching can be isotropic, anisotropic, or a combination of both. In isotropic etching, the etch rate is uniform in all directions, as shown in Fig. 14.2*a*. Agitation of the substrate in isotropic etching can provide a round profile, as shown in Fig. 14.2*b*. In anisotropic etching, the etch rate depends on the crystallographic orientation of the substrate [1]. The crystal planes of silicon are ⟨100⟩, ⟨110⟩, and ⟨111⟩. The crystal plane ⟨111⟩ is at 54.74° to the ⟨100⟩ plane and perpendicular to the ⟨110⟩ plane. Anisotropic etching profiles (using etchants having slower etch rates at ⟨111⟩ planes relative to the other planes) on silicon wafers with surface orientations at ⟨100⟩ and ⟨110⟩ are shown in Fig. 14.2*c* and *d*, respectively. Wet etching processes can be made selective by doping the substrate to be etched with dopants. The heavily doped regions etch very slowly. In addition to wet bulk micromachining, dry etching techniques such as plasma etching, reactive-ion etching (RIE), ion beam etching, and laser etching are also employed for fabrication of MEMS microstructures.

LIGA process allows fabrication of tall three-dimensional microstructures with high aspect ratios—lateral dimensions of a few micrometers and vertical dimensions up to 1000 μm. LIGA involves X-ray lithography, microelectroplating, and micromolding processes. The use of highly collimated X rays for lithography allows the fabrication of structures with submicrometer resolution. The sequence of steps in the LIGA process is shown in Fig. 14.3. The process begins by generating a photoresist (typically polymethylmethacrylate) pattern several hundred micrometers thick by deep X-ray lithography on a conductive

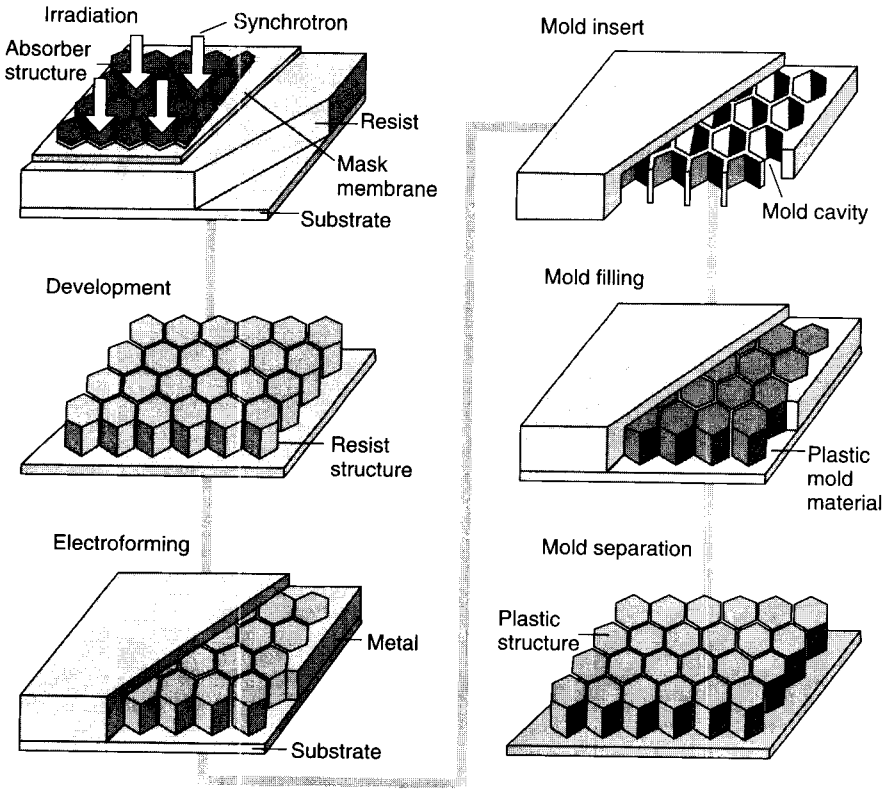


Figure 14.3 Sequence of fabrication steps in LIGA process.

substrate. This is followed by a removal of the exposed resist with a chemical developer and a backfill of the opened areas with metal by electroforming. Finally, the polymer resist is removed from the metal electroformed structure to provide a mold cavity. The metal mold cavity can be used to replicate microstructures by stamping, injection molding, or reaction injection molding. The commonly used materials include low-viscosity polymers such as polyimides, polymethylmethacrylate, other plastic resins, or even ceramic slurries. After curing, the mold is removed, leaving behind microreplicas of the original pattern. The drawback is the cost for the high-energy synchrotron equipment used as the X-ray source. The LIGA process has been used for fabrication of micromachined transmission lines, filters, and couplers [5, 6].

The above three fabrication processes are compared in Table 14.1 [7].

14.2.2 Assembly and Packaging Techniques

Assembly techniques are used for transferring MEMS devices (fabricated by one of the processes discussed in the previous section) to an RF circuit on a different substrate, or for fabrication of nonplanar MEMS structures.

Table 14.1 Comparison of MEMS Fabrication (Micromachining) Technologies

Capability	Bulk[a]	Surface	LIGA
Maximum structural thickness	Substrate thickness	<50 μm	500 μm
Planar geometry	Rectangular	Unrestricted	Unrestricted
Minimum planar feature size	$\sqrt{2} \times$ depth	1 μm	3 μm
Side-wall features	54.74° slope	Limited by dry etch	0.2 μm runout over 400 μm
Surface and edge definitions	Excellent	Mostly adequate	Very good
Material properties	Very well controlled	Mostly adequate	Well controlled
Integration with electronics	Demonstrated	Demonstrated	Difficult
Capital investment and costs	Low	Moderate	High

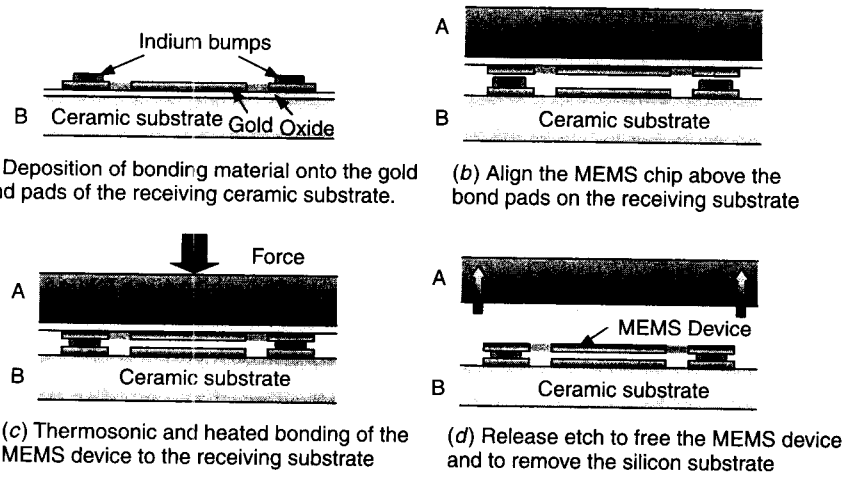
[a]Wet, anisotropic etching ⟨100⟩ silicon wafer.

Source: Tang [7].

The most commonly employed assembly techniques for RF MEMS are flip-chip assembly, solder self-assembly, and wafer-level assembly. These assembly techniques are briefly discussed in this section.

Flip-Chip Assembly. Flip-chip assembly enables transfer of unreleased MEMS devices (with sacrificial layer intact) fabricated by micromachining techniques onto an existing circuitry or onto a high-quality RF substrate such as a ceramic [8, 9]. The process sequence of flip-chip assembly is described in Fig. 14.4. The MEMS devices on substrate A are fabricated with gold bond pads. The RF circuit with bonding pads for MEMS devices is fabricated separately on a substrate B (like alumina). Indium or other solder alloy is deposited to a desired thickness onto the bonding pads on the circuit substrate B. The bond pads on the silicon substrate and on the target substrate are aligned. The entire structure is then thermosonically bonded together by melting and resolidification of the solder. The flip-chip bonded MEMS devices are released to remain in the circuit substrate by etching out the silicon dioxide layer (sacrificial layer) between the MEMS chip and the silicon substrate using hydrofluoric acid. The silicon substrate can then be removed from the MEMS assembly. The flip-chip assembly process has been used for fabrication of RF MEMS switches [10] and RF MEMS tunable capacitors [11, 12].

Solder Self-Assembly. This technique is used for precise assembly of structures using solder. The method utilizes surface tension properties of molten solder to assemble three-dimensional MEMS structures [13]. The solder self-



(*a*) Deposition of bonding material onto the gold bond pads of the receiving ceramic substrate.

(*b*) Align the MEMS chip above the bond pads on the receiving substrate

(*c*) Thermosonic and heated bonding of the MEMS device to the receiving substrate

(*d*) Release etch to free the MEMS device and to remove the silicon substrate

Figure 14.4 Typical steps in flip-chip assembly process.

assembly process is illustrated in Fig. 14.5. In this example, a substrate with metal solder pads and a hinged plate with a specific area metallized to form solder pads are fabricated. A solder ball is deposited at the junction of the hinged edge of the plate and the substrate solder pad and heated to its melting point. The surface tension force (produced by the natural tendency of molten solder to minimize the surface energy) pulls the hinged plate away from the substrate. The final angle between the substrate and the plate is determined by solder volume and solder pad sizes and can be precisely controlled to achieve submicrometer accuracy. Solder self-assembly is capable of large-scale precision alignments of structures in a single batch reflow process. Solder joints provide high-quality mechanical, thermal, and electrical characteristics. Microwave inductors have been fabricated [14] using this process.

Wafer-Level Assembly and Packaging. Wafer-level assembly is useful for packaging of RF MEMS components such as micromachined transmission lines, cavities, switches, and varactors. Wafer-level assembly enables permanent bonding of two or more wafers. The most frequently used wafer-level bonding techniques are silicon fusion bonding [15], anodic bonding, eutectic bonding, and adhesive bonding [16]. *Silicon fusion bonding* is a high-temperature process that provides high bond strength of the order of 10–20 MPa between two silicon wafers. The basic steps are pretreatment, room temperature mating, and thermal annealing. The pretreatment involves polishing and hydrating the wafer surfaces. After pretreatment, the wafers are aligned and mated at room temperature. Subsequently, the completed assembly is annealed at elevated temperature (700–1100°C) to increase the bond strength. *Anodic bonding* is a method of electrostatically bonding two materials, typically silicon and glass. In this process, the two substrates are bonded by heating the assembly on a

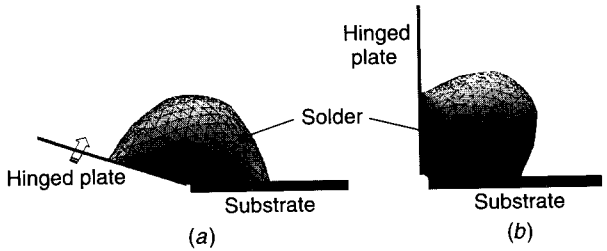


Figure 14.5 Illustration of solder self-assembly: (*a*) wet solder at hinged joint; (*b*) final shape after upward motion due to surface tension forces and solder resolidification.

hot plate, typically to 400°C, and applying electrostatic potential, typically of 1000 V, across the interface. *Eutectic bonding* is based on the phenomenon that the melting temperature of an alloy of two metals can be lower than the melting point of either metal. The lowest melting point of an alloy (eutectic point) depends on the mixing ratio of the metals. The method involves depositing a eutectic film on one of the wafers followed by thermocompression bonding in vacuum. Typically, gold is used to bond silicon to silicon at 363°C (eutectic temperature for a composition of 97.1% Au and 2.9% Si by weight). *Adhesive bonding* uses epoxy resins to bond a large variety of materials such as metals, glasses, and plastics at low process temperatures (between room temperature and 150°C). Other available bonding techniques include soldering and welding.

14.3 RF MEMS ACTUATORS

The movement of the mechanical structure of RF MEMS is achieved by an actuation mechanism such as electrostatic, electrothermal [1, 11], or piezoelectric [1, 12]. The mechanical structure used for the actuation mechanism is called an *actuator*, which converts the electrical (or other) energy into a mechanical movement. Commonly used actuator structures are described in Section 14.3.1. The electrostatic actuators are most widely used in RF MEMS and are discussed in Section 14.3.2.

14.3.1 Actuator Structures

The commonly used actuator structures in RF MEMS are a cantilever beam, a fixed–fixed beam, and a diaphragm. Beams have their lengths greater than other dimensions and have small deflections (perpendicular to their axis) relative to their lengths. The mechanical movement of an RF MEMS structure is achieved by deflection of the beam subjected to a transverse actuation force. When the force is removed, the elasticity property of the beam material causes

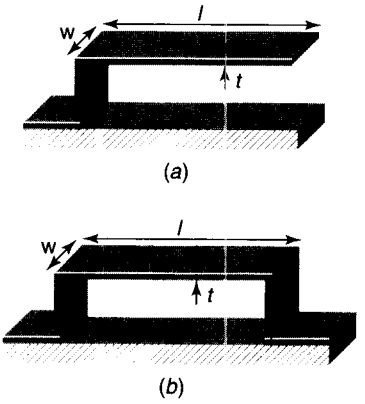


Figure 14.6 Actuation structures used in RF MEMS: (*a*) cantilever; (*b*) fixed–fixed beam. (Sketches not to a scale)

it to regain its original configuration (after having been deformed). The beam deflection is linearly related to the force as long as the elastic limit of the material is not exceeded. This elastic behavior of the beam can be modeled as a *linear spring* of spring constant k (in newtons per meter), which is the ratio of force to deflection.

A *cantilever beam* (shown in Fig. 14.6*a*) is fixed at one end and free at the other end. At the fixed or clamped end, the beam can neither translate nor rotate, whereas at the free end it may do both. A *fixed–fixed beam* or *clamped beam* (shown in Fig. 14.6*b*) is fixed at both ends. A *membrane* or *diaphragm* is similar to a beam except it has a width comparable to its length and typically has fixed boundaries at the edges. The spring constants (under uniformly loaded condition) for the cantilever and fixed–fixed beam [17] are given below:

$$k = \frac{2E'wt^3}{3l^3} \quad \text{(cantilever beam)} \tag{14.1}$$

$$k = \frac{32E'wt^3}{l^3} + \frac{8\sigma_0(1-\nu)wt}{l} \quad \text{(fixed–fixed beam)} \tag{14.2}$$

where $E' = E$ for narrow beams and $E' = E/(1-\nu^2)$ for wide beams (or membranes), E is the Young's modulus of the beam material, ν is Poisson's ratio, σ_0 is the biaxial residual stress, and the dimensions l, w, t are the length, width, and thickness of the beam, respectively.

The spring constant of the actuator structure can be reduced by the design of a compliant flexure between the actuator and the support (or anchor). Various flexure designs have been proposed for fixed–fixed membranes [18, 19] and are shown in Fig. 14.7. Similar flexure designs can also be used for cantilever structures.

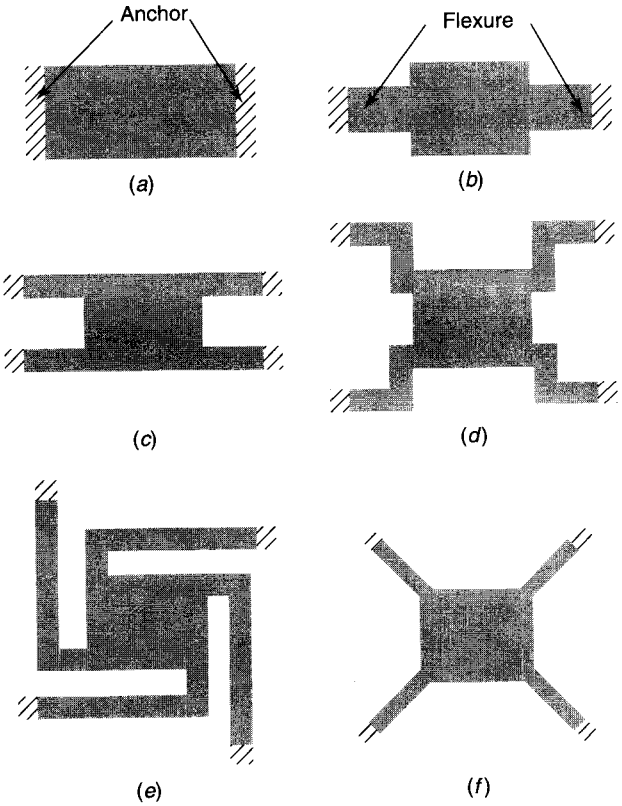


Figure 14.7 Flexure designs: (*a*) fixed–fixed beam; (*b–f*) various compliant flexure designs.

14.3.2 Electrostatic Actuators

Electrostatic actuators are most commonly used in RF MEMS. They have very low power consumption. Two types of electrostatic actuators, parallel-plate and interdigital, are discussed here.

Parallel-Plate Actuator. A parallel-plate actuator with a movable top plate anchored to a support (cantilever or fixed–fixed type) can be modeled by a one-dimensional spring–capacitor model, as shown in Fig. 14.8*a*. This model possesses a single degree of freedom for movement, which is the gap between the movable top plate and the fixed bottom plate. Let us assume that the top plate moves toward the bottom plate from an initial gap height z_0 to a new gap height z $(=z_0 - \Delta z)$. Then, the restoring mechanical force exerted on the plate for a displacement Δz is given by Hooke's law,

$$F_M = k\,\Delta z = k(z_0 - z) \tag{14.3}$$

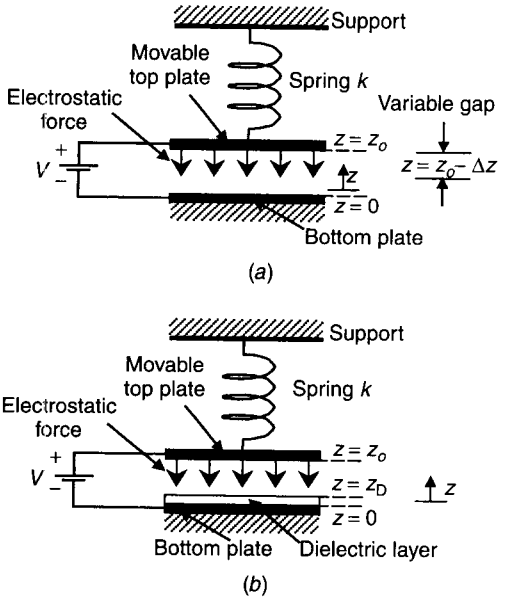


Figure 14.8 One-dimensional spring-capacitor model: (*a*) a parallel-plate actuator; (*b*) parallel-plate actuator with dielectric layer.

where k is the spring constant of the anchored plate and Δz is the displacement of the plate.

The electrostatic energy stored in the capacitor for a DC voltage V applied between the plates is given by

$$U_E(z) = \tfrac{1}{2} C(z) V^2 \tag{14.4}$$

with

$$C(z) = \frac{\epsilon_0 A}{z} \tag{14.5}$$

where $C(z)$ is the parallel-plate capacitance (neglecting the fringing capacitance) changing with the gap height z between the plates.

The energy supplied by the voltage source ($\delta U_b = V\delta Q$) is the sum of the mechanical work done (energy stored in the spring, $\delta W_M = F_E \delta z$) and the energy stored in the electric field of the capacitor (δU_E).

$$\delta U_b = \delta W_M + \delta U_E \tag{14.6a}$$

Using the relationship $Q = CV$ and (14.4) for the capacitor, the electrostatic force F_E (at a constant voltage) can be obtained as

$$F_E = \left.\frac{\delta U_b}{\delta z}\right|_{\delta z \to 0} - \left.\frac{\delta U_E}{\delta z}\right|_{\delta z \to 0} = \frac{dU_b}{dz} - \frac{dU_E}{dz} = \frac{1}{2}\frac{dC(z)}{dz}V^2 \qquad (14.6b)$$

Substituting (14.5) in (14.6) yields

$$F_E = \frac{1}{2}\frac{\partial C(z)}{\partial z}V^2 = -\frac{1}{2}\frac{\epsilon_0 A V^2}{z^2} \qquad (14.7)$$

Equation (14.7) gives the pull-down electrostatic force F_E developed for an applied voltage V given gap height z and plate area A. The negative sign in (14.7) implies that the electrostatic force is in the downward direction ($-z$ direction in Fig. 14.8*a*). The above derivation is based on considering a constant voltage across the capacitor. Derivation for F_E when the charge on the capacitor is constant is given in Seeley [20].

The pull-up mechanical force (14.3) has a linear dependence on the gap height ($F_M \propto z_0 - z$), whereas the pull-down electrostatic force (14.7) has a nonlinear dependence on the gap height ($F_E \propto 1/z^2$). The spring–capacitor system is stable as long as the mechanical force balances the electrostatic force by downward movement of the top plate. Balancing the mechanical restoring force and the electrostatic force in the stable state and solving for V give

$$V = z\sqrt{\frac{2k(z_0 - z)}{\epsilon_0 A}} \qquad (14.8)$$

This equation gives the voltage V required to pull the top plate toward the bottom plate by a distance of Δz ($=z_0 - z$) from an initial gap height z_0.

Equation (14.8) can be expressed as a function of normalized gap height z/z_0:

$$V = V_0\left(\frac{z}{z_0}\right)\sqrt{1 - \frac{z}{z_0}} \qquad (14.9)$$

where

$$V_0 = z_0\sqrt{\frac{2kz_0}{\epsilon_0 A}} \qquad (14.10)$$

Equation (14.9) yields a parabolalike variation of normalized gap height z/z_0 with V, as shown in Fig. 14.9*a*. In the continuous part of this curve the spring-

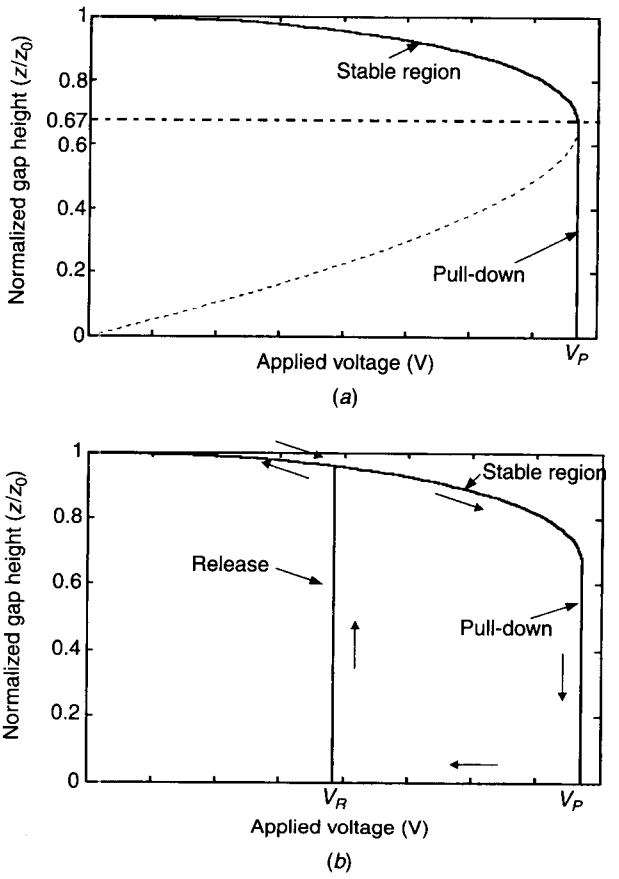


Figure 14.9 Parallel-plate actuator characteristics: (*a*) normalized gap height versus applied voltage; (*b*) hysteresis characteristic.

capacitor system is stable. The dotted part of the curve in the plot represents the unstable state. The transition into this unstable state can be studied by writing the rate of change of V with respect to gap z as

$$\frac{dV}{dz} = \left(\sqrt{\frac{2k}{\epsilon_0 A}}\right)\sqrt{z_0 - z} - \left(\sqrt{\frac{2k}{\epsilon_0 A}}\right)\frac{1}{2}\frac{z}{\sqrt{z_0 - z}} \tag{14.11}$$

It can be noted that $dV/dz = 0$ when

$$z = \tfrac{2}{3}z_0 \tag{14.12}$$

or when

$$V_P = \sqrt{\frac{8kz_0^3}{27A\epsilon_0}} \tag{14.13}$$

The height $\frac{2}{3}z_0$ is also called critical gap height z_c. Equation (14.13) is obtained by substituting (14.12) in (14.8). When the applied voltage $V > V_P$ (dV/dz is negative), F_E exceeds F_M. As a result, the top plate moves down; that is, the gap z decreases and F_E increases further. This positive feedback (due to the inverse dependence of F_E on the gap height) leads to a rapid decrease in the gap height z to zero, as shown in Fig. 14.9*a*. This phenomenon is known as pull-down and the voltage V_P given in (14.13) is termed the *pull-down voltage*. When the applied DC voltage is removed, the restoring mechanical force pulls the plate back into its original position ($z = z_0$). A hysteresis phenomenon in the gap height versus voltage is observed (Fig. 14.9*b*). This phenomenon is discussed later.

It is interesting to note that the pull-down voltage is independent of the width of the parallel-plate actuator for a simple cantilever beam and fixed–fixed beam (zero residual stress) shown in Fig. 14.6. This is because the spring constant $k \propto w$ and the plate area $A \propto w$, which makes the actuation voltage, $V_P \propto (k/A)^{1/2}$, independent of the width of the actuator.

Example Pull-Down Voltage Calculation. Let us consider an example of the pull-down voltage calculation for a parallel-plate actuator. An electrostatic parallel-plate actuator consists of a movable top plate with dimensions 250 μm × 100 μm × 2 μm and is separated from the bottom plate by a gap height of 2 μm. The Young's modulus of the plate material (gold) is $E = 80$ GPa and Poisson's ratio $v = 0.42$. Assume that the top electrode has zero residual stress and is uniformly loaded by the pull-down electrostatic force. Calculate the pull-down voltage when the parallel-plate actuator is (i) a cantilever beam and (ii) a fixed–fixed beam.

In this case, $l = 250$ μm, $w = 100$ μm, $t = 2$ μm, $A = 25000$ (μm)^2, and $z_0 = 2$ μm. The critical gap height (from 14.12) is $z_c = \frac{2}{3}z_0 = \frac{2}{3}(2\ \mu\text{m}) = 1.33$ μm. The width of the beam is comparable to the length of the beam. So, the beam can be considered as a wide beam.

(i) *Cantilever Beam.* The spring constant for a cantilever beam can be obtained using (14.1) as

$$k = \frac{2E'wt^3}{3l^3} = \frac{2 \times 80 \times 10^9/(1 - 0.42^2) \times 100 \times 10^{-6} \times (2 \times 10^{-6})^3}{3 \times (250 \times 10^{-6})^3}$$

$$\cong 3.3\ \text{N/m} \tag{14.14}$$

The pull-down voltage (14.13) is

$$V_P = \sqrt{\frac{8kz_0^3}{27A\epsilon_0}} = \sqrt{\frac{8 \times 3.3 \times (2 \times 10^{-6})^3}{27 \times 250 \times 10^{-6} \times 100 \times 10^{-6} \times 8.854 \times 10^{-12}}}$$

$$\cong 6.0 \text{ V} \tag{14.15}$$

(ii) *Fixed–Fixed Beam.* When the top plate is a fixed–fixed beam with zero residual stress ($\sigma_0 = 0$), the spring constant can be calculated using (14.2) as

$$k = \frac{32E'wt^3}{l^3} = \frac{32 \times 80 \times 10^9/(1 - 0.42^2) \times 100 \times 10^{-6} \times (2 \times 10^{-6})^3}{(250 \times 10^{-6})^3}$$

$$\cong 159 \text{ N/m} \tag{14.16}$$

The pull-down voltage (14.13) is calculated to be

$$V_P = \sqrt{\frac{8kz_0^3}{27A\epsilon_0}} = \sqrt{\frac{8 \times 159 \times (2 \times 10^{-6})^3}{27 \times 250 \times 10^{-6} \times 100 \times 10^{-6} \times 8.854 \times 10^{-12}}}$$

$$\cong 41.3 \text{ V} \tag{14.17}$$

As expected, this value is much higher than the cantilever beam case.

Invariably, a dielectric layer is placed on the bottom plate (as shown in Fig. 14.8*b*) of the actuator to prevent an electrical short in the down position. This provides a capacitive contact when the plate is in the down position. In this case, the capacitance is given by

$$C(z) = \frac{\epsilon_0 A}{z_A + z_D/\epsilon_D} \tag{14.18}$$

where z_D is the thickness of the dielectric layer, ϵ_D is the relative dielectric constant of the dielectric material, and z_A ($=z_0 - z_D$) is the air gap height.

The critical gap height z_c and the pull-down voltage V_P can be obtained by a procedure similar to that described above (or) simply by replacing z_0 by $z_A + z_D/\epsilon_D$ in (14.12) and (14.13). The values of z_c and V_P are now given by

$$z_c = \frac{2}{3}\left(z_A + \frac{z_D}{\epsilon_D}\right) \tag{14.19}$$

$$V_P = \sqrt{\frac{8k}{27A\epsilon_0}\left(z_A + \frac{z_D}{\epsilon_D}\right)^3} \tag{14.20}$$

If $z_A \gg z_D/\epsilon_D$,

$$V_P \cong \sqrt{\frac{8kz_A^3}{27A\epsilon_0}} \tag{14.21}$$

When a dielectric layer (Si_3N_4, $\epsilon_D = 7.6$) of thickness 0.2 μm is placed on the bottom electrode of the parallel-plate actuator example discussed above, the pull-down voltage [using (14.20)] for the fixed–fixed beam case becomes $V_P = 36.0$ V (lower than 41.3 V when there is no dielectric layer).

Gap Height–Voltage Hysteresis Characteristics. When an increasing DC voltage is applied, the top plate moves down steadily as long as the gap height is between z_0 and $2/3z_0$ (in the stable region). At the pull-down voltage, the top plate snaps down as explained earlier. If the voltage is then reduced, the top plate releases back up, but typically at a lower voltage than the pull-down voltage. In the plate-down position, the electrostatic force is higher due to the reduced gap height. Therefore, the voltage required for holding the plate in the down position is lower than the pull-down voltage. The voltage at which the top plate is completely released from the bottom plate is known as the *release voltage* V_R. Thus, the parallel-plate actuator exhibits a hysteresis characteristic, as shown in Fig. 14.9*b*.

Switching Speed. The switching speed depends on the actuator structure, material, and design. Typical values of switching speed of parallel-plate actuators are in the microseconds range, which is slower than that of semiconductor switches discussed in Chapter 12. During the actuator pull-down process, the nonlinearly increasing electrostatic force controls the plate movement, whereas in the release process the linear restoring mechanical force controls the plate movement. This makes the release time T_R (10 μs typical) much longer than the pull-down time T_P (1 μs typical), and this longer release time is usually the switching speed limitation in parallel-plate actuators [21, 22]. Assuming that the parallel-plate actuator is connected to a voltage source (with very low series resistance) that charges the actuator in negligible time, the electrical switching time (Section 12.2.6) is controlled by T_P or T_R depending upon the transition that is being considered.

A one-dimensional dynamic model of the parallel-plate actuator with spring, mass, damper, and capacitor for calculation of pull-down time T_P is shown in Fig. 14.10. The dynamic response of this model is given by

$$m\frac{d^2z}{dt^2} + b\frac{dz}{dt} + k(z_0 - z) = -F_E \tag{14.22}$$

where m is the inertial mass of the actuator, b is its damping coefficient, k is the spring constant, and z is the displacement. The time required to pull down the actuator (in the absence of damping, i.e., $b = 0$) from $z = z_0$ to $z = 0$ is

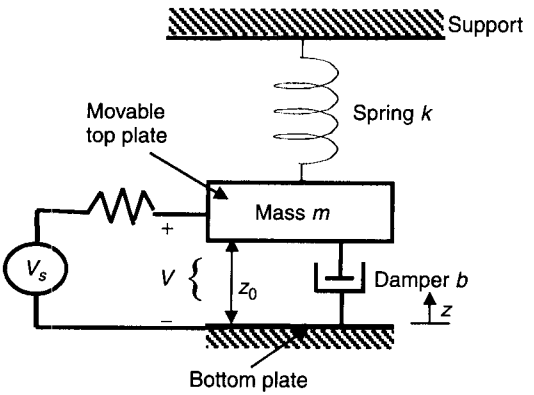


Figure 14.10 One-dimensional dynamic model of parallel-plate actuator.

obtained by using the initial condition $dz/dt = 0$ at $z = z_0$ in (14.22) and is given by [23]

$$T_P = 3.67\frac{V_P}{\omega_0 V_A} \tag{14.23}$$

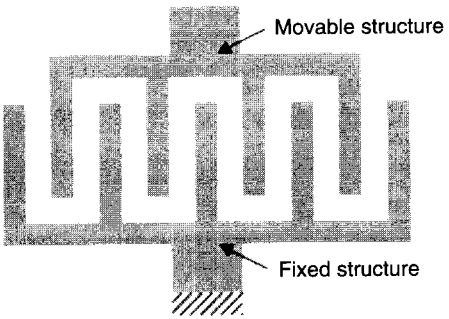
where $\omega_0 = \sqrt{k/m}$ is the mechanical resonance frequency of the actuator, V_A is the applied voltage, and V_P is the pull-down voltage of the actuator. This expression (14.23) is derived by assuming the electrostatic force F_E to remain constant at its initial value ($z = z_0$) and is accurate when V_A is larger than $2V_P$.

Chan et al. [24] consider a nonlinear dynamic model for computing the switching time that includes the effects of electrostatic force, bending force, stretching force, residual stress, inertia, squeeze film damping, Van der Waals forces, and contact forces.

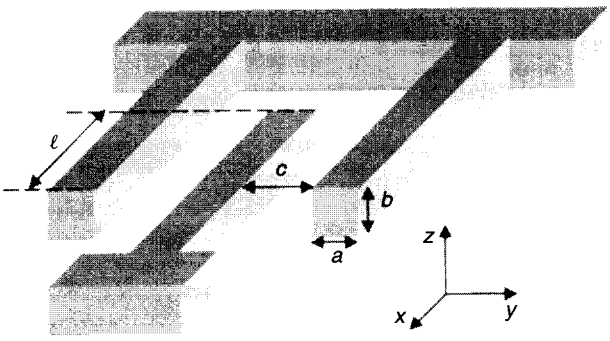
Interdigital Actuator. An interdigital actuator (also known as a comb-drive actuator) consists of a fixed and a movable structure with interdigitated fingers, as shown in Fig. 14.11*a*. A section of an interdigital actuator is shown in Fig. 14.11*b*. The movable structure is usually constrained by mechanical supports in two directions and can move only in the lateral direction (x direction in Fig. 14.11*b*). As the area normal to the direction of movement is smaller than that of a parallel-plate actuator, the damping forces are smaller. An interdigital actuator with $n + 1$ interdigitated fingers will contain n parallel-plate gaps (e.g., nine fingers and eight gaps shown in Fig. 14.11*a*). Neglecting the fringing effect, the total capacitance of an interdigital structure with $n + 1$ fingers is given by

$$C = \frac{n\epsilon_0 lb}{c} \tag{14.24}$$

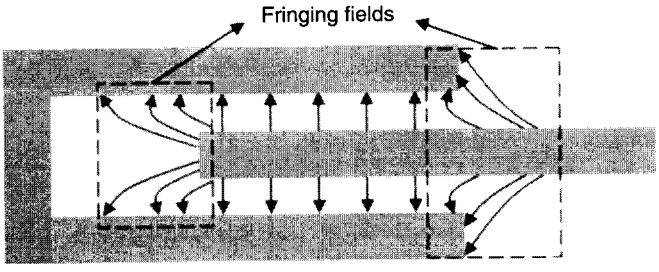
where c is the gap between fingers, b is the thickness of finger, and l is the overlap length of the fixed and movable interdigitated structure. It can be noted



(*a*)



(*b*)



(*c*)

Figure 14.11 Interdigital actuator: (*a*) schematic top view; (*b*) section with three fingers and two gaps; (*c*) fringing fields at finger end.

that, if the fringing field capacitance is neglected, the capacitance between the adjacent fingers has a linear relationship with the lateral displacement in the x direction. This is a useful feature for variable capacitors.

At the ends of the moving and fixed fingers, there exist fringing electrostatic fields with field lines terminated on the sides of adjacent fingers, as shown in

Fig. 14.11*c*. There is an x component of these fringing fields that provides an electrostatic force in the x direction and controls the movement of the finger. By balancing this electrostatic force of the fringing field with a restoring spring force $F_M = -kd$ (where k is the spring constant), dynamics of this kind of interdigital actuators can be investigated. As the length of a finger is much longer than the cross-sectional dimensions, the fringe field and hence the attractive electrostatic force remains constant over the duration of the movement of the finger. This is in contrast to a parallel-plate actuator where the electrostatic force between the two parallel plates increases (inversely to the square of gap height) as the actuator moves down. Consequently, there is no pull-down nonlinear phenomenon observed in interdigital actuators. Thus they are well suited for variable capacitors in RF circuits.

14.4 RF MEMS DEVICES

In this section RF MEMS switches and RF MEMS varactors are discussed.

