*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

# Radio Frequency Circuit Design

## WILEY SERIES IN MICROWAVE AND OPTICAL ENGINEERING

**KAI CHANG,** Editor
*Texas A&M University*

A complete list of the titles in this series appears at the end of this volume.

# Radio Frequency Circuit Design

**W. ALAN DAVIS**
*University of Texas at Arlington*

**KRISHNA AGARWAL**
*Raytheon Systems Company*

A WILEY-INTERSCIENCE PUBLICATION
**JOHN WILEY & SONS, INC.**
NEW YORK / CHICHESTER / WEINHEIM / BRISBANE / SINGAPORE / TORONTO

**ISBN 0-471-20068-9.**

This title is also available in print as ISBN 0-471-35052-4.

For more information about Wiley products, visit our web site at www.Wiley.com.

***Library of Congress Cataloging-in-Publication Data:***

Davis, W. Alan.
Radio frequency circuit design / W. Alan Davis, Krishna Agarwal.
p. cm. — (Wiley series in microwave and optical engineering)
Includes index.
ISBN 0-471-35052-4
1. Radio circuits — Design and construction. I. Agarwal, Krishna K. (Krishna Kumar)
II. Title. III. Series.

TK6560 .D38 2001
621.381'32 — dc21 00-043690

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

*Dedicated to our wives,*
*Margaret Davis, Elisabeth Agarwal*
*and our children:*
*Brent, Nathan, Janelle Davis*
*Sareeta, Sandeep, Suneet Agarwal*

# Contents

# Preface

The cellular telephone has become a symbol for the rapid change in the communications business. Within this plastic container reside the talents of engineers working in the areas of efficient power supplies, digital circuit design, analog circuit design, semiconductor device design, antennas, linear systems, digital signal processing, packaging, and materials science. All these talents are carefully coordinated at a cost that allows a wide cross section of the world's population to have available instant communication. The particular aspect of all these activities that is of primary focus in this text is in the area of analog circuit design, with primary emphasis on radio frequency electronics. Some topics normally considered in electronics courses or in microwave and antenna courses are not covered here. For example, there is no mention of distributed branch line couplers, since at 1 GHz their size would be prohibitive. On the other hand, topics such as transmission line transformers are covered because they fit so well into this frequency range.

This book is meant for readers who have at least advanced standing in electrical engineering. The material in this text has been taught as a senior and graduate-level course in radio frequency circuit design at the University of Texas at Arlington. This class has continued to be popular for the past 20 years under the guidance of at least four different instructors, two of whom are the present authors. Because of the activity in the communications area, there has been ever greater interest in this subject. It is the intent of the authors, therefore, to update the current text offerings while at the same time avoiding simply reworking a microwave text.

The authors gratefully acknowledge the contribution of Michael Black, Raytheon Systems Company, to the phase lock loop discussion in Chapter 12.

W. Alan Davis
Krishna Agarwal

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

# Index

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

## WILEY SERIES IN MICROWAVE AND OPTICAL ENGINEERING

**KAI CHANG,** Editor
*Texas A&M University*

FIBER-OPTIC COMMUNICATION SYSTEMS, Second Edition
*Govind P. Agrawal*

COHERENT OPTICAL COMMUNICATIONS SYSTEMS
*Silvello Betti, Giancarlo De Marchis and Eugenio Iannone*

HIGH-FREQUENCY ELECTROMAGNETIC TECHINQUES: RECENT ADVANCES AND APPLICATIONS
*Asoke K. Bhattacharyya*

COMPUTATIONAL METHODS FOR ELECTROMAGNETICS AND MICROWAVES
*Richard C. Booton, Jr.*

MICROWAVE RING CIRCUITS AND ANTENNAS
*Kai Chang*

MICROWAVE SOLID-STATE CIRCUITS AND APPLICATIONS
*Kai Chang*

RF AND MICROWAVE WIRELESS SYSTEMS
*Kai Chang*

DIODE LASERS AND PHOTONIC INTEGRATED CIRCUITS
*Larry Coldren and Scott Corzine*

MULTICONDUCTOR TRANSMISSION-LINE STRUCTURES: MODAL ANALYSIS TECHNIQUES
*J. A. Brandão Faria*

PHASED ARRAY-BASED SYSTEMS AND APPLICATIONS
*Nick Fourikis*

FUNDAMENTALS OF MICROWAVE TRANSMISSION LINES
*Jon C. Freeman*

OPTICAL SEMICONDUCTOR DEVICES
*Mitsuo Fukuda*

MICROSTRIP CIRCUITS
*Fred Gardiol*

HIGH-SPEED VLSI INTERCONNECTIONS: MODELING, ANALYSIS, AND SIMULATION
*A. K. Goel*

FUNDAMENTALS OF WAVELETS: THEORY, ALGORITHMS, AND APPLICATIONS
*Jaideva C. Goswami and Andrew K. Chan*

ANALYSIS AND DESIGN OF INTERGRATED CIRCUIT ANTENNA MODULES
*K. C. Gupta and Peter S. Hall*

PHASED ARRAY ANTENNAS
*R. C. Hansen*

HIGH-FREQUENCY ANALOG INTEGRATED CIRCUIT DESIGN
*Ravender Goyal (ed.)*

MICROWAVE APPROACH TO HIGHLY IRREGULAR FIBER OPTICS
*Huang Hung-Chia*

NONLINEAR OPTICAL COMMUNICATION NETWORKS
*Eugenio Iannone, Franceso Matera, Antonio Mecozzi, and Marina Settembre*

FINITE ELEMENT SOFTWARE FOR MICROWAVE ENGINEERING
*Tatsuo Itoh, Giuseppe Pelosi and Peter P. Silvester (eds.)*

INFRARED TECHNOLOGY: APPLICATIONS TO ELECTRO-OPTICS, PHOTONIC DEVICES, AND SENSORS
*A. R. Jha*

SUPERCONDUCTOR TECHNOLOGY: APPLICATIONS TO MICROWAVE, ELECTRO-OPTICS, ELECTRICAL MACHINES, AND PROPULSION SYSTEMS
*A. R. Jha*

OPTICAL COMPUTING: AN INTRODUCTION
*M. A. Karim and A. S. S. Awwal*

INTRODUCTION TO ELECTROMAGNETIC AND MICROWAVE ENGINEERING
*Paul R. Karmel, Gabriel D. Colef, and Raymond L. Camisa*

MILLIMETER WAVE OPTICAL DIELECTRIC INTEGRATED GUIDES AND CIRCUITS
*Shiban K. Koul*

MICROWAVE DEVICES, CIRCUITS AND THEIR INTERACTION
*Charles A. Lee and G. Conrad Dalman*

ADVANCES IN MICROSTRIP AND PRINTED ANTENNAS
*Kai-Fong Lee and Wei Chen (eds.)*

OPTICAL FILTER DESIGN AND ANALYSIS: A SIGNAL PROCESSING APPROACH
*Christi K. Madsen and Jian H. Zhao*

OPTOELECTRONIC PACKAGING
*A. R. Mickelson, N. R. Basavanhally, and Y. C. Lee (eds.)*

OPTICAL CHARACTER RECOGNITION
*Shunji Mori, Hirobumi Nishida, and Hiromitsu Yamada*

ANTENNAS FOR RADAR AND COMMUNICATIONS: A POLARIMETRIC APPROACH
*Harold Mott*

INTEGRATED ACTIVE ANTENNAS AND SPATIAL POWER COMBINING
*Julio A. Navarro and Kai Chang*

ANALYSIS METHODS FOR RF, MICROWAVE, AND MILLIMETER-WAVE PLANAR TRANSMISSION LINE STRUCTURES
*Cam Nguyen*

FREQUENCY CONTROL OF SEMICONDUCTOR LASERS
*Motoichi Ohstu (ed.)*

SOLAR CELLS AND THEIR APPLICATIONS
*Larry D. Partain (ed.)*

ANALYSIS OF MULTICONDUCTOR TRANSMISSION LINES
*Clayton R. Paul*

INTRODUCTION TO ELECTROMAGNETIC COMPATIBILITY
*Clayton R. Paul*

ELECTROMAGNETIC OPTIMIZATION BY GENETIC ALGORITHMS
*Yahya Rahmat-Samii and Eric Michielssen (eds.)*

INTRODUCTION TO HIGH-SPEED ELECTRONICS AND OPTOELECTRONICS
*Leonard M. Riaziat*

NEW FRONTIERS IN MEDICAL DEVICE TECHNOLOGY
*Arye Rosen and Harel Rosen (eds.)*

ELECTROMAGNETIC PROPAGATION IN MULTI-MODE RANDOM MEDIA
*Harrison E. Rowe*

ELECTROMAGNETIC PROPAGATION IN ONE-DIMENSIONAL RANDOM MEDIA
*Harrison E. Rowe*

NONLINEAR OPTICS
*E. G. Sauter*

ELECTROMAGNETIC FIELDS IN UNCONVENTIONAL MATERIALS AND STRUCTURES
*Onkar N. Singh and Akhlesh Lakhtakia (eds.)*

FUNDAMENTALS OF GLOBAL POSITIONING SYSTEM RECEIVERS: A SOFTWARE APPROACH
*James Bao-yen Tsui*

InP-BASED MATERIALS AND DEVICES: PHYSICS AND TECHNOLOGY
*Osamu Wada and Hideki Hasegawa (eds.)*

DESIGN OF NONPLANAR MICROSTRIP ANTENNAS AND TRANSMISSION LINES
*Kin-Lu Wong*

FREQUENCY SELECTIVE SURFACE AND GRID ARRAY
*T. K. Wu (ed.)*

ACTIVE AND QUASI-OPTICAL ARRAYS FOR SOLID-STATE POWER COMBINING
*Robert A. York and Zoya B. Popović (eds.)*

OPTICAL SIGNAL PROCESSING, COMPUTING AND NEURAL NETWORKS
*Francis T. S. Yu and Suganda Jutamulia*

SiGe, GaAs, AND InP HETEROJUNCTION BIPOLAR TRANSISTORS
*Jiann Yuan*

ELECTRODYNAMICS OF SOLIDS AND MICROWAVE SUPERCONDUCTIVITY
*Shu-Ang Zhou*

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER ONE

# Communication Channel

## 1.1 BASIC TRANSMITTER–RECEIVER CONFIGURATION

The design of radio frequency (RF) circuits borrows from methods used in low-frequency audio circuits as well as from methods used in design of microwave circuits. Yet there are also important departures from these techniques, so the design of radio frequency circuits requires some specialized techniques not found in these other frequency ranges. The radio frequency range for present purposes will be taken to be somewhere between 300 MHz and 3 GHz. It is this frequency range where much of the present day activity in wireless communication occurs. In this range of frequencies, the engineer must be concerned about radiation, stray coupling, and frequency response of circuit elements that from the point of view of lumped, low-frequency analysis might be expected to be independent of frequency. At the same time the use of common microwave circuit elements such as quarter wave transformers is impractical because of the long line lengths required. The use of monolithic circuits have enabled many high-frequency designs to be implemented with lumped elements, yet the frequency response of these "lumped" elements still must be carefully considered.

Today RF and digital designs have begun to move closer together, so typical communication systems incorporate both of these disciplines in their design. While direct digitizing of RF signals remains a challenge, there are many systems where digital signal processing is playing a larger role than ever in communication systems. A typical radio analog transmitter and receiver is shown in Fig. 1.1. In this system the information source could be an audio or video signal. This information in the process of being converted from, say, sound to an electrical signal by a transducer produces a very low voltage that must be amplified by an audio amplifier.

The modulator is shown schematically as a mixer that represents a wide variety of different modulation schemes. The two major categories are analog and digital modulation. In either case the modulator performs two functions. The first function is that it encodes the message in a certain way so as to meet the communication channel requirements for cost, noise immunity, fading, available

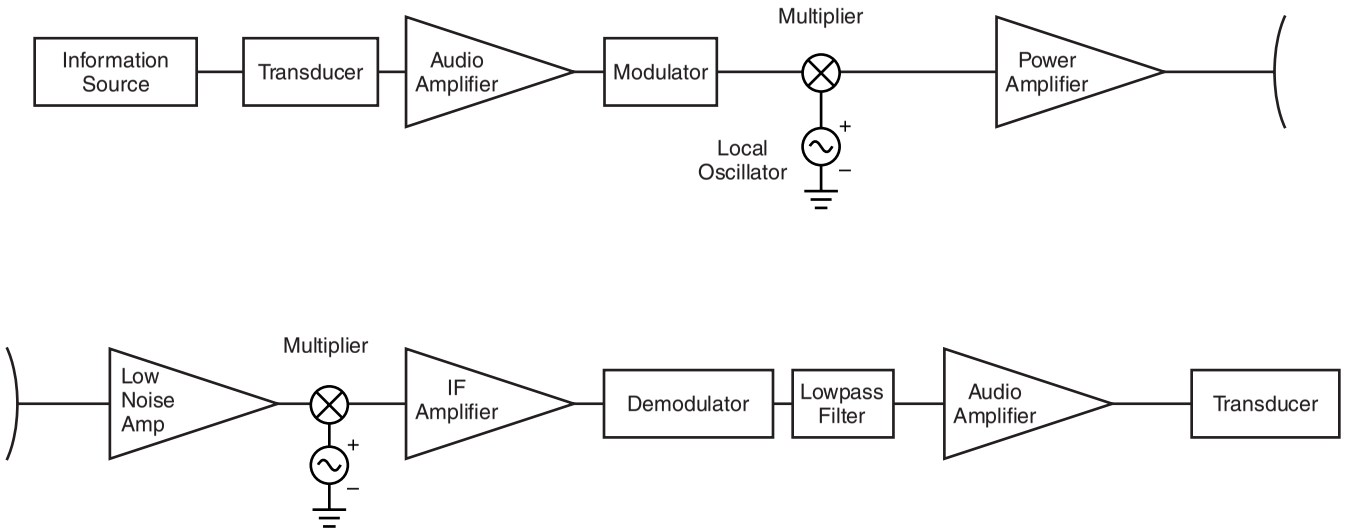



**FIGURE 1.1** Diagram of a communication transmitter and receiver.

bandwidth, bandwidth efficiency (the ratio of the throughput data rate per Hertz in a given bandwidth), power efficiency (which measures the ability of a system to preserve the message under low power conditions), and so on.

For the amplitude modulation (AM) case, the mixer is a multiplier that multiples the information message with the local oscillator frequency. Just as the product of two sine waves produces sum and difference frequencies, so the message frequency is added to the local oscillator frequency. This produces two effects necessary for practical wireless communications. The first is that this enables forming multiple channels, which in the amplitude and frequency modulation (FM) analog systems are separated by different frequency bands. Otherwise, there would be massive interference between different signals. This method of separating signals is called frequency division multiple access (FDMA). Alternate methods are time division multiple access (TDMA) where two or more signals may share the same frequency band but use it at different times. The human receiver is able to integrate over the different time slots so that the message is perceived to be continuous. A third method is the spread spectrum technique known as code division multiple access (CDMA) where a broad bandwidth is used by multiple users continuously. However, each user sends and receives data that is coded in a particular way, different from all the other users. When there is interference between users, it is perceived as low-level noise.

The second function of the modulator is that it translates the message information to a much higher RF signal. For this reason antennas can be made a manageable size, with their mechanical size normally correspondings to the wavelength. A great deal of effort has gone into making smaller antennas, but there are always design compromises.

The final stage of the transmitter before reaching the antenna is the power amplifier. Since this component uses the greatest amount of power, high efficiency becomes an important factor. In FM systems, class C amplifiers are often used because, in practice, they can produce efficiencies as high as 70%. For AM systems, class A or AB amplifiers are often used because of the required linearity of AM signal transmission. However, class A amplifiers typically have efficiencies of 30% to 40%.

As for the receiver, the received signal is sometimes strong enough to be put directly into the mixer. However, as will be seen later (in Chapter 8), the overall noise response of the amplifier is greatly enhanced by using a low-noise amplifier for the front end. The demodulator in the receiver must correspond to the modulator in the transmitter. The subsequent intermediate frequency (IF) amplifier includes the required filtering to provide the desired selectivity for the received signal. The IF frequency is chosen to be sufficiently high to avoid most of the $1/f$ noise ($f$ = frequency) or flicker noise. Since this circuit operates at a fixed frequency, it can be carefully tuned for optimum performance.

## 1.2 INFORMATION AND CAPACITY

RF communication systems provide a means of carrying information from the transmitter to the receiver. Now, what exactly is information? *Webster's*

*Dictionary* states that "information" is "knowledge communicated or received concerning a particular fact or circumstance. . .." A narrower, technical definition that more closely aligns with our focus that "information" is an "indication of the number of possible choices of messages, which are expressible as the value of some monotonic function of the number of choices, usually log to the base 2." "Information" then is a term for data that can be coded for digital processing.

Some examples of data that illustrate the meaning of information is helpful. If a signal were sent through a communication channel that never changed, then it would be conveying no information. There must be change to convey a message. If the signal consisted of 1 0 1 0 1 0 1 0. . . , there would be changes in the signal but still no information is conveyed, because the next bit would be perfectly predictable. So while change is important, it is not the sole criterion for information. There is one last example. If a signal in an amplitude modulation system consists of purely random voltage fluctuations, then again no information can be transmitted. It is simply noise, and the receiver becomes no more knowledgeable upon having heard it.

A communication system consists of a transmitter, a receiver, and a channel. The channel is capable of carrying only a certain limited amount of information. Roughly analogous to an information channel is a water pipe which, because of its diameter, is restricted to carrying only a certain amount of water. This limitation is given the technical term "capacity." It refers to the amount of information that is transmitted over a time interval of $T$ seconds. The time interval can be broken up into shorter time intervals, each of duration $\tau$. Clearly, the more distinct time intervals $\tau$ these are in the total time span $T$, the more information can be transmitted. The minimum size of $\tau$ is determined by how well one pulse in one time frame can be distinguished from a pulse in a neighboring time frame. The limitation on how short a time frame can be is related to the channel bandwidth.

In addition the signal voltage will have a maximum amplitude that is limited by the available power in the system. This voltage range can be divided into many levels, each level representing a bit of information that is distinguished from another bit. The voltage range cannot be infinitely split because of the noise that is always present in the system. Clearly, the more voltage intervals there are in a given time frame $\tau$, the more information capacity there is in the system. Just as the flow of water through a pipe is limited by the amount of pressure on the water, by the friction on the walls of the pipe, and by the diameter of the pipe, so the capacity of a transmission system is limited by the maximum voltage level, by the noise in the system that tends to muddle the distinction between one voltage level and another, and by the bandwidth of the channel, which is related to the rise time of a pulse in the system.

In one of the time intervals, $\tau$, there are $n$ voltage levels. The smaller that $\tau$ is and the larger that $n$ is, the more information there can be transmitted through the channel. In each time interval there are $n$ possible voltage levels. In the next time interval there are also $n$ possible voltage levels. It is assumed that the voltage

level in each time frame is independent of what is going on in other time frames. The amount of information transmitted in a total of $T$ seconds corresponds to the product of the possible levels in each interval:

$$n \cdot n \cdot n \cdot n \ldots n^{T/\tau} \tag{1.1}$$

The total information transmitted intuitively is directly proportional to the total time span $T$, which is defined as the log of the above product. By convention, the base-2 logarithm is used.

$$H = \frac{T}{\tau} \log_2 n \tag{1.2}$$

The system's capacity is simply the maximum *rate* of transmission (in bits/s) through a system:

$$C = \frac{H}{T} = \frac{1}{\tau} \log_2 n \tag{1.3}$$

System capacity is inversely proportional to the minimum time interval over which a unit of information can be transmitted, $\tau$. Furthermore, as the number of voltage levels increases, so does the capacity for more information.

Information can be transmitted through a channel in a variety of different forms, all producing the same amount of information. For example, suppose that a signal can take on any one of eight different voltage levels, 0, 1, ..., 7, in a given time interval $\tau$. But the eight signal levels could also equally be sent with just two levels, 0, 1. For every interval that has eight possible levels, three intervals will be needed for the two-level signal. A convenient conversion between the two systems is shown in Table 1.1.

Clearly, a 16-level signal could be transmitted by a sequence of four binary signals, and a 32-level signal with a sequence of five binary signals, and so on. For $n$ levels, $\log_2 n$ bits are needed. The information content of a signal is defined then to be the number of binary choices, or bits, that are needed

**TABLE 1.1 Eight-Level and Two-Level Systems**

| $n = 8$ | $n = 2$ |
|---|---|
| 0 | 000 |
| 1 | 001 |
| 2 | 010 |
| 3 | 011 |
| 4 | 100 |
| 5 | 101 |
| 6 | 110 |
| 7 | 111 |

for transmission. A system is designed to transmit speech must be designed to have the capacity to transmit the information contained in the speech. While speech is not entirely what humans communicate, in a communication system it is what engineers have to work with. A decision must be made as to with what fidelity the speech is to be transmitted. This translates to the bandwidth requirement of an analog system, or the number of voltage levels available in a given total voltage range. Ultimately this restriction is always present even if sophisticated coding techniques are used. The capacity of the system must be $\geq$, the rate of information that is to be transmitted. Beyond this capacity, system cost, power levels, and available transmission media must be considered.

## 1.3 DEPENDENT STATES

The definitions of the preceding section imply that the voltage level in each time interval, $\tau$, is independent of the voltage level in other time intervals. However, one simple example where this is not the case is the transmission of the English language. It is known that in the English language the letter $e$ appears more frequently than the letter $z$. It is almost certain that the letter $q$ will be followed by the letter $u$. So, in transmitting a typical message in English, less information is actually sent than would be sent if every letter in the alphabet were equally likely to occur. A way to express this situation is in terms of probability. We are interested in the total number of signal combinations that could occur in a message $T$ seconds long if each interval that is independent from the others is $n^T/\tau$. On average, every possible message $T$ seconds long would have a probability of occurrence of $1/n^{T/\tau}$.

The probability takes the form

$$P = \frac{\text{number of occurrences of a particular event}}{\text{total number of events}} \tag{1.4}$$

For information measured in terms of probability, $P = 1/n$ if there are $n$ events specified as $n$ voltage levels and each of these events is equally likely. For any one event, the information transmitted is written $H_1 = -\log_2 P$. For $m$ intervals, each $\tau$ seconds long, there will be $m$ times more information. So for $m$ intervals, the information written in terms of probability is

$$H = \frac{T}{\tau}\log_2 n = -m\log_2 P \quad \text{bits} \tag{1.5}$$

Consider a binary system where a number 0 occurs with probability $p$ and the number 1 occurs with probability $q$. Knowing that $p + q = 1$, the information content of a message consisting of 0's and 1's is to be found. The total information is the sum of the information carried by the 0's and that of the 1's:

$$H = -\frac{T}{\tau}(p \cdot \log_2 p + q \cdot \log_2 q) \quad \text{bits} \tag{1.6}$$

If the probabilities of $p$ and $q$ are each 0.5, then the total information in $T$ seconds is $T/\tau$. If, for example, $p = 0.25$ and $q = 0.75$, then

$$H = -\frac{T}{\tau}(0.25 \cdot \log_2 0.25 + 0.75 \cdot \log_2 0.75) \quad \text{bits}$$

$$H = \frac{T}{\tau}(0.5 + 0.3113) = 0.8113\ \frac{T}{\tau} \quad \text{bits} \tag{1.7}$$

Hence, when there is a greater probability that an expected event will occur, there is less information. As $p$ approaches 1 and $q$ approaches 0, the near certainty of event with probability $p$ will give 0 information. The maximum information occurs when $p = q = 0.5$.

This scenario can be generalized for $n$ signal levels in a given signal interval $\tau$. Assume that each of these $n$ signal levels, $s_i$, has a probability of occurrence of $P_i$ where

$$P_1 + P_2 + \cdots + P_n = \sum P_i = 1 \tag{1.8}$$

Assume further that the probability of a finding a given signal level is independent of the value of the adjacent signal levels. The total information in $T/\tau$ intervals or in T seconds is

$$H = -\frac{T}{\tau}\sum_i^n P_i \log_2 P_i \quad \text{bits} \tag{1.9}$$

The capacity required to transmit this amount of information is then

$$C = -\frac{1}{\tau}\sum_i^n P_i \log_2 P_i \quad \text{bits/s} \tag{1.10}$$

In the case where each level is equally likely, $P_1 = P_2 = P_3 = \cdots P_n = 1/n$, then for the $n$ level signal,

$$H = -\frac{T}{\tau}\sum_i^n P_i \log_2 P_i = \frac{T}{\tau}\log_2 n \quad \text{bits} \tag{1.11}$$

More detail on information transmission can be found in specialized texts; a short introduction is given by Schwartz [1]. In general, this study of radio frequency design, the primary focus will be on fundamental hardware design used in transmitters and receivers. Other topics that are of great interest to communication engineers such as programming digital signal processing chips, modulation schemes, and electromagnetic propagation problems are more fully explored in

specialized texts in those areas. In this book these areas will be referred to only as needed in illustrations of how systems can be implemented.

## PROBLEMS

**1.1** A pulse train is being transmitted through a channel at the maximum channel capacity of $25 \cdot 10^3$ bits/s. The pulse train has 16 levels.

**(a)** What is the pulse width?

**(b)** The pulse width is doubled and sent back on the same channel. What is number of levels required?

## REFERENCES

1. M. Schwartz, *Information Transmission, Modulation, and Noise*, 3rd ed., New York: McGraw-Hill, 1980, Ch. 1.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER TWO

# Resistors, Capacitors, and Inductors

## 2.1 INTRODUCTION

At radio frequencies passive circuit elements must be considered more carefully than in lower-frequency designs. The simple resistor, capacitor, or inductor cannot be counted on to provide pure resistance, capacitance, or inductance in high-frequency circuits. Usually the "lumped" element is best modeled as a combination of these pure elements. In addition, when the size of the element becomes larger than 0.1 wavelength in the circuit medium, the equivalent circuit should include the transmission lines.

## 2.2 RESISTORS

Integrated circuit resistors can be classified into three groups: (1) semiconductor films, (2) deposited metal films, and (3) cermets (a mixture of metal and dielectric materials). Of these, only the first two have found widespread use in high-frequency circuits. Semiconductor films can be fabricated by diffusion into a host semi-insulating substrate, by depositing a polysilcon layer, or by ion implantation of impurities into a prescribed region. Polysilcon, or polycrystalline silicon, consists of many small submicron crystals of silicone with random orientations.

### 2.2.1 Resistor Types

The resistance value of an integrated circuit resistor depends on the conductivity of the channel through which the current is flowing. In the diffused resistors in a semiconductor substrate, the conductivity is a function of the doping concentration

and the carrier mobility. The conductivity is

$$\sigma = q(\mu_n n + \mu_p p) \tag{2.1}$$

It is usually expressed in the units of $(\Omega - \text{cm})^{-1}$. In this expression, $q$ is the electronic charge $(1.602 \cdot 10^{-19}$ C), $\mu_n$ and $\mu_p$ are the electron and hole mobilities ($\text{cm}^2/\text{V} - \text{s}$), and $n$ and $p$ are the number of free electrons and holes, respectively, that are available for conduction ($\text{cm}^{-3}$). At room temperature it may be assumed that all the impurity atoms in the semiconductor are ionized. This means that for an $n$-type semiconductor, the number of available electrons is equal to the donor impurity concentration:

$$n_n \approx N_\text{D} \tag{2.2}$$

Similarly, for a $p$-type semiconductor, the number of holes equals the acceptor impurity concentration:

$$p_p \approx N_\text{A} \tag{2.3}$$

In either an $n$-type or a $p$-type semiconductor, the relationship between the electron and hole concentrations is

$$n\,p = n_i^2 \tag{2.4}$$

where $n_i = 1.45 \cdot 10^{10}$ $\text{cm}^{-3}$ for silicon and $9.0 \cdot 10^6$ for gallium arsenide. This is called the *mass action law*. Thus, for an $n$-type semiconductor, the conductivity is

$$\sigma = q\left(\mu_n N_\text{D} + \mu_p \frac{n_i^2}{N_\text{D}}\right) \approx q\mu_n N_\text{D} \tag{2.5}$$

Typically, in integrated circuits, $n$-channel FETs and NPN bipolar transistors are preferred because of the much larger electron mobility over that of the hole mobility. The total number of processing steps required in a circuit design often dictates the choice of resistor channel type.

Ideally the diffused resistor with conductivity $\sigma$ can be represented by the rectangular block shown in Fig. 2.1. The resistance of the rectangular block is

$$R = \frac{L}{\sigma W T} \tag{2.6}$$

It is often convenient to separate the "processing" aspects from the "layout" aspects of the resistor. This is done by defining the sheet resistance in $(\Omega/\square)$ as

$$R_\square = \frac{1}{\sigma T} \tag{2.7}$$

so that the total resistance is

$$R = R_\square \frac{L}{W} \tag{2.8}$$

The length to width ratio determines the resistance value once the conductivity and layer thickness is set.

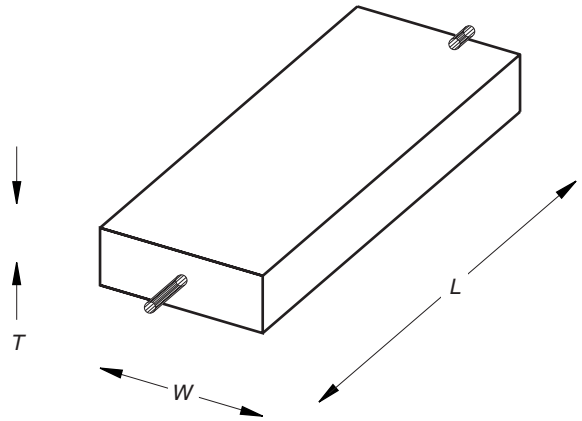



**FIGURE 2.1** Diffused resistor of length $L$, width $W$, height $T$.

**TABLE 2.1 Resistor Materials**

| Resistor Type | Resistance | Temperature Coefficient | Voltage Coefficient |
|---|---|---|---|
| Diffused Si | 10 to 100 Ω/□ | 1500 ppm/°C | 200 ppm/V |
| Diffused GaAs | 300 to 400 Ω/□ | 3000 to 3200 ppm/°C | — |
| Polysilcon | 30 to 200 Ω/□ | 1500 ppm/°C | 100 ppm/V |
| Ion implantation | 0.5 to 2 kΩ/□ | 400 ppm/°C | 800 ppm/V |
| AuGeNi (Alloyed) | 2 Ω/□ | — | — |
| Thin film Cr | 13 μΩ-cm | 3000 ppm/°C | — |
| Thin film Ti | 55 to 135 μΩ-cm | — | — |
| Thin film Ta | 180 to 220 μΩ-cm | — | — |
| Thin film TaN | 280 μΩ-cm | — | — |
| Thin film Ni | 7 μΩ-cm | — | — |
| Thin film NiCr | 60 to 600 μΩ-cm | — | — |

*Sources*: Refs. [1–3].

Metal films are made by evaporation of the desired metal on substrate and the desired pattern determined by photo lithography. Metal films are generally superior to the semiconductor films in that metal films are less sensitive to changes in temperature and voltage. Table 2.1 shows some of the main properties of a variety of methods and materials. The temperature and voltage coefficients are measures of the percentage change in resistance as a function of a change in a given parameter. The definition of temperature coefficient is $(dR/dT)/R$ and the voltage coefficient is $(dR/dV)/R$.

### 2.2.2 Resistance Determination from Layout

The layout shape of a resistor is typically simply a straight rectangular bar, as shown in Fig. 2.1. However, it may at times be better to try different shapes

in order to optimize the overall layout of a circuit. A convenient method for determining the resistance between two points on any shape is the method of curvilinear squares. Of course computer-based numerical methods such as the finite-element technique, can also be used. However, using paper and pencil, in just 20 minutes an answer could be obtained to within 10% to 20% accuracy.

A curvilinear rectangle may be defined "as any area which is bounded on opposite sides by two flux lines, and on the other sides by two equipotential lines. . ." [4]. These rectangles can be divided and subdivided into squares of ever-decreasing size. Then, based on Eq. (2.8), the total resistance can be found by counting the squares.

Rather than estimating the "squareness" of a curvilinear square, circles can be drawn between two flow lines using a compass or a template. Each curvilinear square should have its four sides tangent to the inscribed circle.

The curvilinear square method is illustrated in Fig. 2.2. The procedure takes the following form:

1. Draw flow lines between the two electrodes as if water is to travel between the electrodes in a laminar flow. The spacing between the two flow lines is less important than the shape of the flow lines. The flow lines should intersect the electrodes at right angles.

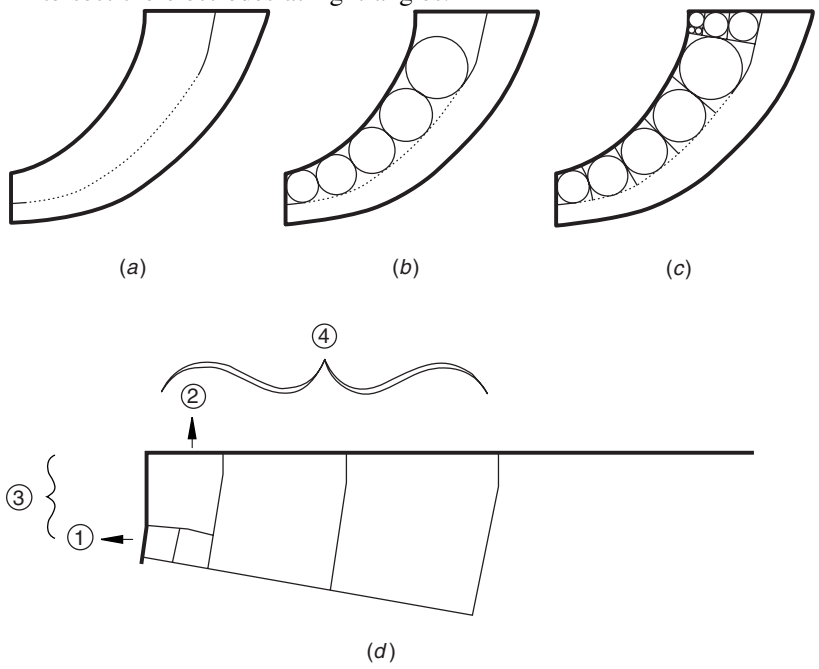



**FIGURE 2.2** (*a*) Resistor shape with a flow line; (*b*) addition of tangential circles; (*c*) drawing best-fit curvilinear squares; (*d*) expansion of the fractional curvilinear square from (*c*).

2. Between two adjacent flow lines, draw a series of circles tangent to the flow lines and to each other.
3. Draw equipotential lines between the circles orthogonal to the flow lines.
4. If there is more rectangle left over than the number of circles, fill the remaining part of the rectangle with circles in the orthogonal direction. Continue this until the last rectangle is sufficiently close to becoming a square.
5. Starting with the smallest square, count all the squares in series. Invert and add to the next largest row of squares going in the orthogonal direction. Continue inverting and adding to the next larger row of squares.

As Fig. 2.2 shows, the first step, in which the smallest squares are added, has the result 2. Step 2 consists in inverting the result of step 1 and adding the remaining series of squares, with the result $\frac{1}{2} + 1 = 1.5$. In step 3 the result of step 2 is inverted and added to the remaining series of squares. At the end of this step, the result is $(1/1.5) + 2 = 2.67$. Finally step 4 gives $(1/2.67) + 5 = 5.375$. The resistance then in the indicated section of the resistor is $5.375 \cdot R_{\square}$. These steps are repeated for the other parallel sections to obtain the total resistance as a parallel combination.

The obvious application of this method to electrical engineers is in finding the resistance of an arbitrarily shaped resistor. However, it can also be applied in finding the magnetic reluctance in a magnetic circuit, capacitance, heat convection, and, of course, laminar fluid flow.

There are a couple of other details that should be considered in predicting resistance values. One is that the rectangular bars of resistance are not really rectangular bars. The bottom is rounded, and a better estimate can be found by taking this characterstic into account. Another complication is that somewhere a semiconductor diffused resistor is going to have to come in contact with a metal. The resulting Schottky barrier can cause an additional voltage drop. Normally an Ohmic contact is used for this interface. An Ohmic contact is formed by heavily doping the semiconductor at the point of contact with the metal. This essentially promotes tunneling of electrons through the barrier. Nevertheless, there is still some residual resistance from the contact. Consequently the previously given expression for resistance, Eq. (2.8), should be modified to incorporate the contact resistance, $R_c$:

$$R = R_{\square}\frac{L}{W} + \frac{2R_c}{W} \tag{2.9}$$

A typical value for $R_c$ is about 0.25 Ω-mm.

Active loads are often used in integrated circuits in place of passive loads where the required resistance value is fairly high. The primary advantage of the active load is its compact size relative to that of a large passive load. These are often used in common emitter NPN transistor amplifiers or FET amplifiers as shown in Fig. 2.3. As the figure shows, the base-collector, the gate-drain of the enhancement mode FET, and the gate-source of the depletion mode FET are

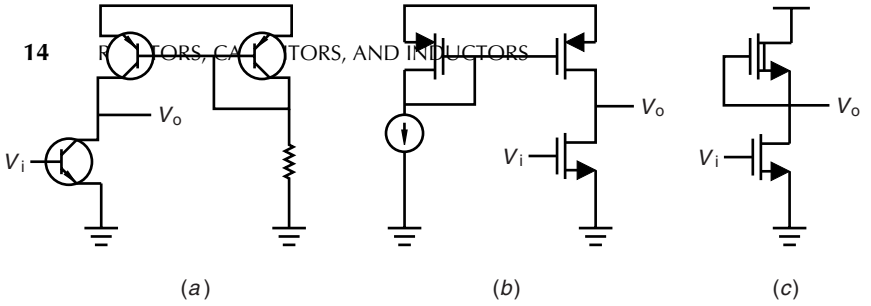



**FIGURE 2.3** Active loads using (*a*) common emitter structure, (*b*) *p*-channel enhancement mode MOSFET load, and (*c*) *n*-channel depletion mode MOSFET load.

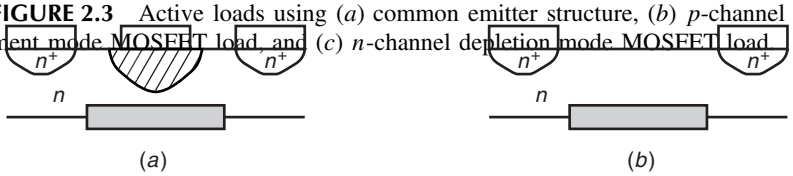



**FIGURE 2.4** Charge distribution for (*a*) shorted gate active load and (*b*) a saturated resistor.

shorted together. An active load can also be made in GaAs with a "saturated resistor" [5]. This structure is essentially a GaAs MESFET without a gate, and it is simpler to construct than the usual depletion mode FET with the gate shorted to the source. The saturation current in GaAs is reached at a rather low saturation field of 3 kV/cm. This means that once saturation has occurred, there is a small increase in the current with each increase in voltage. Consequently a large effective resistance is obtained. The saturated resistor channel depth is effectively greater than that of the MESFET channel as shown in Fig. 2.4. Consequently, for a given resistance value, the width of the saturated resistor would have to be made narrower. Resistance values of 8 to 10 kΩ have been obtained [5]. However, the simpler processing of the saturated resistor has given improved reliability and repeatability of these devices.

## 2.3 CAPACITORS

Some of the most important parameters that need consideration in choosing a capacitance are (1) the capacitance value, (2) capacitance value tolerance, (3) loss or $Q$, (4) temperature stability, (5) mechanical packaging and size, and (6) parasitic inductance. These criteria are interdependent, so often the appropriate compromises depend on the constraints imposed by the particular application. This section will consider both hybrid and monolithic capacitor designs.

### 2.3.1 Hybrid Capacitors

Hybrid capacitors are available in both single-layer capacitors for high-frequency low-capacitance applications and multi-layer capacitors for higher capacitance. Even for multilayer chip capacitors, the self-resonant frequency for a 0.1 pF capacitor is over 10 GHz and for a 1000 pF capacitor the self-resonant frequency of 250 MHz. These capacitors can be attached to hybrid circuit boards to provide high available capacitances with relatively low loss. Unlike low-frequency circuits, certain parasitic circuit elements must be accommodated in the overall design. The parasitic inductance is affected by the packaging, since it is usually associated with the lead attachments to the capacitor and line length effects inside the capacitor. In low-frequency circuits the effect of the inductance is so small that it can safely be neglected. However, at radio frequencies both the inductance and the metal losses often become significant. Consequently the equivalent circuit for a chip capacitor as developed by chip capacitor manufactures is shown in Fig. 2.5 and can sometimes be simplified as simply a series RLC circuit. The additional parallel resistance, $R_p$ is added to this equivalent circuit to model resistive losses caused by dielectric loss. This parameter is the main loss at low frequencies in the hertz to kilohertz range, but at RF it becomes negligible when compared to $R_s$. The impedance of the circuit is

$$Z = R_s + \frac{j}{\omega C}(\omega^2 LC - 1) \tag{2.10}$$

Consequently the effective capacitance is frequency dependent:

$$C_{eff} = \frac{C}{1 - (\omega/\omega_0)^2} \tag{2.11}$$

where $\omega_0 = 1/\sqrt{LC}$ is the self-resonant frequency.

While loss in capacitors is usually less than that in inductors, capacitor loss can still be significant in circuit performance. Loss can be described in terms of dissipation factor (DF), loss tangent ($\tan\delta$), the equivalent series resistance ($R_s$), $R_p$

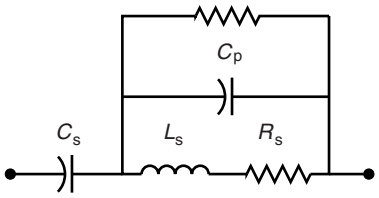



**FIGURE 2.5** Typical equivalent circuit for a chip capacitor.

and $Q_{cap}$. Since the circuit $Q$ is assumed to result from a series RLC configuration,

$$R_s = \frac{X_c}{Q_{cap}} \tag{2.12}$$

The loss terms then are related by

$$\tan\delta = \mathrm{DF} = \frac{1}{Q_{cap}} \tag{2.13}$$

The angle $\delta = 90^\circ - \theta$, where $\theta$ is the angle between the voltage and the current. In a lossless capacitor, $\theta = 90^\circ$.

In a capacitor the dielectric is the primary source of loss. An RF field can cause the dipole molecules in the dielectric to rotate at a rate proportional to the applied frequency and with a force proportional to electric field strength. The rotation of these molecular dipoles is converted to heat loss. When $E$ is the electric field and $f$ is the frequency, the energy dissipation is given by the following empirical expression [6]:

$$\mathcal{E} = E^2 f 55.5 \cdot 10^{-6} \varepsilon_r \tan\delta \quad \mathrm{W/cm^3} \tag{2.14}$$

Some of the most widely used dielectric materials for capacitors are shown in Table 2.2

The $BaTiO_3$, $\varepsilon_r = 8000$, material provides the most compact capacitor. However, it has relatively poor temperature coefficient, $\tan\delta$ shift with voltage, coefficient of expansion in terms of temperature, piezoelectric effects, and aging qualities because of its porosity.

The $BaTiO_3$, $\varepsilon_r = 1200$, capacitance varies by $+15\%$ from $-55^\circ$C to $125^\circ$C. When the $BaTiO_3$ materials are heated to about the Curie point, the value for $\varepsilon_r$ jumps up about 10% to 15%. After cooling and waiting 10 hours, the dielectric constant drops back down only 3% of its peak value, and after 10,000 hours, it drops down only 7% of its peak value. As the voltage changes over a range of 30 V, the loss tangent increases from 0.01 to 0.1 at low frequencies. There are four crystalline phases for $BaTiO_3$ as it is heated up. The crystal changes from orthorhombic to tetragonal to cubic (which is near the Curie point). At each of

**TABLE 2.2 Loss Tangent ($\tan\delta$) of Dielectric Materials**

| Dielectric | At Low Frequency | At 100 MHz | At 1 GHz |
|---|---|---|---|
| $BaTiO_3$, $\varepsilon_r = 8000$ | — | 0.1 | — |
| $BaTiO_3$, $\varepsilon_r = 1200$ | 0.01 | 0.03 | 0.10 |
| Ceramic (NPO), $\varepsilon_r = 30$ | 0.0001 | 0.002 | 0.10 |
| Alumina, $\varepsilon_r = 9.8$ | — | 0.0005 | — |
| Porcelain (ATC100), $\varepsilon_r = 15$ | — | 0.00007 | — |

these changes, there is an abrupt change in the mechanical size of the crystal [7]. This has deleterious implications on solder joints of the capacitor.

The capacitance using NPO material varies with temperature ±30 ppm/°C. It moves in the negative direction, then in the positive direction exceeding the initial capacitance, and finally settling down near the original capacitance as the temperature rises. Hence it gets the name NPO.

The porcelain materials provide high $Q$, no piezoelectric effects, no aging effects (since it is not a porous material), and temperature coefficient of ± 30 ppm/°C up to 125°C. The coefficient of expansion of the porcelain capacitor is the same as alumina ($Al_2O_3$). For this reason, when mounted on an alumina substrate, the two will expand the same amount. The series resistance at 1 GHz varies with the value of capacitance as shown in Table 2.3.

For a 30-pF $BaTiO_3$, $\varepsilon_r = 1200$, capacitor operating at 300 MHz, the resistance can be as high as 1 Ω and result in 0.3-to 3-dB dissipation loss. In solid state circuits that operate in high-current and low-voltage conditions, these losses can be quite significant. The generated heat further degrades the loss tangent, which increases the heat dissipation. Thermal runaway can occur, causing self-destruction. Of the materials shown in Table 2.2, porcelain provides the best loss tangent, especially at frequencies in the 1 to 3 GHz range.

The frequency range of a chip capacitor can be extended by the simple expedient of turning it on its side (Fig. 2.6). Resonances appear to be the result of different path lengths of the path through the lower plates and upper plates of a multi-layer capacitor. Turning the capacitor on its side tends to equalize the path lengths and eliminates all odd-order harmonic resonances [7].

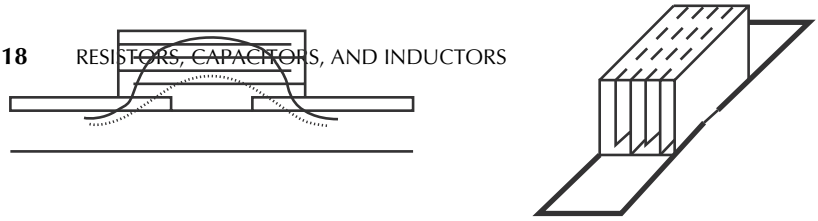
### 2.3.2 Monolithic Capacitors

Capacitors in monolithic circuits are best avoided where possible because of the amount of real estate they occupy. Nevertheless, they are sometimes required. The capacitance tolerance is typically ±10%, and capacitance values range from 0.2 to 100 pF. There are four types of monolithic capacitors that might be used in

**TABLE 2.3 Resistance of Porcelain Capacitors**

| $C$, pF | $R_s$, Ω |
|---|---|
| 5.6 | 0.38 |
| 10 | 0.27 |
| 20 | 0.19 |
| 30 | 0.16 |
| 40 | 0.13 |
| 50 | 0.12 |
| 100 | 0.088 |

*Source*: Ref. [7].



(*a*) (*b*)

**FIGURE 2.6** Metallic conductors in (*a*) horizontal and (*b*) vertical orientation.

monolithic circuit designs: (1) open circuit stub, (2) interdigital line, (3) metal-insulator metal, and (4) varactor diode.

The open circuit stub capacitance is simply an open circuit transmission line whose length is $< \lambda/4$. The capacitive susceptance is obtained from the transmission line equation:

$$B = Y_0 \tan\left(\frac{\omega l}{v_c}\right) \tag{2.15}$$

The value of the susceptance depends on the characteristic admittance, $Y_0$, of the transmission line, the length, $l$, of the transmission line, and the substrate material that governs the velocity of the wave, $v_c$. This open circuit stub provides a shunt capacitance to ground. While the susceptance is not proportional to $\omega$ as in lumped capacitors, it is a good approximation when the argument of the tangent function is small. Line lengths can use a large amount of real estate at low frequencies, so typically the open stub capacitor is most useful at frequencies greater than about 8 GHz.

The interdigital capacitor shown in Fig. 2.7, unlike the open stub, provides series capacitance. It is most useful for capacitances less than 1 pF, and at 12 to 14 GHz it typically has a $Q$ of 35 to 50. The equivalent circuit shown in Fig. 2.7 includes series resistance and inductance, as well as some shunt capacitance to ground. The latter is caused by the metal-insulator-ground metal of the microstrip structure. The main series capacitance can be estimated from

$$C = (N_f - 1)C_g\ell \tag{2.16}$$

where $N_f$ is the number of fingers, $\ell$ is the finger length, and $C_g$ is the static gap capacitance between the fingers.

A third type of capacitor is the metal-insulator-metal capacitor (Fig. 2.8). Of the four monolithic capacitors, this is the most popular and is the most obvious. The dielectric thickness typically used is 0.1 to 0.4 μm. Losses can be reduced if the metal thickness is greater than two skin depths. The metal surface roughness should be as smooth as possible to reduce losses and avoid pin holes in the dielectric. Typically the capacitance ranges from 50 to 300 pF/mm$^2$ [2]. When

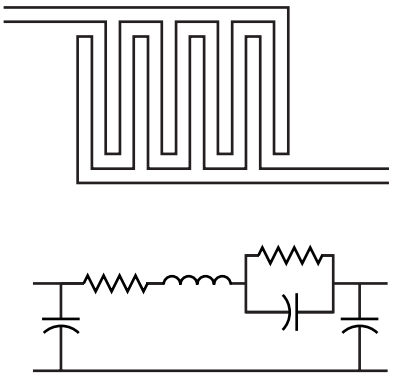


**FIGURE 2.7** Interdigital capacitor layout and equivalent circuit.

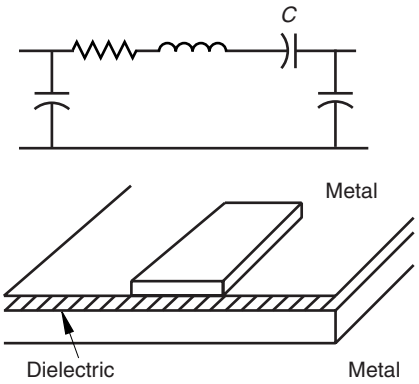



**FIGURE 2.8** Metal-insulator-metal capacitor and equivalent circuit.

the conductor losses prevail over the dielectric losses, the conductor quality factor is [1]

$$Q_c = \frac{3}{2\omega R_s (C/A) l^2} \tag{2.17}$$

In this expression $R_s$ is the surface skin resistivity, $C$ is the capacitance, $A$ is the plate area, and $l$ is length of the plate in the direction the microwave current enters the plate. If the dielectric quality factor is

$$Q_d = \frac{1}{\tan \delta} \tag{2.18}$$

then the total $Q$ is

$$\frac{1}{Q_T} = \frac{1}{Q_d} + \frac{1}{Q_c} \tag{2.19}$$

**TABLE 2.4 Monolithic Capacitor Dielectric Materials**

| Dielectric | Nominal $\varepsilon_r$ | Range of $\varepsilon_r$ | Temperature Coefficient ppm/°C |
|---|---|---|---|
| $SiO_2$ | 5 | 4–5 | 50–100 |
| SiN | 7.5 | 5.5–7.5 | 25–35 |
| $Ta_2O$ | 21 | 20–25 | 200–400 |
| $Al_2O_3$ | 9 | 6–10 | 100–500 |
| Polyimide | 3.5 | 3–4.5 | −500 |

*Source*: Ref. [1].

The dielectric films used in monolithic capacitors tend to be much higher than that obtained in the hybrid capacitors described above. Some typical metal-insulator-metal dielectric materials are shown in Table 2.4. The variableness in the dielectric constant is a result of the variation in deposition methods, uniformity, and thickness.

The fourth way of obtaining capacitance is by means of the junction capacitance of a Schottky diode. This capacitance is

$$C = \frac{C_0}{(1 - V/\phi)^{\gamma}} \tag{2.20}$$

where $\gamma \approx 1/2$ [8, p. 190]. When the applied voltage, $V$, is zero, the capacitance is $C_0$. A major disadvantage of this capacitance is its voltage dependence relative to the built-in potential, $\phi$.

## 2.4 INDUCTORS

Inductors operating at radio frequencies have a variety of practical limitations that require special attention. A tightly wound coil in addition to providing a self inductance also has heat loss due to the nonzero wire resistance, skin effect losses, eddy current losses, and hysteresis losses when a magnetic material is used. Furthermore two conductors close together at two different voltages will also exhibit an interelectrode capacitance. At radio frequencies these effects cannot be neglected as easily as they could at lower frequencies. The equivalent circuit is shown in Fig. 2.9. In this figure the series resistance, $R_s$, represents the conductor loss as well as the skin effect losses. The parallel resistance, $R_p$, represents the effect of eddy current losses and the hysteresis loss in magnetic materials when present. The shunt capacitance, $C$, is the capacitance found between the coils. Straightforward circuit analysis gives the impedance for this equivalent circuit:

$$Z = \frac{R_p R_s + R_p L_s}{s^2 L C R_p + s(R_s C R_p + L) + R_s R_p} \tag{2.21}$$

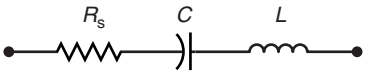



**FIGURE 2.9** Simple equivalent circuit for an inductor.

If $R_p$ is considered so large as to have negligible effect, and if the remaining series circuit $Q = 1/\omega R_s C$ is large, then the effective inductance is approximately

$$L_{\text{eff}} = \frac{L}{1 - \omega^2 LC} \tag{2.22}$$

and the effective resistance is

$$R_{\text{eff}} = \frac{R_s}{1 - \omega^2 LC} \tag{2.23}$$

Clearly, the presence of the capacitance dramatically increases the effective inductance and capacitance near the self-resonant frequency of the inductor. The self-capacitance of the inductor is a function of the coil length to coil diameter ratio, $\ell/D$, and it has an optimum value [9,10]. The following sections will describe in greater detail the origin of the parasitic circuit elements for a practical RF inductor and some design methods for RF inductors.

### 2.4.1 Resistance Losses

The dc current flowing through a wire with a cross-sectional area, $A$, will encounter twice the resistance if the area is doubled. At radio frequencies the ac current tends to flow near the surface of the conductor because of the skin effect. This can be illustrated by an electric field impinging on a conductor whose resistance is not zero. The field will penetrate into the conductor and will exponentially decay as it penetrates deeper:

$$E(x) = E_0 e^{-x/\delta} \tag{2.24}$$

where

$$\delta = \sqrt{\frac{\rho}{\pi f \mu}} \tag{2.25}$$

In this equation $f$ is the frequency, $\rho$ is the resistivity, and $\mu$ is the permeability. Because of this skin depth, the resistance of a given wire with radius $R$ will have a higher resistance at high frequencies than at dc. The ac resistance is given by [9]

$$R_{\text{AC}} = \frac{A_{\text{tot}}}{A_{\text{skin}}} R_{\text{dc}}$$

**TABLE 2.5 Common Conductors**

| Metal | Conductivity $(\Omega\text{-cm})^{-1}$ | Skin Depth (cm) |
|---|---|---|
| Brass | $1.57 \cdot 10^5$ | $12.7 \cdot f^{-1/2}$ |
| Aluminum | $3.54 \cdot 10^5$ | $8.46 \cdot f^{-1/2}$ |
| Gold | $4.27 \cdot 10^5$ | $7.7 \cdot f^{-1/2}$ |
| Copper | $5.8 \cdot 10^5$ | $6.61 \cdot f^{-1/2}$ |
| Silver | $6.14 \cdot 10^5$ | $6.61 \cdot f^{-1/2}$ |
| Mu-metal | $1.58 \cdot 10^8$ | $0.4 \cdot f^{-1/2}$ |

$$= \frac{\pi R^2}{\pi R^2 - \pi(R-\delta)^2} R_{\text{dc}}$$

$$= \left[\frac{\pi R^2}{2\pi\delta R - \pi\delta^2}\right] R_{\text{dc}} \tag{2.26}$$

At high frequencies

$$R_{\text{ac}} \approx \frac{R}{2\delta} R_{\text{dc}}. \tag{2.27}$$

The possibility for $R_{ac}$ to be infinite or even negative clearly indicates that Eq. (2.26) has gone beyond its range of applicability. The problem is that the skin depth has become greater than twice the wire radius. Listed in Table 2.5 are the resistivities and skin depths of a few common metals.

Another important loss mechanism is called the *proximity effect*. When one conductor supporting a changing magnetic field is brought close to another conductor, currents will be induced on the second conductor in conformity with Faraday's law. These currents are called *eddy currents*, and they flow in closed paths so as to produce a magnetic field that is in opposition to the originally applied external field. These currents produce *Joule heating*. This is exactly the condition that occurs in a tightly wound inductive coil. When many wires are close together, the loss problem is compounded, and the eddy current losses can be quite significant. As an illustration of this, consider a coil with length to diameter ratio of 0.7. If this coil is unwound and laid out as a straight wire, the losses will drop by a factor of 6 [9, p. 47].

### 2.4.2 Magnetic Materials

A recurring problem is the need for a large value of inductance. An obvious solution is to increase the flux density within an inductor coil with the addition of a magnetic material with high permeability $\mu_r$. Most magnetic materials introduce losses that are unacceptable at radio frequencies. A variety of ferrite materials however have been found to have low loss at radio and microwave frequencies in comparison with most other magnetic materials. The relative permeability for

ferrites is in the range $10 < \mu_r < 150$. Above the cutoff frequency, $\mu_r$ drops off quickly. The higher the permeability, the lower the cutoff frequency. Typically, for $\mu_r = 10$, $f_{\text{cutoff}} = 1$ GHz. For $\mu_r = 150$, $f_{\text{cutoff}} = 20$ MHz.

### 2.4.3 Solenoid Design up to 2 GHz [11]†

A design procedure for a single layer solenoid is given below. The computer program, SOLENOID, follows the procedure outlined here and is described in Appendix A. The given parameters for the analysis of a solenoid are the form length, number of turns ($n$), and the form diameter. The pitch is defined as

$$\text{Pitch} = \frac{\text{form length}}{n} \quad \text{inches} \tag{2.28}$$

For maximum $Q$, the wire diameter should be 0.6 to 0.7 · pitch. The wire diameter is selected from the standard wire sizes. For a given American wire gauge (AWG), the wire diameter is

$$\text{Wire diameter} = \frac{0.005}{92^{(\text{AWG}-36)/39}} \quad \text{inches} \tag{2.29}$$

Another parameter is the turn diameter. It represents the diameter where the magnetic flux is generated. As shown in Fig. 2.10, it is merely (turn diameter) = (form diameter) + (wire diameter). With these quantities now defined, the analysis recipe can be followed.

$$x = \frac{\text{turn diameter}}{\text{form length}} \tag{2.30}$$

$$K_n = \frac{1}{1 + 0.45x - 0.005x^2} \tag{2.31}$$

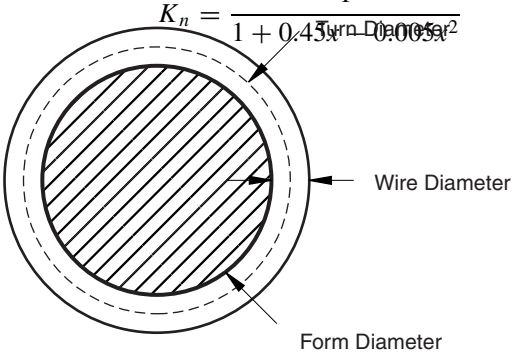



**FIGURE 2.10** Inductor form cross section.

† Reprinted with permission, *Microwave Journal*, Vol 39, pp. 70–76.

When the wire diameter = 0, the current sheet correction factor, $s$, is set to 1. When the wire diameter > 0, the $s$ correction factor is needed. The $s$ factor, and finally the inductance, is found by first finding $a$ and $b$ as shown below:

$$a = 2.3 \cdot \log_{10}\left(1.73 \cdot \frac{\text{wire diameter}}{\text{pitch}}\right) \tag{2.32}$$

$$b = 0.336\left(1 - \frac{2.5}{n} + \frac{3.8}{n^2}\right) \tag{2.33}$$

$$s = 1 - \frac{2 \cdot (\text{form length}) \cdot (a + b)}{\pi \cdot (\text{turn diameter}) \cdot n \cdot K_n} \tag{2.34}$$

$$L_0 = \frac{(n \cdot \pi \cdot \text{turn diameter})^2 \cdot 2.54 \cdot 10^{-9}}{\text{form length}} \tag{2.35}$$

$$L = L_0 \cdot K_n \cdot s \quad (H) \tag{2.36}$$

The value $L_0$ is the inductance of a closely wound coil with a flat strip (wire diameter = 0). The value $K_n$ is the Nagaoka correction factor and is used when the wire length is not much larger than the turn diameter. The value $s$ is the current sheet correction factor and is needed when there is appreciable space between wire turns. Because $L_0$ is not dimensionless, the lengths must be given in terms of inches. An example given in Ref. [11] illustrates the use of these expressions:

Pitch = 0.0267 ($n$ = 15)
Wire diameter = 0.0187 in.
Turn diameter = 0.2187 in.
$x = 0.5467$
$K_n = 0.8035$
$a = 0.1912$
$b = 0.2857$
$s = 0.954$
$L_0 = 674$ nH
$L = 516.8$ nH

A synthesis procedure is also available [11]. In this case a desired inductance is considered. Typically only a finite number of form diameters are available, so the form diameter will also be considered a given quantity. From this, the number of turns $n$ and the form length, *FLEN*, is found.

The inductance is considered to be a function of $n$ and *FLEN*, and $L_d$ is the desired inductance. An iterative procedure is followed where

$$L_1 = L(n_1, FLEN) \tag{2.37}$$

$$L_2 = L(n_2 = n_1 \pm 1, FLEN) \tag{2.38}$$

$$n_3 = n_2 - (L_2 - L_\mathrm{d})\frac{n_2 - n_1}{L_2 - L_1} \tag{2.39}$$

This iteration loop is repeated until

$$L(n, FLEN) = L_\mathrm{d} \tag{2.40}$$

The equality is obtained though with a nonintegral value for $n$. For printed circuit boards, $n$ must be an integer. While the number of form diameters is limited, the form length can be cut to any desired length. Therefore the form length is adjusted to guarantee an integral $n$. The procedure is to increase $n$ to the next higher integer value and adjust $FLEN$ by an iterative scheme much like the previous one:

$$L_1 = L(n, FLEN_1) \tag{2.41}$$

$$L_2 = L(n, FLEN_2) \tag{2.42}$$

$$FLEN_3 = FLEN_2 - (L_2 - L_\mathrm{d})\frac{FLEN_2 - FLEN_1}{L_2 - L_1} \tag{2.43}$$

This iteration loop is repeated until

$$L(n, FLEN) = L_\mathrm{d} \tag{2.44}$$

where $n$ is not an integer value.

Once $L$, $n$, and $FLEN$ are known, the $Q$ factor and the parasitic capacitance can be found using the formulas given in [11]. Using the value for $x$ given in Eq. (2.29), a value for the capacitance is determined:

$$C = (\text{turn diameter})\ \frac{(0.301468x + 0.493025)x + 0.227858}{x} \quad \text{pF} \tag{2.45}$$

The coil resonant frequency is then simply

$$f_r = \frac{1}{2\pi\sqrt{LC}} \tag{2.46}$$

The value for $Q$ is found from the empirical relationship for two cases where the turn diameter is in inches and $f$ is in MHz:

$$A = \begin{cases} [(58.6355x - 171.154)x + 200.674]\,x + 0.089708, & 0.2 < x < 1 \\ [(0.751186x - 9.49018)x + 42.506]\,x + 68.1191, & 1 < x < 5 \end{cases} \tag{2.47}$$

The value for $Q$ is then obtained from the two-step formula below:

$$Q_0 = A \cdot (\text{turn diameter}) \sqrt{f} \tag{2.48}$$

$$Q = Q_0 \left[ 1 - \left( \frac{f}{f_r} \right)^2 \right] \tag{2.49}$$

The procedure described above has been put in the form of the computer algorithm, SOLENOID. An example of the design of 100 nH inductor is found in Appendix A using this program.

### 2.4.4 Monolithic Spiral Inductors

Lumped monolithic inductors have been used in circuit designs as tuned loads for amplifiers, filters to reduce out of band signals and noise, and as a means of enhancing stage gain by tuning out device or parasitic capacitances at the center frequency. Planar inductors have been implemented in practical systems for many years using a variety of different substrates. They were examined early in the development of silicon integrated circuits, but were abandoned because of process limitations and losses in the series resistance and substrate that effectively reduced their operating frequency. Now, however, technological improvements have made them available for mobile communications systems.

Small inductances in the nano Henry range can be fabricated using printed circuit techniques. These have typically been done in either a rectangular or circular spiral shape. Both are widely used, but the circular spiral design seems to provide greater inductance per unit area of real estate. In determining the inductance in either case, the self-inductance of the structure must be supplemented by the mutual inductance of neighboring turns as well as the mutual inductance of its mirror image associated with the ground plane for microstrip. There are, in addition, capacitances between turns of the spiral and capacitances to the ground. These capacitances are calculated from coupled microstrip line theory. A numerical implementation of the rectangular inductance based on [12] is provided. This paper is heavily cited and does provide a comparison between the predicted and the measured inductance.

A comparison is made between square and circular inductors in [13] in which it is stated that square spirals provide less inductance than circular spirals for equivalent sized diameters, although the data seem ambigious. This paper uses a simple lumped element equivalent circuit consisting of a series $R$–$L$ circuit with shunt capacitances on either side to represent a single turn. However, the entire inductor is treated as a distributed circuit.

A design of a square inductor is described in [14], which is modeled like the one in [13] except that an additional resistance is added in series to the parallel capacitors to ground. A comparison is made with measured data, and the design is incorporated into a low-pass filter design.

In an effort to increase the desired $Q$ for an inductor, the ground plane under the square spiral is removed in [15]. Excellent agreement is obtained up to 5 GHz.

An extensive study of over 100 inductors was made in [16]. Comparisons were made using square, octagonal, and circular spirals. Empirically determined equivalent circuits were obtained based on the measured data. The basic conclusions were that the resistance of the circular or octagonal shaped spiral is 10% lower than that of a square spiral inductor for the same inductance. Furthermore it is better to maximize line spacing rather than maximizing line width to achieve high $Q$.