14.4.1 Switches

Microelectromechanical switches were among the first devices developed in the MEMS field [25]. An RF MEMS switch consists of RF electrodes that can be actuated to short circuit or open circuit an RF transmission line. The advantages of RF MEMS switches include low insertion loss, high isolation, and linearity (or absence of intermodulation products). A qualitative comparison of MEMS switches with other switches is presented in Table 14.2. RF MEMS

Table 14.2 Comparison of MEMS Switches with Other Switches

Performance and Characteristics	Ideal Switches	Electromagnetic Relays	Semiconductor Switches		RF MEMS Switches
			RF Frequency	Microwave Frequency	
Insertion loss	Zero	Low	Moderate	High	Low
Isolation	Infinite	High	High	Moderate	High
Switching speed	Infinite	Low	High	High	Low
Power consumption	Zero	High	Moderate	Moderate	Low
Linearity	Perfect	High	Low	Low	High
Operating temperature range	Infinite	High	Low	Low	High
Device density	Infinite	Low	High	High	High
Reliability	Perfect	High	High	Moderate	Moderate
Cost	Zero	High	Low	Moderate	Low
Ease of integration	High	Low	Moderate	Moderate	Moderate

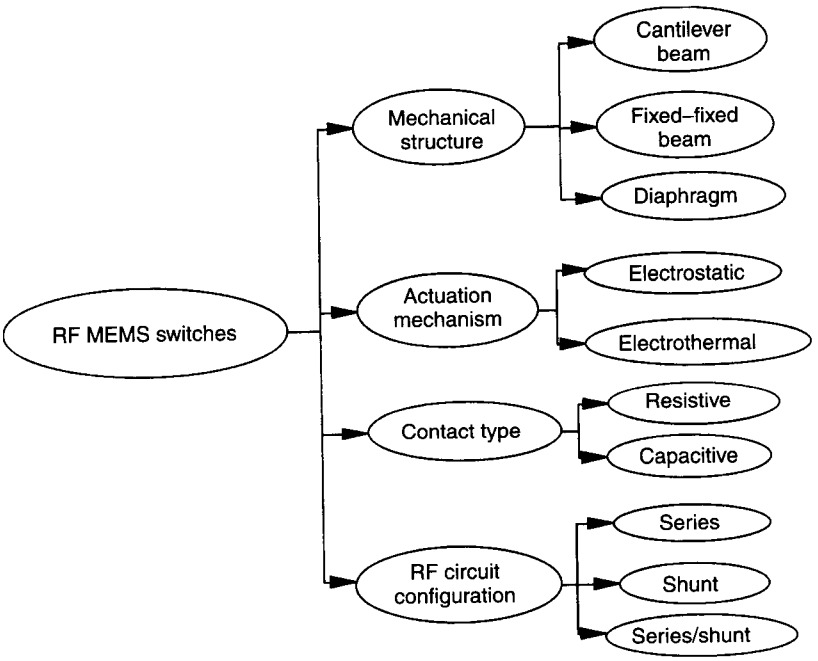


Figure 14.12 Classification of RF MEMS switches.

switches have been developed with various combinations of actuation mechanisms, mechanical structures, contact types, and RF circuit configurations. Various possible classifications of RF MEMS switches are shown in Fig. 14.12. Mechanical structures and actuation mechanisms have been discussed earlier. Electrostatic actuation is the commonly used actuation mechanism for RF MEMS switches.

In this section, we discuss the contact types, circuit configuration, and some equivalent circuit models for RF MEMS switching devices integrated in planar transmission lines.

RF MEMS Structure. A typical RF MEMS device consists of actuation electrodes (where the actuation or control voltage is applied) and RF electrodes (which are integrated with the RF circuit). Actuation electrodes form a part of the actuator structure. The RF electrodes usually become a part of (or discontinuity in) an RF transmission line. The role of actuation is to cause the RF electrodes to short, open, load, or otherwise affect an RF transmission line.

RF Contact Types. Two types of electrode contacts commonly used in RF MEMS devices are resistive contact and capacitive contact. For actuation electrodes, a capacitive contact is usually preferred to avoid the electrical short caused by a resistive contact. On the other hand, a resistive contact is desirable

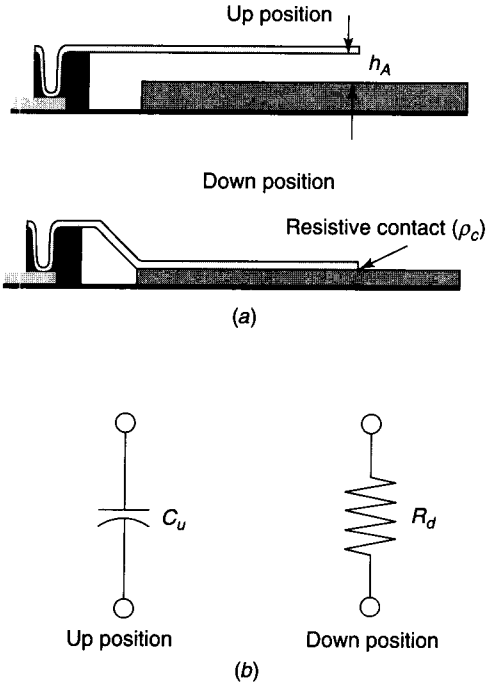


Figure 14.13 Resistive contact MEMS switch: (*a*) cross-sectional view of cantilever switch; (*b*) simplified equivalent-circuit representations.

for RF electrodes to provide an RF short in the actuated state even at low frequencies. These two contact types are discussed here in more details.

Resistive Contact. In this case, there is an ohmic contact between the two electrodes in the down position. The up-position capacitance C_u and the down-position resistance R_d for a *resistive contact* (shown in Fig. 14.13) are given by

$$C_u = \frac{\epsilon_0 A}{h_A} \tag{14.25}$$

$$R_d = \frac{\rho_c}{A} \tag{14.26}$$

where A is the area of the electrodes, h_A is the gap height between the electrodes, ρ_c is the surface resistivity of the electrode material, and ϵ_0 is the permittivity of air. Surface resistivity is defined as resistance for a unit area of the contact. If the contact is approximated as a uniform resistive sheet of thickness t, the surface resistivity $\rho_c = \rho t$, where ρ is the resistivity of the sheet. A figure of merit is defined as a product of the up-position capacitance and down-position resistance as

$$\zeta_R = R_d C_u = \frac{\epsilon_0 \rho_c}{h_A} \tag{14.27}$$

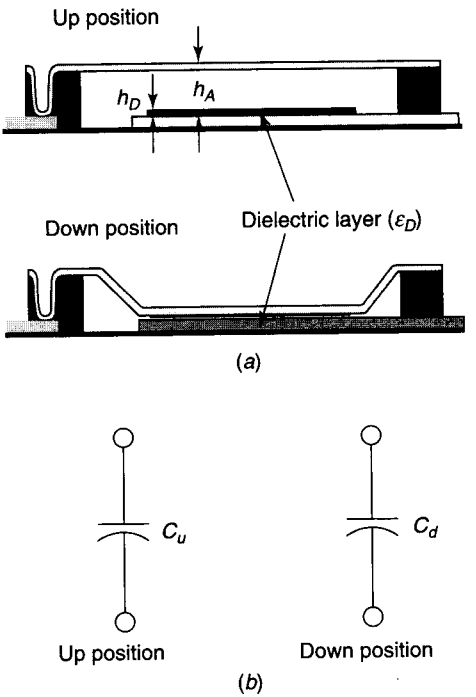


Figure 14.14 Capacitive contact MEMS switch: (*a*) cross-sectional view of bridge; (*b*) simplified equivalent-circuit representations.

This is also expressed as the cutoff frequency [21], defined as

$$f_0 = \frac{1}{2\pi R_d C_u} = \frac{1}{2\pi\zeta_R} \tag{14.28}$$

which needs to be high for good switches.

Capacitive Contact. In this case, a thin dielectric layer is present between the two electrodes in the down position. The up-position capacitance C_u and down-position capacitance C_d for a *capacitive contact* (shown in Fig. 14.14) are given by

$$C_u = \frac{\epsilon_0 A}{h_A + h_D/\epsilon_D} = \frac{\epsilon_0 \epsilon_D A}{\epsilon_D h_A + h_D} \tag{14.29}$$

$$C_d = \frac{\epsilon_0 \epsilon_D A}{h_D} \tag{14.30}$$

where A is the area of the electrodes, h_A is the air gap height between the electrodes, h_D is the thickness of the dielectric layer on the bottom electrode, ϵ_D is

the relative dielectric constant of the dielectric material, and ϵ_0 is the permittivity of air. Since h_A and h_D are much less than the lateral dimensions, the fringing capacitance has been ignored. A figure of merit [21, 22] is defined as the ratio of down-position capacitance to up-position capacitance and is given by

$$\zeta_C = \frac{C_d}{C_u} = \epsilon_D \left(\frac{h_A}{h_D}\right) + 1 \tag{14.31}$$

If $\epsilon_D(h_A/h_D) \gg 1$, Eq. (14.31) can be written as

$$\zeta_C \cong \epsilon_D \left(\frac{h_A}{h_D}\right) \tag{14.32}$$

It can be noted that the capacitance ratio depends only on the up-position gap height h_A, down-position dielectric thickness h_D, and relative dielectric constant of the dielectric material ϵ_D. This capacitance ratio is typically of the order of 100 for switching applications. For example, for a capacitor with a silicon nitride ($\epsilon_D = 7.6$) dielectric layer of thickness $h_D = 0.2$ μm on the bottom electrode and up-position gap height of $h_A = 3$ μm, the capacitance ratio $\zeta_C = 115$. If the area of the plates is 100 μm × 100 μm, the up-position capacitance and down-position capacitance values are 29 fF and 3.4 pF, respectively.

RF Circuit Configuration. RF MEMS switches can also be categorized based on circuit configuration—series and shunt (parallel) configurations. An RF MEMS switching device is usually integrated in a planar transmission line and is used to load, short, or open the line. Shunt and series configurations of RF MEMS on planar transmission lines (such as microstrip line, coplanar waveguide (CPW), coplanar strip (CPS), slot line), and the corresponding equivalent-circuit models are presented in Figs. 14.15 and 14.16, respectively. Selection of series or shunt configuration depends upon the switch application. For example, series configurations are more convenient for integration in microstrip circuits whereas shunt configurations are easy to use in slot-line circuits. For coplanar line (CPW and CPS) circuits both series and shunt configurations can be designed. These circuit configurations and the modeling discussed later are applicable to RF MEMS variable capacitors (discussed in Section 14.4.2).

Shunt Configuration. In the shunt configuration (Fig. 14.15), the MEMS electrode up position allows the signal to propagate, whereas the electrode in the down position causes the signal to be reflected back. Thus, for shunt switch applications the electrode up and down positions correspond to ON and OFF states, respectively.

Four different examples of switches in shunt configuration are shown in Fig. 14.15*a*, with the corresponding lumped equivalent circuits in Fig. 14.15*b*. Physical parts of the device structure that contribute to inductance, capacitance, and resistance are shown in Fig. 14.15*a*. In all these cases, the movable

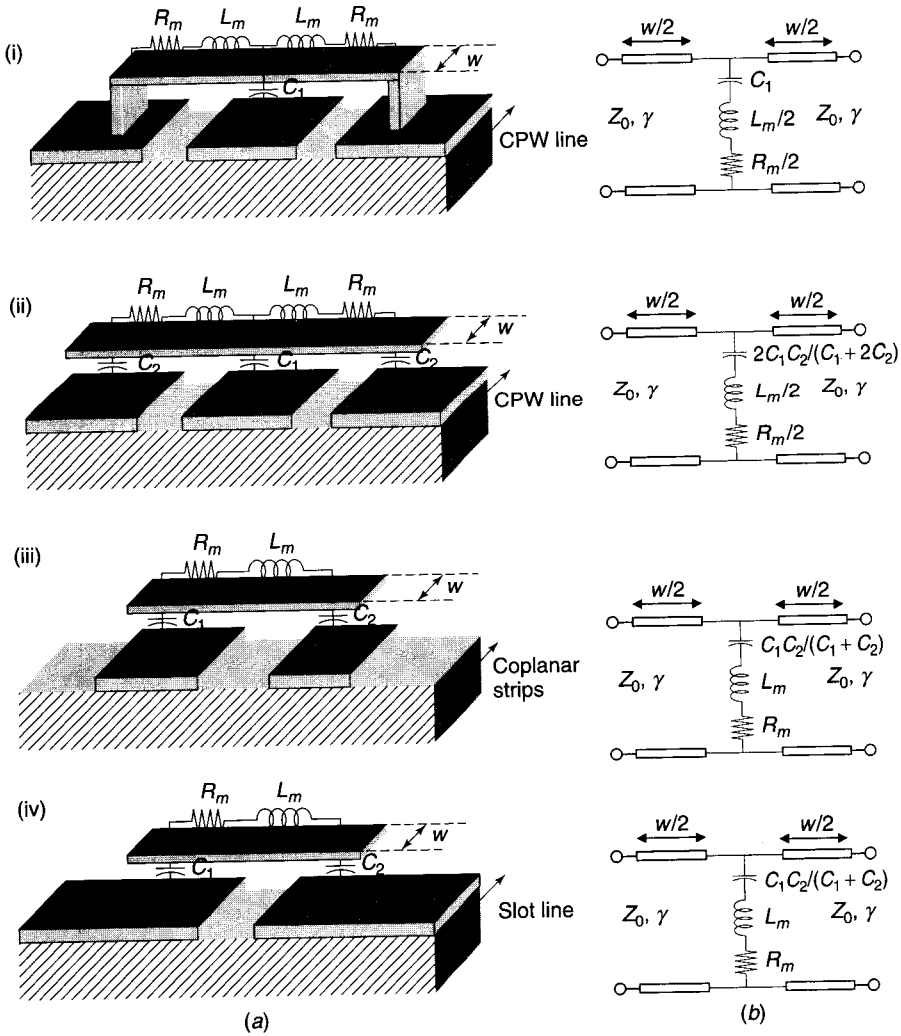


Figure 14.15 Shunt-mounted RF MEMS on planar transmission lines: (*a*) physical configurations; (*b*) corresponding equivalent-circuit models (w = width of top electrode).

MEMS structure contributes to a series combination of a resistance and an inductance. The air gaps between the MEMS structure and the transmission-line conductors contribute to shunt capacitances that appears in series with L and R of the top plate. Values of L, C, and R depend on the geometry selected and can be optimized for desired switch performance. When the moving structure is in the down position, the equivalent circuit becomes different for resistive and capacitive contact devices. For capacitive contact devices, the equivalent circuits are identical to those in Fig. 14.15*b* with capacitance(s) replaced by down-position capacitance(s). The inductance values may also change because

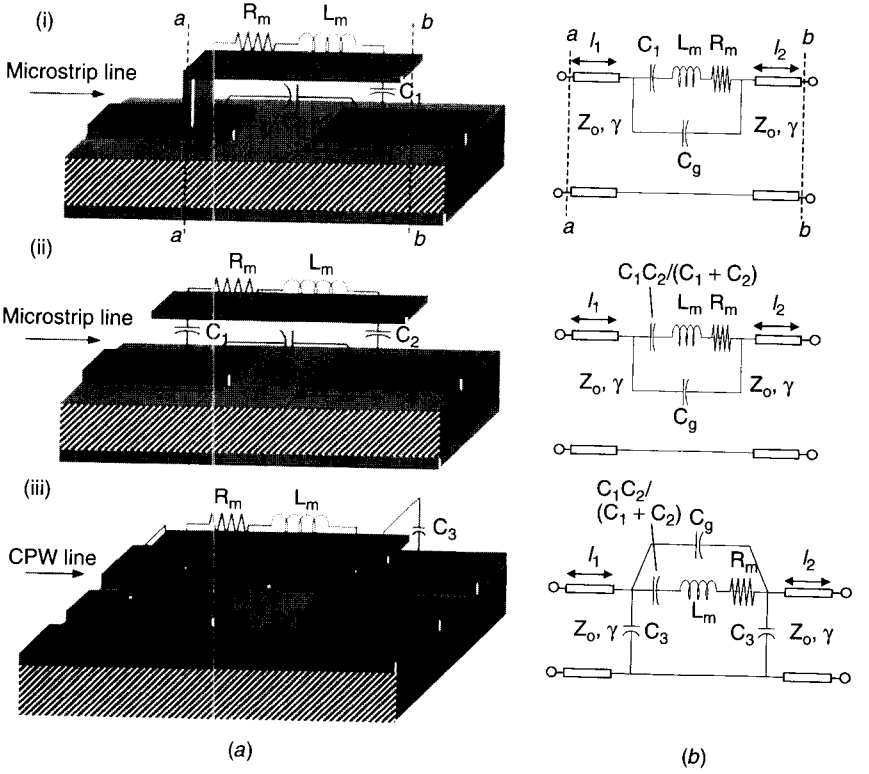


Figure 14.16 Series-mounted RF MEMS on planar transmission lines: (*a*) physical configurations; (*b*) equivalent-circuit models (*l* is overlapping length of top electrode and signal conductor of transmission line).

of the proximity of the MEMS structure to the transmission-line conductors underneath. For *resistive contact* devices, the equivalent circuit in the down position does not contain capacitances and consists of down-position inductance and resistance only.

All the shunt configurations shown in Fig. 14.15 can be modeled as a shunt impedance appearing across the transmission line that can be written as

$$Z = R + jX = R + j\left(\omega L - \frac{1}{\omega C}\right) \tag{14.33}$$

where values of R, L, and C for various configurations are shown in Fig. 14.15*b*. The loss $(1/|S_{21}|^2)$ for a shunt impedance Z across a transmission line of characteristic impedance Z_0 can be written as Eq. (12.5) in Chapter 12:

$$\text{Loss} = \left|\frac{2Z + Z_0}{2Z}\right|^2 \tag{14.34}$$

Substituting for Z from (14.33), we get

$$\text{Loss} = 1 + \frac{RZ_0 + Z_0^2/4}{R^2 + (\omega L - 1/\omega C)^2} \tag{14.35}$$

where $L = L_u$, $C = C_u$, and $R = R_u$ in the up position and $L = L_d$, $C = C_d$, and $R = R_d$ in the down position.

The *insertion loss* (IL) of the switch in the up position (ON state) can be approximated by neglecting inductance L_u and resistance R_u [i.e., $\omega L_u \ll 1/\omega C_u$ (or $f \ll f_0 = 1/2\pi\sqrt{L_u C_u}$), $R_u \ll 1/\omega C_u$, and $R_u \ll \frac{1}{4}Z_0$]:

$$\text{IL} = 1 + \tfrac{1}{4}(\omega^2 C_u^2 Z_0^2) \tag{14.36}$$

For resistive contact devices, C_d is infinite in the down position, and the *isolation* (IS) for a resistive contact switch can be obtained by substituting $C_d = \infty$ in (14.35):

$$\text{IS} = 1 + \frac{R_d Z_0 + Z_0^2/4}{R_d^2 + \omega^2 L_d^2} \tag{14.37}$$

At low frequencies, R_d determines the isolation and at high frequencies ωL_d dominates isolation. Also, in resistive contact RF MEMS devices typically $R_d \ll \frac{1}{4}Z_0$. Therefore, the expression (14.37) can be approximated as

$$\text{IS} \cong \begin{cases} \dfrac{Z_0^2}{4R_d^2} & R_d \gg \omega L_d \text{ (at low frequencies)} \\ 1 + \dfrac{Z_0^2}{4\omega^2 L_d^2} & \omega L_d \gg R_d \text{ (at high frequencies)} \end{cases} \tag{14.38}$$

In capacitive contact switches L_d and C_d can exhibit a series resonance in the down position (OFF state) at $f = f_0 = 1/2\pi\sqrt{L_d C_d}$. Subsequently, the switch can be designed to improve the isolation by locating this resonance at the frequency where maximum isolation is needed. The isolation can be approximated as

$$\text{IS} \cong \begin{cases} \frac{1}{4}(\omega^2 C_d^2 Z_0^2) & f \ll f_0 \\ \dfrac{Z_0^2}{4R_d^2} & f = f_0 \\ \dfrac{Z_0^2}{4\omega^2 L_d^2} & f \gg f_0 \end{cases} \tag{14.39}$$

An ideal switch possesses zero insertion loss in the ON state and infinite isolation in the OFF state. It can be noted that in capacitive contact shunt switches both the insertion loss and isolation increase with increase in fre-

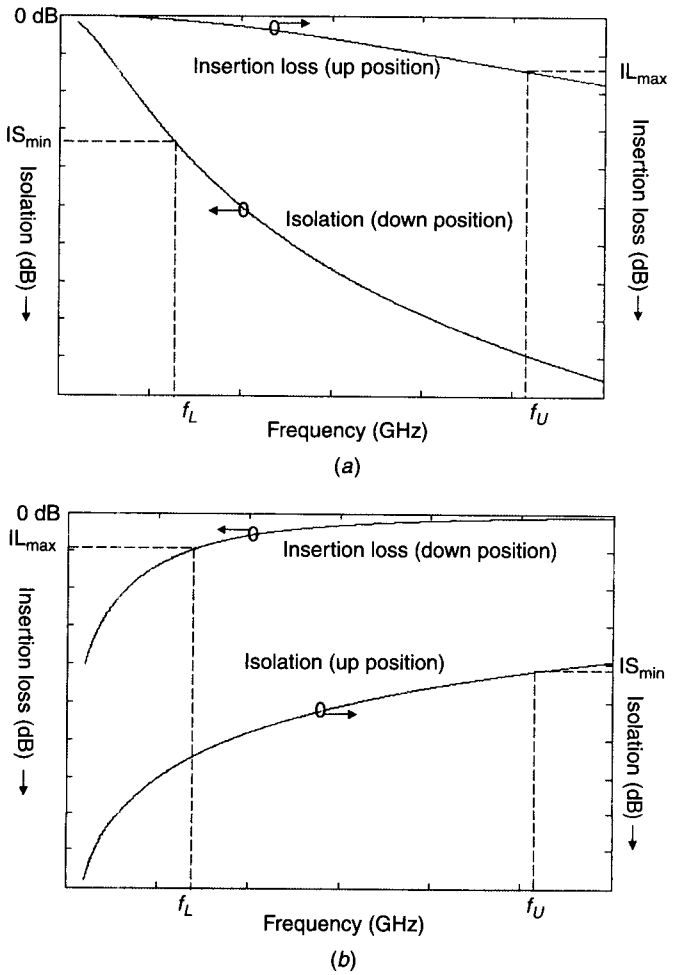


Figure 14.17 Insertion loss and isolation characteristics of RF MEMS capacitive switch: (*a*) shunt switch; (*b*) series switch.

quency, as shown in Fig. 14.17*a*. As a result, the lower operating frequency f_L is limited by the minimum isolation required and the upper operating frequency f_U is limited by the maximum insertion loss acceptable. In resistive contact shunt switches, the insertion loss increases with frequency similar to capacitive shunt switches and the isolation decreases with frequency (in the absence of parasitic capacitance), as shown by Eq. (14.38). Hence, the switch has only an upper operating frequency f_U, limited by the specified maximum insertion loss and/or minimum isolation.

Series Configuration. Examples of series switches in various planar transmission structures are shown in Fig. 14.16*a*. In this case, the electrode up position does not allow the signal to propagate, and the electrode down position allows

the signal to be delivered to the load. In series switch applications, the electrode up and down positions correspond to OFF and ON states, respectively. Equivalent-circuit models for series switches are shown in Fig. 14.16*b*. The important part of these models is a series impedance in a transmission line.

The impedance of the MEMS switches in series configuration in a transmission line can be approximated as

$$Z = R + jX = R + j\left(\omega L - \frac{1}{\omega C}\right) \tag{14.40}$$

The loss $(1/|S_{21}|^2)$ due to this shunt impedance Z can be obtained by using Eq. (12.5) of Chapter 12:

$$\text{Loss} = 1 + \frac{R}{Z_0} + \frac{1}{4}\left(\frac{R}{Z_0}\right)^2 + \frac{1}{4}\left(\frac{X}{Z_0}\right)^2 \tag{14.41}$$

where Z_0 is the characteristic impedance of the transmission line. Substituting Z from (14.40) in the above equation yields

$$\text{Loss} = 1 + \frac{R}{Z_0} + \frac{1}{4}\frac{R^2}{Z_0^2} + \frac{1}{4}\frac{\omega^2 L^2}{Z_0^2} + \frac{1}{4}\frac{1}{\omega^2 C^2 Z_0^2} - \frac{1}{2}\frac{L}{CZ_0^2} \tag{14.42}$$

This general equation can be used to calculate the insertion loss and isolation by substituting the corresponding parameter values. In the up position, $L = L_u$, $C = C_u$, and $R = R_u$; whereas, in the down position, $L = L_d$, $C = C_d$, and $R = R_d$.

The insertion loss for the resistive contact switch in the down position (ON state) is obtained by substituting $C_d = \infty$ in (14.42):

$$\text{IL} = 1 + \frac{R_d}{Z_0} + \frac{R_d^2}{4Z_0^2} + \frac{\omega^2 L_d^2}{4Z_0^2} \tag{14.43}$$

When $R_d \ll Z_0$, the insertion loss (14.43) can be approximated as

$$\text{IL} \cong \begin{cases} 1 + \dfrac{R_d}{Z_0} & R_d \gg \omega L_d \\ 1 + \dfrac{\omega^2 L_d^2}{4Z_0^2} & \omega L_d \gg R_d \end{cases} \tag{14.44}$$

The insertion loss for the capacitive contact switch in the down position (ON state) by neglecting inductance and resistance (i.e., $\omega L_d \ll 1/\omega C_d$, $R_d \ll 1/\omega C_d$, and $R_d \ll Z_0$) is given by

$$\text{IL} = 1 + \frac{1}{4\omega^2 C_d^2 Z_0^2} \tag{14.45}$$

The isolation of the switch can be computed from (14.42) by using the corresponding parameter values in the up position (OFF state). When the inductance and resistance can be neglected, the isolation can be obtained from (14.45) by replacing C_d by C_u.

In a capacitive contact series switch, both the insertion loss and isolation decrease with increase in frequency, as shown in Fig. 14.17*b*. Consequently, the lower operating frequency f_L is limited by the value of maximum insertion loss acceptable and the upper operating frequency f_U is limited by the value of minimum isolation required. In a resistive contact series switch, the insertion loss increases with frequency (unless otherwise affected by parasitic capacitance) and the isolation decreases with increase in frequency similar to a capacitive series switch. Thus, the switch has only an upper operating frequency f_U limited by the required minimum isolation and/or maximum insertion loss.

Example Capacitive Contact Switch Design. Let us consider the design of capacitive contact shunt and series switches on a transmission line of characteristic impedance $Z_0 = 50\ \Omega$ for the following specifications: operating frequency range 5–40 GHz, maximum insertion loss 0.2 dB, and minimum isolation 10 dB. We need to find the required capacitance ratio and up- and down-position capacitances and calculate the required thickness of the Si_3N_4 dielectric layer ($\epsilon_D = 7.6$). The up-position gap height is given to be 3 μm. Also, we can calculate the required plate area (neglecting fringing capacitance) for the configuration of Fig. 14.15*a*(i) and hence the plate length for a width $W = 50$ μm. Assume that the inductance and the resistance are negligible in both up and down positions.

(i) *Shunt Switch Design.* The lower operating frequency is limited by the minimum isolation in the down position [obtained from (14.36)]:

$$10 \log\left|1 + \frac{\omega_L^2 C_d^2 Z_0^2}{4}\right| = \text{IS}_{\min} \qquad \text{(dB)} \tag{14.46}$$

The upper operating frequency is limited by the maximum insertion loss in the up position [also given by (14.36) for capacitive switch] as

$$10 \log\left|1 + \frac{\omega_U^2 C_u^2 Z_0^2}{4}\right| = \text{IL}_{\max} \qquad \text{(dB)} \tag{14.47}$$

The capacitance ratio is obtained using the above equations to be

$$\zeta = \frac{C_d}{C_u} = \left(\frac{f_U}{f_L}\right)\sqrt{\frac{10^{(\text{IS}_{\min}/10)} - 1}{10^{(\text{IL}_{\max}/10)} - 1}} = \left(\frac{40 \times 10^9}{5 \times 10^9}\right)\sqrt{\frac{10^{(10/10)} - 1}{10^{(0.2/10)} - 1}} \cong 111 \tag{14.48}$$

The up-position capacitance C_u is calculated using (14.47):

$$C_u = \frac{\sqrt{10^{(\mathrm{IL_{max}}/10)} - 1}}{\pi f_U Z_0} = \frac{\sqrt{10^{(0.2/10)} - 1}}{\pi \times 40 \times 10^9 \times 50} \cong 35 \text{ fF} \tag{14.49}$$

The down-position capacitance $C_d = \zeta C_u = 3.8$ pF. The dielectric thickness $h_D \cong 0.2$ μm [using (14.32)] and the required electrode area $A \cong 11{,}710$ μm^2 using (14.30). For a given electrode width $W = 50$ μm, the required plate length $L = 234$ μm.

(ii) *Series Switch Design.* The lower operating frequency is limited by the maximum insertion loss in the down position [given in (14.45)] as

$$10 \log \left| 1 + \frac{1}{4\omega_L^2 C_d^2 Z_0^2} \right| = \mathrm{IL_{max}} \quad \text{(dB)} \tag{14.50}$$

The upper operating frequency is limited by the minimum isolation in the up position [also given by (14.45) for capacitive switch] as

$$10 \log \left| 1 + \frac{1}{4\omega_U^2 C_u^2 Z_0^2} \right| = \mathrm{IS_{min}} \quad \text{(dB)} \tag{14.51}$$

The required capacitance ratio can be obtained from the above equations and is

$$\zeta = \frac{C_d}{C_u} = \left(\frac{f_U}{f_L}\right) \sqrt{\frac{10^{(\mathrm{IS_{min}}/10)} - 1}{10^{(\mathrm{IL_{max}}/10)} - 1}} = \left(\frac{40 \times 10^9}{5 \times 10^9}\right) \sqrt{\frac{10^{(10/10)} - 1}{10^{(0.2/10)} - 1}} \cong 111 \tag{14.52}$$

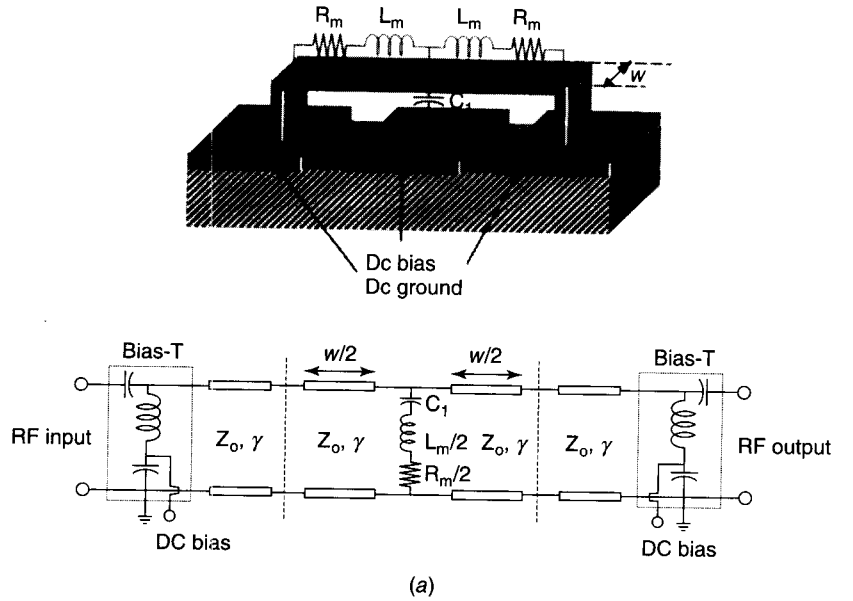
It is interesting to note that the capacitance ratio required for designing a capacitive series or shunt switch is the same for a given specification. The up-position capacitance C_u can be obtained from (14.51) as

$$C_u = \frac{1}{4\pi f_U Z_0} \frac{1}{\sqrt{10^{(\mathrm{IS_{min}}/10)} - 1}} = \frac{1}{4 \times \pi \times 40 \times 10^9 \times 50} \frac{1}{\sqrt{10^{(10/10)} - 1}} \cong 13 \text{ fF} \tag{14.53}$$

Then, the down-position capacitance is calculated as $C_d = \zeta C_u \cong 1.5$ pF. The required dielectric thickness h_D can be calculated using (14.32) to be

$$h_D \cong \frac{\epsilon_D h_A}{\zeta} = \frac{7.6 \times 3 \times 10^{-6}}{111} \cong 0.2 \text{ μm} \tag{14.54}$$

The switch top electrode area can be obtained from (14.30) as



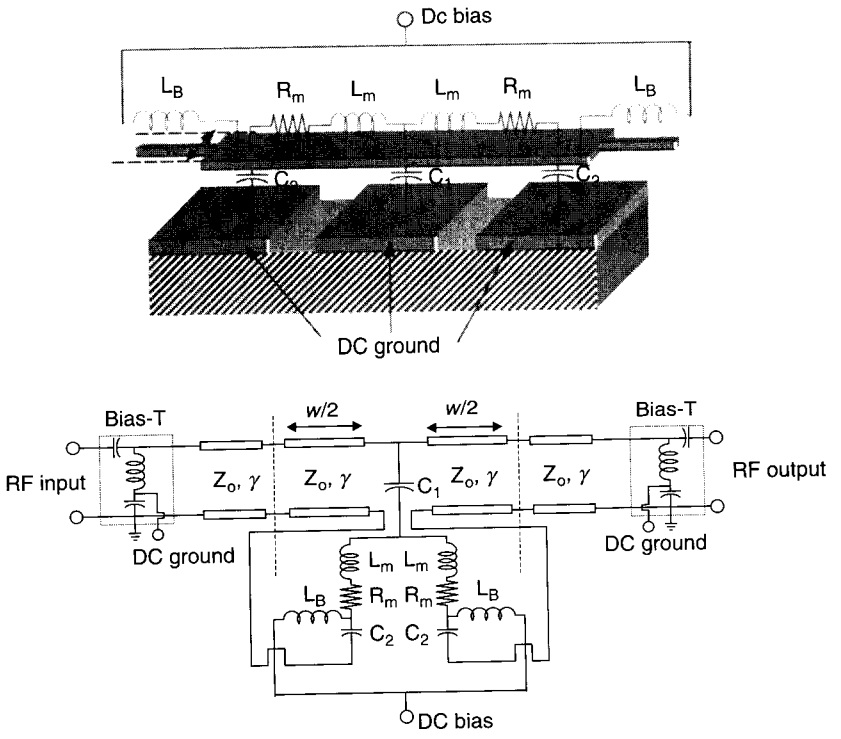
(*a*)

Figure 14.18 Three types of bias arrangements: (*a*) and (*b*) shunt-mounted RF MEMS switches in CPW lines, and (*c*) series-mounted RF MEMS switch in microstrip line. (*w*—width of the top electrode, l_1 and l_2 are line lengths shown).

$$A = \frac{h_D C_d}{\epsilon_0 \epsilon_D} = \frac{0.2 \times 10^{-6} \times 1.5 \times 10^{-12}}{8.854 \times 10^{-12} \times 7.6} \cong 4500\ \mu\text{m}^2 \tag{14.55}$$

The length L must be 90 μm for a plate width $W = 50$ μm.

Biasing Arrangements. A biasing arrangement is needed to apply DC bias voltage for operation of electrostatically actuated RF MEMS devices. Two biasing arrangements are shown in Fig. 14.18 for two configurations of shunt-mounted RF MEMS devices on a coplanar waveguide. Frequently, the same set of electrodes (top and bottom) is used for actuation as well as for RF operation in a configuration such as Fig. 14.15*a*(i) shown again in Fig. 14.18*a*. In this configuration, the center conductor of the CPW line serves as the bottom actuation electrode. The biasing of the top electrode (connected to CPW ground planes) is accomplished by applying a DC voltage between the center conductor and the ground plane of the CPW line. This arrangement needs the use of bias-T's at the input and output ports of the circuit (as shown in Fig. 14.18*a*) to protect the rest of the circuit or measurement system by blocking DC bias voltage from the RF MEMS device. Another type of biasing arrangement for configuration in Fig. 14.15*a*(ii) is shown in Fig. 14.18*b*. In this case, the signal conductor and the ground planes of the CPW line serve as DC ground and the top electrode is DC biased through inductive or resistive con-



(*b*)

Figure 14.18 *(Continued)*

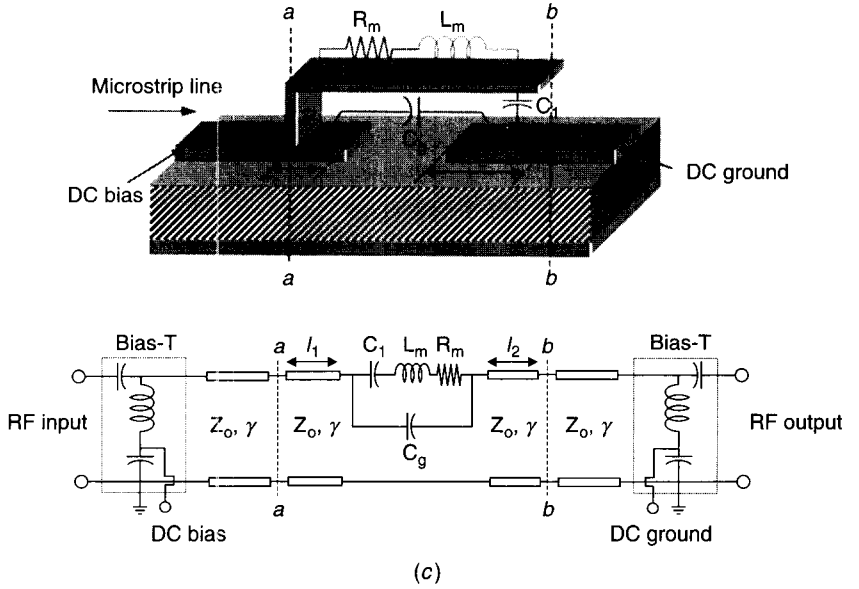
nection (a thin strip of metal). Figure 14.18*c* depicts another biasing configuration suitable for microstrip series switches shown in Fig. 14.16*a*(i). Similar biasing arrangements can be implemented for other configurations shown in Figs. 14.15 and 14.16.

Examples of Switches. Two examples of electrostatically actuated switches are discussed here.

Capacitive Contact Switch [26]. This switch (shown in Fig. 14.19) can be described as follows:

Mechanical structure—bridge (fixed–fixed beam)
Contact type—capacitive
RF circuit configuration—shunt switch mounted on a coplanar waveguide

The switch is manufactured on a high-resistivity (>5000 Ω-cm) silicon wafer using a surface micromachining technique with a total of five masking levels.



(c)

Figure 14.18 *(Continued)*

The details of the fabrication process are described elsewhere [26]. The 50 Ω coplanar waveguide on a high resistivity silicon substrate has a center conductor width of 120 μm and two gaps of 80 μm. The switch has dimensions of 120 μm width and 280 μm length. The gap height between switch membrane and the lower electrode is ~3.4 μm. The pull-down voltage of the switch is about 50 V. The typical switching time is on the order of 3.5–5.3 μs. A capacitance ratio of 80–110 was obtained with an off-capacitance range of 30–40 fF and an on-capacitance range of 3.2–3.5 pF. The insertion loss and isolation values for the switch are 0.15 and 15 dB, respectively, at 10 GHz and 0.28 and 35 dB, respectively, at 35 GHz. A switch lifetime of 5×10^8 cycles has been reported.