The capacitance itself becomes a major part of the inductor model. An effort was made in [17] to predict the distributed capacitances of circular spiral inductors by means of a Green's function analysis. Good agreement between predicited and measured values was obtained.

An actual equivalent circuit model for a spiral inductor was obtained in [18]; the computer program is posted on the web.† The "circular" spiral is a $p$-sided polygon with $n$ turns, with a total of $n \cdot p$ sections. Each section is modeled as shown in Fig. 2.11. The analysis includes the effects of the internal impedance of each section as well as the magnetic and electric coupling to neighboring segments and to the substrate. The primary advantages of using this analysis tool is the speed of computation (unlike the three-dimensional field simulator), optimization, and the ability to analyze spiral transformers as well as inductors with various metalizations and shapes. The geometrical shape of the inductor depends on the area of the spiral, metal width, metal spacing, the number of turns, and frequency of operation. The appropriate choice for these parameters are aided by the above-mentioned program.

An alternate approach approximates an $n$ turn circular spiral as a set of $n$ concentric circular microstrips (Fig. 2.12). Each of these circular microstrips are modeled by an equivalent circuit shown in Fig. 2.11, where $R_L = R_q = 0$ [13]. The total equivalent circuit of the circular spiral is simply the cascade of each of the circular sections. The series resistance, $R$, represents the resistive loss in the conductor. The resistance is proportional to $\sqrt{f}$ because of the skin effect. The capacitances $C_{q1}$ and $C_{q2}$ are the capacitances to the ground plane, and $C_L$ is the total coupling capacitance between neighboring turns. What is lacking here,

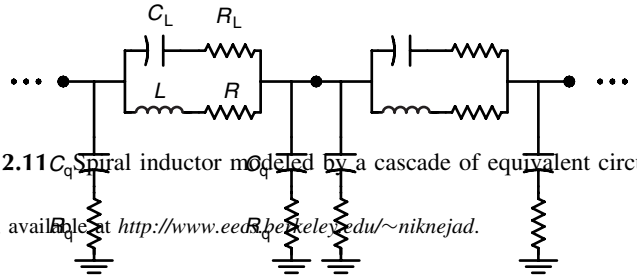



**FIGURE 2.11** Spiral inductor modeled by a cascade of equivalent circuit sections [19].

† ASITIC; available at *http://www.eecs.berkeley.edu/~niknejad*.

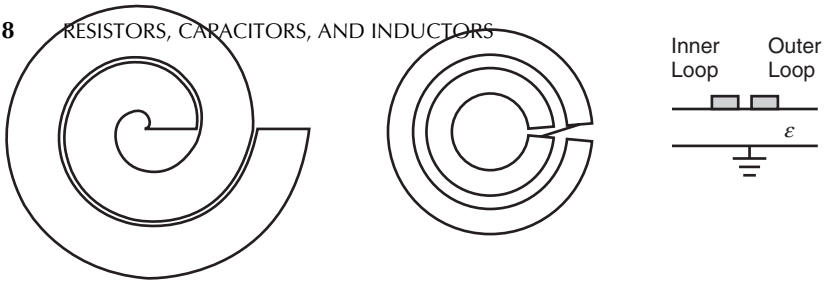



**FIGURE 2.12** Spiral approximated by concentric circular coupled lines.

but considered in [18], is the mutual coupling that occurs for line segments that are not parallel. What is helpful, though, are the closed form equations given in [19] for various types of inductor elements.

The capacitances are determined from coupled line theory in which each line can be excited with the same voltage (even mode) or equal but opposite signed voltages (odd mode). The actual capacitance is a linear combination of the even and odd mode capacitances. The percentage of the even-mode and odd-mode capacitances between two adjacent turns of the spiral may be found based on the following approximations: For a given pair of adjacent concentric circular lines, assume there is a small break between the excitation of one loop and its end, 360° later (Fig. 2.12). There is a small connection from this point to the next loop. If the wave on the inner loop is excited by a 1 volt source $1 \cdot e^{j0^\circ}$, then by the time it reaches the other end of the inner loop, the voltage is $1 \cdot e^{j\Delta\phi}$, where $\Delta\phi$ is the electrical length (circumference) of the inner loop. The outer loop is then excited by the voltage $1 \cdot e^{j\Delta\phi}$. Consequently there is a voltage difference between the inner loop and the outer loop. The percentage of even-mode and odd-mode voltages between the two loops is a function of $\Delta\phi$.

For the purpose of estimating the value for $\Delta\phi$, the circumferences of the two circles will be assumed to be the average of the two circles:

$$\Delta\phi = 2\pi \frac{r_1 + r_2}{2} \beta = 2\pi\beta r_{\text{avg}} \tag{2.50}$$

where $\beta$ is the propagation factor of the line in the given media. If $v_1$ is the voltage at a certain position of the first loop and $v_2$ is the voltage on the second loop adjacent to $v_1$, then the corresponding even- and odd-mode voltages are

$$v_{\text{e}} = \tfrac{1}{2}(v_1 + v_2) \tag{2.51}$$

$$v_{\text{o}} = (v_1 - v_2) \tag{2.52}$$

No information is lost in doing this, since the original voltages $v_1$ and $v_2$ are easily recovered if $v_{\text{e}}$ and $v_{\text{o}}$ are known. The percentage of even- and odd-mode

capacitances are proportional to the even- and odd-mode voltages:

$$\%C_{\mathrm{e}} = \frac{v_{\mathrm{e}}}{v_{\mathrm{e}} + v_{\mathrm{o}}} \tag{2.53}$$

$$\%C_{\mathrm{o}} = \frac{v_{\mathrm{o}}}{v_{\mathrm{e}} + v_{\mathrm{o}}} \tag{2.54}$$

In the equivalent circuit for a single turn of the spiral shown in Fig. 2.13, the percentage of even-mode excitation determines the relative amount of even- and odd-mode capacitance components. For the even mode,

$$C_{\mathrm{L}} = 0 \tag{2.55}$$

$$C_{\mathrm{q}i} = C_{\mathrm{m}} + C'_{\mathrm{fe}} + C'_{\mathrm{f}} \tag{2.56}$$

and for the odd mode,

$$C_{\mathrm{L}} = C_{\mathrm{ga}} + C_{\mathrm{ge}} \tag{2.57}$$

$$C_{\mathrm{q}i} = C_{\mathrm{m}} + C'_{\mathrm{f}} \qquad i = 1, 2 \tag{2.58}$$

In these expressions, $C_{\mathrm{ga}}$ and $C_{\mathrm{ge}}$ represent the gap capacitances between the lines through the air and through the dielectric, respectively. The capacitance, $C_{\mathrm{m}}$, represents the parallel plate capacitance between the spiral conductor and the ground plane. This is modified by the fringing capacitance, $C'_{\mathrm{f}}$, between the two coupled line conductors (which is nonzero only for the even mode excitation) and the fringing [illegible] $C_{\mathrm{f}}$, on the o[illegible]he conductors. The even-

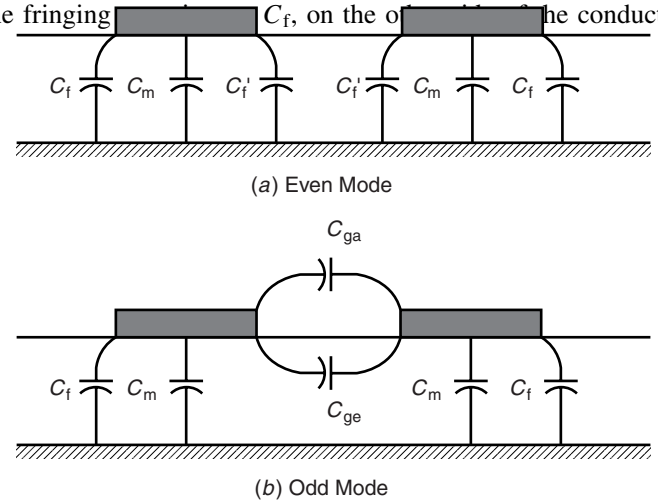



**FIGURE 2.13** Even- and odd-mode excitation of microstrip lines.

and odd-mode capacitances are added together in proportion to their even- and odd-mode voltages:

$$C_{\mathrm{L}} = 0 + \%C_{\mathrm{o}} \cdot (C_{\mathrm{ga}} + C_{\mathrm{ge}}) \tag{2.59}$$

$$C_{\mathrm{q}i} = \%C_{\mathrm{e}} \cdot (C'_{\mathrm{f}} + C_{\mathrm{m}} + C_{\mathrm{f}}) + \%C_{\mathrm{o}} \cdot (C_{\mathrm{m}} + C'_{\mathrm{f}}) \tag{2.60}$$

Detailed formulas for the circuit elements in Fig. 2.13 are found in [19] and some of these are summarized in Appendix B.

Once the equivalent $\pi$ circuit for the spiral section is known, the entire spiral inductor is modeled by cascading each of these sections. The $C_{\mathrm{L}}$, $R$, and $L$ are combined into the single impedance $Z_{\mathrm{p}}$:

$$Z_{\mathrm{p}} = \frac{R + sL}{s^2 L C_{\mathrm{L}} + sC_{\mathrm{L}} R + 1} \tag{2.61}$$

The *ABCD* parameters described in Chapter 4 (Section 4.2) are used to cascade the individual $\pi$ circuits. Thus

$$\boldsymbol{A} = 1 + sC_{\mathrm{q2}} Z_{\mathrm{p}} \tag{2.62}$$

$$\boldsymbol{B} = Z_{\mathrm{p}} \tag{2.63}$$

$$\boldsymbol{C} = s^2 C_{\mathrm{q1}} C_{\mathrm{q2}} Z_{\mathrm{p}} + s(C_{\mathrm{q1}} + C_{\mathrm{q2}}) \tag{2.64}$$

$$\boldsymbol{D} = 1 + sC_{\mathrm{q1}} Z_{\mathrm{p}} \tag{2.65}$$

Each section of the spiral described in terms of an ***ABCD*** matrix may be cascaded together by simply multiplying ***ABCD*** matrices. The $\boldsymbol{C}$ in Eq. (2.64) is a matrix element, not a capacitance. Once the total cascaded ***ABCD*** matrix is found, the input impedance may be determined:

$$Z_{\mathrm{in}} = \frac{\boldsymbol{A}Z_{\mathrm{L}} + \boldsymbol{B}}{\boldsymbol{C}Z_{\mathrm{L}} + \boldsymbol{D}} \tag{2.66}$$

The $Z_{\mathrm{L}}$ is the load impedance on the output side of the spiral. If $Z_{\mathrm{L}}$ is a short to ground, then the effective inductance of the spiral might be estimated by

$$L_{\mathrm{eff}} = \frac{\Im\{Z_{\mathrm{in}}\}}{\omega} \tag{2.67}$$

In the cascaded analysis the capacitance, $C_{\mathrm{q2}}$, from one section is the same as the $C_{\mathrm{q1}}$ of the subsequent section, and hence it ought not to be counted twice in evaluating the cascaded equivalent circuit. One approach is to simply choose $C_{\mathrm{q2}} \rightarrow C_{\mathrm{q2}}/2$ and $C_{\mathrm{q1}} \rightarrow C_{\mathrm{q1}}/2$, except of course for the innermost and outermost coupled line.

## PROBLEMS

**2.1** Calculate the resistance of a 1 m long copper wire over a frequency range of 100 MHz to 1 GHz when (a) the diameter of the wire is 31.2 mils (AWG #20), and (b) when it is 10.0 mils (AWG #30). Plot your results of ac resistance in terms of frequency.

**2.2** You are asked to determine the inductance of a solenoid when the form length is 1.5 in., the form diameter is 0.3 in., and there are 12 turns.

**(a)** What wire diameter would you choose?

**(b)** What is the inductance?

**(c)** What is the self-resonant frequency of the inductor?

## REFERENCES


1. R. A. Pucel, "Design Considerations for Monolithic Microwave Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. **MTT-29**, pp. 513–534, 1981.
2. R. Williams, *Gallium Arsenide Processing Techniques*, Boston: Artech House, 1990.
3. P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design*, New York: Oxford University. Press, 1987, p. 66.
4. A. D. Moore, *Fluid Mapper Patterns*, Ann Arbor: Overbeck, 1956.
5. C. P. Lee, B. M. Welch, and R. Zucca, "Saturated Resistor Load for GaAs Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, Vol. **MTT-30,** pp. 1007–1013, 1982.
6. V. F. Perna, Jr. "A Guide to Judging Microwave, Capacitors," *Microwaves*, Vol. 9, August, pp. 40–42, 1970.
7. V. F. Perna, Jr. "Chip Capacitor Dielectric Effects on Hybrid Microwave Amplifiers," *Proc. International Society for Hybrid Microelectronics*, October 13, 1971.
8. R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed., New York: Wiley, 1986, p. 190.
9. P. L. D. Abrie, *The Design of Impedance-Matching Networks for Radio-Frequency and Microwave Amplifiers*, Boston: Artech House, 1985.
10. R. G. Medhurst, "High Frequency Resistance and Capacity of Single-Layer Solenoids," *Wireless Engineer*, March, p. 35, 1947.
11. P. R. Geffe, "The Design of Single-Layer Solenoids for RF Filters," *Microwave J.*, Vol 39, pp. 70–76, 1996.
12. H. M. Greenhouse, "Design of Planar Rectangular Microelectronic Inductors," *IEEE Trans. Parts, Hybrids, Packaging*, Vol. 10, pp. 101–109, 1974.
13. M. Parisot, Y. Archambault, D. Pavlidis, and J. Magarshack, "Highly Accurate Design of Spiral Inductors for MMIC's with Small Size and High Cut–off Frequency Characteristics," *1984 IEEE MTT-S Digest*, pp. 106–110, 1984.
14. N. M. Nguyen and R. G. Meyer, "Si IC-Compatible Inductors and LC Passive Filters," *IEEE J. Solid State Circuits*, Vol. 25, pp. 1028–1031, 1999.
15. R. G. Arnold and D. J. Pedder, "Microwave Lines and Spiral Inductors in MCM-D Technology," *IEEE Trans. Components, Hybrids, Manufact. Tech.,* Vol. 15, pp. 1038–1043, 1992.

16. S. Chaki, S. Andoh, Y. Sasaki, N. Tanino, and O. Ishihara, "Experimental Study on Spiral Inductors," *1995 IEEE MTT-S Digest*, pp. 753–756, 1995.
17. Z. Jiang, P. S. Excell, and Z. M. Hejazi, "Calculation of Distributed Capacitances of Spiral Resonators," *IEEE Trans. Microwave Theory Tech.*, Vol. 45, pp. 139–142, 1997.
18. A. M. Niknejad and R. G. Meyer, "Analysis, Design, and Optimization of Spiral Inductors and Transformers for Si RF IC's," *IEEE J. of Solid State Circuits*, Vol. 33, pp. 1470–1481, 1998.
19. E. Pettenpaul, H. Kapusta, A. Weisgerber, H. Mampe, J. Luginsland, and I. Wolff, "CAD Models of Lumped Elements on GaAs up to 18 GHz," *IEEE Trans. Microwave Theory Tech.*, Vol. **MTT-36,** pp. 294–304, 1988.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER THREE

# Impedance Matching

## 3.1 INTRODUCTION

A major part of RF design is matching one part of a circuit to another to provide maximum power transfer between the two parts. Even antenna design can be thought of as matching free space to a transmitter or receiver. This chapter describes a few techniques that can be used to match between two real impedance levels. While some comments will be made relative to matching to a complex load, the emphasis will be on real impedance matching. The first part of this chapter will discuss the circuit quality factor, $Q$. The $Q$ factor will be used with some of the subsequent matching circuit designs.

## 3.2 THE $Q$ FACTOR

The the circuit $Q$ factor is defined as the ratio of stored to dissipated power in the following form:

$$Q = \frac{2\pi \cdot (\text{max instantaneous energy stored})}{\text{energy dissipative per cycle}} \tag{3.1}$$

For a typical parallel RLC circuit, the $Q$ becomes

$$Q = \frac{\omega C}{G} \tag{3.2}$$

where $G$ is $1/R$. For a series RLC circuit,

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

It should be emphasized that $Q$ is defined at circuit resonance. If the circuit reactance is plotted as a function of frequency, the slope of the reactance at

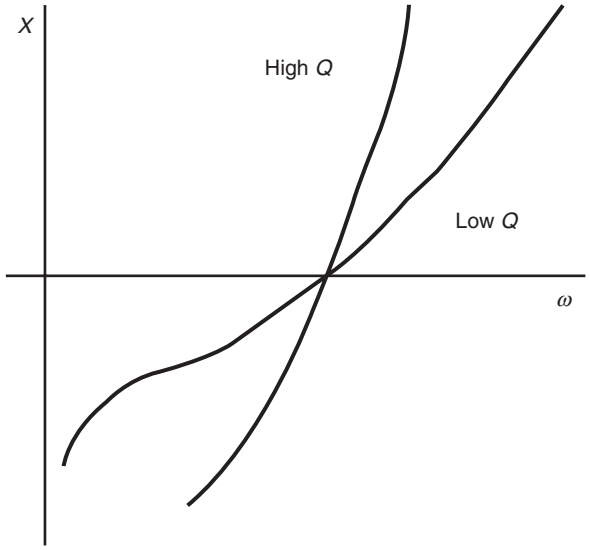



**FIGURE 3.1** Reactance slope related to $Q$.

resonance is a measure of $Q$ (Fig. 3.1). This is explicitly given as

$$Q = \frac{\omega_0}{2G}\frac{dB}{d\omega}\bigg|_{\omega_0} \tag{3.4}$$

where $B$ is the susceptance and $G$ the conductance. Alternately,

$$Q = \frac{\omega_0}{2R}\frac{dX}{d\omega}\bigg|_{\omega_0} \tag{3.5}$$

where $R$ and $X$ are the resistance and reactance of the circuit. For a series RLC circuit this latter formula will result in the solution given by Eq. (3.3). On the other hand, the $Q$ of a complicated circuit can be readily obtained from Eq. (3.4) or Eq. (3.5), even numerically if necessary.

## 3.3 RESONANCE AND BANDWIDTH

The minimum insertion loss or maximum transmission of a parallel RLC circuit occurs at the resonant frequency of the circuit. When this circuit is excited by a current source, and the output is terminated with an open circuit, the transfer function is

$$\frac{V_{\text{out}}}{I_{\text{in}}} = \frac{1}{(1/R) + j\omega C - (j/\omega L)}. \tag{3.6}$$

This is shown in Fig. 3.2. The output voltage, $V_{\text{out}}$, drops from the resonant value by $\sqrt{2}$ (or 3 dB) because the denominator of the transfer function increases from

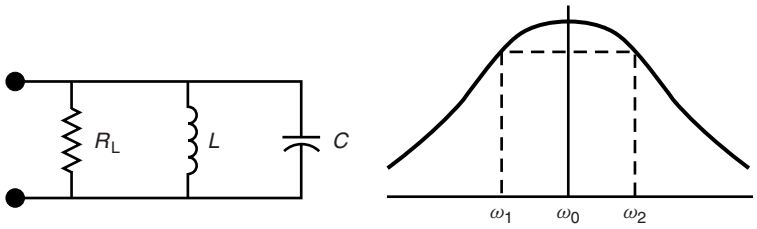



**FIGURE 3.2** Simple parallel resonant circuit.

$1/R$ at resonance to

$$\left|\frac{1}{R} + j\omega C - j\omega L\right| = \frac{\sqrt{2}}{R} \tag{3.7}$$

The resonant frequency is $\omega_0 = 1/\sqrt{LC}$, and the value of $Q$ given in Eq. (3.2) can also be written as $Q = R\sqrt{C/L}$. Using these two facts, Eq. (3.7) becomes explicitly a quadratic equation in $\omega^2$:

$$\omega^4 C^2 L^2 R^2 - \omega^2 (2CLR^2 + L^2) + R^2 = 0 \tag{3.8}$$

The two solutions for $\omega^2$ are

$$\begin{aligned}
\omega^2 &= \omega_0^2 \left\{1 + \left[\frac{1}{2Q^2} \pm \frac{1}{Q}\sqrt{1 + \frac{1}{4Q^2}}\right]\right\} \\
\omega^2 &= \omega_0^2 \left\{1 + \frac{1}{4Q^2} \pm \frac{1}{Q}\sqrt{1 + \frac{1}{4Q^2}} + \frac{1}{4Q^2}\right\} \\
\omega^2 &= \omega_0^2 \left\{\sqrt{1 + \frac{1}{4Q^2}} \pm \frac{1}{2Q}\right\} \cdot \left\{\sqrt{1 + \frac{1}{4Q^2}} \pm \frac{1}{2Q}\right\}
\end{aligned} \tag{3.9}$$

In this expression, $4R^4/\omega_0^4$ is replaced by $4L^4Q^4$. This has been written as a product of two equal terms, so the original quartic equation has two pairs of equal roots. Taking the square root of Eq. (3.9) provides the two 3 dB frequencies of the resonant circuit:

$$\omega_{1,2} = \omega_0 \left\{\sqrt{1 + \frac{1}{4Q^2}} \pm \frac{1}{2Q}\right\} \tag{3.10}$$

The 3 dB bandwidth of the resonant circuit is the difference between the two 3 dB frequencies:

$$\Delta f = \omega_2 - \omega_1 = \frac{1}{RC} \quad \text{rad/s} \tag{3.11}$$

The response is clearly not symmetrical about the resonant frequency, $\omega_0$. The resonant frequency can be found by taking the geometric mean of the two solutions of Eq. (3.10) rather than the arithmetic mean:

$$\omega_1 \cdot \omega_2 = \omega_0^2 \left[\sqrt{1 + \frac{1}{4Q^2}} - \frac{1}{2Q}\right] \cdot \left[\sqrt{1 + \frac{1}{4Q^2}} + \frac{1}{2Q}\right]$$

$$\omega_0 = \sqrt{\omega_1 \omega_2} \tag{3.12}$$

However, for narrow bandwidths, the arithmetic mean of the the two 3 dB frequencies can be used with a small error.

## 3.4 UNLOADED *Q*

In real physical reactive elements there are always some resistive losses. The loss in a capacitor or an inductor can be described in terms of its $Q$. For example, if a lossy inductor is placed in parallel with a lossless capacitor, the $Q$ of the resulting parallel circuit is said to be the circuit $Q$ of the inductor. The inductor $Q_{\text{ind}}$ then is

$$Q_{\text{ind}} = \omega_0 CR = \frac{R}{\omega_0 L} \tag{3.13}$$

or

$$R = X_{\text{L}} Q_{\text{ind}} \tag{3.14}$$

Similarly, for a lossy capacitor, its resistive component could be expressed in terms of the capacitor $Q_{\text{cap}}$. If the inductor, capacitor, and a load resistance, $R_{\text{L}}$ are placed in parallel, then the total resistance is $R_{\text{T}}$:

$$\frac{1}{R_{\text{T}}} = \frac{1}{R_{\text{L}}} + \frac{1}{Q_{\text{ind}} X_{\text{L}}} + \frac{1}{Q_{\text{cap}} X_{\text{C}}} \tag{3.15}$$

At resonance, $X_{\text{L}} = X_{\text{C}}$, so

$$\frac{X_{\text{L}}}{R_{\text{T}}} = \frac{X_{\text{L}}}{R_{\text{L}}} + \left[\frac{1}{Q_{\text{ind}}} + \frac{1}{Q_{\text{cap}}}\right] \tag{3.16}$$

The unloaded $Q$, $Q_{\text{u}}$ is the $Q$ associated with the reactive elements only (i.e., without the load). The bracketed term is the unloaded $Q$:

$$\frac{1}{Q_{\text{u}}} = \frac{1}{Q_{\text{ind}}} + \frac{1}{Q_{\text{cap}}} \tag{3.17}$$

## 3.5 *L* CIRCUIT IMPEDANCE MATCHING

There are four possible configurations that provide impedance matching with only two reactive elements. In each case the design of the matching circuits is based

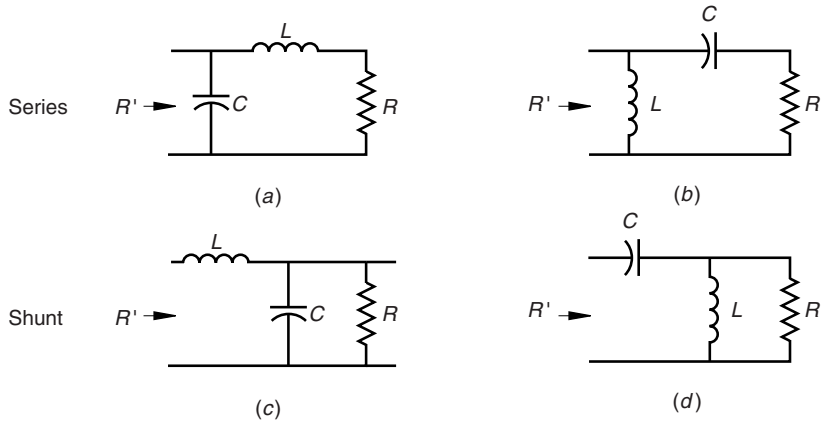



**FIGURE 3.3** Four possible *L* matching circuits.

on the *Q* factor, a concept that will become even more important in designing broadband matching circuits [1]. Two of the circuits will be described as the series connection since the reactive element closest to the load resistance is a series reactance (Fig. 3.3*a*,*b*). The circuits with a shunt reactance closest to the load resistance are called the *shunt connection* (Fig. 3.3*c*,*d*). For the series connection in which the series reactance is an inductance, the total input admittance is given as follows:

$$\begin{aligned} Y_{\text{in}} &= j\omega C + \frac{1}{R + j\omega L} \\ &= \frac{R}{R^2 + (\omega L)^2} + j\left[\omega C - \frac{\omega L}{R^2 + (\omega L)^2}\right] \end{aligned} \tag{3.18}$$

Resonance occurs when the total shunt susceptance, $jB = 0$. Thus

$$C = \frac{L}{R^2 + (\omega_0 L)^2} \tag{3.19}$$

Solution of this for the resonant frequency gives the following expression for the resonant frequency:

$$\omega_0 = \sqrt{\frac{1}{LC} - \frac{R^2}{L^2}} \tag{3.20}$$

The effect of the load resistor is to modify the resonant frequency somewhat.

The conductive part of $Y_{\text{in}}$ at this frequency (where $B = 0$) can be found. Its reciprocal is the input resistance, $R'$, of the circuit:

$$\begin{aligned} R' &= \frac{R^2 + \omega_0^2 L^2}{R} \\ &= R(1 + Q_1^2) \end{aligned} \tag{3.21}$$

The subscript 1 for $Q$ is present to emphasize this is a single resonator circuit. More complicated circuits might have several pertinent $Q$ factors to consider.

At center frequency the reactance of the series part (i.e., not the capacitance part) will change with changing frequency. Its value can be found from the input admittance expression and is the amount the reactance changes because of the series inductance. This reactance change is

$$jX_1' = \frac{jR^2 + (\omega_0 L)^2}{\omega_0 L} \tag{3.22}$$

If $X_1$ represents the series reactance, which in this case is $\omega_0 L$, then the reactance of the series element can be found also in terms of $Q$:

$$jX_1' = jX_1\left(1 + \frac{1}{Q_1^2}\right) \tag{3.23}$$

The second element in the LC section is chosen to resonate out this $X_1'$:

$$\begin{aligned} jX_2 &= -jX_1' = -jX_1\left(1 + \frac{1}{Q_1^2}\right) \\ &= \frac{-jR'}{Q_1} \end{aligned} \tag{3.24}$$

In the typical synthesis problem, $R'$ and $R$ are known. Equation (3.21) gives the necessary value of $Q_1$, Eq. (3.3) gives the required $L$, and Eq. (3.24) gives the required $C$. This procedure is summarized in Table 3.1.

A similar procedure can be applied for the shunt connection in which the capacitance is closest to the load resistance. The input impedance is expressed as follows:

$$\begin{aligned} Z_{\text{in}} &= j\omega L + \frac{1}{G + j\omega C} \\ &= \frac{G}{G^2 + (\omega C)^2} + j\left[\omega L - \frac{\omega C}{G^2 + (\omega C)^2}\right] \end{aligned} \tag{3.25}$$

**TABLE 3.1 *L* Matching Circuit Design Where $X_1$, $B_1$ are the Reactance or Susceptance Closest to the Load R**

| Circuit | $R'$ | $jX_2$ | $Q_1$ |
|---|---|---|---|
| Series | $R(1 + Q_1^2)$ | $-jX_1(1 + 1/Q_1^2)$ or $-jR'/Q_1$ | $X_1/R$ |
| Shunt | $R/(1 + Q_1^2)$ | $-jX_1/(1 + 1/Q_1^2)$ or $-jR'Q_1$ | $B_1/G$ |

For resonance, the series reactance $X = 0$. Solution for the resonant frequency for the shunt connection is as follows:

$$\omega_0 = \sqrt{\frac{1}{LC} - \frac{G^2}{C^2}} \tag{3.26}$$

Substituting this back into the input impedance expression gives the input resistance:

$$\begin{aligned} R' &= \frac{1/G}{1 + (\omega_0 C/G)^2} \\ &= \frac{R}{1 + Q_1^2} \end{aligned} \tag{3.27}$$

The reactance associated with the capacitance is

$$jX_1' = \frac{-j\omega_0 C}{G^2 + (\omega_0 C)^2}$$

Since $jX_1 = 1/j\omega C$,

$$jX_1' = \frac{jX_1}{1 + 1/Q_1^2} \tag{3.28}$$

$$= R'Q_1 \tag{3.29}$$

Since $jX_2 = -jX_1$, the values in Table 3.1 are obtained.

The major feature that should be recognized, whether dealing with elaborate lumped circuits or microwave circuits, if the impedance level needs to be raised, a series connection is needed. If the impedance needs to be lowered, a shunt connection is needed. Furthermore, since the design is based on a resonance condition, the two reactances in the circuit must be of the opposite type. This means two inductors or two capacitors will not work.

## 3.6 $\pi$ TRANSFORMATION CIRCUIT

In the previous $L$ matching circuit, the value for $Q$ is completely determined by the transformation ratio. Consequently there is no independent control over the value of $Q$ which is related to the circuit bandwidth. Addition of a third circuit element gives flexibility to design for bandwidth. If a design begins with a shunt $L$ matching circuit, then addition of another shunt susceptance on the other side of the series element provides the necessary circuit flexibility to be able to choose the circuit $Q$ as a design parameter. The resulting $\pi$ matching

circuit is shown in Fig. 3.4. In this circuit $B_1$ and $X_2$ both act as the impedance transforming elements while the third, $B_3$ is the compensation element that tunes out the excess reactance from the first two elements. As in the $L$ matching circuit, the first shunt element, $B_1$, reduces the resistance level by a factor of $1/(1 + Q_1^2)$, and $X_2$ increases the resistance level by $(1 + Q_2^2)$, where $Q_2$ is a $Q$ factor related to the second element. The final transformation ratio can be $R'' < R$ or $R'' > R$ depending on which $Q$ is larger, as shown in the diagram of Fig. 3.5. To make $R'' < R$, make $Q_1 > Q_2$. The maximum $Q$, $Q_{\max} = Q_1$, will be the major factor that determines the bandwidth.

Now consider design of a circuit where $R'' < R$. Then the first shunt transformation gives

$$R' = \frac{R}{1 + Q_1^2} \tag{3.30}$$

$$X' = -R'Q_1 \tag{3.31}$$

The incremental reactance, $X'$, is added to the series arm. This results in the circuit shown in Fig. 3.6, where $R$ has been transformed to $R'$ with a modified series reactance. This series reactance will act to increase the resistance level from $R'$ to $R''$. The second transformation $Q$ is

$$Q_2 = \frac{X_2 - R'Q_1}{R'} = \frac{R'(X_2/R - Q_1)}{R'}$$

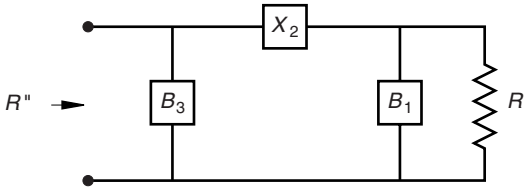



**FIGURE 3.4** $\pi$ impedance transformation circuit.

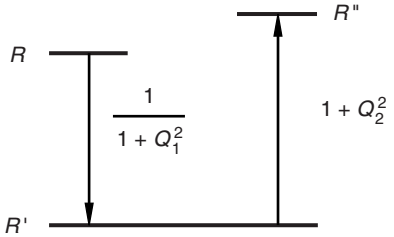



**FIGURE 3.5** Diagram showing two-step transformation.

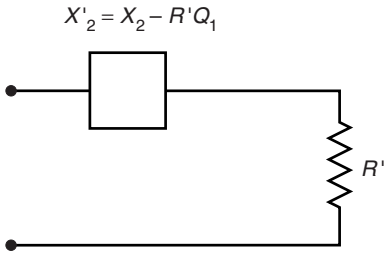



**FIGURE 3.6** Equivalent series reactance after first transformation.

**TABLE 3.2 $\pi$ Matching Circuit Design Formulas**

| Step Number | $R'' < R$ | $R'' > R$ |
|---|---|---|
| 1 | $Q_1 = Q_{\max}$ | $Q_2 = Q_{\max}$ |
| 2 | $R' = R/(1 + Q_1^2)$ | $R' = R''/(1 + Q_2^2)$ |
| 3 | $1 + Q_2^2 = R''/R'$ | $1 + Q_1^2 = R/R'$ |
| 4 | $X_2 = R'(Q_1 + Q_2)$ | $X_2 = R'(Q_1 + Q_2)$ |
| 5 | $B_1 = Q_1/R$ | $B_1 = Q_1/R$ |
| 6 | $B_3 = Q_2/R''$ | $B_3 = Q_2/R''$ |

or

$$\frac{X_2}{R'} = Q_1 + Q_2 \tag{3.32}$$

The $X_2$, $B_3$ combination is a series L section with "load" of $R'$. Consequently

$$R'' = R'(1 + Q_2^2) \tag{3.33}$$

$$X'' = -\frac{R''}{Q_2} \tag{3.34}$$

A summary for the design process is shown below. To make $R'' < R$, make $Q_1 > Q_2$ and $Q_1 = Q_{\max}$, and follow the design steps in the first column of Table 3.2. For $R'' > R$, use the second column.

## 3.7 *T* TRANSFORMATION CIRCUIT

The $T$ transformation circuit is the dual to the $\pi$ transformation circuit and is shown in Fig. 3.7. In this circuit, however, the series reactance $X_1$ first raises the resistance level to $R'$, and the remaining shunt susceptance lowers the resistance level as indicated in Fig. 3.8. The design formulas are derived in the same way as the $\pi$ circuit formulas and are summarized in Table 3.3.

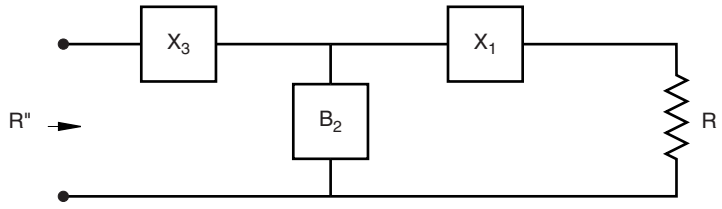



**FIGURE 3.7** $T$ transformation circuit.

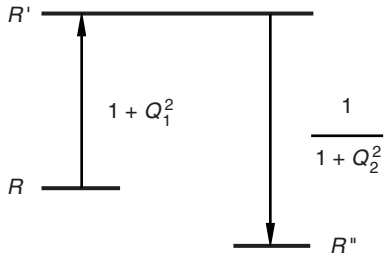



**FIGURE 3.8** Diagram showing impedance transformation for the $T$ circuit.

**TABLE 3.3** ***T* Matching Circuit Design Formulas**

| Step Number | $R'' > R$ | $R'' < R$ |
|---|---|---|
| 1 | $Q_1 = Q_{\max}$ | $Q_2 = Q_{\max}$ |
| 2 | $R' = R(1 + Q_1^2)$ | $R' = R''(1 + Q_2^2)$ |
| 3 | $1 + Q_2^2 = R'/R''$ | $1 + Q_1^2 = R'/R$ |
| 4 | $X_1 = Q_1 R$ | $X_1 = Q_1 R$ |
| 5 | $B_2 = (Q_1 + Q_2)/R'$ | $B_2 = (Q_1 + Q_2)/R'$ |
| 6 | $X_3 = Q_2/R''$ | $X_3 = Q_2/R''$ |

## 3.8 TAPPED CAPACITOR TRANSFORMER

The tapped capacitor circuit is another approximate method for obtaining impedance level transformation. The description of this design process will begin with a parallel RC to series RC conversion. Then the tapped $C$ circuit will be converted to an $L$-shaped matching circuit. The $Q_1$ for an equivalent load resistance $R_{\text{eqv}}$ will be found. Finally, a summary of the circuit synthesis procedure will be given.

### 3.8.1 Parallel to Series Conversion

Shown in Fig. 3.9 is a parallel RC circuit that will be forced to have the same impedance as the series RC circuit, at least at one frequency. The conversion is of

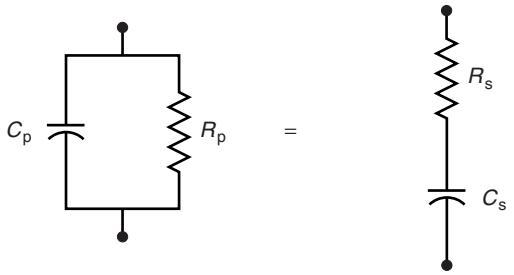



**FIGURE 3.9** Parallel RC to series RC conversion.

course valid for only a narrow frequency range, so this method is fundamentally limited by this approximation.

The impedance of the parallel circuit is

$$Z_p = \frac{R_p}{1 + sC_pR_p} \tag{3.35}$$

The $Q$ for a parallel circuit is $Q_p = \omega C_p R_p$. The equivalent series resistance and reactance in terms of $Q_p$ are

$$R_{seqv} = \frac{R_p}{1 + Q_p^2} \tag{3.36}$$

$$X_{seqv} = -\frac{X_p Q_p^2}{1 + Q_p^2} \tag{3.37}$$

### 3.8.2 Conversion of Tapped *C* Circuit to an *L*-Shaped Circuit

The schematic of the tapped $C$ circuit is shown in Fig. 3.10 where $R'$ is to be matched to $R_2$. The parallel $R_2C_2$ section is converted to a series $R_{eqv}C_{eqv}$, as indicated in Fig. 3.11. Making use of Eqs. (3.36) and (3.37),

$$C_{seqv} = C_2\left(\frac{1 + Q_p^2}{Q_p^2}\right) \approx C_2 \qquad \text{for high } Q_p \tag{3.38}$$

$$R_{seqv} = R = \frac{R_2}{1 + Q_p^2} \tag{3.39}$$

where $Q_p = \omega_0 C_2 R_2$. Considering $R'$ as the load, and using the $L$ circuit transformation for a shunt circuit in Table 3.1,

$$R_{seqv} = \frac{R'}{1 + Q_1^2} \tag{3.40}$$

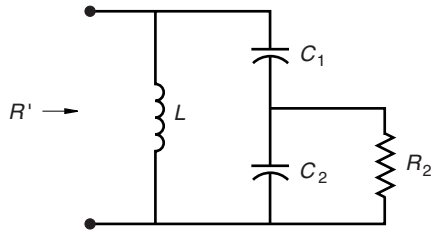



**FIGURE 3.10** Tapped $C$ transformation circuit.

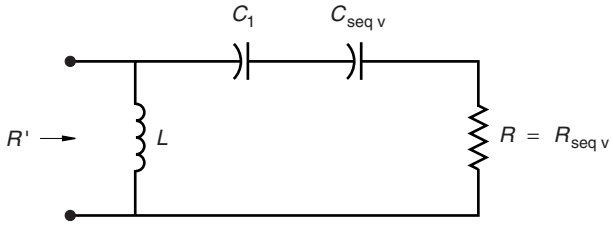



**FIGURE 3.11** Intermediate equivalent transformation circuit.

This is the transformed resistance looking through $C_1$ toward the left. Looking toward the right through $C_{\text{seqv}}$ and again using the parallel to series conversion, Eq. (3.36),

$$R_{\text{seqv}} = \frac{R_2}{1 + Q_{\text{p}}^2} \tag{3.41}$$

These two expressions for $R_{\text{seqv}}$ can be equated and solved for $Q_{\text{p}}$:

$$Q_{\text{p}} = \left[\frac{R_2}{R}(1 + Q_1^2) - 1\right]^{1/2} \tag{3.42}$$

### 3.8.3 Calculation of Circuit *Q*

An approximate value for $Q$ can be found by equating the impedances of the two circuits in Fig. 3.12:

$$Z = \frac{R'\omega^2 L^2 + jR'^2\omega L}{R'^2 + (\omega L)^2} = R'_{\text{eqv}} + j\omega L_{\text{eqv}} \tag{3.43}$$

If the $Q$ of the right-hand circuit is approximately that of the left-hand circuit in Fig. 3.12, then

$$Q_1 = \frac{\omega_0 L_{\text{eqv}}}{R'_{\text{eqv}}} = \frac{\omega_0 R'^2 L}{R'\omega_0^2 L^2} = \frac{R'}{\omega_0 L} \tag{3.44}$$

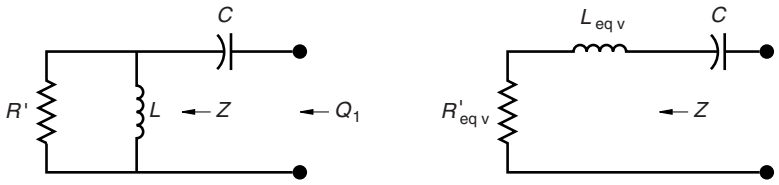



**FIGURE 3.12** Equate the left- and right-hand circuits.

**TABLE 3.4 Tapped *C* Matching Circuit Design Formulas**

| Step Number | Tapped $C$ Formula |
|---|---|
| 1 | $Q_1 = f_0/\Delta f$ |
| 2 | $C = Q_1/\omega_0 R' = 1/2\pi\Delta f R'$ |
| 3 | $L = 1/\omega_0^2 C$ |
| 4 | $Q_p = \left[(R_2/R')(1+Q_1^2) - 1\right]^{1/2}$ |
| 5 | $C_2 = Q_p/\omega_0 R_2$ |
| 6 | $C_{seqv} = C_2(1+Q_p^2)/Q_p^2$ |
| 7 | $C_1 = C_{seqv}C_2/(C_{seqv} - C_2)$ |

The variable $C$ represents the total capacitance of $C_1$ and $C_{seqv}$ in series, as implied in Fig. 3.11 and represented in Fig. 3.12. For a high $Q$ circuit, circuit analysis gives the resonant frequency:

$$\omega_0^2 = \frac{1}{LC - L^2/R'^2} \approx \frac{1}{LC} \tag{3.45}$$

As a result the approximate value for $Q_1$ can be found:

$$Q_1 = \omega_0 R' C = \frac{f_0}{\Delta f} \tag{3.46}$$

Here $\Delta f$ is the bandwidth in Hz and $f_0$ is the resonant frequency.

### 3.8.4 Tapped *C* Design Procedure

The above ideas are summarized in Table 3.4, which provides a design procedure for the tapped $C$ matching circuit. Similar expressions could be found for a tapped inductor transforming circuit, but such a circuit is typically less useful because inductors are more difficult to obtain than capacitors.

## 3.9 PARALLEL DOUBLE-TUNED TRANSFORMER

Each of the above described $T$, $\pi$, or tapped $C$ matching circuits provide some control over the bandwidth. Where precise control over the bandwidth is required, a double tuned circuit allows controlling bandwidth by specifying two different frequencies where maximum transmission occurs. For a small pass band, the

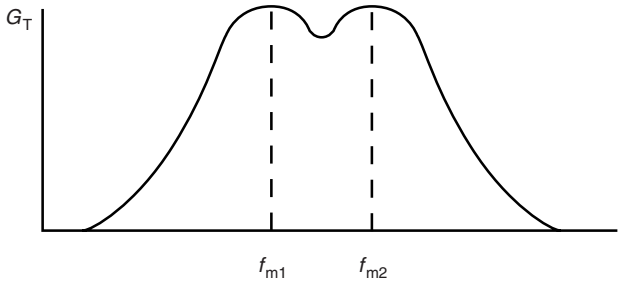



**FIGURE 3.13** Double tuned transformer response.

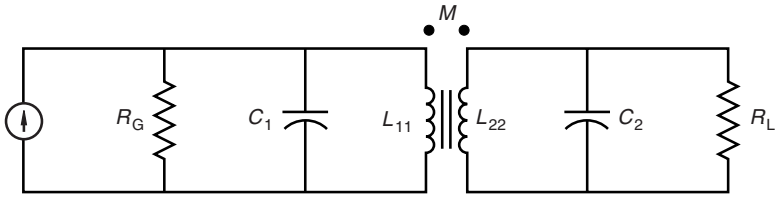



**FIGURE 3.14** Real transformer with resonating capacitances.

midband dip in the transmission can be made small. Furthermore the double-tuned circuit is especially useful when a large difference in impedance levels is desired although its high end frequency range is limited. The filter transmission gain is shown in Fig. 3.13.

The double-tuned circuit consists of a coupled coil transformer with resonating capacitances on the primary and secondary side. This circuit is shown in Fig. 3.14. The transformer is described by its input and output inductance as well as the coupling coefficient $k$. The turns ratio for the transformer is

$$n:1 = \sqrt{\frac{L_{11}}{k^2 L_{22}}}:1 \tag{3.47}$$

The circuit in Fig. 3.14 can be replaced by an equivalent circuit using an ideal transformer (Fig. 3.15*a*). Since an ideal transformer has no inductance, the inductances and coupling factor, $k$, must be added to the ideal transformer. The final circuit topology is shown in Fig. 3.15*b*. Looking toward the right through the ideal transformer, Fig. 3.15*b* shows the circuit values are:

$$L_2' = L_{11}\left(\frac{1}{k^2} - 1\right) \tag{3.48}$$

$$C_2' = \left(\frac{L_{11}}{k^2 L_{22}}\right) C_2 \tag{3.49}$$

$$R_L' = \frac{L_{11}}{k^2 L_{22}} R_L \tag{3.50}$$

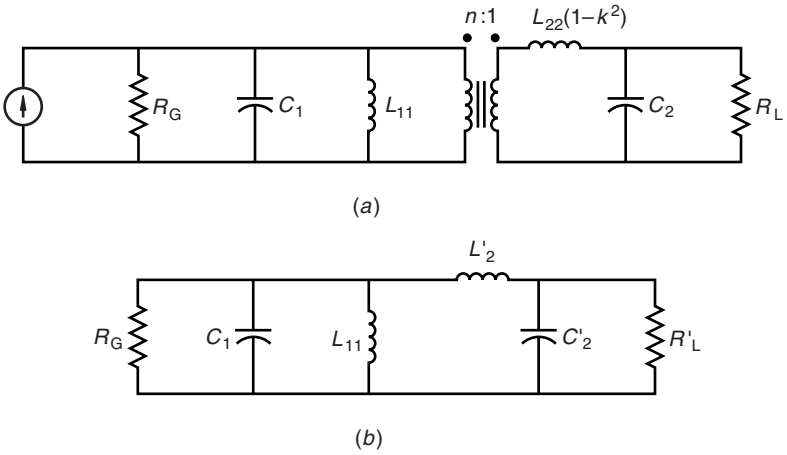



**FIGURE 3.15** (*a*) Alternate equivalent circuit with an ideal transformer, and (*b*) final equivalent circuit.

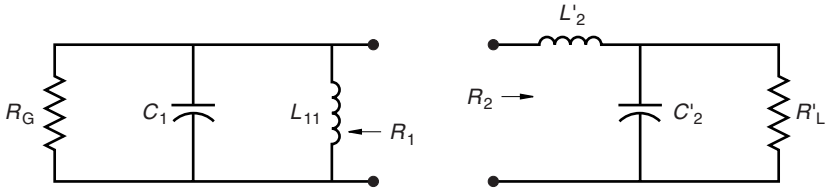



**FIGURE 3.16** Double tuned circuit split into two.

The circuit elements will be chosen to give exact match at the two frequencies, $f_{m1}$ and $f_{m2}$. The circuit in Fig. 3.15*b* can be conceptually split into two (Fig. 3.16). The resistance $R_1$ with the parallel resonant circuit will never be larger than $R_G$. The right-hand side is an $L$ matching circuit with the reactance of the shunt element monotonically decreasing with frequency. Hence $R_2$ monotonically decreases. Consequently, if $R_L$ is small enough, there will be two frequencies where $R_1 = R_2$. This is illustrated in Fig. 3.17.

A design procedure for the parallel double-tuned circuit has been reviewed in [1] and is summarized below. The typical synthesis problem is to design a circuit that will match $R_G$ and $R_L$ over a bandwidth, $\Delta f$, at a center frequency, $f_0$, with a given pass band ripple. The bandwidth and center frequency are approximated by the following:

1. Determine $f_{m1}$ and $f_{m2}$:

$$\Delta f \approx \sqrt{2}(f_{m2} - f_{m1}) \tag{3.51}$$

$$f_0 \approx \sqrt{f_{m1} f_{m2}} \tag{3.52}$$

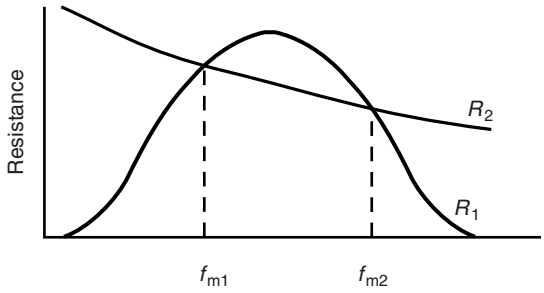



**FIGURE 3.17** Plot of left- and right-hand resistance values versus frequency.

The minimum pass band gain for the filter is dependent on the difference between the match frequencies:

$$G_{\text{Tmin}} = \frac{4f_{m2}/f_{m1}}{(f_{m2}/f_{m1})^2 + 2f_{m2}/f_{m1} + 1} \tag{3.53}$$

2. Determine the actual transducer gain for the given ripple factor:

$$G_{\text{T}} = 10^{-\text{ripple factor (dB)}/10} \tag{3.54}$$

3. Find the resistance ratio if $G_{\text{T}} > G_{\text{Tmin}}$, the pass band ripple specification can be met:

$$r = \frac{1 + |1 - G_{\text{T}}|^{1/2}}{1 - |1 - G_{\text{T}}|^{1/2}} \tag{3.55}$$

4. Calculate the $Q_2$ at the two matching frequencies:

$$Q_{2-m1}^2 = r\frac{f_{m1}}{f_{m2}} - 1 \tag{3.56}$$

$$Q_{2-m2}^2 = r\frac{f_{m2}}{f_{m1}} - 1 \tag{3.57}$$

5. Solve the following simultaneous equations for $L_2'$ and $C_2'$:

$$-\omega_{m1}L_2' + \frac{1}{\omega_{m1}C_2'} = |Q_{2-m1}|\frac{R_{\text{G}}}{1 + Q_{2-m1}^2} \tag{3.58}$$

$$+\omega_{m2}L_2' + \frac{1}{\omega_{m2}C_2'} = |Q_{2-m2}|\frac{R_{\text{G}}}{1 + Q_{2-m2}^2} \tag{3.59}$$

6. Find the value for $R_{\text{L}}'$:

$$R_{\text{L}}' = \frac{1 + Q_{2-m1}^2}{\omega_{m1}^2 C_2'^2 R_{\text{G}}} \tag{3.60}$$

7. Calculate the input susceptance of the right-hand side where $G'_{\mathrm{L}} = 1/R'_L$:

$$B_{m1} = Im\left\{\frac{1}{j\omega_{m1}L'_2 + (1/G'_{\mathrm{L}} + j\omega_{m1}C'_2)}\right\} \tag{3.61}$$

$$B_{m2} = Im\left\{\frac{1}{j\omega_{m2}L'_2 + (1/G'_{\mathrm{L}} + j\omega_{m2}C'_2)}\right\} \tag{3.62}$$

8. Solve the following simultaneous equations for $L_{11}$ and $C_1$:

$$\frac{1}{\omega_{m1}L_{11}} - \omega_{m1}C_1 = |B_{m1}| \tag{3.63}$$

$$\frac{1}{\omega_{m2}L_{11}} - \omega_{m2}C_1 = |B_{m2}| \tag{3.64}$$

9. Find the transformer coupling coefficient, and hence $L_{22}$ and $C_2$:

$$k = \frac{1}{\sqrt{1 + L'_2/L_{11}}} \tag{3.65}$$

$$L_{22} = \frac{L_{11}R_L}{k^2R'_L} \tag{3.66}$$

$$C_2 = \frac{L_{11}}{k^2L_{22}}C'_2 \tag{3.67}$$

This procedure has been coded into the program DBLTUNE, and an example of its use is given in Appendix C.

## PROBLEMS

**3.1** Design an impedance transforming network that matches a generator resistance, $R_{\mathrm{G}} = 400\ \Omega$ to a load resistance $R_{\mathrm{L}} = 20\ \Omega$. The center frequency for the circuit is $f_0 = 6$ MHz. The desired ripple (where appropriate) is to be less than 0.25 dB. In some cases the ripple factor will not be able to be controlled in the design. The problem is to design four different transformation circuits with the specifications above, and for each design do an analysis using SPICE. See Appendix G, Sections G.1, and G.2.

**(a)** Design a two-element $L$ matching circuit and check the results with SPICE.

**(b)** Design a three-element tapped capacitor matching circuit with a bandwidth $\Delta f = 50$ kHz, and check the results with SPICE to determine the actual bandwidth.

**(c)** Design a three-element $\pi$ matching circuit with a bandwidth of $\Delta f = 50$ kHz, and check the results with SPICE to determine the actual bandwidth.

**(d)** Design a double tuned transformer matching circuit with a bandwidth of $\Delta f = 50$ kHz, and check the results with SPICE to determine the actual bandwidth.

**(e)** Repeat part (d) for a 3 dB bandwidth of 2 MHz. Again, check the results using SPICE.

**3.2** The $\pi$ matching circuit shown in Fig. 3.4 is used to match the load $R = 1000\ \Omega$ to $R'' = 80\ \Omega$. If the intermediate resistance level is $R' = 20\ \Omega$, determine the following:

**(a)** What is $Q_1$?

**(b)** What is $Q_2$?

**(c)** What is $B_1$, the first susceptance nearest $R$?

**(d)** What is the estimated 3 dB bandwidth for this circuit in terms of the center frequency, $f_0$?

**3.3** The tapped capacitor transformer is to be used in a narrowband of frequencies around $\omega = 4 \cdot 10^9$ rad/s. In designing the matching circuit, the tapped $C$ circuit is converted to an 'L' matching circuit. If $R_2$ in Fig. 3.10 is 50 $\Omega$, $C_2 = 8$ pF, and $C_1 = 5.0$ pF, then what is the total capacitance for the 'L' matching circuit?

## REFERENCES

1. P. L. D. Abrie, *The Design of Impedance-Matching Networks for Radio-Frequency and Microwave Amplifiers*, Norwood, MA: Artech House, 1985.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER FOUR

# Multiport Circuit Parameters and Transmission Lines

## 4.1 VOLTAGE–CURRENT TWO-PORT PARAMETERS

A linear $n$-port network is completely characterized by $n$ independent excitation variables and $n$ dependent response variables. These variables are the terminal voltages and currents. There are four ways of arranging these independent and dependent variables for a two-port, and they are particularly useful, when considering feedback circuits. They are the impedance parameters ($z$-matrix), admittance parameters ($y$-matrix), hybrid parameters ($h$-matrix), and the inverse hybrid parameters ($g$-matrix). These four sets of parameters are defined as.

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{4.1}$$

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{4.2}$$

$$\begin{bmatrix} v_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ v_2 \end{bmatrix} \tag{4.3}$$

$$\begin{bmatrix} i_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ i_2 \end{bmatrix} \tag{4.4}$$

Two networks connected in series (Fig. 4.1) can be combined by simply adding the $z$ parameters of each network together. This configuration is called the *series–series* connection. In the shunt–shunt configuration shown in Fig. 4.2, the two circuits can be combined by adding their $y$-matrices together. In the series–shunt configuration (Fig. 4.3), the composite matrix for the combination is found by adding the $h$ parameters of each circuit together. Finally, the circuits

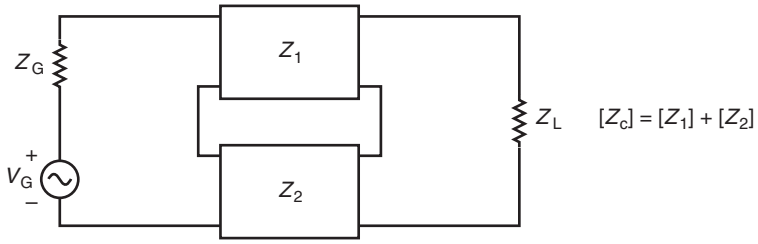



**FIGURE 4.1** Series–series connection.

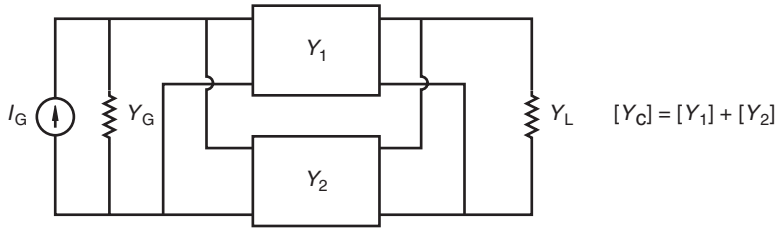



**FIGURE 4.2** Shunt–shunt connection.

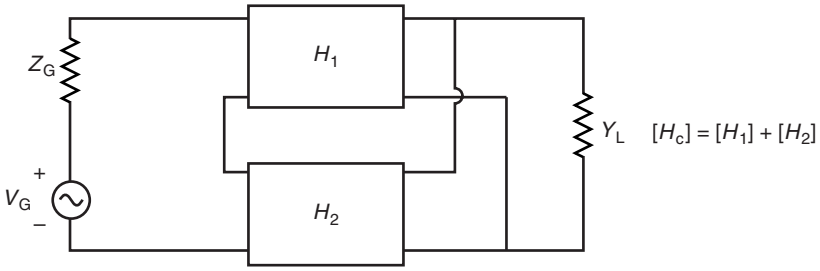



**FIGURE 4.3** Series–shunt connection.

connected in the shunt–series configuration (Fig. 4.4) can be combined by adding the $g$ parameters of the respective circuits. In each type of configuration the independent variables are the same for the individual circuits, so matrix addition is valid most of the time. The case where the matrix addition is not valid occurs when for example in Fig. 4.1 a current going in and out of port 1 of circuit 1 is not equal to the current going in and out of port 1 of circuit 2. These pathological cases will not be of concern here, but further information is found in [1, pp. 188–191] where a description of the Brune test is given.

Any of the four types of circuit parameters described above can be represented by an equivalent circuit with controlled sources. As an example, the impedance (or $z$) parameters can be represented as shown in Fig. 4.5. The input port-1 side is represented by a series resistance of value $z_{11}$ together with a current controlled

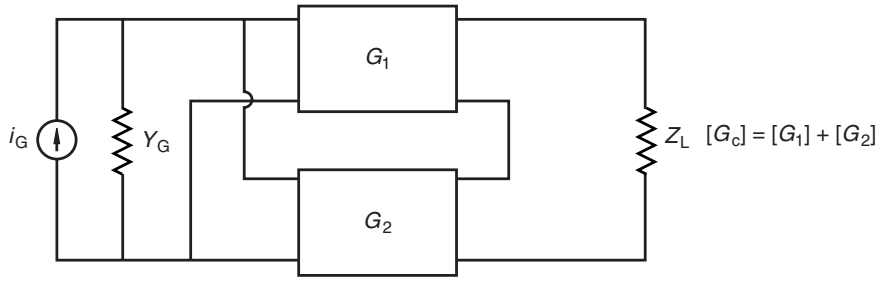



**FIGURE 4.4** Shunt–series connection.

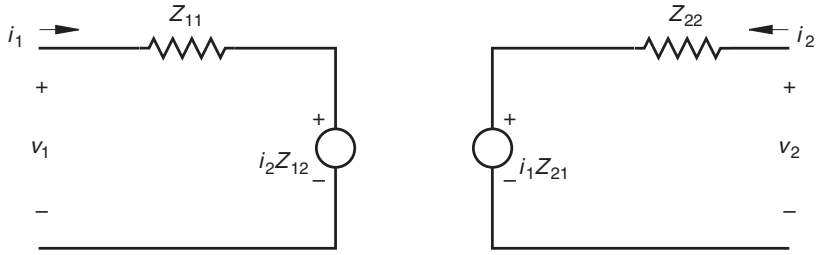



**FIGURE 4.5** Equivalent circuit for the $z$ parameters.

voltage source with gain $z_{12}$ in series. The controlling current is the port-2 current. If the current at port-1 is $i_1$ and the current at port-2 is $i_2$, then the voltage at port-1 is

$$v_1 = i_1 z_{11} + i_2 z_{12}$$

A similar representation is used for the port-2 side.

The individual impedance parameters are found for a given circuit by setting $i_1$ or $i_2$ to 0 and solving for the appropriate $z$ parameter. The $z$ parameters are sometimes termed *open circuit parameters* for this reason. The $y$ parameters are sometimes called *short circuit parameters* because they are found by shorting the appropriate port. These parameters are all summarized in Appendix D where conversions are given for converting them to scattering parameters.

## 4.2 *ABCD* PARAMETERS

Networks are often cascaded together, and it would be useful to be able to describe each network in such a way that the product of the matrices of each individual network would describe the total composite network. The *ABCD* parameters have the property of having the port-1 variables being the independent variables and the port-2 variables being the dependent ones:

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ -i_2 \end{bmatrix} \tag{4.5}$$

This allows the cascade of two networks to be represented as the matrix product of the two circuit expressed in terms of the *ABCD* parameters. The *ABCD* parameters can be expressed in terms of the commonly used $z$ parameters:

$$A = \left.\frac{v_1}{v_2}\right|_{i_2=0} = \frac{z_{11}}{z_{21}} \tag{4.6}$$

$$B = -\left.\frac{v_1}{i_2}\right|_{v_2=0} = \frac{\Delta_z}{z_{21}} \tag{4.7}$$

$$C = \left.\frac{i_1}{v_2}\right|_{i_2=0} = \frac{1}{z_{21}} \tag{4.8}$$

$$D = -\left.\frac{i_1}{i_2}\right|_{v_2=0} = \frac{z_{22}}{z_{21}} \tag{4.9}$$

where

$$\Delta_z \triangleq z_{11}z_{22} - z_{21}z_{12}$$

In addition, if the circuit is reciprocal so that $z_{12} = z_{21}$, then the determinate of the *ABCD* matrix is unity, namely

$$AD - BC = 1 \tag{4.10}$$

## 4.3 IMAGE IMPEDANCE

A generator impedance is said to be matched to a load when the generator can deliver the maximum power to the load. This occurs when the generator impedance is the complex conjugate of the load impedance. For a two-port circuit, the generator delivers power to the circuit, which in turn has a certain load impedance attached to the other side (Fig. 4.6). Consequently maximum power transfer from the generator to the input of the two-port circuit occurs when it has the appropriate load impedance, $Z_L$. The optimum generator impedance

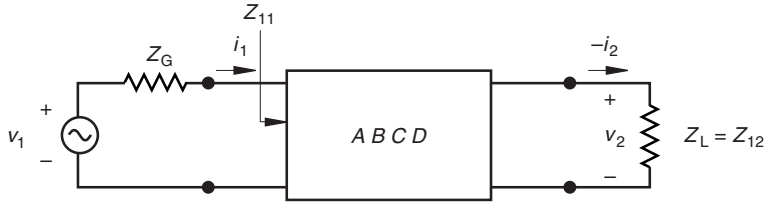



**FIGURE 4.6** Excitation of a two-port at port-1.

depends on both the two-port circuit itself and its load impedance. In addition the matched load impedance at the output side will depend on the two-port as well as on the generator impedance on the input side. Both sides are matched simultaneously when the input side is terminated with an impedance equal to its image impedance, $Z_{I1}$, and the output side is terminated with a load impedance equal to $Z_{I2}$. The actual values for $Z_{I1}$, and $Z_{I2}$ are determined completely by the two-port circuit itself and are independent of the loading on either side of the circuit. Terminating the two-port circuit in this way will guarantee maximum power transfer from the generator into the input side and maximum power transfer from a generator at the output side (if it exists).

The volt–ampere equations for a two-port are given in terms of their *ABCD* parameters as

$$v_1 = Av_2 - Bi_2 \tag{4.11}$$

$$i_1 = Cv_2 - Di_2 \tag{4.12}$$

Now, if the input port is terminated by $Z_{I1} = v_1/i_1$, and the output port by $Z_{I2} = v_2/(-i_2)$, then both sides will be matched. Taking the ratios of Eqs. (4.11) and (4.12) gives

$$\begin{aligned} Z_{I1} = \frac{v_1}{i_1} &= \frac{Av_2/(-i_2) + B}{Cv_2/(-i_2) + D} \\ &= \frac{AZ_{I2} + B}{CZ_{I2} + D} \end{aligned} \tag{4.13}$$

The voltage and current for the output side in terms of these parameters of the input side are found by inverting Eqs. (4.11) and (4.12):

$$v_2 = Dv_1 - Bi_1 \tag{4.14}$$

$$i_2 = Cv_1 - Ai_1 \tag{4.15}$$

If the output port is excited by $v_2$ as shown in Fig. 4.7, then the matched load impedance is the same as the image impedance:

$$Z_{I2} = \frac{v_2}{i_2} = \frac{Dv_1/(-i_1) + B}{Cv_1/(-i_1) + A} = \frac{DZ_{I1} + B}{CZ_{I1} + A} \tag{4.16}$$

Equations (4.13) and (4.16) can be solved to find the image impedances for both sides of the circuit:

$$Z_{I1} = \sqrt{\frac{AB}{CD}} \tag{4.17}$$

$$Z_{I2} = \sqrt{\frac{DB}{AC}} \tag{4.18}$$

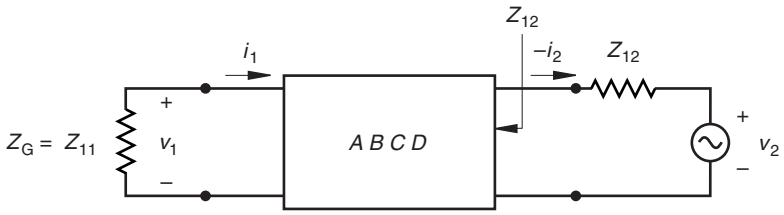



**FIGURE 4.7** Excitation of a two-port at port-2.

When a two-port circuit is terminated on each side by its image impedance so that $Z_G = Z_{I1}$ and $Z_L = Z_{I2}$, then the circuit is matched on both sides simultaneously. The input impedance is $Z_{I1}$ if the load impedance is $Z_{I2}$, and vice versa.

The image impedance can be written in terms of the open circuit $z$ parameters and the short circuit $y$ parameters by making the appropriate substitutions for the $ABCD$ parameters (see Appendix D):

$$Z_{I1} = \sqrt{\frac{z_{11}}{y_{11}}} \tag{4.19}$$

$$Z_{I2} = \sqrt{\frac{z_{22}}{y_{22}}} \tag{4.20}$$

Therefore an easy way to remember the values for the image impedances is

$$Z_{I1} = \sqrt{z_{oc1} z_{sc1}} \tag{4.21}$$

$$Z_{I2} = \sqrt{z_{oc2} z_{sc2}} \tag{4.22}$$

where $z_{oc1}$ and $z_{sc1}$ are the input impedances of the two-port circuit when the output port is an open circuit or a short circuit, respectively.

As an example consider the simple T circuit in Fig. 4.8. The input impedance when the output is an open circuit is

$$z_{oc1} = Z_a + Z_b \tag{4.23}$$

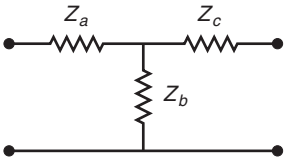



**FIGURE 4.8** Example T circuit.

and the input impedance when the output is a short circuit is

$$z_{sc1} = Z_a + Z_b \| Z_c \tag{4.24}$$

The image impedance for the input port for this circuit is

$$Z_{I1} = \sqrt{(Z_a + Z_b)[Z_c + Z_b \| Z_c]} \tag{4.25}$$

and similarly for the output port

$$Z_{I2} = \sqrt{(Z_c + Z_b)[Z_c + Z_b \| Z_a]} \tag{4.26}$$

The output side of the two-port circuit can be replaced by another two-port whose input impedance is $Z_{I2}$. This is possible if $Z_{I2}$ is the image impedance of the second circuit and the load of the second circuit is equal to its output image impedance, say $Z_{I3}$. A cascade of two-port circuits where each port is terminated by its image impedance would be matched everywhere (Fig. 4.9). A wave entering from the left side could propagate through the entire chain of two-port circuits without any internal reflections. There of course could be some attenuation if the two-port circuits contain lossy elements.

The image propagation constant, $\gamma$, for a two-port circuit is defined as

$$e^{\gamma} = \sqrt{\frac{v_1 i_1}{v_2(-i_2)}} = \frac{v_1}{v_2}\sqrt{\frac{Z_{I2}}{Z_{I1}}} \tag{4.27}$$

If the network is symmetrical so that $Z_{I1} = Z_{I2}$, then $e^{\gamma} = v_1/v_2$. For the general unsymmetrical network, the ratio $v_1/v_2$ is found from Eq. (4.11) as

$$\begin{aligned}\frac{v_1}{v_2} &= \frac{Av_2 - Bi_2}{v_2} \\ &= A + \frac{B}{Z_{I2}} \\ &= A + B\sqrt{\frac{AC}{BD}} \\ &= \sqrt{\frac{A}{D}}\left(\sqrt{AD} + \sqrt{BC}\right)\end{aligned}$$

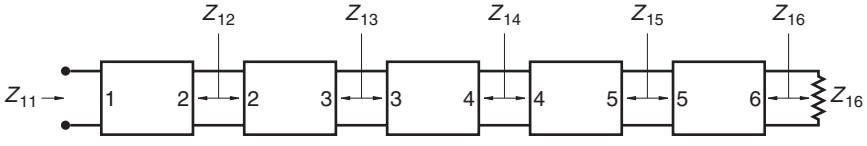



**FIGURE 4.9** Chain of matched two-port circuits.

Similarly

$$\frac{i_1}{(-i_2)} = CZ_{I2} + D$$
$$= \sqrt{\frac{D}{A}}\left(\sqrt{AD} + \sqrt{BC}\right)$$

The image propagation constant is obtained from Eq. (4.27):

$$e^{\gamma} = \sqrt{\frac{v_1 i_1}{v_2(-i_2)}} = \sqrt{AD} + \sqrt{BC} \tag{4.28}$$

Also

$$e^{-\gamma} = \sqrt{AD} - \sqrt{BC} \tag{4.29}$$

When the circuit is reciprocal, $AD - BC = 1$. Now if Eqs. (4.28) and (4.29) are added together and then subtracted from one another, the image propagation constant can be expressed in terms of hyperbolic functions:

$$\cosh\gamma = \sqrt{AD} \tag{4.30}$$

$$\sinh\gamma = \sqrt{BC} \tag{4.31}$$

If $n$ represents the square root of the image impedance ratio, the $ABCD$ parameters can then be written in terms of $n$ and $\gamma$:

$$n \triangleq \sqrt{\frac{Z_{I1}}{Z_{I2}}}$$
$$= \sqrt{\frac{A}{D}} \tag{4.32}$$

$$A = n\cosh\gamma \tag{4.33}$$

$$B = nZ_{I2}\sinh\gamma \tag{4.34}$$

$$C = \frac{\sinh\gamma}{nZ_{I2}} \tag{4.35}$$

$$D = \frac{\cosh\gamma}{n} \tag{4.36}$$

Hence, from the definition of the $ABCD$ matrix (4.5), the terminal voltages and currents can be written in terms of $n$ and $\gamma$:

$$v_1 = nv_2\cosh\gamma - ni_2Z_{I2}\sinh\gamma \tag{4.37}$$

$$i_1 = \frac{v_2}{nZ_{I2}}\sinh\gamma - \frac{i_2}{n}\cosh\gamma \tag{4.38}$$

Division of these two equations gives the input impedance of the two-port circuit when it is terminated by $Z_L$:

$$Z_{in} = \frac{v_1}{i_1} = n^2 Z_{I2} \frac{Z_L + Z_{I2} \tanh \gamma}{Z_L \tanh \gamma + Z_{I2}} \tag{4.39}$$

This is simply the transmission line equation for a lumped parameter network when the output is terminated by $Z_L = v_2/(-i_2)$. A clear distinction should be drawn between the input impedance of the network, $Z_{in}$, which depends on the value of $Z_L$, and the image impedance $Z_{I2}$, which depends only on the two-port circuit itself. For a standard transmission line, $Z_{I1} = Z_{I2} = Z_0$, where $Z_0$ is the characteristic impedance of the transmission line. Just as for the image impedance, the characteristic impedance does not depend on the terminating impedances, but is a function of the geometrical features of the transmission line. When the lumped parameter circuit is lossless, $\gamma = j\beta$ is pure imaginary and the hyperbolic functions become trigonometric functions:

$$Z_{in} = n^2 Z_{I2} \frac{Z_L + jZ_{I2} \tan \beta}{Z_{I2} + jZ_L \tan \beta} \tag{4.40}$$

where $\beta$ is real. For a lossless transmission line of electrical length $\theta = \omega L/v$,

$$Z_{in} = Z_0 \frac{Z_L + jZ_0 \tan \theta}{Z_0 + jZ_L \tan \theta} \tag{4.41}$$

where $\omega$ is the radian frequency, $L$ is the length of the transmission line, and $v$ is the velocity of propagation in the transmission line medium.

## 4.4 THE TELEGRAPHER'S EQUATIONS

A transmission line consists of two conductors that are spaced somewhat less than a wavelength apart. The transmission line is assumed to support only a transverse electromagnetic (TEM) wave. The transmission line might support higher-order modes at higher frequencies, but it is assumed here that only the TEM wave is present. This assumption applies to the vast number of two conductor transmission lines used in practice. A transmission line may take a wide variety of forms; here it will be presented as a two-wire transmission line (Fig. 4.10). This line is represented as having a certain series inductance per unit length, $L$, and a certain shunt capacitance per unit length, $C$ (Fig. 4.11). The inductance for the differential length is thus $Ldz$, and the capacitance is $Cdz$. If the incoming voltage and current wave entering port 1 is $V = v_1$ and $I = i_1$, respectively, then the voltage at port 2 is

$$v_2 = V + \frac{\partial V}{\partial z} dz$$

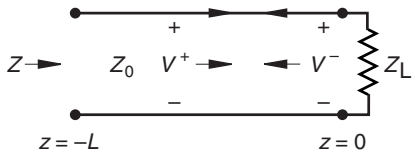



**FIGURE 4.10** Two wire representation of a transmission line.

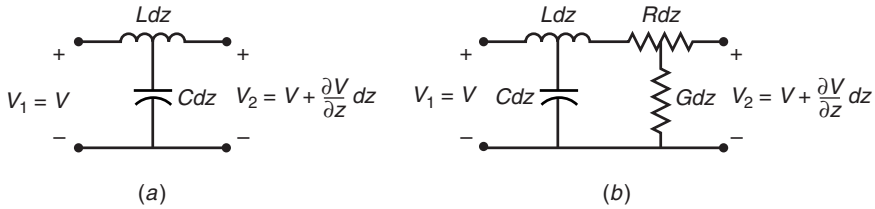



**FIGURE 4.11** Circuit model of a differential length of a transmission line where (*a*) is the lossless line and (*b*) is the lossy line.

so the voltage difference between ports 1 and 2 is

$$v_2 - v_1 = \frac{\partial V}{\partial z} dz = -L dz \frac{\partial I}{\partial t} \tag{4.42}$$

The negative sign for the derivative indicates the voltage is decreasing in going from port-1 to port-2. Similarly the difference in current from port-1 to port-2 is the current going through the shunt capacitance:

$$i_2 - i_1 = \frac{\partial I}{\partial z} dz = -C dz \frac{\partial V}{\partial t} \tag{4.43}$$

The telegrapher's equations are obtained from Eqs. (4.42) and (4.43):

$$\frac{\partial V}{\partial z} = -L \frac{\partial I}{\partial t} \tag{4.44}$$

$$\frac{\partial I}{\partial z} = -C \frac{\partial V}{\partial t} \tag{4.45}$$

Differentiation of Eq. (4.44) with respect to $z$ and Eq. (4.45) with respect to $t$, and then combining, produces the voltage wave equation:

$$\frac{\partial^2 V}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 V}{\partial t^2} \tag{4.46}$$

In similar fashion the current wave equation can be found:

$$\frac{\partial^2 I}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 I}{\partial t^2} \tag{4.47}$$

The velocity of the wave is

$$v = \frac{1}{\sqrt{LC}} \tag{4.48}$$

The solution for these two wave equations given below in terms of the arbitray functions $F_1$ and $F_2$ can be verified by substitution back into Eqs. (4.46) and (4.47):

$$V(z, t) = F_1\left(t - \frac{z}{v}\right) + F_2\left(t + \frac{z}{v}\right) \tag{4.49}$$

$$I(z, t) = \frac{1}{Z_0}\left[F_1\left(t - \frac{z}{v}\right) + F_2\left(t + \frac{z}{v}\right)\right] \tag{4.50}$$

The most useful function for $F_1$ and $F_2$ is the exponential function $\exp[j(\omega t \pm \beta z)]$, where $\beta = \omega/v$. The term $Z_0$ is the same characteristic impedance in Eq. (4.41), which for the telegrapher's equations is

$$Z_0 = \sqrt{\frac{L}{C}} = Lv = \frac{1}{Cv} \tag{4.51}$$

The variables $L$ and $C$ are given in terms of Henries and Farads per unit length and are thus distinguished from $L$ and $C$ used in lumped element circuit theory.

## 4.5 THE TRANSMISSION LINE EQUATION

The transmission line equation was determined in Section 4.3 for a cascade of lumped element matched circuits. It is the input impedance of a transmission line terminated with a load, $Z_L$, and it can also be found directly from analysis of a transmission line itself. The transmission line is characterized by its mechanical length, $L$, and its characteristic impedance, $Z_0$. The characteristic impedance of a transmission line is a function only of the geometry and dielectric constant of the material between the lines and is independent of its terminating impedances. The input impedance of the transmission line depends on $L$, $Z_0$, and $Z_L$. When terminated with a nonmatching impedance, a standing wave is set up in the transmission line where the forward- and backward-going voltages and currents are as indicated in Fig. 4.10. At the load,

$$V_L = V^+ + V^- \tag{4.52}$$

$$I_L = I^+ - I^- \tag{4.53}$$

Since the forward current wave is $I^+ = V^+/Z_0$ and the reverse current wave is $I^- = V^-/Z_0$, the current at the load is

$$I_L = \frac{V^+ - V^-}{Z_0} = \frac{V_L}{Z_L} \tag{4.54}$$

Replacing $V_L$ above with Eq. (4.52), the voltage reflection coefficient can be determined:

$$\Gamma = \frac{V^-}{V^+} = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{4.55}$$

If the transmission line is lossy, the reflection coefficient is actually

$$\Gamma = \frac{V^-}{V^+} = \frac{Z_L - Z_0}{Z_L + Z_0^*} \tag{4.56}$$

The phase velocity of the wave is a measure of how fast a given phase moves down a transmission line. This is illustrated in Fig. 4.12 where $e^{j\omega t}$ time dependence is assumed. If time progresses from $t_1$ to $t_2$, then in order for $e^{j(\omega t - \beta z)}$ to have the same phase at each of these two times, the wave must progress in the forward direction from $z_1$ to $z_2$. Consequently

$$0 = \beta(z_2 - z_1) - \omega(t_2 - t_1)$$

giving the phase velocity

$$v = \frac{\Delta z}{\Delta t} = \frac{\omega}{\beta} \tag{4.57}$$

This is to be distinguished from the group velocity,

$$v_g = \frac{d\omega}{d\beta}$$

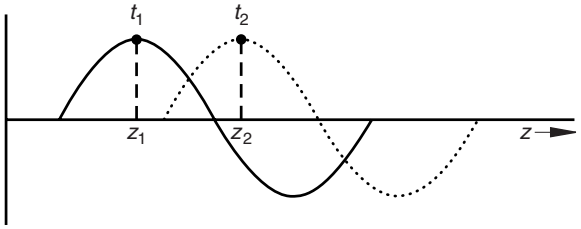



**FIGURE 4.12** Forward-directed propagating wave.

which is a measure of velocity of energy flow. For low-loss media, $v_g v = c^2/\varepsilon$, where $c$ is the velocity of light in a vacuum. The negative-going wave of course has a phase velocity of $-\omega/\beta$.

This traveling wave corresponds to the solution of the lossless telegrapher's equations. The total voltage at any position, $z$, along the transmission line is the sum of the forward- and backward-going waves:

$$V(z) = V^+ e^{-j\beta z} + V^- e^{+j\beta z} \tag{4.58}$$

The total current at any point $z$ is by Kirchhoff's law the difference of the of the two currents:

$$I(z) = \frac{1}{Z_0}(V^+ e^{-j\beta z} - V^- e^{+j\beta z}) \tag{4.59}$$

At the input to the line where $z = -\ell$, the ratio of Eqs. (4.58) and (4.59) gives the input impedance:

$$Z = Z_0 \frac{V^+ e^{-j\beta z} + V^- e^{+j\beta z}}{V^+ e^{-j\beta z} - V^- e^{+j\beta z}} \tag{4.60}$$

$$= Z_0 \frac{e^{-j\beta z} + \Gamma e^{+j\beta z}}{e^{-j\beta z} - \Gamma e^{+j\beta z}} \tag{4.61}$$

At the position $z = -\ell$,

$$Z_{\text{in}} = Z_0 \frac{Z_L + jZ_0 \tan \beta\ell}{Z_0 + jZ_L \tan \beta\ell} \tag{4.62}$$

If the propagation constant is the complex quantity $\gamma = \alpha + j\beta$, then

$$Z_{\text{in}} = Z_0 \frac{Z_L + jZ_0 \tanh \gamma\ell}{Z_0 + jZ_L \tanh \gamma\ell} \tag{4.63}$$

A few special cases illustrates some basic features of the transmission line equation. If $z = 0$, $Z_{\text{in}}(0) = Z_L$ no matter what $Z_0$ is. If $Z_L = Z_0$, then $Z_{\text{in}}(z) = Z_0$ no matter what $z$ is. For a quarter wavelength line, $Z_{\text{in}}(z = \lambda/4) = Z_0^2/Z_L$. The input impedance for any length of line can be readily calculated from Eq. (4.62), or by using the Smith chart.

## 4.6 THE SMITH CHART

The Smith chart, as shown in Fig. 4.13, is merely a plot of the transmission line equation on a set of polar coordinates. The reflection coefficient is really an alternate way of expressing the input impedance relative to some standard value ($Z_0$), which is typically 50 Ω. The reflection coefficient, Γ, has a magnitude

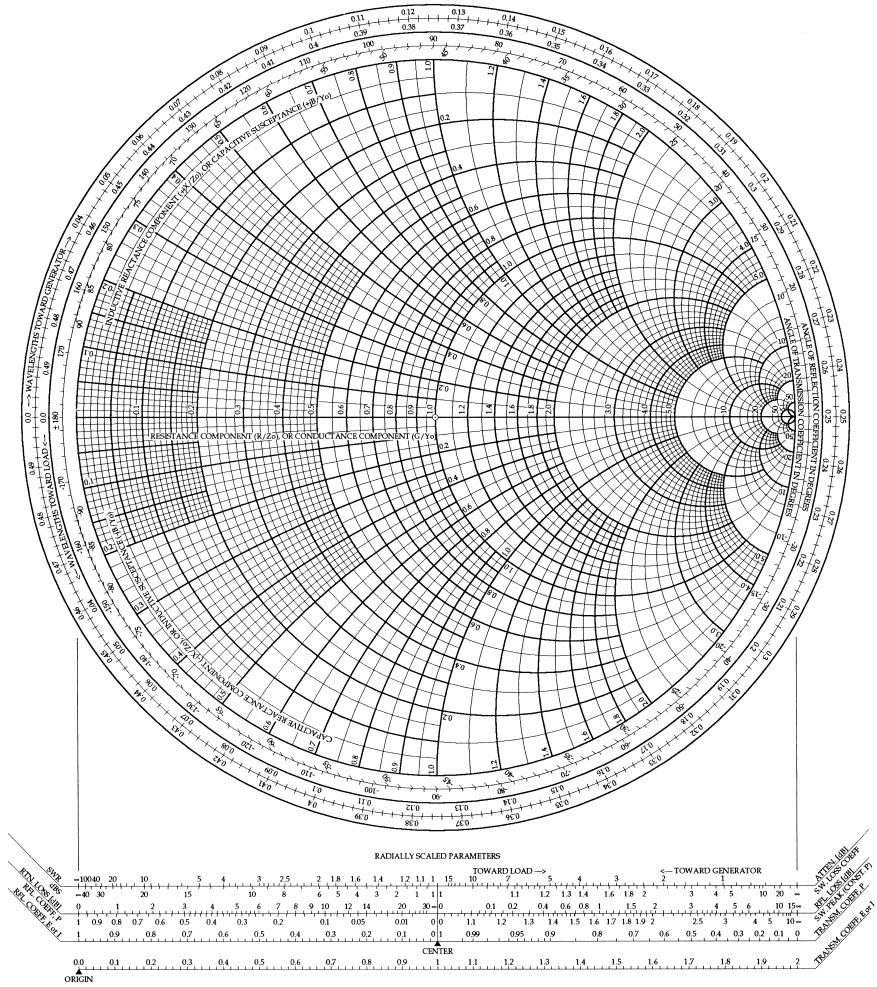



**FIGURE 4.13** The Smith chart.

between 0 and 1 and a phase angle between 0° and 360°. The equations describing the radii and centers of the circles of the coordinates of the Smith chart are found by solving the normalized version of (4.60):

$$\zeta = r + jx = \frac{Z_{\text{in}}}{Z_0} = \frac{1 + \Gamma e^{-2j\beta\ell}}{1 - \Gamma e^{-2j\beta\ell}} \tag{4.64}$$

Solution of the real part of Eq. (4.64) gives the center of the resistance circles as $(r/(1 + r), 0)$ with a radius of $1/(1 + r)$. Solution of the imaginary part gives the center of the reactance circles as $(1, 1/x)$ with a radius of $1/x$ [2, pp. 121–129].

The Smith chart can be used as a computational tool, and it often gives insight where straight equation solving will not. It is also a convenient plotting tool of measured or calculated data, since any passive impedance will fall within its boundaries.

## 4.7 COMMONLY USED TRANSMISSION LINES

Because TEM transmission lines have neither an electric nor a magnetic field component in the direction of propagation, the characteristic impedance can be found from Eq. (4.51) and electrostatics. Since the velocity of propagation in the given media is presumably known all that is necessary is to calculate the electrostatic capacitance between the conductors. When the geometry is particularly nasty and the solution is needed quickly, the field-mapping approach described in Chapter 2 can be used.

### 4.7.1 Two-Wire Transmission Line

The two-wire transmission line, commonly used, for example, between a TV antenna and the receiver, consists of two round conductors each with a radius of $a$ and separated by a distance $b$ (Fig. 4.14). The dielectric surrounding the wires has a dielectric constant of $\varepsilon$. The field theory analysis, such as that given in [2], shows that the characteristic impedance of the two wire line is

$$Z_0 = \frac{\eta}{\pi}\text{Arccosh}\left(\frac{b}{2a}\right) \tag{4.65}$$

where

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \tag{4.66}$$

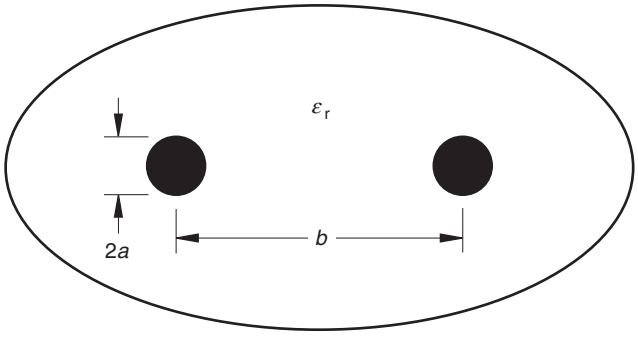



**FIGURE 4.14** The two-wire transmission line.

While the field analysis for a given structure may bring some challenges, the good news is that once $Z_0$ is known, the rest of the problem can be solved from circuit theory. Losses in the two-wire transmission line stem from the lossy dielectric between the conductors and the resistive losses experienced by the current as it flows along the conductor. Since a wave is attenuated as it goes down a line by $\exp(-\alpha z)$, the power loss is proportional to $\exp(-2\alpha z)$, where

$$\alpha = \alpha_{\mathrm{d}} + \alpha_{\mathrm{c}} \tag{4.67}$$

The dielectric and conductor losses are

$$\alpha_{\mathrm{d}} = \frac{\sigma_{\mathrm{d}}\eta}{2} \tag{4.68}$$

$$\alpha_{\mathrm{c}} = \frac{1}{2a}\sqrt{\frac{\omega\mu}{2\sigma_{\mathrm{c}}}}\frac{1}{\eta\,\mathrm{Arccosh}(b/2a)} \tag{4.69}$$

where $\sigma_{\mathrm{d}}$ and $\sigma_{\mathrm{c}}$ are the conductivities of the dielectric and conductor, respectively. The two-wire line is inexpensive and widely used in UHF applications.

### 4.7.2 Two Parallel Plate Transmission Line

The parallel plate transmission line consists of two separate conductors of width $b$ and separated by a distance $a$ (Fig. 4.15). This is a rectangular waveguide without the side walls. It is fundamentally distinct from the rectangular waveguide, which is not a TEM transmission line. The Maxwell equations for a plane wave of this system are an exact analog to the telegrapher's equations:

$$\frac{\partial E_x}{\partial z} = -\mu\frac{\partial H_y}{\partial t} \tag{4.70}$$

$$\frac{\partial H_y}{\partial z} = -\varepsilon\frac{\partial E_x}{\partial t} \tag{4.71}$$

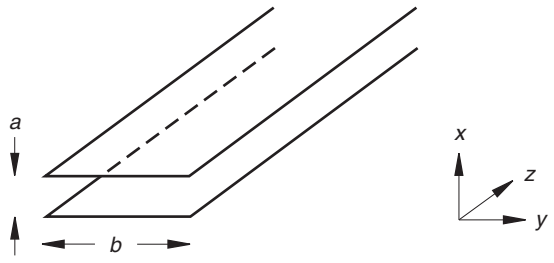



**FIGURE 4.15** The parallel plate transmission line.

The voltage between the plates is the integral of the electric field:

$$V = -\int_0^a E_x dx = -aE_x \tag{4.72}$$

The magnetic field in the $y$ direction will produce a current in the conductor that will travel in the $z$ direction according to Ampère's law:

$$I = -\int_0^b H_y d\,y = -H_y b \tag{4.73}$$

Substitution of Eqs. (4.72) and (4.73) into Eq. (4.44) gives

$$\frac{\partial E_x}{\partial z} = -\frac{Lb}{a}\frac{\partial H_y}{\partial t} \tag{4.74}$$

Comparison of this with Eq. (4.70) indicates that

$$L = \frac{\mu a}{b} \tag{4.75}$$

A similar substitution of Eqs. (4.72) and (4.73) into Eq. (4.45) and comparison with Eq. (4.71) indicates that

$$C = \frac{\varepsilon b}{a} \tag{4.76}$$

so that the characteristic impedance for the parallel plate guide is

$$Z_0 = \sqrt{\frac{L}{C}} = \frac{a}{b}\sqrt{\frac{\mu}{\varepsilon}} \tag{4.77}$$

### 4.7.3 Coaxial Transmission Line

Coaxial transmission line comes in the form of rigid, semirigid, and flexible forms. The end view of a coaxial line, which is shown in Fig. 4.16, consists of an inner conductor and the outer conductor, which is normally grounded. The electric field points from the outer to the inner conductor, and the longitudinal current on the center conductor produces a magnetic field concentric to the inner conductor. The potential between the two conductors is a solution of the transverse form of Laplace's equation in cylindrical coordinates where there is no potential difference in the longitudinal $z$ direction. The notation for the divergence and curl operators follows that given in [3]:

$$\begin{aligned} 0 &= \nabla \nabla_t \Phi \\ 0 &= \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \Phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \Phi}{\partial \phi^2} \end{aligned} \tag{4.78}$$

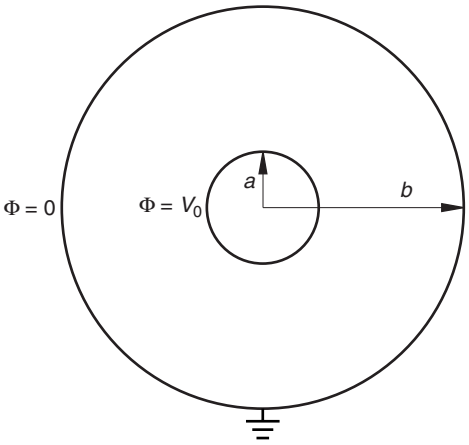



**FIGURE 4.16** The coaxial transmission line.

Because there is no potential variation in the $z$ direction, the $z$ derivative of $\Phi$ is zero. Because of symmetry there is no variation of $\Phi$ in the $\phi$ direction either. Thus Eq. (4.78) simplifies to an ordinary second-order differential equation subject to the boundary conditions that $\Phi = 0$ on the outer conductor and $\Phi = V_0$ on the inner conductor:

$$0 = \frac{1}{r}\frac{d}{dr}\left(r\frac{d\Phi}{dr}\right) \tag{4.79}$$

Integration of Eq. (4.79) twice gives

$$\Phi = C_1 \ln r + C_2 \tag{4.80}$$

which upon applying the boundary conditions gives the potential anywhere between the two conductors:

$$\Phi(r) = \frac{V_0}{\ln(a/b)} \ln\left(\frac{r}{b}\right) \tag{4.81}$$

The electric field is easily obtained by differentiation.

$$\mathbf{E} = -\nabla_t \Phi = \frac{V_0 e^{-j\beta z}}{\ln(b/a)} \frac{\hat{r}}{r} \tag{4.82}$$

The magnetic field is then

$$\begin{aligned} \mathbf{H} &= \hat{z} \times \mathbf{E} \\ &= \frac{V_0 e^{-j\beta z}}{r\eta \ln(b/a)} \hat{\phi} \end{aligned} \tag{4.83}$$

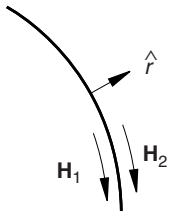



**FIGURE 4.17** The continuity of the magnetic field along the center conductor.

The outward normal vector of the center conductor, $\hat{r}$, is shown in Fig. 4.17. The surface current on the center conductor is determined by the boundary condition for the tangential magnetic field:

$$\mathbf{J}_s = \hat{r} \times (\mathbf{H}_2 - \mathbf{H}_1) = \hat{r} \times \mathbf{H}_2 \tag{4.84}$$

The later result occurs because the magnetic field is zero inside the conductor. The total current flowing in the center conductor is

$$\begin{aligned} I_0\hat{\mathbf{z}} &= \int_0^{2\pi} \frac{\hat{z}V_0}{a\eta \ln(b/a)} a\,d\phi \\ &= \frac{\hat{z}2\pi V_0}{a\eta \ln(b/a)} \end{aligned} \tag{4.85}$$

so that

$$Z_0 = \frac{V_0}{I_0} = \frac{\eta}{2\pi} \ln\left(\frac{b}{a}\right) \tag{4.86}$$

***Coaxial Dielectric Loss.*** The differential form of Ampère's law relates the magnetic field to both the conduction current and the displacement current. In the absence of a conductor,

$$\nabla \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{4.87}$$

$$\approx j\omega\varepsilon\mathbf{E} \tag{4.88}$$

By taking the curl of Eq. (4.88), the Helmholtz wave equation for $\mathbf{H}$ can be found. Solution of the wave equation would give the propagation constant, $\gamma$:

$$\gamma = j\omega\sqrt{\mu\varepsilon} \triangleq jk_0\sqrt{\varepsilon_r} \tag{4.89}$$

where $\varepsilon_r$ is the relative dielectric constant and $k_0$ is the propagation constant in free space. A lossy dielectric is typically represented as the sum of the lossless

(real) and lossy (imaginary) parts:

$$\varepsilon_{\mathrm{r}} = \varepsilon'_{\mathrm{r}} - j\varepsilon''_{\mathrm{r}} \tag{4.90}$$

The revised propagation constant is found by substituting this into Eq. (4.89). The result can be simplified by taking the first two terms of the Taylor series expansion since $\varepsilon''_{\mathrm{r}} \ll \varepsilon'_{\mathrm{r}}$:

$$\gamma = \alpha + j\beta = jk_0\sqrt{\varepsilon'_{\mathrm{r}}}\left(1 - j\frac{\varepsilon''_{\mathrm{r}}}{2\varepsilon'_{\mathrm{r}}}\right) \tag{4.91}$$

so that

$$\alpha_{\mathrm{d}} = \frac{k_0\varepsilon''_{\mathrm{r}}}{2\sqrt{\varepsilon'_{\mathrm{r}}}} \tag{4.92}$$

$$\beta = k_0\sqrt{\varepsilon'_{\mathrm{r}}} \tag{4.93}$$

The power loss is proportional to $\exp(-2\alpha z)$.

***Coaxial Conductor Loss.*** The power loss per unit length, $P_\ell$, is obtained by taking the derivative of the power at a given point along a transmission line:

$$P = P_0 e^{-2\alpha_{\mathrm{c}} z} \tag{4.94}$$

$$P_\ell \triangleq -\frac{dP}{dz} = 2\alpha_{\mathrm{c}} P \tag{4.95}$$

For a low-loss conductor where the dielectric losses are negligible, Eq. (4.87) becomes, with the help of Ohm's law,

$$\nabla \mathbf{H} = \sigma \mathbf{E} \tag{4.96}$$

where $\sigma$ is the metal conductivity. This would be the same as Eq. (4.88) if

$$\varepsilon \Rightarrow \frac{\sigma}{j\omega} \tag{4.97}$$

With this substitution the wave impedance becomes the metal surface impedance:

$$\sqrt{\frac{\mu}{\varepsilon}} \Rightarrow Z_m = (1 + j)\sqrt{\frac{\omega\mu_0}{2\sigma}} \tag{4.98}$$

At the surface there will be a longitudinal electric field of $Z_m J_{\mathrm{s}}$ directed in the $\hat{z}$ direction. Thus, in a lossy line, the fields will no longer be strictly TEM. This longitudinal electric field produces energy flow into the conductor proportional

to $E\hat{z} \times H\hat{\phi}$. This energy is dissipated in the center and outer conductor. The power loss per unit length is found in the following way:

$$\begin{aligned} P_\ell &= \frac{R_m}{2} \oint \mathbf{J}_\mathrm{s} \cdot \mathbf{J}_\mathrm{s}^* d\ell \\ &= \frac{R_m}{2} \oint (\hat{\mathbf{r}} \times \mathbf{H}) \cdot (\hat{\mathbf{r}} \times \mathbf{H}^*) d\ell \\ &= \frac{R_m}{2} \oint \mathbf{H} \cdot \mathbf{H}^* d\ell \\ &= \frac{R_m V_0^2 \pi}{\eta^2 (\ln b/a)^2} \left( \frac{1}{a} + \frac{1}{b} \right) \end{aligned} \tag{4.99}$$

The power, $P$, transmitted down the line is found by the Poynting theorem:

$$\begin{aligned} P &= \frac{1}{2} \Re\{Z_m\} \int_a^b \int_0^{2\pi} \mathbf{E} \times \mathbf{H}^* \cdot \hat{\mathbf{z}} r dr d\phi \\ &= \frac{\pi V_0^2}{\eta \ln(b/a)} \end{aligned} \tag{4.100}$$

The attenuation constant associated with conductor loss is found from Eq. (4.95):

$$\alpha_c = \frac{P_\ell}{2P} = \frac{R_m}{2\eta \ln(b/a)} \frac{a+b}{ab} \tag{4.101}$$

while the dielectric loss found earlier is

$$\alpha_\mathrm{d} = \frac{k_0 \varepsilon_\mathrm{r}''}{2\sqrt{\varepsilon_\mathrm{r}'}} \tag{4.92}$$

The total loss is found from Eq. (4.67).

### 4.7.4 Microstrip Transmission Line

Microstrip has been a popular form of transmission line for RF and microwave frequencies for some time. The microstrip line shown in Fig. 4.18 consists of a conductor strip of width $w$ on a dielectric of thickness $h$ above a ground plane. Part of the electric field between the strip and the ground plane is in the dielectric and part in the air. The field is more concentrated in the dielectric than in the air. Consequently the effective dielectric constant, $\varepsilon_\mathrm{eff}$, is somewhere between $\varepsilon_\mathrm{r}$ and 1, but closer to $\varepsilon_\mathrm{r}$ than 1. A variety of methods have been used to find $\varepsilon_\mathrm{eff}$. However, rather than provide a proof, a simple empirically based procedure for synthesizing a microstrip line will be given. Microstrip line is not strictly a TEM type of transmission line and does have some frequency

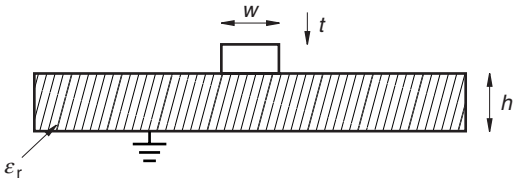



**FIGURE 4.18** The microstrip transmission line.

dispersion. Unless microstrip is being used in a multi-octave frequency range application, the TEM approximation should be adequate. The synthesis problem occurs when the desired characteristic impedance, $Z_0$, and the dielectric constant of the substrate, $\varepsilon_r$, are known and the geometrical quantity $w/h$ is to be found. The synthesis equations, given by [4], are simple, and they give an approximate solution to the microstrip problem:

$$\frac{w}{h} = \frac{8\sqrt{\dfrac{A}{11}\left(7+\dfrac{4}{\varepsilon_r}\right)+\dfrac{1}{0.81}\left(1+\dfrac{1}{\varepsilon_r}\right)}}{A} \tag{4.102}$$

$$A = \exp\frac{Z_0\sqrt{\varepsilon_r+1}}{42.4} - 1 \tag{4.103}$$

The analysis equations given below by [5] are more accurate than the synthesis equations. The value given by Eqs. (4.102) and (4.103) provides an initial value for $w/h$ that can be used in an iterative procedure to successfully solve the synthesis problem. This process depends on knowing $\varepsilon_r$ and the conductor thickness, $t$. The solution results in the effective dielectric constant, $\varepsilon_{eff}$, needed to determine electrical line lengths and $Z_0$. The procedure for the analysis procedure is shown below:

$$\Delta u_a = \frac{t/h}{\pi}\ln\left[1+\frac{4\exp(1)}{t/h+\coth^2\sqrt{6.517w/h}}\right] \tag{4.104}$$

$$\Delta u_r = \frac{1}{2}\left[1+\frac{1}{\cosh\sqrt{\varepsilon_r-1}}\right]\Delta u_a \tag{4.105}$$

$$u_a = \frac{w}{h} + \Delta u_a \tag{4.106}$$

$$u_r = \frac{w}{h} + \Delta u_r \tag{4.107}$$

$$Z_{0a}(x) = \frac{\eta}{2\pi}\ln\left[\frac{f(x)}{x}+\sqrt{1+\left(\frac{2}{x}\right)^2}\right] \tag{4.108}$$

$$f(x) = 6 + (2\pi - 6)\exp[(-30.666/x)^{0.7528}] \tag{4.109}$$

$$\varepsilon_e(x, \varepsilon_r) = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + \frac{10}{x}\right)^{-a(x)b(\varepsilon_r)} \tag{4.110}$$

$$a(x) = 1 + \frac{1}{49}\ln\left[\frac{x^4 + (x/52)^2}{x^4 + 0.432}\right] + \frac{1}{18.7}\ln\left[1 + \left(\frac{x}{18.1}\right)^3\right] \tag{4.111}$$

$$b(\varepsilon_r) = 0.564\left[\frac{\varepsilon_r - 0.9}{\varepsilon_r + 3}\right]^{0.053} \tag{4.112}$$

From the given value of $t$ and trial solutions of $w/h$, Eqs. (4.104) to (4.107) give unique values for $u_a$ and $u_r$. The characteristic impedance and effective dielectric constant are obtained using Eqs. (4.108) through (4.112):

$$Z_0\left(\frac{w}{h},\ t,\ \varepsilon_r\right) = \frac{Z_{0a}(u_r)}{\sqrt{\varepsilon_e(u_r, \varepsilon_r)}} \tag{4.113}$$

$$\varepsilon_{eff}\left(\frac{w}{h},\ t,\ \varepsilon_r\right) = \varepsilon_e\left[\frac{Z_{0a}(u_a)}{Z_{0a}(u_r)}\right]^2 \tag{4.114}$$

Since $w/h$ increases when $Z_0$ decreases, and vice versa, one very simple and effective method for finding the new approximation for $w/h$ is done by using the following ratio:

$$\left(\frac{w}{h}\right)_{i+1} = \left(\frac{w}{h}\right)_i \frac{\text{calculated } Z_0 \text{ from Eq. (4.113)}}{\text{desired } Z_0} \tag{4.115}$$

This procedure has been codified in the program MICSTP. The effect of dielectric and conductor loss has been found [6,7,8]:

$$\alpha_c = \begin{cases} = 0.159A\dfrac{R_m[32 - u_r^2]}{hZ_0[32 + u_r^2]}, & \dfrac{w}{h} \le 1 \\ 7.02 \cdot 10^{-6}A\dfrac{R_m Z_0 \varepsilon_{eff}}{h}\left[u_r + \dfrac{0.667u_r}{u_r + 1.444}\right], & \dfrac{w}{h} \ge 1 \end{cases} \tag{4.116}$$

$$A = 1 + u_r\left[1 + \frac{1}{\pi}\ln\left(\frac{2B}{t}\right)\right] \tag{4.117}$$

$$B = \begin{cases} h, & \dfrac{w}{h} \ge \dfrac{\pi}{2} \\ 2\pi w, & \dfrac{w}{h} \le \dfrac{\pi}{2} \end{cases} \tag{4.118}$$

The value for the loss caused by the dielectric is

$$\alpha_d = \frac{\varepsilon_r}{2\sqrt{\varepsilon_{eff}}}\frac{\varepsilon_{eff} - 1}{\varepsilon_r - 1}\frac{k_0\varepsilon_r''}{\varepsilon_r'} \tag{4.119}$$

The definitions for $R_m$, $\varepsilon_r'$, and $\varepsilon_r''$ were given previously as Eqs. (4.98) and (4.90). There are a variety of other transmission line geometries that could be studied, but these examples should provide information on the most widely used forms. There are cases where one line is located close enough to a neighboring line that there is some electromagnetic coupling between them. There are cases where interactions between discontinuities or nearby structures that would preclude analytic solution. In such cases solutions can be found from $2\frac{1}{2}$ and 3 dimensional numerical Maxwell equation solvers.

## 4.8 SCATTERING PARAMETERS

This chapter began with a discussion of five ways of describing a two-port circuit in terms of its terminal voltages and currents. In principle, any one of these is sufficient. One of these uses *h* parameters and was popular in the early days of the bipolar transistor, since they could be directly measured for a transistor. For a similar reason, scattering parameters, or *S* parameters, were found convenient to use by RF and microwave engineers because the circuits could then be directly measured in terms of them at these frequencies. Scattering parameters represent reflection and transmission coefficients of waves, a quantity that can be measured directly at RF and microwave frequencies. However, these wave quantities can be directly related to the terminal voltages and currents, so there is a relationship between the scattering parameters and the *z*, *y*, *h*, *g*, and *ABCD* parameters.

Consider a one-port circuit excited with a voltage source $E_G$ with an internal impedance, $Z_G$, as shown in Fig. 4.19. The quantity *a* will represent the wave entering into the port. The quantity *b* will represent the wave leaving the port. Both of these quantities are complex and can be related to the terminal voltage and

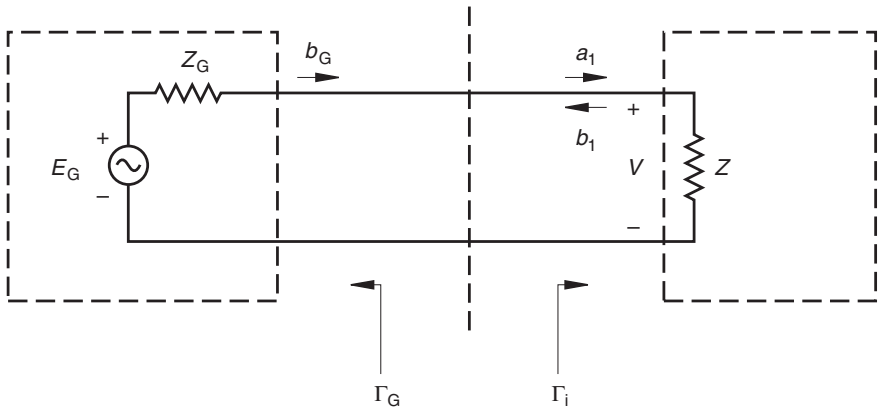



**FIGURE 4.19** Wave reflections from an unmatched generator source.

current. The generator and the load are characterized by a reflection coefficient, $\Gamma_G$ and $\Gamma_i$, respectively. The wave $b_G$ from the generator undergoes multiple reflections until finally the reflected wave from the load, $b_1$, is obtained:

$$\begin{aligned} b_1 &= \Gamma_i b_G + (\Gamma_i b_G)\Gamma_G\Gamma_i + (\Gamma_i^2\Gamma_G b_G) + \cdots \\ &= b_G\Gamma_i[1 + \Gamma_i\Gamma_G + (\Gamma_i\Gamma_G)^2 + \cdots] \\ &= \frac{b_G\Gamma_i}{1 - \Gamma_i\Gamma_G} \end{aligned} \tag{4.120}$$

This last expression is the sum of a geometric series. Since $\Gamma_i = b_1/a_1$, an expression for the wave entering into the load can be found:

$$a_1 = \frac{b_G}{1 - \Gamma_G\Gamma_i} \tag{4.121}$$

The power actually delivered to the load is then

$$P_1 = \tfrac{1}{2}(|a_1|^2 - |b_1|^2) \tag{4.122}$$

From the definition of $\Gamma_i$ and Eq. (4.121), the delivered power can be found:

$$P_1 = \frac{1}{2}\left(\frac{|b_G|^2}{|1 - \Gamma_G\Gamma_i|^2}\right)(1 - |\Gamma_i|^2) \tag{4.123}$$

and when matched so that $\Gamma_i = \Gamma_G^*$:

$$P_1 = P_a = \frac{1}{2}\frac{|b_G|^2}{1 - |\Gamma_G|^2} \tag{4.124}$$

The latter is the available power from the source. Similar expressions could be found for an $n$-port circuit.

Now the wave values will be related to the terminal voltages and currents. With reference to Fig. 4.19, Ohm's law gives

$$E_G = Z_G I + V \tag{4.125}$$

A forward-going voltage wave, $V^+$, can be related to the forward-going current by

$$V^+ = Z_G^* I^+ \tag{4.126}$$

This is based on the reason that if $Z = Z_G^*$, $V = V^+$, because $V^- = 0$. Similarly

$$V^- = Z_G I^- \tag{4.127}$$

The forward-going voltage and current can be expressed by means of Ohm's law as follows:

$$V^+ = \frac{Z_G^* E_G}{Z_G + Z_G^*} = \frac{Z_G^* E_G}{2\Re\{Z_G\}} \tag{4.128}$$

$$I^+ = \frac{E_G}{2\Re\{Z_G\}} \tag{4.129}$$

These voltages and currents represent rms values, so the incident power is

$$\begin{aligned} P_{inc} &= \Re\{V^+ I^{+*}\} = \frac{|E_G|^2}{4\Re\{Z_G\}} \\ &= \frac{|V^+|^2 \cdot \Re\{Z_G\}}{|Z_G|^2} \end{aligned} \tag{4.130}$$

The incident power, $P_{inc}$, is proportional to $|a|^2$, and the reflected power, $P_{ref}$, to $|b|^2$. Taking the square root of a number to get a complex quantity is, strictly speaking, not possible mathematically unless a choice is made regarding the phase angle of the complex quantity. This choice is related to choosing $Z_G^*$ for $a$ and $Z_G$ for $b$:

$$\begin{aligned} a &= \sqrt{P_{inc}} \\ &= \frac{V^+ \sqrt{\Re\{Z_G\}}}{Z_G^*} \end{aligned} \tag{4.131}$$

$$= \frac{I^+ Z_G^* \sqrt{\Re\{Z_G\}}}{Z_G^*} \tag{4.132}$$

and for $b$,

$$\begin{aligned} b &= \sqrt{P_{ref}} \\ &= \frac{V^- \sqrt{\Re\{Z_G\}}}{Z_G} \end{aligned} \tag{4.133}$$

$$= \frac{I^- Z_G \sqrt{\Re\{Z_G\}}}{Z_G} \tag{4.134}$$

From Eqs. (4.131) through (4.134) the forward- and reverse-going voltages and currents can be found in terms of the waves $a$ and $b$. The total voltage and total current are then found:

$$V = V^+ + V^- = \frac{aZ_G^* + bZ_G}{\sqrt{\Re\{Z_G\}}} \tag{4.135}$$

$$I = I^+ - I^- = \frac{a - b}{\sqrt{\Re\{Z_G\}}} \tag{4.136}$$

These are now in the form of two equations where $a$ and $b$ can be solved in terms of $V$ and $I$. Multiplying Eq. (4.136) by $Z_G$ and adding to Eq. (4.135) gives the result

$$V + Z_G I = \frac{2a\Re\{Z_G\}}{\sqrt{\Re\{Z_G\}}}$$

which can be solved for $a$. In similar fashion $b$ is found with the following results:

$$a = \frac{1}{2\sqrt{\Re\{Z_G\}}}(V + Z_G I) \tag{4.137}$$

$$b = \frac{1}{2\sqrt{\Re\{Z_G\}}}(V - Z_G^* I) \tag{4.138}$$

Ordinarily the generator impedance is equal to the characteristic impedance of a transmission line to which it is connected. The common way then to write Eqs. (4.137) and (4.138) is in terms of $Z_0$, which is assumed to be lossless:

$$a = \frac{1}{2\sqrt{Z_0}}(V + Z_0 I) \tag{4.139}$$

$$b = \frac{1}{2\sqrt{Z_0}}(V - Z_0 I) \tag{4.140}$$

The ratio of Eqs. (4.140) and (4.139) is

$$\frac{b}{a} = \frac{V/I - Z_0}{V/I + Z_0} = \Gamma$$

where $\Gamma$ is the reflection coefficient of the wave. The transmission coefficient is defined as the voltage across the load $V$ due to the incident voltage $V^+$:

$$T = \frac{V}{V^+} = 1 + \frac{b}{a} = 1 + \Gamma \tag{4.141}$$

This is to be contrasted with the conservation of power represented by $|T|^2 + |\Gamma|^2 = 1$.

For the two-port circuit shown in Fig. 4.20, there are two sets of ingoing and outgoing waves. These four quantities are related together by the scattering matrix:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{4.142}$$

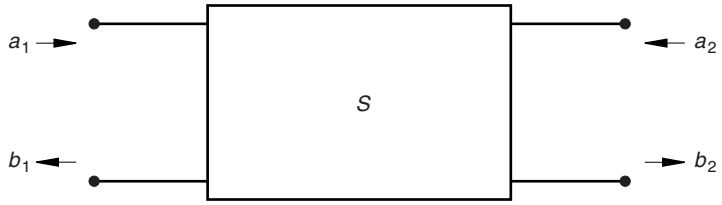



**FIGURE 4.20** The two-port with scattering parameters.

The individual $S$ parameters are found by setting one of the independent variables to zero:

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0}$$

$$S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0}$$

$$S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0}$$

$$S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0}$$

Thus $S_{11}$ is the reflection coefficient at port-1 when port-2 is terminated with a matched load. The value $S_{12}$ is the reverse transmission coefficient when port-1 is terminated with a matched load. Similarly $S_{21}$ is the forward transmission coefficient, and $S_{22}$ is the port-2 reflection coefficient when the other port is matched.

The formulas for converting between the scattering parameters and the volt–ampere relations are discussed in Section 4.1 and given in Appendix D. In each of these formulas there is a $Z_0$ because a reflection or transmission coefficient is always relative to another impedance, which in this case is $Z_0$. This is further corroborated by Eqs. (4.139) and (4.140) where the wave values are related to a voltage, current, and $Z_0$.

## 4.9 THE INDEFINITE ADMITTANCE MATRIX

Typically a certain node in a circuit is designated as being the ground node. Similarly, in an $n$-port network, at least one of the terminals is considered to be the ground node. In an $n$-port circuit in which none of the terminals is considered the reference node, it can be described by the indefinite admittance matrix. This matrix can be used, for example, when changing the $y$ parameters of a common emitter transistor to the $y$ parameters of a common base transistor. The indefinite admittance matrix has the property that the sum of the rows $= 0$ and the sum

of the columns = 0. In this way the indefinite admittance matrix can be easily obtained from the usual definite $y$ matrix, which is defined with at least one terminal connected to ground.

The derivation of this property is based on considering the currents in the indefinite circuit of Fig. 4.21. $J_k$ is the current going into terminal $k$ when all other terminals are connected to ground. It is therefore a result of independent current sources inside the $n$-port or currents resulting from initial conditions. The resulting currents going into each terminal are

$$\begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_n \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} J_1 \\ J_2 \\ \vdots \\ J_n \end{bmatrix} \tag{4.143}$$

The sum of all the equations represented by (4.143) gives the total current going into a node which by Kirchhoff's law must be zero:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} y_{ji} v_i = \sum_{k=1}^{n} i_k - \sum_{k=1}^{n} J_k = 0 \tag{4.144}$$

All the terminal voltages except the $j$th are set to 0 by connecting them to the external ground. Then, since $v_k \neq 0$, the only nonzero left-hand side of Eq. (4.144) would be

$$v_k \sum_{j=1}^{n} y_{jk} = 0. \tag{4.145}$$

Thus the sum of the columns of the indefinite admittance matrix is 0.

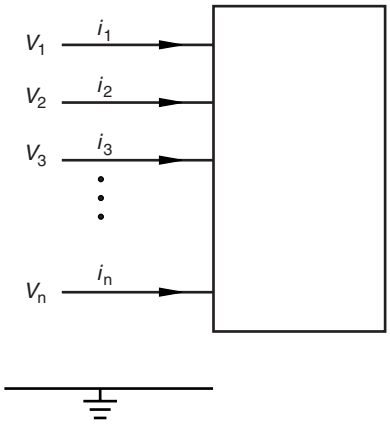



**FIGURE 4.21** An $n$-port indefinite circuit.

The sum of the rows can also be shown to be 0. If the same voltage $v_0$ is added to each of the terminal voltages, the terminal currents would remain unchanged:

$$\begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_n \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{bmatrix} \begin{bmatrix} v_1 + v_0 \\ v_2 + v_0 \\ \vdots \\ v_n + v_0 \end{bmatrix} + \begin{bmatrix} J_1 \\ J_2 \\ \vdots \\ J_n \end{bmatrix} \tag{4.146}$$

Comparison of this with Eq. (4.143) shows

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{bmatrix} \begin{bmatrix} v_0 \\ v_0 \\ \vdots \\ v_0 \end{bmatrix} = 0 \tag{4.147}$$

So that the sum of the rows is 0. A variety of other important properties of the indefinite admittance matrix are described in [1, ch. 2] to which reference should be made for further details.

One of the useful properties of this concept is illustrated by the problem of converting common source hybrid parameters of a FET to common gate hybrid parameters. This might be useful in designing a common gate oscillator with a transistor characterized as a common source device. The first step in the process is to convert the hybrid parameters to the equivalent definite admittance matrix (which contains two rows and two columns) by using the formulas in Appendix E. The definite admittance matrix, which has a defined ground, can be changed to the corresponding $3 \times 3$ indefinite admittance matrix by adding a column and a row such that $\sum$ rows $= 0$ and the $\sum$ columns $= 0$. If the $y_{11}$ corresponds to the gate and $y_{22}$ corresponds to the drain, then $y_{33}$ would correspond to the source:

$$[Y] = \begin{array}{c} \\ g \\ d \\ s \end{array} \begin{array}{c} \begin{array}{ccc} g & d & s \end{array} \\ \begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{pmatrix} \end{array} \tag{4.148}$$

The common gate parameters are found by forcing the gate voltage to be 0. Consequently the second column may be removed, since it is multiplied by the zero gate voltage anyway. At this point the second row represents a redundant equation and can be removed. In this case row 2 and column 2 are deleted, and a new common gate definite admittance matrix is formed. This matrix can then be converted to the equivalent common gate hybrid matrix.

## 4.10 THE INDEFINITE SCATTERING MATRIX

A similar property can be determined for the scattering matrix. The indefinite scattering matrix has the property that the sum of the rows $= 1$ and the sum

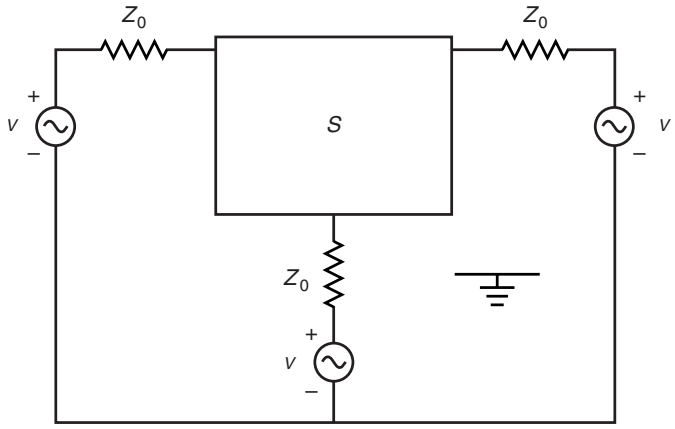



**FIGURE 4.22** The indefinite scattering parameter circuit.

of the columns = 1. For the first property the three-port shown in Fig. 4.22 is excited at all three terminals by the same voltage value. The output wave is

$$b_j = S_{j1}a_1 + S_{j2}a_2 + S_{j3}a_3, \qquad j = 1, 2, 3 \tag{4.149}$$

Under this excitation all the input waves, $a_j$, have the same amplitude, so Eq. (4.149) becomes

$$b_j = (S_{j1} + S_{j2} + S_{j3})a_1, \qquad j = 1, 2, 3 \tag{4.150}$$

Since from Eqs. (4.132) and (4.134) $a_k = \sqrt{Z_0}I_k^+$ and $b_k = \sqrt{Z_0}I_k^-$, Eq. (4.150) can be written in terms of the incident and reflected currents:

$$I_J^- = [S_{j1} + S_{j2} + S_{j3}]I_1^+ \tag{4.151}$$

When all the terminal voltages are set equal, then all the terminal currents must be zero, since there can be no voltage difference between any two ports. Thus $I_j^- = I_1^+$, which means that

$$S_{j1} + S_{j2} + S_{j3} = 1 \tag{4.152}$$

proving that the sum of the rows = 1.

To show that the sum of the columns = 1, only port-1 is excited with a voltage source. This gives $a_1 \neq 0$ and $a_2 = a_3 = 0$. By Kirchhoff's current law the sum of the currents into the three terminal circuit is zero:

$$0 = I_1 + I_2 + I_3 \tag{4.153}$$

$$= (I_1^+ - I_1^-) + (I_2^+ - I_2^-) + (I_3^+ - I_3^-) \tag{4.154}$$

Now, since $I_2^+ = I_3^+ = 0$ because of $a_2$, $a_3$,

$$I_1^+ = I_1^- + I_2^- + I_3^- \tag{4.155}$$

In addition

$$\begin{aligned} b_k &= S_{k1} a_1 \\ I_k^- &= S_{k1} I_1^+ \end{aligned} \tag{4.156}$$

so

$$I_1^+ = [S_{11} + S_{21} + S_{31}] I_1^+ \tag{4.157}$$

which affirms that the sum of the columns for the indefinite scattering matrix is 1.

## PROBLEMS

**4.1** Convert the following scattering parameters (related to 50 Ω) to *ABCD* parameters:

| $\lvert S_{11}\rvert$ | $\angle S_{11}$ | $\lvert S_{21}\rvert$ | $\angle S_{21}$ | $\lvert S_{12}\rvert$ | $\angle S_{12}$ | $\lvert S_{22}\rvert$ | $\angle S_{22}$ |
|---|---|---|---|---|---|---|---|
| 0.49 | −29 | 3.25 | 85 | 0.10 | 65 | 0.65 | −33 |

**4.2** Given the $S$ parameters, derive the $z$ parameters.

**4.3** Two transmission lines are cascaded together. Transmission line 1 has a characteristic impedance of $Z_{01} = 50\ \Omega$, has a length of 30/8 cm, and is terminated on the right-hand side by a resistive load of 25 Ω. The left-hand side is connected to transmission line 2 whose characteristic impedance $Z_{02} = 30\ \Omega$, and its length is $\lambda/4$ at 1 GHz. What is the input impedance at the left-hand side of the 30 Ω line?

**4.4** The transmission line circuit of length $\ell$ and characteristic impedance $Z_0$ is terminated by a resistance $R_L$. Determine the $Q$ for this circuit at the first appropriate nonzero frequency.

## REFERENCES

1. W.-K. Chen, *Active Network and Feedback Amplifier Theory*, New York: McGraw-Hill, 1980.
2. S. R. Seshadri, *Fundamentals of Transmission Lines and Electromagnetic Fields*, Reading MA: Addison-Wesley, 1971, pp. 335–350.
3. C. T. Tai, *Generalized Vector and Dyadic Analysis*, Piscataway, NJ: IEEE Press, 1997.
4. H. A. Wheeler, "Transmission-Line Properties of a Strip on a Dielectric Sheet on a Plane," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-25, pp. 631–647, 1977.

5. E. Hammerstad and O. Jensen, "Accurate Models for Microstrip Computer-Aided Design," *1980 IEEE MTT-S International Microwave Symp. Digest*, pp. 407–409, 1980.
6. K. C. Gupta, R. Garg, and R. Chadha, *Computer Aided Design of Microwave Circuits*, Dedham, MA: Artech House, 1981.
7. R. A. Pucel, D. J. Massé, and C. P. Hartwig, "Losses in Microstrip," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-16, pp. 342–350, 1968.
8. R. A. Pucel, D. J.Massé, and C. P. Hartwig, "Corrections to 'Losses in Microstrip,' "*IEEE Trans. Microwave Theory Tech.*, Vol. MTT-16, p. 1064, 1968.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER FIVE

# Filter Design and Approximation

## 5.1 INTRODUCTION

The subject of network synthesis became highly refined in the twentieth century due to the work of some very capable electrical engineers. To work through that body of knowledge would be beyond the present purpose of reviewing filter synthesis techniques. This chapter will therefore concentrate on basic techniques and solutions to the filter approximation problem and implementations that have come about by special requirements. At this point in time, operational amplifiers have not made a significant inroad to the RF frequency range, so active filters will not be treated here. However, active filter design is discussed by Chen [1]. Filters and impedance matching circuits are vital in the design of transistor amplifiers. The particular values of impedances needed for an amplifier depend on the device type and its orientation. Formulas for these impedances are summarized in Appendix D.

The solutions to the filter approximation problem and implementations that have arisen because of special requirements are treated in this chapter. The treatment is further limited to filters that perform certain tasks on continuous time varying analog signals by way of linear time-invariant circuit elements. Such filters are called *analog filters*. These analog filter functions are frequency selective circuits, and they can be analyzed in the time domain or in the frequency domain using Fourier transforms and Laplace transforms.

## 5.2 IDEAL AND APPROXIMATE FILTER TYPES

The function of a filter is to separate different frequency components of the input signal that passes through the filter network. The characteristics of the network are specified by a transfer function, $H(j\omega)$ or $H(s)$, where $s = +j\omega$ represents

the complex frequency defined for the Laplace transform. The transfer function is the ratio of output signal to input signal, voltage, or current:

$$H(j\omega) = \frac{V_{\text{out}}}{V_{\text{in}}} \tag{5.1}$$

$$= |H(j\omega)|e^{j\phi(\omega)} \tag{5.2}$$

The transfer phase function, $\phi(\omega)$, is related to the transfer group delay through a differential with respect to frequency as follows:

$$\tau_d(\omega) = -\frac{d\phi(\omega)}{d\omega} \tag{5.3}$$

For constant group delay, the phase function must be linear with frequency. In most filters only the magnitude of the transfer function is of interest. However, in modern-day systems using signals with complex modulation schemes, phase and group delay functions are also important.

A filter network passes some of the input signal frequencies and stops others, and being a linear circuit, this function is performed without adding or generating new frequency components. The frequency band that passes, ideally without losses (0 dB insertion loss), defines the pass band, and the band that stops the frequencies, ideally with infinite loss, is called the stop band. Figure 5.1 shows this loss representation of the ideal low-pass filter. It is a low pass because it passes all low-frequency signals from dc to some high frequency, $\omega_c$ and stops all signals above $\omega_c$. The frequency, $\omega_c$, is called the cutoff frequency of the filter. An ideal low-pass filter is physically not realizable as this requires a circuit with an infinite number of elements due to an abrupt change from pass band to stop band. Such a change is not practical.

This raises a practical issue of how does one specify the filters? The concept of the transition band, $\omega_p$ to $\omega_s$, is the frequency range that separates the pass band and stop band where the loss transitions from a minimum to a maximum value. This is shown in Fig. 5.2. The ratio of $\omega_s/\omega_p$ is sometimes referred to as filter selectivity, ratio, or filter steepness. As the selectivity approaches one,

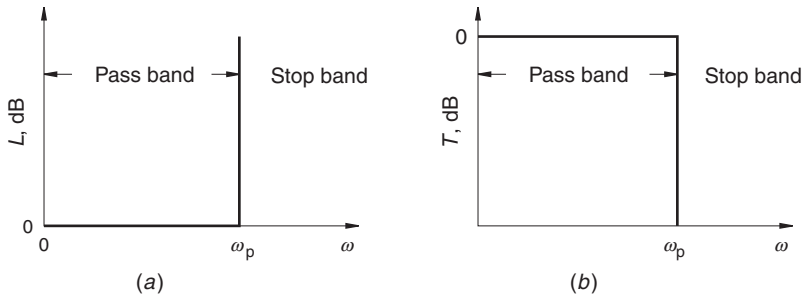



**FIGURE 5.1** Loss or attenuation characteristics of an ideal low-pass filter.

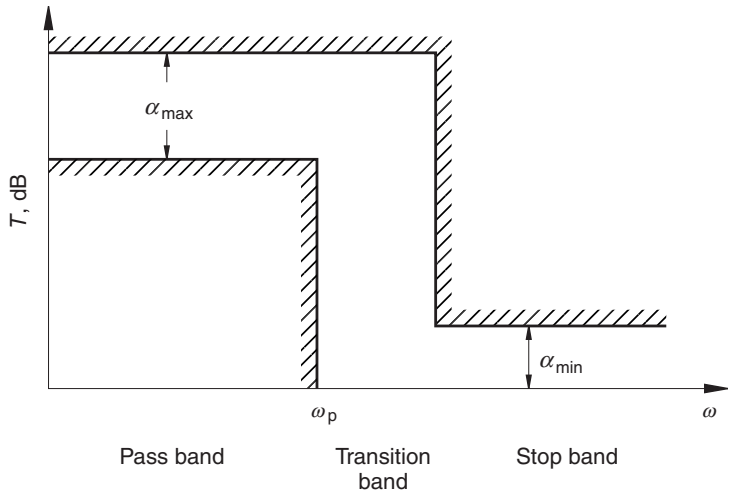



**FIGURE 5.2** Method of specifying a practical low-pass filter.

the more complex and costly the filter becomes. Similar considerations can be applied in the design of filters using phase linearity and/or group delay flatness. The concept of pass band, stop band, and transition band permits specifications of five major types of filters: (1) low pass, (2) high pass, (3) band pass, (4) band stop, and (5) all pass. The transmission behavior of these filters is shown in Fig. 5.3.