Resistive Contact Switch [27]. This switch (Fig. 14.20) can be categorized as follows:

Mechanical structure—cantilever
Contact type—resistive
RF circuit configuration—series mounted switch on a microstrip line

The switch is fabricated on a GaAs substrate by surface micromachining process. The microstrip line was fabricated by the evaporation of gold followed by plasma-enhanced chemical vapor deposition (PECVD) of silicon dioxide. The cantilever was formed by a PECVD nitride–gold–nitride trilayer. A hole was etched in the nitride structural layer, resulting in a gold dimple that contacts

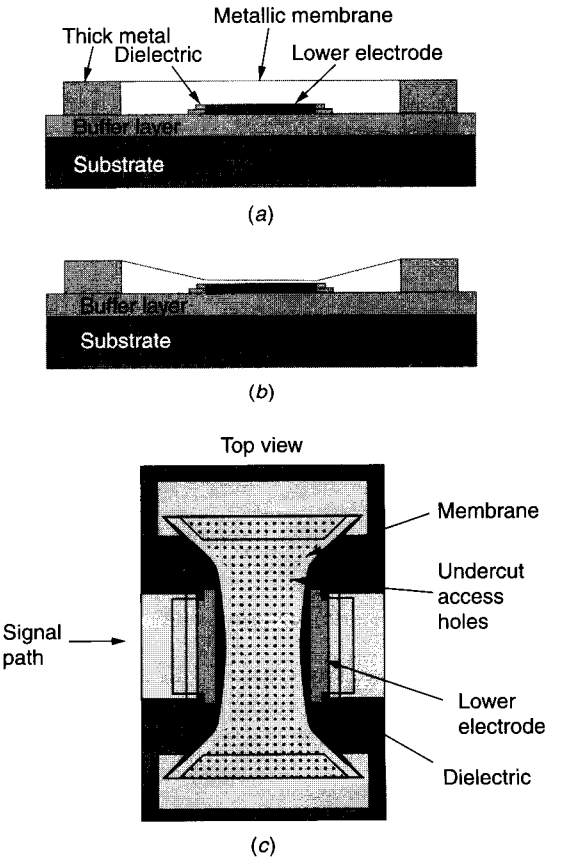


Figure 14.19 RF MEMS capacitive contact switch: (*a*) in up position; (*b*) in down position; (*c*) microphotograph of switch in up position [26].

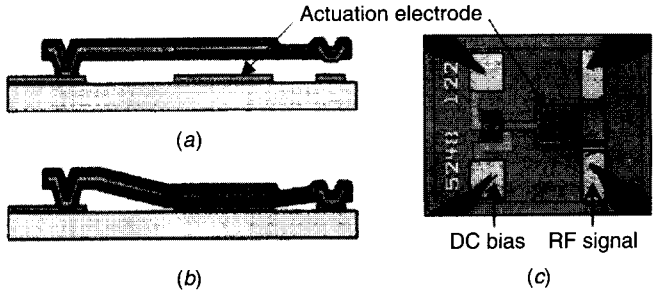


Figure 14.20 RF MEMS resistive contact switch: (*a*) switch in up position; (*b*) switch in down position; (*c*) plan-view microphotograph [27].

the microstrip. The SiO_2 sacrificial layer was removed using hydrofluoric acid followed by a critical point dry. The metal beam RF electrode is mounted on the free end of the cantilever. The insertion loss is less than 0.2 dB from DC to 40 GHz. The isolation is greater than 50 dB at low frequencies and it decreases to 27 dB at 40 GHz. The switch has a pull-down voltage of 30 V, a response time of 20 μs, and a lifetime of 10^9 cycles.

The device characteristics and performance of a selected number of RF MEMS switches are summarized in Table 14.3.

14.4.2 Varactors

The MEMS varactors (or variable capacitor) are widely useful for their tuning functionality in RF communication circuit applications. In MEMS varactors, the capacitance is changed by adjusting the physical dimensions of the device. In these capacitors, the dielectric layer used is usually air, which eliminates the majority of the dielectric losses and results in high Q values. Various possible classifications of RF MEMS varactors based on mechanical structure, actuation mechanism, tuning scheme, capacitance change, and RF circuit configuration are shown in Fig. 14.21. Two types of mechanical structures have been investigated for MEMS varactors: parallel-plate and interdigital structures. Various actuation mechanisms such as electrostatic, electrothermal, and piezoelectric have been explored for MEMS varactors. The MEMS varactors have been designed for analog and digital capacitance variations. In analog designs, the capacitance value can be changed continuously, whereas in digital designs the capacitance value is changed step wise.

As shown in Fig. 14.21, MEMS varactors can be classified based on their tuning schemes: gap tuning and area tuning (referring to the parameter that is being changed). The tuning range (TR) can be defined as

$$\text{TR} = \frac{C_2 - C_1}{C_1} \tag{14.56}$$

where C_1, C_2 are the capacitances in the two states.

The tuning range for a parallel-plate capacitor can be obtained using $C = \epsilon A/d$ (neglecting the fringing effect) for gap- and area-tuning varactors:

$$\text{TR} = \frac{d_1 - d_2}{d_2} \quad \text{(gap tuning)} \tag{14.57}$$

where d_1, d_2 are the distances between the plates in the two states;

$$\text{TR} = \frac{A_2 - A_1}{A_1} \quad \text{(area tuning)} \tag{14.58}$$

where A_1, A_2 are the areas of the capacitor in the two states.

Table 14.3 Characteristics and Performances of Selected RF MEMS Switches[a]

Device Characteristics and Performance Parameters	FET Switch (Typical Values)	Raytheon/TI[b]	HRL Labs[c]	Rockwell RSC[d]	NEC Corporation[e]	University of Michigan[f]	University of California at Berkley[g]
MEMS technology	—	Surface	Surface	Surface	Bonded wafer	Surface	Surface + bonding
Actuation structure	—	Fixed–fixed	Cantilever	Cantilever	Cantilever	Fixed–fixed	Fixed–fixed
Actuation mechanism	—	Electrostatic	Electrostatic	Electrostatic	Electrostatic	Electrostatic	Electrostatic
Contact type	Semiconductor	Capacitive	Resistive	Resistive	Resistive	Capacitive	Resistive
RF configuration	—	Shunt	Series	Series	Series	Shunt	Series, shunt
Device size (μm × μm)	~1000 × 1000	120 × 280	120 × 300	80 × 160	250 × 900	~270 × 300	~160 × 160
Actuation voltage (V)	~1	~50	~60	~60	125	12–25	50–60
Insertion loss (dB)	2 at 6 GHz	0.15 at 10 GHz, 0.28 at 35 GHz	0.2 (DC, 40 GHz)	0.2 (DC, 40 GHz)	0.2 at 30 GHz	<0.2 (1–13 GHz)	Series: <0.3 Shunt: <0.3 100 MHz–50 GHz
Isolation (dB)	22 at 2 GHz 20 at 6 GHz	15 at 10 GHz 35 at 35 GHz	40 at 12 GHz 27 at 40 GHz	32 at 10 GHz 22 at 40 GHz	13 at 30 GHz	>30 (11–13 GHz)	Series: >15 Shunt: >45 100 MHz–50 GHz
Switching time	10 ns	3.5–5.3 μs	20 μs	2–5 μs	—	—	T_{OFF} < 20 μs T_{ON} ~ 200 μs
Lifetime (cycles)	$>10^{11}$	5×10^8	$\sim 4 \times 10^8$	$\sim 10^8$	—	—	2×10^5

[a] Partially based on Yao [28].

[b] From Goldsmith et al. [26].

[c] From Hyman et al. [27].

[d] From Yao and Chang [28, 29].

[e] From Suzuki et al. [30].

[f] From Muldavin and Rebeiz [31, 32].

[g] From Veljko et al. [10].

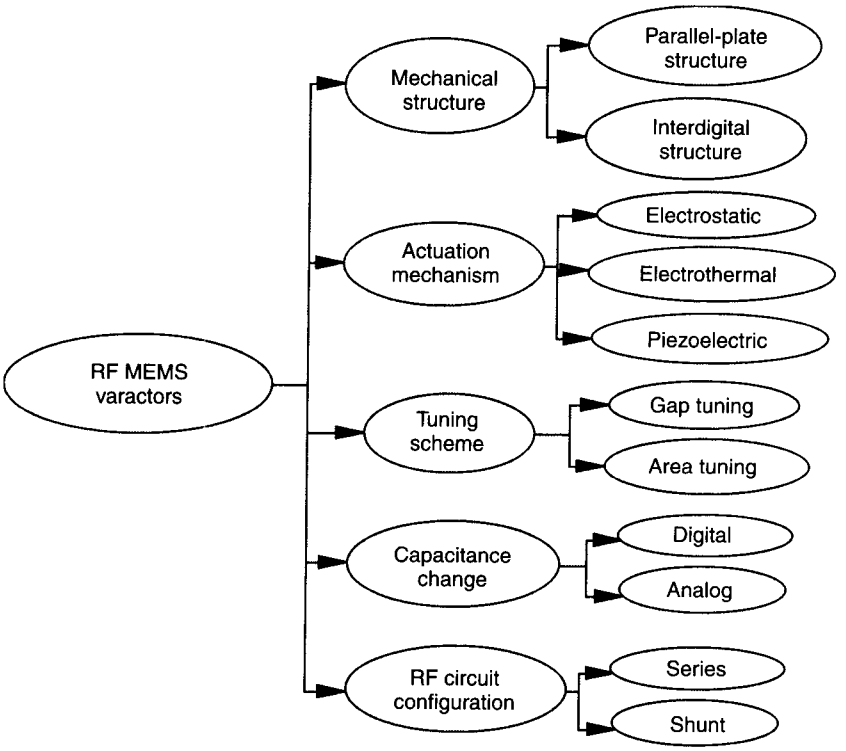


Figure 14.21 Classification of RF MEMS varactors.

Parallel-Plate Varactors. MEMS parallel-plate varactors have been developed using (i) two parallel plates and (ii) three parallel plates.

(i) *Two-Parallel-Plate Varactor.* A parallel-plate actuator with an initial gap height of z_0 is shown in Fig. 14.22*a*. As discussed in Section 14.1.1, the parallel actuator becomes unstable for gap height less than the critical gap height $2/3z_0$. So, the maximum tuning range is limited by the critical gap height and is calculated to be 50% using (14.57). A parallel-plate system based gap-tuning capacitor fabricated using surface micromachining technique, shown in Fig. 14.22*b*, was reported in 1996 [33]. A stationary bottom electrode is fabricated on the substrate, and the movable top electrode is suspended over the stationary electrode by flexures (springs) at each corner. The top electrode can move in the vertical direction normal to the substrate. The gap height can be controlled (electrostatically) by applying a voltage between the electrodes, resulting in a change in the device capacitance. The fabrication details are described elsewhere [33]. The electrode size and gap height are 200×200 μm^2 and 1.5 μm, respectively. The device exhibited a tuning range of 16% by a change of capacitance from 2.11 pF at 0 V to 2.46 pF at 5.5 V.

(ii) *Three-Parallel-Plate Varactor.* In the three-parallel-plate varactor (shown in Fig. 14.23*a*), the bottom and the top plates are fixed and the middle

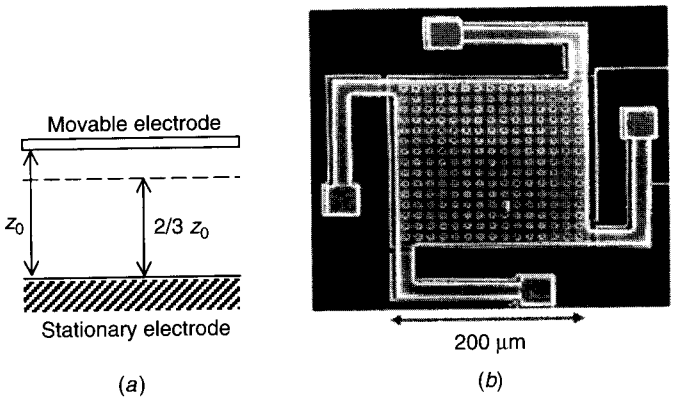


Figure 14.22 Parallel-plate varactor: (*a*) cross section showing critical gap height; (*b*) top view of surface micromachined varactor [33].

plate moves up and down. If z_0 is the initial gap height between the plates, the movement of the middle plate must be within the range $z_1 = \frac{4}{3}z_0$ to $z_2 = \frac{2}{3}z_0$ for stable operation. Thus, the maximum tunable range of capacitance between the moving electrode and any one of the two stationary electrodes is 100% from (14.57).

A three-parallel-plate varactor, shown in Fig. 14.23*b*, has been reported in the literature [34]. The capacitances C_1 (between the top plate and the movable plate) and C_2 (between the bottom plate and the movable plate) can be controlled either by varying voltages V_1 or V_2, separately. The application of V_1 will increase the C_1 value, and that of V_2 will decrease C_1. The tuning range of C_1 with respect to V_1, as well as that of C_2 with respect to V_2, still remains at 50%. However, the application of V_2 also reduces C_1 by approximately 50%,

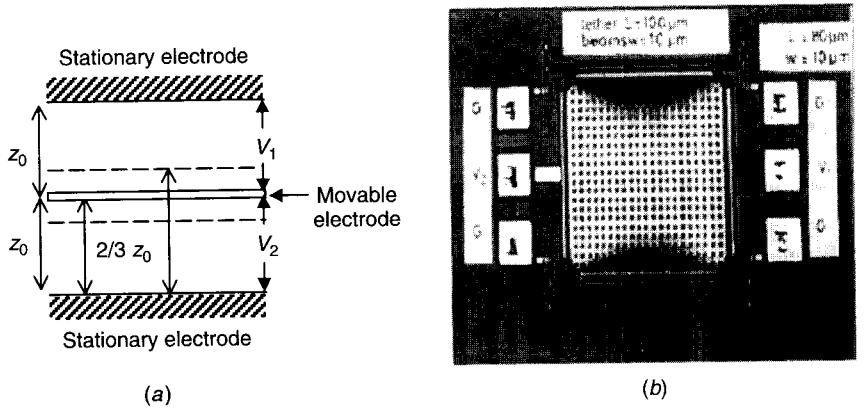


Figure 14.23 Three parallel-plate varactor: (*a*) schematic showing critical gap heights; (*b*) top-view image of surface micromachined varactor [34].

and the combined effect of changes in C_1 and C_2 results in a total 100% tuning range for C_1. The device was fabricated using a multi-user MEMS polysilicon process (MUMPs) [4]. The capacitor plates are 426 × 426 μm^2, with air gaps of the order of 0.75 μm. The capacitance of the signal capacitor C_1 varied from 3.5 pF with $V_1 = 0$ V and $V_2 = 1.8$ V to 4.39 pF with $V_1 = 0.7$ V and $V_2 = 0$ V, corresponding to a tuning range of 25%. The Q of the device was reported to be between 9.6 and 13.2 at 1 GHz.

Interdigital Varactor. In the interdigital varactor shown in Fig. 14.11, two sets of interdigital (or comb) structures are interlaced to form a variable capacitor. The tuning range can be obtained substituting capacitance C from (14.24) in (14.58) and is given by $\text{TR} = (l_2 - l_1)/l_1$, where l_1 and l_2 are the overlapping length of the fixed and movable interdigital structures in two positions. It can be noted that in this configuration the length of the fingers limits the maximum tuning range, and a pull-down instability similar to that in the case of parallel-plate capacitors can be avoided.

In 1998, Yao et al. [35] reported the interdigital MEMS area-tuning varactor shown in Fig. 14.24. The capacitor consists of many parallel interdigitated structures fabricated using a deep reactive-ion etch of single-crystal silicon substrate followed by coating metal thin films to form electrodes. The size of a typical tunable capacitor is 1 mm^2. The dimension of an interdigital finger is 2 μm (width) × 2 μm (spacing) and 20–30 μm (height). The supporting structures are typically 1 mm in length and have the same width and height as the interdigital fingers. The details of the fabrication process are described elsewhere [35]. A tuning range of about 100% was demonstrated by a change of the total device capacitance (including all parasitic effects) from 5.19 pF at 0 V to 2.48 pF at 5 V.

Digital Varactors. The n-bit MEMS digital varactors have been designed by two methods: (i) using only n basic capacitor elements and (ii) using $2^n - 1$ basic capacitor elements to obtain a linear capacitance–voltage characteristic.

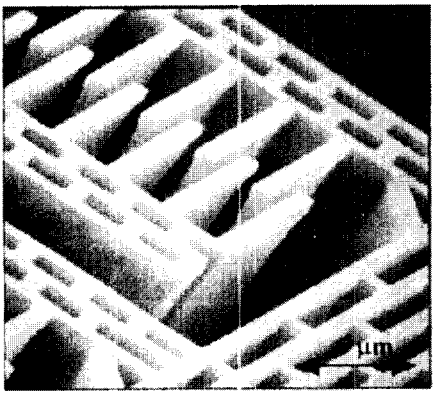


Figure 14.24 Microphotograph of MEMS interdigital varactor [28].

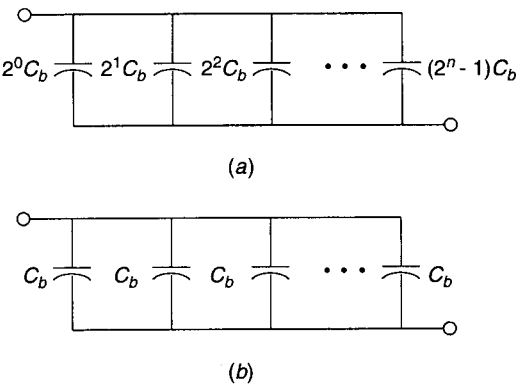


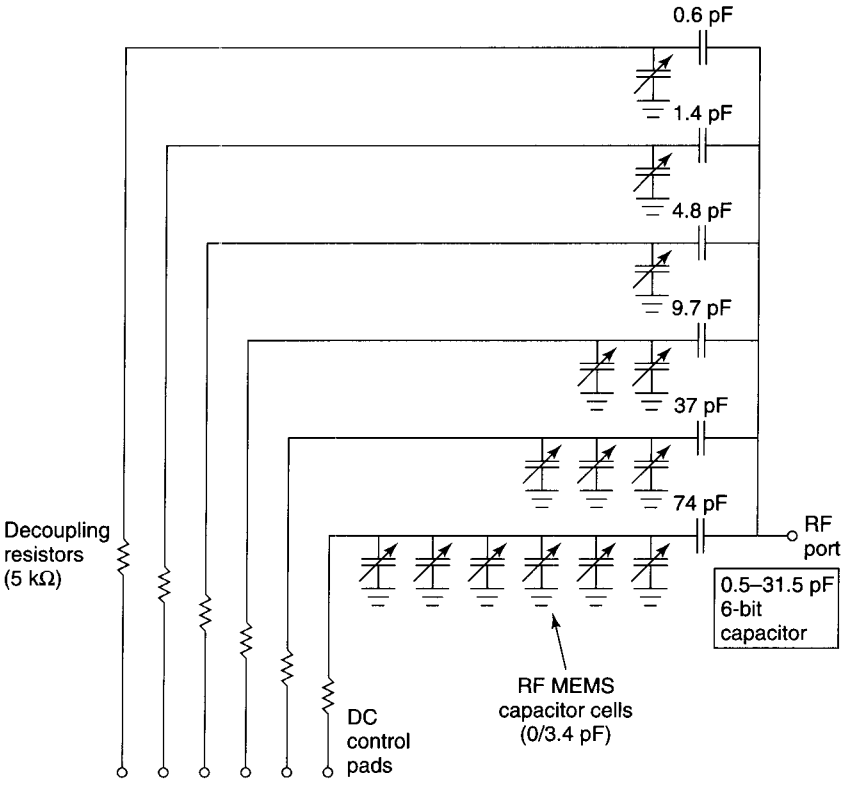
Figure 14.25 Two configurations for realization of n-bit digital MEMS varactor: (*a*) using n capacitors of different values; (*b*) using $2^n - 1$ capacitor elements of equal values.

Method I. An n-bit digital capacitor can be realized by a parallel combination arrangement of n basic capacitor elements (as shown in Fig. 14.25*a*), each representing a bit. The total capacitance is given by

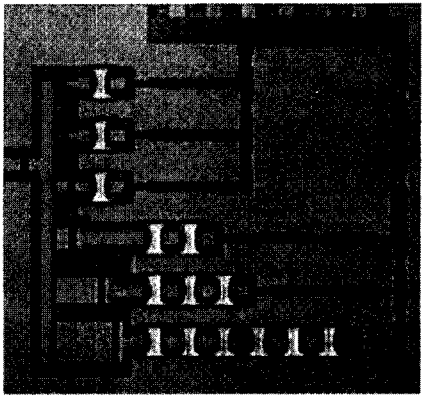
$$C = C_0 + \eta_1 2^0 C_b + \eta_2 (2^1 C_b) + \cdots + \eta_n (2^n C_b) \qquad \eta_n = \begin{cases} 0 & \text{unactuated state} \\ 1 & \text{actuated state} \end{cases} \tag{14.59}$$

where η_n represents the state of the nth element (actuated or unactuated), C_b is the capacitance contributed by the smallest bit, and C_0 is the capacitance of this arrangement when all the elements are in the unactuated state. The various combinations of n element states in this arrangement yield 2^n capacitor states. For example, a 3-bit digital capacitor can be designed using C_b, $2C_b$, and $4C_b$ representing 001, 010, and 100 digital bits, respectively. A parallel combination of these three basic bits can provide 8 ($=2^3$) states. A minimum capacitance of $C = C_0$, and a maximum capacitance of $C = 7C_b + C_0$ is obtained when all the elements are actuated. The value of tuning ratio TR obtainable in this case is $(7C_b + C_0)/C_0 = 1 + (7C_b/C_0)$.

In 1999 Goldsmith et al. [36] reported a voltage-controlled 6-bit variable capacitor constructed using switchable MEMS capacitors as the control element. In this design, each capacitor bit is constructed as a single branch containing a series combination of one or more MEMS capacitors and a fixed capacitor, as shown in Fig. 14.26*a*. For example, the largest bit consists of six switchable MEMS capacitors at 3.4 pF each for a total of 20 pF in series with a 74-pF fixed capacitor to yield 16 pF. Similarly, smaller bits are constructed with total capacitances of 8, 4, 2, 1, and 0.5 pF in respective branches. For bits 1–3 that have a total capacitance of less than 3.4 pF, a single MEMS capacitor in series with a fixed capacitor is used. A microphotograph of a fabricated device of size 3.2 mm × 3.2 mm × 0.53 mm is shown in Fig. 14.26*b*. A tuning



(*a*)



(*b*)

Figure 14.26 Six-bit digital varactor: (*a*) schematic; (*b*) microphotograph of fabricated device [36].

range of 1:22 was demonstrated by tuning the total device capacitance from 1.5 to 33.2 pF. Bistable MEMS capacitors used in this work operate up to 40 GHz and exhibit individual tuning ranges of 70:1–100:1. Switch pull-down voltages in the range 30–55 V and switching speeds less than 10 μS were reported.

Method II. An n-bit digital capacitor can be designed to obtain a linear capacitance–actuation voltage characteristic by using $2^n - 1$ basic capacitor elements (as shown in Fig. 14.25*b*). In this design, the capacitance at the mth state is given by

$$C_m = mC_b + C_0 \tag{14.60}$$

where $m = 0, 1, \ldots, 2^n - 1$, C_b is the actuated state capacitance of each element, and C_0 is the total capacitance in unactuated states of all the elements. In this design [37], a single control voltage is used for actuation and different capacitance elements are designed to pull down at different values of this single control voltage by designing these elements to have varying values of spring constant k. The linear relationship between overall capacitance and the actuation voltage for this design can be written as

$$C = \alpha V + \beta \tag{14.61}$$

where α and β are constants. Using the condition $C = C_0$ (up-position device capacitance) when $V = 0$ in (14.61) gives $\beta = C_0$. Substituting $\beta = C_0$ in (14.61) yields

$$C = \alpha V + C_0 \tag{14.62}$$

The mth-state capacitance is given by

$$C_m = \alpha V_m + C_0 \tag{14.63}$$

Rearranging the above equation to obtain an expression for V_m and substituting C_m from (14.60) give

$$V_m = \frac{C_m - C_0}{\alpha} = \frac{mC_b}{\alpha} \tag{14.64}$$

The ratio between the pull-down voltages at the $(m + 1)$th state and mth state can be obtained from (14.64),

$$\frac{V_{m+1}}{V_m} = \frac{m+1}{m} \quad \text{(or)} \quad V_{m+1} = (m+1)V_1 \tag{14.65}$$

where $m = 1, 2, \ldots, 2^n - 1$ and V_1 is the pull-down voltage for the first nonzero state.

If the area of the plate A and the gap height z_0 are the same to realize C_b in each capacitor state, the ratio between the pull-down voltages at the $(m + 1)$th

state and mth state is obtained using (14.13),

$$\frac{V_{m+1}}{V_m} = \sqrt{\frac{k_{m+1}}{k_m}} \tag{14.66}$$

where $m = 1, 2, \ldots, 2^n - 1$. From (14.20) and (14.66), the relationship between the spring constants at the $(m+1)$th state and mth state (to achieve a linear capacitance–actuation voltage characteristic) can be expressed as

$$k_{m+1} = \left(\frac{m+1}{m}\right)^2 k_m \quad \text{(or)} \quad k_{m+1} = (m+1)^2 k_1 \tag{14.67}$$

where $m = 1, 2, \ldots, 2^n - 1$ and k_1 is the spring constant for the first nonzero state.

In Hoivik et al. [37] a digitally controllable, electrostatically actuated, variable MEMS capacitor constructed using a flip-chip bonding technique (shown in Fig. 14.27) has been reported. The capacitor consists of an array of individual plates of equal area that are connected to the bonding pads by flexures (springs) of different widths. This design exhibits a cascading snap-down effect when the applied voltage is increased. The capacitor plates are fabricated using the MUMPS process [4]. This process includes three layers of polysilicon, two layers of oxide, and one layer of gold. The top capacitor plate was constructed using polysilicon(1) layer (2.0 μm thick), an oxide(2) layer (0.75 μm thick), polysilicon(2) layer (1.5 μm thick), and a gold layer (0.5 μm thick). The oxide(1) layer is used for release that separates the substrate and polysilicon(0) layer from the capacitor plates. The top electrode of the capacitor consists of an array of 30 individual plates of equal area connected to the bonding pads by springs with different widths (Fig. 14.27*c*). The device is transferred onto an alumina substrate by the flip-chip bonding technique described in Section 14.2, and the silicon MUMPS substrate is released [by dissolving the oxide(1) layer]. The gap height between the top and bottom plates is determined by the initial solder bump height and the bonding conditions. A typical gap height is 2 μm. The size of the device is 0.5×1.0 mm^2 and it is flip-chip bonded onto an alumina substrate of size 5×5 mm^2. The measured Q factor of the device is 140 for a capacitance value of 0.8 pF at 745 MHz and the tuning ratio is approximately 3:1. The total up-position capacitance is in the range of 0.8–1.0 pF.

The device characteristics and performance of a selected number of RF MEMS varactors are summarized in Table 14.4.

14.5 RF MEMS CIRCUIT APPLICATIONS

Various applications of RF MEMS devices and components in circuits and antennas are shown in Fig. 14.28. Different RF MEMS circuit applications

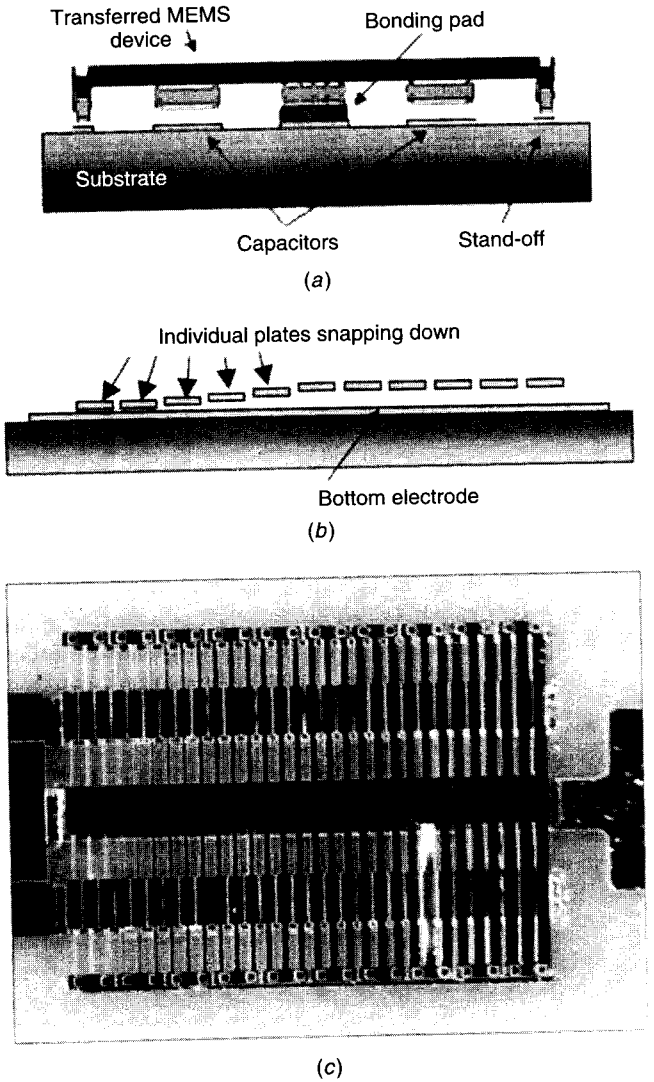


Figure 14.27 Digitally controllable variable capacitor: (*a*) cross section of MEMS capacitor flip-chip mounted on substrate; (*b*) cross-sectional view of device; (*c*) photograph of typical device [37].

such as phase shifters, tunable filters, impedance tuners, oscillators, and micromachined components are discussed in this section.

14.5.1 Phase Shifters

The low insertion loss and high isolation performance exhibited by RF MEMS switches have been used for the design of low-loss phase shifters.

Table 14.4 Characteristics and Performances of Selected RF MEMS Varactors[a]

Device Characteristics and Performance Parameters	Varactor Diodes	University of California[b]	Columbia University[c]	Rockwell RSC[d]	Raytheon Systems[e]	University of Colorado[f]	University of Colorado[g]	LG Electrical Institute of Technology[h]	University of Illinois[i]
MEMS technology	—	Surface	Surface	Bonded wafer	Surface	Surface + bonding	Surface + bonding	Surface + bonding	Surface
Structural material	Silicon	Aluminum	Polysilicon	Silicon	Gold	Polysilicon	Polysilicon	Metal/PZT/metal	Permalloy
Actuation structure	—	Parallel-plate	Parallel-plate	Interdigital	Parallel-plate	Parallel-plate	Parallel-plate	Parallel-plate	Parallel-plate
Actuation mechanism	HBT junction	Electrostatic	Electrostatic	Electrostatic	Electrostatic	Electrostatic	Electrothermal	Piezoelectric	Electrostatic
Tuning scheme	Analog	Gap tuning	Gap tuning	Area tuning	Gap tuning	Area tuning	Gap tuning	Gap tuning	Gap tuning
Capacitance change	—	Analog	Analog	Analog	Digital	Digital	Analog	Analog	Analog
Device size (μm × μm)	—	~800 × 800	~400 × 400	~1000 × 2000	3200 × 3200	500 × 1000	~500 × 800	~150 × 150	~500 × 500
Actuation voltage (V)	4	3	3.3	2–14	30–55	35	2.5	6	15
Min. capacitance (pF)	2 at 4 V	2	2.05	1.8	1.5	0.2–0.5	0.331	—	~0.085
Quality factor (Q)*	350 at 50 MHz	60 at 1 GHz	20 at 1 GHz	34 at 500 MHz	<20 at 1 GHz	140 at 750 MHz	108 at 1 GHz	210 at 1 GHz	30 at 5 GHz
Tuning range	335	15%	1.5:1	200	22:1	4:1	2.7:1	3.1:1	69.8%
Series resistance (Ω)	2.8	1.2 at 1 GHz	—	1.8	0.3–0.5	—	~4.0	—	—
Series inductance (nH)	1.5	—	—	0.19	—	—	0.12–0.16	—	—
Self-resonance (GHz)	~1.3	—	—	5	—	—	14–21 GHz	—	>5 GHz

* At minimum capacitance value.

[a] Partially based on Yao [28].

[b] From Young et al. [33, 51].

[c] From Dec and Suyama [34].

[d] From Yao et al. [35].

[e] From Goldsmith et al. [36].

[f] From Hoivik et al. [37].

[g] From Feng et al. [11].

[h] From Park et al. [12].

[i] From Zou et al. [38].

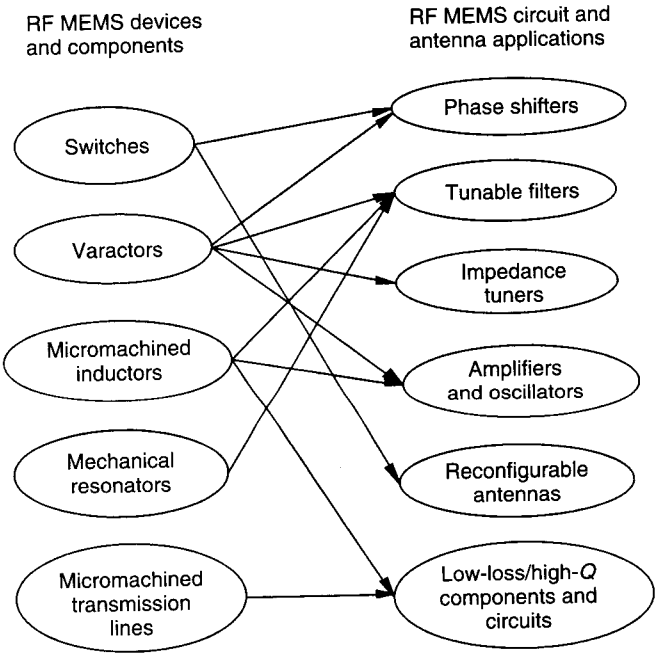


Figure 14.28 Applications of RF MEMS devices and components in circuits and antennas.

Periodically Loaded-Line Phase Shifter. RF MEMS capacitors have been utilized to realize phase shifters by periodically loading a high-impedance transmission line, known as a distributed MEMS transmission line (DMTL) (shown in Fig. 14.29). The transmission line loaded periodically by shunt capacitors exhibits a low-pass filter behavior with the equivalent line impedance Z_l [39] given by

$$Z_l = \sqrt{\frac{sL_t}{sC_t + C_M}} \sqrt{1 - \left(\frac{\omega}{\omega_C}\right)^2} \tag{14.68}$$

and the phase constant β is given by

$$\beta \cong \frac{1}{s}\omega\sqrt{sL_t(sC_t + C_M)}\left(1 + \frac{\omega^2}{6\omega_C^2}\right) \tag{14.69}$$

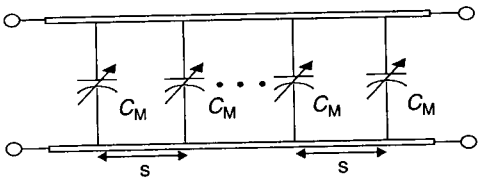


Figure 14.29 Transmission line periodically loaded with MEMS capacitors.

where ω is the operation frequency and ω_C is the cutoff frequency given by

$$\omega_C = \frac{2}{\sqrt{sL_t(sC_t + C_M)}} \tag{14.70}$$

where L_t and C_t are the per-unit length inductance and capacitance of the unloaded transmission line, respectively; C_M is the loading capacitance of the MEMS capacitor; and s is the periodic spacing between the MEMS capacitors. The phase velocity can be obtained using (14.69) as

$$v = \frac{\omega}{\beta} \cong \frac{s}{\sqrt{sL_t(sC_t + C_M)}(1 + \omega^2/(6\omega_C^2))} \tag{14.71}$$

Derivation of relations (14.68)–(14.71) is given in the Appendix to this chapter.

The phase shift per unit length is obtained from the change in phase constant and is given by

$$\Delta\phi = \beta_1 - \beta_2 = \omega\left(\frac{1}{v_1} - \frac{1}{v_2}\right) \tag{14.72}$$

where v_1 and v_2 are the phase velocities in the transmission line for the two different MEMS capacitance values.

Hence, by increasing the loading capacitance and thereby decreasing the phase velocity, a phase shift can be achieved. Both analog and digital phase shifters based on this concept have been reported.

Example DMTL Phase Shifter Design. A transmission line (with characteristic impedance $Z_0 = 100\ \Omega$ and effective dielectric constant $\epsilon_{\text{eff}} = 2.5$) is periodically loaded by electrostatically actuated MEMS capacitors located along the line length (as shown in Fig. 14.29) with a spacing $s = 155\ \mu\text{m}$. Ignoring the fringing capacitance, each MEMS capacitor can be approximated as a parallel-plate capacitor with an up-position capacitance $C_M = 20$ fF. The maximum capacitance ratio of each MEMS capacitor is limited to 1.5 because of the pull-down phenomenon described in Section 14.3.2. Let us calculate the cutoff frequency of the loaded line, impedance of the loaded line (at the two extreme loading cases), and maximum obtainable phase shift at 20 GHz for the loaded line of length 10 mm.

The capacitance and the inductance of the unloaded transmission line (from Chapter 2, Problem 2.10) are calculated to be

$$C_t = \frac{\sqrt{\epsilon_{\text{eff}}}}{cZ_0} = \frac{\sqrt{2.5}}{3 \times 10^8 \times 100} \cong 52.71\ \text{pF/m} \tag{14.73}$$

$$L_t = C_t Z_0^2 = 52.7 \times 10^{-12} \times 100^2 = 527.1\ \text{nH/m} \tag{14.74}$$

The cutoff frequency is obtained using (14.70) as

$$f_C = \frac{1}{3.14\sqrt{155 \times 10^{-6} \times 527.1 \times 10^{-9}(155 \times 10^{-6} \times 52.71 \times 10^{-12} + 20 \times 10^{-15})}}$$
$$\cong 210 \text{ GHz} \tag{14.75}$$

The frequency factor appearing in the expressions of the loaded-line parameters is calculated to be $\omega^2/\omega_C^2 = 0.01$. Therefore, the impedance and phase velocity of the loaded line can be approximated as

$$Z_l \cong \sqrt{\frac{sL_t}{sC_t + C_M}} \qquad v_l \cong \frac{s}{\sqrt{sL_t(sC_t + C_M)}} \tag{14.76}$$

In the up position of the MEMS capacitors, the impedance and the phase velocity of the loaded line are calculated using the above expressions and are given as

$$Z_{l1} = \sqrt{\frac{155 \times 10^{-6} \times 527.1 \times 10^{-9}}{155 \times 10^{-6} \times 52.71 \times 10^{-12} + 20 \times 10^{-15}}} = 53.9\ \Omega \tag{14.77}$$

$$v_{l1} = \frac{155 \times 10^{-6}}{\sqrt{155 \times 10^{-6} \times 527.1 \times 10^{-9}(155 \times 10^{-6} \times 52.71 \times 10^{-12} + 20 \times 10^{-15})}}$$
$$\cong 1.02 \times 10^8 \text{ m/s} \tag{14.78}$$

The maximum capacitance ratio of MEMS capacitors is 1.5. When the value of MEMS capacitance is $1.5C_M$ (=30 fF), the impedance and the phase velocity of the loaded line are calculated to be $Z_{l2} = 46.3\ \Omega$ and $v_{l2} = 8.78 \times 10^7$ m/s, respectively. The maximum obtainable phase shift per unit length (at 20 GHz) is obtained using (14.72) as

$$\Delta\phi = \omega\left(\frac{1}{v_{l1}} - \frac{1}{v_{l2}}\right) = 2 \times 3.14 \times 20 \times 10^9\left(\frac{1}{1.02 \times 10^8} - \frac{1}{8.78 \times 10^7}\right)$$
$$= 201.75 \text{ rad/m} \tag{14.79}$$

Therefore, the maximum obtainable phase shift for the loaded line of length 10 mm is 2.02 rad, or 115.6°.