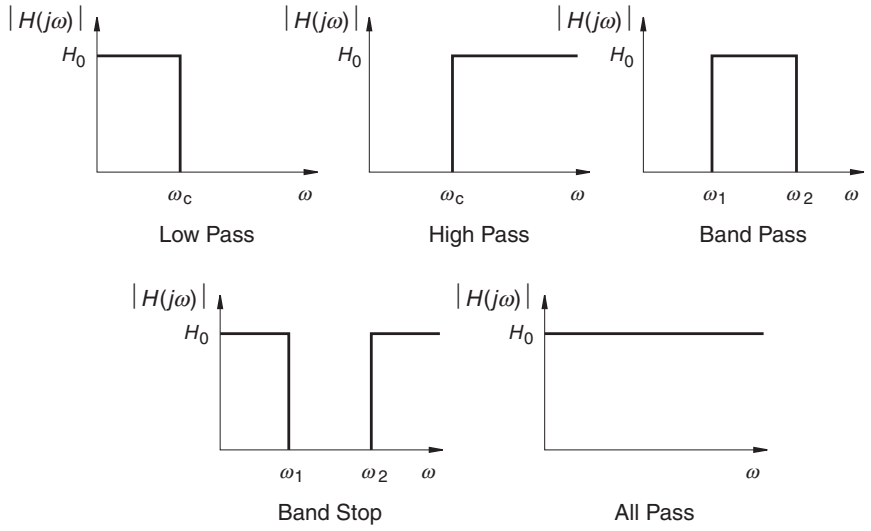



**FIGURE 5.3** The five major filter types using transfer function amplitude response.

### 5.2.1 Low-Pass Filter

Low-pass filter networks are realized by using a cascade of series inductors and shunt capacitors. The number of these elements determines the steepness of the filter with the larger number resulting in a more complex and steep filter. The penalties are the complexity of the filter, the in-band loss, higher cost, and larger size. At low frequencies, series inductances produce low impedance, and shunt capacitors produce high impedance, thus allowing the signal to appear at the output of the filter. Above the cutoff frequency, the series inductors behave as large impedances and shunt capacitors as low impedances, thereby impeding the signal transfer to the load.

### 5.2.2 High-Pass Filter

The high-pass filter shown in Fig. 5.3*b* allows signal frequencies higher than the cutoff frequency to pass through the filter to the load with a minimum loss and stops all frequencies below the cutoff frequency. This behavior is the reverse of the low-pass filter, and sometimes the high-pass filter is referred to as the complement of the low-pass filter. High-pass filter networks are realized by using a cascade of series capacitors and shunt inductors. Capacitors at high frequencies have low impedance, and inductors have high impedance. Thus the high-frequency signal passes through the filter to the output load with a minimum loss. Just the opposite happens at low frequencies, resulting in a high attenuation of the low frequencies.

### 5.2.3 Band-Pass Filter

The band-pass filter shown in Fig. 5.3*c* shows the signal is transferred to the load in a band of frequencies between the lower cutoff frequency, $\omega_{c1}$, and the upper cutoff frequency, $\omega_{c2}$. Between the lower and upper cutoff frequency is the center frequency, $\omega$, defined by the geometric mean of $\omega_{c1}$ and $\omega_{c2}$.

### 5.2.4 Band-Stop Filter

The band-stop filter is a complement of the band-pass filter and is shown in Fig. 5.3*d*. The signal in a band-stop filter is transferred to the load in two frequency bands, one from a low frequency to a low cutoff frequency, $\omega_c$, and the other from the upper cutoff frequency, $\omega_{c2}$, to infinite frequency. The signal experiences high loss between $\omega_{c1}$ to $\omega_{c2}$, hence the name band stop or band reject.

### 5.2.5 All-Pass Filter

The all-pass filter allows the signal amplitude for all frequencies to pass through the network without any significant loss (Fig. 5.3*e*). This network has no frequency selective pass band or stop band. The transmitted signal ideally experiences a linear phase shift or constant group delay with frequency.

Unfortunately, minimum phase networks do not have constant group delay: rather there are peaks near the corner frequency. All passive ladder networks, such as filters that have frequency selectivity, are minimum phase. In the design there is a trade-off between flat group delay and filter selectivity. However, a network that is nonminimum phase can be cascaded with a minimum phase network to achieve both flat group delay and selectivity. All pass networks with nonminimum phase are used as group delay compensation devices.

## 5.3 TRANSFER FUNCTION AND BASIC FILTER CONCEPTS

Before proceeding with the design of filters, it is important to understand the transfer function in the complex frequency domain, $s = \sigma + j\omega$. The transfer function can easily be transformed from the time to the frequency domain when losses are small so that $\sigma = 0$ and $s = j\omega$. As described in Section 5.2, the filter transfer function is the ratio of the output signal voltage to the input signal voltage. One could also easily select the ratio of currents. The transfer function, in general, can be written as a ratio of two polynomials:

$$H(s) = \frac{P(s)}{Q(s)} = \frac{a_0 + a_1 s + a_2 s^2 + \cdots + a_{m-1} s^m}{b_0 + b_1 s + b_2 s^2 + \cdots + b_{n-1} s^n} \tag{5.4}$$

where polynomials $P(s)$ and $Q(s)$ in general are of order $m$ and $n$. These polynomials are Hurwitz stable, which requires that the order of the numerator polynomial $m$ be equal to or less than the denominator polynomial $n$, $m < n$. The order of polynomial $Q(s)$ is the order of the filter as well. Polynomials $P(s)$ and $Q(s)$ can be factored and rewritten in the form

$$H(s) = \frac{(s - z_1)(s - z_2)(s - z_3) \ldots (s - z_m)}{(s - p_1)(s - p_2)(s - p_3) \ldots (s - p_n)} \tag{5.5}$$

The values $z_1, z_2, z_3, \ldots, z_m$, are called the zeros of the transfer function, or simply transmission zeros. The roots of $Q(s)$, $p_1, p_2, p_3, \ldots, p_n$, are the poles of the transfer function. The poles and zeros can be real or complex, but complex poles and zeros must occur in conjugate pairs. That is, if $-2 + j3$ is a pole, then $-2 - j3$ must be a pole as well. The magnitude plot of voltage transfer function represents the loss or attenuation of the filter circuit, and in dB is given by

$$L_{\text{dB}} = 20 \log |H(s)| \tag{5.6}$$

Poles and zeros of realizable passive networks must follow certain rules:

1. All poles of a transfer function occur in the left half $s$-plane. The left half $s$-plane includes the imaginary $j\omega$-axis.
2. Complex poles and zeros occur in complex conjugate pairs. However, on the imaginary axis, poles and zeros may exist singly.

## 5.4 LADDER NETWORK FILTERS

The class of minimum phase filters are those filters in which the zeros of the impedance function are chosen to be in the left half-plane. This is a way of saying that the topology of a low-pass filter is a ladder network. The filter is assumed to be lossless and terminated on each side by a real resistance. As indicated in Fig. 5.4, there are four possible choices for the first and last reactive elements. The final reactive element depends on whether the number of elements is even or odd. The basic procedure is to develop a design for a low-pass filter whose terminating resistors are 1 Ω and whose cutoff frequency is $\omega_c = 1$ rad/s. Once this normalized low-pass filter is designed, the impedance level is adjusted to the desired value, the cutoff frequency is adjusted, and circuit topology is transformed to a high-pass, band-pass, or band-stop filter as desired.

The notation for the low-pass filter prototype filter with $g$ values, shown in Fig. 5.4, is widely used. In this way convenient recursion formulas can be used for finding the filter values. A $g$ beside an inductor stands for inductance, a $g$ beside a capacitor stands for capacitance, and a $g$ beside a resistor stands for resistance. For the normalized prototype circuit, $g_0 = 1$.

### 5.4.1 The Butterworth Filter

The transducer power gain for a two port circuit is

$$G_T = \frac{\text{power delivered to the load}}{\text{power available from the source}} \tag{5.7}$$

which for a passive filter is a quantity $\leq 1$. A filter with many reactive elements would be expected to more closely approximate an ideal filter with rectangular shape (infinitely steep band edge skirts and flat-pass band) than one with few reactive elements. For a filter with $n$ poles ($n$ reactive elements), the low-pass Butterworth approximation provides the maximum flatness in its pass band near

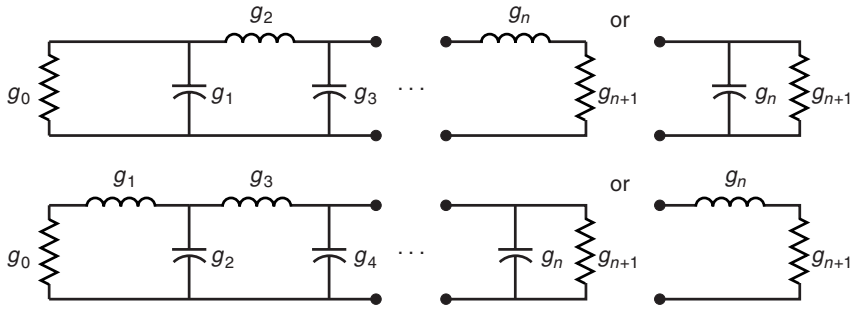



**FIGURE 5.4** The lumped element prototype low-pass filter.

$\omega = 0$. The gain function for this type of filter is given by

$$|H(j\omega)|^2 = G_{\mathrm{T}} = \frac{H_0}{1 + (\omega/\omega_{\mathrm{c}})^{2n}} \tag{5.8}$$

where $H_0 \leq 1$. The first $2n - 1$ derivatives of the denominator of this function are all zero at $\omega = 0$, implying that it is maximally flat. The poles of this function all have a magnitude of 1 and are separated from one another on the unit circle by $\pi/n$ radians. Furthermore there are no poles on the $j\omega$-axis. At the edge of the pass band, the filter attenuates the power by $\frac{1}{2}$ or $-3$ dB. A recursion formula for the filter elements that would produce this response can be found in a variety of references, one of which is [1]:

$$g_0 = g_{n+1} = 1 \tag{5.9}$$

$$g_k = 2 \sin\left[\frac{(2k-1)\pi}{2n}\right], \qquad k = 1, 2, 3, \ldots, n \tag{5.10}$$

A set of filter values that is worth remembering is the three-pole Butterworth filter, where $g_1, g_2, g_3 = 1, 2, 1$. The low-pass prototype starting with a shunt $C$ has $C_1 = 1$ F, $L_2 = 2$ H, and $C_3 = 1$ F.

Often minimum requirements are placed on the shape of the pass band. In this instance the minimum number of poles needed to produce a desired specification is

$$n = \frac{\log[(10^{\alpha_{\min}/10} - 1)(10^{\alpha_{\max}/10} - 1)]}{2\log(\omega_{\mathrm{s}}/\omega_{\mathrm{p}})} \tag{5.11}$$

In this expression the maximum attenuation in the pass band $0 \leq \omega \leq \omega_{\mathrm{p}}$ is $\alpha_{\max}$. The minimum attenuation in the stop band, $\omega_{\mathrm{s}} \leq \omega < \infty$, is $\alpha_{\min}$.

### 5.4.2 The Chebyshev Filter

The slope of the pass-band skirts for a given number of poles of a filter can be improved by allowing small ripples in the pass band. In antenna theory the Dolph-Chebyshev amplitude weighting of the array elements provides the minimum beam width for a specified side lobe level. Similarly in filter design the Chebyshev function provides the maximum possible bandwidth for a given pass-band ripple or the minimum possible pass-band ripple for a given bandwidth. The Chebyshev (equal ripple) low-pass filter transducer gain function is

$$|H(j\omega)|^2 = G_T = \frac{H_0}{1 + \varepsilon^2 T_n^2(\omega/\omega_{\mathrm{c}})} \tag{5.12}$$

where $\omega_{\mathrm{c}}$ is the low-pass cutoff frequency. The value $\varepsilon$ is a number $<1$ and is a measure of the pass-band ripple. The Chebyshev function, $T_n(x)$, oscillates

between +1 and −1 when its argument is less than 1. The poles of this transfer function lie on an ellipse with no $j\omega$ axis poles.

For $x > 1$, $T_n(x)$ rapidly becomes large. The Chebyshev function can be written in a form that clearly shows this characteristic:

$$T_n(x) = \cos[n \arccos(x)], \qquad 0 \le x \le 1 \tag{5.13}$$

$$T_n(x) = \cosh[n \ \text{arccosh}(x)], \qquad x > 1 \tag{5.14}$$

Since $T_n(x) < 1$ in the pass band, the pass-band transfer function is

$$\frac{1}{1+\varepsilon^2} \le |H(j\omega)|^2 \le 1$$

Outside the pass band, $T_n(x)$ increases approximately exponentially.

The Chebyshev functions can be found in terms of a polynomial of its argument from a recursion formula:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{5.15}$$

The formula begins by setting $T_0(x) = 1$ and $T_1(x) = x$. Furthermore for $n$ odd

$$T_n(0) = 0 \quad \text{and} \quad T_n(\pm 1) = \pm 1 \tag{5.16}$$

while for $n$ even

$$T_n(0) = (-1)^{n/2} \quad \text{and} \quad T_n(\pm 1) = 1 \tag{5.17}$$

The next few Chebyshev functions are shown below:

$$T_2(x) = 2x^2 - 1 \tag{5.18}$$

$$T_3(x) = 4x^3 - 3x \tag{5.19}$$

$$T_4(x) = 8x^4 - 8x^2 + 1 \tag{5.20}$$

$$T_5(x) = 16x^5 - 20x^3 + 5x \tag{5.21}$$

$$T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1 \tag{5.22}$$

If the maximum pass-band frequency is $\omega_c$ and the minimum stop-band frequency beyond which the attenuation is always greater than $\alpha_{\min}$ is $\omega_s$, then the number of poles required in the function is $n$ [1]:

$$n = \frac{\text{arccosh}\left[\dfrac{1}{\varepsilon}(10^{\alpha_{\max}/10} - 1)^{-1/2}\right]}{\text{arccosh}(\omega_s/\omega_c)} \tag{5.23}$$

Just as in the Butterworth approximation, there is a set of recursion formulas for the Chebyshev filter. Finding expressions for the $g$ values for the filter requires first expanding the Chebyshev functions by its own set of recursion formulas. The low-pass prototype filter structure (for a given number $n$ of reactive elements) is then equated to the $n$th-order filter function so that a correlation is made between the circuit and the function. Fortunately the hard work has been done in network synthesis books (e.g., [1,2]). The final recursion formulas are given below:

$$g_0 = 1 \tag{5.24}$$

$$g_{n+1} = \begin{cases} 1, & n \text{ odd} \\ \tanh^2(\beta/4), & n \text{ even} \end{cases} \tag{5.25}$$

$$g_1 = \frac{2a_1}{\gamma} \tag{5.26}$$

$$g_k = \frac{4a_{k-1}a_k}{b_{k-1}g_{k-1}}, \qquad k = 2, 3, \ldots, n \tag{5.27}$$

$$a_k = \sin\left[\frac{(2k-1)\pi}{2n}\right], \qquad k = 1, 2, \ldots, n \tag{5.28}$$

$$b_k = \gamma^2 + \sin^2\left(\frac{k\pi}{n}\right), \qquad k = 1, 2, \ldots, n \tag{5.29}$$

$$\beta = \ln\left[\coth\left(\frac{A_m}{17.32}\right)\right] \tag{5.30}$$

$$A_m = 10\ \log(\varepsilon^2 + 1) \tag{5.31}$$

$$\gamma = \sinh\left(\frac{\beta}{2n}\right) \tag{5.32}$$

One important difference between the Butterworth and Chebyshev approximations is the value for $g_{n+1}$. The unequal impedance levels for the even-order Chebyshev termination impedances is often avoided by simply restricting the choices of $n$ for the Chebyshev function to odd values. The circuit element values for these two filter functions were found by using network synthesis techniques after determining the poles of the transfer function. Other filter functions are available and are briefly mentioned here.

### 5.4.3 The Inverse Chebyshev Filter

In this filter, equal ripples are found in the stop band while the pass band is smooth. The inverse Chebyshev function is useful when it is necessary to control the attenuation over the entire stop band. The transfer function and the minimum number of poles needed to guarantee a maximum attenuation in the pass band

for the inverse Chebyshev function is given below:

$$|H(j\omega)|^2 = \frac{H_0\varepsilon^2 T_n^2(\omega_c/\omega)}{1+\varepsilon^2 T_n^2(\omega_c/\omega)} \tag{5.33}$$

$$n = \frac{\operatorname{arccosh}\left[\frac{1}{\varepsilon}(10^{\alpha_{\max}/10}-1)^{-1/2}\right]}{\operatorname{arccosh}(\omega_c/\omega_p)} \tag{5.34}$$

The actual circuit is generated using network synthesis techniques described in Section 5.6.1. To do this, the polynomial expansion of the Chebyshev functions will be needed.

### 5.4.4 The Thompson-Bessel Filter

The previous functions were designed to provide a specific magnitude in the transfer response, while the phase was left uncontrolled. The Thompson-Bessel filter is designed to provide a maximally flat time delay response. The ideal Thompson-Bessel response would have a flat magnitude response, $|H(j\omega)| = 1$, and a phase response proportional to frequency, $[\arg H(j\omega)] = -\omega T$. The normalized time delay $D$ is

$$D(\omega T) = \frac{d[\arg H(j\omega T)]}{d\omega T} \tag{5.35}$$

where $T$ is time delay of the filter. The transfer function designed to provide maximally flat time delay is

$$H(sT) = \frac{B_n(0)}{B_n(sT)} \tag{5.36}$$

Letting $y = sT$, the recursion formula for the Bessel polynomials is

$$B_n(y) = (2n-1)B_{n-1}(y) + y^2 B_{n-2}(y) \tag{5.37}$$

where $B_0(y) = 1$ and $B_1(y) = y + 1$. The first few polynomials are as follows:

$$B_2(y) = y^2 + 3y + 3 \tag{5.38}$$

$$B_3(y) = y^3 + 6y^2 + 15y + 15 \tag{5.39}$$

$$B_4(y) = y^4 + 10y^3 + 45y^2 + 105y + 105 \tag{5.40}$$

Again, once the polynomials are known, network synthesis techniques can be used to derive a low-pass prototype filter [1,2,3].

## 5.5 THE ELLIPTIC FILTER

The low-pass filter can be characterized as having a pass band from $\omega = 0$ to $\omega = \omega_p$ with an attenuation no greater than $H_0$ plus a small ripple. In addition it is characterized as having a stop band from $\omega_s$ to $\infty$ with an insertion loss no less than some high value, $\alpha_{min}$ (Fig. 5.5). In the Chebyshev filter the pass-band ripple is fixed to a certain maximum, but small, value while the attenuation in the stop band increases monotonically with $\omega$. The inverse Chebyshev filter produces an equal ripple in the stop band and a monotonically decreasing insertion loss for $\omega$ going from $\omega_p$ toward $\omega = 0$. The elliptic function filter equal ripple response in both the pass band and in the stop band. This design provides a way of not throwing away excess stop-band attenuation at high frequencies by allowing redistribution of the attenuation over the whole stop band. As a consequence the rate of cutoff may be increased by putting some of the transmission zeros near the pass band. The cost for having equal ripple response in both the pass band and in the stop band is a slightly more complicated circuit topology for the elliptic filter (Fig. 5.5).

There is no simple recursion formula for the design of elliptic function filters. Typically tables of values are derived numerically [4,5] and are used for the low-pass prototype filter. These tabulated values have been incorporated in a program called ELLIPTIC. In this program the desired maximum attenuation level in the pass band, minimum attenuation in the stop band, the frequencies where the pass band ends and the stop band begins, and finally the number of poles, $n$, are balanced against each other to provide an elliptic filter design. If so desired, the program will produce a SPICE net list that can be used to analyze the design. In the PSPICE version of SPICE, the voltage is plotted using V(21) or VDB(21) to display the insertion loss on a linear or log scale, respectively.

## 5.6 MATCHING BETWEEN UNEQUAL RESISTANCES

For a low-pass filter, perfect match cannot in principle be achieved when impedance matching is used. In the preceding Butterworth and Chebyshev

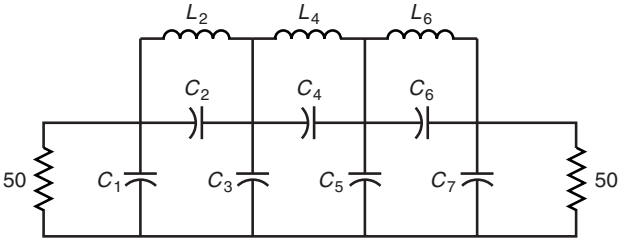



**FIGURE 5.5** A seven-pole low-pass elliptic filter topology. When $f_p = 0.8$ GHz and $f_s = 1$ GHz, $C_1 = 3.285$ pF, $C_2 = 0.547$ pF, $L_2 = 12.653$ nH, $C_3 = 5.459$ pF, $C_4 = 2.682$ pF, $L_4 = 9.947$ nH, $C_5 = 4.846$ pF, $C_6 = 2.040$ pF, $L_6 = 8.963$ nH, and $C_7 = 2.231$ pF.

functions, the constant, $H_0$, is $\leq 1$, since a passive filter cannot produce gain greater than 1. When the input and output resistance levels are equal, then $H_0$ is 1. The ratio of the load to generator resistances introduces a constraint on $H_0$. For Butterworth filters this constraint is

$$\frac{R_L}{R_G} = \left( \frac{1 + \sqrt{1 - H_0}}{1 - \sqrt{1 - H_0}} \right)^{\pm 1} \tag{5.41}$$

while for Chebyshev filters this constraint is given as follows [1]:

$$\frac{R_L}{R_G} = \left( \frac{1 + \sqrt{1 - H_0}}{1 - \sqrt{1 - H_0}} \right)^{\pm 1}, \qquad n \text{ odd} \tag{5.42}$$

$$\frac{R_L}{R_G} = \left( \frac{\sqrt{1 + \varepsilon^2} + \sqrt{1 + \varepsilon^2 - H_0}}{\sqrt{1 + \varepsilon^2} - \sqrt{1 + \varepsilon^2 - H_0}} \right)^{\pm 1}, \qquad n \text{ even} \tag{5.43}$$

One might wonder if the generator and load consisted of complex impedances, what technique might be used for matching. Without getting too involved with that issue, it is known that such matching is not always possible. The impedances must be "compatible" for matching to occur. One thing a designer can do, though, is try to incorporate the reactive part of the load into the filter as much as possible.

### 5.6.1 The Darlington Procedure

A doubly terminated filter can be designed for any physically realizable transfer function. A variety of different circuit realizations may be possible, but only one will be described. However, this particular realization method is widely used and provides practical filter design. Approximation theory determines the transfer function $|H(j\omega)|^2$ that comes closest to the ideal filter characteristics. In a lossless, low-pass circuit with possibly unequal termination resistances (Fig. 5.6), the reflected power $|\Gamma(j\omega)|^2$ can be found:

$$1 - |H(j\omega)|^2 = |\Gamma(j\omega)|^2 = \Gamma(j\omega)\Gamma(-j\omega) \tag{5.44}$$

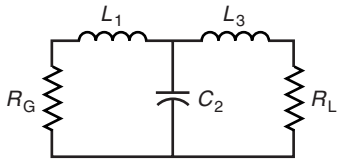



**FIGURE 5.6** Butterworth low-pass filter with unequal resistance terminations. When $R_G = 20\ \Omega$ and $R_L = 80\ \Omega$, $L_1 = 127.75$ H, $C_2 = 0.01804$ F, and $L_3 = 43.400$ H.

The final expression results from the magnitude being the product of the reflection coefficient and its complex conjugate. This can be generalized by replacing $j\omega$ with the complex frequency $s$:

$$\Gamma(s)\Gamma(-s) = 1 - |H(s)|^2 \tag{5.45}$$

The right-hand side is a known function that is given in the form of a ratio of polynomials in $s$. A requirement for realizability of an impedance or reflection coefficient is that it be positive real. All the poles of the function must lie in the left half side of the complex plane in order to avoid unrealizable growing exponentials. Half of the poles of $|\Gamma(s)|^2$ lie in the left half side and half in the right half side of the complex frequency plane. The function $\Gamma(s)$ can be extracted from $\Gamma(s)\Gamma(-s)$ by choosing only those poles in the left half side. The zeros of the function need some further consideration. The choice of which zeros to choose is more arbitrary, since there is not the same realizability restrictions on the zeros. If the choice is made to use only the left half-plane zeros, the resulting reflection coefficient and the corresponding driving point impedance is the minimum phase function. The $j\omega$-axis zeros are even multiples of complex conjugate pairs and are divided equally between $\Gamma(s)$ and $\Gamma(-s)$.

The problem of actually finding the poles and zeros requires finding the roots of the denominator and numerator polynomials. While these roots can be found analytically for the Butterworth and Chebyshev filters, the roots for other functions such as the Thompson-Bessel filter function must be found numerically. The transfer function takes the form

$$|H(s)|^2 = \frac{H_0}{F(s)} \tag{5.46}$$

where $F(0) = 1$.

The dc transfer function is

$$|\Gamma(0)|^2 = 1 - H_0 \tag{5.47}$$

$$\Gamma(0) = \pm\sqrt{1 - H_0} \tag{5.48}$$

Since at dc in a low-pass circuit the series reactive elements are short circuits and the shunt reactive elements are open circuits, the reflection coefficient is

$$\Gamma(0) = \frac{R_L - R_G}{R_L + R_G} \tag{5.49}$$

Consequently

$$H_0 = \frac{4R_L/R_G}{(R_L/R_G + 1)^2} \tag{5.50}$$

which of course equals 1 when both sides of the filter have equal terminations.

Once the reflection coefficient is determined, the Darlington synthesis procedure is used to obtain the circuit elements. The input impedance to the filter at any frequency is given in terms of the reflection coefficient:

$$Z_{\text{in}} = R_{\text{G}} \frac{1 + \Gamma(s)}{1 - \Gamma(s)} \tag{5.51}$$

The Cauer extraction technique for a ladder network can now be used. The polynomials in the numerator and denominator are arranged in descending powers of $s$. It will always be the case for a lossless transfer function that the highest power of the numerator and denominator will differ by at least 1. If the numerator is the higher-order polynomial, then an impedance pole at $s = \infty$ (i.e., a series inductor) can be extracted from the impedance function. This is done by synthetic division. The fractional remainder is now inverted, and synthetic division again carried out to extract an admittance pole at $s = \infty$ (i.e., a shunt capacitor). The process continues until only the load resistance or conductance remains.

As an example, consider a three-pole Butterworth filter with a 3 dB cutoff frequency at 1 rad/s. The input resistance $R_{\text{G}} = 20\ \Omega$ and the output resistance is $R_{\text{L}} = 80\ \Omega$. The Butterworth transfer function is therefore

$$|H(\omega)| = \frac{H_0}{1 + \omega^6} \tag{5.52}$$

where from Eq. (5.50),

$$H_0 = \frac{16}{25} \tag{5.53}$$

and

$$|\Gamma(\omega)|^2 = 1 - |H(\omega)|^2 = \frac{1 + \omega^6 - H_0}{1 + \omega^6} \tag{5.54}$$

Now replace $\omega$ with $-js$, factor the denominator into the six roots of 1, and recombine into two cubic factors where one factor contains the left half-plane roots and the other the right half-plane roots. This is the standard Butterworth polynomial:

$$\Gamma(s)\Gamma(-s) = \frac{9/25 - s^6}{(-s^3 + 2s^2 - 2s + 1)(s^3 + 2s^2 + 2s + 1)} \tag{5.55}$$

In this case the denominator is readily factored analytically, but the roots of the numerator when $H_0 \neq 1$ must be found numerically. The program POLY can provide the complex roots of a polynomial with complex coefficients. In this example all values are calculated using double precision arithmetic, though for clarity only three or four significant figures are shown.

The reflection coefficient containing only left half-plane poles and zeros is

$$\Gamma(s) = \frac{s^3 + 1.687s^2 + 1.423s + 0.599}{s^3 + 2s^2 + 2s + 1} \tag{5.56}$$

The input impedance is found from Eq. (5.51):

$$Z_{\text{in}} = 20\frac{2s^3 + 3.687s^2 + 3.423s + 1.599}{0s^3 + 0.313s^2 + 0.577s + 0.400s} \tag{5.57}$$

Extraction of the impedance pole at $s = \infty$ is done by synthetic division:

$$\begin{array}{r} \quad\quad 6.387\text{s} \\ 0.313s^2 + 0.577s + 0.400 \,\big|\, 2s^3 + 3.687s^2 + 3.423s + 1.599 \\ \quad\quad \underline{2s^3 + 3.687s^2 + 2.555s} \\ \quad\quad 0s^3 + 0s^2 + 0.868s + 1.599 \end{array}$$

The remainder is inverted, and an admittance pole at $s = \infty$ is extracted:

$$\begin{array}{r} \quad\quad 0.361s \\ 0.868s + 1.599 \,\big|\, 0.313s^2 + 0.577s + 0.400 \\ \quad\quad \underline{0.313s^2 + 0.577s} \\ \quad\quad 0s^2 + 0s + 0.868s + 0.400 \end{array}$$

By inversion again and performing synthetic division, once more another impedance pole at $s = \infty$ is removed:

$$\begin{array}{r} \quad\quad 2.170s \\ 0.400 \,\big|\, 0.868s + 1.599 \end{array}$$

The final remainder, $1.599/0.400 = 4.000$ represents the normalized load resistance, which is the expected value. Hence $L_1' = 6.387$ H, $C_2' = 0.361$ F, and $L_3' = 2.170$ H. The impedance level of the circuit is now adjusted from 1 Ω to $R_{\text{G}} = 20$ Ω by multiplying all the inductances and dividing all capacitances by 20 Ω. Thus $L_1'$ becomes $L_1 = 127.75$ H, $C_2'$ becomes $C_2 = 0.01804$ F, and $L_3'$ becomes $L_3 = 43.400$ H. The final circuit is shown in Fig. 5.6. Verification of this circuit is shown by a SPICE analysis found in Fig. 5.7. Near zero frequency the insertion loss is 0.8 or $-1.938$ dB and at 1 rad/s (0.159 Hz); the loss has increased by 3 dB.

Easier analytical methods are available for the Chebyshev filter, and these are in fact used in the Chebyshev impedance transforming circuit described in

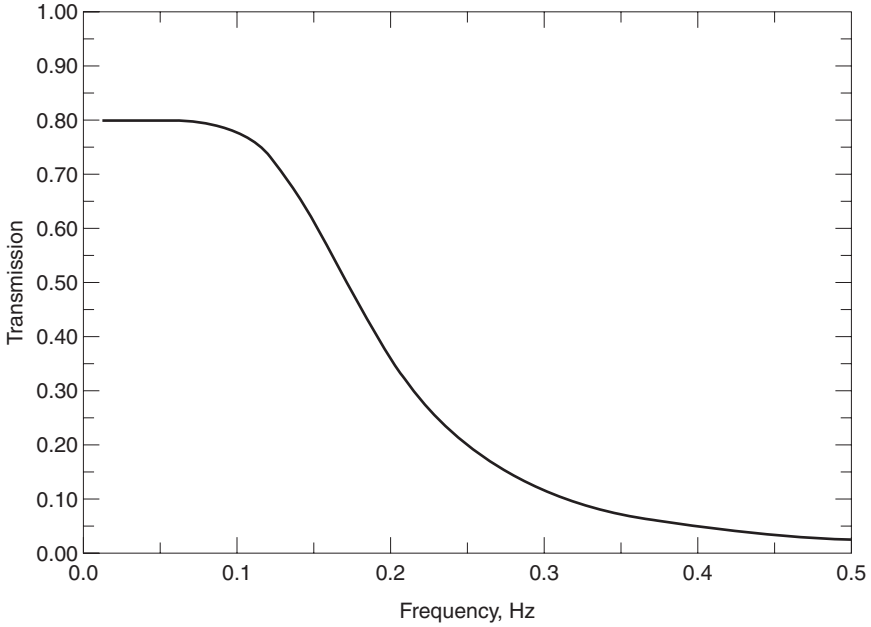



**FIGURE 5.7** Frequency response of the Butterworth low-pass filter.

Section 5.6.3. The Darlington method shown here can be used where a closed form solution for the roots is not available.

### 5.6.2 Filter Type Transformation

Filter design is based on the design of a low-pass prototype circuit whose impedance level is 1 Ω and whose low-pass cutoff frequency is $\omega_c = 1$ rad/s. If the desired impedance level is to be changed from 1 to $R_L$, then all inductors and resistors should be multiplied by $R_L$ and all capacitors should be divided by $R_L$. If the circuit elements of the low-pass prototype are denoted by a "p" subscript, then the new adjusted values can be found:

$$L = R_L L_p \tag{5.58}$$

$$C = \frac{C_p}{R_L} \tag{5.59}$$

$$R = R_L R_p \tag{5.60}$$

To adjust the cutoff frequency from 1 rad/s to $\omega_c$, the low-pass circuit elements are further modified in the following way:

$$L' = \frac{L}{\omega_c} \tag{5.61}$$

$$C' = \frac{C}{\omega_c} \tag{5.62}$$

$$R' = R \tag{5.63}$$

Transformation of the low-pass filter to a high-pass filter can be accomplished by another frequency transformation. The normalized complex frequency variable for the low-pass prototype circuit is $s_n$. On the $j\omega$-axis the pass band of the low-pass filter occurs between $\omega = -1$ and $+1$. If the cutoff frequency for the high-pass filter is $\omega_c$, then the high-pass frequency variable is

$$s = \frac{\omega_c}{s_n} \tag{5.64}$$

Applying this transformation will transform the pass-band frequencies of the low-pass filter to the pass band of the high-pass filter. This is illustrated in Fig. 5.8. The reactance of an inductor, $L$, in the low-pass filter becomes a capacitance, $C_h$, in the high-pass filter:

$$Ls_n = \frac{L\omega_c}{s} = \frac{1}{C_h s} \tag{5.65}$$

or

$$C_h = \frac{1}{L\omega_c} \tag{5.66}$$

Similarly application of the frequency transformation Eq. (5.64) will convert a capacitor in the low-pass filter to an inductor in the high-pass filter:

$$L_h = \frac{1}{C\omega_c} \tag{5.67}$$

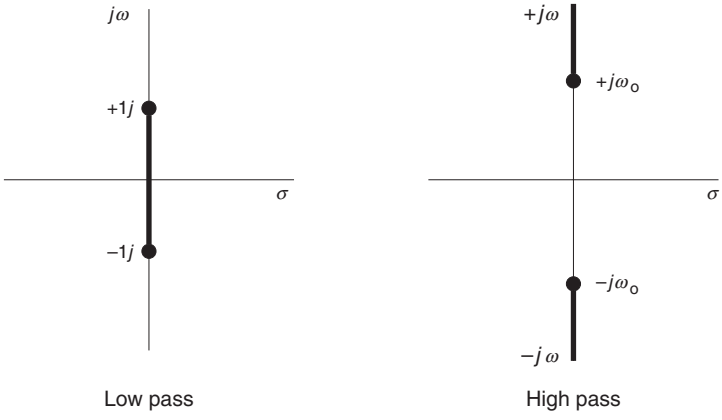



**FIGURE 5.8** Low-pass to high-pass transformation.

A band-pass filter is specified to have a pass band from $\omega_1$ to $\omega_2$. The "center" of the pass band is the geometric mean of the band edge frequencies, $\omega_0 = \sqrt{\omega_1\omega_2}$. The fractional bandwidth is $w = (\omega_2 - \omega_1)/\omega_0$. A band-pass circuit can be formed from the low-pass prototype by using a frequency transformation that will map the pass band of the low-pass filter to the pass band of the band-pass filter. The desired frequency transformation is

$$s_n = \frac{1}{w}\left(\frac{s}{\omega_0} + \frac{\omega_0}{s}\right) \tag{5.68}$$

where $s$ is the frequency variable for the bandpass circuit. To verify this expression for the $j\omega$-axis, Eq. (5.58) is rewritten as

$$\omega_n = \frac{1}{w}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \tag{5.69}$$

A short table of specific values for the normalized low-pass prototype circuit and the corresponding band-pass frequencies are shown in Table 5.1.

A graphic illustration of the frequency transformation is shown in Fig. 5.9. A consequence of this transformation is that an inductor $L$ in the low-pass prototype filter becomes a series $LC$ circuit in the band-pass circuit:

$$Ls_n = \frac{Ls}{w\omega_0} + \frac{L\omega_0}{ws} \tag{5.70}$$

Similarly a capacitance in the low-pass filter is transformed to a parallel $LC$ circuit:

$$Cs_n = \frac{Cs}{w\omega_0} + \frac{C\omega_0}{ws} \tag{5.71}$$

Finally, the low-pass to band-stop filter frequency transformation is the reciprocal of Eq. (5.68):

$$s_n = w\left(\frac{s}{\omega_0} + \frac{\omega_0}{s}\right)^{-1} \tag{5.72}$$

All these transformations from the low-pass prototype filter are summarized in Fig. 5.10.

**TABLE 5.1 Low-Pass to Band-pass Mapping**

| Bandpass $\omega$ | Low-pass $\omega_n$ |
|---|---|
| $\omega_2$ | $+1$ |
| $\omega_0$ | 0 |
| $\omega_1$ | $-1$ |
| $-\omega_1$ | $+1$ |
| $-\omega_0$ | 0 |
| $-\omega_2$ | $-1$ |

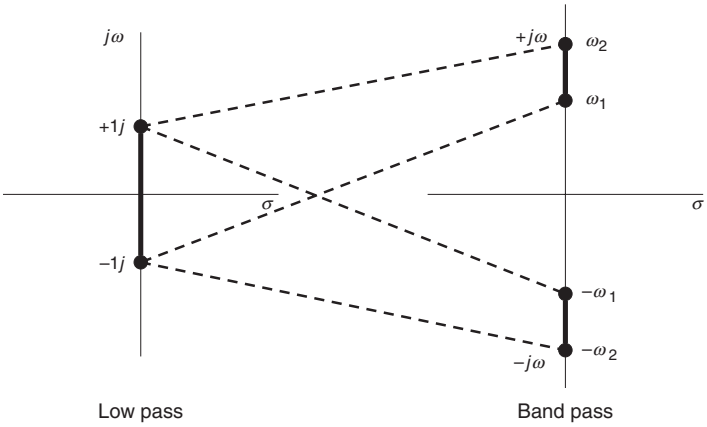



**FIGURE 5.9** Low-pass to band-pass transformation.

| Low pass | High pass | Band pass | Band stop |
|---|---|---|---|
| $L$ | $\frac{1}{\omega_c L}$ | $\frac{L}{\omega_o W}$ $\frac{w}{L\omega_o}$ | $\frac{Lw}{\omega_o}$ $\frac{1}{L\omega_o w}$ |
| $C$ | $\frac{1}{\omega_c C}$ | $\frac{C}{\omega_o W}$ $\frac{w}{\omega_o C}$ | $\frac{1}{C\omega_o w}$ $\frac{Cw}{\omega_o}$ |

**FIGURE 5.10** Filter conversion chart.

### 5.6.3 Chebyshev Bandpass Filter Example

The analytical design technique for a Chebyshev filter with two unequal resistances has been implemented in the program called CHEBY. As an example of its use, we will consider the design of a Chebyshev filter that matches a 15 to a 50 Ω load resistance. It will have $n = 3$ poles, center frequency of 1.9 GHz, a fractional bandwidth $w = (f_2 - f_1)/f_0 = 20\%$. The program CHEBY is used

to find the filter circuit elements. The program could have used the Darlington procedure described in Section 5.6.1, but instead it used the simpler analytical formulas [1]. The following is a sample run of the program:

```
Generator AND Load resistances    15.,50.
Passband ripple   (dB)    0.2
Bandpass Filter?  ⟨Y/N⟩    Y
Specify stopband attenuation OR n,   ⟨A/N⟩    N
Number of transmission poles n =    3
L(1) =  .62405E+02  C(2) =  .25125E−01  L(3) =  .36000E+02
Number of poles = 3 Ripple =  .20000E+00 dB
Center Frequency, Fo (Hz), AND Fractional Bandwidth,
   w  1.9E9,.2
Through series LC. L1( 1) =  .261370E−07 C1( 1)
   =  .268458E−12
```

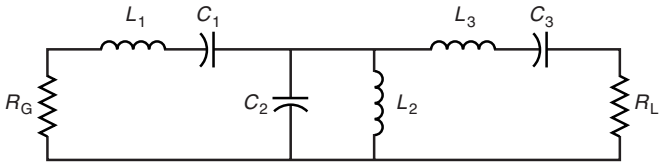



**FIGURE 5.11** A 15 : 50 Ohm Chebyshev band-pass filter, where $L_1 = 26.14$ nH, $C_1 = 0.2685$ pF, $L_2 = 0.6668$ nH, $C_1 = 10.52$ pF, $L_3 = 15.08$ nH, and $C_3 = 0.4654$ pF.

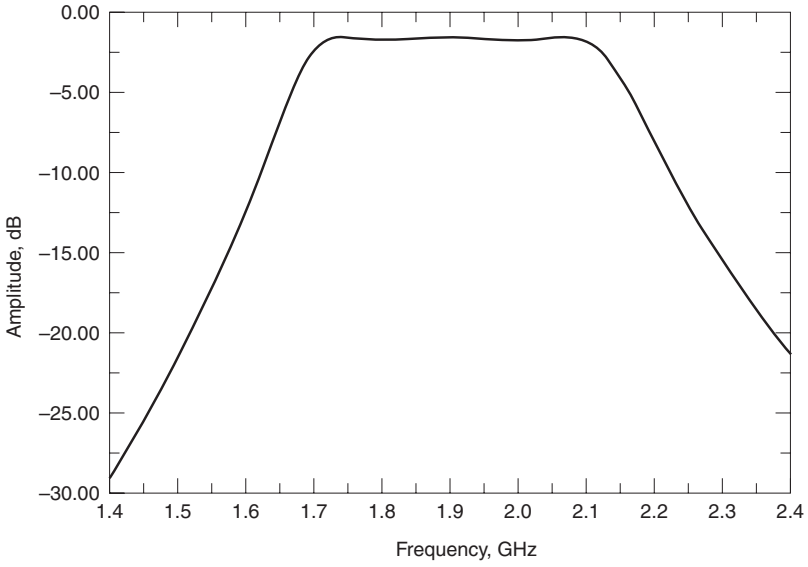



**FIGURE 5.12** SPICE analysis of a Chebyshev filter.

```
Shunt parallel LC. L2( 2) =  .666805E-09 C2( 2)
   =  .105229E-10
Through series LC. L3( 3) =  .150777E-07 C3( 3)
   =  .465370E-12
```

The resulting circuit shown in Fig. 5.11 can be analyzed using the SPICE template described in Appendix G. The results in Fig. 5.12 show that the minimum loss in the pass band is $-1.487$ dB, which corresponds to $\sqrt{H_0} \leq 0.7101$.

## PROBLEMS

**5.1** Design a band-pass filter with center frequency 500 MHz, fractional bandwidth $w = 5\%$, and pass band ripple of 0.1 dB. The out-of-band attenuation is to be 10 dB 75 MHz from the band edge. The terminating impedances are each 50 Ω. Using SPICE, plot the return loss (reflection coefficient in dB) and the insertion loss over the pass band.

**5.2** Design a band-pass filter with center frequency 500 MHz, fractional bandwidth $w = 5\%$, and pass band ripple of 0.1 dB. The out-of-band attenuation is to be 10 dB 75 MHz from the band edge, and it is to transform a 50 Ω source impedance to a 75 Ω load impedance. Using SPICE, plot the return loss (reflection coefficient in dB) and the insertion loss over the pass band.

**5.3** Design an elliptic function filter with the same specifications as in Problem 5.1, and plot the results using SPICE.

**5.4** Design a high-pass three-pole Butterworth filter with cutoff frequency of 900 MHz.

## REFERENCES

1. W.-K. Chen, *Passive and Active Filters*, New York: Wiley, 1986.
2. E. A. Guillemin, *Synthesis of Passive Networks*, New York: Wiley, 1957.
3. F. F. Kuo, *Network Analysis and Synthesis*, New York: Wiley, 1962.
4. A. Zverev, *Handbook of Filter Synthesis*, New York: Wiley, 1967.
5. H. Howe, *Stripline Circuit Design*, Norwood, MA: Artech House, 1974.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER SIX

# Transmission Line Transformers

## 6.1 INTRODUCTION

The subject matter of Chapter 3 was impedance transformation. This subject is taken up here again, but now with more careful attention given to the special problems and solutions required for RF frequency designs. The discrete element designs described previously can be used in RF designs with the understanding that element values will change as frequency changes. The alternative to discrete element circuits are transmission line circuits. The classical microwave quarter wavelength transformer can be used up to hundreds of GHz in the appropriate transmission line medium. However, at 1 GHz, a three-section quarter wavelength transformer would be a little less than a meter long! The solution lies in finding a transformation structure that may not work at 100 GHz but will be practical at 1 GHz.

The conventional transformer consists of two windings on a high-permeability iron core. The flux, $\phi$, is induced onto the core by the primary winding. By Faraday's law the secondary voltage is proportional to $d\phi/dt$. For low-loss materials, the primary and secondary voltages will be in phase. Ideal Transformers have perfect coupling and no losses. The primary-to-secondary voltage ratio is equal to the turns ratio, $n$, between the primary and secondary windings, namely $V_p/V_s = n$. The ratio of the primary to secondary current is $I_p/I_s = 1/n$. This implies conservation of power, $V_pI_p = V_sI_s$. As a consequence the impedance seen by the generator or primary side in terms of the load impedance is

$$Z_G = n^2 Z_L$$

When the secondary side of the ideal transformer is an open circuit, the input impedance of the transformer on the primary side is $\infty$.

In a physical transformer the ratio of the leakage inductances on primary and secondary sides is $L_p/L_s = n$. For the ideal transformer, $L_p$ and $L_s$ approach

$\infty$, but their ratio remains finite at $L_p/L_s = n$. The physical transformer has an associated mutual inductance, $M = k\sqrt{L_p L_s}$, where $k$ is the coupling coefficient. The leakage inductance together with the interwire capacitances limits the high-frequency response. The transmission line transformer avoids these frequency limitations.

## 6.2 IDEAL TRANSMISSION LINE TRANSFORMERS

It was found earlier, in Chapter 2, that inductive coils always come with stray capacitance. It was this capacitance that restricted the frequency range for a standard coupled coil transformer. The transmission line transformer can be thought of as simply tipping the coupled coil transformer on its side. The coil inductance and stray capacitance now form the components for an artificial transmission line whose characteristic impedance is

$$Z_0 = \sqrt{\frac{L}{C}} \tag{6.1}$$

The transmission line can be used, in principle, up to very high frequencies, and in effect it reduces the deleterious effects of the parasitic capacitance. The transmission line transformer can be made from a variety of forms of transmission lines such as a two parallel lines, a twisted pair of lines, a coaxial cable, or a pair of wires on a ferrite core. The transmission line transformer can be defined as having the following characteristics:

1. The transmission line transformer is made up of interconnected lines whose characteristic impedance is a function of such mechanical things as wire diameter, wire spacing, and insulation dielectric constant.
2. The transmission lines are designed to suppress even mode currents and allow only odd-mode currents to flow (Fig. 6.1).
3. The transmission lines carry their own "ground," so transmission lines relative to true ground are unintentional.
4. All transmission lines are of equal length and typically $< \lambda/8$.
5. The transmission lines are connected at their ends only.
6. Two different transmission lines are not coupled together by either capacitance or inductance.

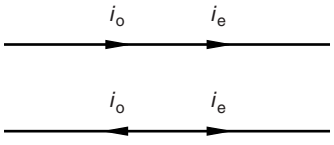



**FIGURE 6.1** A two-wire transmission line showing the odd- and even-mode currents.

7. For a short transmission line, the voltage difference between the terminals at the input port is the same as the voltage difference at the output port.

Some explanation of these points is needed to clarify the characteristics of the transmission line transformer. In property 2, for a standard transmission line the current going to the right must be equal to the current going to the left in order to preserve current continuity (Fig. 6.1). Since only odd-mode currents are allowed, the external magnetic fields are negligible. The net current driving the magnetic field outside of the transmission line is low. The third point is implied by the second. The transmission line is isolated from other lines as well as the ground. The equality of the odd mode currents in the two lines of the transmission line as well as the equivalence of the voltages across each end of the transmission line is dependent on the transmission line being electrically short in length. The analysis of transmission line transformers will be based on the given assumptions above.

As an example consider the transmission line transformer consisting of one transmission line with two conductors connected as shown in Fig. 6.2. The transformation ratio will be found for this connection. Assume first that $v_1$ is the voltage across $R_G$ and $i_1$ is the current leaving the generator resistance:

1. $i_1$ is the current through the upper conductor of the transmission line.
2. The odd-mode current $i_1$ flows in the opposite direction in the lower conductor of the transmission line.
3. The sum of the two transmission line currents at the output node is $2i_1$.
4. The voltage at the output node is assumed to be $v_o$. Consequently the voltage at left side of the lower conductor in the transmission line is $v_o$ above ground.
5. On the left-hand side, the voltage difference between the two conductors is $v_1 - v_o$.

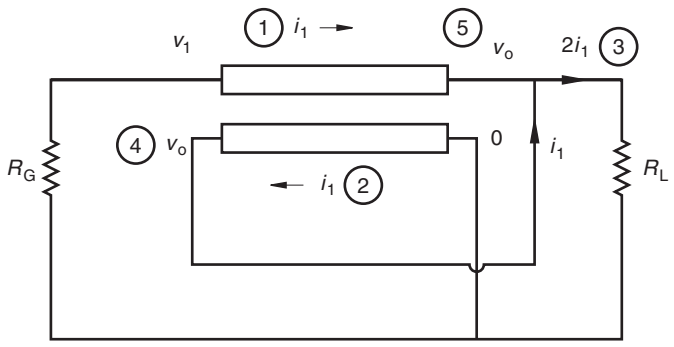



**FIGURE 6.2** Analysis steps for a transmission line transformer.

This is the same voltage difference on the right hand side. Consequently,

$$v_o - 0 = v_1 - v_o$$

$$v_o = \frac{v_1}{2}$$

If $R_G = v_1/i_1$, then

$$R_L = \frac{v_o}{2i_1} = \frac{v_1/2}{2i_1} = \frac{R_G}{4} \tag{6.2}$$

This 4 : 1 circuit steps down the impedance level by a factor of 4.

A physical connection for this transformer is shown in Fig. 6.3 where the transmission line is represented as a pair of lines. In this diagram the nodes in the physical representation are matched to the corresponding nodes of the formal representation. The transmission line is bent around to make the $B-C$ distance a short length. The transmission line, shown here as a two-wire line, can take a variety of forms such as coupled line around a ferromagnetic core, flexible microstrip line, or coaxial line. If the transformer is rotated about a vertical axis at the center, the circuit shown in Fig. 6.4 results. Obviously this results in a 1 : 4 transformer where $R_L = 4R_G$. Similar analysis to that given above verifies this result. In addition multiple two-wire transmission line transformers may be tied together to achieve a variety of different transformation ratios. An example of three sections connected together is shown in Fig. 6.5. In this circuit the current from the generator splits into four currents going into the transmission lines. Because of the equivalence of the odd-mode currents in each line, these four currents are all equal. The voltages on the load side of each line pair build up from ground to 4 $\times$ the input voltage. As a result, for match to occur, $R_L = 16R_G$.

The voltages and currents for a transmission line transformer (TLT) having a wide variety of different interconnections and numbers of transmission lines can

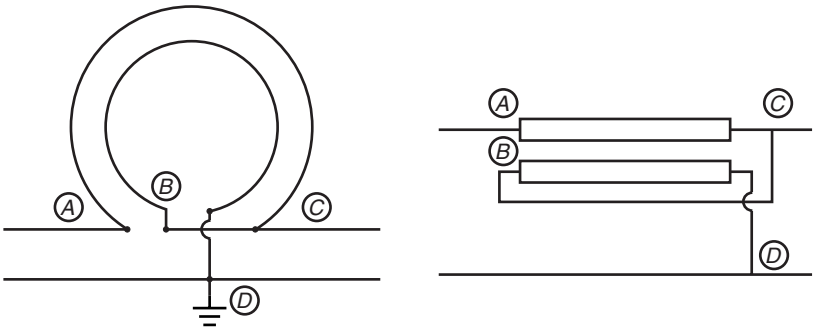



**FIGURE 6.3** A physical two-wire transmission line transformer and the equivalent formal representation.

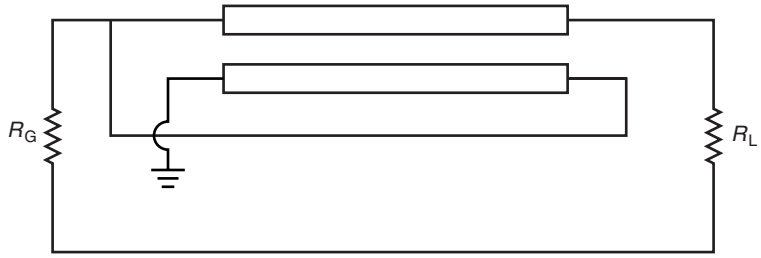



**FIGURE 6.4** An alternate transmission line transformer connection.

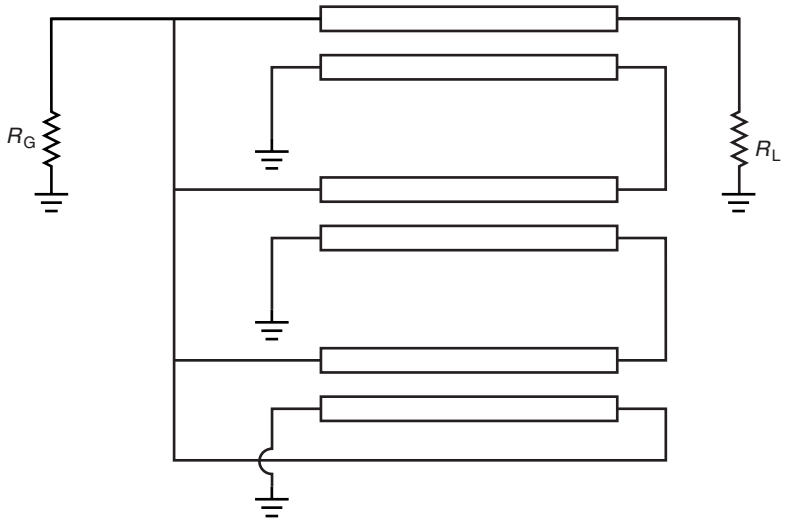



**FIGURE 6.5** A 16 : 1 transmission line transformer.


**FIGURE 6.6** Symbol for general transmission line transformer.

be represented by the simple diagram in Fig. 6.6 where $x$ and $y$ are integers. The impedance ratios, $R_G = (x/y)^2 R_L$, range from (1 : 1) for a one-transmission line circuit to (1 : 25) for a four-transmission line circuit with a total of 16 different transformation ratios [1]. A variety of transmission line transformer circuits are found in [1] and [2].

## 6.3 TRANSMISSION LINE TRANSFORMER SYNTHESIS

All the transmission lines in the transmission line transformer shown in Fig. 6.5 have their left-hand sides near the generator connected in parallel and all their right-hand sides near the load connected in series. In this particular circuit there are three transmission lines, and analysis shows that $V_{in} : V_{out} = 1 : 4$, and $R_G : R_L = 1 : 16$. The number of transmission lines, $m$, is the order of the transformer, so that when all the transmission lines on the generator side are connected in shunt and on the load side in series, the voltage ratio is $V_{in} : V_{out} = 1 : (m + 1)$. Synthesis of impedance transformations of 1 : 4, 1 : 9, 1 : 16, 1 : 25, and so on, are all obvious extensions of the transformer shown in Fig. 6.5. The allowed voltage ratios, which upon being squared, gives the impedance ratios as shown in Table 6.1. To obtain a voltage ratio that is not of the form $1 : (m + 1)$, there is a simple synthesis technique [3]. The voltage ratio is $V_{in} : V_{out} = H : L$, where $H$ is the high value and $L$ the low value. This ratio is decomposed into an $V_{in} : V_{out} = H - L : L$. If now $H - L < L$, this procedure is repeated where $H' = L$ and $L' = H - L$. This ratio is now $V_{out} : V_{in}$, which in turn can be decomposed into $H' - L' : L'$. These steps are repeated until a 1 : 1 ratio is achieved, all along keeping track which ratio that is being done, $V_{in} : V_{out}$ or $V_{out} : V_{in}$.

An example given in [3] illustrates the procedure. If an impedance ratio of $R_G : R_L = 9 : 25$ is desired, the corresponding voltage ratio is $V_{in} : V_{out} = 3 : 5$

$$\text{Step 1} \quad H : L = V_{out} : V_{in} = 5 : 3 \rightarrow (5 - 3) : 3 = 2 : 3$$

$$\text{Step 2} \quad H : L = V_{in} : V_{out} = 3 : 2 \rightarrow (3 - 2) : 2 = 1 : 2$$

$$\text{Step 3} \quad H : L = V_{out} : V_{in} = 2 : 1 \rightarrow (2 - 1) : 1 = 1 : 1$$

Now working backward from step 3, a $V_{in} : V_{out} = 1 : 2$ transmission line transformer is made by connecting two transmission lines in shunt on the input side and series connection on the output side (Fig. 6.7*a*). From step 2, the $V_{out}$ is already 2, so another transmission line is attached to the first pair in shunt on the output side and series on the input side (Fig. 6.7*b*). Finally from step 1, $V_{in} = 3$

**TABLE 6.1 Voltage Ratios for Transmission Line Transformers**

| Number of Lines | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| | 1 : 1 | 2 : 3 | 3 : 4 | 4 : 5 |
| | 1 : 2 | 1 : 2 | 3 : 5 | 5 : 7 |
| | — | 1 : 3 | 2 : 5 | 5 : 8 |
| | — | — | 1 : 4 | 4 : 7 |
| | — | — | — | 3 : 7 |
| | — | — | — | 3 : 8 |
| | — | — | — | 2 : 7 |
| | — | — | — | 1 : 5 |

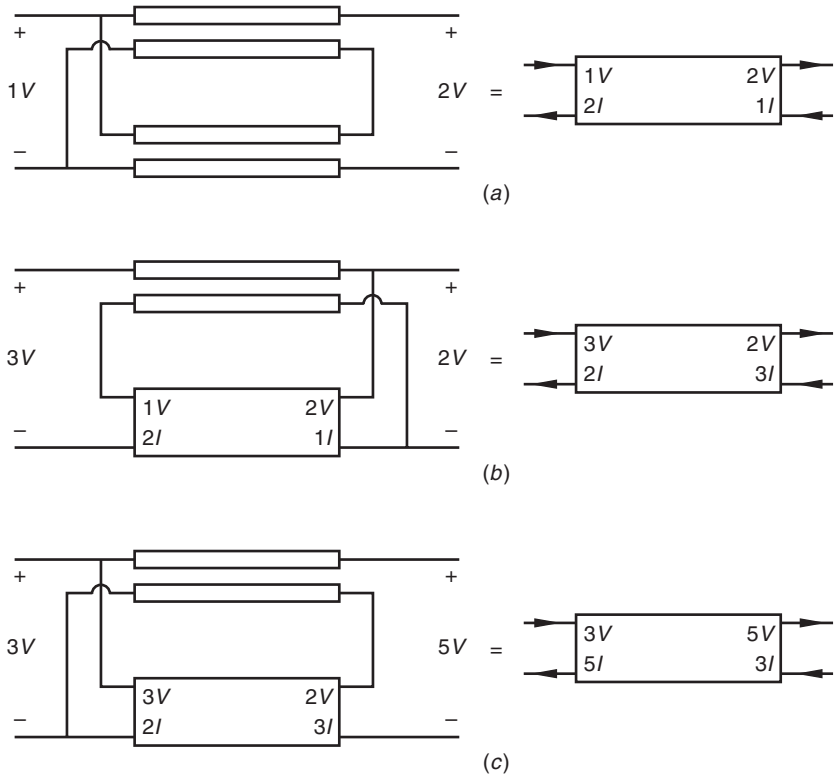



**FIGURE 6.7** Step-by-step procedure for synthesis for a desired impedance ratio.

already, so the input is connected in shunt with the another added transmission line and the outputs connected in series (Fig. 6.7*c*). The final design has $V_{\text{in}} : V_{\text{out}} = 3 : 5$ as desired.

## 6.4 ELECTRICALLY LONG TRANSMISSION LINE TRANSFORMERS

One of the assumptions given in the previous section was that the electrical length of the transmission lines was short. Because of this the voltages and currents at each end of an individual line could be said to be equal. However, as the the line becomes electrically longer (or the frequency increases), this assumption ceases to be accurate. It is the point of this section to provide a means of determining the amount of error in this assumption. Individual design goals would dictate whether a full frequency domain analysis is needed.

As was pointed out in Chapter 4, the total voltage and current on a transmission line are each expressed as a combination of the forward and backward terms (Fig. 6.8). In this case let $V_2$ and $I_2$ represent the voltage and current at the load

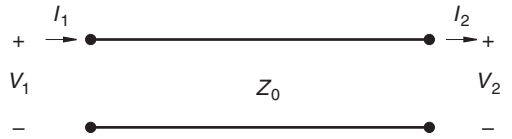



**FIGURE 6.8** An electrically long transmission line.

end, where $V^+$ and $V^-$ are the forward- and backward-traveling voltage waves:

$$V_2 = V^+ + V^- \tag{6.2}$$

$$I_2 = \frac{V^+}{Z_0} - \frac{V^-}{Z_0} \tag{6.3}$$

Assuming that the transmission line is lossless, the voltage and current waves at the input side, 1, are modified by the phase associated with the electrical length of the line:

$$V_1 = V^+ e^{j\theta} + V^- e^{-j\theta} \tag{6.4}$$

$$I_1 = \frac{V^+}{Z_0} e^{j\theta} - \frac{V^-}{Z_0} e^{-j\theta} \tag{6.5}$$

The sign associated with the phase angle, $+\theta$, for $V^+$ is used because the reference is at port 2 while a positive phase is associated with traveling from left to right. The Euler formula is used in converting the exponentials to sines and cosines. The voltage at the input, $V_1$, is found in terms of $V_2$ and $I_2$ with the help of Eqs. (6.2) and (6.3):

$$V_1 = V_2 \cos\theta + jZ_0 I_2 \sin\theta \tag{6.6}$$

Similarly $I_1$ can be expressed in terms of the voltage and current at port 2:

$$I_1 = I_2 \cos\theta + j\frac{V_2}{Z_0} \sin\theta \tag{6.7}$$

The 1 : 4 transmission line transformer shown in Fig. 6.4 is now reconsidered in Fig. 6.9 to determine its frequency response. The generator voltage can be expressed in terms of the transmission line voltages and currents:

$$V_g = (I_1 + I_2)R_G + V_1 \tag{6.8}$$

The nontransmission line connections are electrically short. Therefore the output voltage across $R_L$ is $V_o = V_1 + V_2$, and

$$V_g = (I_1 + I_2)R_G + I_2 R_L - V_2 \tag{6.9}$$

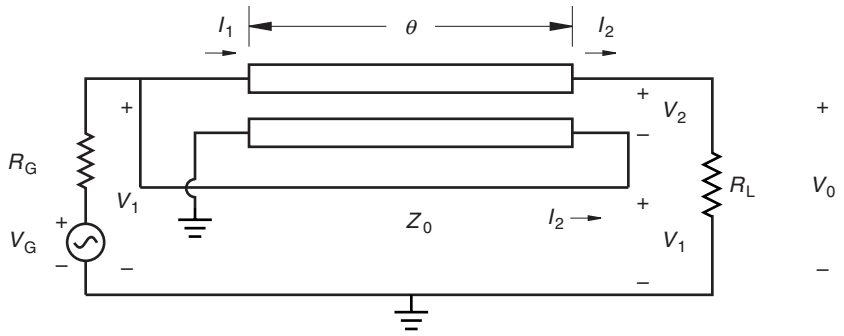



**FIGURE 6.9** An electrically long 1 : 4 transmission line transformer.

In Eqs. (6.9), (6.6), and (6.7), $V_1$ is replaced by $I_2R_L - V_2$ to give three equations in the three unknowns $I_1$, $I_2$, and $V_2$:

$$V_G = I_1R_G + I_2(R_G + R_L) - V_2 \tag{6.10}$$

$$0 = 0 + I_2(jZ_0 \sin\theta - R_L) + V_2(1 + \cos\theta) \tag{6.11}$$

$$0 = -I_1 + I_2 \cos\theta + j\frac{V_2}{Z_0} \sin\theta \tag{6.12}$$

The determinate of these set of equations is

$$\Delta = -2R_G(1 + \cos\theta) - R_L \cos\theta + j \sin\theta \left( \frac{-R_GR_L}{Z_0} - Z_0 \right) \tag{6.13}$$

and the current $I_2$ is

$$I_2 = \frac{-V_G(1 + \cos\theta)}{\Delta} \tag{6.14}$$

Consequently the power delivered to the load from the source voltage is

$$P_o = \frac{1}{2}|I_2|^2R_L = \frac{1}{2} \frac{|V_g|^2(1 + \cos\theta)^2R_L}{[2R_G(1 + \cos\theta) + R_L \cos\theta]^2 + [(R_GR_L + Z_0^2/Z_0^2]) \sin^2\theta} \tag{6.15}$$

Now the particular value of $R_L$ that guarantees maximum power transfer into the load is found by maximizing Eq. (6.15). Let $D$ represent the denominator in Eq. (6.15):

$$\frac{dP_o}{dR_L} = 0 = \frac{1}{2}|V_G|^2 \frac{(1 + \cos\theta)^2}{D} \times \left\{ 1 - \frac{R_L}{D} \left[ 2[2R_G(1 + \cos\theta) + R_L \cos\theta] \cos\theta + [\cdots] \sin^2\theta \right] \right\} \tag{6.16}$$

In the low-frequency limit where $\theta \to 0$, $R_L = 4R_G$. The optimum characteristic impedance is found by maximization $P_o$ with respect to $Z_0$, while this time keeping the line length $\neq 0$. The result is not surprising, as it is the geometric mean between the generator and load resistance:

$$Z_0 = 2R_G \tag{6.17}$$

The output power then when $Z_0 = 2R_G$ and $R_L = 4R_G$ is

$$P_o = \frac{1}{2} \frac{|V_G|^2 (1 + \cos\theta)^2}{R_G (1 + 3\cos\theta)^2 + 4R_G \sin^2\theta} \tag{6.18}$$

This reduces to the usual form for the available power when $\theta \to 0$.