Analog DMTL Phase Shifter. The first DMTL phase shifter (shown in Fig. 14.30) was demonstrated [39] by periodic loading of CPW lines using MEMS capacitors. A single analog control voltage is applied to reduce the gap height of the MEMS capacitor, thereby increasing the capacitive loading and decreasing the phase velocity. To avoid membrane snap-down, the maximum applied voltage is maintained to be less than the pull-down voltage. In this

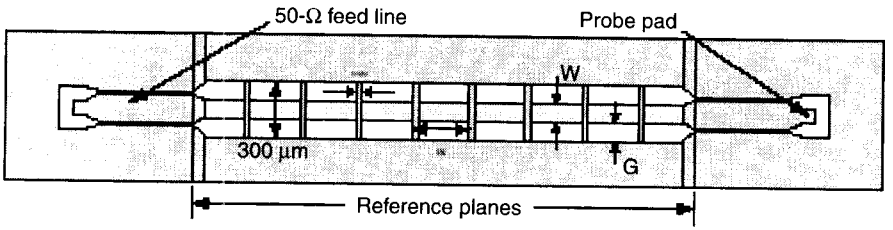


Figure 14.30 Schematic layout of distributed MEMS transmission-line phase shifter [39].

work [39], a typical DMTL consists of 32 MEMS capacitors that are 30 μm wide and are spaced at 306-μm spacings. A pull-down voltage of 23 V was measured for these MEMS capacitors; thus, the maximum voltage that could be applied without the MEMS capacitors becoming unstable was 22 V. By changing the applied voltage from 0 to 22 V, an impedance change from 59 to 56 Ω was achieved. The increased loading corresponds to an increase in bridge capacitance from 27 fF at 0 V to 31 fF at 22 V. The cutoff frequency ω_C at 22 V (31 fF) is calculated to be 119 GHz. The phase shift is linear with frequency up to 40 GHz, with 1.6 dB loss and −67° phase shift. This phase shifter exhibited a maximum phase shift of 118° at 60 GHz with 2.1 dB insertion loss. Phase shifts of −6.6°/mm at 40 GHz and −11.7°/mm at 60 GHz were reported. Phase shifts of 12°, 36°, and 62° at 40 GHz were obtained for 8, 16, and 32 MEMS devices (width 30 μm and spacing 306 μm) with insertion losses of 0.51, 0.75, and 1.56 dB, respectively.

Digital DMTL Phase Shifter. A 3-bit digital DMTL phase shifter (shown in Fig. 14.31) has been demonstrated using three 1-bit phase shifters designed for 180°, 90°, and 45° phase shifts respectively [40]. Each 1-bit phase shifter consists of a CPW transmission line loaded periodically with shunt MEMS capacitive switches. In this design, capacitive MEMS switches are used as a 1-bit capacitor with two capacitance values corresponding to the switch up and down positions. One-bit capacitors are used to periodically load CPW line to realize 1-bit phase shifters. The desired phase shift is obtained by selecting an appropriate length of the loaded transmission line. The central conductor width and ground-to-ground spacing of the CPW transmission line are 100 μm and 190 μm, respectively. The CPW line is loaded by switches of size 200 × 30 μm^2

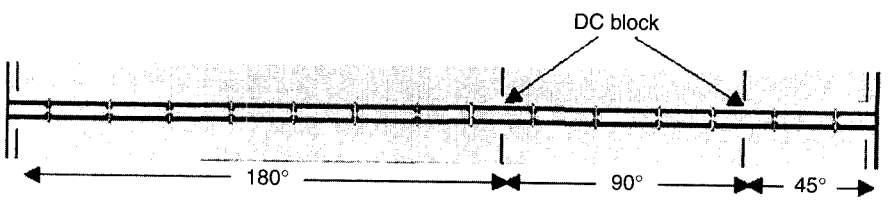


Figure 14.31 Schematic layout of 3-bit distributed MEMS transmission-line phase shifter [40].

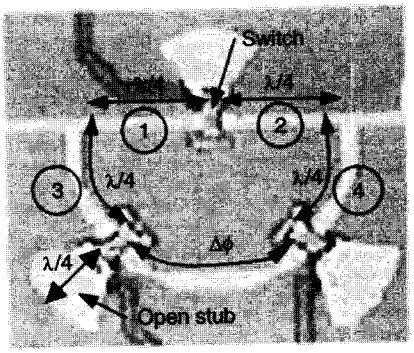
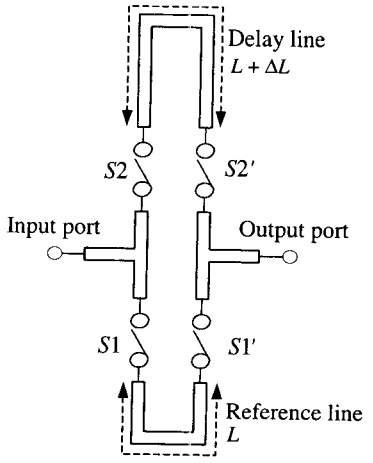


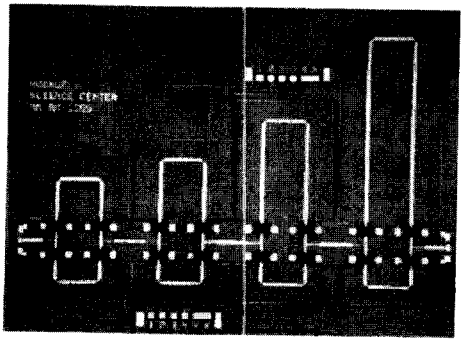
Figure 14.32 Single-bit section of switched-line phase shifter [41].

spaced 780 μm. The total length of the circuit is 11 mm. This structure provides a high-impedance transmission line ($Z_{up} = 67\ \Omega$) and a low-impedance transmission line ($Z_{down} = 37\ \Omega$) in the capacitor up and down positions, respectively. The phase shifter demonstrates an average 1.7 dB insertion loss at 26 GHz with a return loss of better than 7 dB. A phase shift from 0° to 315° in 45° phase steps was obtained with a measured phase error less than 8.5° for all of the switching states.

Switched-Line Phase Shifter. One of the switched-line-type digital phase shifters that has been demonstrated uses RF MEMS capacitive switches on microstrip lines fabricated on a high-resistivity silicon substrate [41]. Each bit (shown in Fig. 14.32) consists of four quarter-wave sections (1–4), three quarter-wave resonant open stubs, and three switches. Each bit contains two paths: a reference path and a delay path. The reference path is designed to be half wavelength in length with a switch connected from the center of this line to a quarter-wave resonant open stub. The delay path is designed to be a specific length longer than the reference path to provide the desired phase shift ($\Delta\phi$). In the delay path, resonant open stubs are connected a quarter-wavelength away from the T junction (formed by the delay path connected to the reference path) through switches. When the switches in the delay path are actuated, the quarter-wave resonant open stubs present a short at the locations of the two switches. Then the quarter-wave sections (3 and 4) transform the short at the location of the switch to an open at the T junction. Hence, the signal sees an open in the delay path and passes through the reference path. On the other hand, when the switch in the reference path is actuated, the quarter-wave resonant open stub presents a short at the location of the switch and the quarter-wave sections (1 and 2) transform the short to an open at the T junction. In this case, the reference path is effectively open and the signal passes through the delay path. Four-bit sections designed to provide phase shifts $\Delta\phi$ of 22.5°, 45°, 90°, and 180° have been cascaded to realize a 4-bit phase shifter to shift phase from 0° to 337.5° in 22.5° steps. This phase shifter exhibited an average insertion loss of 2.25 dB and a return loss of better than 15 dB for all 16 states. In



(a)



(b)

Figure 14.33 Four-bit switched-line-type phase shifter: (*a*) circuit layout of single-bit phase shifter; (*b*) microphotograph of fabricated 4-bit phase shifter [42].

this circuit, the typical capacitance values for the RF MEMS switches in the unactuated and actuated states are reported to be 33 fF and 3 pF, respectively.

Another switched-line phase shifter [42] is a 4-bit microtstrip line phase shifter using resistive contact switches. A single-bit section (shown in Fig. 14.33*a*) consists of an input T junction and an output T junction. One branch of the input and output T junctions is connected by a reference line of length L through two switches S_1 and S_1'. The other branches of the T junctions are connected by a delay line of length $L + \Delta L$ through two switches S_2 and S_2'. When only switches S_1 and S_1' are turned on, the signal passes through the reference line to the output port with a phase shift corresponding to the electrical length from input to output. On the other hand, when only switches S_2 and S_2' are turned on, the signal appears at the output with an additional phase shift (relative to the reference line) corresponding to the additional path length (ΔL) of the delay line. A 4-bit phase shifter (shown in Fig. 14.33*b*) has been realized

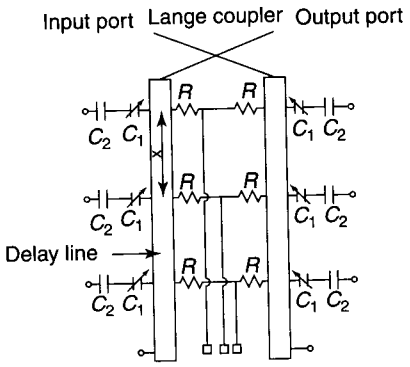


Figure 14.34 Schematic of 4-bit MEMS reflection-type phase shifter: C_1 = MEMS switch; C_2 = 10-pF capacitor.

at 10.8 GHz by cascading four single-bit sections to obtain 360° phase shift in steps of 22.5°. The overall size of the device shown in Fig. 14.33*b* is 6 × 5 mm^2. An inductive (narrow-width) microstrip section (not shown in the figure) is added in series with the input to the T junction to reduce the mismatch and the phase dispersion caused by the capacitances of the short-length transmission-line sections between the open switches and the T junctions. Insertion losses of 2.6 dB at 10 GHz and 4.3 dB at 30 GHz were measured for the phase shifter. The return loss was better than 15 dB (over most of the frequency range) for all 16 states. Nearly flat time delay response over 40 GHz bandwidth has been reported for this phase shifter.

Reflection-Type Phase Shifter. A reflection-type digital phase shifter using microstrip lines constructed on high-resistivity silicon has been reported [43]. The phase shifter (shown in Fig. 14.34) is designed using RF MEMS capacitive switches and a Lange coupler. The RF signal input to the Lange coupler is divided between two coupled ports. The coupled ports are terminated on delay lines that contain RF MEMS switches at appropriate locations. When actuated, the switches provide an RF short at the switch location on the delay line. The signal propagates through the delay lines and reflects from one of the actuated shunt switches or the short at the end of the delay line (if all the switches are unactuated). As long as the coupled ports are balanced (terminated by identical loads), the reflected signals add in phase at the output port and no signal is reflected at the input port. Using this design concept, a 4-bit reflection phase shifter has been constructed using two 2-bit phase shifters. The first 2-bit phase shifter is designed to provide phase shifts of 0°, 90°, 180°, and 270°. A reference phase of 0° is obtained by shorting the delay line using the first pair of switches. The subsequent switches are separated by line lengths equivalent to 45° phase shift (at 10 GHz design frequency). When the second pair of switches is actuated, the signal propagates 45° along the delay line until it gets reflected from the shorted switch and provides a total of 90° phase shift. A phase shift of 180° is obtained by actuating the third pair of switches. A phase shift of 270° is obtained when all the switches are unactuated and the RF

signal reflects from the shorted end of the delay line. Similarly, the second 2-bit phase shifter is designed to provide phase shifts of 0°, 22.5°, 45°, and 67.5°. These two 2-bit phase shifters are combined to produce a 4-bit phase shifter that shifts from 0° to 337.5° in 22.5° steps. The average insertion loss is 1.4 dB and return loss is better than 11 dB for all 16 states.

14.5.2 Filters

RF MEMS variable capacitors have been used to fabricate RF MEMS tunable filters. A few examples are described here.

Continuously Tunable Bandpass Filters. A schematic of a continuously tunable bandpass filter reported in the literature [44] is shown in Fig. 14.35*a*.

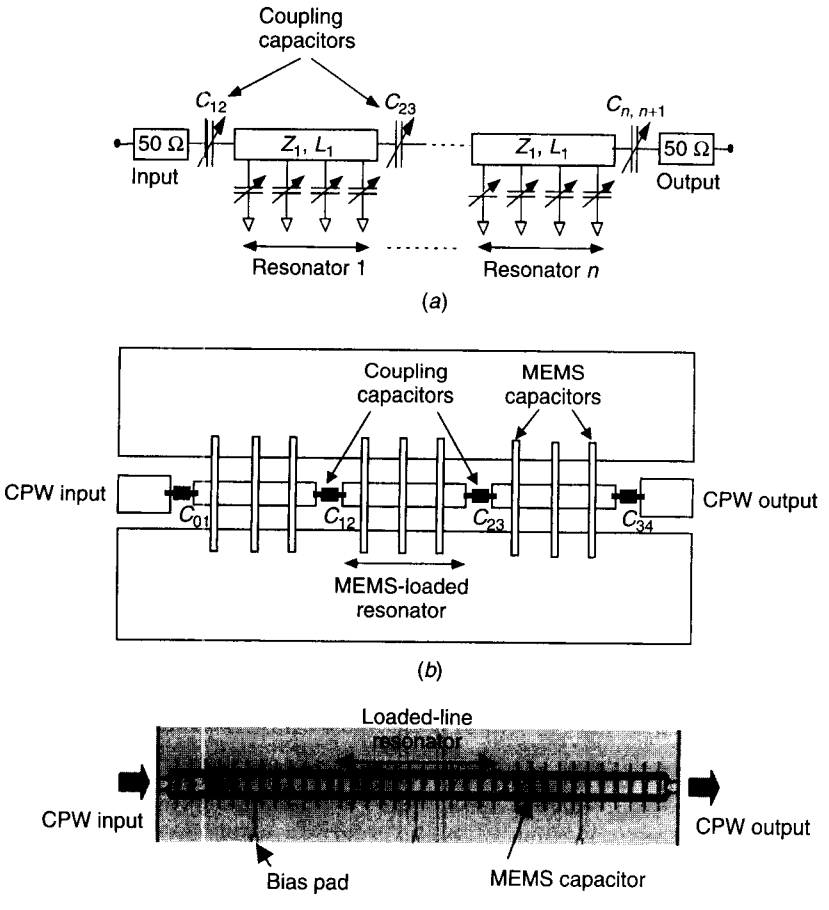


Figure 14.35 RF MEMS tunable bandpass filter based on capacitively coupled DMTL resonators: (*a*) schematic; (*b*) layout of three-sections in CPW; (*c*) photograph of fabricated filter [44].

This filter is a modification of the capacitive gap-coupled transmission-line filter. Each filter section consists of a DMTL line that is approximately a half wavelength at the passband center frequency. The adjacent resonant filter sections are connected through coupling capacitors whose values are designed to provide the desired bandwidth. A three-section bandpass filter is designed in a CPW line (shown in Fig. 14.35*b*) for a center frequency of 20 GHz. This design is similar to direct-coupled resonators filter discussed in Chapter 6 (Fig. 6.28). The three-pole filter was fabricated on a 700-μm-thick glass substrate ($\epsilon_r = 5.7$) using surface micromachining. A photograph of the fabricated *K*-band filter is shown in Fig. 14.35*c*, and the overall filter dimensions are 8.3 mm × 1.1 mm. The MEMS loading capacitance is 12–13 fF when no bias voltage is applied. A resistive biasing network is connected to the central strip of each CPW resonator for applying DC bias voltages to MEMS capacitors. A 3.8% continuous tuning range (with 60 V bias) at 20 GHz with 3.6 dB minimum insertion loss has been demonstrated. Out-of-band rejections of −36.3 and −30.1 dB were measured at 15 and 25 GHz, respectively. A return loss variation from −36 to −8 dB in the passband at all tuning ranges has been reported.

In Kim et al. [45], millimeter-wave tunable filters using cantilever-type MEMS variable capacitors are reported. Two types of bandpass filters–two-pole lumped-element filter and two-pole resonator filter–at the *Ka* band are described. The schematics and photographs of these two filter types are shown in Fig. 14.36. A tuning range of 4.2% was measured for the lumped-element filter with a center frequency shift of 1.1 GHz from 26.6 GHz at zero bias to 25.5 GHz at 65 V bias. For the two-pole resonator filter, a tuning range of 2.5% with a center frequency shift of 0.8 GHz from 32 GHz at zero bias to 31.2 GHz at 50 V bias was measured. Another paper [46] describes a design scheme for tunable coplanar waveguide RF MEMS components with applications to compact lumped-element reconfigurable filters.

Digitally Tunable Bandpass Filter. A tunable bandpass filter using variable capacitors is reported in the literature [47]. In this design, the variable capacitors have been realized by a series combination of fixed MIM (metal–insulator–metal) capacitors and RF MEMS capacitive switches (as shown in Fig. 14.37*a*). This combination yields a two-state capacitor with the minimum value limited by the off capacitance of the MEMS and the maximum value limited by the MIM capacitance. The filter uses 4-bit tunable capacitors, each consisting of five fixed capacitors, four of which are in series with RF MEMS capacitive switches, as shown in Fig. 14.37*a*. In this design, using 4-bit MEMS capacitors, admittance inverters (discussed in Section 6.3.4) are used to convert a filter with both series and shunt resonators to a filter that has only shunt *LC* resonators. The inverter-based design creates an independent design variable and allows all inductors to remain fixed at a constant value as the center frequency of the filter is varied. This enables the design of tunable filters by varying only RF MEMS capacitors. A schematic of a capacitively coupled *LC* resonator based five-pole bandpass filter is shown in Fig. 14.37*b*. The component

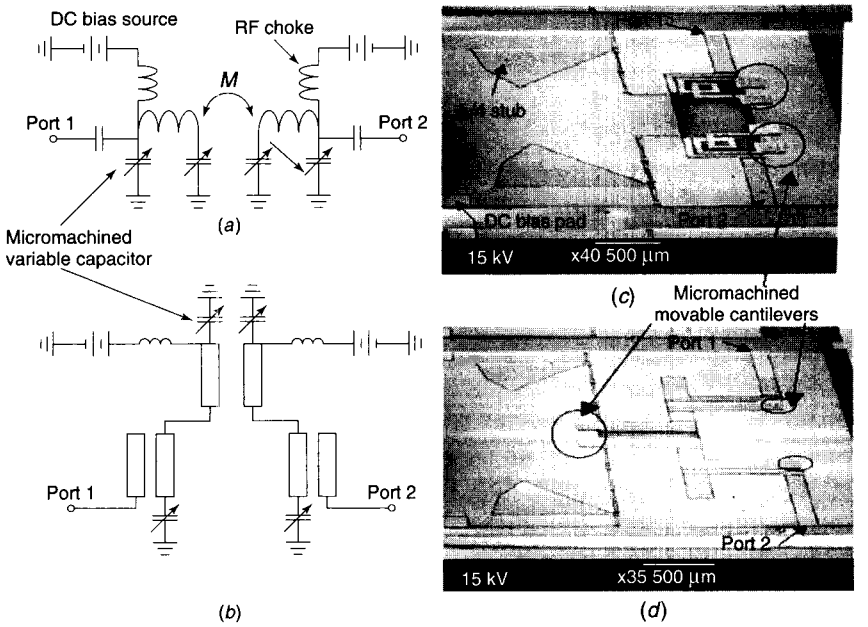


Figure 14.36 Micromachined variable capacitor millimeter-wave tunable filters: (*a*) schematic of two-pole lumped-element filter; (*b*) schematic of two-pole resonator filter; (*c*) photograph of two-pole lumped-element filter; (*d*) photograph of two-pole resonator filter [45].

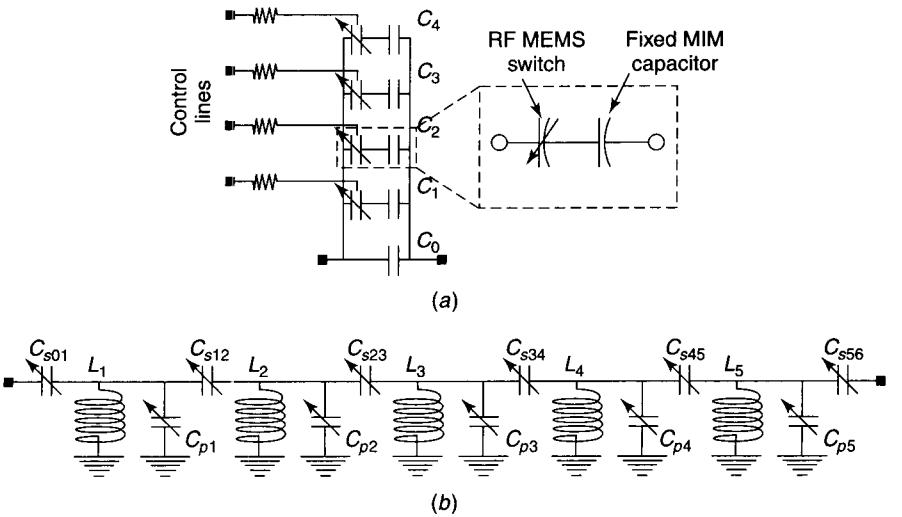


Figure 14.37 RF MEMS tunable bandpass filter: (*a*) 4-bit variable capacitor; (*b*) schematic of capacitively coupled five-pole bandpass filter using 4-bit RF MEMS capacitor shown in (*a*) [47].

values are calculated from the given filter requirements, such as tuning frequency range, bandwidth, insertion loss, and tunable states. A five-pole Chebyshev bandpass filter tunable from 880 to 992 MHz in 16 discrete center-frequency values and having a bandwidth between 168 and 174 MHz has been reported. The filter exhibited a center-frequency insertion loss between 6.6 and 7.3 dB MHz across all 16 tuning states. Also, a six-pole Chebyshev bandpass filter tunable from 110 to 160 MHz with a variable bandwidth from 37 to 58 MHz and an insertion loss between 3.7 and 4.2 dB has been demonstrated.

14.5.3 Impedance Tuners

A variable impedance tuner using MEMS capacitors has been demonstrated [48]. The schematic of the impedance tuner is shown in Fig. 14.38*a*. The tuner consists of two tunable MEMS capacitors (C_1 and C_3) and two frequency-tunable parallel resonant sections. In parallel resonant sections, effective inductances are created using capacitances C_5 and C_6 sandwiched between J inverters (admittance inverters discussed in Section 6.3.4), as shown in Fig. 14.38*b*. Frequency tuning of the resonant sections is achieved by adjusting the capacitances C_2 and C_4 together with the capacitances C_5 and C_6. The series capacitor C_1 is used to change the angle of the reflection coefficient and the parallel resonance sections are used to change the magnitude of the reflection coefficients. The bias circuit used for each of these tunable capacitors is designed using a quarter-wave line and an MIM capacitor. This impedance

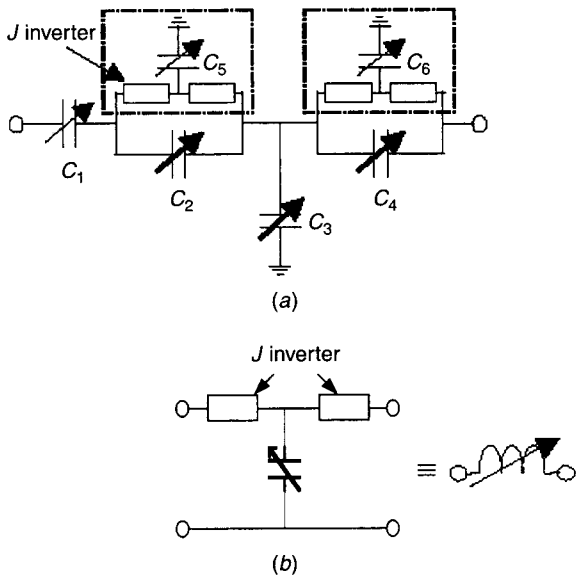


Figure 14.38 RF MEMS variable impedance tuner: (*a*) circuit schematic of impedance tuner (C_1–C_5 are MEMS capacitors); (*b*) J inverter.

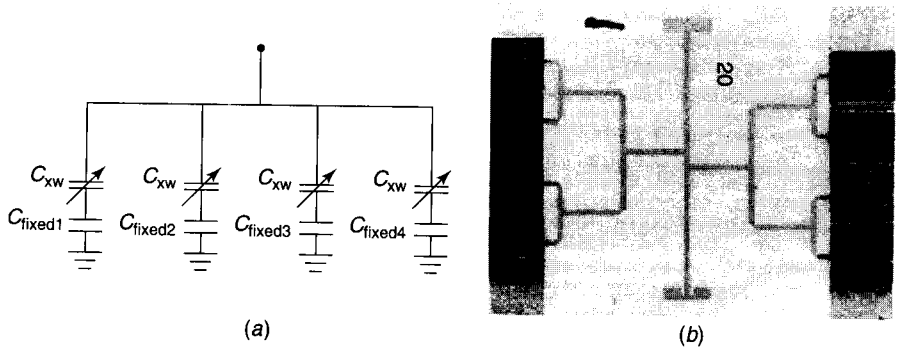


Figure 14.39 RF MEMS double-stub impedance tuner: (*a*) schematic of 4-bit stub; (*b*) photograph of 4 × 4-bit (256-state) tuner [49].

tuner circuit was fabricated on a quartz substrate using surface micromachining technology. The overall size of the tuner is 3.7×2.0 mm^2. The tuner operating at 25 GHz is capable of generating the impedances in the second and third quadrants of the Smith chart, which correspond to the optimum load range of the power transistors. The MEMS capacitors needed a maximum DC bias voltage of 40 V.

A planar reconfigurable double-stub tuner using MEMS switches (shown in Fig. 14.39) has been developed [49]. The double-stub impedance matching design (discussed in Chapter 4) is used to determine the susceptances of each stub and the length between them to match the desired range of loads. The desired match impedance was taken to be 50 Ω and the desired range of load impedances to be matched was chosen to be 20 to 80 Ω for the real part and −150 to +150 Ω for the imaginary part. The distance of the first stub to the load was taken to be zero for simplicity, and the distance between the two stubs was selected to be 0.1λ. Then, the susceptance of each stub required to match the range of loads to 50 Ω was determined using standard procedures. The circuit shown in Fig. 14.39*a* was designed to provide a large range of capacitances. It consists of four parallel sets of a series combination of RF MEMS capacitive switches (described in Section 14.4) and a fixed "stub" capacitor. The series combination of the switch capacitance (C_s) and the fixed capacitance (C_f) represents 1 bit of equivalent capacitance [$C_{eq} = C_s C_f/(C_s + C_f)$]. When the switch is off, $C_s \ll C_f$ and $C_{eq} \approx C_s$, and when the switch is on, $C_{eq} \approx C_f$. The four capacitance values corresponding to 4 bits are calculated to cover the range of capacitances required to match the desired range of load impedances. The switch capacitances (C_s) in the off and on states are 35 fF and 3 pF, respectively. The required fixed capacitance value for various bits was calculated using $C_f = C_{eq}C_s/(C_s - C_{eq})$ and these values ranged from 45 to 1155 fF. The fixed capacitors were realized using open-circuited stubs. Four 4-bit sections were used to provide a total of 256 tuning states of the double-stub tuner. The measurement results for the double-stub tuner have shown a matching

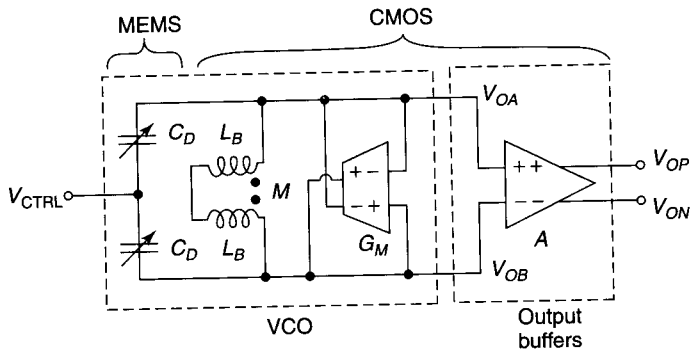


Figure 14.40 Block diagram of MEMS-based voltage-controlled oscillator [50].

load impedance range of 5–108 Ω for the real part and −60 to +48 Ω for the imaginary part at 20 GHz.

14.5.4 Oscillators

RF MEMS oscillators using MEMS tunable capacitors for frequency tuning (shown in Fig. 14.40) have been reported [50]. A MEMS capacitor (described in Section 14.4.2) with a tunable capacitance from 1.4 pF at 0 V to 1.9 pF at 5 V and Q values of 12 and 14 at 1 and 2 GHz, respectively, has been used. The oscillator is fabricated using CMOS technology and assembled in a ceramic package, where the CMOS and MEMS chips are bonded. The resonant circuit is designed using two MEMS tunable capacitors (C_D) and bonding wire inductors (L_B). A cross-coupled transconductance amplifier (G_M) is used to provide a negative conductance, shunting the resonant circuit. An output buffer (A) is used to isolate the resonant circuit from the load. An oscillation frequency of 2.4 GHz was obtained using a capacitance of 1.4 pF (MEMS capacitor at 0 V) and 2.2-nH bonding wire inductors with a mutual inductance (M) of 0.6 nH. A tuning range of 3.4% has been demonstrated by tuning the oscillation frequency from 2.40 GHz at 0 V to 2.32 GHz at 6 V bias voltage at the MEMS capacitor. This oscillator has phase noise values of −93 and −122 dBc/Hz at 0.1 and 1 MHz offset, respectively, from the 2.4 GHz carrier center frequency. Another low-phase-noise voltage-controlled oscillator operating at 855–863 MHz has been reported [51].

14.5.5 Amplifiers

An up-converter parametric effect amplifier using a MEMS capacitor (shown in Fig. 14.41) has been reported [52]. The time-varying capacitor needed for parametric amplification is realized by pumping the MEMS capacitor by a large-signal voltage at a pump frequency $f_P = 1.64$ MHz. The MEMS capacitor is also fed by a low-frequency small-signal source at frequency $f_s = 200$ KHz.

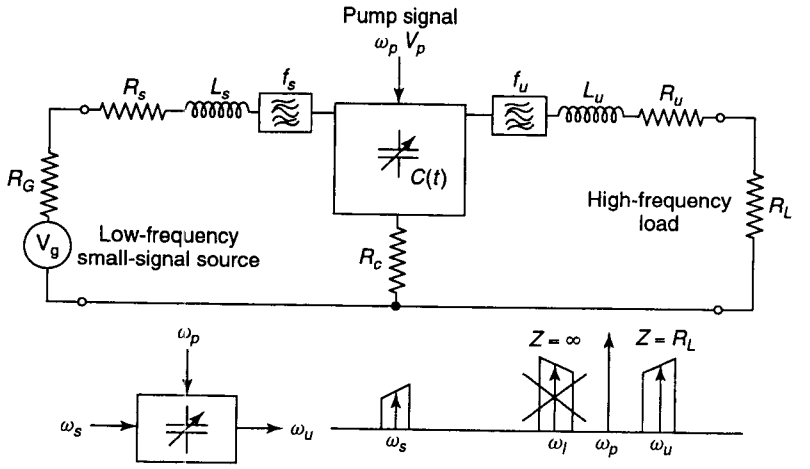


Figure 14.41 Schematic of up-converter parametric amplifier (pump signal ω_p, input frequency ω_s, and up-converter frequency $\omega_u = \omega_p + \omega_s$).

The nonlinear reactance of the capacitor produces harmonic frequencies of $mf_P \pm nf_s$ $(m, n = 0, 1, 2, \ldots)$. The time-varying capacitor is terminated in an ideal filter such that signals at only the three frequencies f_P, $f_u = f_P + f_s$, and $f_l = f_P - f_s$ are allowed to exist. An up-conversion frequency ratio of 9:1 by amplification of 200 kHz input to 1.84 MHz output with a gain of 6 dB has been demonstrated.

This approach has a high-frequency limitation, as the pump frequency should be less than the mechanical resonance frequency of the MEMS structure. However, it is reported [52] that in the future it should be possible to build amplifiers up to the 1–5-MHz range using smaller area membranes.

MEMS capacitors and inductors are likely to find applications also as RF components in usual transistor amplifiers at microwave and millimeter-wave frequencies.

14.5.6 MEMS Mechanical Resonators and Filters

A mechanical filter is composed of mechanically coupled multiple lumped mechanical resonators. Mechanical filters transform electrical signals into mechanical signals, perform a filtering function, and then transform the processed mechanical signals back into an output electrical signal. The characteristics of mechanical filters include low loss, narrow bandwidth, and high-temperature stability. Mechanical filters were first developed in 1946 [53] for use in telephone systems. Recently, MEMS technology has been applied to miniaturize mechanical resonators and filters [54].

Theory of Operation. A single mechanical resonator can be modeled as a spring–mass system (shown in Fig. 14.42*a*) with a resonance frequency given

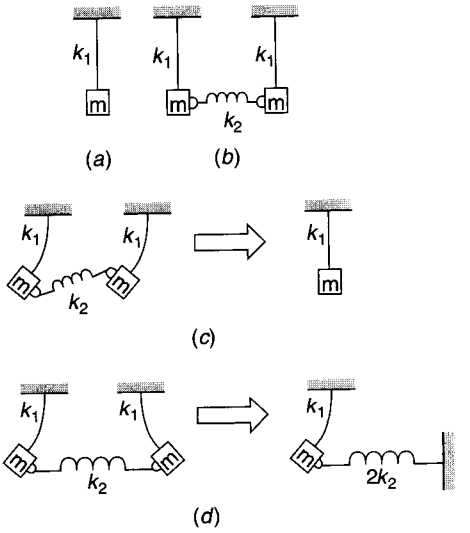


Figure 14.42 Spring–mass system model of mechanical resonators: (*a*) single resonator; (*b*) two identical spring–mass resonators connected by coupling spring; (*c*) two identical spring–mass resonators in phase; (*d*) two identical spring–mass resonators out of phase.

by

$$f_1 = \frac{1}{2\pi}\sqrt{\frac{k_1}{m}} \tag{14.80}$$

where k_1 is the stiffness of the spring without mass and m is the resonating mass without compliance. The two identical spring–mass systems (shown in Fig. 14.42*b*) connected side by side through a third spring hinged between two resonating masses exhibit two resonance frequencies. The first resonance occurs when the two resonators are in phase, as shown in Fig. 14.42*c*, and there is no compression or tension exerted (i.e., no energy is stored) on the coupling spring k_2. In this case, coupling spring does not interfere with the overall system and the resonance frequency is the same as that of a single-resonator system (f_1). The second resonance occurs when the two resonators are 180° out of phase (shown in Fig. 14.42*d*) and the coupling spring k_2 undergoes compression and tension cycles. Considering the symmetry of the system with respect to a reference plane at the center of the coupling spring k_2, where there is no displacement throughout the cycle, the out-of-phase case can be modeled as a single resonator with two springs of constants k_1 and $2k_2$, as shown in Fig. 14.42*d*. The second resonance frequency is given by

$$f_2 = \frac{1}{2\pi}\sqrt{\frac{k_1 + 2k_2}{m}} \tag{14.81}$$

This characteristic can be used to design a bandpass filter by coupling two or more mechanical resonators. When all resonators are in phase, we get the

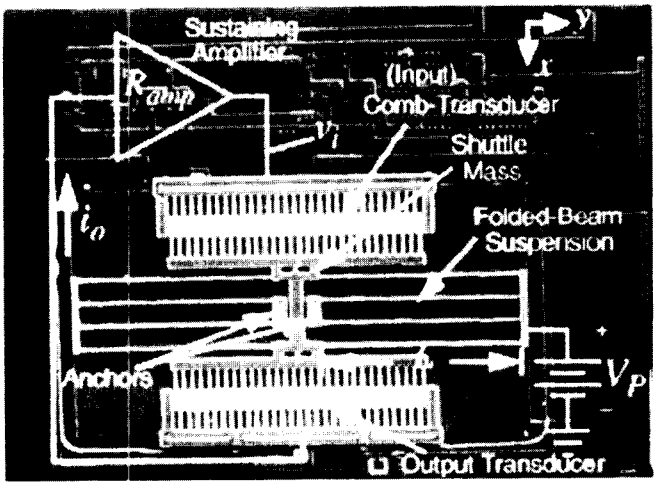


Figure 14.43 Double-folded beam micromechanical resonator along with input and output transducers [55].

lowest resonance frequency, and when all adjacent resonators are 180° out of phase, the highest resonance frequency is obtained.

The structures investigated for MEMS-based mechanical resonators include double-folded beam, clamped–clamped beam, and free–free beam [55]. These structural configurations are discussed briefly.

Double-Folded Beam Resonator and Filter. A double-folded beam micromechanical resonator using interdigital transducers, shown in Fig. 14.43, has been demonstrated [55]. The device consists of two interdigital transducers (at input and output) connected back to back by a double-folded beam resonator. The input voltage, consisting of a DC bias V_p and an AC excitation v_i, is applied at the input transducer and the AC current generated is measured at the output transducer. The frequency response is obtained from the transconductance spectrum of the resonator. A resonance frequency of 16.5 kHz (under 20-mTorr vacuum and at 20 V DC bias) and an extracted Q value of 50,000 have been reported.