More complicated transmission line transformers might benefit from using SPICE to analyze the circuit. The analysis above gives a clue to how the values of $Z_0$ and the relative values of $R_G$ and $R_L$ might be chosen with the help of a low frequency analysis.

As an example consider the circuit in Fig. 6.9 again where $R_G = 50\ \Omega$ so that $R_L = 200\ \Omega$ and $Z_0 = \sqrt{50 \cdot 200} = 100\ \Omega$, and the electrical length of the transformer is 0.4 wavelength long at 1.5 GHz. The plot in Fig. 6.10 is the return loss (= 20 log of the reflection coefficient) as seen by the generator voltage $V_G$.

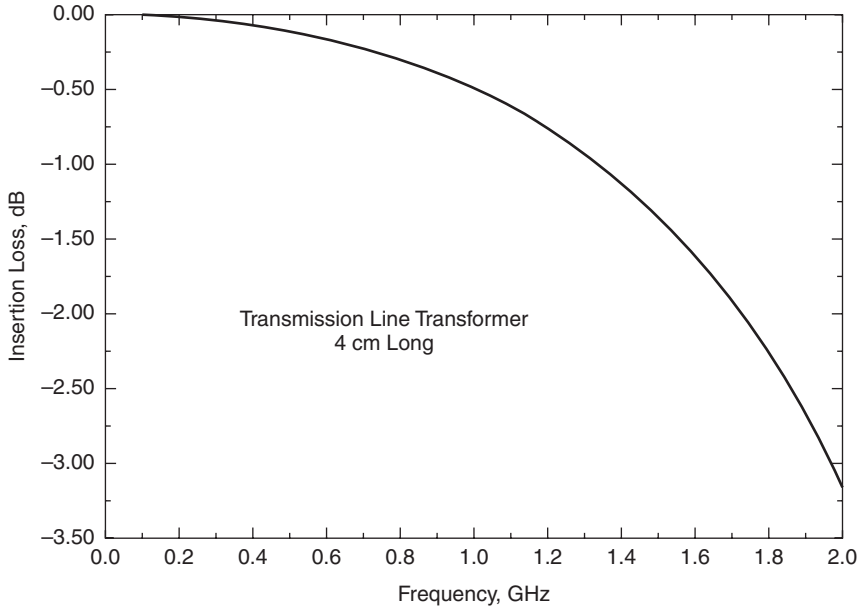



**FIGURE 6.10** Return loss for the frequency dependent transmission line transformer of Fig. 6.9.

The SPICE net list used to analyze this circuit makes use of the conversion of voltages to *S* parameters:

```
Analysis of a circuit for S11 and S21
*
* RO1 and RO2 are input and output resistance levels.
* RL is the load resistance. The load may be
  supplemented
* with additional elements.
*.PARAM RO1=50, RO2=50. RLOAD=50. IN1=-1/RO1
.PARAM RO1=50, RO2=200. RLOAD=200. IIN=-1/RO1
.FUNC N(RO1,RO2) SQRT(RO2/RO1)
RO1   1    0   RO1
VIN   10   11  AC   1
GI1   1    0   VALUE=-V(10,11)/RO1
*GI1  1    0   10  11  "-1/RO1"
E11   10   0   1   0   2
R11   11   0   1
Xcircuit   1   2   TLTCKT
RL    2    0   RLOAD
E21   21   0   VALUE=V(2)*2/N(RO1,RO2)
*  n = SQRT(RO2/RO1)
*E21   21   0   2   0   "2/n"
R21   21   0   1
*
.SUBCKT TLTCKT   1  4
* Input side
* 4 cm = .1333 wavelength at 1 GHz
TLT4  1    0   4   1   ZO=200  F=1GHz  NL=.1333
* Output side
.ENDS   TLTCKT
* Code for S11 and S21
*.AC DEC "num"  "f1"   "f2"
.AC  LIN  301  .1MEG   2GHZ
.PROBE V(11)  V(21)
.END
```

## 6.5 BALUNS

A balun (balanced–unbalanced) is a circuit that transforms a balanced transmission line to an unbalanced transmission. An example of a balanced line is the two-wire transmission line. An unbalanced line is one where one of the lines is grounded, such as in coaxial line or microstrip. One situation where this is important is in feeding a dipole antenna with a coaxial line where the antenna

is balanced and the coaxial line is unbalanced. One simple structure is shown in Fig. 6.11 where the difference between the inputs of the antenna is forced to be 180° by addition of a half wavelength line between them. At RF frequencies, a more practical way to perform this same function is to use a transmission line transformer as shown in the example of the 1 : 1 balun in Fig. 6.12*a*. There is no specified ground on the right-hand side of this circuit, but since the voltage difference on the input side is $V$, the voltage across the load must also be $V$.

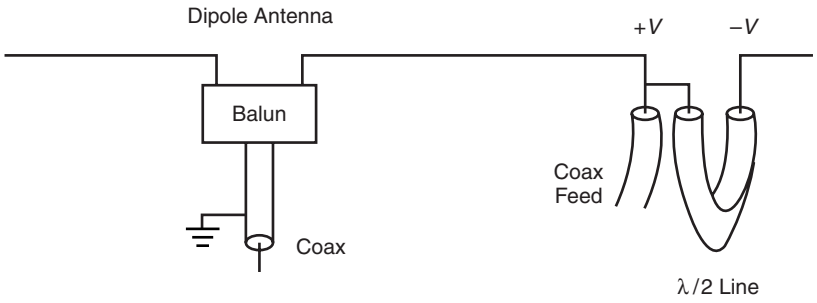
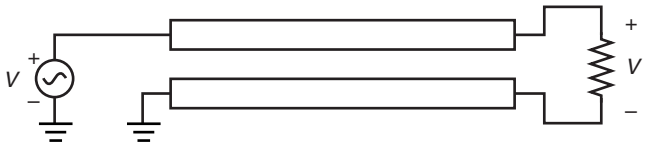
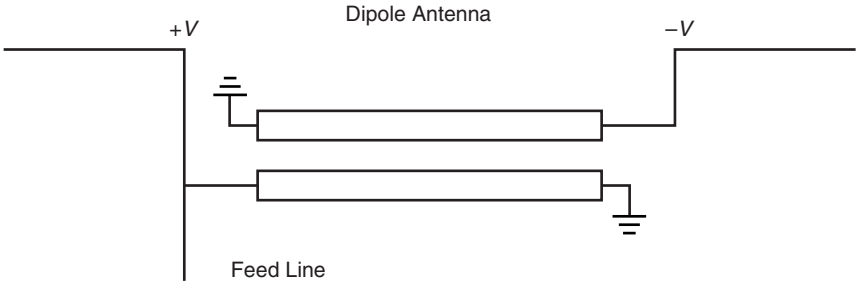



**FIGURE 6.11** Balun example used for dipole antenna.




(*a*)




(*b*)

**FIGURE 6.12** (*a*) Transmission line transformer implementation of a (1 : 1) balun, and (*b*) grounding one side gives a $+V$ and $-V$ to the two sides of the dipole antenna.

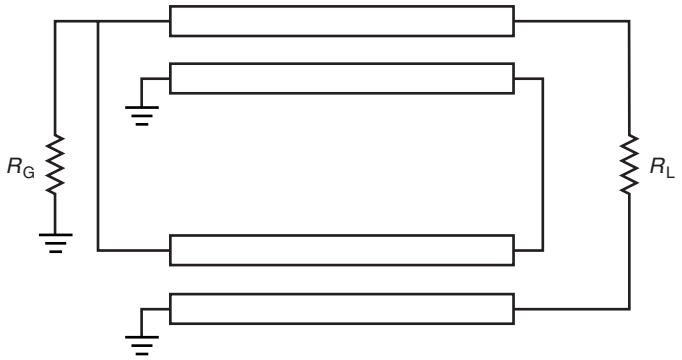



**FIGURE 6.13** A balun with a $R_G : R_L = 1 : 4$ impedance ratio.

For the dipole application, where a $+V$ is needed on one side and $-V$ on the other side, one of the output sides can be grounded as indicated in Fig. 6.12*b*. The ($R_G : R_L = 1 : 4$) balun in Fig. 6.13 shows that impedance matching and changing to a balanced line can be accomplished with a balun. Analysis of this circuit may be aided by assuming some voltage, $V_x$, at the low side of $R_L$. When the voltage at the upper side of $R_L$ is found, it also contains $V_x$. The difference between the lower and upper sides of $R_L$ removes the $V_x$.

## 6.6 DIVIDERS AND COMBINERS

Transmission lines can be used to design power dividers and power combiners. These are particularly important in design of high-power solid state RF amplifiers where the input can be split between several amplifiers or where the outputs of several amplifiers may be effectively combined into one load. A very simple two-way power divider is shown in Fig. 6.14. In this circuit $R_L = 2R_G$, and the

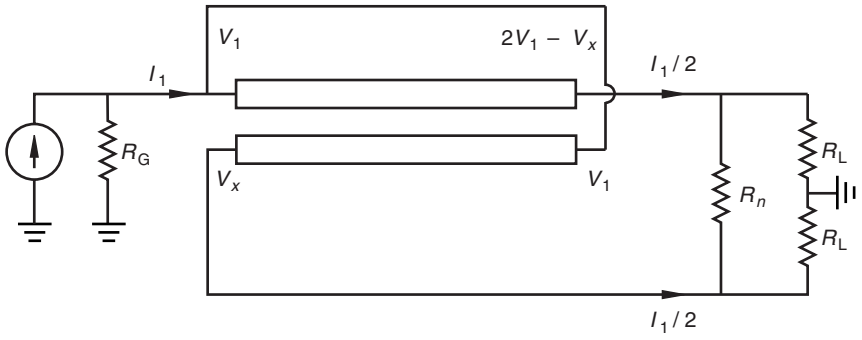



**FIGURE 6.14** A two-way power divider.

transmission line characteristic impedance is designed to be $Z_0 = \sqrt{2}R_G$. The current in $R_n$ ordinarily would be 0 because of equal voltages on either side of that resistance. Under unbalanced load conditions, $R_n$ can absorb some of the unbalanced power and thus protect whatever the load is. When the two loads are both $2R_G$. The input voltage is $V_1$ on the top conductor, and the voltage on the lower conductor is $V_x$ on the left side. On the right-hand side the lower conductor is $V_1$, and so the top conductor must be $2V_1 - V_x$ to ensure that both sides of the transmission line have the same voltage across the terminals, that is, $V_1 - V_x$. Since the current flowing through the upper load resistor and the lower load resistor must be the same, the voltage on either side of $R_n$ is the same. Consequently $2V_1 - V_x = V_x$ or $V_x = V_1$, so the voltage to current ratio at the load is

$$R_L = \frac{V_1}{I_1/2} = 2R_G \tag{6.19}$$

A two-way 180° power combiner shown in Fig. 6.15 makes use of a hybrid coupler and a balun. The resistor $R_n$ is used to dissipate power when the two inputs are not exactly equal amplitude or exactly 180° out of phase so that matched loading for the two sources is maintained. For example, consider when $I_1 = I_2$, as shown in Fig. 6.15, so that $I_1$ is entering the hybrid and $I_2$ is leaving the hybrid. The current flowing through the load, $R_L$, is $I_0$. The current flowing into the hybrid transmission line from the top is $I_1 - I_0$, while the current

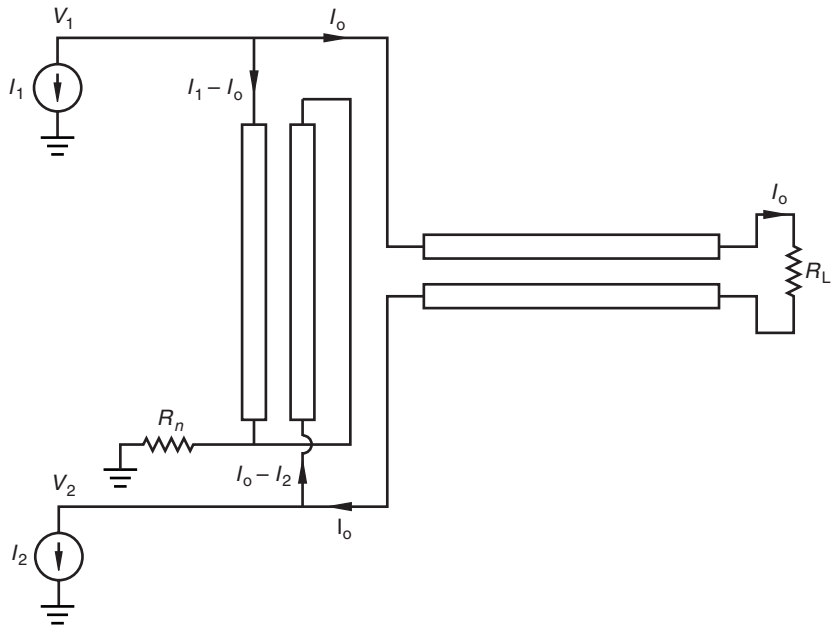



**FIGURE 6.15** A two-way 180° power combiner.

flowing from the bottom is $I_0 - I_2$. The odd-mode current in the transmission line forces is

$$I_1 - I_0 = I_0 - I_2$$

or

$$I_0 = I_1 \tag{6.20}$$

All the current goes through the balun, and no current flows through the hybrid. The current through $R_n$ is therefore 0 leading to $V_x = 0$. The voltage difference between the two ends of the transmission lines of the hybrid is the same, which implies that

$$V_1 - V_x = V_x - V_2$$

or

$$V_1 = -V_2 \tag{6.21}$$

and

$$V_0 = V_1 - V_2 = 2V_1 \tag{6.22}$$

The matching load resistance is then

$$\frac{V_0}{I_0} = R_{\mathrm{L}} = 2R_{\mathrm{G}} \tag{6.23}$$

When $I_1$ and $I_2$ are both entering the circuit so that $I_1 = -I_2$, and $V_1 = V_2$, then voltages across the top and bottom of the transmission line in the hybrid circuit of Fig. 6.15 are

$$V_1 - V_x = V_x - V_2$$

or

$$V_x = V_1 \tag{6.24}$$

The voltage across the load is $V_0 = 0$ and $I_0 = 0$. The current in the hybrid transmission line is $I_1$, so the current flowing through $R_n$ is $2I_1$:

$$R_n = \frac{V_x}{2I_1} = \frac{V_1}{2I_1} = \frac{1}{2}R_{\mathrm{G}} \tag{6.25}$$

The choices for $R_{\mathrm{L}}$ and $R_n$ ensure impedance matching for an arbitrary phase relationship between $I_1$ and $I_2$. Optimum performance would be expected if the characteristic impedances of the transmission lines were

$$Z_{0\text{-balun}} = \sqrt{2}R_{\mathrm{G}} \tag{6.26}$$

$$Z_{0\text{-hybrid}} = \frac{R_{\mathrm{G}}}{\sqrt{2}} \tag{6.27}$$

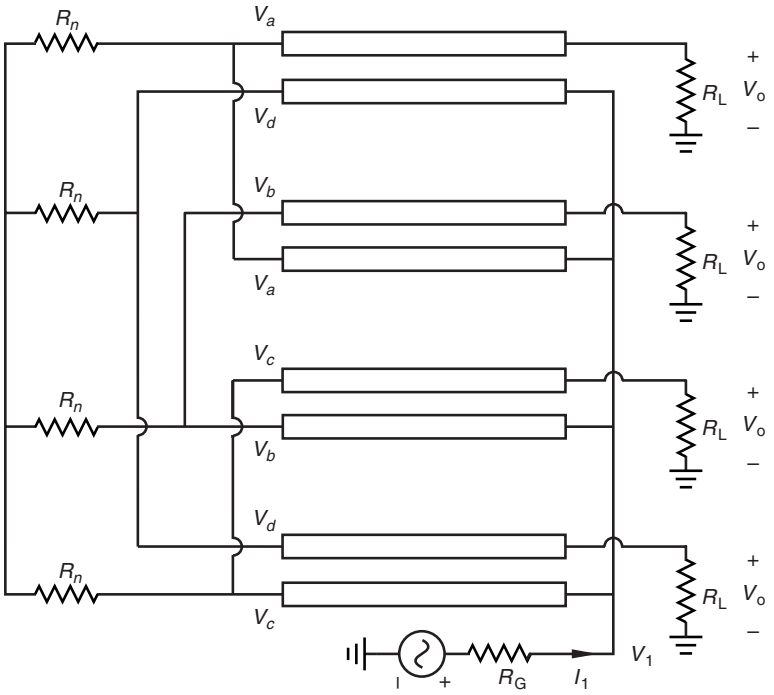



**FIGURE 6.16** A four-way power divider.

The four-way power divider illustrated in Fig. 6.16 has some similarities with the Wilkinson power divider used at microwave frequencies. In the Wilkinson divider, matching impedances between the input and output is done by choosing the quarter wavelength transmission lines to have a characteristic impedance $Z_0 = \sqrt{N}R_G$, where $N$ is the power division ratio, and $R_n = R_G$. In the present circuit, impedance matching is done using an impedance transformer at the voltage source (not shown in Fig. 6.16). If it is desired that all the output loads and voltages be equal to one another, then it follows that the current in the $R_n$ resistors is 0. This can be shown easily as follows: The voltage difference between the conductors on the right-hand side in each of the transmission lines is $V_o - V_1$. Then, for the left-hand side,

$$V_o - V_1 = V_a - V_d = V_b - V_a = V_c - V_b = V_d - V_c \tag{6.28}$$

Combining the second and third expressions, then the third and fourth expressions, and so on, leads to the following:

$$2V_a = V_b + V_d \tag{6.29}$$

$$2V_b = V_c + V_a \tag{6.30}$$

$$2V_c = V_b + V_d \tag{6.31}$$

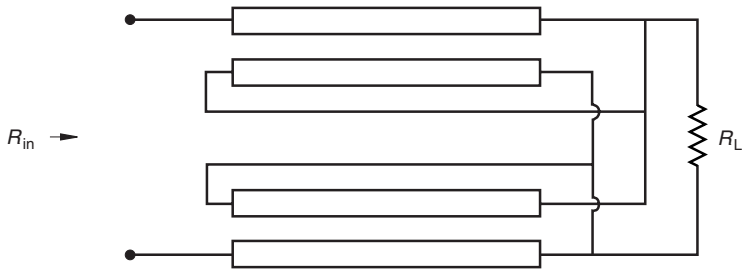



**FIGURE 6.17** Transmission line transformer for Problem 6.1.

Equations (6.29) and (6.31) clearly show that $V_a = V_c$, and Eq. (6.30) shows that $V_b = V_a$ and finally that $V_d = V_a$. This means that there is no current flowing in the $R_n$ resistors and that on the right-hand side, $V_o = V_1$. The current entering each transmission line must then be $I_1/4$, where $I_1$ is the input current from the source. The load currents are also $I_1/4$, so the impedance transformation at the input requires that $R_G = R_L/4$.

## PROBLEMS

**6.1** Indicate the direction of the currents in the transmission line transformer shown in Fig. 6.17. Determine the value of $R_{in}$ in terms of $R_{out}$.

**6.2** For the 4 : 1 transformer shown in Fig. 6.2, calculate the output power, $P_o = 1/2|I_2|^2 R_L$, where the frequency dependence of the transmission lines is used. You will have three equations in the three unknowns $I_1$, $I_2$, and $V_2$. The final answer was given by Eq. (6.15).

**6.3** Design a transmission line transformer that matches a 200 Ω load to a 50 Ω source impedance. The transmission lines are to be 4 cm long, but the transmission line characteristic impedance can be chosen to give an acceptable match by not deviating from 50 Ω by more than 25 Ω to at least 2.5 GHz. Using SPICE, plot the return loss at the input side as a function of frequency. What is the return loss at 1 GHz?

**6.4** Repeat Problem 6.3 for a transmission line transformer that matches 800 Ω to 50 Ω. The SPICE analysis should again show the return loss versus frequency. For this circuit, what is the return loss at 1 GHz?

## REFERENCES

1. P. L. D. Abrie, *The Design of Impedance-Matching Networks for Radio-Frequency and Microwave Amplifiers*, Dedham, MA: Artech House, 1985.
2. H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: Wiley, pp. 371–382, 1980.
3. E. Rotholz, "Transmission Line Transformers," *IEEE Trans. Microwave Theory Tech.* Vol. MTT-29, pp. 327–331, 1981.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER SEVEN

# Class A Amplifiers

## 7.1 INTRODUCTION

The class A amplifier is typically used as the first amplification stage of a receiver or transmitter where minimum distortion is desired. This comes with a cost of relatively low efficiency. Since the first stages in an amplifier chain handle low-power levels, the low efficiency of these amplifiers actually wastes little power. The variety of amplifier classes are described in [1] and will be covered more extensively in a later chapter. The primary properties of importance to class A amplifier design are gain, bandwidth control, stability, and noise figure. These are the topics that will be considered here.

## 7.2 DEFINITION OF GAIN [2]

In low-frequency circuits, gain is often thought of in terms of voltage or current gain, such as the ratio of the output voltage across the load to the input applied voltage. At radio frequencies it is difficult to directly measure a voltage, so typically some form of power gain is used. But once the notion of power is introduced, there are several definitions of power gain that might be used.

1. *Power gain*. This is the ratio of power dissipated in the load, $Z_L$, to the power delivered to the input of the amplifier. This definition is independent of the generator impedance, $Z_G$. Certain amplifiers, especially negative resistance amplifiers, are strongly dependent on $Z_G$.
2. *Available gain*. This is the ratio of the amplifier output power to the available power from the generator source. This definition depends on $Z_G$ but is independent of $Z_L$.
3. *Exchangeable gain*. This is the ratio of the output exchangeable power to the input exchangeable power. The exchangeable power of the source is

defined as

$$P = \frac{|V|^2}{4\Re\{Z_G\}}, \qquad \Re\{Z_G\} \neq 0 \tag{7.1}$$

For negative resistance amplifiers $P < 0$! Furthermore this definition is independent of $Z_L$.

4. *Insertion gain*. This is the ratio of output power to the power that would be dissipated in the load if the amplifier were not present. There is a problem in applying this definition to mixers or parametric upconverters where the input and output frequencies differ.
5. *Transducer power gain*. This is the ratio of the power delivered to the load to the available power from the source. This definition depends on both $Z_G$ and $Z_L$. It gives positive gain for negative resistance amplifiers as well. Since the characteristics of real amplifiers change when either the load or generator impedance is changed, it is desirable that the gain definition reflect this characteristic. Thus the transducer power gain definition is found to be the most useful.

## 7.3 TRANSDUCER POWER GAIN OF A TWO-PORT

The linear two-port circuit in Fig. 7.1 can be analyzed with the help of Fig. 7.2 and is characterized by its impedance parameters:

$$V_1 = z_{11}I_1 + z_{12}I_2 \tag{7.2}$$

$$V_2 = z_{21}I_1 + z_{22}I_2 \tag{7.3}$$

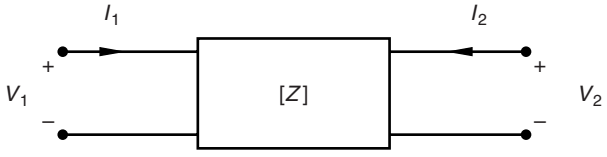



**FIGURE 7.1** Two-port circuit expressed in impedance parameters.

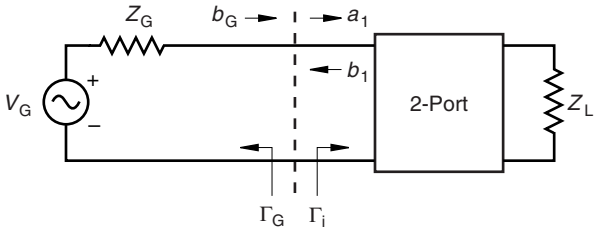



**FIGURE 7.2** Equivalent circuit to determine the input available power.

But the relationship between the port-2 voltage and current is determined by the load impedance:

$$V_2 = -I_2 Z_\mathrm{L} \tag{7.4}$$

Substitution of this for $V_2$ in Eq. (7.3) gives the input impedance. This is dependent on both the contents of the two-port itself and the load:

$$Z_\mathrm{in} = \frac{V_1}{I_1} = z_{11} - \frac{z_{12}z_{21}}{z_{22} + Z_\mathrm{L}} \tag{7.5}$$

This will be used to determine the transducer power gain. The power delivered to the load is $P_2$:

$$P_2 = \tfrac{1}{2}|I_2|^2 \Re\{Z_\mathrm{L}\} \tag{7.6}$$

Since the power available from the source is

$$P_\mathrm{1a} = \frac{|V_\mathrm{G}|^2}{8\Re\{Z_\mathrm{G}\}} \tag{7.7}$$

the transducer power gain is

$$G_\mathrm{T} = \frac{P_2}{P_\mathrm{1a}} \tag{7.8}$$

$$= \frac{4\Re\{Z_\mathrm{L}\}\Re\{Z_\mathrm{G}\}|z_{21}|^2}{|(Z_\mathrm{G} + z_{11})(Z_\mathrm{L} + z_{22}) - z_{21}z_{21}|^2} \tag{7.9}$$

Similar expressions can be obtained for $y$, $h$, or $g$ parameters by simply replacing the corresponding $z_{ij}$ with the desired matrix elements and by replacing the $Z_\mathrm{G}$ and $Z_\mathrm{L}$ with the appropriate termination. However, for radio frequency and microwave circuits, scattering parameters are the most readily measured quantities. The transducer power gain will be found in terms of the scattering parameters in the following section.

## 7.4 POWER GAIN USING S PARAMETERS

The available power, $P_a$, when the input of the two-port circuit is matched with $\Gamma_\mathrm{i} = \Gamma_\mathrm{G}^*$, was given by Eq. (4.124) in Chapter 4.

$$P_a = \frac{\frac{1}{2}|b_\mathrm{G}|^2}{1 - |\Gamma_\mathrm{G}|^2} \tag{7.10}$$

At the output side of the circuit, the power delivered to the load is given by the following:

$$P_\mathrm{L} = \tfrac{1}{2}|b_2|^2(1 - |\Gamma_\mathrm{L}|^2) \tag{7.11}$$

The transducer gain is simply the ratio of Eq. (7.11) to Eq. (7.10):

$$G_\mathrm{T} = \frac{|b_2|^2}{|b_\mathrm{G}|^2}(1 - |\Gamma_\mathrm{L}|^2)(1 - |\Gamma_\mathrm{G}|^2) \tag{7.12}$$

As this stands, $b_2$ and $b_G$ are not very meaningful. However, this ratio can be expressed entirely in terms of the known $S$ parameters of the two-port circuit. From the description of the $S$ parameters as a matrix corresponding to forward- and backward-traveling waves, the two-port circuit can be represented in terms of a flow graph. Each branch of the flow graph is unidirectional and the combination describes the $S$ matrix completely. The presumption is that the circuit is linear. The problem of finding $b_2/b_G$ can be done using either algebra or some flow graph reduction technique. The classical method developed for linear systems is Mason's nontouching loop rules. The method shown below is easier to remember, but it is more complicated to administer to complex circuits that require a computer analysis. For the relatively simple graph shown in Fig. 7.3, the simpler method works well. This method of flow graph reduction is based on four rules:

Rule 1. The cascade of two branches in series can be reduced to one branch with the value equal to the product of the two original branches (Fig. 7.4*a*).

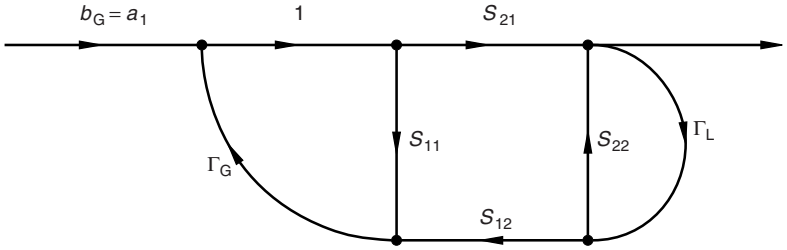



**FIGURE 7.3** Flow graph equivalent of the two-port circuit in Fig. 7.2.

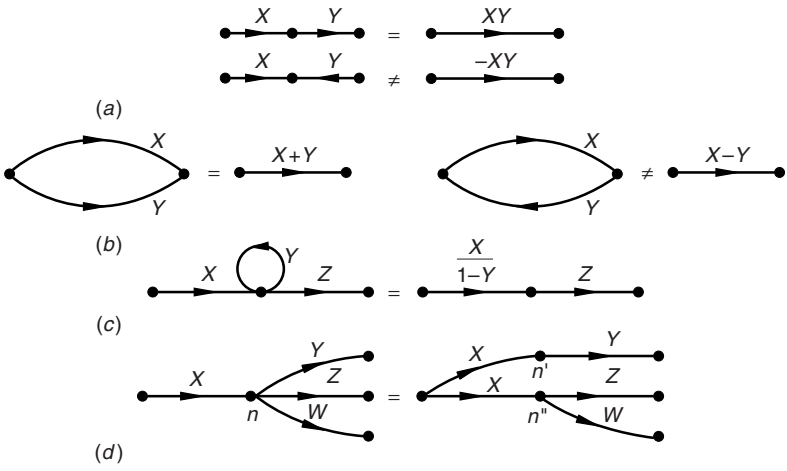



**FIGURE 7.4** Flow graph reduction rules for (*a*) two-series branches, (*b*) two-shunt branches, (*c*) a self-loop, and (*d*) splitting a node.

Rule 2. Two parallel branches can be reduced to one branch whose value is the sum of the two original branches (Fig. 7.4*b*).

Rule 3. As illustrated in Fig. 7.4*c*, a self-loop with value $Y$ with an incoming branch $X$ can be reduced to a single line of value

$$\frac{X}{1-Y} \tag{7.13}$$

Rule 4. The transfer function remains unchanged if a node with one input branch and $N$ output branches can be split into two nodes. The input branch goes to each of the new nodes. Similarly the transfer function remains unchanged if a node with one output branch and $N$ input branches can be split into two nodes. The output branch goes to each of the new nodes (Fig. 7.4*d*).

These rules can be used to finish the calculation of the transducer power gain of Eq. (7.12) by finding $b_2/b_G$. The first step in this reduction is the splitting of two nodes shown in Fig. 7.5*a* by use of rule 4. This forms a self-loop in the right-hand side of the circuit. The lower left-hand node is also split into two nodes (Fig. 7.5*b*). The incoming branches to the self-loop on the right-hand side are modified by means of rule 3 (Figs. 7.5*c*). In the same figure another self-loop is made evident on the left-hand side. In this case there are two incoming branches modified by the self-loop. Use of rule 3 produces Fig. 7.5*d*. Splitting the node by means of rule 4 results in Fig. 7.5*e*. The resulting self-loop modifies the incoming branch on the left-hand side (rule 3). The result is three branches in series (rule 1), so the transfer function can now be written by inspection:

$$b_2 = \frac{\dfrac{b_G}{1-\Gamma_G S_{11}}}{1-\dfrac{S_{21}S_{12}\Gamma_L\Gamma_G}{(1-\Gamma_L S_{22})(1-\Gamma_G S_{11})}} \cdot \frac{S_{21}}{1-\Gamma_L S_{22}}$$

$$\frac{b_2}{b_G} = \frac{S_{21}}{(1-\Gamma_L S_{22})(1-\Gamma_G S_{11}) - S_{12}S_{21}\Gamma_G\Gamma_L} \tag{7.14}$$

This ratio can be substituted into the transducer power gain expression (7.12). Thus the transducer power gain is known in terms of scattering parameters of the two-port and the terminating reflection coefficients:

$$G_T = \frac{|S_{21}|^2(1-|\Gamma_G|^2)(1-|\Gamma_L|^2)}{|(1-\Gamma_L S_{22})(1-\Gamma_G S_{11}) - S_{12}S_{21}\Gamma_G\Gamma_L|^2} \tag{7.15}$$

This is the full equation for the transducer power gain. Other expressions making use of approximations are strictly speaking a fiction, though this fiction is sometimes used to characterize certain transistors. For example, unilateral power gain is found by setting $S_{12} = 0$. In real transistors $S_{12}$ should be small, but it is never

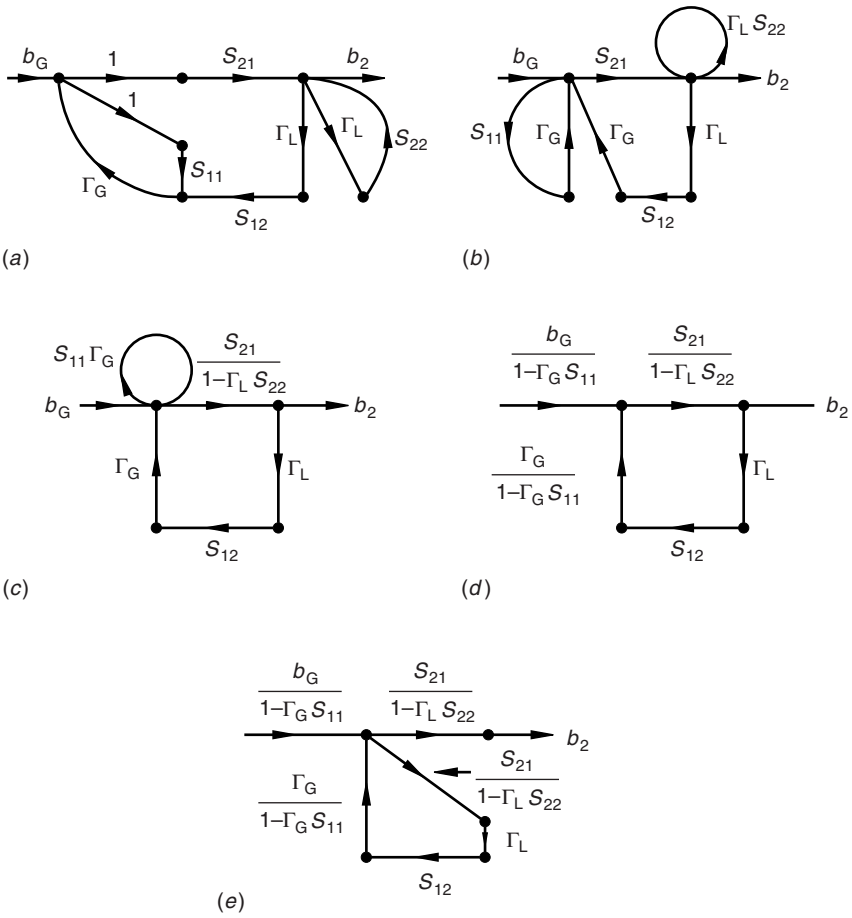



**FIGURE 7.5** Demonstration of the amplifier flow graph.

actually 0. The maximum unilateral power gain is found by setting $S_{12} = 0$, $\Gamma_G = S_{11}^*$, and $\Gamma_L = S_{22}^*$:

$$G_{\text{umax}} = \frac{|S_{21}|^2}{(1 - |S_{11}|^2)(1 - |S_{22}|^2)} \tag{7.16}$$

## 7.5 SIMULTANEOUS MATCH FOR MAXIMUM POWER GAIN

Maximum gain is obtained when both the input and output ports are simultaneously matched. One way to achieve this is to guess at a $\Gamma_L$ and calculate $\Gamma_i$ (Fig. 7.6). The generator impedance then is made to match the complex conjugate of $\Gamma_i$. With this new value of $\Gamma_G$, a new value of $\Gamma_O$ is found. Matching this to

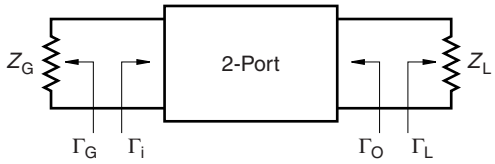



**FIGURE 7.6** The definition of the reflection coefficients for the two-port circuit.

$\Gamma_L$ means the that $\Gamma_L$ changes. This iterative process continues until both sides of the circuit are simultaneously matched.

A better way is to recognize this as basically a problem with two equations and two unknowns. Simultaneous match forces the following two requirements:

$$\Gamma_i = \Gamma_G^* = S_{11} + \frac{S_{21}S_{12}\Gamma_L}{1 - \Gamma_L S_{22}} \tag{7.17}$$

$$\Gamma_O = \Gamma_L^* = S_{22} + \frac{S_{21}S_{12}\Gamma_G}{1 - \Gamma_G S_{11}} \tag{7.18}$$

Since both of these equations have to be satisfied simultaneously, finding $\Gamma_G$ and $\Gamma_L$ requires solution of two equations with two unknowns. These can be written in terms of the determinate of the $S$ matrix $\Delta$ as follows:

$$\Gamma_G^* = \frac{S_{11} - \Gamma_L S_{11}S_{22} + S_{12}S_{21}\Gamma_L}{1 - \Gamma_L S_{22}}$$

$$= \frac{S_{11} - \Gamma_L \Delta}{1 - \Gamma_L S_{22}} \tag{7.19}$$

$$\Gamma_L^* = \frac{S_{22} - \Gamma_G \Delta}{1 - \Gamma_G S_{11}} \tag{7.20}$$

Substitution of Eq. (7.20) into Eq. (7.19) eliminates $\Gamma_L$:

$$\Gamma_G^* = \frac{S_{11}(1 - \Gamma_G^* S_{11}^*) - \Delta(S_{22}^* - \Gamma_G^* \Delta^*)}{1 - \Gamma_G^* S_{11}^* - |S_{22}|^2 + S_{22}\Delta^*\Gamma_G^*} \tag{7.21}$$

This expression can be rearranged in the usual quadratic form. After taking the complex conjugate, this yields the following:

$$\Gamma_G^2(S_{22}^*\Delta - S_{11}) + \Gamma_G(1 - |S_{22}|^2 + |S_{11}|^2 - ||\Delta|^2) - S_{11}^* + \Delta^* S_{22} = 0 \tag{7.22}$$

This equation can be rewritten in the form

$$0 = -\Gamma_G^2 C_1 + \Gamma_G B_1 - C_1^* \tag{7.23}$$

where

$$C_1 = S_{11} - \Delta S_{22}^* \tag{7.24}$$

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 \tag{7.25}$$

The required generator reflection coefficient for maximum gain can be found:

$$\Gamma_{\mathrm{Gm}} = \frac{C_1^*}{2|C_1|^2}\left[B_1 \pm \sqrt{B_1^2 - 4|C_1|^2}\right] \tag{7.26}$$

In similar fashion the load reflection coefficient (impedance) for maximum gain is

$$\Gamma_{\mathrm{Lm}} = \frac{C_2^*}{2|C_2|^2}\left[B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}\right] \tag{7.27}$$

where

$$C_2 = S_{22} - \Delta S_{11}^* \tag{7.28}$$

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2 \tag{7.29}$$

The parameters $B_\mathrm{i}$ and $C_\mathrm{i}$ are determined solely from the scattering parameters of the two-port. The $-$ sign is used when $B_\mathrm{i} > 0$, and the $+$ sign is used when $B_\mathrm{i} < 0$. Once the terminating reflection coefficients are known, the corresponding impedances may be determined:

$$Z_\mathrm{G} = Z_0\frac{1+\Gamma_\mathrm{G}}{1-\Gamma_\mathrm{G}} \tag{7.30}$$

$$Z_\mathrm{L} = Z_0\frac{1+\Gamma_\mathrm{L}}{1-\Gamma_\mathrm{L}} \tag{7.31}$$

## 7.6 STABILITY

A stable amplifier is an amplifier where there are no unwanted oscillations anywhere. Instability outside the operating band of the amplifier can still cause unwanted noise and even device burnout. Oscillations can only occur when there is some feedback path from the output back to the input. This feedback can result from an external circuit, from external feedback parasitic circuit elements, or from an internal feedback path such as through $C_\mu$ in a common emitter bipolar transistor. Of these three sources, the last is usually the most troublesome. The following sections describe a method for determining transistor stability and some procedures to stabilize an otherwise unstable transistor.

### 7.6.1 Stability Circles

The criteria for unconditional stability require that $|\Gamma_\mathrm{i}| \leq 1$ and $|\Gamma_o| \leq 1$ for any passive terminating loads. A useful amplifier may still be made if the terminating loads are carefully chosen to stay out of the unstable regions. It is helpful to find the borderline between the stable and the unstable regions. For the input side, this is done by finding the locus of points of $\Gamma_\mathrm{L}$ that will give $|\Gamma_\mathrm{i}| = 1$. The

borderline between stability and instability is found from Eq. (7.17) and (7.19) when $|\Gamma_i| = 1$:

$$1 = \left| \frac{S_{11} - \Delta\Gamma_L}{1 - \Gamma_L S_{22}} \right| \tag{7.32}$$

This equation can be squared and then split up into its complex conjugate pairs:

$$(1 - \Gamma_L S_{22})(1 - \Gamma_L^* S_{22}^*) = (S_{11} - \Delta\Gamma_L)(S_{11}^* - \Delta^*\Gamma_L^*) \tag{7.33}$$

The coefficients of the different forms of $\Gamma_L$ are collected together:

$$\begin{aligned} &|\Gamma_L|^2(|S_{22}|^2 - |\Delta|^2) + \Gamma_L(\Delta S_{11}^* - S_{22}) + \Gamma_L^*(S_{11}\Delta^* - S_{22}^*) \\ &\quad = |S_{11}|^2 - 1 \end{aligned} \tag{7.34}$$

$$\begin{aligned} &|\Gamma_L|^2 + \Gamma_L\left(\frac{\Delta S_{11}^* - S_{22}}{|S_{22}|^2 - |\Delta|^2}\right) + \Gamma_L^*\left(\frac{S_{11}\Delta^* - S_{22}^*}{|S_{22}|^2 - |\Delta|^2}\right) \\ &\quad = \frac{|S_{11}|^2 - 1}{|S_{22}|^2 - |\Delta|^2} \end{aligned} \tag{7.35}$$

Division of Eq. (7.35) by the coefficient of $|\Gamma_L|^2$ shows that this equation can be put in a form that can be factored by completing the square. The value $|m|^2$ is added to both sides of the equation:

$$(\Gamma_L + m^*)(\Gamma_L^* + m) = |\Gamma_L|^2 + \Gamma_L m + \Gamma_L^* m^* + |m|^2 + \frac{|S_{11}|^2 - 1}{|S_{22}|^2 - |\Delta|^2} \tag{7.36}$$

where

$$m \triangleq \frac{\Delta S_{11}^* - S_{22}}{|S_{22}|^2 - |\Delta|^2} \tag{7.37}$$

Substitution of Eq. (7.37) into Eq. (7.36) upon simplification yields the following factored form:

$$\left(\Gamma_L + \frac{\Delta^* S_{11} - S_{22}^*}{|S_{22}|^2 - |\Delta|^2}\right)\left(\Gamma_L^* + \frac{\Delta S_{11}^* - S_{22}}{|S_{22}|^2 - |\Delta|^2}\right) = \frac{|S_{12}S_{21}|^2}{(|S_{22}|^2 - |\Delta|^2)^2} \tag{7.38}$$

This is the equation of a circle whose center is

$$C_L = \frac{S_{11}\Delta^* - S_{22}^*}{|\Delta|^2 - |S_{22}|^2} \tag{7.39}$$

The radius of the load stability circle is

$$r_L = \left| \frac{S_{21}S_{12}}{|\Delta|^2 - |S_{22}|^2} \right| \tag{7.40}$$

The center and radius for the generator stability circle can be found in the same way by analogy:

$$C_G = \frac{S_{22}\Delta^* - S_{11}^*}{|\Delta|^2 - |S_{11}|^2} \tag{7.41}$$

$$r_G = \left| \frac{S_{21}S_{12}}{|\Delta|^2 - |S_{11}|^2} \right| \tag{7.42}$$

These two circles, one for the load and one for the generator, represent the borderline between stability and instability. These two circles can be overlayed on a Smith chart. The center of the circle is located at the vectorial position relative to the center of the Smith chart. The "dimensions" for the center and radius are normalized to the Smith chart radius (whose value is unity).

The remaining issue is which side of these circles is the stable region. Consider first the load stability circle shown in Fig. 7.7. If a matched $Z_0 = 50\,\Omega$

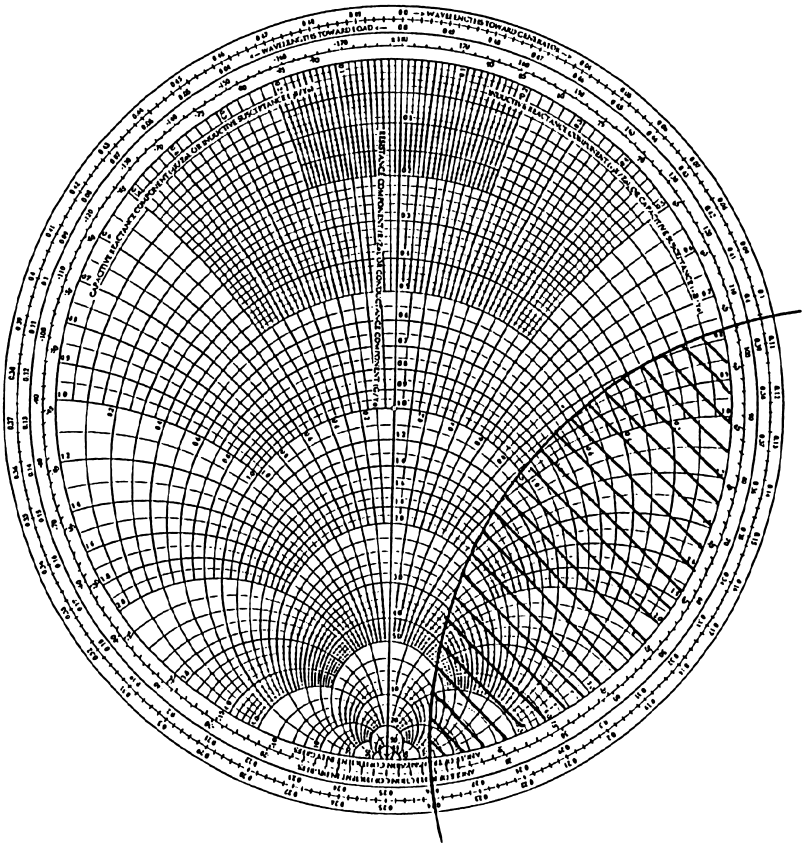


**FIGURE 7.7** Illustration of the stability circles where the shaded region is unstable.

transmission line were connected directly to the output port of the two-port circuit, then $\Gamma_L = 0$. This load would be located in the center of the Smith chart. Under this condition, Eq. (7.17) indicates that $\Gamma_i = S_{11}$. If the known value of $|S_{11}| < 1$, then $|\Gamma_i| < 1$ when the load is at the center of the Smith chart. If one point on one side of the stability circle is known to be stable, then all points on that side of the stability circle are also stable. The same rule would apply to the generator side when it is replaced by a matched load $= Z_0$.

Unconditional stability requires that both $|\Gamma_i| < 1$ and $|\Gamma_o| < 1$ for any passive load and generator impedance attached to the ports. In this case, if $|S_{11}| < 1$ and $|S_{22}| < 1$, the stability circles would lie completely outside the Smith chart. Conditional stability occurs when at least one of the stability circles intersects the Smith chart. As long as the load and source impedances are on the stable side of the stability circle, stable operation occurs. This choice may not, and usually will not, coincide with the generator and load impedance for maximum transducer power gain as given by Eqs. (7.26) and (7.27). Avoiding unstable operation will usually require compromising the maximum gain for a slightly smaller but often acceptable gain. Clearly, using an impedance too close to the edge of the stability circle can result in unstable operation because of manufacturing tolerances.

### 7.6.2 Rollett Criteria for Unconditional Stability

It is often useful to determine if a given transistor is unconditionally stable for any pair of passive impedances terminating the transistor. The two conditions necessary for this are known as the Rollett stability criteria [3] and are given as follows:

$$k = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} \geq 1 \tag{7.43}$$

$$|\Delta| \leq 1 \tag{7.44}$$

Rollett's original derivation was done using any one of the volt–ampere immittance parameters, $z$, $y$, $h$, or $g$. Subsequently his stability equations were expressed in terms of $S$ parameters as shown in Eqs. (7.43) and (7.44). Others arrived at stability conditions that appeared different from these, but it was pointed out that most of these alternate formulations were equivalent to those in Eqs. (7.43) and (7.44) [4]. The derivation of these two quantities will be given in this section.

The first of these equations is based on unconditional stability occurring when the load stability circle lies completely outside the Smith chart when $|S_{11}| < 1$, that is,

$$|C_L| - r_L \geq 1 \tag{7.45}$$

or

$$r_L - |C_L| \geq 1 \tag{7.46}$$

where Eq. (7.46) describes the case where the stability circle contains the entire Smith chart within it. Substitution of Eqs. (7.39) and (7.40) into Eq. (7.45) gives

$$\frac{|S_{22} - S_{11}^{*}\Delta| - |S_{12}S_{21}|}{||\Delta|^2 - |S_{22}|^2|} \geq 1 \tag{7.47}$$

Squaring Eq. (7.47) gives the following:

$$[|S_{22} - S_{11}^{*}\Delta| - |S_{12}S_{21}|]^2 \geq ||\Delta|^2 - |S_{22}|^2|^2 \tag{7.48}$$

$$|S_{22} - S_{11}^{*}\Delta|^2 - 2|S_{12}S_{21}||S_{22} - S_{11}^{*}\Delta| + |S_{12}S_{21}|^2 \geq ||\Delta|^2 - |S_{22}|^2|^2 \tag{7.49}$$

$$2|S_{12}S_{21}||S_{22} - S_{11}^{*}\Delta| \leq -||\Delta|^2 - |S_{22}|^2|^2 + |S_{12}S_{21}|^2 + |S_{22} - S_{11}^{*}\Delta|^2 \tag{7.50}$$

The last term on the right-hand side of Eq. (7.50) can be expanded:

$$\begin{aligned} |S_{22} - S_{11}^{*}\Delta|^2 &= (S_{22} - S_{11}^{*}\Delta)(S_{22}^{*} - S_{11}\Delta^{*}) \\ &= |S_{22}|^2 - S_{11}S_{22}\Delta^{*} - S_{11}^{*}S_{22}^{*}\Delta + |S_{11}|^2|\Delta|^2 \qquad (7.51) \\ &= |S_{22}|^2 + |\Delta|^2|S_{11}|^2 - |S_{11}S_{22}|^2 \\ &\quad + (S_{11}S_{22}S_{12}^{*}S_{21}^{*} - |S_{11}S_{22}|^2 + S_{11}^{*}S_{22}^{*}S_{12}S_{21}) \qquad (7.52) \end{aligned}$$

Now expansion of $|\Delta|^2$ gives the following:

$$\begin{aligned} |\Delta|^2 &= (S_{11}S_{22} - S_{12}S_{21})(S_{11}^{*}S_{22}^{*} - S_{12}^{*}S_{21}^{*}) \\ &= |S_{11}S_{22}|^2 + |S_{12}S_{21}|^2 - S_{11}S_{22}S_{12}^{*}S_{21}^{*} - S_{11}^{*}S_{22}^{*}S_{12}S_{21} \end{aligned} \tag{7.53}$$

By subtracting $|S_{12}S_{21}|$ inside the parenthesis in Eq. (7.52) and adding the same value outside the parenthesis, the quantity inside the parenthesis is equivalent to $|\Delta|^2$ given in Eq. (7.53). Thus Eq. (7.52) can be factored as shown below:

$$\begin{aligned} |S_{22} - S_{11}^{*}\Delta|^2 &= |S_{22}|^2 + |\Delta|^2|S_{11}|^2 - |S_{11}S_{22}|^2 + |S_{12}S_{21}|^2 - |\Delta|^2 \\ &= |S_{12}S_{21}|^2 + (1 - |S_{11}|^2)(|S_{22}|^2 - |\Delta|^2) \qquad (7.54) \\ &= \delta + \alpha\beta \qquad (7.55) \end{aligned}$$

The expression (7.55) is based on the definitions

$$\alpha \triangleq (1 - |S_{11}|^2) \tag{7.56}$$

$$\beta \triangleq (|S_{22}|^2 - |\Delta|^2) \tag{7.57}$$

$$\delta \triangleq |S_{12}S_{21}|^2 \tag{7.58}$$

The original inequality, Eq. (7.50), written in terms of these new variables is given below:

$$2\sqrt{\delta}\sqrt{\alpha\beta+\delta} \le (\alpha\beta+\delta)+\delta-\beta^2 \tag{7.59}$$

By first squaring both sides and then canceling terms, Eq. (7.59) can be greatly simplified:

$$\begin{aligned}
4\delta(\alpha\beta+\delta) &\le \left[(\alpha\beta+2\delta)-\beta^2\right]^2 \\
4\delta(\alpha\beta+\delta) &\le (\alpha\beta+2\delta)^2-2\beta^2(\alpha\beta+2\delta)+\beta^4 \\
4\delta(\alpha\beta+\delta) &\le (\alpha\beta)^2+4\delta(\alpha\beta)+4\delta^2-2\beta^2(\alpha\beta+2\delta)+\beta^4 \\
0 &\le (\alpha\beta)^2-2\beta^2(\alpha\beta+2\delta)+\beta^4 \\
&\le (\alpha-\beta)^2-4\delta \\
1 &\le \frac{(\alpha-\beta)^2}{4\delta}
\end{aligned} \tag{7.60}$$

Taking the square root of Eq. (7.60) yields

$$1 \le \frac{\alpha-\beta}{2\sqrt{\delta}} = \frac{1-|S_{11}|^2-|S_{22}|^2+|\Delta|^2}{2|S_{12}S_{21}|} = k \tag{7.43}$$

Since the value of $k$ is symmetrical on interchange of ports 1 and 2, the same result would occur with the generator port stability circle.

The second condition for unconditional stability, Eq. (7.44) can also be demonstrated based on the requirement that the $|\Gamma_i| < 1$. The second term of the right-hand side of Eq. (7.17) can be modified by multiplying it by 1 ($= S_{22}/S_{22}$) and adding 0 ($= S_{12}S_{21} - S_{12}S_{21}$) to the numerator. This results in the following:

$$\begin{aligned}
|\Gamma_i| &= \left| S_{11} + \frac{\Gamma_L S_{12}S_{21}S_{22} + (S_{12}S_{21} - S_{12}S_{21})}{(1-\Gamma_L S_{22})S_{22}} \right| \\
&= \frac{1}{|S_{22}|}\left| \frac{S_{11}S_{22}(1-\Gamma_L S_{22}) - S_{12}S_{21}(1-\Gamma_L S_{22}) + S_{12}S_{21}}{1-\Gamma_L S_{22}} \right| \\
&= \frac{1}{|S_{22}|}\left| \Delta + \frac{S_{12}S_{21}}{1-\Gamma_L S_{22}} \right| < 1
\end{aligned} \tag{7.61}$$

The complex quantity, $(1-\Gamma_L S_{22})$ can be written in polar form as $(1-|\Gamma_L S_{22}|e^{j\theta})$. Any passive load must lie within the unit circle $|\Gamma_L| < 1$, so $|\Gamma_L|$ is set to 1. As described in [5], the quantity

$$\frac{1}{1-|S_{22}|e^{j\theta}}$$

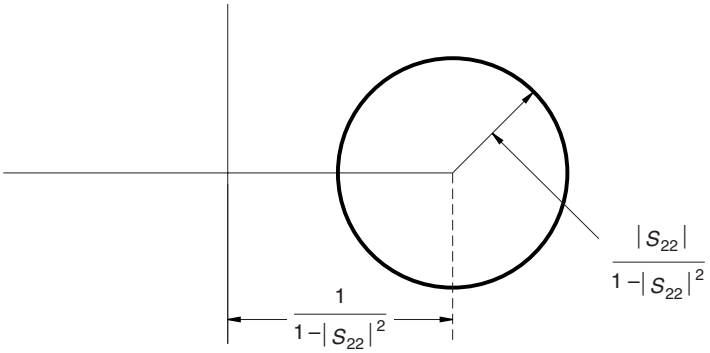



**FIGURE 7.8** Representation of the circle with $|\Gamma_L| = 1$.

which appears in Eq. (7.61) is a circle, as pictured in Fig. 7.8, centered at

$$\frac{1}{2}\left[\frac{1}{1-|S_{22}|}+\frac{1}{1+|S_{22}|}\right]=\frac{1}{1-|S_{22}|^2}$$

and with radius

$$\frac{1}{2}\left[\frac{1}{1-|S_{22}|}-\frac{1}{1+|S_{22}|}\right]=\frac{|S_{22}|}{1-|S_{22}|^2}$$

Equation (7.61) is expressed in terms of this circle:

$$\frac{1}{|S_{22}|}\left|\Delta+\frac{S_{12}S_{21}}{1-|S_{22}|^2}+\frac{S_{12}S_{21}|S_{22}e^{j\theta}|}{1-|S_{22}|^2}\right|<1 \tag{7.62}$$

The phase of the load is chosen so that it maximizes the left-hand side of Eq. (7.62). However, it must still obey the stated inequality. This means that Eq. (7.62) can be written as the sum of the two magnitudes without violating the inequality condition:

$$\frac{1}{|S_{22}|}\left|\Delta+\frac{S_{12}S_{21}}{1-|S_{22}|^2}\right|+\frac{|S_{12}S_{21}|}{1-|S_{22}|^2}<1$$

$$0<\frac{1}{|S_{22}|}\left|\Delta+\frac{S_{12}S_{21}}{1-|S_{22}|^2}\right|<1-\frac{|S_{12}S_{21}|}{1-|S_{22}|^2}$$

Comparison of the far right-hand side of this expression with 0 results in the following inequality:

$$1-|S_{22}|^2>|S_{12}S_{21}| \tag{7.63}$$

If the process had begun with the condition that $|\Gamma_o| < 1$, then the result would be the same as that of Eq. (7.63) with the 1's and 2's interchanged:

$$1-|S_{11}|^2>|S_{12}S_{21}| \tag{7.64}$$

When Eqs. (7.63) and (7.64) are added together,

$$2 - |S_{11}|^2 - |S_{22}|^2 > 2|S_{12}S_{21}|^2 \tag{7.65}$$

However, from the definition of the determinate of the $S$ parameter matrix,

$$|\Delta| = |S_{11}S_{22} - S_{12}S_{21}| < |S_{11}S_{22}| + |S_{12}S_{21}| \tag{7.66}$$

When the term $|S_{12}S_{21}|$ in Eq. (7.66) is replaced with something larger as given in Eq. (7.65), the inequality is still true:

$$|\Delta| < |S_{11}S_{22}| + 1 - \tfrac{1}{2}(|S_{11}|^2 + |S_{22}|^2)$$

$$|\Delta| < 1 - \tfrac{1}{2}(|S_{11}| - |S_{22}|)^2 < 1 \tag{7.44}$$

An alternate, but equivalent set of requirements for stability, is [4]

$$k > 1 \tag{7.67}$$

and either

$$B_1 > 0 \tag{7.68}$$

or

$$B_2 > 0 \tag{7.69}$$

### 7.6.3 Stabilizing a Transistor Amplifier

There are a variety of approaches to stabilizing an amplifier. In Section 7.6.1 it was suggested that stability could be achieved from a potentially unstable transistor by making sure that the chosen amplifier terminating impedances remain inside the stable regions at all frequencies as determined by the stability circles.

Another method would be to load the amplifier with an additional shunt or series resistor on either the generator or load side. The resistor is incorporated as part of the two-port parameters of the transistor. If the condition for unconditional stability is achieved for this expanded transistor model, then optimization can be performed for the other circuit elements to achieve the desired gain and bandwidth. It is usually better to try loading the output side rather than the input side in order to minimize increasing the amplifier noise figure.

A third approach that is sometimes useful is to introduce an external feedback path that can neutralize the internal feedback of the transistor. The most widely used scheme is the shunt–shunt feedback circuit shown in Fig. 7.9. The $y$ parameters for the composite circuit are simply the sum of the $y$ parameters of the amplifier and feedback two-port circuits:

$$[Y_c] = [Y_a] + [Y_f] \tag{7.70}$$

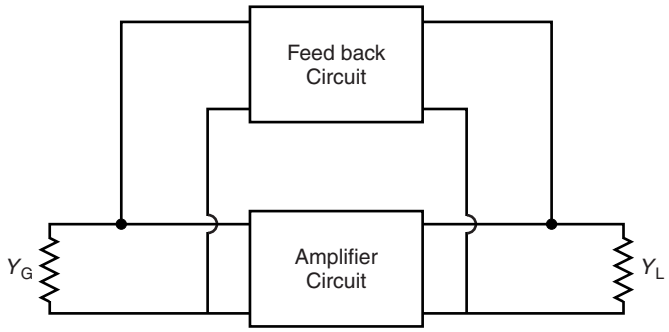



**FIGURE 7.9** Shunt–shunt feedback for stabilizing a transistor.

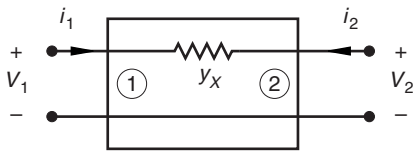



**FIGURE 7.10** Two-port representation of the feedback circuit.

To use this method, the transistor scattering parameters must be converted to admittance parameters (Appendix D). The $y$ parameters for a simple series admittance, $y_{fb}$ can be found from circuit theory (Fig. 7.10):

$$y_{11f} = y_{22f} = \left.\frac{i_1}{v_1}\right|_{v_2=0} = y_{fb} \tag{7.71}$$

$$y_{12f} = y_{21f} = \left.\frac{i_2}{v_1}\right|_{v_2=0} = -y_{fb} \tag{7.72}$$

Consequently the composite $y$ parameters are

$$y_{11c} = y_{11a} + y_{11f} = y_{11a} + y_{fb} \tag{7.73}$$

$$y_{12c} = y_{12a} + y_{12f} = y_{12a} - y_{fb} \tag{7.74}$$

$$y_{21c} = y_{21a} + y_{21f} = y_{21a} - y_{fb} \tag{7.75}$$

$$y_{22c} = y_{22a} + y_{22f} = y_{22a} + y_{fb} \tag{7.76}$$

If $y_{12c}$ could be made to be zero, then $S_{12c}$ would also be zero and unconditional stability could be achieved:

$$g_{12a} + jb_{12a} = g_{fb} + jb_{fb} \tag{7.77}$$

Since the circuit parameter $g_{12a} < 0$, the value $g_{fb} < 0$ must be true also. Since it is not possible to have a negative passive conductance, complete removal of

the internal feedback is not possible. However, the susceptance, $b_{12a}$, can be canceled by a passive external feedback susceptance. Although total removal of $y_{12a}$ cannot be achieved, yet progress toward stabilizing the amplifier can often be achieved. There is no guarantee that neutralization will provide a composite $y$ matrix that is unconditionally stable. In addition neutralization of the feedback susceptance occurs at only one frequency.

As an example consider a transistor to have the following $S$ parameters at a given frequency:

$$\begin{aligned} S_{11a} &= 0.73\angle -102^\circ \\ S_{21a} &= 2.21\angle 104^\circ \\ S_{12a} &= 0.10\angle 48^\circ \\ S_{22a} &= 0.47\angle -48^\circ \end{aligned} \tag{7.78}$$

For this transistor, $k = 0.752$ and $|\Delta| = 0.294$ as found from Eqs. (7.43) and (7.44). Conversion of Eq. (7.78) to $y$ parameters gives

$$\begin{aligned} y_{11a} &= 5.5307 \cdot 10^{-3} + j1.9049 \cdot 10^{-2} \text{ S} \\ y_{12a} &= 3.9086 \cdot 10^{-4} - j2.3092 \cdot 10^{-3} \text{ S} \\ y_{21a} &= 4.7114 \cdot 10^{-2} - j2.1376 \cdot 10^{-2} \text{ S} \\ y_{22a} &= 5.4445 \cdot 10^{-3} + j5.1841 \cdot 10^{-3} \text{ S}. \end{aligned} \tag{7.79}$$

Nothing can be done about $g_{12a}$, but $b_{12a}$ can be removed by setting $b_{fb} = b_{12a} = -2.3092$. The composite admittance matrix becomes

$$\begin{aligned} y_{11c} &= 5.5307 \cdot 10^{-3} + j1.6739 \cdot 10^{-2} \text{ S} \\ y_{12c} &= 3.9086 \cdot 10^{-4} - j(0) \text{ S} \\ y_{21c} &= 4.7114 \cdot 10^{-2} - j1.9067 \cdot 10^{-2} \text{ S} \\ y_{22c} &= 5.4445 \cdot 10^{-3} + j2.8750 \cdot 10^{-3} \text{ S} \end{aligned} \tag{7.80}$$

The composite scattering parameters can now be found and the stability factor calculated yielding $k = 2.067$ and $|\Delta| = 0.4037$. The transistor with the feedback circuit is unconditionally stable at the given frequency. This stability has been achieved by adding inductive susceptance in shunt with the transistor input and output ports.

Broadband stability can be achieved by replacing the feedback inductor with an inductor and resistor as shown in Fig. 7.11. A starting value for the inductor can be found as described for the single frequency analysis. The resistor is typically in the 200 to 800 $\Omega$ range, but optimum values for $R$ and $L$ are best found by computer optimization.

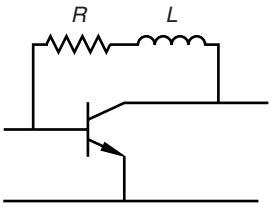



**FIGURE 7.11** Broadband feedback stabilization.

## 7.7 CLASS A POWER AMPLIFIERS

Class A amplifiers, whether for small signal or large signal operation, are intended to amplify the incoming signal in a linear fashion. This type of amplifier will not introduce significant distortion in the amplitude and phase of the signal. A linear class A power amplifier will introduce low harmonic frequency components and low intermodulation distortion (IMD). An example of intermodulation distortion can be described in terms of a double sideband suppressed carrier wave, which is represented as

$$\frac{V}{2}\cos(\omega_c + \omega_m)t + \frac{V}{2}\cos(\omega_c - \omega_m)t \tag{7.81}$$

where $\omega_c$ is the high-frequency carrier frequency and $\omega_m$ is the low-frequency modulation frequency. Intermodulation distortion would result in frequencies at $\omega_c \pm n\omega_m$, and harmonic distortion would cause frequency generation at $k\omega_\mathrm{c} \pm n\omega_\mathrm{m}$. The later harmonic distortion can usually be filtered out, but the intermodulation distortion is more difficult to handle because the distortion frequencies are near if not actually inside the system pass band. Clearly, this distortion in a class A amplifier is a greater problem for power amplifiers than for small signal amplifiers. Reduction of IMD depends on efficient power combining methods and careful design of the transistors themselves.