An example of a filter using three double-folded beam resonators (coupled mechanically) has also been reported [55]. The entire filter structure is fabricated using a doped polysilicon (2-μm-thick) layer suspended 2 μm above a ground plane. The folded beam length and width are 32.8 and 2 μm, respectively. The folded beam resonators exhibited a Q-factor value of 25,000 at 40 mTorr pressure. The filter performance results are: passband 510 Hz, center frequency 300 kHz, $Q = 590$, insertion loss 3 dB, stopband rejection 38 dB, frequency tuning 2% at 50 V, and DC bias 150 V.

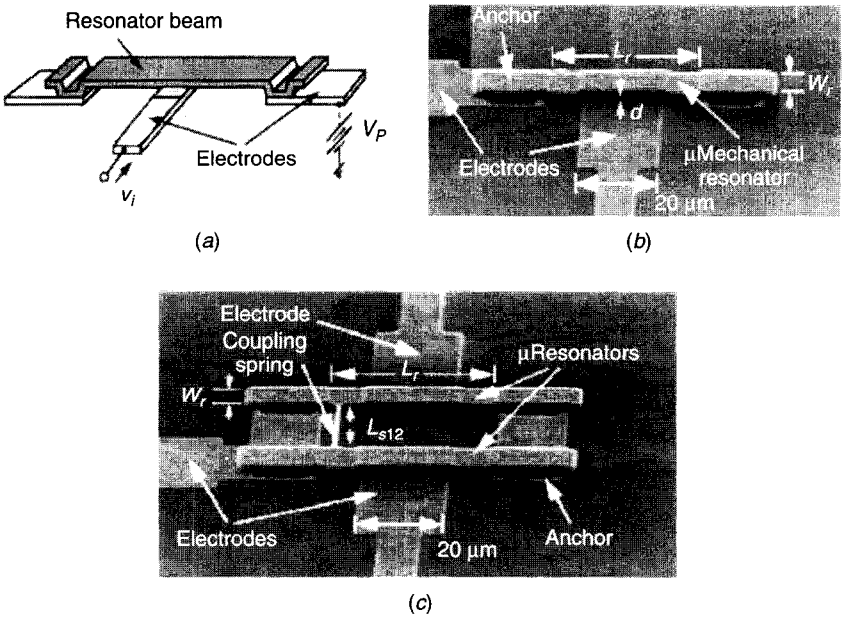


Figure 14.44 Fixed–fixed beam micromechanical resonator: (*a*) schematic; (*b*) photograph; (*c*) photograph of filter [55].

Fixed–Fixed Beam Resonator and Filter. A micromechanical resonator based on a fixed–fixed beam for high-frequency applications has been reported [55]. A parallel-plate capacitive transducer is formed between the resonator beam and an underlying electrode, as shown in Fig. 14.44*a*. An input voltage, comprised of a DC bias voltage V_p and an AC signal v_i, is applied across the transducer and the conductance spectrum is measured. A photograph of a fabricated micromechanical resonator is shown in Fig. 14.44*b*, and the dimensions of the resonator beam are length $L_r = 40.8$ μm, width $W_r = 8$ μm, thickness $h = 2$ μm, and gap height $d = 0.1$ μm. A resonance frequency of 8.5 MHz (at a pressure of 70 mTorr) with a DC bias $V_p = 10$ V and an AC input voltage $v_i = 3$ mV was measured.

A mechanical filter using the fixed–fixed beam resonator, shown in Fig. 14.44*c*, has been demonstrated. The dimensions of the polysilicon resonator beams are 8 μm wide, 22 μm long, and 2 μm thick. The length of the mechanical coupling beam between the two resonators is 11.8 μm. The gap between the resonating beams and the bottom electrodes is 25 nm. Measured performance results of this mechanical filter at a DC bias of 15 V are central frequency 34.5 MHz, insertion loss <6 dB, rejection 25 dB, and bandwidth 460 KHz.

Free–Free Beam Resonator. A mechanical resonator using an effective free–free beam, shown in Fig. 14.45, has been developed [55] for very high frequency applications. The device consists of a free–free beam resonator sus-

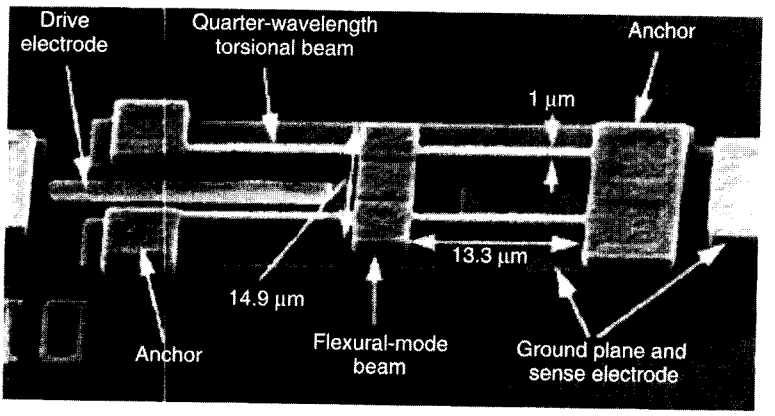


Figure 14.45 SEM of free–free beam micromechanical resonator (design dimensions shown for 71 MHz resonance frequency) [55].

pended by four torsional beams attached at the flexural node points of the resonator. The torsional beams are anchored to the substrate at each end and are designed to have a length equal to a quarter wavelength of the free–free beam resonance frequency, so that the impedance presented to the beam by the supports is zero. This configuration virtually levitates the free–free beam resonator and minimizes the clamping loss at the anchoring points of the supporting torsional beams. A resonant frequency as high as 92.25 MHz has been demonstrated with a Q of nearly 8000. This device required a bias voltage of about 76 V.

It has been predicted that micromechanical resonators operating in the gigahertz frequency range can be fabricated [55]. For example, the dimensions of a fixed–fixed beam resonator (shown in Fig. 14.44*a*) required to attain a frequency of 1 GHz are length 4 μm, width 2 μm, and height 2 μm. This resonance frequency can also be obtained by longer beams vibrating in higher modes.

14.5.7 Micromachined Components and Circuits

Bulk micromachining technology (described in Section 14.2.1) enables removal of the lossy substrate material surrounding microwave and millimeter-wave components. Also, micromachining allows three-dimensional integration for higher functionality. Advantages of micromachined RF components include superior performance, low volume and weight, and potentially lower cost. Typical microwave and millimeter-wave components fabricated by micromachining include inductors, waveguides, transmission lines, couplers, and power dividers.

Micromachined Inductors. In modern wireless and mobile communication applications, high-performance inductors are in increasingly higher demand.

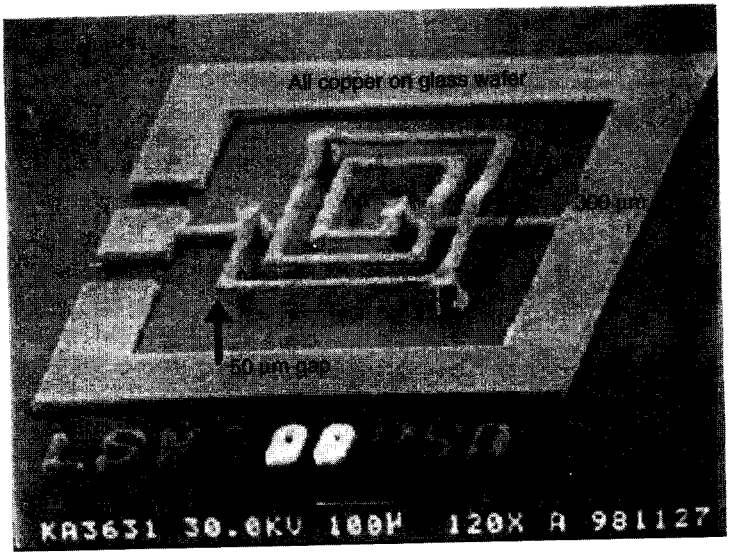


Figure 14.46 Three-dimensional spiral inductor made on glass substrate [56].

Micromachining technology enables the manufacturing of three-dimensional inductors, which have the advantages of being suspended and having very low parasitic capacitance.

In 1999, Yoon et al. [56] demonstrated a three-dimensional spiral inductor using copper metal lines on a glass substrate, as shown in Fig. 14.46. In this three-dimensional fabrication method, plated nickel is used as the molding material for copper plating. A multiple-layered structure can thus be made with the sacrificial nickel mold removed at the end of the process. Inductors made on a glass substrate have shown a Q of 57 at 10 GHz with an inductance of 1.75 nH. Young et al. [57] demonstrated a three-dimensional coil inductor, shown in Fig. 14.47, with a low-loss alumina core. Copper traces (5 μm in thickness) were plated around the alumina core, which is approximately 500 μm in width and height. A Q of 16 was measured at 1 GHz for a two-turn inductor with an inductance of 8.2 nH.

Zhou et al. [58] have described a switched inductor network by combining micromachined relays with spiral inductors, as shown in Fig. 14.48. The relays are made of a $TaSi_2$–SiO_2 bimorph cantilever beam driven by combined thermal and electrostatic actuation. The inductors are made using a spiral design on a membrane with the portion of silicon substrate directly underneath the inductor area removed. A 4-bit switched inductor network has been demonstrated with a measured total inductance ranging from 2.5 to 324.8 nH. The self-resonance frequency is 1.9 GHz. This inductor has a Q factor of 1.7 at a frequency of 530 MHz for the switching configuration that yields an inductance of 324.8 nH. This digitally tuned inductance can be monolithically integrated

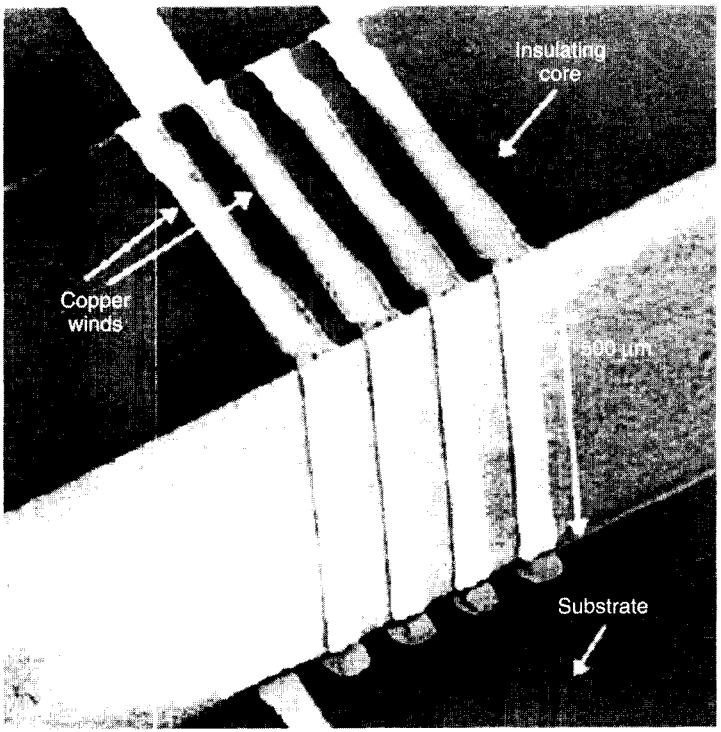


Figure 14.47 Three-dimensional Cu coil inductor with low-loss alumina core [57].

with other ICs to provide a fully integrated radio on a chip or other circuit/subsystem for RF applications.

Dahlmann et al. [14] reported microwave inductors of 1.5–2.5 nH fabricated by a solder self-assembly process (shown in Fig. 14.49). Using a planar process, copper structures are fabricated with solder pads being placed between an

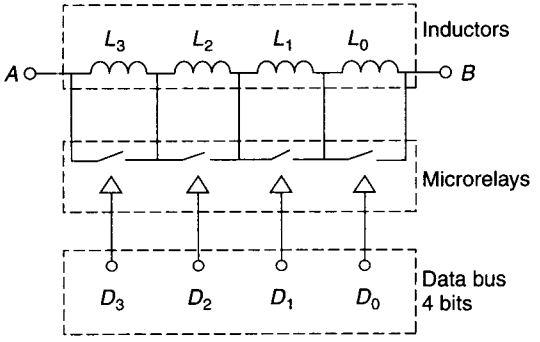


Figure 14.48 Schematic of digitally controllable variable inductor.

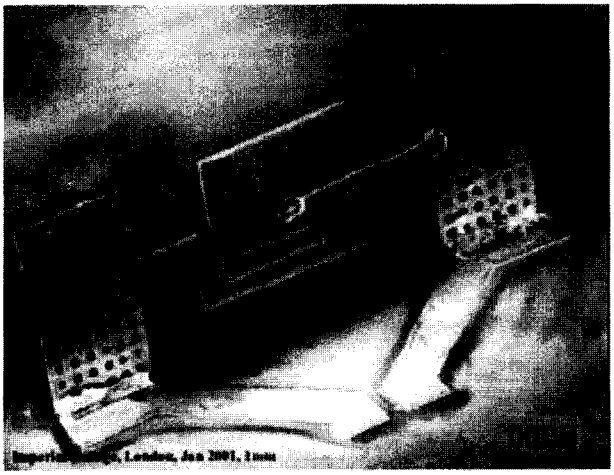


Figure 14.49 Inductor fabricated by solder self-assembly process [14].

anchored and a released portion of the devices. When the wafer is heated, the solder pads melt, and the resulting force rotates the structure out of the substrate plane. The details of solder self-assembly are discussed in Section 14.2. The wafer is then cooled, the solder pads resolidify, and the structure remains assembled. This approach allows batch processing of a large number of components at the same time. A very substantial separation between the coil and the substrate (several hundreds of micrometers) has been obtained. Moreover, further reduction of capacitance and inductive substrate coupling is achieved by rotating the coil into a plane perpendicular to the substrate. The decoupling of the inductor (by solder self-assembly) from the substrate results in an improvement in the Q value from 4 to 20 and in the frequency of the maximum Q value from 0.5 to 3 GHz on low-resistivity silicon substrates.

Micromachined Waveguides. Air-filled rectangular [59, 60] and diamond-shaped [61] micromachined waveguides for millimeter-wave applications have been reported. McGrath et al. [59] developed a micromachined rectangular waveguide operating at 75–100 GHz. Process steps for fabrication of a half section of waveguide are shown in Fig. 14.50. Two of these sections (shown in Fig. 14.50*d*) are mated to form the complete waveguide. The measured insertion loss is about 0.04 dB per wavelength at 100 GHz. A typical loss value for a commercially available waveguide in this frequency range is about 0.024 dB per wavelength. A diamond-shaped waveguide fabricated by anisotropic etching of $\langle 100 \rangle$ Si for operation in the 80–183-GHz range is shown in Fig. 14.51. Dispersion curves have been measured.

Micromachined Transmission Lines. Micromachining technology has been used for removing some portions of substrates underneath transmission-line

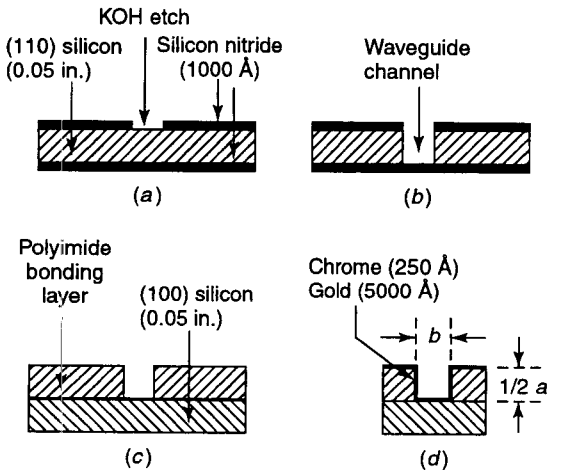


Figure 14.50 Fabrication of one half-section of waveguide (a = waveguide width, b = height): (*a*) Si_3N_4 mask defines waveguide height; (*b*) wafer is etched completely through; (*c*) wafer with waveguide channel is bonded to unetched wafer which forms waveguide side wall; (*d*) completed half section of waveguide with gold plating.

components (such as transmission lines, filters, and couplers) to reduce propagation loss, frequency dispersion, and non-TEM modes. Membrane-supported transmission lines such as microstrip lines (Fig. 14.52*a*), microshield lines (Fig. 14.52*b*), shielded membrane microstrip lines (Fig. 14.52*c*), and shielded strip lines (Fig. 14.52*d*) have been proposed [62]. A 1.4-μm three-layer structure of SiO_2, Si_3N_4, and SiO_2 of thicknesses 7000, 3000, and 4000 Å, respectively, is deposited on a high-resistivity ⟨100⟩ silicon substrate using thermal oxidation and high-temperature chemical vapor deposition. Then transmission lines are

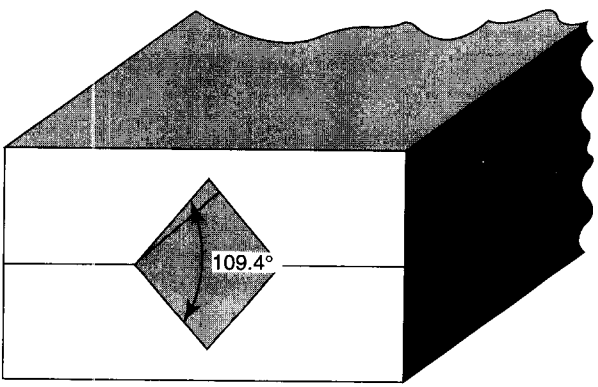


Figure 14.51 Diamond-shaped waveguide formed from two etched V grooves.

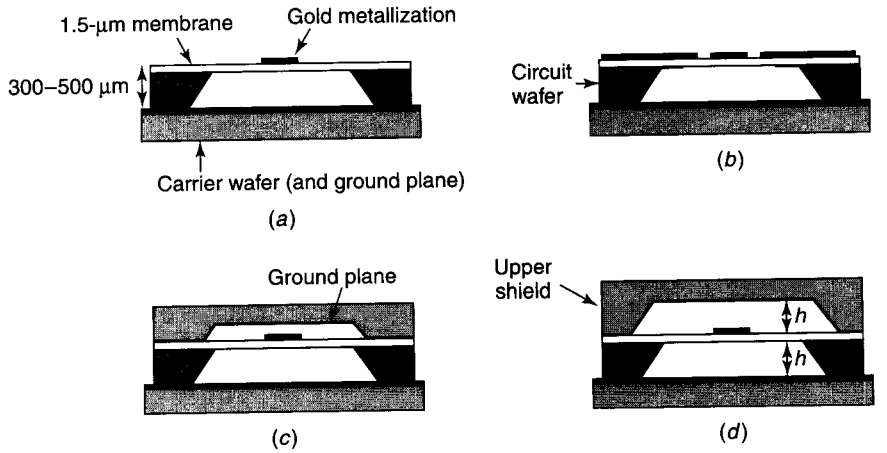


Figure 14.52 Micromachined transmission lines: (*a*) microstrip line; (*b*) microshield line; (*c*) shielded microstrip line; (*d*) suspended strip line.

defined on top of the membrane using lithography, gold evaporation, and 2–3 μm gold electroplating process. The back side of the substrate underneath the transmission lines is etched anisotropically by bulk micromachining technique until the transparent dielectric membrane appears. Top cavity shielding (for strip lines) or RF ground plane shielding (for shielded microstrip lines) is achieved by bonding an etched silicon wafer electroplated with 2–3 μm of gold. The wafer bonding technique is described in Section 14.2.2.

Micromachined Resonators. Suspended strip lines mentioned earlier have been used to fabricate an end-coupled resonator on a 350-μm-thick high-resistivity silicon wafer on a 1.4-μm dielectric membrane layer of dimensions 7.1×5.3 mm^2 [62]. The resonator is fed by a 50-Ω grounded CPW line. The strip-line resonator dimensions are 1.1 cm long (half wavelength at 13.5 GHz), 500 μm wide, and 3 μm thick. The impedance of the strip-line resonator was calculated to be 80.8 Ω. The measured values of resonance frequencies are 13.5, 27.4, and 39.6 GHz.

A micromachined cavity resonator fed by CPW feed lines (shown in Fig. 14.53) has been demonstrated [63]. A cavity is constructed from two silicon wafers (525 μm) etched 450 μm down. The dimensions of the cavity are 8.84×8.84 mm^2 designed for a TE_{011} mode resonant frequency at 24 GHz. A CPW feed line is fabricated on the edge of the cavity and a wire bond is placed from the end of the CPW feed line to the base of the cavity to trigger the dominant TE_{011} mode in the cavity. By changing the length of the wire bond and the number of bonds placed, the amount of coupling can be controlled. The measured TE_{011} resonant mode frequency was 23.97 GHz with $|S_{21}|$ of

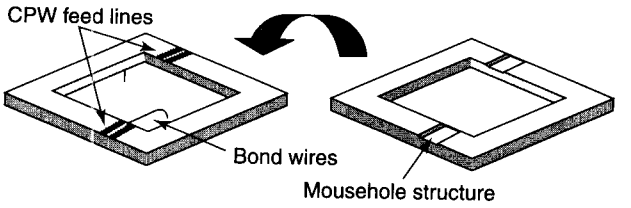


Figure 14.53 Micromachined cavity resonator fed by CPW lines.

−15.2 dB. An unloaded Q of 1117 (after subtracting the losses of feed lines) was reported.

Micromachined Filters [62]. A 20-GHz interdigitated bandpass filter using suspended strip line has been demonstrated. The filter was fabricated on a 1.4-μm-thick membrane with dimensions of 7.9×3.4 mm^2. The input line to the filter is a 2.6-mm-long CPW line on a high-resistivity silicon wafer. Port-to-port insertion loss of 1.7 dB at 20.3 GHz (with a 3-dB bandwidth of 3.1 GHz) and a return loss better than 15 dB in the passband have been measured.

A microshield-line-based low-pass filter (shown in Fig. 14.54) designed at 30 GHz using a five-section stepped-impedance implementation has been reported. The membrane dimensions are 6×1.8 mm^2 with a cavity height of 350 μm. The high- and low-impedance sections of the five-section filter are 265 and 45 Ω, respectively. The feed-line impedance is designed to be 75 Ω. A rejection of greater than 20 dB up to 75 GHz (about 1.5 octaves above the 3-dB point at 26 GHz) and an insertion loss between 0.25 and 0.9 dB in the frequency range 10–23 GHz were reported.

Micromachined Coupler. A four-port quarter-wave CPW coupled-line coupler (shown in Fig. 14.55) fabricated on 420-μm-thick quartz substrate ($\epsilon_r = 3.81$) using the LIGA process has been reported [5]. The coupling section

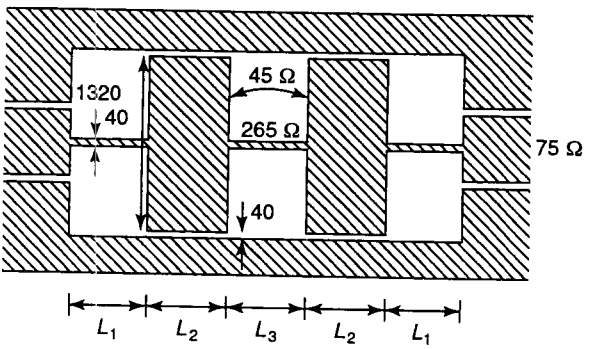


Figure 14.54 Microshield low-pass filter. Lengths of different sections are $L_1 = 745$ μm, $L_2 = 1200$ μm, and $L_3 = 1189$ μm.

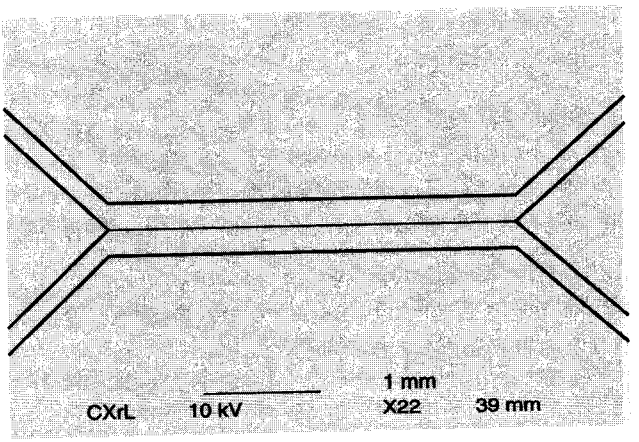


Figure 14.55 Micromachined four-port CPW coupled-line coupler [6].

was designed for an input impedance of 50 Ω with a coupling equal to −6 dB. The length of the coupling section was designed to be quarter-wave length (arithmetic average of the odd and even modes) at 15 GHz. The coupling $|S_{31}|$ was measured to be −6.1 dB. In the passband, reflection loss values between −15 and −25 dB and isolation $|S_{41}| < -20$ dB were measured.

Micromachined Power Divider. A microshield-line-based Wilkinson power divider (shown in Fig. 14.56) operating at 33 GHz has been reported [64]. It makes use of a lumped resistor of value 212 Ω placed between the two signal lines) that serves to provide an input matching at the ports. The impedance of the lines at the two outputs of the power divider is chosen to be 106 Ω, which is matched to the 50-Ω terminations through 73-Ω matching sections. Isolation of better than 15 dB and insertion loss (single divider) of 0.2 dB were measured.

As the MEMS technology matures and reliability investigations for MEMS devices are completed, we can look forward to increasingly frequent applications of RF MEMS devices in microwave circuits and subsystems.

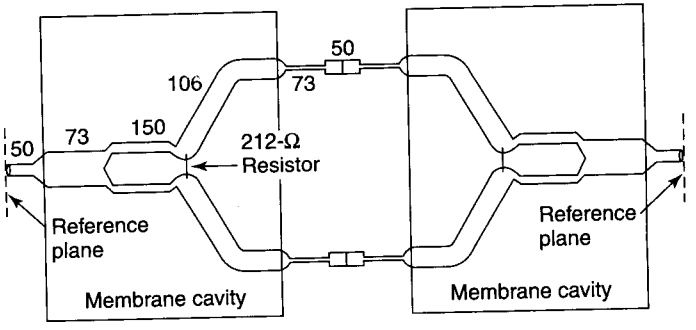


Figure 14.56 Schematic layout of micromachined Wilkinson power divider.

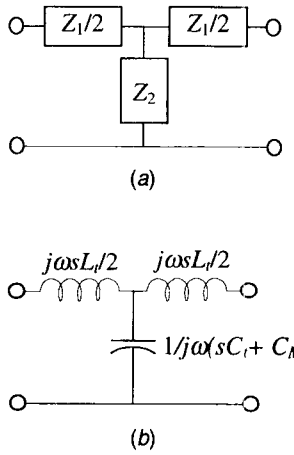


Figure 14.57 (*a*) Symmetrical T-network section; (*b*) T-network model for section of periodically loaded transmission line.

APPENDIX: PARAMETERS OF A SYMMETRICAL T-NETWORK

The two basic parameters that characterize the signal propagation in a periodic lumped-element symmetrical T-network (shown in Fig. 14.57*a*) are the characteristic impedance and the propagation delay. The characteristic impedance of a symmetrical T-network is defined as the impedance Z_{0T} for which the network presents an input impedance of Z_{0T} (at the input port) when terminated by a load impedance of Z_{0T} at the output port. The propagation delay γ_T is defined by $V_2 = V_1 e^{-\gamma_T}$, where V_1 and V_2 are the port voltages of the network for matched termination.

The characteristic impedance Z_{0T} and the propagation delay γ_T of a symmetrical T network [65] shown in Fig. 14.57*a* are given by

$$Z_{0T} = \sqrt{Z_1 Z_2 \left(1 + \frac{Z_1}{4Z_2}\right)} \tag{14.82}$$

$$\cosh(\gamma_T) = 1 + \frac{Z_1}{2Z_2} \tag{14.83}$$

A lumped-element T model for a section of a periodically loaded line (shown in Fig. 14.29) is shown in Fig. 14.57*b*. Comparing with the symmetric T-network model shown in Fig. 14.57*a*, $Z_1 = j\omega sL_t$ and $Z_2 = 1/[j\omega(sC_t + C_M)]$. The characteristic impedance Z_l of the loaded line can be obtained using (14.82) as

$$Z_l = \sqrt{\frac{sL_t}{sC_t + C_M}\left(1 - \frac{\omega^2}{4/(sL_t(sC_t + C_M))}\right)} \tag{14.84}$$

The characteristic impedance becomes zero at a frequency given by

$$\omega_C = \frac{2}{\sqrt{sL_t(sC_t + C_M)}} \tag{14.85}$$

At low frequencies, when $\omega < \omega_C$, the characteristic impedance Z_l has a real value, whereas at high frequencies, when $\omega > \omega_C$, the characteristic impedance Z_l is imaginary. Thus, the loaded line exhibits a low-pass filter behavior with a cutoff frequency given by ω_C.

The characteristic impedance Z_l (14.84) can be written in terms of ω/ω_C as

$$Z_l = \sqrt{\frac{sL_t}{sC_t + C_M}\left(1 - \frac{\omega^2}{\omega_C^2}\right)} \tag{14.86}$$

The propagation delay in a section of periodically loaded transmission line of length s is $\gamma_T = j\beta s$ (lossless case). The phase constant can be obtained using $Z_1 = j\omega sL_t$ and $Z_2 = 1/[j\omega(sC_t + C_M)]$ in (14.83) as

$$\cosh(j\beta s) = 1 - \frac{\omega^2}{2/(sL_t(sC_t + C_M))} \tag{14.87}$$

Using $\cosh(j\beta s) = \cos(\beta s)$, the above equation can be expressed as

$$\beta = \frac{1}{s}\cos^{-1}\left(1 - \frac{2\omega^2}{\omega_C^2}\right) \tag{14.88}$$

Now, using the approximation

$$\cos^{-1}(1 - x) = \sqrt{2}x^{1/2} + \frac{1}{12}\sqrt{2}x^{3/2} + \cdots \quad \text{when } x < 1, \tag{14.89}$$

the phase constant can be written as

$$\beta \cong \frac{1}{s}\frac{2\omega}{\omega_C}\left(1 + \frac{\omega^2}{6\omega_C^2}\right) \quad \text{or} \quad \frac{1}{s}\omega\sqrt{sL_t(sC_t + C_M)}\left(1 + \frac{\omega^2}{6\omega_C^2}\right) \tag{14.90}$$

which is the same as (14.69).

The phase velocity can be obtained as

$$v = \frac{\omega}{\beta} \cong \frac{s\omega_C}{2(1 + \omega^2/(6\omega_C^2))} \quad \text{or} \quad \frac{s}{\sqrt{sL_t(sC_t + C_M)}(1 + \omega^2/6\omega_C^2)} \tag{14.91}$$

which is the same as (14.71).

REFERENCES

1. Kovacs, G. T. A., *Micromachined Transducers Sourcebook*, McGraw-Hill, Boston, 1998.
2. Bryzek, J., K. E. Petersen, and W. McCulley, "Micromachines on the March," *IEEE Spectr.*, May 1994, pp. 20–31.
3. Tai, Y.-C., and C.-J. Kim, "Microelectromechanical Systems (MEMS)," in I. J. David (Ed.), *The Industrial Electronics Handbook*, CRC Press, Boca Raton, FL, 1997, Chap. 114.
4. Koester, D. A., R. Mahadevan, A. Shishkoff, and K. W. Markus, *MUMPs Design Handbook*, rev. 6, Cronos Integrated Microsystems (A JDS Uniphase company), San Jose CA 2001.
5. Willke, T. L., and S. S. Gearhart, "LIGA Micromachined Planar Transmission Lines and Filters," *IEEE Trans. Microwave Theory Tech.*, Vol. 45, No. 10, Oct. 1997, pp. 1681–1688.
6. Willke, T. L., E. Onggosanusi, and S. S. Gearhart, "Micromachined Thick-Metal Coplanar Coupled-Line Filters and Couplers," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1998, pp. 115–118, Baltimore MD.
7. Tang, W. C., "MEMS Applications in Space Exploration," *Proc. SPIE Int. Soc. Opt. Eng.*, Vol. 3224, 1997, pp. 202–211.
8. Harsh, K. F., W. Zhang, V. Bright, and Y. C. Lee, "Flip-Chip Assembly for Si-Based RF MEMS," *Proceedings of the 12th IEEE International Conference on Microelectromechanical Systems*, MEMS '99, Jan. 1999, pp. 273–276.
9. Harsh, K. F., B. Su, W. Zhang, V. M. Bright, and Y. C. Lee, "The Realization and Design Considerations of a Flip-Chip Integrated MEMS Tunable Capacitor," *Sensors Actuators*, Vol. 80, 2000, pp. 108–118.
10. Milanovic, V., M. Maharbiz, and K. S. J. Pister, "Batch Transfer Integration of RF Microrelays," *IEEE Microwave Guided Wave Lett.*, Vol. 10, No. 8, Aug. 2000, pp. 313–315.
11. Feng, Z., H. Zhang, K. C. Gupta, W. Zhang, V. M. Bright, and Y. C. Lee, "MEMS-Based Series and Shunt Variable Capacitors for Microwave and Millimeter-Wave Frequencies," *Sensors Actuators A*, Vol. 91, 2001, pp. 256–265.
12. Park, J. Y., Y. J. Yee, H. H. Nam, and J. U. Bu, "Micromachined RF MEMS Tunable Capacitors Using Piezoelectric Actuators," *IEEE MTT-S Int. Microwave Symp. Dig.*, Phoenix AZ, May 2001, pp. 2111–2114.
13. Harsh, K. F., V. M. Bright, and Y. C. Lee, "Solder Self-assembly for Three-dimensional Microelectromechanical Systems," *Sensors Actuators*, Vol. 77, 1999, pp. 237–244.
14. Dahlmann, G. W., E. M. Yeatman, P. R. Young, I. D. Robertson, and S. Lucyszyn, "MEMS High Q Microwave Inductors Using Solder Surface Tension Self-Assembly," *IEEE MTT-S Int. Microwave Symp. Dig.*, May 2001, pp. 329–332, Phoenix AZ.
15. Harendt, C., H.-G. Graf, B. Hofflinger, and E. Penteker, "Silicon Fusion Bonding and its Characterization," *J. Micromech. Miroeng.*, Vol. 2, 1992, pp. 113–116.
16. Schomburg, W. K., D. Maas, W. Bacher, B. Bustgens, J. Fahrenberg, W. Menz,

and D. Seidel, "Assembly for Micromechanics and LIGA," *J. Micromech. Mirroeng.*, Vol. 5, 1995, pp. 57–63.

17. Osterberg, P., H. Yie, X. Cai, J. White, and S. Senturia, "Self-Consistent Simulation and Modeling of Electrostatically Deformed Diaphragms," *Proc. 1994 IEEE Microelectromechanical Syst. Conf.*, Osio, Japan, Jan. 25–28, 1994, pp. 28–32.
18. Wang, C., R. Ramadoss, S. Lee, K. C. Gupta, V. M. Bright, and Y. C. Lee, "Flexible Circuit-Based RF MEMS Switches," *Proc. 2001 ASME Int. Mech. Eng. Cong. Exposition*, New York, Nov. 11–16, 2001.
19. Huang, J.-M., K. M. Liew, C. H. Wong, S. Rajendran, M. J. Tan, and A. Q. Liu, "Mechanical Design and Optimization of Capacitive Micromachined Switch," *Sensors Actuators A*, Vol. 93, No. 3, 2001, pp. 273–285.
20. Seeley, S., *Electro-Mechanical Energy Conversion*, McGraw-Hill, New York, 1962, Section 2.3.
21. Yao, Z., S. Chen, S. Eshelman, D. Denniston, and C. Goldsmith, "Micromachined Low-Loss Microwave Switches," *J. Microelectromech. Syst.*, Vol. 8, No. 2, June 1999, pp. 129–134.
22. Brown, E. R., "RF-MEMS Switches for Reconfigurable Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. 46, No. 11, Nov. 1998, pp. 1868–1880.
23. Muldavin, J. B., and G. M. Rebeiz, "Nonlinear Electro-Mechanical Modeling of MEMS Switches," *IEEE MTT-S Int. Microwave Symp. Dig.*, Phoenix AZ, May 2001, pp. 2119–2122.
24. Chan, E. K., E. C. Kan, and R. W. Datton, "Nonlinear Dynamic Modeling of Micromachined Microwave Switches," *IEEE MTT-S Int. Microwave Symp. Dig.*, Denver CO, June 1997, pp. 1511–1514.
25. Petersen, K., "Micromechanical Membrane Switches," *IBM J. Res. Devel.*, Vol. 23, No. 4, July 1979, pp. 376–385.
26. Goldsmith, C. L., Z. Yao, S. Eshelman, and D. Denniston, "Performance of Low-Loss RF MEMS Capacitive Switches," *IEEE Microwave Guided Wave Lett.*, Vol. 8, No. 8, Aug. 1998, pp. 269–271.
27. Hyman, D., J. Lam, B. Warneke, A. Schmitz, T. Y. Hsu, J. Brown, J. Schaffner, A. Walston, R. Y. Loo, M. Mehregany, and J. Lee, "Surface-Micromachined RF MEMS Swithces on GaAs Substrates," *Int. J. RF Microwave CAE*, Vol. 9, No. 4, Aug. 1999, pp. 348–361.
28. Yao, J. J., "RF MEMS from a Device Perspective," *J. Micromech. Microeng.*, Vol. 10, 2000, pp. R9–R38.
29. Yao, J. J., and M. F. Chang, "A Surface Micromachined Miniature Switch for Telecommunications Applications with Signal Frequencies from DC up to 4 GHz," *8th Int. Conf. Solid-State Sensors Actuators*, Stockholm (Sweden), June 1995, pp. 384–387.
30. Suzuki, K., S. Chen, T. Marumoto, Y. Ara, and R. Iwata, "A Micromachined RF Microswitch Applicable to Phased-Array Antenna," *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1999, Anaheim, CA, pp. 1923–1926.
31. Muldavin, J. B., and G. M. Rebeiz, "High-Isolation CPW MEMS Shunt Switches—Part 1: Modeling," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 6, June 2000, pp. 1045–1052.

32. Muldavin, J. B., and G. M. Rebeiz, "High-Isolation CPW MEMS Shunt Swithces—Part 2: Design," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 6, June 2000, pp. 1053–1056.

33. Young, D. J., and B. E. Boser, "A Micromachined Variable Capacitor for Monolithic Low-Noise VCOs," *IEEE Solid State Sensor Actuator Workshop Dig.*, June 1996, pp. 86–89, Hilton Head SC.

34. Dec, A., and K. Suyama, "Micromachined Electro-mechanically Tunable Capacitors and Their Applications to RF IC's," *IEEE Trans. Microwave Theory Tech.*, Vol. 46, No. 12, Dec. 1998, pp. 2587–2596.