A transistor acting in the class A mode remains in its active state throughout the complete cycle of the signal. Two examples of common emitter class A amplifiers are shown in Fig. 7.12. The maximum efficiency of the class A amplifier in Fig. 7.12*a* has been shown to be 25%, for example [6]. However, if an RF coil can be used in the collector (Fig. 7.12*b*), the efficiency can be increased to almost 50%. This can be shown by recognizing first that there is no ac current flow in the bias source and no dc current flow in the load, $R_\mathrm{L}$. The total current flowing in the transistor collector is

$$i_\mathrm{c} = I_\mathrm{Q} - I_\mathrm{o}\sin\omega t \tag{7.82}$$

and the total collector voltage is

$$V_\mathrm{ce} = V_\mathrm{CC} + V_\mathrm{o}\sin\omega t \tag{7.83}$$

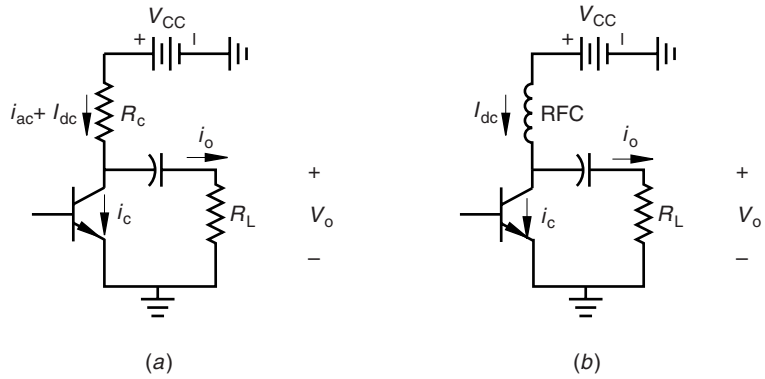



**FIGURE 7.12** Class A amplifiers with (*a*) collector resistor and (*b*) collector inductor.

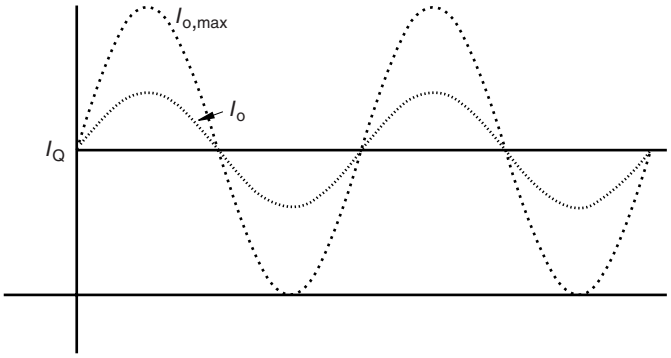



**FIGURE 7.13** Magnitude of the output current and quiescent current of the class A amplifier.

The quiescent current, $I_Q$, and the output current, $I_o$, is defined in Fig. 7.13. When the load is drawing the maximum instantaneous power,

$$I_{o,max} = I_Q = I_{dc}. \tag{7.84}$$

At this point the maximum output voltage is

$$V_{o,max} = I_{o,max} R_L \tag{7.85}$$

and

$$|V_{o,max}| = V_{CC} = I_{o,max} R_L \tag{7.86}$$

The dc power source supplies

$$P_{dc} = I_{dc} V_{CC} = \frac{V_{CC}^2}{R_L} \tag{7.87}$$

The maximum average power delivered to the load can now be written in terms of the supply voltage:

$$P_o = \frac{|V_{o,max}|^2}{2R_L} = \frac{V_{CC}^2}{2R_L} \tag{7.88}$$

If the RF input power is $P_i$, the power added efficiency is

$$\eta = \frac{P_o - P_i}{P_{dc}} \tag{7.89}$$

For high-gain amplifiers, $P_i \ll P_o$ and the maximum efficiency is $\eta \approx \frac{1}{2}$. However, it should be noted that many times high-power amplifiers do not have high gain, so the power added efficiency given by Eq. (7.89) offers a more useful quality factor for a transistor than if $P_i$ were neglected.

## 7.8 POWER COMBINING OF POWER AMPLIFIERS

Design of power FET amplifiers requires use of large gate periphery devices. However, eventually, the large the gate periphery causes other problems such as impedance matching especially at RF and microwave frequencies. Bandwidth improvement can be obtained by combining several transistors, often on a single chip. An example of combining two transistors is shown in Fig. 7.14 [7,8]. The separation of the transistors may induce odd-order oscillations in the circuit, even if the stability factor of the individual transistors (even-order stability) indicate they are stable. This odd-order instability can be controlled by adding $R_{odd}$ between the two drains to damp out such oscillations. This resistor is typically less than 400 Ω. Symmetry indicates no power dissipation when the outputs of the two transistors are equal and in phase. An example of a four transistor combining circuit is shown in Fig. 7.15, which now includes resistors $R_{odd1}$ and $R_{odd2}$ to help suppress odd-order oscillations.

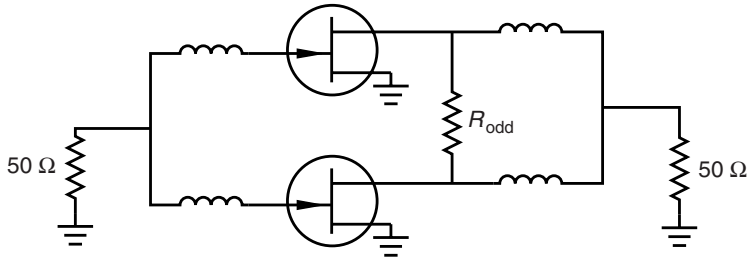



**FIGURE 7.14** Power combining two transistors [7,8].

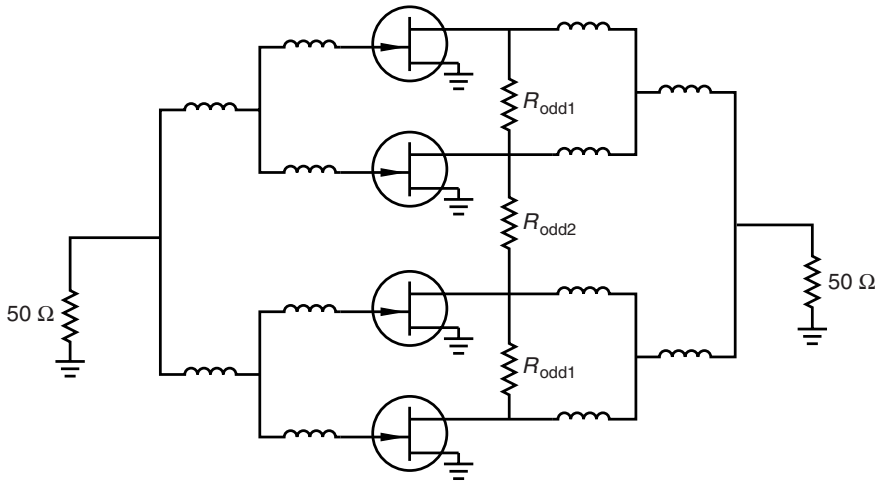



**FIGURE 7.15** Power combining four transistors [7,8].

## PROBLEMS

**7.1** Using the flow graph reduction method, verify the reflection coefficient found in Eq. (7.17).

**7.2** The measured scattering parameters of a transistor in an amplifier circuit are found to be the following:

| $\lvert S_{11}\rvert$ | $\angle S_{11}$ | $\lvert S_{21}\rvert$ | $\angle S_{21}$ | $\lvert S_{12}\rvert$ | $\angle S_{12}$ | $\lvert S_{22}\rvert$ | $\angle S_{22}$ |
|---|---|---|---|---|---|---|---|
| 0.85 | −32 | 3.8 | −145 | 0.04 | 74 | 0.92 | −15 |

**(a)** Determine the stability factor, $k$, for this transistor.

**(b)** Determine the $y$ parameters for this circuit.

**(c)** Determine the circuit that would neutralize (almost unilateralize) the circuit. While this procedure does not guarantee stability in all cases, it always helps lead toward greater stability.

**(d)** Determine the new scattering parameters for the neutralized circuit.

**(e)** Determine the generator and load impedances that would give maximum transducer power gain (not unilateral power gain).

**(f)** What is the value for the maximum transducer power gain.

**7.3** Determine the transfer function for the flow graph in Fig. 7.16.

**7.4** A certain transistor has the following $S$ parameters:

$$S_{11} = 1.2, \quad S_{21} = 4.0, \quad S_{12} = 0, \quad S_{22} = 0.9$$

Determine whether this transistor is unconditionally stable.

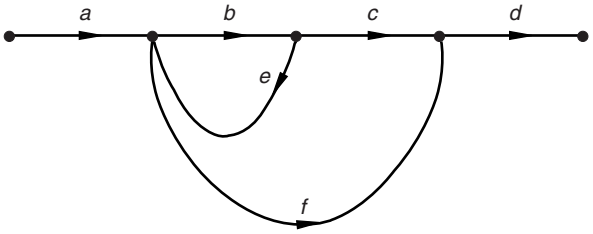



**FIGURE 7.16** Flow graph for Problem 7.3.

## REFERENCES

1. H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: Wiley, 1980, pp. 371–382.
2. P. J. Khan, Private communication, 1971.
3. J. M. Rollett, "Stability and Power-Gain Invariants of Linear Twoports," *IRE Trans. Circuit Theory*, Vol. CT-9, pp. 29–32, 1962.
4. D. Woods, "Reappraisal of the Unconditional Stability Criteria for Active 2-Port Networks in Terms of *S* Parameters," *IEEE Trans. Circuits and Systems*, Vol. CAS-23, pp. 73–81, 1976.
5. T. T. Ha, *Solid State Microwave Amplifier Design*, New York: Wiley, 1981.
6. S. G. Burns and P. R. Bond, *Principles of Electronic Circuits*, Boston: PWS Publishing, 1997.
7. R. G. Reitag, S. H. Lee, and D. M. Krafcsik, "Stability and Improved Circuit Modeling Considerations for High Power MMIC Amplifiers," *1988 IEEE Microwave Theory Tech. Symp. Digest*, pp. 175–178, 1988.
8. R. G. Freitag "A Unified Analysis of MMIC Power Amplifier Stability," *1992 IEEE Microwave Theory Tech. Symp. Digest*, pp. 297–300, 1992.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER EIGHT

# Noise

## 8.1 SOURCES OF NOISE

The dynamic range of a communication transmitter or receiver circuit is usually limited at the high-power point by nonlinearities and at the low-power point by noise. Noise is the random fluctuation of electrical power that interferes with the desired signal. There can be interference with the desired signal by other unwanted deterministic signals, but at this point only the interference caused by random fluctuations will be considered. There are a variety of physical mechanisms that account for noise, but probably the most common is thermal (also referred to as *Johnson noise* or *Nyquist noise*). This can be illustrated by simply examining the voltage across an open circuit resistor (Fig. 8.1). The resulting voltage is not zero! The average voltage is zero, but not the instantaneous voltage. At any temperature above absolute 0 K, the Brownian motion of the electrons will produce random instantaneous currents. These currents will produce random instantaneous voltages, and this leads to noise power.

Noise arising in electron tubes, semiconductor diodes, bipolar transistors, or field effect transistors come from a variety of mechanisms. For example, for tubes, these include random times of emission of electrons from a cathode (called *shot noise*), random electron velocities in the vacuum, nonuniform emission over the surface of the cathode, and secondary emission from the anode. Similarly, for diodes, a random emission of electrons and holes produces noise. In a bipolar transistor, there is in addition partition noise. This represents the fluctuation in the path that charge carriers take between the base and the collector after leaving the emitter. There is in addition $1/f$, or flicker noise (where $f$ is frequency), that is caused by surface recombination of base minority carriers at the base-emitter junction [1]. Clearly, as the frequency approaches dc, the flicker noise increases dramatically. As a consequence intermediate amplifier stages are designed to operate well above the frequency where $1/f$ noise is a significant contributor to the total noise. Typically this frequency ranges from 100 Hz to 10 kHz. In a field effect transistor, there is thermal noise arising from channel resistance, $1/f$

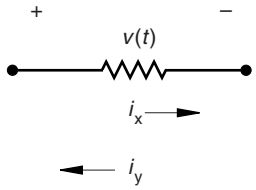



**FIGURE 8.1** Voltage across an open circuit resistor.

noise, and a coupling of the channel noise back to the gate where it is of course amplified by the transistor gain. Noise also arises from reverse breakdown in the avalanching of electrons in such devices as Zener diodes and IMPATT diodes. At RF frequencies the two most common noise sources are the thermal noise and the shot noise.

## 8.2 THERMAL NOISE

The random fluctuation of electrons in a resistance would be expected to rise as the temperature increases, since the electron velocities and the number of collisions per second increases. The noise voltage is expressed as an auto correlation of the instantaneous voltage over a time period $T$:

$$\langle v^2 \rangle = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} v^2(t) dt \tag{8.1}$$

The expression for thermal noise voltage has been derived in a variety of ways. Harry Nyquist first solved the problem based on a transmission line model. Other approaches included using a lumped element circuit, the random motion of electrons in a metal conductor, and the radiation from a black body. These are all basically thermodynamic models, and each method resulted in the same expression. The black body method is based on quantum mechanics and therefore provides a solution for noise sources at both cryogenic and room temperatures.

### 8.2.1 Black Body Radiation

Classical mechanics is based on the continuity of energy states. When this theory was applied to calculation of the black body radiation, it was found that the radiation increased without limit. This so-called ultraviolet catastrophe was clearly not physical. However, Max Planck was able to correct the situation by postulating that energy states are not continuous but are quantized in discrete states. These energy values are obtained by solving the Schrödinger equation for the harmonic oscillator. The actual derivation is found in most introductory texts on quantum mechanics [2]:

$$\mathcal{E} = \left( n + \frac{1}{2} \right) hf, \qquad n = 0, 1, 2, \ldots \tag{8.2}$$

In this equation $h = 6.547 \cdot 10^{-34}$ J · s is Planck's constant. If energy were continuous, then the average energy could be obtained from the Boltzmann probability distribution function, $P(\mathcal{E})$, by the following integral:

$$\overline{\mathcal{E}} = \frac{\int_0^\infty \mathcal{E}P(\mathcal{E})d\mathcal{E}}{\int_0^\infty P(\mathcal{E})d\mathcal{E}} \tag{8.3}$$

where

$$P(\mathcal{E}) = C\exp(-\beta\mathcal{E}) \tag{8.4}$$

$$\beta = \frac{1}{kT}$$

and

$$C = \frac{1}{\sum e^{-\beta\mathcal{E}}}$$

The value $k = 1.380 \cdot 10^{-23}$ J/°K is the Boltzmann constant and is essentially the proportionality constant between energy measured in terms of Joules and energy measured in terms of absolute temperature. Planck replaced the continuous integrals in Eq. (8.3) with summations of the discrete energy levels [3]:

$$\overline{\mathcal{E}} = \frac{\sum_{n=0}^{\infty} \mathcal{E}P(\mathcal{E})}{\sum_{n=0}^{\infty} P(\mathcal{E})} \tag{8.5}$$

$$= \frac{\sum_{n=0}^{\infty} (n + 1/2)hf e^{-\beta(n+1/2)hf}}{\sum_{n=0}^{\infty} e^{-\beta(n+1/2)hf}} \tag{8.6}$$

It may be easily verified by differentiation that

$$\frac{d}{d\beta} \ln \sum e^{-\beta(n+1/2)hf} = -\overline{\mathcal{E}} \tag{8.7}$$

The argument of the logarithm can be evaluated by recognizing it as an infinite geometric series:

$$e^{-\beta hf/2} \sum_{n=0}^{\infty} e^{-n\beta hf} = \frac{e^{-\beta h/2}}{1 - e^{-\beta hf}} \tag{8.8}$$

If Eq. (8.8) is substituted back into Eq. (8.7), the average energy can be found:

$$\mathcal{E} = \frac{d}{d\beta}[\ln(1 - e^{-\beta hf}) - \ln e^{-\beta hf/2}] \tag{8.9}$$
$$= \frac{hf e^{-\beta hf}}{1 - e^{-\beta hf}} + \frac{hf}{2}$$

or

$$\mathcal{E} = \frac{hf}{e^{hf/kT} - 1} + \frac{h}{2} \tag{8.10}$$

This will be used as the starting point for finding the noise power.

### 8.2.2 The Nyquist Formula

The thermal noise power in a given bandwidth $\Delta f$ is obtained directly from Eq. (8.10):

$$N_T = \frac{hf\Delta f}{e^{hf/kT} - 1} + \frac{hf\Delta f}{2} \tag{8.11}$$

At room temperature the second term, $hf\Delta f/2$, plays no role, but it may be essential in finding the minimum noise figure for cyrogenically cooled devices [4]. An approximation for the noise power can be found by expanding Eq. (8.11) into a Taylor series:

$$N_T \approx hf\Delta f \left[1 + \frac{hf}{kT} - 1\right]^{-1} + \frac{hf\Delta f}{2} = kT\Delta f \left[1 + \frac{hf}{2kT}\right] \tag{8.12}$$

At room temperature, $hf/kT \ll 1$, so this reduces to the usual practical formula for noise power as given by Nyquist [5]:

$$N_T = kT\Delta f \tag{8.13}$$

If this is the available power, the corresponding mean-squared voltage is obtained by multiplying this by four times the resistance, $R$:

$$\langle v^2 \rangle = 4RN_T$$
$$= 4RkT\Delta f \tag{8.14}$$

The mean-squared noise current is

$$\langle i^2 \rangle = 4GkT\Delta f \tag{8.15}$$

where $G$ is the associated conductance.

## 8.3 SHOT NOISE

Shot noise arises from random variations of a dc current, $I_0$, and is especially associated with current carrying active devices. Shot noise is most apparent in a current source with zero-shunt source admittance. For the purpose of illustration, consider a current source feeding a parallel RLC circuit (Fig. 8.2). The inductor provides a dc current path and is open to ac variations of the current. Hence the resulting noise voltage appears across the resistor (which is presumed free of any thermal noise). If an instrument could measure the current produced by randomly arriving electrons, the instrument would record a series of current impulses for each electron. If $n$ is the average number of electrons emitted by the source in a given time interval $\Delta t$, then the dc current is

$$I_0 = \frac{qn}{\Delta t} \tag{8.16}$$

where $q$ is the charge of an electron. Each current pulse provides an energy pulse to the capacitor with the value of

$$\mathcal{E} = \frac{q^2}{2C} \tag{8.17}$$

The average shot noise power delivered to the load is then

$$N_s = \frac{n\mathcal{E}}{\Delta t}$$

which in the light of Eqs. (8.16) and (8.17) becomes

$$\begin{aligned} N_s &= \frac{nq^2}{2C\Delta t} \\ &= \frac{qI_0}{2C} \end{aligned} \tag{8.18}$$

The equipartition theorem, as found in thermodynamics textbooks, states that the average energy of a system of uniform temperature is equally divided among the degrees of freedom of the system. If there are $N$ degrees of freedom, then

$$\overline{\mathcal{E}} = \frac{N}{2}kt \tag{8.19}$$

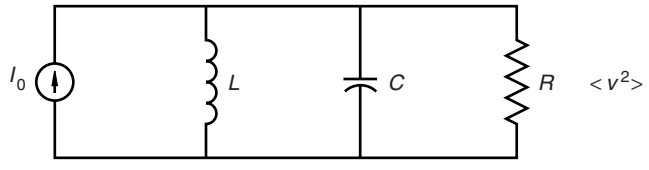



**FIGURE 8.2** Equivalent circuit for shot noise and for certain thermal noise calculations.

A system with $N$ degrees of freedom can be described uniquely by $N$ variables. The circuit in Fig. 8.2 has two energy storage elements, each containing an average energy of $kT/2$. For the capacitor this average energy is

$$\overline{\mathcal{E}} = \tfrac{1}{2}C\langle v^2\rangle = \tfrac{1}{2}kT \tag{8.20}$$

But it was found that the Nyquist noise formula predicted that $\langle v^2\rangle = 4RkT\Delta f$. Consequently

$$C = \frac{1}{4R\Delta f} \tag{8.21}$$

Using Eq. (8.21) to replace the value of the capacitance in Eq. (8.18) gives the desired formula for the shot noise power:

$$N_s = 2qRI_0\Delta f \tag{8.22}$$

The corresponding shot noise current is found by dividing by $R$:

$$\langle i^2\rangle = 2qI_0\Delta f \tag{8.23}$$

The shot noise current is directly proportional to the dc current as has been verified experimentally.

## 8.4 NOISE CIRCUIT ANALYSIS

When a circuit contains several resistors, the total noise power can be calculated by suitable combination of the resistors. Two resistors in series each produce a mean-squared voltage, $\langle v^2\rangle$. Since the individual noise voltage sources are uncorrelated, the total $\langle v^2\rangle$ is the sum of the $\langle v^2\rangle$ of each of the two resistors. Similarly two conductances in parallel each produce a mean-squared noise current, $\langle i^2\rangle$, that may be added when the two conductances are combined, since the noise currents are uncorrelated. It should be emphasized that two noise voltages $\langle v\rangle$ cannot be added together, only the mean-squared values can be added. The use of an arrow in the symbol for a noise current source is used to emphasize that this is a current source. The use of $+$ and $-$ signs in the symbol for a noise voltage source are used to emphasize that this is a voltage source. They do not imply anything about the phase of the noise sources. When both series and parallel resistors are present as shown in Fig. 8.3, then Thévenin's theorem provides an equivalent circuit and associated noise voltage. The output resistance is $(R_1 + R_2)||R_3$, and the corresponding noise voltage delivered to the output is

$$\langle v^2\rangle = 4kT(R_1 + R_2)||R_3 \tag{8.24}$$

When there is reactive element in the circuit such as that shown in the simple RLC circuit in Fig. 8.4, the output noise voltage would be attenuated by the magnitude of total admittance. If the admittance is constant over the bandwidth,

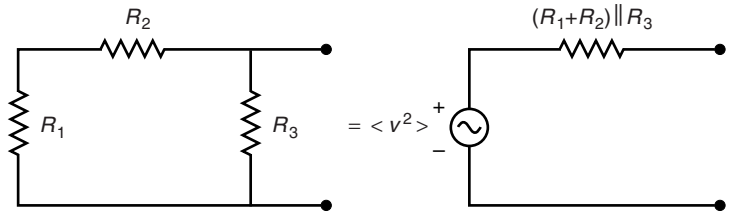



**FIGURE 8.3** Noise voltage from series and parallel resistors.

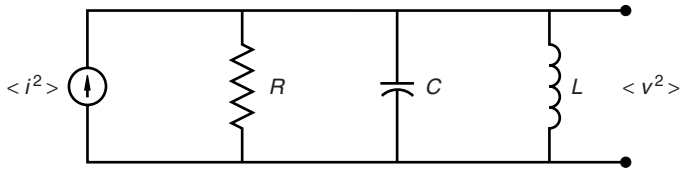



**FIGURE 8.4** Noise voltage from an RLC circuit.

$\Delta f$, which would be typically the case when the measured noise frequency, $f$, is approximately a sinusoid where $f \gg \Delta f$, then

$$\begin{aligned} \langle v^2 \rangle &= \frac{\langle i^2 \rangle}{|G + j(\omega C - 1/\omega L)|^2} \\ &= \frac{4kT\Delta f G}{|Y|^2} \end{aligned} \tag{8.25}$$

If the output resistance varies appreciably over the range of the noise bandwidth $\Delta f$, then the individual noise "sinusoids" must be summed over the bandwidth, resulting in the following integral:

$$\langle v^2 \rangle = 4kT \int_{\Delta f} \frac{G}{|Y|^2}\, df \tag{8.26}$$

As a simple example, consider the noise generated from a shunt *RC* circuit that would result from removing the inductance in Fig. 8.4:

$$\langle v^2 \rangle = \int_0^\infty \frac{4kTG df}{G^2 + (\omega C)^2} \tag{8.27}$$

$$\begin{aligned} &= \frac{4kTG}{2\pi G^2} \cdot \frac{G}{C} \int_0^\infty \frac{d(\omega C/G)}{1 + (\omega C/G)^2} \\ &= \frac{2kT}{\pi C} \cdot \frac{\pi}{2} = \frac{kT}{C} \end{aligned} \tag{8.28}$$

This expression does not say that the capacitor is the source of the noise voltage. Indeed, experiments have shown that changing the temperature of the resistor is what changes the output noise. When the 3 dB frequency point of the circuit output impedance ($f_{3\text{dB}} = 2\pi/RC$) is considered, the noise voltage in Eq. (8.28) becomes

$$\langle v^2 \rangle = 2\pi f_{3\text{dB}} kTR$$

This looks similar to the Nyquist formula in its original form, Eq. (8.14).

## 8.5 AMPLIFIER NOISE CHARACTERIZATION

One important quality factor of an amplifier is a measure of how much noise it adds to the signal while it amplifies it. The "actual noise figure," $F$, is a convenient measure of how the amplifier affects the total output noise. The noise figure, which by the IEEE standards was considered analogous to "noise factor," has been defined [7] as the ratio of (1) the total noise power per unit bandwidth at a corresponding output port when the standard noise temperature of the input termination is 290 K to (2) that portion of the total noise power engendered at the input frequency by the input termination. The standard 290 K noise temperature approximates the actual noise temperature of most input terminations as follows:

$$\begin{aligned} F &= \frac{\text{actual noise output power at } T_0}{\text{available noise input power}} \cdot \frac{1}{G_\text{T}} \\ &= \frac{N_\text{Tout}}{kT_0 G_\text{T} \Delta f} \end{aligned} \tag{8.29}$$

In this expression $G_\text{T}$ is the transducer power gain, and $T_0 = 290$ K. The noise figure is a measure of the total output noise after it leaves the amplifier divided by the input noise power entering the amplifier and amplified by an ideal noiseless gain $G_\text{T}$. In an analog amplifier, the amplifier can only add noise, so $F$ must always be greater than one. The noise figure can also be expressed in terms of the signal-to-noise ratio at the input to that at the output. If $P$ represents the input signal power, then

$$\begin{aligned} F &= \frac{P/kT_0\Delta f}{G_\text{T} P/N_\text{Tout}} \\ &= \frac{S_\text{in}/N_\text{Tin}}{S_\text{out}/N_\text{Tout}} \end{aligned} \tag{8.30}$$

Since the signal-to-noise ratio will always be degraded as the signal goes through the amplifier, again $F > 1$. The expression (8.30) is strictly true only if the input temperature is 290 K. This is called the *spot noise figure*.

The portion of the total thermal noise output power contributed by the amplifier itself is

$$N_\text{a} = N_\text{Tout} - kT_0 G_\text{T} \Delta f \tag{8.31}$$

$$= (F-1) kT_0 G_\text{T} \Delta f \tag{8.32}$$

The factor, $F - 1$, is used in two alternative measures of noise. One of these is noise temperature, which is particularly useful when dealing with very low noise amplifiers where the dB scale typically used in describing noise figure becomes too compressed to give insight. In this case the equivalent noise temperature is defined as

$$T_e = T_0(F - 1) \tag{8.33}$$

This is the temperature of the source resistance that when connected to the noise-free two-port circuit will give the same output noise as the original noisy circuit.

Another useful parameter for the description of noise is the noise measure [6]:

$$M = \frac{F - 1}{1 - (1/G)} \tag{8.34}$$

This is particularly useful for optimizing a receiver in which, for example, a trade-off has to be made between a low-gain low-noise amplifier and a high-gain high-noise amplifier.

## 8.6 NOISE MEASUREMENT

Measurement of noise figure can be accomplished by using a power meter and determining the circuit bandwidth and gain. However, it is inconvenient to determine gain and bandwidth each time a noise measurement is to be taken. The $Y$ factor method for determining noise figure is an approach where these two quantities need not be determined explicitly. Actual noise measurements are done over a range of frequencies. The average noise figure over a given bandwidth is [7]

$$\overline{F} = \frac{\int F(f)G_T(f)df}{\int G_T(f)df} \tag{8.35}$$

This represents a more realistic expression for an actual measurement of noise figure than the spot noise figure.

An equivalent noise bandwidth $\Delta f_0$ can be defined in terms of the maximum gain over the band as

$$\int G_T(f)df = G_0 \Delta f_0 \tag{8.36}$$

so that

$$\overline{F} = \frac{N_{Tout}}{kT_0G_0\Delta f_0} \tag{8.37}$$

A measurement system that can be used to measure the noise figure of an amplifier is shown in Fig. 8.5. This excess noise source in this circuit is gated on and off to produce two values of noise measured at the output power detector, $N_1$, and $N_2$:

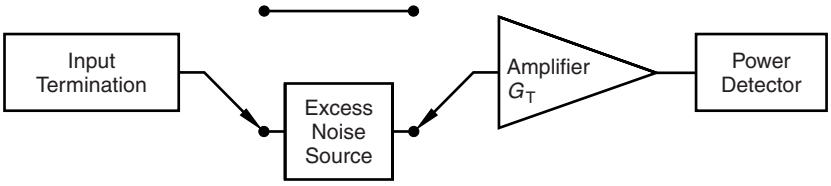



**FIGURE 8.5** Noise measurement using the $Y$ factor method.

$N_{ex}$ = calibrated excess noise source at $T_2 - T_0$
$N_1 = N_{Tout}$ when excess noise source is off
$N_2 = N_{Tout}$ when excess noise source is on
$N_{in}$ = noise from input termination
$N_a$ = noise added by the amplifier itself

The $Y$ factor as the ratio of $N_2$ to $N_1$ is easily obtained:

$$Y = \frac{N_2}{N_1} = \frac{G_0 N_{in} + G_0 N_{ex} + N_a}{G_0 N_{in} + N_a} \tag{8.38}$$

$$= \frac{G_0 k T_0 \Delta f_0 + G_0 k (T_2 - T_0)\Delta f_0 + (\overline{F} - 1) k T_0 G_0 \Delta f_0}{G_0 k T_0 \Delta f_0 + (\overline{F} - 1) k T_0 G_0 \Delta f_0}$$

$$= \frac{T_2 - T_0 + \overline{F} T_0}{\overline{F} T_0}$$

When solved for $\overline{F}$,

$$\overline{F} = \frac{T_2 - T_0}{T_0 (Y - 1)} \tag{8.39}$$

Since a calibrated noise source is used, $(T_2 - T_0)/T_0$ is known. Also $Y$ is known from the measurement. The amplifier noise figure is then obtained. Modern noise measurement instruments implicitly use the Y factor.

## 8.7 NOISY TWO-PORTS

The noise delivered to the output of a two-port circuit depends on the two-port circuit itself and the impedance of the input excitation source. The noise figure for a two-port circuit is given by the following:

$$F = F_{min} + \frac{R_n}{G_G}[(G_G - G_{opt})^2 + (B_G - B_{opt})^2] \tag{8.40}$$

$F_{\min}$ = minimum noise figure
$R_n$ = equivalent noise resistance (usually device data are given in terms of a normalized resistance, $r_n = R_n/50$)
$Y_G = G_G + jB_G$ = the excitation source admittance
$Y_{opt} = G_{opt} + jB_{opt}$ = optimum source admittance where the minimum noise figure occurs

While a designer can choose $Y_G$ to minimize the noise figure, such a choice will usually reduce the gain somewhat. Sometimes the noise figure is expressed in terms of reflection coefficients, where

$$\Gamma_G = \frac{Y_0 - Y_G}{Y_0 + Y_G} = \frac{Z_G - Z_0}{Z_G + Z_0} \tag{8.41}$$

where $Y_0$ and $Z_0$ are the characteristic admittance and impedance, respectively. Then the noise figure is given as follows:

$$F = F_{\min} + 4r_n \frac{|\Gamma_G - \Gamma_{opt}|^2}{(1 - |\Gamma_G|^2)|1 + \Gamma_{opt}|^2} \tag{8.42}$$

The noise figure expression (8.40) and its equivalent (8.42) are the basic expressions used to optimize transistor amplifiers for noise figure. The derivation of Eq. (8.40) is the subject of the following section. Readers not wishing to pursue these details at this point may proceed to Section 8.9 without loss of continuity.

## 8.8 TWO-PORT NOISE FIGURE DERIVATION

The work described here is based on the IRE standards published between 1956 and 1960 [8,9]. A noisy resistor can be modeled as a noiseless resistor in series with a voltage noise source. In similar fashion a two-port can be represented as a noiseless two-port and two noise sources. These two noise sources are represented in Fig. 8.6*a* as a voltage $v_n$ and a current $i_n$. The two-port circuit can be described in terms of its *ABCD* parameters and internal noise sources as

$$\begin{aligned} v_1 &= Av_2 + Bi_2 + v_n \\ i_1 &= Cv_2 + Di_2 + i_n \end{aligned} \tag{8.43}$$

or, as shown in Fig. 8.6, as a noiseless circuit and the noise sources referred to the input side. If the input termination, $Y_G$, produces a noise current, $i_G$, then the circuit is completed. The polarity markings on the symbols for these sources merely point out the distinction between voltage and current sources. Being noise sources, the polarities are actually random. The Thévenin equivalent circuit in

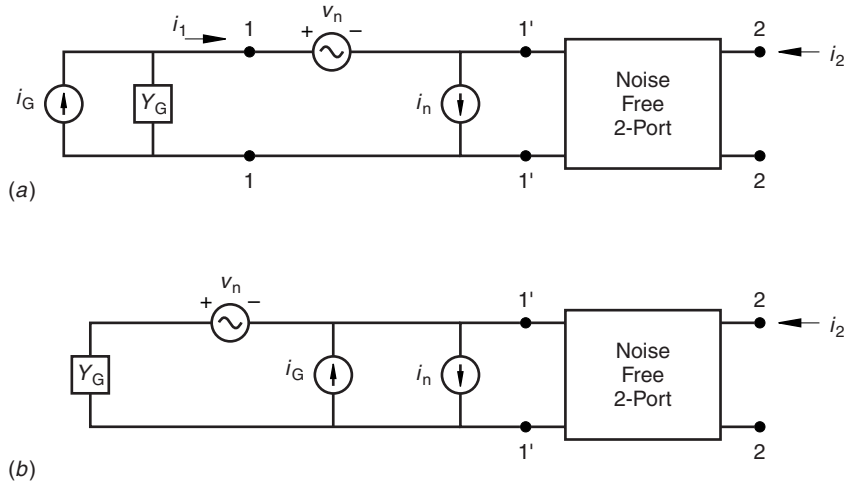



**FIGURE 8.6** Equivalent circuit (*a*) for two-port noise calculation, and (*b*) the equivalent Theévenin circuit.

Fig. 8.6*b* shows that the short circuit current at the $1'-1'$ port is

$$\langle i_{sc}^2 \rangle = \langle i_G^2 \rangle + \langle |i_n + Y_G v_n|^2 \rangle \tag{8.44}$$

$$= \langle i_G^2 \rangle + \langle i_n^2 \rangle + |Y_G|^2 \langle v_n^2 \rangle + Y_G^* \langle v_n^* i_n \rangle + Y_G \langle i_n^* v_n \rangle \tag{8.45}$$

The total output noise power is proportional to $\langle i_{sc}^2 \rangle$, and the noise caused by the input termination source alone is $\langle i_G^2 \rangle$. The part between $1'-1'$ and $2-2$ is noise free; that is, it adds no additional noise to the output. All the noise sources are referred to the input side, so the noise figure is

$$F = \frac{\langle i_{sc}^2 \rangle}{\langle i_G^2 \rangle} \tag{8.46}$$

Part of the noise current source, $i_n$, is correlated and part is uncorrelated with the noise voltage $v_n$. The uncorrelated current is $i_u$. The rest of the current is correlated with $v_n$ and is given by $(i_n - i_u)$. This correlated noise current must be proportional to $v_n$. The proportionality constant is the correlation admittance given by $Y_c = G_c + jB_c$ and is defined so that

$$i_n = i_u + Y_c v_n \tag{8.47}$$

While this defines $Y_c$, its explicit value in the end will not be needed. The mean value of the product of the correlated and uncorrelated current is of course 0. By definition, the average of the product of the noise voltage, $v_n$, and the uncorrelated noise current, $i_u$, must also be 0. Using the complex conjugate of

the current (which is a fixed phase shift) will not change this fact:

$$\langle v_n i_u^* \rangle = 0 \tag{8.48}$$

Rearranging Eq. (8.47) gives

$$\frac{i_n - i_u}{Y_c} = v_n \tag{8.49}$$

The product of the noise voltage and the uncorrelated current in Eq. (8.48) can be expressed by substitution of Eq. (8.49) into Eq. (8.48):

$$\langle (i_n - i_u) i_u^* \rangle = 0 \tag{8.50}$$

Because $\langle v_n i_u^* \rangle = 0$ from Eq. (8.48), the product of the noise voltage and the correlated current can be found using Eq. (8.47):

$$\langle v_n i_n^* \rangle = \langle v_n (i_n - i_u)^* \rangle = Y_c^* \langle v_n^2 \rangle \tag{8.51}$$

The noise sources are determined by their corresponding resistances:

$$\langle v_n^2 \rangle = 4kT_0 R_n \Delta f \tag{8.52}$$

$$\langle i_u^2 \rangle = 4kT_0 G_u \Delta f \tag{8.53}$$

$$\langle i_G^2 \rangle = 4kT_0 G_G \Delta f \tag{8.54}$$

The resistance, $R_n$, is the equivalent noise resistance for $\langle v_n^2 \rangle$, and $G_u$ is the equivalent noise conductance for the uncorrelated part of the noise current, $\langle i_u^2 \rangle$. The total noise current is the sum of the uncorrelated current and the remaining correlated current:

$$\begin{aligned} \langle i_n^2 \rangle &= \langle i_u^2 \rangle + \langle |i_n - i_u|^2 \rangle \\ &= \langle i_u^2 \rangle + |Y_c|^2 \langle v_n^2 \rangle \qquad (8.55) \\ &= 4kT_0 \Delta f (G_u + R_n |Y_c|^2) \qquad (8.56) \end{aligned}$$

Now the expression for the short circuit current in Eq. (8.45) can be modified by Eq. (8.51):

$$\langle i_{sc}^2 \rangle = \langle i_G^2 \rangle + \langle i_n^2 \rangle + |Y_G|^2 \langle v_n^2 \rangle + Y_G^* Y_c \langle v_n^2 \rangle + Y_G Y_c^* \langle v_n^2 \rangle \tag{8.57}$$

Furthermore $\langle i_n^2 \rangle$ can be replaced by Eq. (8.55):

$$\langle i_{sc}^2 \rangle = \langle i_G^2 \rangle + \langle i_u^2 \rangle + |Y_c|^2 \langle v_n^2 \rangle + |Y_G|^2 \langle v_n^2 \rangle + Y_G^* Y_c \langle v_n^2 \rangle + Y_G Y_c^* \langle v_n^2 \rangle \tag{8.58}$$

The noise figure, given by Eq. (8.46), can now be put in more convenient form:

$$F = 1 + \frac{\langle i_u^2 \rangle + \langle v_n^2 \rangle (|Y_c|^2 + |Y_G|^2 + Y_G^* Y_c + Y_G Y_c^*)}{\langle i_G^2 \rangle} \tag{8.59}$$

$$= 1 + \frac{4kT_0 G_u \Delta f + 4kT_0 R_n \Delta f (|Y_G| + |Y_c|)^2}{4kT_0 G_G \Delta f} \tag{8.60}$$

$$= 1 + \frac{G_u}{G_G} + \frac{R_n}{G_G}[(G_G + G_c)^2 + (B_G + B_c)^2] \tag{8.61}$$

The noise figure, $F$, is a function of the input termination admittance, $Y_G$, and reaches a minimum when the source admittance is optimum. In particular, the optimum susceptance is $B_G = B_{opt} = -B_c$. The value for $F_{min}$ is found by setting the derivative of $F$ with respect to $G_G$ to zero and setting $B_G = -B_c$. This will determine the a value for $G_G = G_{opt}$ in terms of $G_u$, $R_u$, and $G_c$:

$$\frac{dF}{dG_G} = 0 = -\frac{G_u}{G_G^2} - \frac{R_n}{G_G^2}(G_G + G_c)^2 + \frac{2R_n}{G_G}(G_G + G_c) \tag{8.62}$$

Solution for $G_G$ yields

$$G_G = G_{opt} = \sqrt{\frac{G_u + R_n G_c^2}{R_n}} \tag{8.63}$$

or

$$G_c^2 = G_{opt}^2 - \frac{G_u}{R_n} \tag{8.64}$$

Substituting this into Eq. (8.61) provides the minimum noise figure, $F_{min}$:

$$F_{min} = 1 + \frac{1}{G_{opt}}\left[G_u + R_n\left(G_{opt}^2 + 2G_{opt}\sqrt{G_{opt}^2 - \frac{G_u}{R_n}} + G_{opt}^2 - \frac{G_u}{R_n}\right)\right] \tag{8.66}$$

$$= 1 + 2R_n\left[G_{opt} + \sqrt{G_{opt}^2 - \frac{G_u}{R_n}}\right] \tag{8.67}$$

The correlation conductance, $G_c$, can be replaced from the total noise figure expression in Eq. (8.61) by Eq. (8.64) to give the following expression for $F$:

$$F = 1 + \frac{G_u}{G_G} + \frac{R_n}{G_G} \times \left[\left(G_G^2 + 2G_G\sqrt{G_{opt}^2 - \frac{G_u}{R_n}} + G_{opt}^2 - \frac{G_u}{R_n}\right) + (B_G - B_{opt})^2\right] = 1 + \frac{R_n}{G_G}$$

$$\times \left[\left(G_G^2 + 2G_G\sqrt{G_{opt}^2 - \frac{G_u}{R_n}} + G_{opt}^2 - 2G_G G_{opt} + 2G_G G_{opt}\right) + (B_G - B_{opt})^2\right]$$

$$= 1 + \frac{R_n}{G_G}\left(2G_G G_{opt} + 2G_G\sqrt{G_{opt}^2 - \frac{G_u}{R_n}}\right)$$

$$+ \frac{R_n}{G_G}[(G_G - G_{opt})^2 + (B_G - B_{opt})^2] \tag{8.68}$$

Noting Eq. (8.67), the desired expression is obtained:

$$F = F_{min} + \frac{R_n}{G_G}[(G_G - G_{opt})^2 + (B_G - B_{opt})^2] \tag{8.69}$$

## 8.9 THE FUKUI NOISE MODEL FOR TRANSISTORS

Fukui found an empirically based model that accurately describes the frequency dependence of the noise for high-frequency field effect transistors [10]. This model reduces to predicting the four noise parameters, $F_{min}$, $R_n$, $R_{opt}$, and $X_{opt}$ where the later two parameters are formed from the reciprocal of $Y_{opt}$. For the circuit shown in Fig. 8.7, the Fukui relationships are as follows:

$$F_{min} = 1 + k_1 f C_{gs}\left(\frac{R_g + R_s}{g_m}\right)^{1/2} \tag{8.70}$$

$$R_n = \frac{k_2}{g_m} \tag{8.71}$$

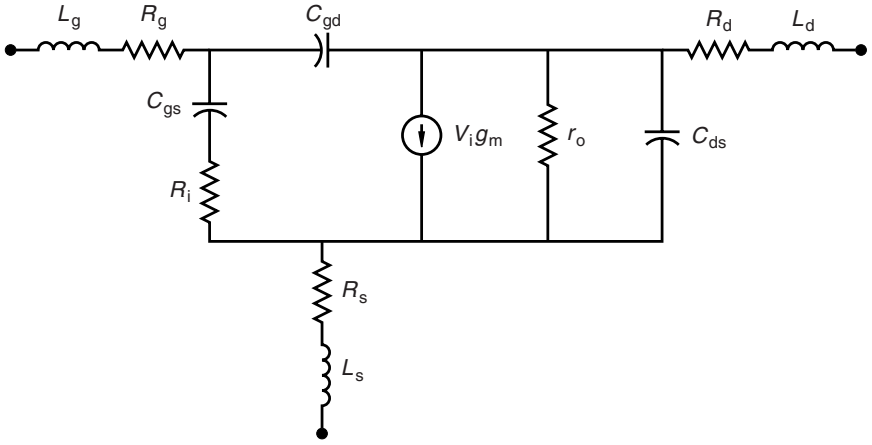



**FIGURE 8.7** Equivalent circuit for noise calculation for a FET.

$$R_{\text{opt}} = \frac{k_3}{f}\left(\frac{1}{4g_{\text{m}}} + R_{\text{s}} + R_{\text{g}}\right) \tag{8.72}$$

$$X_{\text{opt}} = \frac{k_4}{f C_{\text{gs}}} \tag{8.73}$$

In these expressions, $f$ is the operating frequency in GHz, the capacitance is in pF, and the transconductance in Siemens. The constants $k_1$, $k_2$, $k_3$, and $k_4$ are empirically based fitting factors. The expression for $R_{\text{opt}}$ in Eq. (8.72) differs from that originally given by Fukui, as modified by Golio [11]. The circuit elements of the equivalent FET model in Fig. 8.7 can be extracted at a particular bias level. The resistance, $R_{\text{i}}$, is often difficult to obtain, but for the purpose of the noise estimation, it may be incorporated with the $R_{\text{g}}$. The empirically derived fitting factors should be independent of frequency. They are not quite constant, but over a range of 2 to 18 GHz average values for these are shown below [11]:

$$k_1 = 0.0259$$
$$k_2 = 2.966$$
$$k_3 = 14.51$$
$$k_4 = 162.6$$

These values can be used for approximate estimates of noise figure for both metal semiconductor field effect transistors (MESFETs) as well as high-electron mobility transistors (HEMTs).

The transistor itself can be modified to provide either improved noise characteristics or improved power-handling capability by adjusting the gate width, $W$. The drain current, $I_{\text{ds}}$, increases with the base width $W$. Consequently those equivalent circuit parameters determined by derivatives of $I_{\text{ds}}$ will also be proportional to $W$. Also the capacitance between the gate electrode and the source electrode or between the gate electrode and the drain electrode will be also proportional to $W$. This is readily seen from the layout of a FET shown in Fig. 8.8. The gate resistance, $R_{\text{g}}$, scales differently, since the gate current flows in the direction of the width. Also the number of gate fingers, $N$, will reduce the effective gate resistance. The gate resistance is then proportional to $W/N$. These relationships may be summarized as follows:

$$g_{\text{m}} \propto W$$
$$R_{\text{ds}} \propto \frac{1}{W}$$
$$C_{\text{gs}} \propto W$$
$$C_{\text{gd}} \propto W$$
$$R_{\text{g}} \propto \frac{W}{N}$$

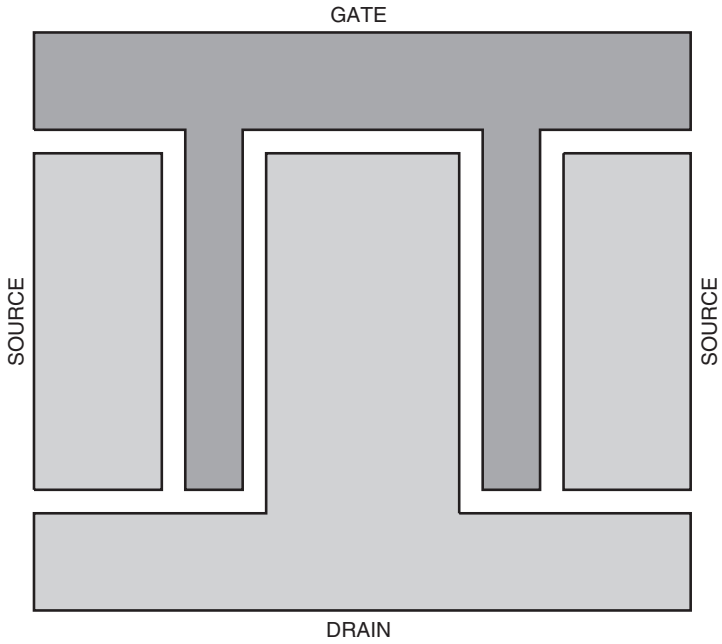



**FIGURE 8.8** Typical FET layout.

These circuit elements can clearly be adjusted by scaling the transistor geometry. This scaling will in turn change the noise characteristics. If a transistor with a given geometry has a known set of noise parameters, then the noise characteristics of a new modified transistor can be predicted. The scaling factors between the new and the old transistor are

$$s_1 = \frac{W'}{W}$$
$$s_2 = \frac{W'/N'}{W/N} \tag{8.74}$$

As a result the new equivalent circuit parameters can be predicted [11]:

$$g'_{\mathrm{m}} = g_{\mathrm{m}} s_1 \tag{8.75}$$
$$R'_{\mathrm{s}} = \frac{R_{\mathrm{s}}}{s_1} \tag{8.76}$$
$$R'_{\mathrm{d}} = \frac{R_{\mathrm{d}}}{s_1} \tag{8.77}$$
$$C'_{\mathrm{gs}} = C_{\mathrm{gs}} s_1 \tag{8.78}$$
$$R'_{\mathrm{g}} = R_{\mathrm{g}} s_2 \tag{8.79}$$

The Fukui equations, (8.70) to (8.73), for the newly scaled equivalent circuit parameters are given as follows:

$$F'_{\min} = 1 + k_1 f C'_{gs} \left( \frac{R'_g + R'_s}{g'_m} \right)^{1/2}$$

$$= 1 + (F_{\min} - 1) \left( \frac{s_1 s_2 R_g + R_s}{R_g + R_s} \right)^{1/2} \tag{8.80}$$

$$R'_n = \frac{R_n}{s_1} \tag{8.81}$$

$$R'_{opt} = \frac{R_{opt}}{s_1} \left[ \frac{1 + 4g_m(R_s + R_g s_1 s_2)}{1 + 4g_m(R_s + R_g)} \right] \tag{8.82}$$

$$X'_{opt} = \frac{X_{opt}}{s_1} \tag{8.83}$$

Reference should be made to [11] for a much fuller treatment of modeling MESFETs and HEMTs.

The bipolar transistor has a much different variation of noise with frequency than does the FET type of device. An approximate value for $F_{\min}$ for the bipolar transistor at high frequencies is [12]

$$F_{\min} \approx 1 + h \left[ 1 + \sqrt{1 + \frac{2}{h}} \right] \tag{8.84}$$

where

$$h \triangleq \frac{q I_c r_b}{kT} \left( \frac{\omega}{\omega_T} \right)^2 \tag{8.85}$$

In this equation, $I_c$ is the dc collector current, $r_b$ is the base resistance, and $\omega_T$ is the frequency where the short circuit current gain is 1. Values for $Y_{opt}$ and $R_n$ are also given in [12] but are rather lengthy. A somewhat more accurate expression is given in [13].

Comparison of Eq. (8.84) with the corresponding expression for FETs, Eq. (8.70), indicates that the bipolar transistor minimum noise figure increases with $f^2$, while that for the FET increases only as $f$. Consequently designs of low-noise amplifiers at RF and microwave frequencies would tend to favor use of FETs.

## 8.10 PROPERTIES OF CASCADED AMPLIFIERS

Ideally amplifiers are completely unilateral so that there is no feedback signal returning to the input side. Under this condition analysis of cascaded amplifiers

results in some interesting properties related to noise figure and efficiency. The results obtained will be approximately valid for almost unilateral amplifiers, even if some of the "amplifiers" are attenuators. The following two sections deal with the total noise figure and total efficiency respectively of a cascade of unilateral amplifiers.

### 8.10.1 Friis Noise Formula

The most critical part for achieving low noise in a receiver is the noise figure and gain of the first stage. This is intuitively clear, since the magnitude of the noise in the first stage will be a much larger percentage of the incoming signal than it will be in subsequent stages where the signal amplitude is much larger. For a receiver with $n$ unilateral stages, the total noise figure for all $n$ stages is [14]

$$F_{Tn} = \frac{N_{Tn}}{kT_0 \Delta f G_1 G_2 \ldots G_n} \tag{8.86}$$

where $N_{Tn}$ is the total noise power delivered to the load. This can be expressed in terms of the sum of the noise added by the last stage, $N_n$, and that of all the previous stages multiplied by the gain of the last stage:

$$N_{Tn} = N_n + G_n N_{T(n-1)} \tag{8.87}$$

If the $n$th stage were removed, and its noise figure measured alone, then its noise figure would be

$$F_n = \frac{kT_0 G_n \Delta f + N_n}{kT_0 G_n \Delta f} \tag{8.88}$$

or

$$F_n - 1 = \frac{N_n}{kT_0 G_n \Delta f} \tag{8.89}$$

By substituting Eq. (8.87) into Eq. (8.86) an expression for the noise figure is obtained that separates the contributions of the noise coming from the last stage only from the previous $n - 1$ stages:

$$F_{Tn} = \frac{N_n}{(kT_0 \Delta f G_n) G_1 G_2 \ldots G_{n-1}} + \frac{G_n N_{T(n-1)}}{kT_0 \Delta f G_1 G_2 \ldots G_{n-1} G_n} \tag{8.90}$$

Canceling the $G_n$ in the second term and substituting Eq. (8.89) yields

$$F_{Tn} = \frac{F_n - 1}{G_1 G_2 \ldots G_{n-1}} + \frac{N_{\mathrm{T(n-1)}}}{kT_0 \Delta f G_1 G_2 \ldots G_{n-1}} \tag{8.91}$$

The second term in Eq. (8.91) is the same as Eq. (8.86) except that $n$ has been reduced to $n - 1$. This process is repeated $n$ times giving what is known as the

Friis formula for the noise figure for a cascade of unilateral gain stages:

$$F_{Tn} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_n - 1}{G_1 G_2 \ldots G_{n-1}} \tag{8.92}$$

Clearly the noise figure of the first stage is the most important contributor to the overall noise figure of the system. If the first stage has reasonable gain, the subsequent stages can have much higher noise figure without affecting the overall noise figure of the receiver.

### 8.10.2 Multistage Amplifier Efficiency

For a multistage amplifier, the overall power efficiency can be found that will correspond in some way with the overall noise figure expression. Unlike the noise figure, however, the efficiency of the last stage will be found to be most important. Again, this would appear logical since the last amplifying stage handles the greatest amount of power so that poor efficiency here would waste the most amount of power. For the $k$th stage of an $n$ stage amplifier chain, the power added efficiency is

$$\eta_k = \frac{P_{ok} - P_{ik}}{P_{dk}} \tag{8.93}$$

where

$P_{ok}$ = output power of the $k$th stage
$P_{ik}$ = input power to the $k$th stage
$P_{dk}$ = source of power which is typically the dc bias for the $k$th stage

If the input power to the first stage is $P_{i1}$, then

$$P_{ik} = P_{i1} G_1 G_2 G_3 \ldots G_{k-1} \tag{8.94}$$

and for the $k$th stage alone

$$P_{ok} = P_{ik} G_k \tag{8.95}$$

When Eq. (8.94) and Eq. (8.95) are substituted into the efficiency equation, Eq. (8.93), the power from the power source can be found:

$$P_{dk} = \frac{G_1 G_2 \ldots G_{k-1}(G_k - 1)}{\eta_k} P_{i1} \tag{8.96}$$

The total added power for a chain of $n$ amplifiers is

$$P_{on} - P_{i1} = P_{i1}(G_1 G_2 G_3 \ldots G_n - 1) \tag{8.97}$$

The efficiency for the whole amplifier chain is clearly given by the following:

$$\eta_T = \frac{P_{on} - P_{i1}}{\sum_{i=1}^{n} P_{dk}} \tag{8.98}$$

Replacing the power levels in Eq. (8.98) with their explicit expression gives the value for the overall efficiency of a chain of unilateral amplifiers:

$$\eta_T = \frac{G_1 G_2 G_3 \ldots G_n - 1}{\dfrac{G_1 - 1}{\eta_1} + \dfrac{G_1(G_2 - 1)}{\eta_2} + \cdots + \dfrac{G_1 G_2 \ldots G_{n-1}(G_n - 1)}{\eta_n}} \tag{8.99}$$

When each amplifier stage has a gain sufficiently greater than one, the overall efficiency becomes

$$\eta_T \approx \frac{1}{\dfrac{1}{G_2 G_3 \ldots G_n \eta_1} + \cdots + \dfrac{1}{G_n \eta_{n-1}} + \dfrac{1}{\eta_n}} \tag{8.100}$$

This final equation illustrates that it is important to make the final stage the most efficient one

## 8.11 AMPLIFIER DESIGN FOR OPTIMUM GAIN AND NOISE

The gain for a nonunilateral amplifier was previously given as Eq. (7.15) is repeated here:

$$G_T = \frac{|S_{21}|^2(1 - |\Gamma_G|^2)(1 - |\Gamma_L|^2)}{|(1 - \Gamma_G S_{11})(1 - \Gamma_L S_{22}) - S_{12} S_{21} \Gamma_G \Gamma_L|^2} \tag{8.101}$$

If $S_{12}$ is set to zero, thereby invoking the unilateral approximation for the amplifier gain, then

$$G_T \approx \frac{1 - |\Gamma_G|^2}{|1 - \Gamma_G S_{11}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|(1 - \Gamma_L S_{22}|^2} \tag{8.102}$$

This approximation of course removes the possibility of analytically determining the transistor stability conditions. Using this expression, a set of constant gain circles can be drawn on a Smith chart for a given transistor. That is, for a given set of device scattering parameters and for a fixed load impedance, a set of constant gain circles can be drawn for a range of generator impedances expressed here in terms of $\Gamma_G$.

The noise figure was previously found in Eq. (8.40). As was done for the constant gain circles, constant noise figure circles can be drawn for a range of values for the generator admittance, $Y_G = G_G + jB_G$. The optimum gain occurs when $\Gamma_G = S_{11}^*$, and the minimum noise figure occurs when $Y_G = Y_{opt}$. These two source impedances are rarely the same, but a procedure is available that at least optimizes both of these parameters simultaneously [15]. As seen on the Smith chart in Fig. 8.9, it appears that the least damage to either the gain or the noise figure is obtained if the actual chosen generator impedance lies on a

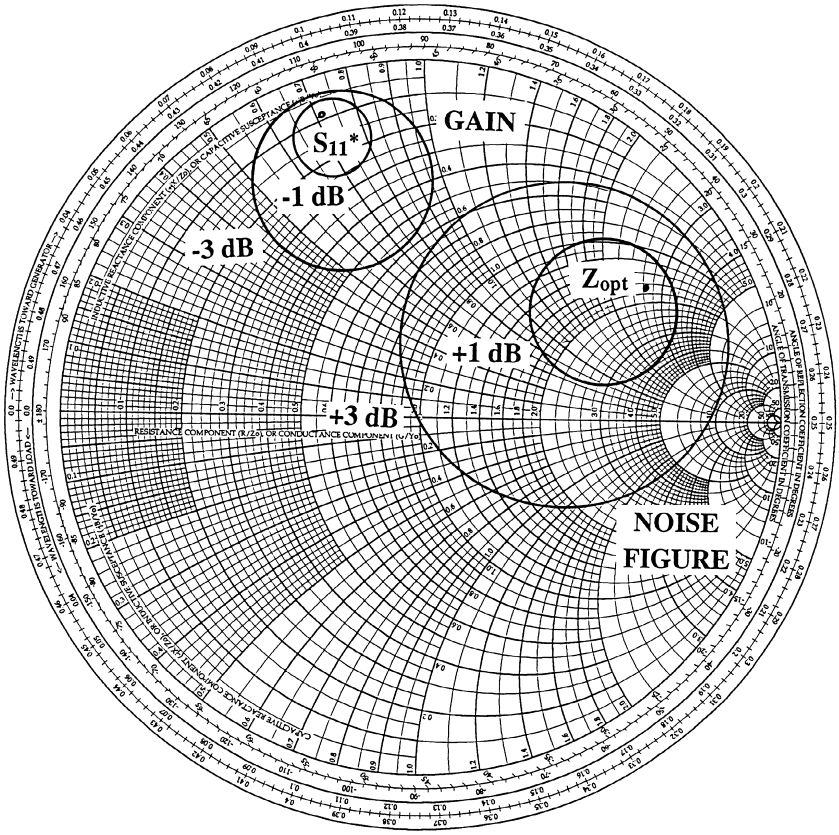



**FIGURE 8.9** Constant gain and noise figure circles.

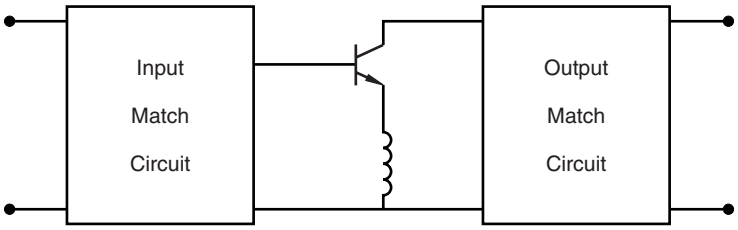



**FIGURE 8.10** Series inductive feedback can be used to lower noise figure.

line between $S_{11}^*$ and $Y_{\text{opt}}$. It has been found that addition of series inductance, such as shown in Fig. 8.10, will lower the minimum noise figure of the circuit because it lowers the effective $F_{\text{min}}$ and $r_{\text{n}}$. This series inductance also increases the real part of the input impedance. The output impedance does not affect the noise figure, but it can be manipulated to adjust the gain.

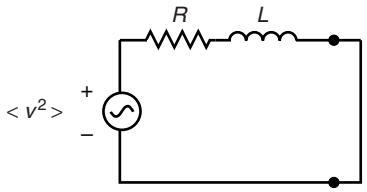



**FIGURE 8.11** Noise current developed by a series *RL* circuit.

## PROBLEMS

**8.1** What is the noise voltage at the output of Fig. 8.2 when the capacitor is removed from the circuit?

**8.2** What is the noise current from a noise voltage source in a series *RL* circuit shown in Fig. 8.11.

**8.3** Derive Eqs. (8.80) to (8.83).

**8.4** A MESFET has a base width $W = 300$ μm and at 2 GHz with a given bias is found to have $g_m = 75$ mS, $R_g = 1\ \Omega$, $R_d = 5\ \Omega$, $R_s = 3\ \Omega$, and $C_{gs} = 0.4$ pF. What are the four noise parameters $F_{min}$, $R_n$, $R_{opt}$, and $X_{opt}$? If the base width is changed to $W' = 200$ μm and the number of base fingers remains unchanged, what are the four noise parameters?

**8.5** A three-stage amplifier consists of three individual unilateral amplifiers. The first one (the input stage) has a gain $G_1 = 10$ dB, a noise figure $F_1 = 1.5$, and an efficiency $\eta_1 = 1\%$. For stage 2, $G_2 = 20$ dB, $F_2 = 10$, and $\eta_2 = 5\%$. For stage 3, $G_3 = 5$ dB, $F_3 = 15$, and $\eta_3 = 50\%$. What is the overall total noise figure for the cascaded amplifier? What is the overall total efficiency for the cascaded amplifier?

## REFERENCES

1. J. L. Plumb and E. R. Chenette, "Flicker Noise in Transistors," *IEEE Trans. Electron Devices*, Vol. 10, pp. 304–308, 1963.
2. A. Messiah, *Quantum Mechanics*, Amsterdam: North-Holland 1964, ch. 12.
3. W. A. Davis, *Microwave Semiconductor Circuit Design*, New York: Van Nostrand, 1984, chs. 8, 9.
4. A. E. Siegman, "Zero-Point Energy as the Source of Amplifier Noise," *Proc. IRE*, Vol. 49, pp. 633–634, 1961.
5. H. Nyquist, "Thermal Agitation of Electronic Charge in Conductors," *Physical Rev.*, Vol. 32, p. 110, 1928.
6. K. Kurokawa, "Actual Noise Measure of Linear Amplifiers," *Proc. IRE*, Vol. 49, pp. 1391–1397, 1961.
7. H. A. Haus, "IRE Standards on Methods of Measuring Noise in Linear Twoports," *Proc. IRE*, Vol. 47, pp. 66–68, 1959.

8. H. Rohte and W. Danlke, "Theory of Noisy Fourpoles," *Proc. IRE*, Vol. 44, pp. 811–818, 1956.
9. H. A. Haus, "Representation of Noise in Linear Twoports," *Proc. IRE*, Vol. 48, pp. 69–74, 1960.
10. H. Fukui, "Design of Microwave GaAs MESFETs for Broad-Band Low-Noise Amplifiers," *IEEE Trans. Electron Devices*, Vol. 26, pp. 1032–1037, 1979.
11. J. M. Golio, *Microwave MESFETs and HEMTs*, Norwood, MA: Artech House, 1991, ch. 2.
12. H. Fukui, "The Noise Performance of Microwave Transistors," *IEEE Trans. Electron Devices*, Vol. 13, pp. 329–341, 1966.
13. R. J. Hawkins, "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 Electronics*, Vol. 20, pp. 191–196, 1977.
14. H. T. Friis, "Noise Figures of Radio Receivers," *Proc. IRE*, Vol. 32, pp. 419–422, 1944.
15. R. E. Lehmann and D. D. Heston, "X-Band Monolithic Series Feedback LNA," *IEEE Trans. Microwave Theory Tech.*, Vol. 33, pp. 1560–1566, 1985.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER NINE

# RF Power Amplifiers

## 9.1 TRANSISTOR CONFIGURATIONS

Earlier in Chapter 7, class A amplifiers were treated. Some discussion was given to its application as a power amplifier. While class A amplifiers are used in power applications where linearity is of primary concern, they do so at the cost of efficiency. In this chapter a description of power amplifiers that provide higher efficiency than the class A amplifier. Before describing these in detail, it should be recalled that a single transistor amplifier can be installed in one of four different ways: common emitter, common base, common collector (or emitter follower), and common emitter with emitter degeneracy. Although there are always exceptions, the common emitter circuit is used in amplifiers where high voltage gain is required. The common base amplifier is used when low input impedance and high output impedance is desired. This is accompanied with a current gain $\approx 1$. The emitter follower is used when high-input impedance and low-output impedance is desired. This is accompanied with a voltage gain $\approx 1$. The common emitter with emitter degeneracy is used when improved stability is needed with respect to differences in the transistor short circuit current gain ($\beta$) with some degradation in the voltage gain. These are illustrated in Fig. 9.1 in which the bias supplies are not shown. These properties are described in detail in most electronics texts.