35. Yao, J. J., S. Park, and J. F. DeNatale, "High Tuning Ratio MEMS-Based Tunable Capacitors for RF Communications Applications," *IEEE Solid-State Sensor and Actuator Workshop*, pp. 124–127, Hilton Head, SC, June 1998.

36. Goldsmith, C. L., A. Malczewski, Z. J. Yao, S. Chen, J. Ehmke, and D. H. Hinzel, "RF MEMs Variable Capacitors for Tunable Filters," *Int. J. RF Microwave CAE*, Vol. 9, No. 4, Aug. 1999, pp. 362–374.

37. Hoivik, N., M. A. Michalicek, Y. C. Lee, K. C. Gupta, and V. M. Bright, "Digitally Controllable Variable High-Q MEMS Capacitor for RF Applications," *IEEE MTT-S Int. Microwave Symp. Dig.*, May 2001, pp. 2115–2118, Phoenix AZ.

38. Zou, J., C. Liu, and J. E. Schutt-Aine, "Development of a Wide-Tuning-Range Two-Parallel-Plate Tunable Capacitor for Integrated Wireless Communication Systems," *Int. J. RF Microwave CAE*, Vol. 11, No. 5, pp. 322–329, 2001.

39. Barker, N. S., and G. M. Rebeiz, "Distributed MEMS True-Time Delay Phase Shifters and Wide Band Switches," *IEEE Trans. Microwave Theory Tech.*, Vol. 46, No. 11, Nov. 1998, pp. 1881–1890.

40. Lin, Y., A. Borgioli, A. S. Nagra, and R. A. York, "K-Band 3-Bit Low-Loss Distributed MEMS Phase Shifter," *IEEE Microwave Guided Wave Lett.*, Vol. 10, No. 10, Oct. 2000, pp. 415–417.

41. Pillans, B., S. Eshelman, A. Malczewski, J. Ehmke, and C. Goldsmith, "Ka-Band RF MEMS Phase Shifters," *IEEE Microwave Guided Wave Lett.*, Vol. 9, No. 12, Dec. 1999, pp. 520–522.

42. Kim, M., J. B. Hacker, R. E. Mihailovich, and J. F. DeNatale, "A DC-to-40 GHz Four-Bit RF MEMS True-Time Delay Network," *IEEE Microwave Wireless Comp. Lett.*, Vol. 11, No. 2, Feb. 2001, pp. 56–58.

43. Malczewski, A., S. Eshelman, B. Pillmans, J. Ehmke, and C. L. Goldsmith, "X-Band RF MEMS Phase Shifters for Phased Array Applications," *IEEE Microwave Guided Wave Lett.*, Vol. 9, No. 12, Dec. 1999, pp. 517–519.

44. Liu, Y., A. Borgioli, A. S. Nagra, and R. A. York, "Distributed MEMS Transmission Lines for Tunable Filter Application," *Int. J. RF Microwave CAE*, Vol. 11, No. 5, 2001, pp. 254–260.

45. Kim, H.-T., J.-H. Park, Y.-K. Kim, and Y. Kwon, "Millimeter-Wave Micromachined Tunable Filters," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. 3, 1999, pp. 1235–1238.

46. Peroulis, D., S. Pacheco, K. Sarabandi, and L. P. B. Katehi, "Tunable Lumped Components with Applications to Reconfigurable MEMS Filters," *IEEE MTT-S Int. Microwave Symp. Dig.*, May 2001, pp. 341–344, Phoenix AZ.

47. Brank, J., J. Yao, M. Eberly, A. Malczewski, K. Varian, and C. Goldsmith, "RF

MEMS-Based Tunable Filters," *Int. J. RF Microwave CAE*, Vol. 11, No. 5, 2001, pp. 276–284.

48. Jung, S., K. Kang, J.-H. Park, K.-W. Chung, Y.-K. Kim, and Y. Kwon, "Micromachined Frequency-Variable Impedance Tuners Using Resonant Unit Cells," *IEEE MTT-S Int. Microwave Symp. Dig.*, May 2001, pp. 333–336, Phoenix AZ.
49. Lange, K. L., J. Papapolymerou, C. L. Goldsmith, A. Malczeswski, and J. Kleber, "A Recofigurable Double-Stub Tuner Using MEMS Devices," *IEEE MTT-S Int. Microwave Symp. Dig.*, May 2001, pp. 337–340, Phoenix AZ.
50. Dec, A., and K. Suyama, "Microwave MEMS-Based Voltage-Controlled Oscillators," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 11, Nov. 2000, pp. 1943–1949.
51. Young, D. J., B. E. Boser, V. Malba, and A. F. Bernhardt, "A Micromachined RF Low Phase Noise Voltage-Controlled Oscillator for Wireless Communications," *Int. J. RF Microwave CAE*, Vol. 11, No. 5, 2001, pp. 285–300.
52. Raskin, J.-P., A. R. Brown, B. T. Khuri-Yakub, and G. M. Rebeiz, "A Novel Parametric-Effect MEMS Amplifier," *IEEE/ASME J. Microelectromech. Syst.*, Vol. 9, No. 4, Dec. 2000, pp. 528–537.
53. Adler, R., "Compact Electromechanical Filters," *Electronics*, Vol. 20, 1947, pp. 100–105.
54. Lin, L., C. T.-C. Nguyen, R. T. Howe, and A. P. Pisano, "Microelectromechanical Filters for Signal Processing," *Proc. IEEE Microelectromechanical Systems Workshop*, pp. 226–231, Travemünde (Germany), Feb. 1992.
55. Nguyen, C. T.-C., "Frequency-Selective MEMS for Miniaturized Low-Power Communication Devices," *IEEE Trans. Microwave Theory Tech.*, Vol. 47, No. 8, Aug. 1999, pp. 1486–1503.
56. Yoon, J.-B., C.-H. Han, E. Yoon, and C.-K. Kim, "High Performance Three Dimensional On-Chip Inductors Fabricated by Novel Micromachining Technology for RF MMIC," *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 1523–1526, June 1999, Anaheim, CA.
57. Young, D. J., V. Malba, J. J. Ou, A. F. Bernhardt, and B. E. Boser, "A Low-Noise RF Voltage-Controlled Oscillator Using On-Chip High-Q Three-Dimensional Coil Inductor and Micromachined Variable Capacitor," *IEEE Solid-State Sensor and Actuator Workshop*, pp. 128–131, Hilton Head SC, June 1998.
58. Zhou, S., X.-Q. Sun, and W. N. Carr, "Monolithic Variable Inductor Network Using Microrelays with Combined Thermal and Electrostatic Actuation," *J. Micromech. Microeng.*, Vol. 9, Mar. 1999, pp. 45–50.
59. McGrath, W. R., C. Walker, M. Yap, and Y.-C. Tai, "Silicon Micromachined Waveguides for Millimeter-Wave and Submillimeter-Wave Frequencies," *IEEE Microwave Guided Wave Lett.*, Vol. 3, Mar. 1993, pp. 61–63.
60. Digby, J. W., C. E. McIntosh, G. M. Parkhurst, B. M. Towlson, S. Hadjiloucas, J. W. Bowen, J. M. Chamberlain, R. D. Pollard, R. E. Miles, D. P. Steenson, L. S. Karatzas, N. J. Cronin, and S. R. Davies, "Fabrication and Characterization of Micromachined Rectangular Waveguide Components for Use at Millimeter-Wave and Terahertz Frequencies," *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 8, Aug. 2000, pp. 1293–1302.
61. Shenouda, B. A., and L. Wilson Pearson, "Micromachined Waveguide for Milli-

meter Wave Applications," *Proc. Int. Conf. on Microwave and Millimeter wave Technology (ICMMT)*, Aug. 1998, pp. 615–618, Beijing (China).

62. Rebeiz, G. M., L. P. Katehi, T. M. Weller, C. Y. Chi, and S. V. Robertson, "Micromachined Filters for Microwave and Millimeter-Wave Applications," *Int. J. Microwave Millimeter-Wave Computer Aided Eng.*, Vol. 7, Feb. 1997, pp. 149–166.
63. Brown, A. R., P. Blondy, and G. M. Rebeiz, "Microwave and Millimeter-Wave High-Q Micromachined Resonators," *Int. J. RF Microwave CAE*, Vol. 9, No. 4, Aug. 1999, pp. 326–337.
64. Weller, T. M., L. P. B. Katehi, M. I. Herman, and P. D. Wanhof, "Membrane Technology (MIST-T) Applied to Microstrip: A 33 GHz Wilkinson Power Divider," *IEEE MTT-S Int. Microwave Symp. Dig.*, May 1994, pp. 911–914, San Diego CA.
65. Ryder, J. D., *Networks, Lines and Fields*, NJ, Prentice-Hall, Englewood Cliffs, 1996, Chap. 4.

PROBLEMS

14.1 An electrostatically actuated parallel-plate actuator consists of a movable gold top-electrode with length, width, and thickness of 150, 100, and 2 μm, respectively. The top electrode is anchored at the longer (150 μm) side to form a cantilever beam and is separated from the fixed bottom-electrode by a gap height of 3 μm. Young's modulus $E =$ 80 GPa and Poisson's ratio $v = 0.42$ for gold. Calculate the spring constant, critical gap height, and pull-down voltage.

14.2 A parallel-plate electrostatic actuator consists of a movable aluminum top-electrode with length, width, and thickness of 300, 200, and 2 μm, respectively. The up-position gap height of the actuator top electrode is 1.5 μm. Young's modulus $E = 70$ GPa and Poisson's ratio $v = 0.33$ for aluminum. The top electrode is anchored at both longer (300 μm) sides to form a fixed–fixed beam and has a residual stress of 120 MPa. Calculate the spring constant, critical gap height, and pull-down voltage. Also, calculate the critical gap height and pull-down voltage when a dielectric layer (Si_3N_4, $\epsilon_D = 7.6$) of thickness 0.1 μm is placed on the fixed bottom-electrode.

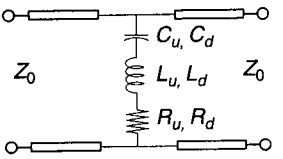
14.3 A capacitive contact RF MEMS device consists of a movable top-electrode of area 150×150 μm^2. The top-electrode up-position gap height $h_A = 3.5$ μm and the thickness of dielectric layer (SiO_2, $\epsilon_D = 3.9$) on the fixed bottom-electrode is $h_D = 0.1$ μm. Neglecting the fringing capacitance, calculate the up-position capacitance, down-position capacitance, and figure-of-merit capacitance ratio.

14.4 Design a capacitive contact series switch in a transmission line of characteristic impedance $Z_0 = 50$ Ω for the following specifications:

operating frequency range 10–30 GHz, maximum insertion loss 0.2 dB, and minimum isolation 15 dB. Find the up-position gap height required for this design when a dielectric layer ($\epsilon_D = 7.6$) of thickness 0.2 μm is deposited on the fixed bottom-electrode. Assuming that the switch electrodes form a parallel-plate configuration, calculate the required plate area (neglecting fringing capacitance) and the plate width for a length $L = 100$ μm. Ignore the inductance and resistance values for this design.

14.5 Design a capacitive contact shunt switch in a transmission line of characteristic impedance $Z_0 = 50\ \Omega$ for the following specifications: operating frequency range, K band (18–27 GHz); maximum insertion loss 0.1 dB; and minimum isolation 20 dB. Find the up-position gap height required for this design when a dielectric layer ($\epsilon_D = 3.9$) of thickness 0.1 μm is deposited on the bottom-electrode. If the switch electrodes constitute a parallel-plate configuration, calculate the required plate area (ignoring fringing capacitance) and the plate width for a length $L = 100$ μm. Assume that the inductance and resistance are negligible.

14.6 The equivalent circuit of an RF MEMS switch mounted on a transmission line ($Z_0 = 50\ \Omega$) in the shunt configuration is shown in the figure below. The up-position (ON-state) capacitance $C_u = 30$ fF (L_u and R_u may be ignored) and the down-position (OFF-state) parameters are $C_d = 3$ pF, $R_d = 0.25\ \Omega$, and $L_d = 25$ pH. Calculate the operating frequency range for a maximum insertion loss of 0.1 dB and a minimum isolation of 20 dB. Also, calculate the frequency at which the maximum isolation occurs and the value of this maximum isolation.



14.7 Find the switch down-position (OFF-state) inductance value required for X-band (8–12-GHz) operation of the switch in Problem 14.6 with minimum isolation of 20 dB at 8 GHz. Assume that the down-position capacitance and the resistance values remain the same. For this design, calculate the frequency at which the maximum isolation occurs, the isolation value at 12 GHz, and the frequency range where the isolation is better than 15 dB.

14.8 A transmission line with characteristic impedance $Z_0 = 115\ \Omega$ and effective dielectric constant $\epsilon_{\text{eff}} = 7.3$ is periodically loaded by electrostatically actuated MEMS capacitors located along the line length

(as shown in Fig. 14.29) with a spacing s = 200 μm. Each MEMS capacitor can be approximated as a parallel-plate capacitor with an up-position capacitance $C_M = 50$ fF (ignoring the fringing capacitance). The MEMS capacitors are controlled by a single DC bias voltage and are operated in the stable region. Calculate the cutoff frequency of the loaded line, impedance of the loaded line at two extreme operating positions of MEMS capacitors in the stable region, and maximum obtainable phase shift at 10 GHz for the loaded line of length 8 mm.

14.9 A 3-bit digital varactor is realized by a parallel combination of seven $(2^3 - 1)$ electrostatically actuated parallel-plate bistable capacitors. The up-position capacitance of the MEMS capacitor $C_u = 40$ fF and each capacitor has a capacitance ratio $\varsigma_c = 100$. Ignoring the parasitic capacitance of interconnects, calculate the capacitance value corresponding to each digital bit and the capacitance ratio of the digital varactor. Also, calculate the capacitance ratio of the varactor in the presence of interconnect parasitic capacitance $C_p = 1$ pF.

14.10 Design a 2-bit digital varactor by a parallel combination of three electrostatically actuated parallel-plate bistable capacitors to provide linear capacitance–actuation voltage characteristics. The up-position capacitance and capacitance ratio of each of the MEMS capacitors are $C_u = 0.1$ pF and $\varsigma_c = 10$, respectively. Calculate the capacitance values corresponding to each digital bit and the capacitance ratio of the digital varactor (ignore the interconnect parasitic capacitance in both cases). Also, calculate the actuation voltages for all other states if the actuation voltage for the first nonzero state is 20 V. If the capacitors have the same top electrode plate area and up-position gap height, find the spring constants for the other two capacitors in terms of the spring constant k_1 of the lowest value capacitor.

15

CIRCUIT FABRICATION TECHNOLOGIES

Inder Bahl

15.1 INTRODUCTION

The evolution of microwave circuits started with waveguides and coaxial lines. Still a large variety of commercially available passive components, including couplers, filters, baluns, isolators, attenuators, and detectors, are manufactured using conventional waveguides and coaxial lines. Since the early 1950s after the invention of planar transmission lines such as striplines and microstrip lines, the work on microwave planar printed circuits kicked off. This brought about a revolution in the microwave industry because several microwave functions could be batch fabricated for large-volume production. The early work on planar printed circuits served as the seeds for the successful introduction of hybrid microwave integrated circuits (MICs). Several factors contributed to the success of MIC technology, including the availability of good-quality polished alumina substrates, the evolution of cost-effective processes for thin-film metallization and high-resolution photoetching, the development of alumina drilling and cutting tools, and the proliferation of good-quality GaAs metal–semiconductor field-effect transistors (MESFETs), or simply FETs. Microwave IC technology has steadily improved in the areas of modeling, automatic manufacturing, multilayer production, and cost effectiveness. During the past 30 years, MIC technology has played a key role in the growth of microwave applications. Several MIC manufacturing technologies [1–4] are being used to reduce product size, component count, and cost. These include thin- and thick-film hybrid and low- and high-temperature cofired ceramic technologies.

Microwave Solid State Circuit Design, 2nd Edition, Edited by Inder Bahl and Prakash Bhartia
ISBN 0-471-20755-1

Building upon the success of MIC technology, a new microwave GaAs semiconductor-based technology known as monolithic microwave integrated circuit (MMIC) was introduced in the mid-1970s. Unlike MIC, in MMICs all active and passive circuit elements and interconnections are formed together on the surface of a semi-insulating substrate (usually gallium arsenide). Monolithic MICs provide low cost, improved reliability, reproducibility, small size, low weight, broadband performance, circuit design flexibility, and multifunction performance on a single chip.

Over the past 10 years, microwave technology has gone through a significant evolution to meet necessary requirements for lower cost solutions, circuit miniaturization, higher levels of integration, improved reliability, lower power consumption, low-voltage operation, and high-volume applications. Component size and weight are prime factors in the design of electronic systems for satellite communications, phased-array radar (PAR), electronic warfare, and other airborne applications, while high volume and low cost drive the PAR and consumer electronics market. Monolithic MICs are the key to meeting the above requirements.

This chapter briefly describes circuit fabrication technologies for RF and microwave applications, including printed-circuit board, microwave printed circuit, thin, thick and cofired ceramic, and monolithic IC. Their salient features such as materials, fabrication, and cost are reviewed and compared.

15.1.1 Materials

The basic materials for fabricating printed circuits and MICs, in general, are divided into the four categories, with some examples of each category:

1. Substrate materials—sapphire, alumina, beryllia, ferrite/garnet, silicon, RT/duroid, FR-4, quartz, GaAs, InP
2. Conductor materials—copper, gold, silver, aluminum
3. Dielectric films—SiO, SiO_2, Si_3N_4, Ta_2O_5
4. Resistive films—NiCr, Ta, Ti, TaN, WN, Cermet, GaAs, Si.

Substrate Materials. The substrate selected for MICs must have the following general characteristics [5–7]:

1. The cost of the substrate must be justifiable for the application.
2. The choice of thickness and permittivity determines the achievable impedance range as well as the usable frequency range.
3. The loss tangent should be low, for negligible dielectric loss.
4. The substrate surface finish should be good (~0.05–0.1 μm finish), with relative freedom from voids, to keep conductor loss low and yet maintain good metal–film adhesion.
5. There should be good mechanical strength and thermal conductivity.
6. No deformation should occur during processing of the circuit.

7. There should be a matched thermal expansion coefficient with solid state devices and package materials attached onto or to such materials to avoid susceptible to large temperature variations for improved reliability.

The dielectric constant of the substrate should be high while meeting the other criteria to keep the circuit size small. A variety of substrate materials with their properties are listed in Table 15.1. The data provided here for materials are of a general nature and must be used with care. For accurate data, refer to the manufacturer's data sheets.

Use of a high-dielectric-constant substrate $\epsilon_r \cong 10$ is highly desirable. However, the substrate thickness is limited by the presence of higher order modes. High-impedance lines on thin substrates require very narrow conductors, which become lossy, and the definition of these narrow conductors can be difficult. The temperature dependence of the dielectric constant of substrates (such as rutile) can lead to problems in certain applications, where the temperature variations are large.

For low frequencies, up to about 4–6 GHz for circuits and up to and beyond 20 GHz for array antennas, plastic substrates ($\epsilon_r \cong 2$–4) are often used. Alumina (Al_2O_3) is one of the most suitable substrate materials for use up to 20 GHz. The grade of the Al_2O_3 used depends upon the fabrication technology employed: thin or thick film. Alumina with 85% purity has high dielectric loss and poor reproducibility and is normally not used. The dielectric constant of alumina may be too high for millimeter-wave circuits because the high-impedance lines with required tolerances are difficult to fabricate and are lossy. Quartz with a dielectric constant of 4 is more suitable and widely used for high-frequency (>20 GHz) microwave and millimeter-wave ICs.

Beryllia and aluminum nitride (AlN) have excellent thermal conductivity and are suitable for power applications where heat dissipation from active devices mounted on the substrate is large and a low thermal resistance is required. Gallium arsenide is one of the most suitable substrates for MMICs, since most of the active devices, such as low-noise MESFETs, power MESFETs, and Schottky diodes, are fabricated on GaAs substrates, and the material has semi-insulating properties.

Conductor Materials. The conductor material used for microwave ICs should have high conductivity, a low temperature coefficient of resistance, low RF resistance, good adhesion to the substrate, and good etchability and solderability and be easy to deposit or electroplate [5, 7]. The resistance is determined by the RF surface resistivity and skin depth, and thus, the skin depth determines the thickness required. The conductor thickness should be at least three to four times skin depth, to include 98% of the current density within the conductor.

Table 15.2 shows the properties of some widely used conductor materials for MICs. These materials have good conductivity and can be deposited by a number of methods. Some conductors have good electrical conductivity but poor substrate adhesion, whereas others have poor electrical conductivity and

Table 15.1 Properties of Substrates for MICs

Material	Surface Roughness (μm)	Loss Tangent ($\tan\delta$) at 10 GHz (10^{-4})	Relative Dielectric Constant (ϵ_r)	Thermal Conductivity (W/cm · °C)	Dielectric Strength (kV/cm)	MIC Applications
Alumina						
99.5%	2–8	1–2	10	0.37	4×10^3	Microstrip, suspended substrate
96%	20	6	9	0.28	4×10^3	
85%	50	15	8	0.2	4×10^3	
Sapphire	1	1	9.3–11.7	0.4	4×10^3	Microstrip, lumped element
AlN	1–2	5	8.8	2.3	—	Compound substrate, package
Glass	1	20	5	0.01	—	Lumped element
Beryllia (BeO)	2–50	1	6.6	2.5	—	Compound substrate, package
Rutile	10–100	4	100	0.02	—	Microstrip
Ferrite/garnet	10	2	13–16	0.03	4×10^3	Microstrip, coplanar
GaAs (high resistivity)	1	6	12.9	0.46	350	High-frequency, microstrip, monolithic MIC
Si (high resistivity)	1	10–100	11.7	1.45	300	MMICs
Quartz	1	1	3.8	0.01	10×10^3	Microstrip, high frequency
Polyolefin	1	1	2.3	0.001	~300	
InP	—	—	14	0.68	—	MMICs

Table 15.2 Properties of Conductors for MICs

Material	Surface Resistivity (Ω/square $\times$ $10^{-7}\sqrt{f}$)	Skin Depth δ at 2 GHz (μm)	Coefficient of Thermal Expansion (α_t/°C $\times$ 10^6)	Adherence to Dielectrics	Deposition Technique
Ag	2.5	1.4	21	Poor	Evaporation
Cu	2.6	1.5	18	Poor	Evaporation plating
Au	3.0	1.7	15	Poor	Evaporation plating
Al	3.3	1.9	26	Poor	Evaporation
Cr	4.7	2.7	9.0	Good	Evaporation
Ta	7.2	4.0	6.6	Good	Electron-beam sputtering
Ti	—	—	—	Good	Evaporation sputtering
Mo	4.7	2.7	6	Fair	Electron-beam sputtering evaporation
W	4.7	2.6	4.6	Fair	Sputtering, vapor phase, electron-beam, evaporation
Pt	—	3.6	9	—	Sputtering, electron beam
Pd	—	3.6	11	—	Evaporation, sputtering, electroplating

good substrate adhesion. To obtain a good adhesion with high-conductivity material, a very thin layer (100–500 Å) of a poor conductor is deposited between the substrate and the good conductor. Some examples of typical conductor combinations to obtain good adhesion and good conductivity material are Cr/Au, Pd/Au, and Ta/Au for hybrid MICs and Cr/Au, Ti/Pt/Au, and Ti/Pd/Au for MMICs. The selection of the conductors is determined by compatibility with other materials required in the circuit and the process required. A typical adhesion layer may have a surface resistivity ranging from 500 to 1000 Ω/square but does not contribute to any loss because of its extremely small thickness.

Dielectric Film Materials. Dielectric films in MICs are used as insulators for capacitors, protective layers for active devices, and insulating layers for passive circuits. The desirable properties of these dielectric materials are reproducibility, high breakdown voltage, low-loss tangent, and the ability to undergo processing without developing pin holes [7]. Table 15.3 shows some of the properties of commonly used dielectric films in MICs. Silicon oxide is not very stable and can be used in noncritical applications, such as bypass and DC

Table 15.3 Properties of Dielectric Films for MICs

Material	Method of Deposition	Relative Dielectric Constant (ϵ_r)	Dielectric Strength (V/cm)	Microwave Q
SiO	Evaporation	6–8	4×10^5	30
SiO_2	Deposition	4	10^7	100–1000
Si_3N_4	Vapor phase sputtering	7.6	10^7	
Al_2O_3	Anodization evaporation	7–10	4×10^6	
Ta_2O_5	Anodization sputtering	22–25	6×10^6	100

blocking capacitors. A quality factor Q of more than 100 can be obtained for capacitors using SiO_2, Si_3N_4, and Ta_2O_5 materials. These materials can be deposited by sputtering or plasma-enhanced chemical vapor deposition (CVD). For high-power applications, high breakdown voltage in excess of 200 V is required. Such capacitors can be obtained with fairly thick dielectric films ($\sim$1 μm) with low probability of pin holes.

Resistive Films. Resistive films in MICs are required for fabricating resistors for terminations, attenuators, and bias networks. The properties required for a resistive material are good stability, low temperature coefficient of resistance (TCR), and sheet resistance in the range of 10–2000 Ω/square [7, 8]. Table 15.4 lists some of the thin-film resistive materials used in MICs. Evaporated Ni-chrome and tantalum nitride are the most commonly used materials.

15.1.2 Mask Layouts

Any MIC design starts with a schematic diagram for the circuit. After the circuit is finalized, a rough layout is drawn. The next step is to obtain an accurate mask layout for producing a single mask layer for hybrid MICs or a set of masks for miniature MICs and MMICs. Finally, hybrid MIC substrates are etched using these masks for the required pattern, and for miniature and monolithic MICs various photolithographic steps are carried out using a set of masks.

Table 15.4 Properties of Resistive Films for MMICs

Material	Method of Deposition	Resistivity (Ω/square)	TCR (%/°C)	Stability
Cr	Evaporation	10–1000	−0.100 to +0.10	Poor
NiCr	Evaporation	40–400	+0.001 to +0.10	Good
Ta	Sputtering	5–100	−0.010 to +0.01	Excellent
Cr–SiO	Evaporation or cement	Up to 600	−0.005 to −0.02	Fair
Ti	Evaporation	5–2000	−0.100 to +0.10	Fair

For MICs the layout is carefully prepared keeping in mind the chip or packaged devices (active and passive), crosstalk considerations, microstrip and layout discontinuities, and tuning capability. There are several techniques that have been used to produce accurate layouts for MICs. In addition to manually prepared printed-circuit taping and rubylith methods [9, 10], digitally controlled methods are being used. Both microwave CAD interactive and standalone IC layout tools are used to translate the ciricuit descriptions into mask layouts (single layer for hybrid MICs or multilayer for low-temperature cofired ceramic (LTCC)/monolithic MICs). The output is in the form of a coordinate printout, pen plot of the circuit, and the complete circuit on a magnetic tape that can be given to a mask manufacturer.

15.1.3 Mask Fabrication

Optical masks are usually used for both hybrid MICs and MMICs. However, in MMICs, new lithography techniques (considered very important for good process yield and fast turnaround) are headed in the direction of beam writing, including electron beam, focused ion beam, and laser beam. However, except for a small percentage of direct writing on the wafers (only critical geometries), optical masks are widely used. These masks are usually generated using optical techniques or electron-beam lithography.

Masks consist of sheets of glass or quartz (also called blanks) with the desired pattern defined on them in thin-film materials such as photoemulsion (silver halide based), chromium, or iron oxide. Emulsion mask coatings are still the most widely used for hybrid MICs and for noncritical working plates. Silver-halide-based emulsions have numerous advantages such as low cost, high photosensitivity, good image resolution and contrast, and reversal processing. Their major disadvantages are scratch sensitivity and higher image-defect density. Shiny chrome is the most popular hard-surface coating on glass blanks and has been proven successful for high-resolution work when used with positive optical photoresists. The main difficulty with chromium is its high reflectivity, which is solved by using an antireflection layer of chromium oxide. Iron oxide is another hard-surface coating material that has very low reflectivity and is used commonly to make "see-through" masks. Iron oxide is transparent at longer wavelengths, allowing the operator to see through the entire mask when aligning it to the pattern on the wafer. Shorter-wavelength light, at which the photoresist is sensitive and the iron oxide mask is opaque, is then used to make the exposure.

There are many different ways available for transferring digital pattern data onto mask plates [11]. The magnetic tape on which the pattern data are stored is loaded into the console, and a light-field emulsion reticle, typically at 10X, is obtained through computer control of the exposure shapes and placement. This reticle is then contact printed to yield a dark-field emulsion reticle. The next step is to make a 10X reticle on a hard-surface blank and step and repeat it into

1X emulsion master masks for the complete die. Finally these emulsion masters are contact printed to make hard-surface working plates.

15.2 PRINTED-CIRCUIT BOARDS [12, 13]

Printed-circuit boards (PCBs) or printed-wiring boards (PWBs) are extensively used for electronic packaging and RF front-end circuit boards. In these applications, the primary function of PCBs is to provide mechanical support and multilevel electrical interconnections for packaged solid state devices, resistors, capacitors, and inductors. For RF/microwave applications, there is a need for high-performance and low-cost PCB materials that can provide low-loss finer lines ($\cong$5 mils wide) and narrower spacings ($\cong$5 mils) for high-density circuits as well as provide limited impedance matching capability. Also high-speed data processing by using digital circuits requires higher performance low-dielectric-constant PCB materials. All these materials have low-loss copper conductors capable of carrying high-current densities. The PCB can be single-sided, double-sided, or consist of multilayer substrates. Multilayer PCBs have two or more layers of dielectric and metalization layers, with the latter being interconnected by plated-through via holes. Substrates may be rigid or flexible.

Substrate manufacturers have tried to combine the characteristics of various basic materials to obtain the desired electrical and mechanical properties. The resulting material is called a composite. By adding fiberglass, quartz, or ceramic in suitable proportion to the organic or synthetic materials, the mechanical properties are modified and the dielectric constant value is adjusted. A very wide variety of products are now available with dielectric constant range of 2.1–10 and $\tan\delta$ values from 0.0004 to 0.01. Table 15.5 shows important electrical and thermal parameters of several PCB materials

Table 15.5 Electrical Properties and Thermal Expansion Characteristics of Selected Dielectric Materials

Material	Dielectric Constant	Dissipation Loss	CTE, xy (ppm/°C)	CTE, z (ppm/°C)
FR-4/glass	4.5	0.03	16–20	50–70
Driclad/glass	4.1	0.01	16–18	55–65
BT/epoxy/glass	4.0	0.01	17	55–65
Epoxy/PPO/glass	3.9	0.01	12–18	150–170
Cyanate ester/glass	3.5	0.01	16–20	50–60
Polyimide/glass	4.5	0.02	12–16	65–75
Ceramic fill thermoset	3.3	0.0025	15	50
EPTFE with thermoset	2.8	0.004	50–70	50–70
Silica filled PTFE	2.9	0.003	16	24–30
PTFE/glass	2.4	0.001	12–20	140–280
PTFE	2.1	0.0004	70–90	70–90

that are currently in use. The FR-4 (fire retardant) is an epoxy-based glass substrate, widely used and has the lowest cost, while PTFE (polytetrafluoroethylene) gives the highest performance and can be operated above 300°C. FR-4, BT/epoxy, and polyimides, called thermoset materials, are hard and elastic. These materials become soft above the glass transition temperature (T_g). The glass transition temperatures of FR-4, BT/epoxy, and polyimides are about 150, 210, and 250°C, respectively. Materials such as PTFE/glass, known as thermoplastic, become soft and melt if heated. The melting temperature (T_m) of PTFE/glass is about 325°C. The coefficient of thermal expansion (CTE), given for several materials in Table 15.5, measures the dimensional stability with temperature. Thermal conductivity of these materials is quite poor, typical value is about 0.2 W/m · °C. Glass-reinforced epoxy laminates offer the lowest cost while PTFE-based laminates have the lowest dielectric constant and loss. In addition, PTFE substrates provide better protection from moisture and have ultrahigh adhesion strength. The high loss tangent of FR-4 and relatively variable ϵ_r limit its usage below 3 GHz. The values of parameters of composite materials vary slightly from manufacturer to manufacturer.

15.2.1 PCB Fabrication

Salient steps in the fabrication of such PCBs are shown in Fig. 15.1. In a basic multilayer PCB fabrication process, first copper foil is laminated to dielectric sheets and the required interconnect/wiring patterns are etched on all substrates by using a photolithography technique. The substrates are then stacked and laminated under heat and pressure to make a monolithic board. Next via holes are drilled in the board to make interlayer metallic connections, the connections are catalyzed, and the whole board is electroless copper plated. This increases the thickness of the surface conductor pattern and provides the copper layer in the via holes. The board is then tinned for soldering or nickel and gold plated for gold wire bonding. Finally, the board is cut into required small sizes.

15.2.2 Example of a PCB

The RF prototyping PCB is generally made from multilayer FR-4. The top dielectric layer is 10 mils thick. The top metal layer is made from 1 oz Cu (1.4 mils thick). The ground plane 10 mils below is made from 2 oz Cu (2.8 mils). The 10 mils thickness between the RF layer and the ground layer sets the width of a 50-Ω microstrip line to 17.5 mils wide. The total board thickness is set to 62 mils thick to be compatible with standard RF connectors. The 62-mil-thick FR-4 is very rigid for withstanding bench-top tuning.

The actual prototyping test board is shown in Fig. 15.2. The board is solder plated for ease of assembly. The exposed backside paddle of the plastic package lead frame is solder pasted to the prototype board. The FR-4 board has a land

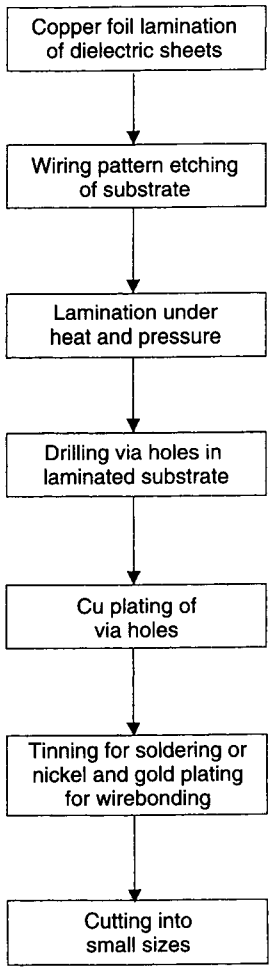


Figure 15.1 Flow diagram for multilayer PCB process.

area filled with 14-mil-diameter vias to help heat conduction away from the lead frame for power devices. This helps keep the chip cooler within the package. The open-faced packaged power amplifier (PA) is shown in Fig. 15.2.

15.3 MICROWAVE PRINTED CIRCUITS

Microwave printed-circuit (MPC) technology is widely used for microwave passive circuits and printed antennas. Substrate choice and evaluation are essential parts of the design procedure. Many substrate properties may be involved in these considerations: Dielectric constant and loss tangent and their variation with temperature and frequency, homogeneity, isotropicity, thermal

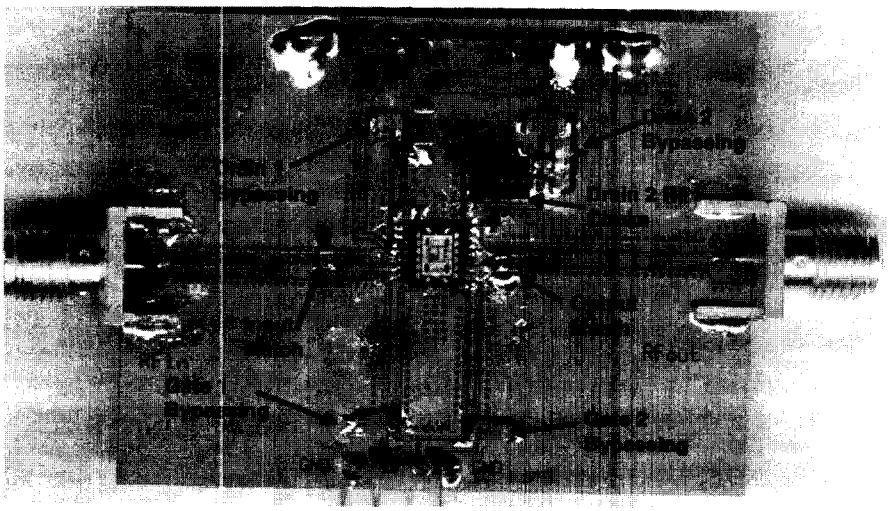


Figure 15.2 Example prototype PCB with power amplifier into TSSOP 16-pin plastic package for RF testing.

coefficient and temperature range, dimensional stability with processing, humidity, aging, and thickness uniformity of the substrate are all of importance. Similarly, other physical properties, such as resistance to chemicals, tensile and structural strengths, flexibility, machinability, impact resistance, strain relief, formability, bondability, and substrate characteristics when clad are important in fabrication.

The principal microstrip substrates currently used are listed in Table 15.6. Most types are available from several manufacturers. The large range of PTFE, hydrocarbon, and polyester composite substrates available permits considerable flexibility in the choice of a substrate for particular applications. There is no one ideal substrate and the choice rather depends on the application. For instance, conformal microwave printed circuits require flexible substrates, while low-frequency applications require high dielectric constants to keep size small. In terms of high-power operation, moisture absorption, processability, and cost, substrates such as hydrocarbon and PTFE ceramic, PTFE glass, polyester glass and hydrocarbon, and polyester glass, respectively, are more suitable.

A wide range of substrate materials is available, clad with copper, aluminum, or gold. Most of these substrates use 0.5–2 oz electrodeposited (ED) or rolled copper. Laminates are usually available in $\frac{1}{32}$, $\frac{1}{16}$, or $\frac{1}{8}$ in. thicknesses, and more recently in 10, 25, 50, 75, and 100 mils thicknesses or thicknesses in increments of 5 mils. The cladding material usually is designated in terms of weight per square yard, such as 14 g ($\frac{1}{2}$ oz), 28 g (1 oz), 57 g (2 oz), and so on. Typical cladding thicknesses for these ounce designations are given in Table 15.7. Low cladding thicknesses simplify fabrication of the MPCs to required tolerances, whereas thicker clads ease soldering. For high-power applications, a

Table 15.6 Dielectric Properties of MPC Substrate Materials at Room Temperature

Material	Trade Name	Supplier	ϵ_r	$\tan\delta$	TC of ϵ_r	k (W/m · K)	Density (g/cm^3)	CTE, X/Y	CTE, Z	T_g (°C)
Hydrocarbon glass	RO4003	Rogers	3.38	0.0025	+40	0.64	1.8	13	46	280
Hydrocarbon glass	RO4350	Rogers	3.48	0.0040	+50	0.62	1.9	15	50	280
PTFE ceramic	RO3003	Rogers	3.0	0.0013	13	0.50	2.1	17	24	325
PTEE ceramic	RO3006	Rogers	6.15	0.0025	−169	0.61	2.6	17	24	325
PTFE ceramic	RO3010	Rogers	10.2	0.0035	−295	0.66	3.0	17	24	325
PTFE glass fiber	5880	Rogers	2.2	0.0009	−125	0.2	2.2	40	237	—
PTFE glass	TLC-32	Taconic	3.2	0.0030	−125	—	—	10	70	325
PTFE glass	AR320	Arlon	3.2	0.0030	−125	—	—	10	71	325
Thermoset ceramic glass	25N	Arlon	3.25	0.0024	0	—	—	17	70	100
Polyester glass	GML 1000	Glasteel	3.05	0.0040	—	—	—	40	60	140

Note: Units of TC and CTE are ppm/°C.

Table 15.7 Standard Copper Foil Weights and Foil Thickness (t)

Foil weight				
g	14	28	57	142
oz	0.5	1	2	4
Foil thickness				
mm	0.01778	0.03556	0.07112	0.14224
in.	0.0007	0.0014	0.0028	0.0056

thick cladding is desirable. These substrates are easily machined by punching, drilling, and milling.

Over the past decade, the explosive growth in wireless RF and microwave applications has generated a significant market for lightweight, compact, and low-cost passive components such as couplers, filters, and baluns. These components must be manufactured without tuning. It is well known that the wavelength of a signal is inversely proportional to the square root of the dielectric constant of the medium in which the signal propagates. Hence, increasing the dielectric constant of the medium a hundred-fold will reduce the circuit dimensions by a factor of 10. This simple concept is being exploited extensively as distributed circuit technology is being adopted in the S band and below for cellular telephony, global positioning system (GPS) receivers, and mobile SATCOM.

A number of very high dielectric constant ceramic substrates with ϵ_r = 20–95, very low dielectric loss (Q factor = 5000–20,000), and high-temperature stability (3 ppm/°C) are currently available. They are composed of solid solutions of various titanates and are relatively inexpensive. A list of such materials with their properties is given in Table 15.8.

15.3.1 MPC Fabrication

Microwave printed-circuits are fabricated like conventional PCBs using a photoetching process. Figure 15.3 shows a flow diagram for MPC fabrication. The

Table 15.8 Dielectric Properties of High-K Ceramic Materials at Room Temperature

Ceramic	Dielectric Constant @ 1 GHz	Dielectric Loss @ 1 GHz ($\times 10^{-4}$)	Thermal Coefficient of ϵ_r (ppm/°C)	Thermal Coefficient of Expansion (ppm/°C)
TiO_2	86	2.0	−800	7–9 (rutile)
$SrTiO_3$	232	1.0	−3000	9.4
$CaTiO_3$	165	2.4	−1300	14
$BaTiO_3$	800	3.0	Almost flat nonlinear	17
$BaNdTiO_3$	92	1.0	−20 (<25°C), 20 (>25°C)	9.0

Source: From Walpita et al. [14].

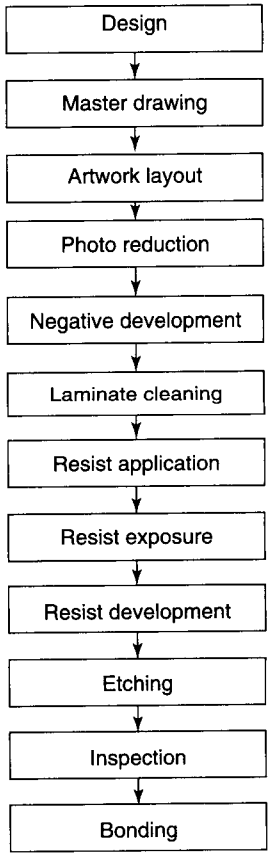


Figure 15.3 Flow diagram for microwave printed circuits.

first step is to generate the artwork from design. The enlarged artwork is then photoreduced using a high-precision camera to produce a high-resolution negative (also known as a mask), which is used for exposing the photoresist spinned over the substrate. The laminate/substrate is properly cleaned as per the manufacturer recommended procedure to ensure proper adhesion to the photoresist, which is applied to both sides of the substrate. The mask is placed on the substrate and held using a vacuum frame or other technique to assure the fine-line resolution required. With exposure to the proper wavelength light, a polymerization of the exposed photoresist occurs, making it insoluble in the developer solution. The backside of the substrate is exposed completely without a mask, since the copper foil is retained to act as a ground plane. The substrate is developed in a developer that removes the soluble photoresist material. Visual inspection is used to assure proper development.

When these steps have been completed, the substrate is ready for etching. This is a critical step and requires considerable care so that proper etch rates are achieved. After etching, the excess photoresist is removed using a stripping

solution. Visual and optical inspections should be carried out to ensure a good product and to ensure conformance with dimensional tolerances. The substrate is rinsed in water and dried.

If desired, a thermal cover bonding may be applied by placing a bonding film between the laminates to be bonded and placing these between tooling plates. Dowel pins can be used for alignment and the assembly is then heated under pressure until the bondline temperature is reached. The assembly is allowed to cool under pressure below the melting point of the bonding film and then removed for inspection.

The above procedure outlines the general steps necessary in producing a microstrip printed circuit. For example, the substances used for the various processes (e.g., cleaning and etching) or the tools used for machining depend on the substrate chosen. Most manufacturers provide informative brochures on the appropriate choice of chemicals, cleaners, and etchants for their substrates.

15.3.2 MPC Examples

Microwave printed circuit technology is exclusively applied to a wide variety of microwave passive components, including manifolds for power distribution, filters, couplers, baluns, and printed antennas. Both strip lines and microstrip lines are used. Among MPC components, directional couplers and filters are the most popular Figure 15.4 shows popular configurations of these components.

15.4 HYBRID INTEGRATED CIRCUITS

Hybrid MICs have been used almost exclusively in the frequency range of 1–20 GHz for wireless, space, and military applications, because they meet the

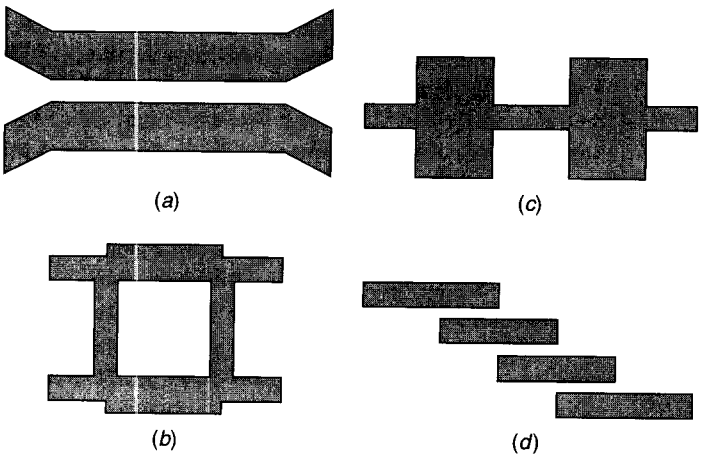


Figure 15.4 Printed-circuit components: (*a*) transmission line proximity coupler; (*b*) branch-line coupler; (*c*) low-pass filter; (*d*) bandpass filter.

requirements for shock, temperature conditions, and severe vibration. This section is intended to provide a brief introduction of several hybrid technologies such as thin film, thick film, and cofired ceramic. The most commonly used ceramic for MICs is alumina (Al_2O_3). There are a number of other ceramic materials available with ϵ_r ranging between 20 and 150. A high dielectric constant is useful for important circuit size reduction at RF and low microwave frequencies.

15.4.1 Thin-Film MICs

The thin-film fabrication technology used for MICs is constantly developing to meet the requirements of increasing frequency of operation, higher yield, and reduced costs. This can be achieved by a thin-film manufacturing process based on high-resolution photoetching that is carefully controlled, accurate, and repeatable in a clean-room environment. The first step in the fabrication process is the deposition of a first layer (seed layer) of metal film on the substrate. The selection of the film is made based on the criteria of good adhesion to the substrate and is one of the important factors in the selection of conductor material for the first layer of metal film. Some precautions specific to MIC conductors should be mentioned with regard to the deposition techniques. At RFs the electromagnetic fields are confined to several skin depths of the conductors. In order to achieve low loss, the layer of high-resistance materials such as chromium must be extremely thin. The main conductor must have a low-bulk DC resistivity for low-loss propagation. Improper processing techniques can result in high RF loss for a low-sheet-resistance material made of thin chromium and a thicker gold structure. In particular, as a result of very high substrate temperatures (>300°C) sometimes encountered during sputtering, this thin sputtered chromium layer will diffuse into an overlaying gold film. This results in a high RF loss, even though the sheet resistance may be low with a thick gold layer. Therefore, techniques such as sputtering must be used with care for MIC materials. Metal films are deposited on the substrates by three methods: vacuum evaporation, electron-beam evaporation, and sputtering.

A typical metal combination for alumina substrate is Cr/Cu/Au or NiCr/Ni/Au. A very thin seed layer of suitable metal is deposited on the substrate by one of the preceding techniques and then the bulk conductor metal is deposited by electroplating techniques. The seed layers of metal provide mechanical and electrical foundation layers on which to electroplate a good-quality bulk conductor metal. The circuit definition can be accomplished by a plate-through technique or by an etchback technique. The techniques that are used to define patterns in metal layers can influence the deposition choice. Figure 15.5 illustrates the two fabrication techniques. The plate-through technique begins with a substrate coated with a thin layer of evaporated metal. This is followed by an application of a thick photoresist, as shown in Fig. 15.5*a*. The thickness of this photoresist is similar to the thickness of the final metal film required. After defining a pattern in the photoresist, the second metal layer is plated up to the

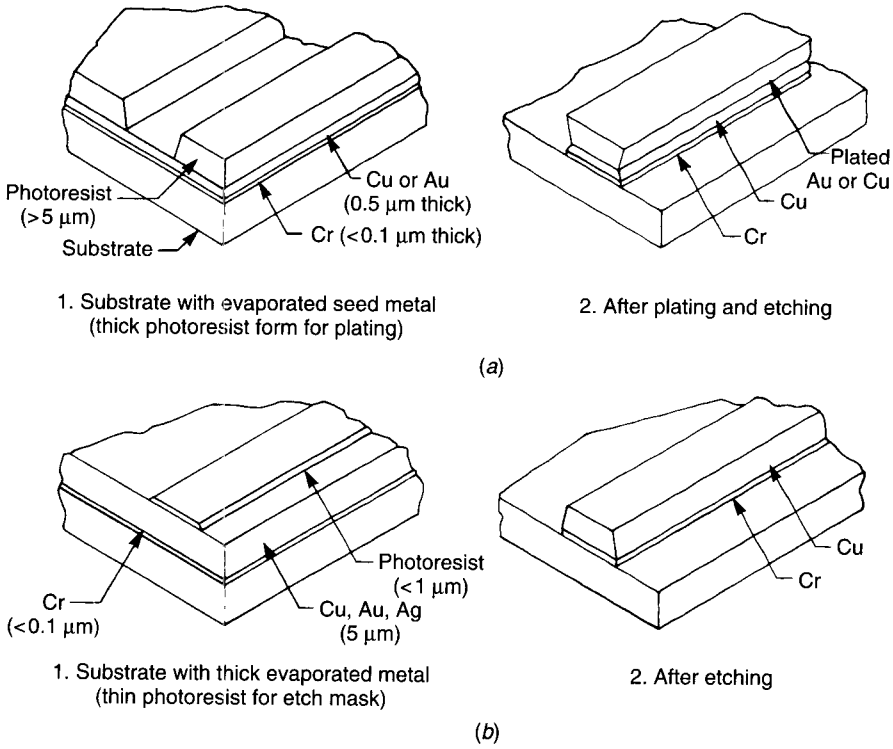


Figure 15.5 Techniques for defining conductor pattern in hybrid MICs: (*a*) fabrication by plating and etching; (*b*) fabrication by etching thin plated metal.

desired thickness with precise definition, and only in the areas where metal is required. The photoresist layer is then washed away and the thick seed metal is etched with very little undercut from the undesired areas. This technique is also suitable for fabricating lines that are 25–50 μm wide and/or when the separation between them is 25–50 μm.

The second technique is the etchback technique. This technique, as illustrated in Fig. 15.5*b*, utilizes a thick metal layer obtained either completely by evaporation or by a combination of a thin evaporated layer and a thicker plated layer. A thin photoresist layer is used as a mask to define the circuit pattern. The undesired areas of metal are then removed by etching. This technique results in undercutting the metal film by about twice the line conductor thickness. The plate-through technique not only permits better definition for thick conductors but also saves on cost in that only the required material is deposited.

In the early 1980s, a thin-film technology variant was introduced called miniature hybrid [15]. Miniature hybrid MIC technology is based on thin-film technology in which the multilevel passive circuits are batch fabricated on the

substrate and solid state devices are externally attached to these circuits. The advantages of this circuit technology are small size, light weight, excellent heat dissipation, and broadband performance.

15.4.2 Thick-Film Technology

Thick-film MICs are manufactured using various inks pressed through patterned silk screens. Thick-film MICs are inexpensive and generally limited to the lower end of the microwave spectrum. In a conventional thick-film technology, the multilayer interconnects are formed by successive screen printing of conductors, dielectric layers, and resistor patterns on a base substrate. The materials are in the form of inks or pastes. After screen printing, each layer is dried at about 150°C for 15 min and fired at about 850°C for 30–60 min. Figure 15.6 shows the first-layer screen printing process using a paste through a mesh. The mesh is designed according to the pattern to be made. The printing, drying, and firing steps are repeated to fabricate the multilayer circuitry in a fully automated way to produce high-volume, cost-effective components.

The commonly used base substrate materials are alumina (Al_2O_3), beryllia (BeO), and aluminum nitride (AlN). The dielectric pastes are typically glass–ceramic compositions having low dielectric constant and loss tangent, high breakdown voltage, and a coefficient of thermal expansion (CTE) matched to the substrate material. The conductors may be gold, copper, silver, palladium–silver/gold, and platinum–silver/gold. Properties of various conductor materials are given in Table 15.2. The commonly used resistor material is ruthenium (RuO_2) doped glass.

Recently [16] this technology has been improved by using a photoimageable thick-film process that is capable of producing 1-mil lines and gaps and 3-mil vias. In this process, both Cu and Au conductors up to 10 layers can be used. Earlier, thick-film technology was used for interconnect of discrete components; however, improved technology is also capable of printing conductor patterns for low-loss passive circuits at RF and low microwave frequencies.

15.4.3 Cofired Ceramic and Glass–Ceramic Technology

Around the time of the introduction of hybrid miniature MICs, a thick-film variant entitled low-temperature cofired ceramic (LTCC) was also introduced [17]. The LTCC manufacturing process is similar to the thick-film process except that it does not use a base substrate. Dielectric layers are in the form of unfired ceramic tapes (also called green tapes) instead of paste. This technology also enables the printing of reliable capacitors and resistors. The process as shown in Fig. 15.7 consists of blanking, punching vias, conductor screen printing, collating, laminating, and firing. The vias are punched in the green tape and filled with conducting paste. At the same time conductor patterns are screen printed. This process is carried over for each dielectric layer and finally the composite structure is fired to obtain a monolithic substrate. The firing

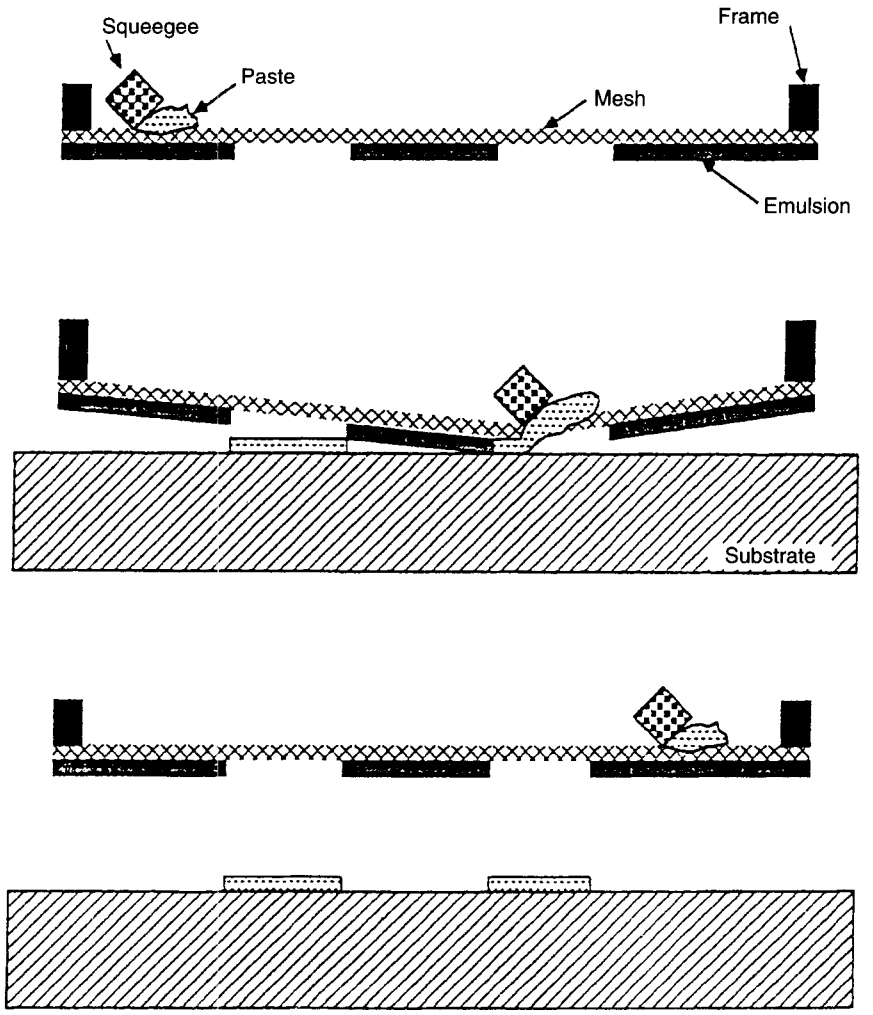


Figure 15.6 Screen printing process for material deposition onto substrate.

temperature for the glass–ceramic substrates is 850–900°C, and this technology is known as low-temperature cofired ceramic technology. Low-temperature firing allows one to use high-conductivity metals such as Ag, Cu, and gold. The dielectric tapes use a glass–ceramic composite optimized for a better CTE match with base metal and the semiconductor chips. When ceramic tapes are used, they are fired at 1500–1600°C, and the technology is known as high-temperature cofired ceramic (HTCC) technology. Commonly used conductors in this case are tungsten (W) and molybdenum (Mo). Dielectric properties of cofired glass–ceramic are compared with cofired alumina ceramic, alumina, BeO, and AlN in Table 15.9.

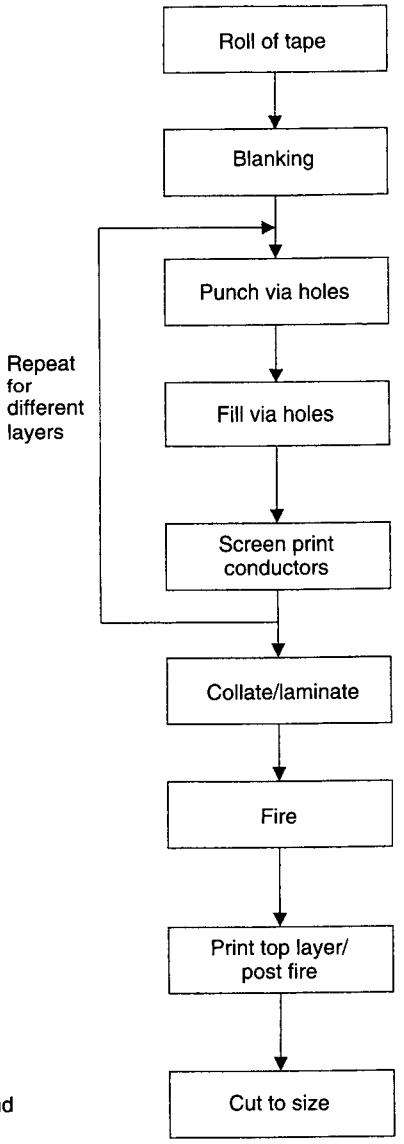


Figure 15.7 Basic steps for cofired ceramic and cofired glass–ceramic process.

Due to its multilayer process, LTCC technology offers several advantages over conventional thin-film, thick-film, and HTCC technologies. These advantages include a higher level of integration of components (e.g., capacitors, resistors, inductors, transmission lines, and bias lines) and greater design flexibility by enabling the realization of different types of transmission-line media such as microstrip, strip-line, coplanar waveguide, and rectangular coax. Passive components, matching networks, bias lines, and shielding of RF lines can

Table 15.9 Typical Electrical and Thermal Properties of Ceramic Materials

Property	Al_2O_3	HTCC	LTCC	BeO	AlN
Relative dielectric constant at 1 MHz	9.8	9.5	5.0	6.4	8.8
Loss tangent at 1 MHz	0.0002	0.0004	0.0002	0.0003	<0.001
Coefficient of thermal expansion, $10^{-6}/°C$	6.5	7.1	3.0	7.2	4.4
Thermal conductivity, $W/m \cdot °C$	37	25	2	250	230
Dielectric strength, kV/m	25	23	1.5	26	14
Density, g/cm^3	3.8	3.9	2.6	2.8	3.3

Note: Some materials have slightly different parameter values as given in Table 15.1 due to different manufacturers.

be combined in LTCC technology using several available ceramic and metal layers. Finally, solid state low-power devices are attached on the top surface to realize active or passive circuits. High-power devices can be integrated with LTCC by attaching the devices directly to the next level assembly chassis through holes fabricated in the LTCC MIC. Figure 15.8 shows the three-dimensional view of the LTCC module with embedded passive components and bias lines.

Microwave IC technology is very diverse in its application of materials and processes to implement a broad array of functions. Table 15.10 lists some of these materials and processes. The circuit functions enabled by MIC technology include oscillators, doublers, amplifiers, mixers, receivers, transmitters, and transmit/receive (T/R) modules. Figure 15.9 shows a photograph of an MIC-based 6–18 GHz four-channel T/R module. Each channel had amplitude and phase adjustment and produced 5 W of output power.

The current trend in MIC technology is to reduce the system cost by integrating as many components and circuit functions on a single substrate.

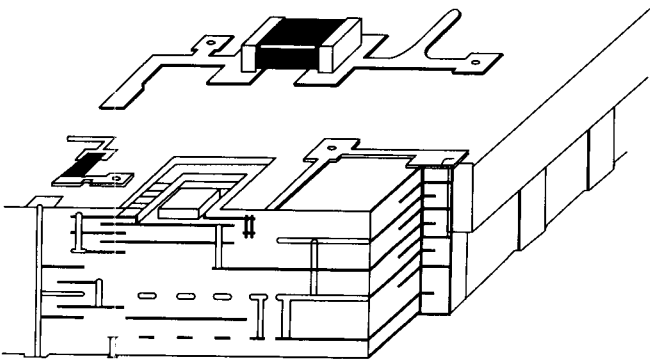


Figure 15.8 Three-dimensional view of LTCC module.

Table 15.10 Summary of Typical Materials and Processes Used to Fabricate Microwave Integrated Circuits

Materials/ Processes	Microwave Printed Circuit	Thin Film	Thick Film	Cofired Glass–Ceramic (LTCC)	Cofired Ceramic (HTCC)
Base substrates	PTFE glass fiber, PTFE ceramic, hydrocarbon ceramic, polyester glass	Al_2O_3, AlN, BeO, quartz, glass/ceramic	Al_2O_3, AlN, BeO	Not applicable	Not applicable
Conductors	Cu	Au, Al, Cu	Au, PdAu, PtAu, Ag, PdAg, PtAg, PtPdAg, Cu	Au, Ag, PdAgCu	W, Mo
Dielectrics	Not available	SiO_2, polyimide, benzocyclobutene (BCB)	Glass–ceramics, recrystallizing glasses	Glass–ceramic tape	Ceramic (Al_2O_3) tape
Resistors	Not available	NiCr, TaN	RuO_2-doped glass	RuO_2-doped glass	Not available
Processes	Photolithography, etch, collate sheets, bonding	Sequentially vacuum deposit, spin coat, and/or plate conductors, dielectrics, and resistors; photolithography; etch	Sequentially print, dry, and fire conductor, dielectric, and resistor pastes	Punch vias, print and dry conductors on tape, collate layers, laminate, cofire	Punch vias, print and dry conductors on tape, collate layers, laminate, cofire

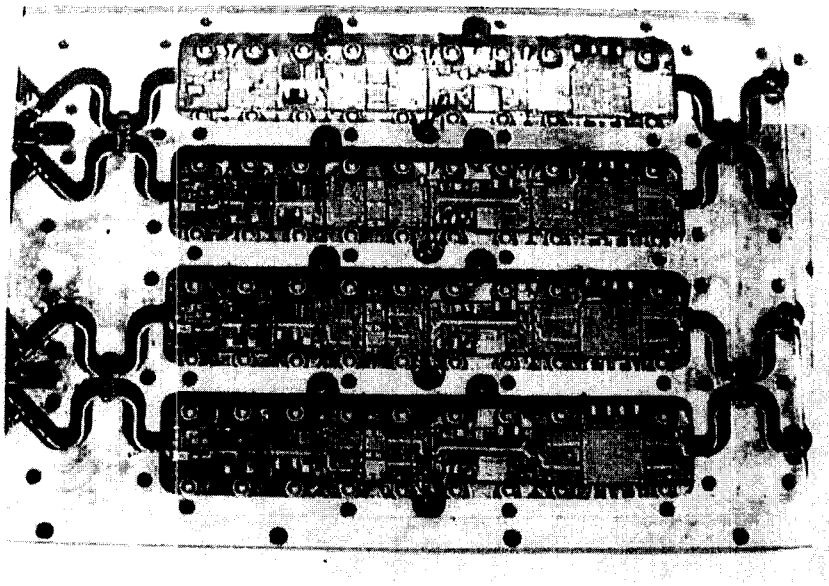


Figure 15.9 Four-channel 6–18-GHz transmit/receive MIC module designed for airborne radar appliation. (Courtesy, A. S. Virdee, Filtronic).

Various MIC technology contenders emerging for such applications are thick film, LTCC, and multilayer system-on-a-package (SOP) [3, 4, 18]. In general, thick-film applications are suitable for low frequency, LTCC for high integration, and SOP for high-performance applications.

15.5 MONOLITHIC INTEGRATED CIRCUITS

While most MMICs currently in production operate in the 0.5–30-GHz microwave range, there are increasing applications covering the millimeter wave (mmW) spectrum from 30 to 300 GHz. Monolithic technology is particularly beneficial to mmW applications through the elimination of parasitic effect of bond wires that connect discrete components in conventional hybrid MICs. In MMIC-based mmW subsystems the cost can be lowered by a factor of 10 or more as compared to hybrid solutions.

Advantages of MMICs include low cost, small size, light weight, circuit design flexibility, broadband performance, elimination of circuit tweaking, high-volume manufacturing capability, package simplification, improved reproducibility, radiation hardness, improved reliability, and multifunction performance on a single chip. Indeed, the concept of implementing a "subsystem on a chip" became a reality through monolithic microwave technology.

Typically MMICs use microstrip and metal–insulator–metal (MIM) capacitors for the matching networks, whereas at low microwave frequencies

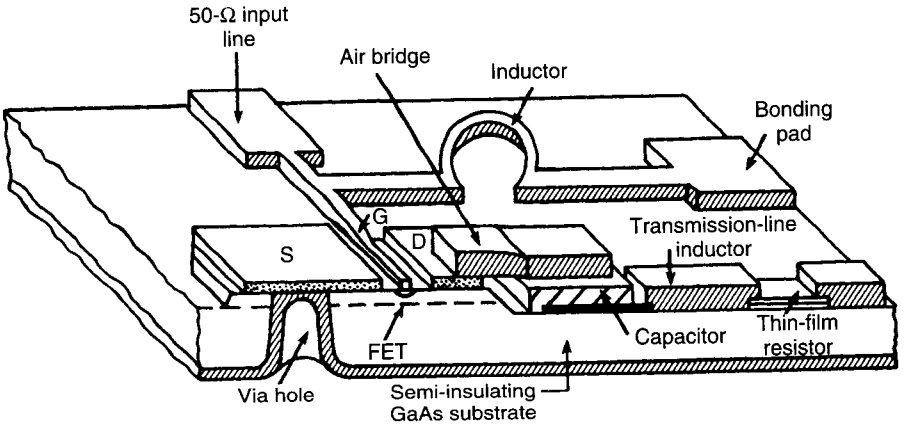


Figure 15.10 Three-dimensional view of MMIC.

lumped inductors and MIM capacitors are commonly used. Via holes, metal-filled holes from the bottom of the substrate (ground plane) to the top surface of MMICs, provide low-loss and low-inductance ground connections. Figure 15.10 shows a three-dimensional view of an MMIC.

There are many ways to fabricate MMICs [19–27]. Monolithic MICs using MESFETs and high-electron-mobility transistors (HEMTs) are most commonly fabricated with a recessed-gate process, but the self-aligned gate (SAG) process is gaining popularity because of its ability to efficiently fabricate devices optimized for different functions, such as microwave small signal, microwave power, and digital signal processing, on the same wafer at the same time. The self-aligned gate process has demonstrated superior performance uniformity in a manufacturing environment.

15.5.1 MMIC Fabrication

To give the reader an understanding of the relative complexity of GaAs MMIC manufacturing, a process flow chart for the SAG process [26] is given is Fig. 15.11. The process for recessed-gate MMICs has many similarities. It includes the fabrication of active devices, resistors, capacitors, inductors, distributed matching networks, air bridges, and via holes for ground connections through substrate. The basic process steps are similar for any MMIC technology.

It should be noted that GaAs MMIC processing is less complex than silicon processing for devices operating at the low end of the microwave spectrum. Since silicon has inherently lower frequency capability and poorer isolation properties for integration purposes, a more exotic processing is required to compete in the frequency region of overlap with GaAs applicability (~1–2 GHz). For example, a silicon BiCMOS process for such IC applications may require two or three times as many mask layers, adding significantly to the cost.

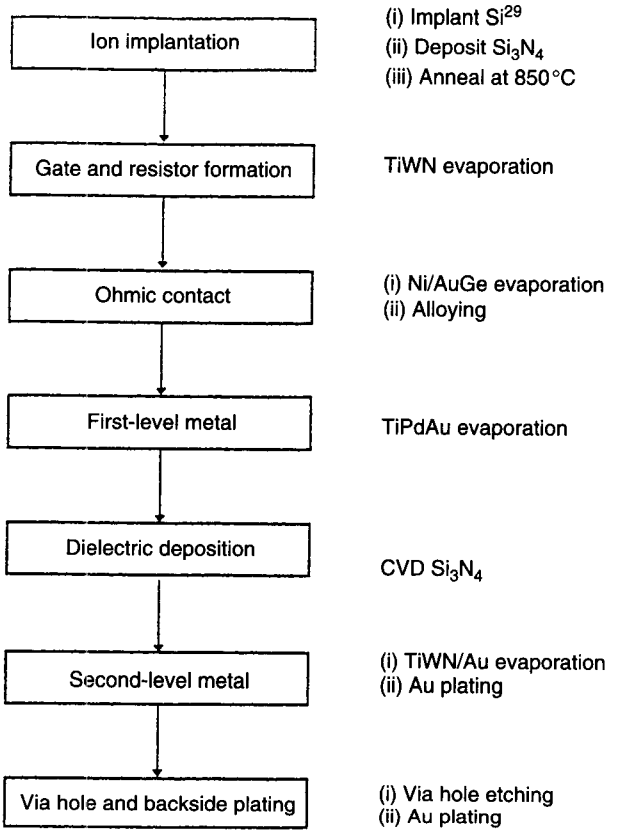


Figure 15.11 MMIC process flow chart for multifunction self-aligned gate (MSAG) process.

The MMIC process starts with the formation of an active layer on/into a qualified semi-insulating GaAs substrate. There are basically two methods of forming an *n*-type active layer: ion implantation and epitaxy. Next a gate is formed using titanium tungsten nitride (TiWN) composition. The quality and placement of the gate metal is critical to device performance in both low-noise and high-power applications. The choice of the gate metal is generally based on good adhesion to GaAs, electrical conductivity, and thermal stability.

The device ohmic contacts are made next. The purpose of an ohmic contact on a semiconductor material is to provide a good contact between the interconnect metal and the active channel at the semiconductor surface. The most common approach in industry is to fabricate ohmic contacts on GaAs by alloying gold and germanium (88% Au and 12% Ge by weight, having a melting point of 360°C). Next, a thick TiPdAu metal is overlayed on the gate by evaporation and liftoff. The metal reduces the gate resistance and also serves as a first-level metallization for MMIC fabrication, for example, as a capacitor

bottom plate or the interconnect metal under air bridges or crossovers. Dielectric films are used in GaAs MMICs for passivation of active areas of devices and resistors, for MIM capacitors, and for crossover isolation. Silicon nitride (Si_3N_4) is commonly used as a dielectric material that is easily deposited either by plasma-assisted chemical vapor deposition or sputtering. The thickness of the dielectric film determines the capacitance per unit area of the MIM capacitor. Typical values for the film thickness, capacitance, and breakdown voltage are 0.2 μm, 300 pF/mm^2, and 60 V, respectively. Interconnection of components, air bridges, and the top plate of MIM capacitors is formed with the second-level layer TiWN/Au metal system. Backside processing consisting of thinning by grinding or lapping, via hole etching, and ground contact metallization and plating is an important and cost-sensitive part of the processing. In a production environment, a significant investment has been made in the wafer by the time the frontside processing is completed and the backside processing is started. After the frontside process, the wafer is thinned by a lapping technique from ~600 μm to the required thickness, typically 100–125 μm for small-signal MMICs and 75 μm for power MMICs (to maximize heat conduction). High-performance MMICs require low-inductance ground connections to the FET source and other passive components and good thermal dissipation paths from the FET to its ground. In via hole technology, holes are etched through the GaAs substrate under each FET source connection as well as under other pads where ground connections are needed. Then the backside and the via hole sidewalls are metallized. This provides a good connection from the frontside devices and components to the backside ground plane. This also eliminates the need for separate wire bonds to ground each FET and other RF ground connections. The first check for a good circuit is automatic testing on wafer with microwave probes. After identifying RF good ICs, the wafer is diced into chips.

15.5.2 MMIC Example

Figure 15.12 shows the photograph of an MMIC 12-W *X*-band high-power amplifier. The chip uses three stages designed to operate with $V_{DD} = 10$ V at a quiescent operating point of 25% of I_{dss}. The output stage employs 20 mm gate periphery of multifunction self-aligned gate (MSAG) MESFETs [28]. Typical measured output power and power-added efficiency (PAE) were over 12 W and 35%, respectively, in the 8.5–10.5-GHz frequency range.

15.6 TECHNOLOGY COMPARISON AND CHOICES

The relative merits of MMICs vis-à-vis hybrid MICs and conventional microwave components are in the areas of cost, size and weight, high-volume production capability, design flexibility, broadband performance, high level of integration, reproducibility, and reliability. These merits are compared in Table 15.11. The performance advantages of GaAs devices for microwave circuits have been realized for over three decades in the form of hybrid circuits utilizing

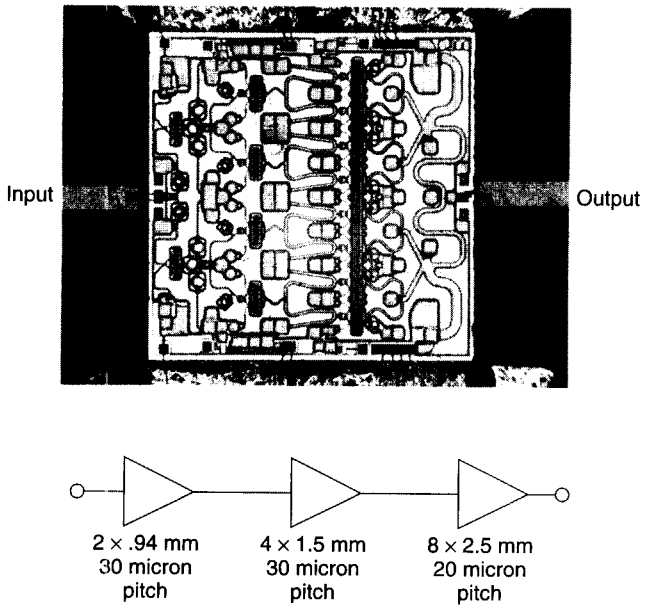


Figure 15.12 12-W HPA: layout and schematic. Chip is 4.6 × 4.6 mm.

Table 15.11 Comparison between Monolithic and Hybrid MICs

Feature	Monolithic	Hybrid
Substrate	Semi-insulator	Insulator
Interconnections	Deposited	Wire bonded/deposited
Distributed elements	Microstrip or coplanar waveguide	Microstrip and/or coplanar lines
Lumped elements	Deposited	Discrete/deposited
Solid state devices	Deposited	Discrete
Controlled parasitics	Yes	No
Labor intensive	No	Yes
Repairability	No	Yes
Equipment costs	High	Low
Mass production	Yes	No
Debugging	Difficult	Easy
Integration with digital and electrooptic ICs	Possible	Impossible
NRE Cost	Very high	Low
Production cost in high volume	Low	High
Size and weight	Small	Large
Design flexibility	Very good	Good
Circuit tweaking	Impractical	Practical
Broadband performance	Relatively good	Limited
Reproducibility	Excellent	Fair to good
Reliability	Excellent	Fair to good

discrete GaAs transistors. The further advantages of smaller size, lower weight, multifunctionality, higher reliability, and lower cost in high-volume production are now realized routinely through MMIC technology. The MESFET has become the workhorse of the MMIC industry while more complex and costly devices, such as heterojunction bipolar transistors (HBTs) and HEMTs, are beginning to play important roles in single-power-supply operation and high-performance niches, respectively.

Several solid state devices are being used to develop wireless circuits, including silicon bipolar junction transistors (BJTs), silicon metal–oxide–semiconductor field-effect transistors (MOSFETs), laterally diffused metal–oxide–semiconductor (LDMOS) transistors, MESFETs, both GaAs- and InP-based HEMTs, and both silicon germanium (SiGe) and GaAs based HBTs. Each device technology has its own merits, and an optimum technology choice for RF/microwave applications depends not only on technical issues but also on economic issues such as cost, power supply requirements, time to develop a product, time to market a product, and existing or new markets. The purpose of this section is to give an overview of various device technologies described above and compare their advantages and disadvantages.