The transistor itself can be in one of four different states: saturation, forward active, cutoff, and reverse active. It is in the forward active region, when for the bipolar transistor, the base–emitter junction is forward biased and the base–collector junction is reverse biased. These states are illustrated in Fig. 9.2 for a npn transistor. An actual bipolar transistor requires a base–emitter voltage greater than 0.6 to 0.7 volts for it to go into the active state.

The voltage swing of a class A amplifier will remain in the forward active region throughout its entire cycle. If the signal current is given as follows:

$$i_o(\omega t) = \hat{I}_C \sin \omega t \tag{9.1}$$

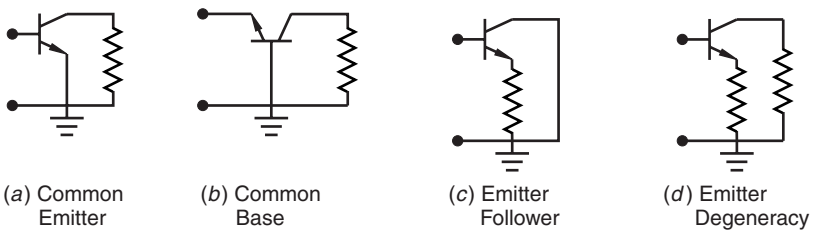



**FIGURE 9.1** (*a*) Common emitter, (*b*) common base, (*c*) common collector or emitter follower, and (*d*) common emitter with emitter degeneracy.

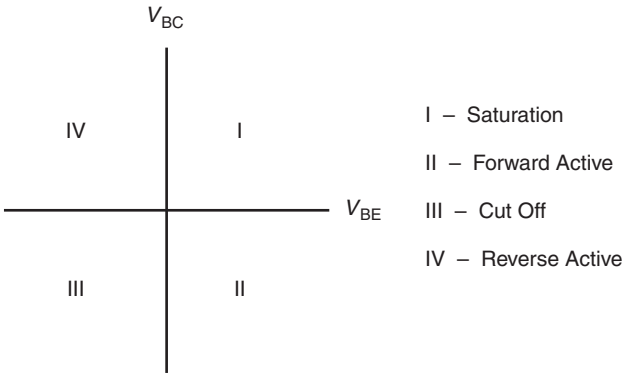



**FIGURE 9.2** The four bias regions for a npn bipolar transistor.

and the dc bias current is $I_{dc}$, then the total instantaneous current is

$$I_{dc} + \hat{I}_C \sin \omega t \tag{9.2}$$

For the class A amplifier, $\hat{I}_C < I_C$, so the entire waveform of the ac signal is amplified without distortion. The conduction angle is 360°. For the amplifiers under consideration in this chapter, the transistor(s) will be operating during part of their cycle in either cutoff or saturation, or both.

## 9.2 THE CLASS B AMPLIFIER

The class B amplifier is biased so that the transistor is on only during half of the incoming cycle. The other half of the cycle is amplified by another transistor so that at the output the full wave is reconstituted. This is illustrated in Fig. 9.3. While each transistor is clearly operating in a nonlinear mode, the total input wave is directly replicated at the output. The class B amplifier is therefore classed as a linear amplifier. In this case the bias current, $I_C = 0$. Since only one of the

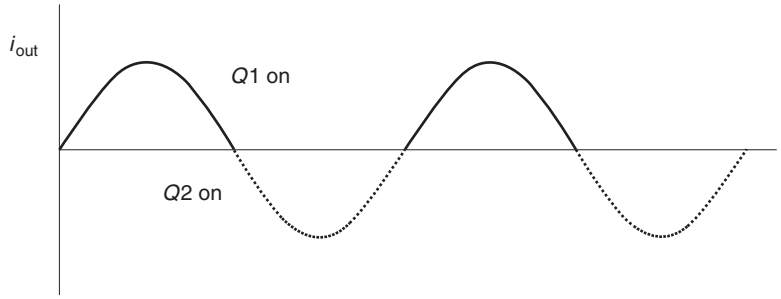



**FIGURE 9.3** The reconstituted waveform of a class B amplifier.

transistors is cut off when the total voltage is less than 0, only the positive half of the wave is amplified. The conduction angle is 180°. The term, class AB amplifier, is sometimes used to describe the case when the dc bias current is much smaller than the signal amplitude, $\hat{I}_C$, but still greater than 0. In this case,

$$180^\circ < \text{conduction angle} \ll 360^\circ$$

### 9.2.1 Complementary (npn/pnp) Class B Amplifier

Figure 9.4*a* shows a complementary type of class B amplifier. In this case transistor $Q1$ is biased so that it is in the active mode when the input voltage, $v_{in} > 0.7$ and cut off when the input signal $v_{in} < 0.7$. The other half of the signal

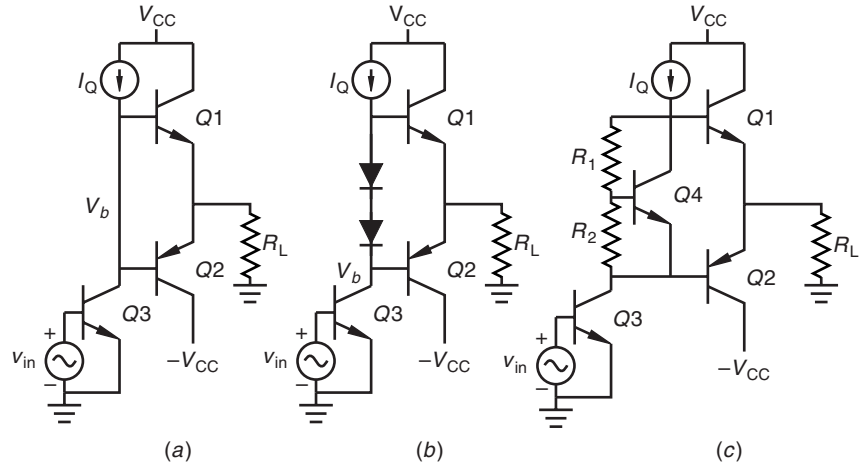



**FIGURE 9.4** (*a*) The basic complementary class B amplifier, (*b*) class B amplifier with diode compensation to reduce crossover distortion, and (*c*) class B amplifier with a $V_{BE}$ multiplier to reduce crossover distortion.

is amplified by transistor $Q2$ when $v_{in} < -0.7$. When no input signal is present, no power is drawn from the bias supply through the collectors of $Q1$ or $Q2$, so the class B operation is attractive when low standby power consumption is an important consideration. There is a small region of the input signal for which neither $Q1$ nor $Q2$ is on. The resulting output will therefore suffer some distortion.

The npn transistor $Q1$ in the class B circuit in Fig. 9.4*a* has its collector connected to the positive power supply, $V_{CC}$, and its emitter connected to the load, $R_L$. The collector of the pnp output transistor, $Q2$, has its collector connected to the negative supply voltage $V_{EE}$, which is often equal to $-V_{CC}$, and its emitter also connected to the load, $R_L$. The bases of $Q1$ and $Q2$ are connected together and are driven by the collector of the input transistor $Q3$. The input transistor, $Q3$, has a bias current source, $I_Q$ feeding its collector, which also provides base current for $Q1$. The input voltage, $v_{in}$ to the input transistor $Q3$ is what drives the output stage. It is tempting when doing hand or SPICE calculations to start with $v_{in}$. However, because a small change in base voltage of $Q3$ makes a large change in the collector voltage of $Q3$, it is easier to start the analysis at the base of $Q2$. This base voltage can be called $V_X$.

When $V_X = 0$, both the output transistors $Q1$ and $Q2$ are turned off because the voltage is less than the 0.7 volts necessary to turn the transistors on. If $V_X > 0.7$, then $Q1$ (npn) is on and $Q2$ (pnp) is off. Current is then drawn from the power supply, $V_{CC}$, through $Q1$ to produce the positive half-wave of the signal in the load. If $V_X < 0.7$, then $Q1$ (npn) is off and $Q2$ (pnp) is on. The voltage $V_X$ is made negative by turning $Q3$ on thus bringing the collector voltage of $Q3$ closer to $V_{EE}$ which is less than zero. An extreme positive or negative input voltage puts the turned on output transistor (either $Q1$ or $Q2$) into saturation. The maximum positive output voltage is

$$V_O^+ = V_{CC} - V_{CE1(sat)} \tag{9.3}$$

and the maximum negative output voltage is

$$V_O^- = -V_{EE} + V_{EC2(sat)} \tag{9.4}$$

Typically the value for $V_{CE(sat)} \approx 0.2$ volts for a bipolar transistor. More design details are available from a variety of sources, such as [1].

### 9.2.2 Elimination of the Dead Band

The 1.2 to 1.4 volt range in the base voltages of $Q1$ and $Q2$ can be substantially compensated by addition of two diodes in series between the bases of $Q1$ and $Q2$ (Fig. 9.4*b*). These diodes are named, respectively, $D4$ and $D5$. For purposes of calculation, let $V_X$ stand for the voltage at the collector of the driver transistor $Q3$, which is the same as the base voltage for the pnp output transistor $Q2$. To get to the base of $Q1$ from $V_X$ now requires going through the two series-connected diodes "backwards," from cathode to anode. If $V_X > 0$ but not so high as to turn

off the diodes $D4$ and $D5$, then $Q1$ is on as described Fig. 9.4*a*. The voltage across the load is

$$V_{\mathrm{O}}^{+} = V_X + V_{D4} + V_{D3} - V_{\mathrm{BE1}} \tag{9.5}$$

To make $V_X < 0$, the input voltage to the driver $Q3$ must be a positive voltage. The npn output transistor $Q1$ is turned off, and the excess bias current from $I_Q$ flows through the diodes $D4$ and $D5$ and then through the now turned on $Q3$. The output voltage is not affected directly by the diodes now:

$$V_{\mathrm{O}}^{-} = V_{\mathrm{EB2}} + V_X \tag{9.6}$$

Under this condition, the value of $V_X$ is actually a negative number. In the middle when $V_X = 0$, the output voltage across $R_{\mathrm{L}}$ is

$$V_{\mathrm{O}}^{+} = V_{D4} + V_{D5} - V_{\mathrm{BE1}} \approx V_{\mathrm{BE}} \tag{9.7}$$

and

$$V_{\mathrm{O}}^{-} = V_{\mathrm{BE2}} = V_{\mathrm{BE}} \tag{9.8}$$

If the forward diode voltage drops are equal to the base–emitter drops of the transistors, there is no discontinuity in $V_{\mathrm{O}}$ in going from negative to positive input voltages.

In actual production circuits, tight specifications are needed on diodes $D4$ and $D5$, since they are in the base circuit of the output transistors and consequently carry much less current than the output power devices. The discrepancy between the high-power and low-power devices can be alleviated by using the $V_{\mathrm{BE}}$ multiplier shown in Fig. 9.4*c*. In this circuit the base–emitter voltage of $Q4$ sets the current through $R_2$:

$$I_{R2} = \frac{V_{\mathrm{BE4}}}{R_2} \tag{9.9}$$

Assuming the base current of $Q4$ is negligible, the voltage drop between the bases of the output transistors $Q1$ and $Q2$ is

$$V_{\mathrm{CE4}} = I_{R2}(R_1 + R_2) = V_{\mathrm{BE4}}\left(\frac{R_2}{R_1} + 1\right) = V_{\mathrm{BE1}} + V_{\mathrm{EB2}} \tag{9.10}$$

When the voltage at the base of $Q2$ is positive, the load voltage is

$$V_{\mathrm{O}}^{+} = V_X + V_{\mathrm{BE4}}\left(\frac{R_2}{R_1} + 1\right) - V_{\mathrm{BE1}} \tag{9.11}$$

and where $V_X < 0$,

$$V_{\mathrm{O}}^{-} = V_{\mathrm{EB2}} + V_X \tag{9.12}$$

In the middle where $V_X = 0$ the $V_O^+$ and $V_O^-$ can be forced to be equal by adjustment of the resistors $R_1$ and $R_2$,

$$V_O = V_{BE4}\left(\frac{R_2}{R_1} + 1\right) - V_{BE1} = V_{EB2} \tag{9.13}$$

In addition to reducing or eliminating the dead band zone, the compensation circuits in Fig. 9.4*b* and 9.4*c* also provide for temperature stability, since a change in the temperature changes the transistor $V_{BE}$ value. The compensation circuit and the power transistors vary in the same way with temperature, since they are physically close together.

Another aspect that deserves attention is the actual value of the current source, $I_Q$. Since this supplies the base current for the npn output transistor $Q1$, $I_Q$ must be large enough to not "starve" $Q1$ when it is drawing the maximum current through its collector. This means that $I_Q \geq I_{C1}/\beta_1$.

### 9.2.3 The Composite pnp Transistor

One of the primary problems in using this type of class B amplifier is the requirement for obtaining two equivalent complementary transistors. The problem fundamentally arises because of the greater mobility of electrons by over a 3 : 1 factor over that of holes in silicon. The symmetry of the gain in this circuit depends on the two output transistors having the same short circuit base to collector current gain, $\beta = i_c/i_b$. When it is not possible to obtain a high $\beta$ pnp transistor, it is sometimes possible to use a composite transistor connection. A high-power npn transistor, $Q1$, is connected to a low-power low $\beta$ pnp transistor, $Q2$, as shown in Fig. 9.5. Normally the base–emitter junctions of the composite and single pnp transistor are forward biased so that the Shockley diode equation may be used to describe the bias currents. For $Q2$ in the composite circuit,

$$I_{C2} = -I_S e^{qV_{EB}/kT} \tag{9.14}$$

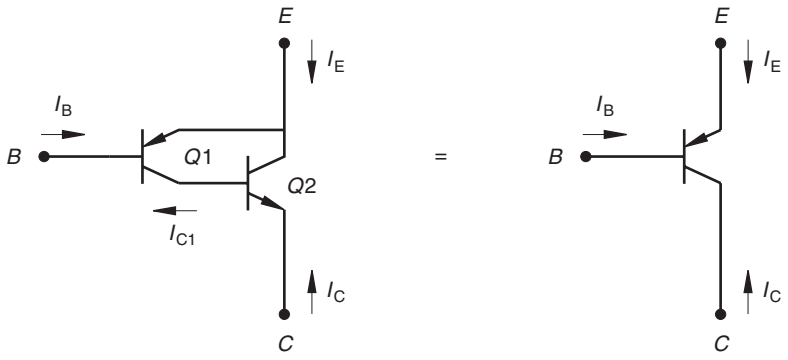



**FIGURE 9.5** A composite connection for a pnp transistor.

The collector current for $Q1$ in the composite circuit is the same as the collector current for the single pnp transistor:

$$I_C = (\beta_1 + 1)I_{C2} = -(\beta_1 + 1)I_S e^{V_{BE}/kT} \tag{9.15}$$

The composite circuit has the polarity of a pnp transistor with the gain of an npn transistor.

### 9.2.4 Small Signal Analysis

The three fundamental parameters that characterize an amplifier are its voltage gain, $A_v$, input resistance, $R_{in}$, and output resistance, $R_{out}$. In the circuit shown in Fig. 9.4*a*, neither $Q1$ nor $Q2$ are on simultaneously. If $Q1$ is on, $Q2$ is an open circuit and need not be considered as part of the ac analysis. A small signal hybrid $\pi$ model (Fig. 9.6) for a bipolar transistor consists of a base resistance, $r_b$, base–emitter resistance, $r_\pi$, collector–emitter resistance, $r_o$, transconductance, $g_m$, and short circuit current gain $\beta = g_m r_\pi$. There are in addition high-frequency effects caused by reactive parasitic elements within the device. Since the voltage gain of an emitter follower is $\approx 1$, the voltage gain of the $Q3$ and $Q1$ combination is

$$A_v = -g_{m3} R_{L(eff)} \tag{9.16}$$

The effective load resistance $R_{L(eff)}$ seen by the first transistor, $Q3$, is the same as the input resistance of the emitter follower circuit $Q1$. Circuit analysis of the low-frequency transistor hybrid model would give (Appendix F)

$$R_{L(eff)} = r_{\pi 1} + r_{b1} + (\beta_1 + 1)(r_{o1} || R_L) \tag{9.17}$$

$$\approx \beta_1 R_L \tag{9.18}$$

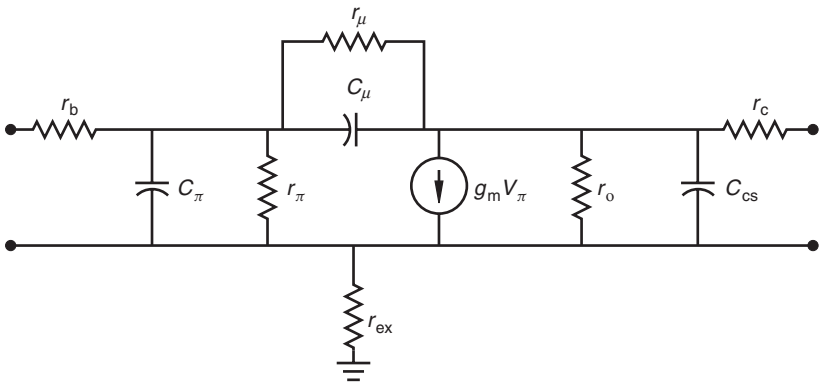



**FIGURE 9.6** The small signal hybrid $\pi$ model of a bipolar transistor.

The voltage gain is then found by substitution:

$$A_v^+ = -g_{m3}[r_{\pi 1} + r_{b1} + (\beta_1 + 1)(r_{o1} || R_L)] \tag{9.19}$$

$$\approx -g_{m3}\beta_1 R_L \tag{9.20}$$

The low-frequency input resistance to the actual class B amplifier is given by $R_{L(eff)}$ in Eq. (9.17), and the output resistance is

$$R_{out} = \frac{r_\pi + R_{bb} + r_b}{1 + \beta} \tag{9.21}$$

Thus the input resistance is high and the output resistance is low for a class B amplifier, which enables it to drive a low-impedance load with high efficiency.

### 9.2.5 All npn Class B Amplifier

The complementary class B amplifier shown in Fig. 9.4 needs to have symmetrical npn and pnp devices. In addition this circuit also requires complementary power supplies. These two problems can be alleviated by using the totem pole or all npn transistor class B amplifier. This circuit requires only one power supply, and it has identical npn transistors that amplify both the positive and negative halves of the signal. However, it requires that the two transistors operate with an input phase differential of 180°. This circuit is illustrated in Fig. 9.7. Clearly, the cost of the all npn transistor amplifier is the added requirement of two center-tapped transformers. These are necessary to obtain 180° phase difference between $Q1$ and $Q2$. The center tapped transformer also provides dc isolation for the load. When the input voltage is positive, $Q1$ is on and $Q2$ is off. When the input voltage is negative, the input transformer induces a positive voltage at the "undotted" secondary winding which turns $Q2$ on. The output of $Q2$ will induce on the output transformer a positive voltage on the "undotted" terminal and a negative voltage on the "dotted" terminal. The negative input voltage swing is thus replicated as a negative voltage swing at the output. The transformer turns

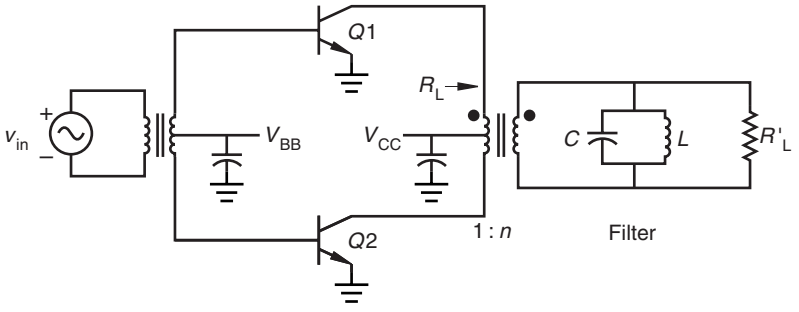



**FIGURE 9.7** The all npn class B amplifier.

ratio can be used for impedance matching. The output filter is used to filter out any harmonics caused by crossover or other sources of distortion. The filter is not necessary to achieve class B operation, but it can be helpful.

### 9.2.6 Class B Amplifier Efficiency

The maximum efficiency of a class B amplifier is found by finding the ratio of the output power delivered to the load to the required dc power from the bias voltage supply. In determining efficiency in this way, power losses caused by nonzero base currents and crossover distortion compensation circuits used in Fig. 9.4*b* and 9.4*c* are neglected. Furthermore the power efficiency rather than the power added efficiency is calculated so as to form a basis for comparison for alternative circuits. It is sufficient to do the calculation during the part of the cycle when $Q1$ is on and $Q2$ is off. The load resistance in Fig. 9.7 is transformed through to the primary side of the output transformer, loading the transistors with a value of $R_L$.

The magnitude of the collector current that flows into $R_L$ is $\hat{I}_C$. The ac current is

$$i_o(\omega t) = \hat{I}_C \sin(\omega t) \tag{9.22}$$

and the voltage is

$$v_o(\omega t) = \hat{I}_C R_L \sin(\omega t) \tag{9.23}$$

Since the collector–base voltage must remain positive to avoid the danger of burning out the transistor, $\hat{V}_C = \hat{I}_C R_L \leq V_{CC}$. The maximum allowable output power delivered to the load is

$$P_o = \frac{\hat{V}_C^2}{2R_L} \tag{9.24}$$

Now a determination of the dc current supplied by the bias supply is needed. The magnitude of the current delivered by the bias supply to the load by $Q1$ is

$$i_{BB1} = \hat{I}_C \sin(\omega t), \qquad 0 < \omega t < \pi \tag{9.25}$$

and for $Q2$,

$$i_{BB2} = -\hat{I}_C \sin(\omega t), \qquad \pi < \omega t < 2\pi \tag{9.26}$$

The total current is then $\hat{I}_C |\sin(\omega t)|$, which is shown in Fig. 9.8. The dc current from the bias source(s) is found by finding the average current:

$$\begin{aligned} I_{dc} &= \frac{1}{T}\int_0^{T/2} \hat{I}_{C1} \sin \omega t dt \\ &= -\frac{\hat{I}_{C1}}{\omega T} \cos \omega t \Big|_0^{T/2} \end{aligned}$$

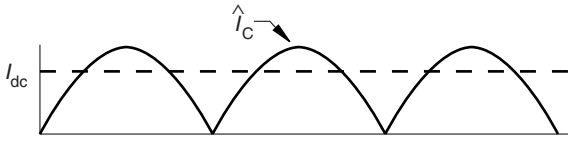



**FIGURE 9.8** Waveform for finding the average dc current from the power supply.

$$= -\frac{\hat{I}_{C1}}{(2\pi/T)T}\left[\cos\left(\frac{2\pi}{T}\frac{T}{2}\right) - 1\right]\Bigg|_0^{T/2}$$

$$= -\frac{\hat{I}_{C1}}{2\pi}[-1-1] \tag{9.27}$$

$$I_{dc} = \frac{\hat{I}_{C1}}{\pi} = \frac{1}{\pi}\frac{V_O}{R_L} \tag{9.28}$$

The power drawn from both of the power supplies by both of the output transistors is

$$P_{supply} = 2V_{CC}I_{dc} = \frac{2}{\pi}\frac{V_{CC}}{R_L}\cdot V_O \tag{9.29}$$

Thus the output power is proportional to $V_O$, and is the average power drawn from the power supply. The power delivered to the load is

$$P_L = \frac{|V_O|^2}{2R_L} \tag{9.30}$$

The efficiency is the ratio of these latter two values:

$$\eta = \frac{P_L}{P_{supply}} = \frac{|V_O|^2}{2R_L}\frac{\pi}{2}\frac{R_L}{V_{CC}V_O} \tag{9.31}$$

$$\eta = \frac{\pi}{4}\frac{V_O}{V_{CC}} \tag{9.32}$$

The maximum output power occurs when the output voltage is $V_{CC} - V_{CE(sat)}$:

$$P_{L(max)} = \frac{1}{2}\frac{(V_{CC} - V_{CE(sat)})^2}{R_L} \tag{9.33}$$

$$\eta_{max} = \frac{\pi}{4}\frac{V_{CC} - V_{CE(sat)}}{V_{CC}} \approx 78.5\% \tag{9.34}$$

This efficiency for the class B amplifier should be compared with the maximum efficiency of a class A amplifier, where $\eta_{max} = 25\%$ when the bias to the collector

is supplied through a resistor and $\eta_{max} = 50\%$ when the bias to the collector is supplied through an RF choke.

## 9.3 THE CLASS C AMPLIFIER

The class C amplifier is useful for providing a high-power continuous wave (CW) or frequency modulation (FM) output. When it is used in amplitude modulation schemes, the output variation is done by varying the bias supply. There are several characteristics that distinguish the class C amplifier from the class A or B amplifiers. First of all it is biased so that the transistor conduction angle is $<180^{\circ}$. Consequently the class C amplifier is clearly nonlinear in that it does not directly replicate the input signal like the class A and B amplifiers do (at least in principle). The class A amplifier requires one transistor, the class B amplifier requires two transistors, and the class C amplifier uses one transistor. Topologically it looks similar to the class A except for the dc bias levels. It was noted that in the class B amplifier, an output filter is used optionally to help clean up the output signal. In the class C amplifier, such a tuned output is necessary in order to recover the sine wave. Finally, class C operation is capable of higher efficiency than either of the previous two classes, so for the appropriate signal types they become very attractive as power amplifiers.

The class C amplifier shown in Fig. 9.9 gives the output circuit for a power bipolar transistor (BJT) with the required tuned circuit. An *N*-channel enhancement-mode metal oxide semiconductor field effect transistor (MOSFET) can be used in place of the BJT. The $Q$ of the tuned circuit will determine the bandwidth of the amplifier. The large inductance RF coil in the collector voltage supply ensures that only dc current flows there. During that part of the input cycle when the transistor is on, the bias supply current flows through the transistor and the output voltage is approximately 90% of $V_{CC}$. When the transistor is off, the supply current flows into the blocking capacitor. The current waveform at the

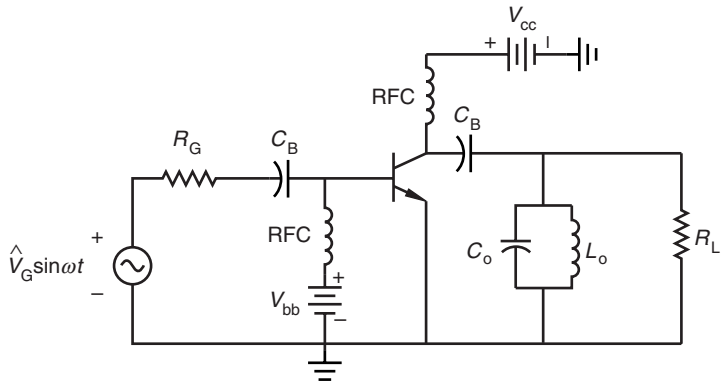



**FIGURE 9.9** A simple class C amplifier where $V_{BB}$ determines the conduction angle.

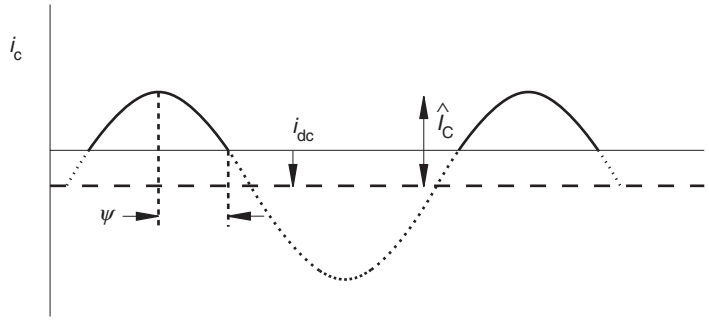



**FIGURE 9.10** The collector current waveform for class C operation.

collector can be modeled as the waveform shown in Fig. 9.10:

$$i_C(\omega t) = \begin{cases} I_C - \hat{I}_C \sin(\omega t), & -\psi \le \omega t \le +\psi \\ 0, & \text{otherwise} \end{cases} \tag{9.35}$$

For class C operation, the magnitude of the quiescent current is $|I_C| < \hat{I}_C$. The point where quiescent current equals the total current is

$$i_C\left(\frac{3\pi}{2} \pm \psi\right) = 0 = I_C - \hat{I}_C \sin\left(\frac{3\pi}{2} \pm \psi\right)$$

$$0 = I_C + \hat{I}_C \cos\psi \tag{9.36}$$

This determines the value of the quiescent current in terms of the conduction angle $2\psi$:

$$I_C = -\hat{I}_C \cos\psi \tag{9.37}$$

The dc current from the power supply is the average of the total collector current $i_C(\theta)$:

$$\begin{aligned} I_{dc} &= \frac{1}{2\pi}\int_0^{2\pi} i_C(\theta)d\theta \\ &= \frac{1}{2\pi}\int_{(3\pi/2)-\psi}^{(3\pi/2)+\psi} (I_C - \hat{I}_C \sin\theta)d\theta \\ &= \frac{1}{\pi}(\psi I_C + \hat{I}_C \sin\psi) \end{aligned} \tag{9.38}$$

Evaluation of the integral makes use of the trigonometric identity, $\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$. From Eq. (9.36) the dc current is

$$I_{\mathrm{dc}} = \frac{\hat{I}_{\mathrm{C}}}{\pi}(\sin\psi - \psi\cos\psi) \tag{9.39}$$

This gives the dc current from the power supply in terms of $\hat{I}_{\mathrm{C}}$ and the conduction angle $\psi$, so power supplied by the source is

$$P_{\mathrm{in}} = V_{\mathrm{CC}} I_{\mathrm{dc}} \tag{9.40}$$

The ac component of the current flows through the blocking capacitor and into the load. Harmonic current components are shorted to ground by the tuned circuit. The magnitude of the output voltage at the fundamental frequency is found using the Fourier method:

$$\hat{V}_{\mathrm{O}} = -\frac{1}{\pi}\int_0^{2\pi} i_{\mathrm{C}}(\theta) R_{\mathrm{L}} \sin\theta d\theta \tag{9.41}$$

$$= -\frac{R_{\mathrm{L}}}{\pi}\int_{(3\pi/2)-\psi}^{(3\pi/2)+\psi} (I_{\mathrm{C}} - \hat{I}_{\mathrm{C}}\sin\theta)\sin\theta d\theta$$

$$= \frac{R_{\mathrm{L}}}{\pi}\left[2I_{\mathrm{C}}(-\cos\theta)\Big|_{(3\pi/2)-\psi}^{(3\pi/2)+\psi} + \frac{\hat{I}_{\mathrm{C}}}{2}\left(\theta - \frac{\sin 2\theta}{2}\right)\Big|_{(3\pi/2)-\psi}^{(3\pi/2)+\psi}\right] \tag{9.42}$$

$$= \frac{R_{\mathrm{L}}}{\pi}\left[2I_{\mathrm{C}}\sin\psi + \frac{\hat{I}_{\mathrm{C}}}{2}\left\{(2\psi) - \frac{1}{2}[\sin(3\pi + 2\psi) - \sin(3\pi - 2\psi)]\right\}\right] \tag{9.43}$$

The quiescent current term, $I_{\mathrm{C}}$, is replaced by Eq. (9.37) again, and the trigonometric identity for $\sin\alpha\cos\beta$ is used:

$$\hat{V}_{\mathrm{O}} = \frac{R_{\mathrm{L}}\hat{I}_{\mathrm{C}}}{\pi}\left[-\sin 2\psi + \psi - \frac{1}{4}(-\sin 2\psi - \sin 2\psi)\right] \tag{9.44}$$

$$\hat{V}_{\mathrm{O}} = \frac{R_{\mathrm{L}}\hat{I}_{\mathrm{C}}}{2\pi}(2\psi - \sin 2\psi) \tag{9.45}$$

The ac output power delivered to the load is

$$P_{\mathrm{O}} = \frac{\hat{V}_{\mathrm{O}}^2}{2R_{\mathrm{L}}} \tag{9.46}$$

The efficiency (neglecting the input power) is simply the ratio of the output ac power to the input dc power. The maximum output power occurs when $\hat{V}_{\mathrm{O}} =$

$V_{CC}$. The maximum efficiency is then [2]

$$\eta_{max} = \frac{P_{Omax}}{P_{dc}} = \left(\frac{V_{CC}^2}{2R_L}\right)\left(\frac{1}{V_{CC}I_{dc}}\right) \tag{9.47}$$

$$= \frac{2\psi - \sin 2\psi}{4(\sin\psi - \psi\cos\psi)} \tag{9.48}$$

A plot of this expression (Fig. 9.11) clearly illustrates the efficiency in terms of the conduction angle for class A, B, and C amplifiers. The increased efficiency of the class C amplifier is a result of the collector current flowing for less than a half-cycle. When the collector current is maximum, the collector voltage is minimum, so the power dissipation is inherently lower than class B or class A operation.

Another important parameter for the power amplifier is the ratio of the maximum average output power where $\hat{V}_O = V_{CC}$, to the peak instantaneous output power:

$$r = \frac{P_{Omax}}{V_{Cmax}i_{Cmax}} \tag{9.49}$$

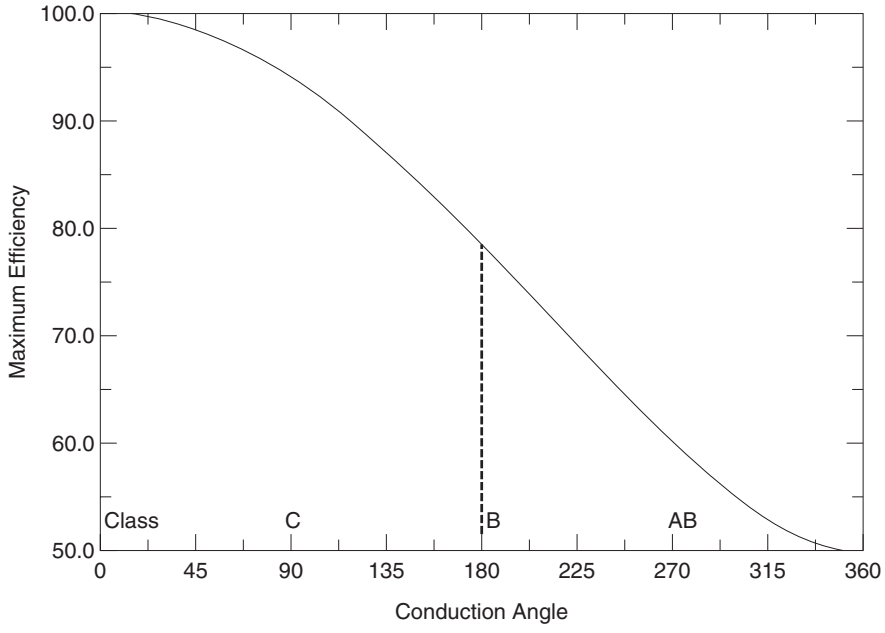



**FIGURE 9.11** Power efficiency for class A, B, and C amplifiers.

The maximum average output power occurs when $\hat{V}_O = V_{CC}$ and is given by

$$P_{Omax} = \frac{V_{CC}^2}{2R_L} \tag{9.50}$$

$$= \frac{\hat{I}_C^2 R_L^2}{4\pi^2} \frac{(2\psi - \sin 2\psi)^2}{2R_L} \tag{9.51}$$

The maximum voltage at the collector is the output voltage ac voltage swing plus the bias voltage:

$$V_{Cmax} = 2V_{CC} = \frac{\hat{I}_C R_L}{2\pi}(2\psi - \sin 2\psi) \tag{9.52}$$

The maximum current is

$$i_{Cmax} = I_C + \hat{I}_C = -\hat{I}_C \cos\psi + \hat{I}_C. \tag{9.53}$$

The ratio of the maximum average power to the peak power from (9.49) is [2]

$$r = \frac{2\psi - \sin 2\psi}{8\pi(1 - \cos\psi)} \tag{9.54}$$

A plot of this ratio as a function of conduction angle in Fig. 9.12 shows that maximum efficiency of the class C amplifier occurs when there is no output

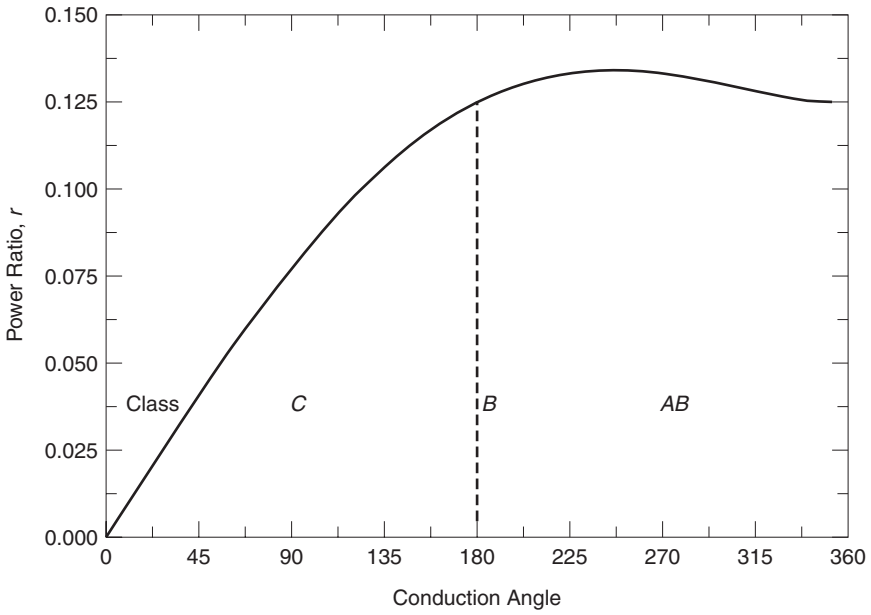



**FIGURE 9.12** Maximum output power to the peak power ratio.

power. Nevertheless, Figs. 9.11 and 9.12 indicate the trade-offs in choosing the appropriate conduction angle for class C operation.

## 9.4 CLASS C INPUT BIAS VOLTAGE

Device SPICE models for RF power transistors are relatively rare. Manufacturers often do supply optimum generator and load impedances that have been found to provide the rated output power for the designated frequency. The circuit shown in Fig. 9.9 is a generic example of a 900 MHz amplifier with a bandwidth of 18 MHz in which the manufacturer has determined empirically the optimum $Z_G$ and $Z_L$. The $Q$ of the output tuned circuit then is

$$Q = \frac{f_0}{\Delta f} \tag{9.55}$$

The $Q$ determines the inductance and capacitance of the output resonant circuit:

$$C = \frac{Q}{\omega_0 R_L} \tag{9.56}$$

$$L = \frac{R_L}{\omega_0 Q} \tag{9.57}$$

Furthermore, if the desired output power is $P_{Omax}$, the collector voltage source, $V_{CC}$, and the maximum collector current is $i_{cmax}$, then the average to peak power ratio, $r$, is found from Eqs. (9.49) and (9.54). Iterative solution of Eq. (9.54) gives the value for the conduction angle, $\psi$. This will then allow for estimation of the maximum efficiency from Eq. (9.48). Alternatively, for a given desired efficiency, the conduction angle, $\psi$, can be obtained by iterative solution of Eq. (9.48). Numerically it is useful to take the natural logarithm of Eq. (9.48) before searching iteratively for a solution:

$$\ln \eta_{max} = \ln(2\psi - \sin 2\psi) - \ln[4(\sin \psi - \psi \cos \psi)] \tag{9.58}$$

The efficiency expression can be modified to account for the non zero saturation collector–emitter voltage.

$$\eta_{max} = \frac{2\psi - \sin 2\psi}{4(\sin \psi - \psi \cos \psi)} \left( \frac{V_{CC} - V_{sat}}{V_{CC}} \right) \tag{9.59}$$

To achieve the desired conduction angle, the emitter–base junction must be biased so that the transistor will be in conduction for the desired portion of the input signal. Collector current flows when $V_{BE} > 0.7$. First it is necessary to determine the required generator voltage amplitude, $\hat{V}_G$, that will produce the desired maximum output current. This is illustrated in Fig. 9.13 where the input ac signal

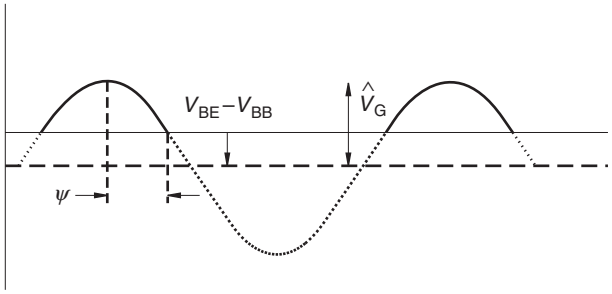



**FIGURE 9.13** Conduction angle dependence on $V_{BB}$.

is superimposed on the emitter bias voltage. The input voltage commences to rise above the turn on voltage of the transistor at

$$\hat{V}_{BB} = V_{BE} - \hat{V}_G \cos\psi \tag{9.60}$$

In this way the base bias voltage is determined.

## 9.5 THE CLASS D POWER AMPLIFIER

Inspection of the efficiency and output power of a class C amplifier reveals that 100% efficiency only occurs when the output power is zero. A modification of class B operation shown in [3] indicates that judicious choice of bias voltages and circuit impedances provide a clipped voltage waveform at the collector of the BJT while, in the optimum case, retaining the half sine wave collector current. In the limit the clipped waveform becomes a square wave. This is no longer linear, and thus is distinguished from the class B amplifier.

The class D amplifier shown in Fig. 9.14 superficially looks like a class B amplifier except for the input side bias. In class D operation the transistors act as

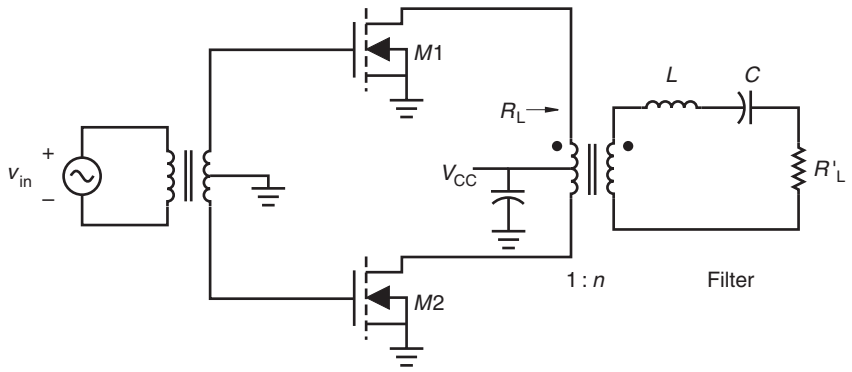



**FIGURE 9.14** Class D power amplifier.

near ideal switches that are on half of the time and off half of the time. The input ideally is excited by a square wave. If the transistor switching time is near zero, then the maximum drain current flows while the drain–source voltage, $V_{DS} = 0$. As a result 100% efficiency is theoretically possible. In practice the switching speed of a bipolar transistor is not sufficiently fast for square waves above a few MHz, and the switching speed for field effect transistors is not adequate for frequencies above a few tens of MHz.

## 9.6 THE CLASS F POWER AMPLIFIER

"A class F amplifier is characterized by a load network that has resonances at one or more harmonic frequencies as well as at the carrier frequency" [2, p. 454]. The class F amplifier was one of the early methods used to increase amplifier efficiency and has attracted some renewed interest recently. The circuit shown in Fig. 9.15 is a third harmonic peaking amplifier where the shunt resonator is resonant at the fundamental and the series resonator at the third harmonic. More details on this and higher-order resonator class F amplifiers are found in [4]. When the transistor is excited by a sinusoidal source, it is on for approximately half the time and off for half the time. The resulting output current waveform is converted back to a sine wave by the resonator, $L_1$, $C_1$. The $L_3$, $C_3$ resonator is not quite transparent to the fundamental frequency, but blocks the third harmonic energy from getting to the load. The drain or collector voltage, it would seem, will range from 0 to twice the power supply voltage with an average value of $V_{CC}$. The third harmonic voltage on the drain or collector, if it has the appropriate amplitude and phase, will tend to make this device voltage more square in shape. This will make the transistor act more like a switch with the attendant high efficiency. It is found that maximum "squareness" is achieved if the third harmonic voltage is 1/9th of the fundamental voltage. This will give a maximum collector efficiency of 88.4% [2, pp. 454–456].

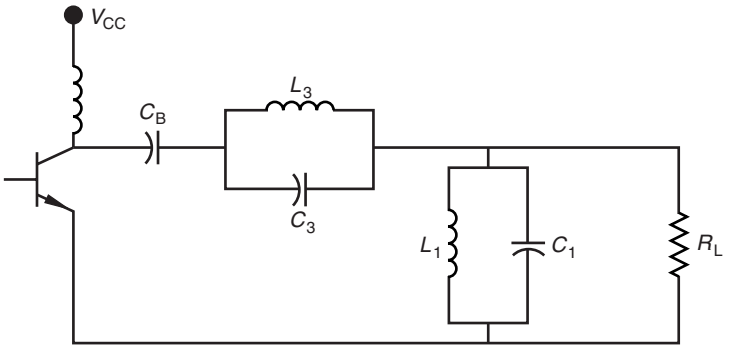



**FIGURE 9.15** Class F third harmonic peaking power amplifier.

The Fourier expansion of a square wave with amplitude from +1 to −1 and period $2\pi$ is

$$\frac{4}{\pi}\left(\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \cdots\right)$$

Consequently, to produce a square wave voltage waveform at the transistor terminal, the load must be a short at even harmonics and large at odd harmonics. Ordinarily only the fundamental second harmonic and third harmonic impedances are determined. In the typical class F amplifier shown in Fig. 9.15, the $L_1C_1$ tank circuit is resonant at the output frequency, $f_0$, and the $L_3C_3$ tank circuit is resonant at $3f_0$. It has been pointed out [5] that the blocking capacitor, $C_B$, could be used to provide a short to ground at $2f_0$ rather than simply acting as a dc block.

The design of the class F amplifier final stage proceeds by first determining $C_1$ from the desired amplifier bandwidth. The circuit $Q$ is assumed to be determined solely by the $L_1$, $C_1$, and $R_L$. Thus

$$Q = \omega_0 C_1 R_L = \frac{\omega_0}{\Delta\omega}$$

or

$$C_1 = \frac{1}{R_L \Delta\omega} \tag{9.61}$$

Once $C_1$ is determined, the inductance must be that which resonates the tank at $f_0$:

$$L_1 = \frac{1}{\omega_0^2 C_1} \tag{9.62}$$

At $2f_0$ the $L_1C_1$ tank circuit has negative reactance, and the $L_3C_3$ tank circuit has positive reactance. The capacitances $C_B$ and $C_3$ can be set to provide a short to ground:

$$-\frac{1}{2\omega_0 C_B} + \frac{2\omega_0 L_3}{1-(2\omega_0)^2 L_3 C_3} + \frac{2\omega_0 L_1}{1-(2\omega_0)^2 L_1 C_1} = 0 \tag{9.63}$$

On multiplying through Eq. (9.63) by $\omega_0$ and recognizing that $L_3C_3 = 1/9\omega_0^2$, $L_1C_1 = 1/\omega_0^2$, this equation reduces to

$$0 = -\frac{1}{C_B} + \frac{2/(9C_3)}{1-4/9} + \frac{2/C_1}{1-4}$$

or

$$\frac{1}{C_B} = \frac{4}{5C_3} - \frac{4}{3C_1} \tag{9.64}$$

which is the requirement for series resonance at $2f_0$. In addition, at the fundamental frequency, $C_B$ and the $L_3C_3$ tank circuit can be tuned to provide no

reactance between the transistor and the load, $R_L$. This eliminates the approximation that the $L_3C_3$ has zero reactance at the fundamental:

$$0 = -\frac{1}{\omega_0 C_B} + \frac{\omega_0 L_3}{1 - \omega_0^2 L_3 C_3} \tag{9.65}$$

$$C_B = 8C_3 \tag{9.66}$$

This value for $C_B$ can be substituted back into Eq. (9.64) to give a relationship between $C_3$ and $C_1$:

$$C_3 = \frac{81}{160} C_1 \tag{9.67}$$

In summary, $C_1$ is determined by the bandwidth Eq. (9.61), $L_1$ by Eq. (9.62), $C_3$ from Eq. (9.67), $L_3$ from its requirement to resonate $C_3$ at $3f_0$, and finally $C_B$ from Eq. (9.66). These equations are slightly different than those given by [5], but numerically they give very similar results. In addition interstage networks are presented in [5] that aim at reducing the spread in circuit element values, and hence this helps make circuit design practical.

Additional odd harmonics can be controlled by adding resonators that make the collector voltage have a more square shape. In effect an infinite number of odd harmonic resonators can be added if a $\lambda/4$ transmission line at the fundamental frequency replaces the lumped element third harmonic resonator (Fig. 9.16). This of course is useful only in the microwave frequency range where the transmission line length is not overly long.

At the fundamental the admittance seen by the collector is

$$Y'_L = \frac{Y_0^2}{(1/R_L) + sC_1 + (1/sL_1)} \tag{9.68}$$

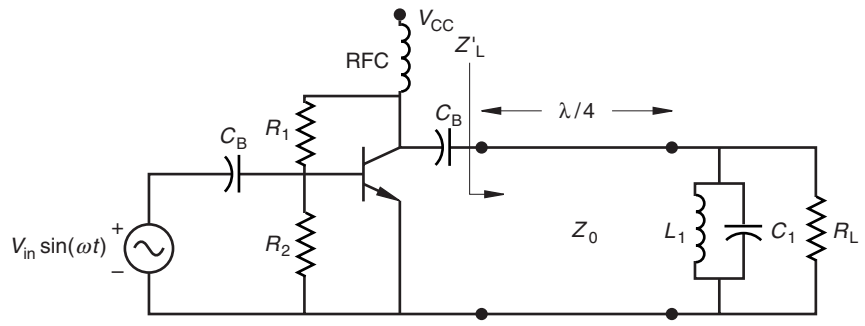



**FIGURE 9.16** Class F transmission line power amplifier. Here $C_B = 1$ μF, $Z_0 = 20$ Ω, $C_1 = 936.6$ pF, $L_1 = 33.39$ pH, $R_L = 42.37$ Ω, $V_{CC} = 24$, $R_1 = 5$ kΩ, $R_2 = 145$ kΩ.

The $\lambda/4$ transmission line in effect converts the shunt load to a series load:

$$Z'_{\mathrm{L}} = \frac{Z_0^2}{R_{\mathrm{L}}} + sC_1Z_0^2 + \frac{Z_0^2}{sL_1} \tag{9.69}$$

in which

$$R'_{\mathrm{L}} = \frac{Z_0^2}{R_{\mathrm{L}}}$$
$$L' = C_1Z_0^2$$
$$C' = \frac{Z_0^2}{sL_1}$$

At the second harmonic, the transmission line is $\lambda/2$ and the resonator $(L, C)$ is a short, so $Z'_{\mathrm{L}}(2\omega_0) = 0$. The effective load for all the harmonics can easily found at each of the harmonics:

$$Z'_{\mathrm{L}}(2\omega_0) = 0, \qquad \lambda/2$$
$$Z'_{\mathrm{L}}(3\omega_0) = \infty, \qquad 3\lambda/4$$
$$Z'_{\mathrm{L}}(4\omega_0) = 0, \qquad \lambda$$
$$Z'_{\mathrm{L}}(5\omega_0) = \infty, \qquad 5\lambda/4$$
$$\vdots$$

While this provides open and short circuits to the collector, it is not obvious that these impedances, which act in parallel with the output impedance of the transistor, will provide the necessary amplitude and phase that would produce a square wave at the collector.

An example of a class F amplifier design using the idealistic default SPICE bipolar transistor model illustrates what these waveforms might look like. As in the class C amplifier example, assume that the center frequency is 900 MHz, that the bandwidth is 18 MHz, and consequently that the circuit $Q = 50$. Furthermore, as in the class C amplifier example, assume that the collector looks into a load resistance of $R'_{\mathrm{L}} = 9.441\ \Omega$ and that $Z_0 = 20\ \Omega$ is chosen. For the series resonant effective load, the load at the end of the transmission line can be found:

$$R'_{\mathrm{L}} = \frac{Z_0^2}{R_{\mathrm{L}}} = 42.37\ \Omega$$
$$Q = 50 = \frac{\omega_0 L'}{R'_{\mathrm{L}}}$$

or

$$L' = 374.6\ \mathrm{nH}$$

and

$$C' = \frac{1}{\omega_0^2 L'} = 83.47 \text{ fF}$$

$$L = Z_0^2 C' = 33.39 \text{ pH}$$

$$C = \frac{L'}{Z_0^2} = 0.9366 \text{ nH}$$

Even with all the assumptions regarding the transistor and lossless, dispersionless elements, the results are still not pretty. The transistor is biased to provide 0.8 volts at the base (Fig. 9.16). When the ac input voltage amplitude at the base of the transistor is 0.11 volts, the resulting collector current is shown in Fig. 9.17. This is hardly a half sine wave as one might expect from an over simplified analysis. The graph in Fig. 9.18 shows at least the rudiments of a square wave on the collector. The places where the voltage exceeds $2V_{CC}$ is a result of constructive interference of various traveling waves. Nevertheless, an average output power on the load, $R_L$, of approximately 5.5 W is achieved as seen from the instantaneous output power in Fig. 9.19. The power-added efficiency for this circuit is found from the SPICE analysis. The dc input power is 5.656 W, and the ac input power is 2.363 mW. The power-added efficiency is then

$$\eta = \frac{P_{out} - P_{in-ac}}{P_{dc}} \tag{9.70}$$

$$= 97.4\%$$

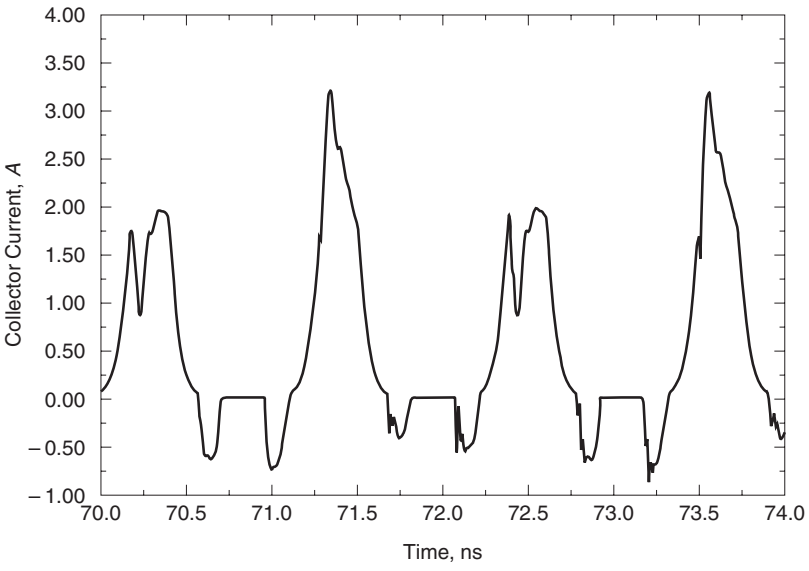



**FIGURE 9.17** Class F collector current when the ac $V_G = 0.11$ volts.

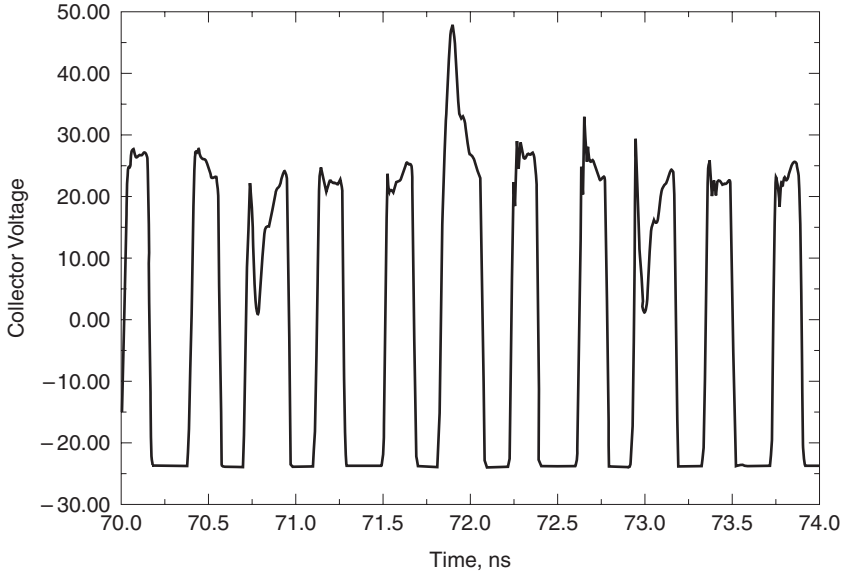



**FIGURE 9.18** Class F collector voltage when the ac $V_G = 0.11$ volts.

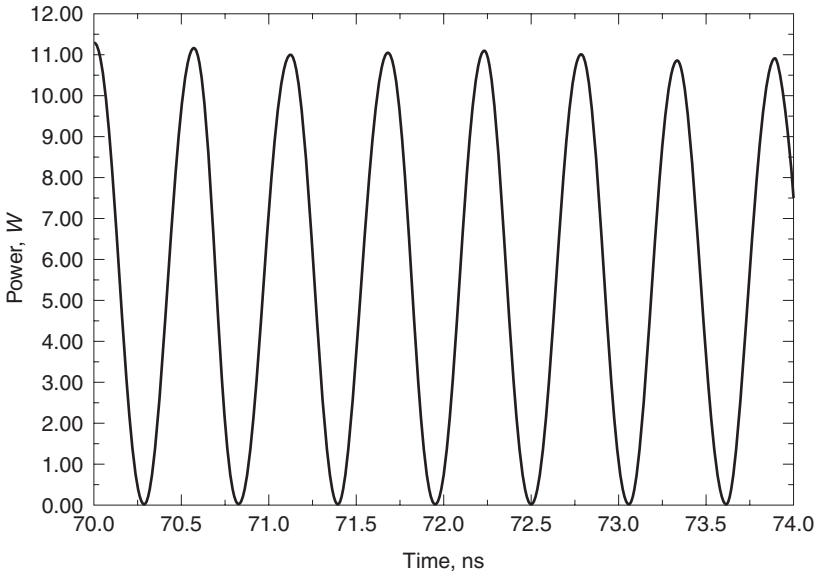



**FIGURE 9.19** Class F collector load power when the ac $V_G = 0.11$ volts.

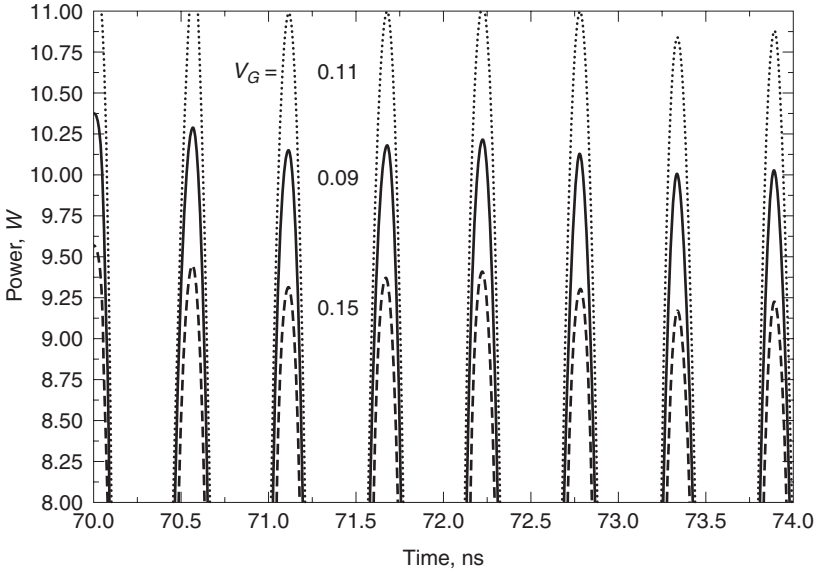



**FIGURE 9.20** Class F collector load power as a function of $V_G$.

The bad news is that the output power is very sensitive to the amplitude of the ac input voltage, $V_G$, as demonstrated in Fig. 9.20. A more extensive harmonic balance analysis of a physics based model for a metal semiconductor field effect transistor (MESFET) showed that a power added efficiency of 75% can be achieved at 5 GHz [6].

## 9.7 FEED-FORWARD AMPLIFIERS

The concept of feed-forward error control was conceived in a patent disclosure by Harold S. Black in 1924 [7]. This was several years prior to his more famous concept of feedback error control. An historical perspective on the feed-forward idea is found in [8]. The feedback approach is an attempt to correct an error after it has occurred. A 180° phase difference in the forward and reverse paths in a feedback system can cause unwanted oscillations. In contrast, the feed-forward design is based on cancellation of amplifier errors in the same time frame in which they occur. Signals are handled by wideband analog circuits, so multiple carriers in a signal can be controlled simultaneously. Feed-forward amplifiers are inherently stable, but this comes at the price of a somewhat more complicated circuit. Consequently feed-forward circuitry is sensitive to changes in ambient temperature, input power level, and supply voltage variation. Nevertheless, feed-forward offers many advantages that have brought it increased interest.

The major source of distortion, such as harmonics, intermodulation distortion, and noise, in a transmitter is the power amplifier. This distortion can be greatly

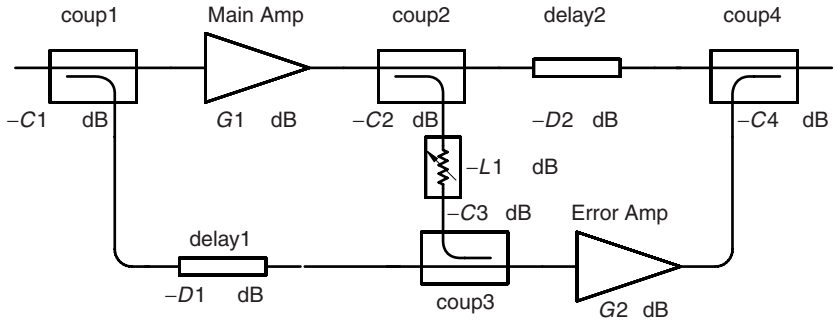



**FIGURE 9.21** Linear feed-forward amplifier.

reduced using feed-forward design. The basic idea is illustrated in Fig. 9.21, where it is seen that the circuit consists basically of two loops. The first one contains the main power amplifier, and the second loop contains the error amplifier. In the first loop, a sample of the input signal is coupled through "coup1" reducing the signal by the coupling factor $-C1$ dB. This goes through the delay line with insertion loss of $-D1$ dB into the comparator coupler "coup3." At the same time the signal passing through the main amplifier with gain $G1$ dB is sampled by coupler "coup2" reducing the signal by $-C2$ dB, the attenuator by $-L1$ dB, and the coupler "coup3" by $-C3$ dB. The delay line is adjusted to compensate for the time delay in the main amplifier as well as the passive components so that two input signals for "coup3" are 180° out of phase but synchronized in time. The amplitude of the input signal when it arrives at the error amplifier is

$$-C1 - D1 - [G1 - C2 - L1 - C3] \tag{9.71}$$

which should be adjusted to be zero. What remains is the distortion and noise added by the main amplifier which is in turn amplified by the error amplifier by $G2$ dB. At the same time the signal from the main amplifier with its distortion and noise is attenuated by $D2$ dB in the second loop delay line. The second delay line is adjusted to compensate for the time delay in the error amplifier. The relative phase and amplitude of the input signals to "coup4" are adjusted so that the distortion terms cancel. The output distortion amplitude

$$-D2 - [-C2 - L1 - C3 + G2 - C4] \tag{9.72}$$

should be zero for complete cancellation to occur.

The error amplifier will also add distortion and noise to its input signal so that perfect error correction will not occur. Nevertheless, a dramatic improvement is possible, since the error amplifier will be operating on a smaller signal (only distortion) that will likely lie in the linear range of the amplifier. Further improvement may be accomplished by treating the entire amplifier in Fig. 9.21 as the main amplifier and adding another error amplifier with its associated circuitry [8].

A typical implementation of a feed-forward system is described in [9] for an amplifier operating in the frequency range of 2.1 to 2.3 GHz with an RF gain of 30 dB, and an output power of 1.25 W. This amplifier had intermodulation products at least 50 dB below the carrier level. Their design used a 6 dB coupler for "coup1," a 13 dB coupler for "coup2," a 10 dB coupler for "coup3," and an 8 dB coupler for "coup4." In some designs the comparator coupler, "coup3," is replaced by a power combiner.

The directional coupler itself can be implemented using microstrip or stripline coupled lines at higher frequencies [10] or by a transmission line transformer. A variety of feed forward designs have been implemented, some using digital techniques [11,12].

## PROBLEMS

**9.1** If the crossover discontinuity is neglected, is a class B amplifier considered a linear amplifier or a nonlinear amplifier? Explain your answer.

**9.2** A class B amplifier such as that shown in Fig. 9.7 is biased with an 18 volt power supply, but the maximum voltage amplitude across each transistor is 16 volts. The remaining 2 volts is dissipated as loss in the output transformer. If the amplifier is designed to deliver 12 W of RF power, find the following:

**(a)** The maximum RF collector current

**(b)** The total dc current from the power supply

**(c)** The collector efficiency of this amplifier.

**9.3** The class C amplifier shown in Fig. 9.9 has a conduction angle of 60°. It is designed to deliver 75 W of RF output power. The saturated collector–emitter voltage is known to be 1 volt, and the power supply voltage is 26 volts. What is the maximum peak collector current.

## REFERENCES

1. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, New York: Wiley, 1993.
2. H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: Wiley, 1980.
3. D. M. Snider, "A Theoretical Analysis and Experimental Confirmation of the Optimally Loaded and Overdriven RF Power Amplifier," *IEEE Trans. on Electron Devices*, Vol. ED-14, pp. 851–857, 1967.
4. F. H. Raab, "Class-F Power Amplifiers with Maximally Flat Waveforms," *IEEE Trans. Microwave Theory Tech.*, Vol. 45, pp. 2007–2012, 1997.
5. C. Trask, "Class-F Amplifier Loading Networks: A Unified Design Approach," *1999 IEEE MTT-S International Symp. Digest*, Piscataway, NJ: IEEE Press, 1999, pp. 351–354.