Depending on applications, devices are required to have some of the following features:

1. Maximum power gain bandwidth
2. Minimum noise figure
3. Maximum PAE
4. Low thermal resistance
5. High temperature of operation and reliability
6. Low on-resistance/high off-resistance
7. High linearity
8. Low power dissipation
9. Low leakage current under cutoff operation
10. Low $1/f$ noise
11. Multifunctionality
12. Low single power supply
13. Semi-insulating substrate
14. Mature technology
15. Low cost

Different RF/microwave circuits require different transistor parameters. For example, power amplifiers use transistors with higher power densities, low-noise amplifiers employ transistors with low-noise characteristics, and switches use transistors having low "on-resistance" and small "off-capacitance" features. As discussed in Chapters 7 and 8, various figure-of-merit terms are used to evaluate and compare transistor characteristics, including maximum avail-

Table 15.12 Comparison of Monolithic Integrated Circuit Substrates: Pure Materials at Room Temperature

Property	Silicon	SiC	GaAs	InP	GaN
Semi-insulating	No	Yes	Yes	Yes	Yes
Resitivity (Ω-cm)	10^3–10^5	$>10^{10}$	10^7–10^9	$\sim 10^7$	$>10^{10}$
Dielectric constant	11.7	40	12.9	14	8.9
Electron mobility (cm^2/V-s)	1450	500	8500	6000	800
Saturation electrical velocity (cm/s)	9×10^6	2×10^7	1.3×10^7	1.9×10^7	2.3×10^7
Radiation hardness	Poor	Excellent	Very good	Good	Excellent
Density (g/cm^3)	2.3	3.1	5.3	4.8	—
Thermal conductivity (W/cm-°C)	1.45	4.3	0.46	0.68	1.3
Operating temperature (°C)	250	>500	350	300	>500
Energy gap (eV)	1.12	2.86	1.42	1.34	3.39
Breakdown field (kV/cm)	≈300	≥2000	400	500	≥5000

able gain, cutoff frequency (f_T or f_c), maximum frequency of oscillations (f_{max}), minimum noise figure (F_{min}), output power density, and power-added efficiency (PAE). These figures of merit for various transistors are basic requirements in addition to previously mentioned device characteristics.

Fabrication of any solid state device starts with the selection of a wafer type or substrate. Various substrate materials used for active devices are silicon, silicon carbide, GaAs, InP, and GaN. Their electrical and physical properties are compared in Table 15.12. Except for Si, all other substrate materials are called compound semiconductors. Silicon dominates the marketplace. Gallium arsenide is a distant second with less mature technologies such as InP, SiC, and GaN only now emerging. The semi-insulating property of the substrate material is crucial to providing higher device isolation and lower dielectric loss for MMICs. For example, while bipolar silicon devices are capable of operating up to about 10 GHz, the relatively low resistivity of bulk silicon precludes monolithic integration for frequencies above the *S* band (2–4 GHz). Gallium arsenide semi-insulating substrates provide isolation up to about 100 GHz. For millimeter-wave HEMTs, InP has been used. Pseudomorphic HEMTs fabricated on InP substrate exhibit much higher performance in terms of gain, noise figure, and power than a GaAs-based PHEMT of similar geometry. In this case the InP substrate supports higher two-dimensional electron gas densities, resulting in high current and transconductance values. The high value of transconductance in InP-based PHEMTs is responsible for characteristics such as ultralow noise figure, high gain, and high frequency of operation. For high-power and high-temperature applications, wide-bandgap materials with relatively high thermal conductivity such as SiC and GaN play a significant role.

Recent advancements in epitaxial techniques have made it possible to develop active devices on these substrates. Table 15.13 compares various active device technologies [29–48].

The GaAs FETs and HEMTs have the highest frequency of operation, lowest noise figure, excellent switch characteristics, and high-power and PAE capability at lower operating voltages. These salient features of GaAs FETs have enabled the introduction of power amplifiers in portable communication products such as cellular phones. They have high $1/f$ noise corner frequency (10–100 MHz), which precludes their use in ultralow phase noise oscillators. Due to the semi-insulating property of GaAs substrates, the matching networks and passive components fabricated on GaAs have lower loss than on Si. The GaAs FET as a single discrete transistor has been widely used in hybrid amplifiers (low noise, broadband, medium power, high power, high efficiency), mixers, multipliers, switching circuits, and gain control circuits. This wide utilization of GaAs FETs can be attributed to their high frequency of operation and versatility. However, increasing emphasis is being placed on new devices for better performance and higher frequency operation. The HEMT and HBT devices offer potential advantages in microwave and millimeter-wave IC applications, arising from the use of heterojunctions to improve charge transport properties (as in HEMTs) or *pn*-junction injection characteristics (as in HBTs). The HEMTs appear to have a performance edge in ultra low-noise, low loss switches, high linearity and high frequency applications. The MMICs produced using novel structures such as pseudomorphic and lattice matched HEMTs have significantly improved the noise performance and high-frequency (up to 200 GHz) operation. The PHEMTs, which utilize multiple epitaxial III–V compound layers, have shown excellent millimeter-wave power performance from the *Ku* through the *W* bands. Heterojunction bipolar transistors are vertically oriented heterostructure devices and are gaining popularity as power devices when operated using a single power supply. They offer better linearity and lower phase noise than FETs and HEMTs.

On the other hand, bipolar transistors require only a single power supply, have low leakage, low $1/f$ noise, and are produced much more inexpensively on Si. The SiGe HBTs have low cost potential of silicon BJTs and electrical performance similar to GaAs HBTs. Thus, discrete silicon BJTs, SiGe HBTs, and MOSFETs have an edge over GaAs FETs, HEMTs, and HBTs in terms of cost at low microwave frequencies. For highly integrated RF front ends, GaAs FETs and HEMTs are superior to bipolar transistors and Si substrate devices due to high-performance multifunction devices and lower capacitive loss, respectively. The electrical performance and cost trade-offs between Si and GaAs generally favor silicon devices below 2 GHz due to single-power-supply operation and lower cost, whereas above 3 GHz, GaAs-based devices are preferred due to high performance at low-voltage operation.

Figure 15.13 shows the different applications and the technology options available to designers. As technology matures and manufacturing yields improve, semiconductor costs will come down significantly and packaging,

Table 15.13 Comparison of Selected Semiconductor Device Technologies

Capability	D Mode		E Mode		GaAs HBT	SiGe HBT	Si BJT	Si MOSFET
	FET	HEMT	FET	HEMT				
Linearity	+	+	0	0	++	0	−	−
Noise figure	+	++	+	++	−	0	−	−
Power	++	++	+	+	++	++	++	++
PAE	++	++	0	0	+	+	+	+
Control circuits	++	++	+	+	0	+	−	−
Passive components	+	+	+	+	+	−	−	−
Integration on a single chip	++	+	−	−	0	0	0	0
Single-polarity supply	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Turn-on voltage control	0	0	0	0	++	++	++	+
Multiple thresholds	++	−	+	−	N/A	N/A	N/A	++
Low-voltage operation	++	++	+	+	+	0	0	−
Maximum operating temperature	++	++	++	++	+	0	0	0
$T_{j(\max)}$, °C	150	150	150	150	125	125	125	125
Wafer size, mm	100	100	100	100	100	200	200	200
Technology maturity	+	0	+	0	+	−	++	++
Cost	Moderate	High	Moderate	High	Moderate-High	Low	Low	Low
Maximum frequency of operation	+	++	+	++	+	+	0	0

Note: ++-excellent; +-good; 0-fair; −-poor.

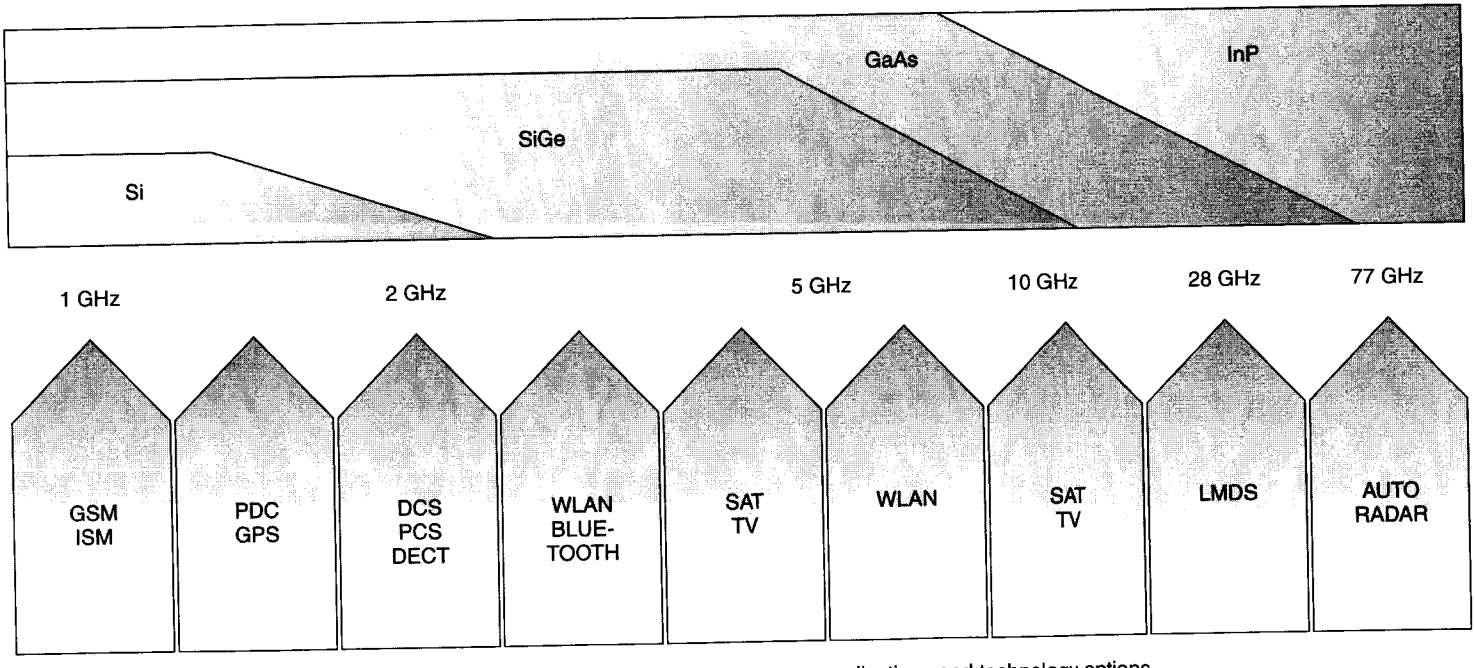


Figure 15.13 RF, microwave, and millimeter-wave applications and technology options.

testing, and other nonsemiconductor costs will become a significant part of the total product cost.

REFERENCES

1. Tummala, R. R., and E. J. Rayaszewski (Eds.), *Microelectronic Packaging Handbook*, Van Nostrand Reinhold, New York, 1989.
2. Manzione, L. T., *Plastic Packing of Microelectronic Devices*, Van Nostrand Reinhold, New York, 1990.
3. Sergent, J. E., and C. A. Harper (Eds.), *Hybrid Microelectronic Handbook*, 2nd ed., McGraw-Hill, New York, 1995.
4. Garrou, P. E., and I. Turlik, *Multichip Module Technology Handbook*, McGraw-Hill, New York, 1998.
5. Kesiter, F. Z., "An Evaluation of Materials and Processes for Integrated Microwave Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-16, July 1978, pp. 469–475.
6. Caulton, M., and H. Sobol, "Microwave Integrated Circuit Technology—A Survey," *IEEE J. Solid-State Circuits*, Vol. SC-5, Dec. 1970, pp. 292–303.
7. Sobol, H., "Applications of Integrated Circuit Technology to Microwave Frequencies," *Proc. IEEE*, Vol. 59, Aug. 1971, pp. 1200–1211.
8. Sobol, H., "Technology and Design of Hybrid Integrated Circuits," *Solid State Technol.*, Vol. 13, Feb. 1970, pp. 49–57.
9. Laverghetta, T. S., *Microwave Materials and Fabrication Techniques*, Artech House, Dedham, MA, 1984.
10. Jones, R. D., *Hybrid Circuit Design and Manufacture*, Dekker, New York, 1982.
11. Elliott, D. J., *Integrated Circuit Mask Technology*, McGraw-Hill, New York, 1985.
12. Farquhar, D. S., A. M. Seman, and M. D. Paliks, "Manufacturing Experience with High Performance Mixed Dielectric Circuit Board," *IEEE Trans. Advanced Packaging*, Vol. 22, May 1999, pp. 153–159.
13. Aguayo, A., "PTFE and Low Cost, High Performance Materials for High Speed and High Frequency Applications," *Proc. 6th Annual Wireless Symp.*, Penton Publishing, Hasbrouck Heights, NJ, 1998, pp. 243–267.
14. Walpita, L. M., et al., "Temperature-Compensated Thermoplastic High Dielectric-Constant Microwave Laminates," *IEEE Trans. Microwave Theory Tech.*, Vol. 47, Aug. 1999, pp. 1577–1583.
15. Sechi, F., et al., "Miniature Beryllia Circuits—A New Technology for Microwave Power Circuits," *RCA Rev.*, Vol. 43, 1982, p. 363.
16. Barnwell, P., and J. Wood, "A Novel Thick Film on Ceramic MCM Technology Offering MCM-D Performance," paper presented at the Sixth International Conference on Multichip Modules, Denver, CO, April 1997.
17. Brown, R. L., P. W. Polinski, and A. S. Shaikh, "Manufacturing of Microwave Modules Using Low-Temperature Cofired Ceramics," *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. 3, 1994, pp. 1727–1730.

18. Carchon, G., et al., "Multilayer Thin-Film MCM-D for the Integration of High Performance Wireless Front-End Systems," *Mirowave J.*, Vol. 44, Feb. 2001, pp. 96–110.
19. Williams, R. E., *Gallium Arsenide Processing Techniques*, Artech House, Norwood, MA, 1984.
20. Pucel, R. A. (Ed.), *Monolithic Microwave Integrated Circuits* (Reprint Volume), IEEE Press, Piscataway, NJ, 1985.
21. Ferry, D. K. (Ed.), *Gallium Arsenide Technology*, Howards, Sames, Indianapolis, IN, 1985.
22. Einspruch, N. G., and W. R. Wisseman, *GaAs Microelectronics*, Academic, New York, 1985.
23. Mun, J. (Ed.), *GaAs Integrated Circuits: Design and Technology*, McMillan, New York, 1988.
24. Bahl, I. J., and P. Bhartia, *Microwave Solid State Circuit Design*, Wiley, New York, 1988.
25. Goyal, R. (Ed.), *Monolithic Microwave Integrated Circuits: Technology and Design*, Artech House, Norwood, MA, 1989.
26. Fisher, D., and I. Bahl, *Gallium Arsenide IC Applications Handbook*, Academic, San Diego, CA, 1995.
27. Bahl, I. J., "Monolithic Microwave Integrated Circuits," in *Wiley Encyclopedia of Electrical and Electronics*, Vol. 13, Wiley, New York, 1999, pp. 490–513.
28. Griffin, E. L., "Application of Loadline Simulation to Microwave High Power Amplifiers," *IEEE Microwave Mag.*, Vol. I, June 2000, pp. 58–66.
29. Nakajima, S., et al., "Enhancement-Mode GaAs MESFET Technology for Low Consumption Power and Low Noise Applications," *IEEE Trans. Microwave Theory Tech.*, Vol. 42, Dec. 1994, pp. 2517–2523.
30. Burns, L. M., "Applications for GaAs and Silicon Integrated Circuits in Next Generation Wireless Communication Systems," *IEEE J. Solid-State Circuits*, Vol. 30, Oct. 1995, pp. 1088–1095.
31. Eason, S. D., "SiGe Stretches Limits of Silicon Applications," *Microwaves RF*, Vol. 35, Dec. 1996, pp. 89–96.
32. Halchin, D., and M. Golio, "Trends and Portable Wireless Applications," *Microwave J.*, Vol. 40, Jan. 1997, pp. 62–78.
33. Bailey, M. J., "PHEMT Devices Offer High Power Density and Efficiency," *Microwaves RF*, Vol. 36, Feb. 1997, pp. 61–70.
34. Case, M., "SiGe MMICs and Flip-Chip MICs for Low Cost Microwave Systems," *Microwave J.*, Vol. 40, May 1997, pp. 264–276.
35. O'Conner, C., "RFIC Receiver Technology for Digital Mobile Phones," *Microwave J.*, Vol. 40, July 1997, pp. 64–75.
36. Lum, E. J., "GaAs Technology Rides the Wireless Wave," *IEEE GaAs IC Symp. Dig.*, 1997, pp. 11–14.
37. Moniz, J. M., "Is SiGe the Future of GaAs for RF Applications," *IEEE GaAs IC Symp. Dig.*, 1997, pp. 229–232.
38. Wang, N. L., "Transistor Technologies for RFICs in Wireless Applications," *Microwave J.*, Vol. 41, Feb. 1998, pp. 98–110.

39. Larson, L. E., "Integrated Circuit Technology Options for RFICs—Present Status and Future Directions," *IEEE J. Solid-State Circuits*, Vol. 33, Mar. 1998, pp. 387–399.
40. Kobayashi, K. W., et al., "InP-Based HBT Technology for Next-Generation Lightwave Communications," *Microwave J.*, Vol. 41, June 1998, pp. 22–38.
41. Mickanin, W., "Don't Rule Out the MESFET," *Microwave J.*, Vol. 41, July 1998, pp. 112–120.
42. Heftman, G., "Wireless Communications Chips Span the Semiconductor Gamut," *Microwaves RF*, Vol. 35, Oct. 1998, pp. 31–39.
43. Oki, A. K., et al., "Future Technologies for Commercial and Defense Telecommunication Electronics," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1999, pp. 1069–1072.
44. Larson, L. E., "Microwave MEMS Technology for Next Generation Wireless Communications," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1999, pp. 1073–1076.
45. Larson, L. E., and M. J. Delaney, "Applications of Si/SiGe Technology for High-Speed Communications Systems," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1998, pp. 1077–1080.
46. Binari, S. C., "GaN Electronic Devices for Future Systems," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1999, pp. 1081–1084.
47. Allen, S. T., "Progress in High Power SiC Microwave MESFETs," *IEEE MTT-S Int. Microwave Symp. Dig.*, 1999, pp. 321–324.
48. Iwai, H., "CMOS Technology—Year 2010 and Beyond," *IEEE J. Solid-State Circuits*, Vol. 34, Mar. 1999, pp. 357–366.

APPENDIX A
UNITS AND SYMBOLS

A.1 SI UNITS AND THEIR SYMBOLS

In 1960 the International System of Units was established as a result of a long series of international discussions. This modernized metric system, called SI, from the French name, Systéme International d'Unités, is now, as a general world trend, to replace all former systems of measurement, including former versions of the metric system.

In the SI system, four physical quantities are classified as fundamental: length, mass, time, and charge. For practical purposes, temperature is included here as a basic unit. In Table A.1 the first five are basic quantities and the rest

Table A.1 SI Units and Their Symbols

Quantity	Unit	Symbol	Dimensions
Length	meter	m	basic
Mass	kilogram	kg	basic
Charge	coulomb	C	basic
Time	second	s	basic
Temperature	kelvin	K	basic
Frequency	hertz	Hz	1/s
Energy	joule	J	$kg \times m^2/s^2$
Force	newton	N	$kg \times m/s^2$
Power	watt	W	J/s
Pressure	pascal	Pa	N/m^2
Electric current	ampere	A	C/s
Electric potential (voltage)	volt	V	J/C
Electric field	volts/meter	V/m	J-m/C
Resistance	ohm	Ω	V/A
Resistivity	ohm-meter	Ω-m	V-m/A
Conductance	siemen	S	A/V
Capacitance	farad	F	C/V
Permittivity	farads/meter	F/m	F/m
Magnetic field	amperes/meter	A/m	A/m
Inductance	henry	H	$V \times s/A$
Permeability	henrys/meter	H/m	H/m

Table A.2 SI Prefixes

Prefix	Symbol	Factor by Which Unit Is Multiplied
exa	E	10^{18}
peta	P	10^{15}
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
hecto	h	10^{2}
deca	da	10^{1}
		10^{0}
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}

are derived quantities, that is, their dimensions can be expressed as a combination of the first five.

A.2 METRIC PREFIXES

The nomenclature in this decimal structure is derived from a system of prefixes, which are attached to units of all sorts. For example, the prefix "kilo" means 1000, hence kilometer, kilogram, and kilowatt mean 1000 meters, 1000 grams, and 1000 watts, respectively. Most of our everyday experiences with metric units will involve some of the prefixes listed in Table A.2.

A.3 DECIBEL UNITS

The ratio of signals between the output and input ports of a network is expressed in decibels and its absolute powers are measured in dBm or dBW.

The Decibel (dB). The decibel is a logarithmic unit of power ratio, although it is commonly also used for current ratio and voltage ratio. If the input power P_i and the output power P_o of a network are expressed in the same units, then the network insertion gain or loss is

$$G = 10 \log \frac{P_o}{P_i} \quad \text{dB} \tag{A.1}$$

For example, if $P_i = 5$ W and $P_o = 20$ W, then $G = 10 \log 4 = 6$ dB, that is, a power gain of 6 dB. If $P_i = 5$ W and $P_o = 2.5$ W, then $G = 10 \log 0.5 = -3$ dB, and the network is said to have a power loss of 3 dB.

The dBm and dBW. The absolute power levels of a network are expressed in dBm, which is defined as the power level P in reference to 1 mW, that is,

$$P(\text{dBm}) = 10 \log \frac{P(\text{mW})}{1 \text{ mW}} \tag{A.2}$$

Thus $P = 1$ mW $= 0$ dBm, $P = 100$ mW $= 20$ dBm, and $P = 0.5$ mW $= -3$ dBm. If power unit reference is 1 W, the decibels are expressed in dBW.

APPENDIX B

PHYSICAL CONSTANTS AND OTHER DATA

Permittivity of vacuum, $\epsilon_0 = 8.854 \times 10^{-12} \simeq (1/36\pi) \times 10^{-9}$ F/m

Permeability of vacuum, $\mu_0 = 4\pi \times 10^{-7}$ H/m

Impedance of free space, $\eta_0 = 376.7 \simeq 120\pi\ \Omega$

Velocity of light, $c = 2.998 \times 10^8$ m/s

Charge of electron, $e = 1.602 \times 10^{-19}$ C

Mass of electron, $m = 9.107 \times 10^{-31}$ kg

$\eta = e/m = 1.76 \times 10^{11}$ C/kg

Mass of proton, $M = 1.67 \times 10^{-27}$ kg

Boltzmann's constant, $k = 1.380 \times 10^{-23}$ J/K

Planck's constant, $h = 6.547 \times 10^{-34}$ J-s

10^7 erg = 1 J

1 joule = 0.6285×10^{19} eV

1 electron volt = energy gained by an electron in accelerating through a potential of 1 V

Energy of 1 electron volt = equivalent electron temperature of 1.15×10^4 K

Electron plasma frequency,

$$f_p = \frac{e}{2\pi}\left(\frac{N}{m\epsilon_0}\right)^{1/2} = 8.97\sqrt{N} \quad \text{Hz}$$

where N is the number of electrons per cubic meter

Electron cyclotron frequency, $f_c = eB/2\pi m = 28.000B$ MHz for B in webers per square meter; $f_c = 2.8B$ MHz for B in gauss

10^4 gauss = 1 Wb/m^2

Conductivity of copper, $\sigma = 5.8 \times 10^7$ S/m

Conductivity of gold, $\sigma = 4.1 \times 10^7$ S/m

APPENDIX C

ABCD AND *S*-PARAMETERS

At microwave frequencies, the generalized circuit constants matrix (*ABCD*) and scattering matrix (*S*) methods of circuit analysis are exclusively used. Calculations are most easily made using *ABCD* parameters because (1) lumped elements and transmission-line elements are related to the matrix elements by simple expressions, and (2) elements are cascaded simply by multiplying their matrices. On the other hand, scattering matrix formulation is a more general method of representing microwave networks and can handle three or more ports.

C.1 *ABCD* PARAMETERS

ABCD parameters for a two-port network, such as shown in Fig. C.1, are defined

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \tag{C.1}$$

Note that I_2 is shown to flow outward and becomes I_1 of the next two-port network in a cascaded chain. The *ABCD* matrices for commonly used microwave circuit elements are listed in Table C.1. The operations for combining *ABCD* matrices in series and in parallel are given in Fig. C.2. *ABCD* matrices exhibit the following characteristics.

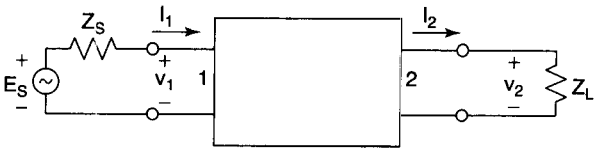
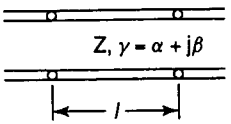
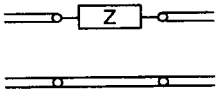
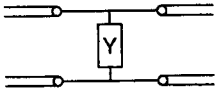
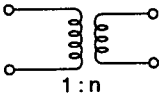
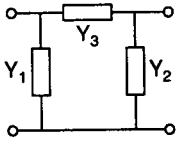
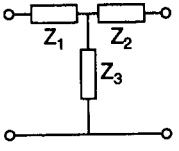
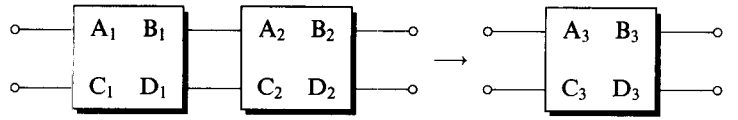


Figure C.1 A two-port network.

Table C.1 *ABCD* Matrices of Commonly Used Two-port Ladder Type Networks

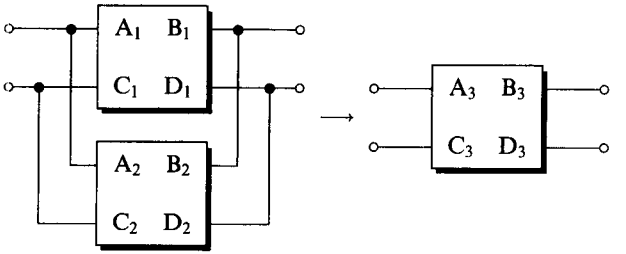
Network	ABCD Matrix
1. A transmission line section 	$\begin{bmatrix} \cosh \gamma\ell & Z \sinh \gamma\ell \\ \dfrac{\sinh \gamma\ell}{Z} & \cosh \gamma\ell \end{bmatrix}$
2. A series impedance 	$\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$
3. A shunt admittance 	$\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$
4. An ideal transformer 	$\begin{bmatrix} \dfrac{1}{n} & 0 \\ 0 & n \end{bmatrix}$
5. π-network 	$\begin{bmatrix} 1+\dfrac{Y_2}{Y_3} & \dfrac{1}{Y_3} \\ Y_1+Y_2+\dfrac{Y_1Y_2}{Y_3} & 1+\dfrac{Y_1}{Y_3} \end{bmatrix}$
6. T-network 	$\begin{bmatrix} 1+\dfrac{Z_1}{Z_3} & Z_1+Z_2+\dfrac{Z_1Z_2}{Z_3} \\ \dfrac{1}{Z_3} & 1+\dfrac{Z_2}{Z_3} \end{bmatrix}$

CASCADE:



$$\begin{bmatrix} A_3 & B_3 \\ C_3 & D_3 \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} = \begin{bmatrix} A_1A_2 + B_1C_2 & A_1B_2 + B_1D_2 \\ C_1A_2 + D_1C_2 & C_1B_2 + D_1D_2 \end{bmatrix}$$

PARALLEL:



$$\begin{bmatrix} A_3 & B_3 \\ C_3 & D_3 \end{bmatrix} = \begin{bmatrix} \dfrac{A_1B_2 + A_2B_1}{B_1 + B_2} & \dfrac{B_1B_2}{B_1 + B_2} \\ (C_1 + C_2) + \dfrac{(A_2 - A_1)(D_1 - D_2)}{B_1 + B_2} & \dfrac{D_1B_2 + D_2B_1}{B_1 + B_2} \end{bmatrix}$$

Figure C.2 *ABCD* matrix operations.

1. For reciprocal networks,

$$AD - BC = 1$$

2. For symmetrical networks, which remain unaltered when the two ports are interchanged, we have $A = D$.

C.2 S-PARAMETERS

The use of *ABCD* parameters at microwave frequencies is not very convenient from the measurements point of view. Also, its main advantage of cascading network components does not hold when the network consists of components with three or more ports and when the topology is different. Scattering matrix formulation is a more general method for representing microwave networks.

A scattering matrix represents the relationship between variables a_n (proportional to the incoming wave at the nth port) and variables b_n (proportional to the outgoing wave at the nth port) defined in the following manner

Figure C.3 *S* parameter representation.

$$a_n = \frac{v_n^+}{\sqrt{Z_{0n}}} \tag{C.2}$$

$$b_n = \frac{v_n^-}{\sqrt{Z_{0n}}} \tag{C.3}$$

where v_n^+ and v_n^- represent voltages corresponding to the incoming and the outgoing waves in the transmission line (or the waveguide) connected to the nth port, and Z_{0n} is the characteristic impedance of the line (or waveguide). Knowledge of v_n^+ and v_n^- is not required to evaluate coefficients of the scattering matrix. Relationships between b_n and a_n for the two-port network shown in Fig. C.3 may be written as

$$b_1 = S_{11}a_1 + S_{12}a_2 \tag{C.4}$$

and

$$b_2 = S_{21}a_1 + S_{22}a_2 \tag{C.5}$$

In general, for an n-port network, we have

$$[b] = [S][a]$$

Some of the important characteristics of S-matrices are listed below.

1. For a reciprocal network the S-matrix is symmetrical, that is,

$$S = S^t$$

   where the superscript t indicates the transpose of a matrix.

2. For a lossless passive network

$$\sum_{n=1}^{N} |S_{ni}|^2 = \sum_{n=1}^{N} S_{ni}S_{ni}^* = 1 \tag{C.6}$$

   for all $i = 1, 2, \ldots, N$.

3. Again for lossless passive networks, the power conservation condition

yields an orthogonality constraint given by

$$\sum_{n=1}^{N} S_{ns} S_{nr}^{*} = 0 \tag{C.7}$$

for all $s, r = 1, 2, \ldots, N$, $s \neq r$.

The relationships between $ABCD$ parameters and S-parameters are given in the following.

From S-Matrix to ABCD *Matrix*

$$A = \frac{(1 + S_{11} - S_{22} - \Delta S)\sqrt{Z_{01}/Z_{02}}}{(2S_{21})} \tag{C.8}$$

$$B = \frac{(1 + S_{11} + S_{22} + \Delta S)\sqrt{Z_{01}Z_{02}}}{(2S_{21})} \tag{C.9}$$

$$C = (1 - S_{11} - S_{22} + \Delta S)(2S_{21}\sqrt{Z_{01}Z_{02}}) \tag{C.10}$$

$$D = \frac{(1 - S_{11} + S_{22} - \Delta S)\sqrt{Z_{02}/Z_{01}}}{(2S_{21})} \tag{C.11}$$

where Z_{01} and Z_{02} are normalizing impedances for S-parameters at ports 1 and 2, respectively, and

$$\Delta S = (S_{11}S_{22} - S_{21}S_{12}) \tag{C.12}$$

From ABCD *Matrix to* S-*Matrix*

$$S_{11} = \frac{AZ_{02} + B - CZ_{01}Z_{02} - DZ_{01}}{AZ_{02} + B + CZ_{01}Z_{02} + DZ_{01}} \tag{C.13}$$

$$S_{12} = \frac{2(AD - BC)\sqrt{Z_{01}Z_{02}}}{AZ_{02} + B + CZ_{01}Z_{02} + DZ_{01}} \tag{C.14}$$

$$S_{21} = \frac{2\sqrt{Z_{01}Z_{02}}}{AZ_{02} + B + CZ_{01}Z_{02} + DZ_{01}} \tag{C.15}$$

and

$$S_{22} = \frac{-AZ_{02} + B - CZ_{01}Z_{02} + DZ_{01}}{AZ_{02} + B + CZ_{01}Z_{02} + DZ_{01}} \tag{C.16}$$

APPENDIX D

TRANSFER FUNCTION RESPONSES

This appendix presents two popular and practical responses used in matching networks, couplers, filters, namely, the Butterworth and Chebyshev responses. Both responses can be expressed mathematically as

$$H(j\omega) = \frac{K_n}{[1 + R_n^2(\omega)]^{1/2}}, \quad 0 \leq K_n \leq 1 \tag{D.1}$$

where $R_n(\omega) = \omega^n$ for Butterworth and $R_n(\omega) = \epsilon T_n(\omega)$ for Chebyshev. Here ϵ is the equial-ripple amplitude and $T_n(\omega)$ is the Chebyshev polynomial of order n.

D.1 BUTTERWORTH RESPONSE

A simple normalized Butterworth response is given by

$$H(j\omega) = \frac{1}{[1 + \omega^{2n}]^{1/2}} \tag{D.2}$$

Near $\omega = 0$,

$$[1 + \omega^{2n}]^{1/2} = 1 + \tfrac{1}{2}\omega^{2n} - \tfrac{1}{8}\omega^{4n} + \tfrac{1}{16}\omega^{6n} + \cdots \tag{D.3}$$

This expansion shows that the first $2n - 1$ derivatives are zero at $\omega = 0$. Since $R_n(\omega)$ is an nth-order polynomial, $2n - 1$ are the maximum number of derivatives that can be made zero. Thus the slope is as flat as possible near $\omega = 0$. For this reason the Butterworth response is also known as the maximally flat response.

The poles of Butterworth response $H(s)$ are the zero of the polynomial

$$1 + (-1)^n s^{2n} = 0 \tag{D.4}$$

which are given by

$$s_k = \exp \frac{j\pi(2k-1+n)}{2n}, \quad k = 1, 2, \ldots, 2n \tag{D.5}$$

These poles are located on the unit circle in the s-plane.

D.2 CHEBYSHEV RESPONSE

Instead of a maximally flat characteristic, an equial-ripple response is obtained by making the reflection in the network vary according to a Chebyshev polynomial; hence the name Chebyshev response. A Chebyshev polynomial of degree n is defined as

$$T_n(\omega) = \cos(n \cos^{-1} \omega), \qquad \omega < 1 \tag{D.6a}$$

$$= \cosh(n \cosh^{-1} \omega), \quad \omega > 1 \tag{D.6b}$$

The first seven Chebyshev polynomials may be written as

$$\begin{aligned}
T_0(\omega) &= 1 \\
T_1(\omega) &= \omega \\
T_2(\omega) &= 2\omega^2 - 1 \\
T_3(\omega) &= 4\omega^3 - 3\omega \\
T_4(\omega) &= 8\omega^4 - 8\omega^2 + 1 \\
T_5(\omega) &= 16\omega^5 - 20\omega^3 + 5\omega \\
T_6(\omega) &= 32\omega^6 - 48\omega^4 + 18\omega^2 - 1
\end{aligned} \tag{D.7}$$

A recurrence relation useful for finding the nth-degree polynomial from $(n-1)$th- and $(n-2)$th-degree polynomials is

$$T_n(\omega) = 2\omega T_{n-1} - T_{n-2} \tag{D.8}$$

Also

$$T_{-n}(\omega) = T_n(\omega) \tag{D.9}$$

For $\omega < 1$, we note that $-1 \leq T_n(\omega) \leq 1$ and increase in magnitude monotonically for $\omega > 1$.

INDEX

The new edition of an essential guide to MMIC

Monolithic microwave integrated circuits (MMICs) based on gallium arsenide (GaAs) technology are increasingly important in applications where component size and performance are prime factors. These include electronic systems for satellite communications, phased-array radar systems, electronic warfare, and other military applications, as well as consumer electronics. The new *Second Edition* of *Microwave Solid State Circuit Design* presents a comprehensive discussion of the most current trends in RF and microwave circuits technologies.

This contributed volume brings together a team of experts to provide state-of-the-art coverage of network theory basics, the design of passive circuits, solid state devices, and microwave solid state circuits. Richly supported by extensive references and problems, the book examines transmission lines and lumped elements, resonators, impedance matching networks, hybrids and couplers, filters, active and passive solid state devices, oscillators, amplifiers, detectors and mixers, microwave control circuits, frequency multipliers and dividers, MEMS, and circuit fabrication technologies. Appendixes cover S-parameters and ABCD parameters, transfer functions, including Butterworth and Chebyshev, units and symbols, as well as physical constants. Features include:

- Comprehensive coverage of passive and active RF and microwave circuit design
- Treatment of practical aspects of microwave circuits including fabrication technologies
- An overview of MEMS technology
- Treatment of heterostructure and wide-band gap devices
- Inclusion of compact and low-cost circuit design methodologies

Thorough and up to date, this *Second Edition* of a key reference remains a valuable resource for researchers, engineers, and graduate students in RF and microwave engineering.

INDER BAHL, PHD, received a PhD in electrical engineering from the Indian Institute of Technology, Kanpur, India in 1975. Currently a Distinguished Fellow of Technology at M/A-COM, Dr. Bahl has over thirty years of experience in microwave technology, including twenty years in the GaAs MMIC–related topics. He has authored over 140 research papers, ten books, and holds fifteen patents. Dr. Bahl is an IEEE Fellow and a member of the Electromagnetic Academy. **PRAKASH BHARTIA, PHD,** obtained his MSc and PhD degrees from the University of Manitoba, Winnipeg. He is currently Director General of the DRDC Ottawa laboratory in Ottawa. He has published extensively in refereed journals and has many books and patents to his credit. Dr. Bhartia is a Fellow of the Royal Society of Canada, a Fellow of the IEEE, IETE, Canadian Academy of Engineers, and the Engineering Institute of Canada. In 2002, he was awarded IEEE Canada's highest award, the A. G. L. McNaughton Gold Medal for Exceptional Contributions to Engineering.

WILEY-INTERSCIENCE
wiley.com

ISBN 0-471-20755-1
90000
9 780471 207559

BAHL
BHARTIA

Microwave Solid State Circuit Design

Second Edition

WILEY-INTERSCIENCE