6. L. C. Hall and R. J. Trew, "Maximum Efficiency Tuning of Microwave Amplifiers," *1991 MTT-S International Symp. Digest*, Piscataway, NJ: IEEE Press, 1991, pp. 123–126.
7. H. S. Black, U.S. Patent 1,686,792, issued October 9, 1929.
8. H. Seidel, H. R. Beurrier, and A. N. Friedman, "Error-Controlled High Power Linear Amplifiers at VHF," *Bell Sys. Tech. J.*, Vol. 47, pp. 651–722, 1968.
9. C. Hsieh and S. Chan, "A Feedforward S-Band MIC Amplifier System," *IEEE J. Solid State Circuits*, Vol. SC-11, pp. 271–278, 1976.
10. W. A. Davis *Microwave Semiconductor Circuit Design*, 1984, Ch. 4.
11. S. J. Grant, J. K. Cavers, and P. A. Goud, "A DSP Controlled Adaptive Feed Forward Amplifier Linearizer," Annual International Conference on Universal Personal Communications, pp. 788–792, September 29–October 2, 1996.
12. G. Zhao, F. M. Channouchi, F. Beauregard, and A. B. Kouki, "Digital Implementations of Adaptive Feedforward Amplifier Linearization Techniques," *IEEE Microwave Theory Tech. Symp. Digest*, June 1996.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER TEN

# Oscillators and Harmonic Generators

## 10.1 OSCILLATOR FUNDAMENTALS

An oscillator is a circuit that converts energy from a power source (usually a dc power source) to ac energy. In order to produce a self-sustaining oscillation, there necessarily must be feedback from the output to the input, sufficient gain to overcome losses in the feedback path, and a resonator. There are number of ways to classify oscillator circuits, one of those being the distinction between one-port and two-port oscillators. The one-port oscillator has a load and resonator with a negative resistance at the same port, while the two-port oscillator is loaded in some way at the two ports. In either case there must be a feedback path, although, in the case of the one-port, this path might be internal to the device itself.

An amplifier with positive feedback is shown in Fig. 10.1. The output voltage of this amplifier is

$$V_o = aV_i + a\beta V_o$$

which gives the closed loop gain

$$A = \frac{V_o}{V_i} = \frac{a}{1 - a\beta} \tag{10.1}$$

The positive feedback allows an increasing output voltage to feedback to the input side until the point is reached where

$$a\beta = 1 \tag{10.2}$$

This is called the *Barkhausen criterion* for oscillation and is often described in terms of its magnitude and phase separately. Hence oscillation can occur

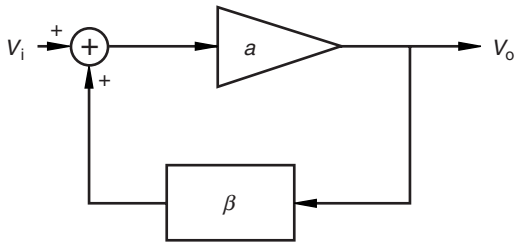



**FIGURE 10.1** Circuit with positive feedback.

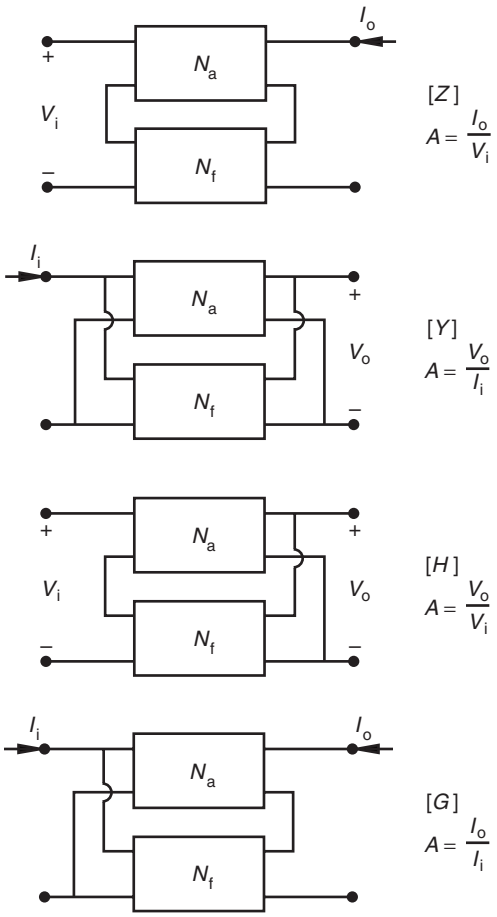



**FIGURE 10.2** Four possible ways to connect the amplifier and feedback circuit. The composite circuit is obtained by adding the designated two-port parameters. The units for "gain" are as shown.

when $|a\beta| = 1$ and $\angle a\beta = 360^\circ$. An alternate way of determining conditions for oscillation is determining when the value $k < 1$ for the stability circle as described in Chapter 7. Still a third way will be considered later in Section 10.4.

## 10.2 FEEDBACK THEORY

The active amplifier part and the passive feedback part of the oscillator can be considered as a pair of two two-port circuits. Usually the connection of these two-ports occurs in four different ways: series–series, shunt–shunt, series–shunt, and shunt–series (Fig. 10.2). A linear analysis of the combination of these two two-ports begins by determining what type of connection exists between them. If, for example, they are connected in series–series, then the best way to describe each of the two-ports is in terms of their $z$ parameters. A description of the composite of the two two-ports is found by simply adding the $z$ parameters of the two circuits together. Thus, if $[z_a]$ and $[z_f]$ represent the amplifier and feedback circuits connected in series–series, then the composite circuit is described by $[z_c] = [z_a] + [z_f]$. The form of the feedback circuit itself can take a wide range of forms, but being a linear circuit, it can always be reduced to a set of $z$, $y$, $h$ or $g$ parameters, any one of which can be represented by $k$ for the present. The term that feeds back to the input of the amplifier is $k_{12f}$. The $k_{12f}$ term, though small, is a significant part of the small incoming signal, so it cannot be neglected. The open loop gain, $a$, of the composite circuit is found by setting $k_{12f} = 0$. Then, using the normal circuit analysis, the open loop gain is determined. The closed loop gain is found by including $k_{12f}$ in the closed loop gain given by Eq. (10.1). The Barkhausen criterion for oscillation is satisfied when $ak_{12f} = a\beta = 1$.

## 10.3 TWO-PORT OSCILLATORS WITH EXTERNAL FEEDBACK

There are a wide variety of two-port oscillator circuits that can be designed. The variety of oscillators results from the different ways the feedback circuit is connected to the amplifier and the variety of feedback circuits themselves. Five of these shown in Fig. 10.3 are known as the Colpitts, Hartley, Clapp-Gouriet [1,2], Armstrong, and Vackar [2,3] oscillators. The Pierce oscillator is obtained by replacing the inductor in the Colpitts circuit with a crystal that acts like a high $Q$ inductor. As shown the first four of these feedback circuits are drawn in a series–series connection, while the Vackar is drawn as a series–shunt configuration. Of course a wide variety of connections and feedback circuits are possible. In each of these oscillators, there is a relatively large amount of energy stored in the resonant reactive circuit. If not too much is dissipated in the load, sustained oscillations are possible.

The Colpitts is generally favored over that of the Hartley, because the Colpitts circuit capacitors usually have higher $Q$ than inductors at RF frequencies and come in a wider selection of types and sizes. In addition the inductances in the Hartley circuit can provide a means to generate spurious frequencies because it

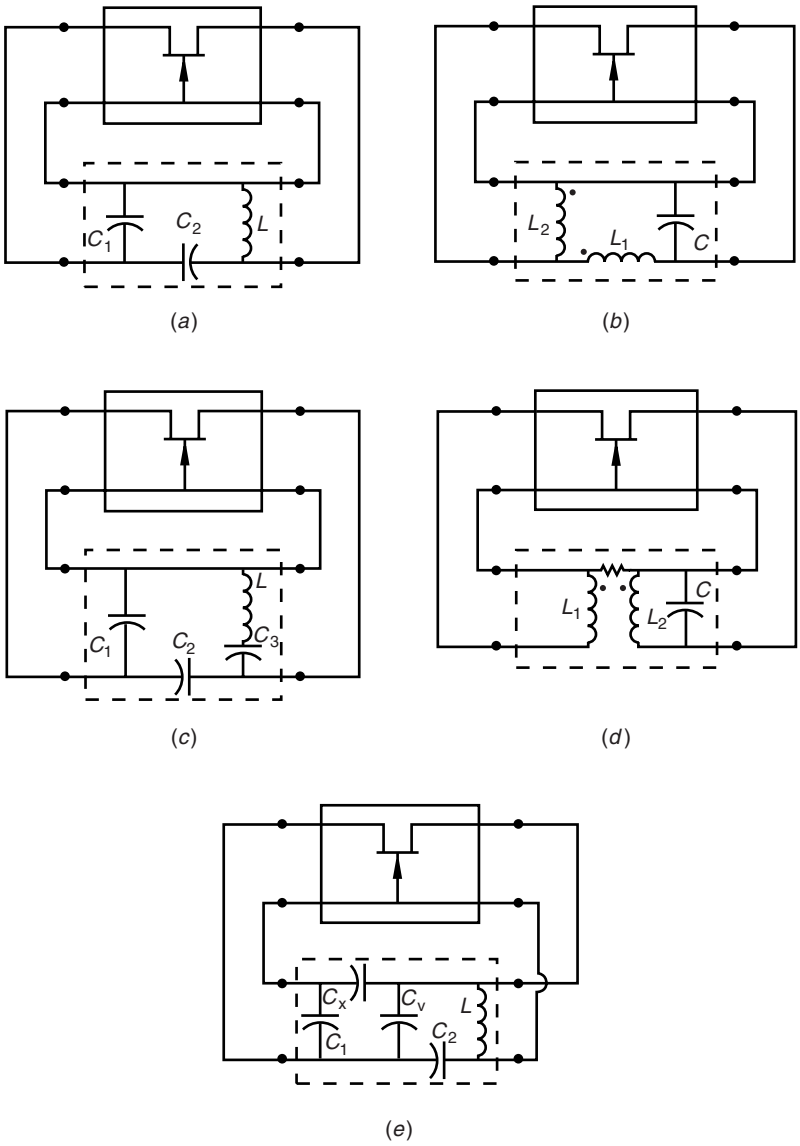



**FIGURE 10.3** Oscillator types: (*a*) Colpitts, (*b*) Hartley, (*c*) Clapp-Gouriet, (*d*) Armstrong, and (*e*) Vackar.

is possible to resonate the inductors with parasitic device capacitances. Because the first element in the Colpitts circuit is a shunt capacitor, it can be said to be a low-pass circuit. For similar reasons the Hartley is a high-pass circuit and the Clapp-Gouriet is a bandpass circuit. There is an improvement in the frequency stability of the tapped capacitor circuit over that of a single LC tuned circuit [1].

In a voltage-controlled oscillator application, it is often convenient to vary the capacitance to change the frequency. This can be done using a reverse biased varactor diode as the capacitor. If the capacitance shown in Fig. 10.4*a* changes because of say a temperature shift, the frequency will change by

$$\frac{df}{f} = -\frac{dC_0}{2C_0} \tag{10.3}$$

However, the tapped circuit in Fig. 10.4*b* in which $C_2$ is used for tuning a Colpitts circuit has a frequency stability given by

$$\frac{df}{f} = -\frac{C_0}{C_2}\frac{dC_2}{C_2} \tag{10.4}$$

This has an improved stability by the factor of $C_0/C_2$. Furthermore, by increasing $C_0$ so that $C_1$ and $C_2$ are increased by even more while adjusting the inductance to maintain the same resonant frequency, the stability can be further enhanced. The Clapp-Gouriet circuit exhibits even better stability than the Colpitts [2]. In this circuit, $C_1$ and $C_2$ are chosen to have large values compared to the tuning capacitor $C_3$. The minimum transistor transconductance, $g_m$, required for oscillation for the Clapp-Gouriet circuit increases $\propto \omega^3/Q$. While the $Q$ of a circuit often rises with frequency, it would not be sufficient to overcome the cubic change in frequency. For the Vackar circuit, the required minimum $g_m$ to maintain oscillation is $\propto \omega/Q$. This would tend to provide a slow drop in the amplitude of the oscillations as the frequency rises [2].

The oscillator is clearly a nonlinear circuit, but nonlinear circuits are difficult to treat analytically. In the interest of trying to get a design solution, linear analysis is used. It can be said that a circuit can be treated by small signal linear mathematics to just prior to it's breaking into oscillation. In going through the transition between oscillation and linear gain, the active part of the circuit does not change appreciably. As a justification for using linear analysis, the previous statement certainly has some flaws. Nevertheless, linear analysis does give remarkably close answers. More advanced computer modeling using methods like harmonic balance will give more accurate results and provide predictions of output power.

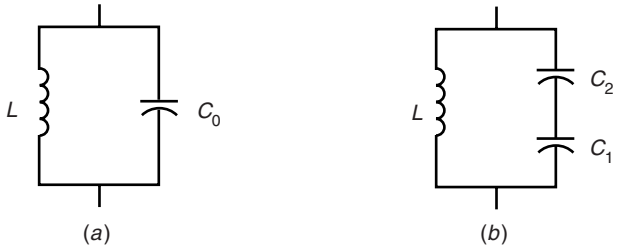



**FIGURE 10.4** (*a*) Simple *LC* resonant circuit and (*b*) tapped capacitor *LC* circuit used in the Colpitts oscillator.

As an example, consider the Colpitts oscillator in Fig. 10.5. Rather than drawing it as shown in Fig. 10.3*a* as a series–series connection, it can be drawn in a shunt–shunt connection by simply rotating the feedback circuit 180° about its *x*-axis. The *y* parameters for the feedback part are

$$y_{11f} = sC_1 + \frac{1}{sL} \tag{10.5}$$

$$y_{22f} = sC_2 + \frac{1}{sL} \tag{10.6}$$

$$y_{12f} = \frac{-1}{sL} \tag{10.7}$$

The equivalent circuit for the *y* parameters now may be combined with the equivalent circuit for the active device (Fig. 10.6). The open loop gain, *a*, is found by setting $y_{12f} = 0$.

$$\frac{v_o}{v_{gs}} = \frac{g_m + y_{12f}}{(1/R_D) + y_{22f}} \tag{10.8}$$

In the usual feedback amplifier theory described in electronics texts, the $y_{21f}$ term would be considered negligible, since the forward gain of the feedback circuit would be very small compared to the amplifier. This cannot be assumed here.

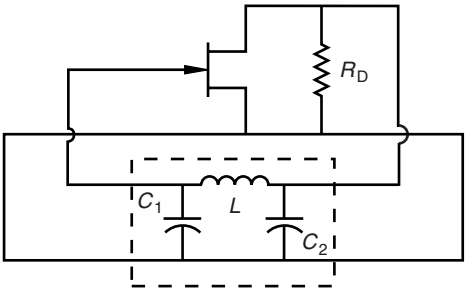



**FIGURE 10.5** Colpitts oscillator as a shunt–shunt connection.

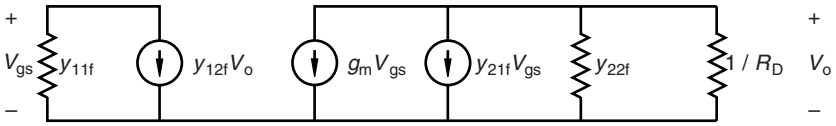



**FIGURE 10.6** Equivalent circuit of the Colpitts oscillator.

The open loop gain, $a$, for the shunt–shunt configuration is

$$a = \frac{v_o}{i_i} = \frac{v_o}{-v_{gs} y_{11f}} = -\frac{g_m + y_{12f}}{y_{11f}[(1/R_D) + y_{22f}]} \tag{10.9}$$

The negative sign introduced in getting $i_i$ is needed to make the current go north rather than south as made necessary by the usual sign convention. Finally, by the Barkhausen criterion, oscillation occurs when $\beta a = 1$:

$$1 = a y_{12f} = \frac{y_{12f}(g_m + y_{12f})}{y_{11f}[(1/R_D) + y_{22f}]} \tag{10.10}$$

Making the appropriate substitutions from Eqs. (10.5) through (10.7) results in the following:

$$-\left(g_m - \frac{1}{sL}\right) = sL\left[\left(sC_1 + \frac{1}{sL}\right)\frac{1}{R_D} + \left(sC_1 + \frac{1}{sL}\right)\left(sC_2 + \frac{1}{sL}\right)\right] \tag{10.11}$$

Both the real and imaginary parts of this equation must be equal on both sides. Since $s = j\omega_0$ at the oscillation frequency, all even powers of $s$ are real and all odd powers of $s$ are imaginary. Since $g_m$ in Eq. (10.11) is associated with the real part of the equation, the imaginary part should be considered first:

$$\frac{1}{sL} = sL\left(s^2 C_1 C_2 + \frac{C_1}{L} + \frac{C_2}{L} + \frac{1}{s^2 L^2}\right) \tag{10.12}$$

When this is solved, the oscillation frequency is found to be

$$\omega_0 = \sqrt{\frac{C_1 + C_2}{LC_1C_2}} \tag{10.13}$$

Solving the real part of Eq. (10.11) with the now known value for $\omega_0$ gives the required value for $g_m$:

$$g_m = \frac{C_1}{R_D C_2} \tag{10.14}$$

The value for $g_m$ found in Eq. (10.14) is the minimum transconductance the transistor must have in order to produce oscillations. The small signal analysis is sufficient to determine conditions for oscillation assuming the frequency of oscillation does not change with current amplitude in the active device. The large signal nonlinear analysis would be required to determine the precise frequency of oscillation, the output power, the harmonic content of the oscillation, and the conditions for minimum noise.

An alternative way of looking at his example involves simply writing down the node voltage circuit equations and solving them. The determinate for the two nodal equations is zero, since there is no input signal:

$$\Delta = \begin{vmatrix} sC_1 + \dfrac{1}{sL} & \dfrac{-1}{sL} \\ \dfrac{-1}{sL} + g_m & sC_2 + \dfrac{1}{sL} + \dfrac{1}{R_D} \end{vmatrix} \tag{10.15}$$

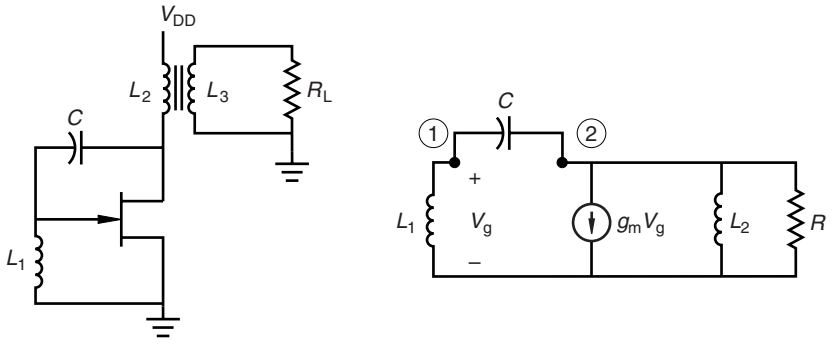
This gives the same equation as Eq. (10.11) and of course the same solution. Solving nodal equations can become complicated when there are several amplifying stages involved or when the feedback circuit is complicated. The method shown here based on the theory developed for feedback amplifiers can be used in a wide variety of circuits.

## 10.4 PRACTICAL OSCILLATOR EXAMPLE

The Hartley oscillator shown in Fig. 10.7 is one of several possible versions for this circuit [4]. In this circuit the actual load resistance is $R_L = 50\ \Omega$. Directly loading the transistor with this size resistance would cause the circuit to cease to oscillate. Hence the transformer is used to provide an effective load to the transistor of

$$R = R_L \left(\frac{n_2}{n_3}\right)^2 \tag{10.16}$$

and at the same time $L_2$ acts as one of the tapped inductors. By solving the network in Fig. 10.7*b* in the same way describe for the Colpitts oscillator, the




(*a*) (b)

**FIGURE 10.7** (*a*) Practical Hartley oscillator and (*b*) equivalent circuit.

frequency of oscillation and minimum transconductance can be found:

$$\omega_0 = \frac{1}{\sqrt{C(L_1 + L_2)}} \tag{10.17}$$

$$g_m = \frac{L_2}{L_1 R} \tag{10.18}$$

For a 10 MHz oscillator biased with $V_{DD} = 15$ V, the inductances, $L_1$ and $L_2$ are chosen to be both equal to 1 μH. The capacitance from Eq. (10.17) is 126.6 pF. If the minimum device transconductance for a 2N3819 JFET is 3.5 mS, then from Eq. (10.18), $R > 285\ \Omega$. Choosing the resistance $R$ to be 300 Ω will require the transformer turns ratio to be

$$\frac{n_2}{n_3} = \sqrt{\frac{R}{R_L}} = 2.45$$

and

$$L_3 = L_2 \left(\frac{n_3}{n_2}\right)^2 = 1 \cdot \left(\frac{1}{2.45}\right)^2 = 0.1667\ \mu\text{H}$$

These circuit values can be put into SPICE to check for the oscillation. However, SPICE will give zero output when there is zero input. Somehow a transient must be used to start the circuit oscillating. If the circuit is designed correctly, oscillations will be self-sustaining after the initial transient. One way to initiate a start up transient is to prevent SPICE from setting up the dc bias voltages prior to doing a time domain analysis. This is done by using the SKIPBP (skip bias point) or UIC (use initial conditions) command in the transient statement. In addition it may be helpful to impose an initial voltage condition on a capacitance or initial current condition on an inductance. A second approach is to use the PWL (piecewise linear) transient voltage somewhere in the circuit to impose a short pulse at $t = 0$ which forever after is turned off. The first approach is illustrated in the SPICE net list for the Hartley oscillator.

```
Hartley Oscillator Example. 10 MHz, RL = 50.
* This will take some time.
L1      1   16  1u
VDC     16  0   DC   -1.5
C       1   2   126.7p  IC=-15
L2      3   2   1u
L3      4   0   .1667u
K23     L2  L3  .9999  ;Unity coupling not allowed
RL      4   0   50.
J1      2   1   0   J2N3819
.LIB    EVAL.LIB
VDD     3   0   15
```

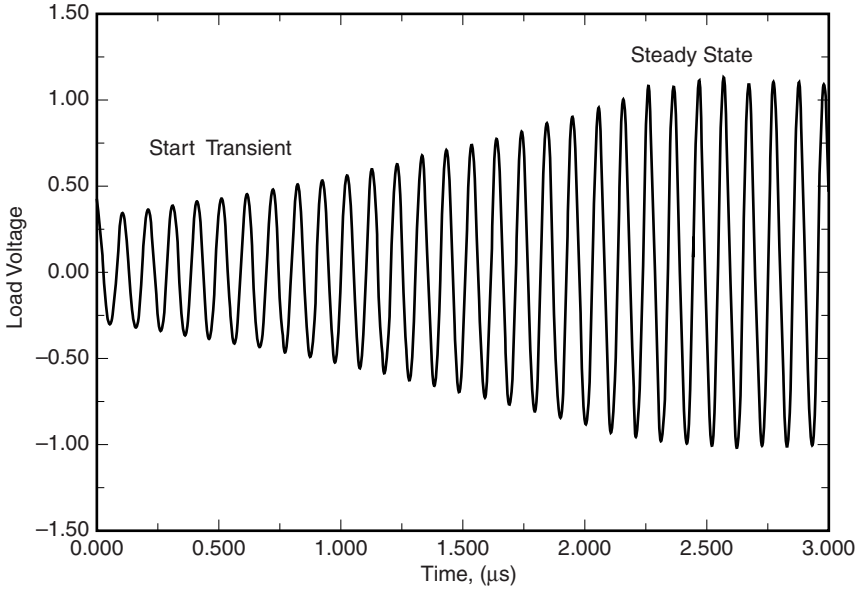



**FIGURE 10.8** 10 MHz Hartley oscillator time domain response.

```
*.TRAN  <print step> <final time> <no print>
>step ceiling> SKIPBP
.TRAN   2n  3uS  0    7nS   SKIPBP
.PROBE
.OP
.END
```

The result of the circuit analysis in Fig. 10.8 shows the oscillation building up to a steady state output after many oscillation periods.

## 10.5 MINIMUM REQUIREMENTS OF THE REFLECTION COEFFICIENT

The two-port oscillator has two basic configurations: (1) a common source FET that uses an external resonator feedback from drain to gate and (2) a common gate FET that produces a negative resistance. In both of these the dc bias and the external circuit determine the oscillation conditions. When a load is connected to an oscillator circuit and the bias voltage is applied, noise in the circuit or start up transients excites the resonator at a variety of frequencies. However, only the resonant frequency is supported and sent back to the device negative resistance. This in turn is amplified and so the oscillation begins building up.

Negative resistance is merely a way of describing a power source. Ohm's law says that the resistance of a circuit is the ratio of the voltage applied to the current

flowing out of the positive terminal of the voltage source. If the current flows back into the positive terminal of the voltage source, then of course it is attached to a negative resistance. The reflection coefficient of a load, $Z_L$, attached to a lossless transmission line with characteristic impedance, $Z_0$, is

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{10.19}$$

Just like viewing yourself in the mirror, the wave reflected off a positive resistance load would be smaller than the incident wave. It is not expected that an image in the mirror would be brighter than the incident light. However, if the $\Re\{Z_L\} < 0$, then it would be possible for $\Gamma$ in Eq. (10.19) to be greater than 1. The "mirror" is indeed capable of reflecting a brighter light than was incident on it. This method is sometimes used to provide amplifier gain, but it can also produce oscillations when the denominator of Eq. (10.19) approaches 0.

The conditions for oscillation then for the two-port in Fig. 10.9 are

$$k < 1 \tag{10.20}$$

and

$$Z_G = -Z_i \tag{10.21}$$

where $k$ is the amplifier stability factor and $Z_i$ is the input impedance of the two-port when it is terminated by $Z_L$. The expression for oscillation in terms of reflection coefficients is easily found by first determining the expressions for $\Gamma_i$ and $\Gamma_G$:

$$\Gamma_i = \frac{R_i - Z_0 + jX_i}{R_i + Z_0 + jX_i} \tag{10.22}$$

$$\Gamma_G = \frac{R_G - Z_0 + jX_G}{R_G + Z_0 + jX_G} \tag{10.23}$$

If $Z_G$ is now replaced by $-Z_i$ in Eq. (10.23),

$$\Gamma_G = \frac{-R_i - Z_0 - jX_i}{-R_i + Z_0 - jX_i} = \frac{1}{\Gamma_i} \tag{10.24}$$

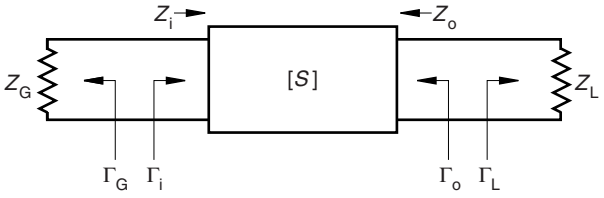



**FIGURE 10.9** Doubly terminated two-port circuit.

Thus Eqs. (10.21) and (10.24) are equivalent conditions for oscillation. In any case, the stability factor, $k$, for the composite circuit with feedback must be less than one to make the circuit unstable and thus capable of oscillation.

An equivalent condition for the load port may be found from Eq. (10.24). From Eq. (7.17) in Chapter 7, the input reflection coefficient for a terminated two-port was found to be

$$\begin{aligned}\Gamma_{\mathrm{i}} &= S_{11} + \frac{S_{12}S_{21}\Gamma_{\mathrm{L}}}{1 - S_{22}\Gamma_{\mathrm{L}}} \\ &= \frac{S_{11} - \Delta\Gamma_{\mathrm{L}}}{1 - S_{22}\Gamma_{\mathrm{L}}} = \frac{1}{\Gamma_{\mathrm{G}}}\end{aligned} \tag{10.25}$$

where $\Delta$ is the determinate of the $S$-parameter matrix. Solving the right-hand side of Eq. (10.25) for $\Gamma_{\mathrm{L}}$ gives

$$\Gamma_{\mathrm{L}} = \frac{1 - S_{11}\Gamma_{\mathrm{G}}}{S_{12} - \Delta\Gamma_{\mathrm{G}}} \tag{10.26}$$

But from Eq. (7.21),

$$\begin{aligned}\Gamma_{\mathrm{o}} &= S_{22} + \frac{S_{12}S_{21}\Gamma_{\mathrm{G}}}{1 - S_{11}\Gamma_{\mathrm{G}}} \\ &= \frac{S_{22} - \Delta\Gamma_{\mathrm{G}}}{1 - S_{11}\Gamma_{\mathrm{G}}} = \frac{1}{\Gamma_{\mathrm{L}}}\end{aligned} \tag{10.27}$$

The last equality results from Eq. (10.26). The implication is that if the conditions for oscillation exists at one port, they also necessarily exist at the other port.

## 10.6 COMMON GATE (BASE) OSCILLATORS

A common gate configuration is often advantageous for oscillators because they have a large intrinsic reverse gain ($S_{12\mathrm{g}}$) that provides the necessary feedback. Furthermore feedback can be enhanced by putting some inductance between the gate and ground. Common gate oscillators often have low spectral purity but wide band tunability. Consequently they are often preferred in voltage-controlled oscillator (VCO) designs. For a small signal approximate calculation, the scattering parameters of the transistor are typically found from measurements at a variety of bias current levels. Probably the $S$ parameters associated with the largest output power as an amplifier would be those to be chosen for oscillator design. Since common source $S$ parameters, $S_{ij}$, are usually given, it is necessary to convert them to common gate $S$ parameters, $S_{ij\mathrm{g}}$. Once this is done, the revised $S$ parameters may be used in a direct fashion to check for conditions of oscillation.

The objective at this point is to determine the common gate $S$ parameters with the possibility of having added gate inductance. These are derived from the common source $S$ parameters. The procedure follows:

1. Convert the two-port common source $S$ parameters to two-port common source $y$ parameters.
2. Convert the two-port $y$ parameters to three-port indefinite $y$ parameters.
3. Convert the three-port $y$ parameters to three-port $S$ parameters.
4. One of the three-port terminals is terminated with a load of known reflection coefficient, $r$.
5. With one port terminated, the $S$ parameters are converted to two-port $S$ parameters, which could be, among other things, common gate $S$ parameters.

At first, one might be tempted to convert the $3 \times 3$ indefinite admittance matrix to a common gate admittance matrix and convert that to $S$ parameters. The problem is that "common gate" usually means shorting the gate to ground, which is fine for $y$ parameters, but it is not the same as terminating the gate with a matched load or other impedance, $Z_g$, for the $S$ parameters.

The first step, converting the $S$ parameters to $y$ parameters, can be done using the formulas in Table D.1 or Eq. (D.10) in Appendix D. For example, if the common source $S$ parameters, $[S_s]$, are given the $y$ parameter matrix is

$$[Y_s] = Y_0 \begin{bmatrix} \dfrac{(1 - S_{11s})(1 + S_{22s}) + S_{12s}S_{21s}}{D_s} & \dfrac{-2S_{12s}}{D_s} \\ \dfrac{-2S_{21s}}{D_s} & \dfrac{(1 + S_{11s})(1 - S_{22s}) + S_{12s}S_{21s}}{D_s} \end{bmatrix} \tag{10.28}$$

where

$$D_s \triangleq (1 + S_{11s})(1 + S_{22s}) - S_{12s}S_{21s} \tag{10.29}$$

Next the $y$ parameters are converted to the $3 \times 3$ indefinite admittance matrix. The term "indefinite" implies that there is no assumed reference terminal for the circuit described by this matrix [5]. The indefinite matrix is easily found from its property that the sum of the rows of the matrix is zero, and the sum of the columns is zero. Purely for convenience, this third row and column will be added to the center of the matrix. Then the $y_{11}$ will represent the gate admittance, the $y_{22}$ the source admittance, and the $y_{33}$ the drain admittance. The new elements for the indefinite matrix are then put in between the first and second rows and in between the first and second columns of Eq. (10.28). For example, the new $y_{12}$ is

$$y_{12} = Y_0 \left[ \frac{2S_{12s} - (1 - S_{11s})(1 + S_{22s}) + S_{12s}S_{21s} - S_{12s}S_{21s}}{D_s} \right] \tag{10.30}$$

The values for $y_{21}$, $y_{23}$, and $y_{32}$ are found similarly. The new $y_{22}$ term is found from $y_{22} = -y_{21} - y_{23}$. The indefinite admittance matrix is then represented as follows:

$$[Y] = \begin{matrix} \\ g \\ s \\ d \end{matrix} \begin{matrix} \begin{matrix} g & s & d \end{matrix} \\ \begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{pmatrix} \end{matrix} \tag{10.31}$$

The $S$ parameter matrix for $3 \times 3$ or higher order can be found from Eq. (D.9) in Appendix D:

$$S = F(I - G^*Y)(I + GY)^{-1}F^{-1} \tag{10.32}$$

In this equation $I$ is the identity matrix, while $G$ and $F$ are defined in Appendix D. When the measurement characteristic impedances, $Z_0$, are the same in all three ports, the $F$ and the $F^{-1}$ will cancel out. Determining $S$ from Eq. (10.32) is straightforward but lengthy. At this point the common terminal is chosen. To illustrate the process, a common source connection is used in which the source is terminated by a load with reflection coefficient, $r_s$, as shown in Fig. 10.10. If the source is grounded, the reflection coefficient is $r_s = -1$. The relationship between the incident and reflected waves is

$$\begin{aligned} b_1 &= S_{11}a_1 + S_{12}a_2 + S_{13}a_3 \\ b_2 &= S_{21}a_1 + S_{22}a_2 + S_{23}a_3 \\ b_3 &= S_{31}a_1 + S_{32}a_2 + S_{33}a_3 \end{aligned} \tag{10.33}$$

Solution for $S_{11s}$ is done by terminating the drain at port 3 with $Z_0$ so that $a_3 = 0$. The source is terminated with an impedance with reflection coefficient

$$r_s = \frac{a_2}{b_2} \tag{10.34}$$

or for any port

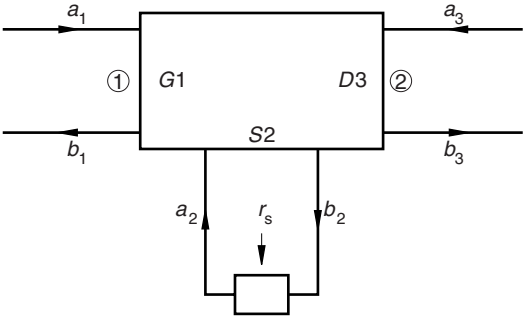



**FIGURE 10.10** Three-port with source terminated with $r_s$.

$$r_{\mathrm{i}} = \frac{Z_{\mathrm{i}} - Z_{\mathrm{ref}}}{Z_{\mathrm{i}} + Z_{\mathrm{ref}}} \tag{10.35}$$

The reflection coefficient is determined relative to the reference impedance which is the impedance looking back into to the transistor. With Eq. (10.34), $b_2$ can be eliminated in Eq. (10.33) giving a relationship between $a_1$ and $a_2$:

$$\frac{a_2}{r_{\mathrm{s}}} = S_{21}a_1 + S_{22}a_2$$

$$a_2 = \frac{S_{21}a_1}{1/r_{\mathrm{s}} - S_{22}} \tag{10.36}$$

The ratio between $b_1$ and $a_1$ under these conditions is

$$S_{11\mathrm{s}} = \frac{b_1}{a_1} = S_{11} + \frac{S_{12}S_{21}}{1/r_{\mathrm{s}} - S_{22}} \tag{10.37}$$

This represents the revised $S_{11\mathrm{s}}$ scattering parameter when the source is terminated with an impedance whose reflection coefficient is $r_{\mathrm{s}}$. In similar fashion the other parameters can be easily found, as shown in Appendix E. The numbering system for the common source parameters is set up so that the input port (gate side) is port-1 and the output port (drain) is port-2. Therefore the subscripts of the common source parameters, $S_{ij\mathrm{s}}$, range from 1 to 2. In other words, after the source is terminated, there are only two ports, the input and output. These are written in terms of the three-port scattering parameters, $S_{ij}$, which of course have subscripts that range from 1 to 3.

The common gate connection can be calculated using this procedure. The explicit formulas are given in Appendix E. For a particular RF transistor, in which the generator is terminated with a 5 nH inductor, the required load impedance on the drain side to make the circuit oscillate is shown in Fig. 10.11 as obtained from the program SPARC (*S*-parameters conversion). Since a passive resistance must be positive, the circuit is capable of oscillation only for those frequencies in which the resistance is above the 0 Ω line. An actual oscillator would still require a resonator to force the oscillator to provide power at a single frequency. A numerical calculation at 2 GHz that illustrates the process is found in Appendix E.

When the real part of the load impedance is less than the negative of the real part of the device impedance, then oscillations will occur at the frequency where there is resonance between the load and the device. For a one-port oscillator, the negative resistance is a result of feedback, but here the feedback is produced by the device itself rather than by an external path. Specific examples of one-port oscillators use a Gunn or IMPATT diode as the active device. These are normally used at frequencies above the band of interest here. On the surface the one-port oscillator is in principle no different than a two-port oscillator whose opposite side is terminated in something that will produce negative resistance at the other end. The negative resistance in the device compensates for positive resistance in the

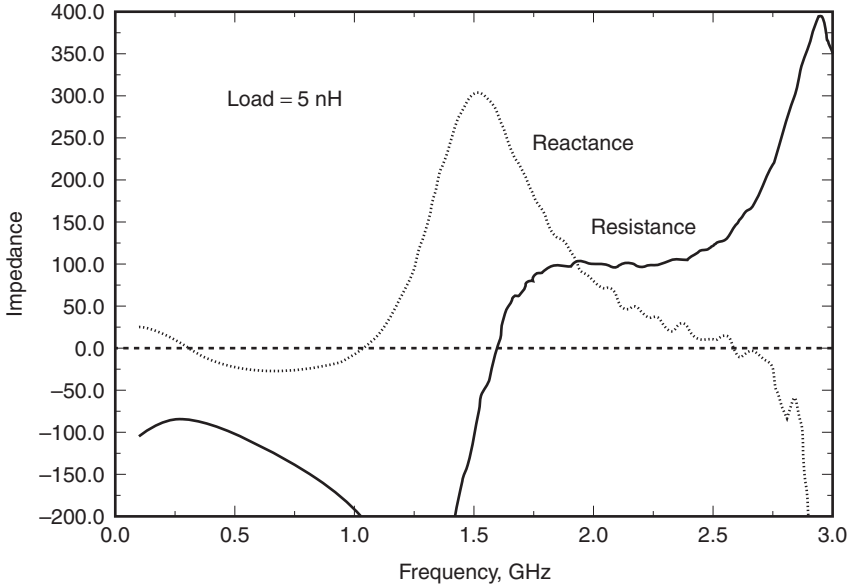



**FIGURE 10.11** Plot of load impedance required for oscillation when generator side is terminated with a 5 nH inductor.

resonator. Noise in the resonator port or a turn on transient starts the oscillation going. The oscillation frequency is determined by the resonant frequency of a high-$Q$ circuit.

## 10.7 STABILITY OF AN OSCILLATOR

In the previous section, a method has been given to determine whether a circuit will oscillate or not. What is yet to be addressed is whether the oscillation will remain stable in the face of a small current transient in the active device. The simple equivalent circuit shown in Fig. 10.12 can be divided into the part with the active device, and the passive part with the high-$Q$ resonator. The current flowing through the circuit is

$$i(t) = A(t)\cos(\omega t + \phi(t)) = \Re\{A(t)e^{j\omega t+\phi(t)}\} \tag{10.38}$$

where $A$ and $\phi$ are slowly varying functions of time. The part of the circuit with the active device is represented by $Z_{\mathrm{d}}(A, \omega)$ and the passive part by $Z(\omega)$. The condition for oscillation requires that the sum of the impedances around the loop to be zero:

$$Z_{\mathrm{d}}(A, \omega) + Z(\omega) = 0 \tag{10.39}$$

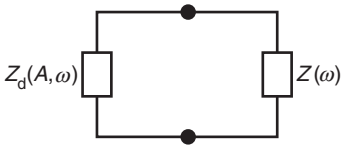



**FIGURE 10.12** Oscillator model when the passive impedance $Z(\omega)$ is separated from the active device $Z_d(A, \omega)$.

Ordinarily the passive circuit selects the frequency of oscillation by means of a high-$Q$ resonator. The relative frequency dependence of the active device is small, so Eq. (10.39) can be approximated by

$$Z_d(A) + Z(\omega) = 0 \tag{10.40}$$

In phaser notation the current is

$$I = Ae^{j\phi} \tag{10.41}$$

and

$$Z(\omega) = R(\omega) + jX(\omega) \tag{10.42}$$

so that the voltage drop around the closed loop in Fig. 10.12 is

$$\begin{aligned} 0 &= \Re\{[Z(\omega) + Z_d(A)]I\} \\ &= [R(\omega) + R_d(A)]A\cos(\omega t + \phi) - [X(\omega) + X_d(A)]A\sin(\omega t + \phi) \end{aligned} \tag{10.43}$$

The time rate of change of the current is found by taking the derivative of Eq. (10.38):

$$\begin{aligned} \frac{di}{dt} &= -A\left(\omega + \frac{d\phi}{dt}\right)\sin(\omega t + \phi) + \frac{dA}{dt}\cos(\omega t + \phi) \\ &= \Re\left\{\left[j\left(\omega + \frac{d\phi}{dt}\right) + \frac{1}{A}\frac{dA}{dt}\right]Ae^{j\omega t + \phi}\right\} \end{aligned} \tag{10.44}$$

Ordinarily, in ac circuit analysis, $d/dt$ is equivalent to $j\omega$ in the frequency domain. Now, with variation in the amplitude and phase, the time derivative is equivalent to

$$\frac{d}{dt} \rightarrow j\omega' = j\left[\omega + \frac{d\phi}{dt} - j\frac{1}{A}\frac{dA}{dt}\right] \tag{10.45}$$

The Taylor series expansion of $Z(\omega')$ about $\omega_0$ is

$$Z\left(\omega + \frac{d\phi}{dt} - j\frac{1}{A}\frac{dA}{dt}\right) \approx Z(\omega_0) + \frac{dZ}{d\omega}\left(\frac{d\phi}{dt} - j\frac{1}{A}\frac{dA}{dt}\right) \tag{10.46}$$

Consequently an expression for the voltage around the closed loop can be found:

$$\Re\{(Z+Z_{\mathrm{d}})I\} = \left[R(\omega_0)+R_{\mathrm{d}}(A)+\frac{dR}{d\omega}\frac{d\phi}{dt}+\frac{dX}{d\omega}\frac{1}{A}\frac{dA}{dt}\right]A\cos(\omega t+\phi) - \left[X(\omega_0)+X_{\mathrm{d}}(A)+\frac{dX}{d\omega}\frac{d\phi}{dt}-\frac{dR}{d\omega}\frac{1}{A}\frac{dA}{dt}\right]A\sin(\omega t+\phi) \tag{10.47}$$

Multiplying Eq. (10.47) by $\cos(\omega t+\phi)$ and then by $\sin(\omega t+\phi)$ and finally integrating will produce, by the orthogonality property, the following two equations:

$$0 = R(\omega)+R_{\mathrm{d}}(A)+\frac{dR}{d\omega}\frac{d\phi}{dt}+\frac{dX}{d\omega}\frac{1}{A}\frac{dA}{dt} \tag{10.48}$$

$$0 = -X(\omega)-X_{\mathrm{d}}(A)-\frac{dX}{d\omega}\frac{d\phi}{dt}+\frac{dR}{d\omega}\frac{1}{A}\frac{dA}{dt} \tag{10.49}$$

Multiplying Eq. (10.48) by $dX/d\omega$ and Eq. (10.49) by $dR/d\omega$ and adding will eliminate the $d\phi/dt$ term. A similar procedure will eliminate $dA/dt$. The result is

$$0 = [R(\omega)+R_{\mathrm{d}}(A)]\frac{dX}{d\omega}-[X(\omega)+X_{\mathrm{d}}(A)]+\left|\frac{dZ(\omega)}{d\omega}\right|^2\frac{1}{A}\frac{dA}{dt} \tag{10.50}$$

$$0 = [X(\omega)+X_{\mathrm{d}}(A)]\frac{dX}{d\omega}+[R(\omega)+R_{\mathrm{d}}(A)]+\left|\frac{dZ(\omega)}{d\omega}\right|^2\frac{d\phi}{dt} \tag{10.51}$$

Under steady state conditions the time derivatives are zero. The combination of Eqs. (10.50) and (10.51) gives

$$\frac{dR/d\omega}{dX/d\omega} = \frac{R(\omega)+R_{\mathrm{d}}(A)}{X(\omega)+X_{\mathrm{d}}(A)} = -\frac{X(\omega)+X_{\mathrm{d}}(A)}{R(\omega)+R_{\mathrm{d}}(A)} \tag{10.52}$$

The only way for this equation to be satisfied results in Eq. (10.40). However, suppose that there is a small disturbance in the current amplitude of $\delta A$ from the steady state value of $A_0$. Based on Eq. (10.40) the resistive and reactive components would become

$$R(\omega_0)+R_{\mathrm{d}}(A) = R(\omega_0)+R_{\mathrm{d}}(A_0)+\delta A\frac{dR_{\mathrm{d}}(A)}{dA} = \delta A\frac{dR_{\mathrm{d}}(A)}{dA} \tag{10.53}$$

$$X(\omega_0)+X_{\mathrm{d}}(A) = \delta A\frac{dX_{\mathrm{d}}(A)}{dA} \tag{10.54}$$

The derivatives are of course assumed to be evaluated at $A = A_0$. Substituting these into Eq. (10.50) gives the following differential equation with respect to time:

$$0 = \delta A \frac{dR_d(A)}{dA}\frac{dX(\omega)}{d\omega} - \delta A \frac{dX_d(A)}{dA}\frac{dR(\omega)}{d\omega} + \left|\frac{dZ(\omega)}{d\omega}\right|^2 \frac{1}{A_0}\frac{d\delta A}{dt} \quad (10.55)$$

or

$$0 = \delta A S + \alpha \frac{d\delta A}{dt} \quad (10.56)$$

where

$$S \triangleq \frac{\partial R_d(A)}{\partial A}\frac{dX(\omega)}{d\omega} - \frac{\partial X_d(A)}{\partial A}\frac{dR(\omega)}{d\omega} > 0 \quad (10.57)$$

and

$$\alpha \triangleq \left|\frac{dZ(\omega)}{d\omega}\right|^2 \frac{1}{A_0} \quad (10.58)$$

The solution of Eq. (10.56) is

$$\delta A = Ce^{-St/\alpha}$$

which is stable if $S > 0$. The Kurokawa stability condition for small changes in the current amplitude is therefore given by Eq. (10.57) [6]. As an example, consider the stability of a circuit whose passive circuit impedance changes with frequency as shown in Fig. 10.13 and whose device impedance changes with current amplitude as shown in the third quadrant of Fig. 10.13. As the current amplitude increases, $R_d(A)$ and $X_d(A)$ both increase:

$$\frac{\partial R_d(A)}{\partial A} > 0$$
$$\frac{\partial X_d(A)}{\partial A} > 0$$

As frequency increases, the passive circuit resistance, $R(\omega)$, decreases and the circuit reactance, $X(\omega)$, increases:

$$\frac{\partial R(\omega)}{\partial \omega} < 0$$
$$\frac{\partial X(\omega)}{\partial \omega} > 0$$

From Eq. (10.57) this would provide stable oscillations at the point where $Z(\omega)$ and $-Z_d(A)$ intersect. If there is a small change in the current amplitude, the circuit tends to return back to the $A_0$, $\omega_0$ resonant point.

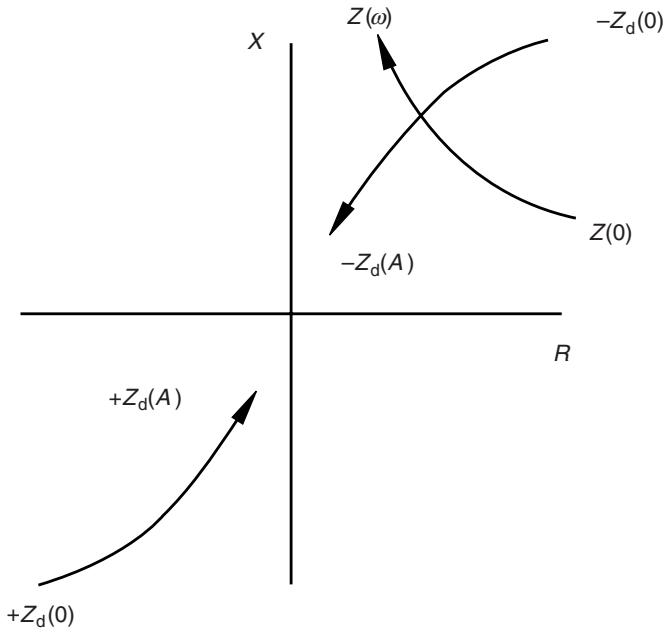



**FIGURE 10.13** Locus of points for the passive and active oscillator impedances.

If there is a small perturbation in the phase rather than the amplitude of the of the current, the stability criterion is

$$S' \triangleq \frac{\partial X_d(\phi)}{\partial \phi}\frac{dX(\omega)}{d\omega} + \frac{\partial X_d(\phi)}{\partial \phi}\frac{dR(\omega)}{d\omega} > 0 \tag{10.59}$$

This is found by substituting into Eq. (10.51) with the appropriate Taylor series approximation for a change in phase.

## 10.8 INJECTION-LOCKED OSCILLATORS

A free running oscillator frequency can be modified by applying an external frequency source to the oscillator. Such injection-locked oscillators can be used as high-power FM amplifiers when the circuit $Q$ is sufficiently low to accommodate the frequency bandwidth of the signal. If the injection signal voltage, $V$, is at a frequency close to but not necessarily identical to the free running frequency of the oscillator, is placed in series with the passive impedance, $Z(\omega)$, in Fig. 10.12, then the loop voltage is

$$[Z(\omega) + Z_d(A)]I = V \tag{10.60}$$

The amplitude of the current at the free running point is $A_0$ and the relative phase between the voltage and current is $\phi$. Hence

$$Z(\omega) = -Z_{\mathrm{d}}(A) + \frac{|V|}{A_0} e^{-j\phi} \tag{10.61}$$

Up to this point the passive impedance has been left rather general. As a specific example, the circuit can be considered to be a high-$Q$ series resonant circuit determined by its inductance and capacitance together with some cavity losses, $R_{\mathrm{C}}$, and a load resistance, $R_{\mathrm{L}}$:

$$Z(\omega) = j\left(\omega L - \frac{1}{\omega C}\right) + R_{\mathrm{C}} + R_{\mathrm{L}} \tag{10.62}$$

Since $\omega$ is close to the circuit resonant frequency $\omega_0$,

$$\begin{aligned} Z(\omega) &= j\frac{L}{\omega}(\omega^2 - \omega_0^2) + R_{\mathrm{C}} + R_{\mathrm{L}} \\ &\approx j2L\Delta\omega_{\mathrm{m}} + R_{\mathrm{C}} + R_{\mathrm{L}} \end{aligned} \tag{10.63}$$

where $\Delta\omega_{\mathrm{m}} = \omega - \omega_{\mathrm{m}}$.

Equation (10.61) represented in Fig. 10.14 is a modification of that shown in Fig. 10.13 for the free running oscillator case. If the magnitude of the injection voltage, $V$, remains constant, then the constant magnitude vector, $|V|/A_0$, which must stay in contact with both the device and circuit impedance lines,

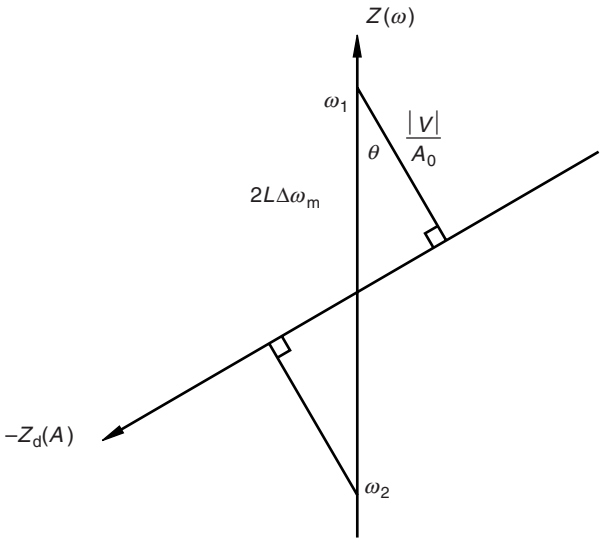



**FIGURE 10.14** Injection-locked frequency range.

will change its orientation as the injection frequency changes (thereby changing $Z(\omega)$). However, there is a limit to how much the $|V|/A_0$ vector can move because circuit and device impedances grow too far apart. In that case the injection lock ceases. The example in Fig. 10.14 is illustrated the simple series-resonant cavity where the circuit resistance is independent of frequency. Furthermore the $|V|/A_0$ vector is drawn at the point of maximum frequency excursion from $\omega_0$. Here $|V|/A_0$ is orthogonal to the $Z_d(A)$ line. If the frequency moves beyond $\omega_1$ or $\omega_2$, the oscillator loses lock with the injected signal. At the maximum locking frequency,

$$|2\Delta\omega_m L\cos\theta| = \frac{|V|}{A_0} \tag{10.64}$$

The expressions for the oscillator power delivered to the load, $P_0$, the available injected power, and the external circuit $Q_{ext}$ are

$$P_0 = \frac{1}{2}R_L A_0^2 \tag{10.65}$$

$$P_i = \frac{|V|^2}{8R_L} \tag{10.66}$$

$$Q_{ext} \approx \frac{\omega_0 L}{R_L} \tag{10.67}$$

When these are substituted into Eq. (10.64), the well-known injection locking range is found [7]:

$$\Delta\omega_m = \frac{\omega_0}{Q_{ext}}\sqrt{\frac{P_i}{P_0}}\frac{1}{\cos\theta} \tag{10.68}$$

The total locking range is from $\omega_0 + \Delta\omega_m$ to $\omega_0 - \Delta\omega_m$. The expression originally given by Adler [8] did not included the $\cos\theta$ term. However, high-frequency devices often exhibit a phase delay of the RF current with respect to the voltage. This led to Eq. (10.68) where the device and circuit impedance lines are not necessarily orthogonal [7]. In the absence of information about the value of $\theta$, a conservative approximation for the injection range can be made by choosing $\cos\theta = 1$. The frequency range over which the oscillator frequency can be pulled from its free-running frequency is proportional to the square root of the injected power and inversely proportional to the circuit $Q$ as might be expected intuitively.

## 10.9 HARMONIC GENERATORS

The nonlinearity of a resistance in a diode can be used in mixers to produce a sum and difference of two input frequencies (see Chapter 11). If a large signal is applied to a diode, the nonlinear resistance can produce harmonics of the input

voltage. However, the efficiency of the nonlinear resistance can be no greater than $1/n$, where $n$ is the order of the harmonic. Nevertheless, a reverse-biased diode has a depletion elastance (reciprocal capacitance) given by

$$\frac{dv}{dq} = S = S_0 \left(1 - \frac{v}{\phi}\right)^{\gamma} \tag{10.69}$$

where $\phi$ is the built-in voltage and typically is between 0.5 and 1 volt positive. The applied voltage $v$ is considered positive when the diode is forward biased. The exponent $\gamma$ for a varactor diode typically ranges from 0 for a step recovery diode to $\frac{1}{3}$ for a graded junction diode to $\frac{1}{2}$ for an abrupt junction diode. Using the nonlinear capacitance of a diode theoretically allows for generation of harmonics with an efficiency of 100% with a loss free diode. This assertion is supported by the Manley-Rowe relations which describe the power balance when two frequencies, $f_1$ and $f_2$, along with their harmonics are present in a lossless circuit:

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{mf_1 + nf_2} = 0 \tag{10.70}$$

$$\sum_{n=0}^{\infty} \sum_{m=-\infty}^{\infty} \frac{nP_{m,n}}{mf_1 + nf_2} = 0 \tag{10.71}$$

These equations are basically an expression of the conservation of energy. From (10.70)

$$P_1 = -\sum_{m=2}^{\infty} P_{m,0}, \qquad n = 0 \tag{10.72}$$

The depletion elastance given by Eq. (10.69) is valid for forward voltages up to about $v/\phi = \frac{1}{2}$. Under forward bias, the diode will tend to exhibit diffusion capacitance that tends to be more lossy in varactor diodes than the depletion capacitance associated with reverse-biased diodes. Notwithstanding these complexities, an analysis of harmonic generators will be based on Eq. (10.69) for all applied voltages up to $v = \phi$. This is a reasonably good approximation when the minority carrier lifetime is long relative to the period of the oscillation. The maximum elastance (minimum capacitance) will occur at the reverse break down voltage, $V_B$. The simplified model for the diode then is defined by two voltage ranges:

$$\frac{S}{S_{\max}} = \left(\frac{\phi - v}{\phi - V_B}\right)^{\gamma}, \qquad v \le \phi \tag{10.73}$$

$$\frac{S}{S_{\max}} = 0, \qquad v > \phi \tag{10.74}$$

Integration of Eq. (10.69) gives

$$-\int_v^{\phi} \frac{\phi \, d(1 - v/\phi)}{(1 - v/\phi)} = S_0 \int_q^{q_\phi} dq \tag{10.75}$$

$$\frac{(\phi - v)^{1-\gamma}}{1 - \gamma} = S_0(q_\phi - q) \tag{10.76}$$

This can be evaluated at the breakdown point where $v = V_B$ and $q = Q_B$. Taking the ratio of this with Eq. (10.76) gives the voltage and charge relative to that at the breakdown point:

$$\frac{\phi - v}{\phi - V_B} = \left(\frac{q_\phi - q}{q_\phi - Q_B}\right)^{1/(1-\gamma)} \tag{10.77}$$

For the abrupt junction diode where $\gamma = \frac{1}{2}$, it can be that it is possible to produce power at $mf_1$ when the input frequency is $f_1$ except for $m = 2$ [9]. Higher-order terms require that the circuit support intermediate frequencies called *idlers*. While the circuit allows energy storage at the idler frequencies, no external currents can flow at these idler frequencies. Thus multiple lossless mixing can produce output power at $mf_1$ with high efficiency when idler circuits are available.

Design of a varactor multiplier consists in predicting the input and output load impedances for maximum efficiency, the value of the efficiency, and the output power. A quantity called the *drive*, $D$, may be defined where $q_{\max}$ represents the maximum stored charge during the forward swing of the applied voltage:

$$D = \frac{q_{\max} - Q_B}{q_\phi - Q_B} \tag{10.78}$$

If $q_{\max} = q_\phi$, then $D = 1$. An important quality factor for a varactor diode is the cutoff frequency. This is related to the series loss, $R_s$, in the diode:

$$f_c = \frac{S_{\max} - S_{\min}}{2\pi R_s} \tag{10.79}$$

When $D \geq 1$, $S_{\min} = 0$. When $f_c/nf_1 > 50$, the tabulated values given in [10]† provide the necessary circuit parameters. These tables have been coded in the program MULTIPLY. The efficiency given by [10] assumes loss only in the diode where $f_{\text{out}} = mf_1$:

$$\eta = e^{\alpha f_{\text{out}}/f_c} \tag{10.80}$$

The output power at $mf_1$ is found to be

$$P_m = \beta \frac{\omega_1(\phi - V_B)^2}{S_{max}} \tag{10.81}$$

The values of $\alpha$ and $\beta$ are given in [9,10]. If the varactor has a dc bias voltage, $V_o$, then the normalized voltage is

$$V_{o,norm} = \frac{\phi - V_o}{\phi - V_B} \tag{10.82}$$

This value corresponds to the selected drive level. Finally, the input and load resistances are found from the tabulated values. The elastances at all supported harmonic frequencies up to and including $m$ are also given. These values are useful for knowing how to reactively terminate the diode at the idler and output frequencies. A packaged diode will have package parasitic circuit elements, as shown in Fig. 10.15, that must be considered in design of a matching circuit. When given these package elements, the program MULTIPLY will find the appropriate matching impedances required external to the package. Following is an example run of MULTIPLY in the design of a 1–2–3–4 varactor quadrupler with an output frequency of 2 GHz. The **bold** numbers are user input values.

```
Input frequency, GHz. =
 0.5
Diode Parameters
Breakdown Voltage =
 60
Built-in Potential phi =
 0.5
Specify series resistance or cutoff frequency,
Rs OR fc. <R/F>
 f
Zero Bias cutoff frequency (GHz), fc =
 50.
Junction capacitance at 0 volts (pF), Co =
 0.5
```

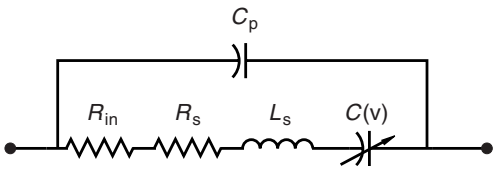



**FIGURE 10.15** Intrinsic varactor diode with package.

```
Package capacitance (pF), Series inductance (nH) =
 0.1, 0.2
For a Doubler Type A
For a 1-2-3 Tripler Type B
For a 1-2-4 Quadrupler Type C
For a 1-2-3-4 Quadrupler Type D
For a 1-2-4-5 Quintupler Type E
For a 1-2-4-6 Sextupler Type F
For a 1-2-4-8 Octupler Type G
For a 1-4 Quadrupler using a SRD, Type H
For a 1-6 Sextupler using a SRD, Type I
For a 1-8 Octupler using a SRD, Type J
Ctrl C to end
 d
Type G for Graded junction (Gamma = .3333)
Type A Abrupt Junction (Gamma=.5)
Choose G or A
 g
Drive is  1.0< D < 1.6.
Linear extrapolation done for D outside this range.
Choose drive.
 2.0
Input Freq =   0.5000 GHz, Output Freq =   2.0000 GHz,
fc =   50.0000 GHz, Rs =  31.4878 Ohms.
Pout =   78.50312 mWatt, Efficiency =   75.47767%
At Drive  2.00, DC Bias Voltage =  -7.76833
Harmonic elastance values
S0( 1) =  0.197844E+13
S0( 2) =  0.313252E+13
S0( 3) =  0.296765E+13
S0( 4) =  0.263791E+13
Total Capacitance with package cap.
CT0( 1) =  0.605450E-12
CT0( 2) =  0.419232E-12
CT0( 3) =  0.436967E-12
CT0( 4) =  0.479087E-12
Inside package, Rin =   643.400     RL =   346.470
Diode model Series Ls, Rin+Rs, S(v) shunted by Cp
Required impedances outside package.
Zin =   456.218     + j -606.069
Zout =   208.267     + j -242.991
Match these impedances with their complex conjugate
Match idler  2 with conjugate  of 0 + j  -379.181
Match idler  3 with conjugate  of 0 + j  -242.125
```

## PROBLEMS

**10.1** In Appendix D derive (D.9) from (D.10).

**10.2** In Appendix E derive the common gate $S$ parameters from the presumably known three-port $S$ parameters.

**10.3** Prove the stability factor $S'$ is that given in Eq. (10.59).

**10.4** The measurements of a certain active device as a function of current give $Z_d(10\text{ mA}) = -20 + j30\ \Omega$ and $Z_d(50\text{ mA}) = -10 + j15\ \Omega$. The passive circuit to which this is connected is measured at two frequencies: $Z_c(800\text{ MHz}) = 12 - j10\ \Omega$ and $Z_c(1000\text{ MHz}) = 18 - j40\ \Omega$. Determine whether the oscillator will be stable in the given ranges of frequency and current amplitude. Assume that the linear interpolation between the given values is justified.

## REFERENCES

1. J. K. Clapp, "An Inductance-Capacitance Oscillator of Unusual Frequency Stability," *Proc. IRE*, Vol. 36, pp. 356–358, 1948.
2. J. K. Clapp, "Frequency Stable LC Oscillators" *Proc. IRE*, Vol. 42, pp. 1295–1300, 1954.
3. J. Vackar, "LC Oscillators and Their Frequency Stability," *Telsa Tech. Reports*, Czechoslovakia, pp. 1–9, 1949.
4. A. R. Hambley, *Electronics: A Top–Down Approach to Computer-Aided Circuit Design*, New York: Macmillan, 1994, p. 959.
5. W. K. Chen, *Active Network and Feedback Amplifier Theory*, New York: McGraw-Hill, 1980.
6. K. Kurokawa, "Some Basic Characteristics of Broadband Negative Resistance Oscillator Circuits," *Bell Syst. Tech. J.*, Vol. 48, pp. 1937–1955, 1969.
7. R. Adler, "A Study of Locking Phenomena in Oscillators," *Proc. IRE.*, Vol. 22, pp. 351–357, 1946.
8. K. Kurokawa, "Injection Locking of Microwave Solid-State Oscillators," *Proc. IEEE*, Vol. 61, pp. 1386–1410, 1973.
9. M. Uenohara and J. W. Gewartowski, "Varactor Applications," in H. A. Watson, ed., *Microwave Semiconductor Devices and Their Circuit Applications*, New York: McGraw-Hill, 1969, pp. 194–270.
10. C. B. Burckhardt, "Analysis of Varactor Frequency Multipliers for Arbitrary Capacitance Variation and Drive Level," *Bell Syst. Tech. J.*, Vol. 44, pp. 675–692, 1965.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER ELEVEN

# RF Mixers

## 11.1 NONLINEAR DEVICE CHARACTERISTICS

A typical mixer is a three-port circuit that accepts two signals at two different frequencies and produces at the third port a signal that is the sum or difference of the two input frequencies. Production of a new frequency or frequencies requires a nonlinear device. The two most common semiconductor nonlinear characteristics are of the form $e^{qV(t)/kT}$ as found in *pn* junction diodes or bipolar junction transistors and of the form $I_{\mathrm{DSS}}(1 - V(t)/V_{\mathrm{T}})^2$ as found in field effect transistors. Schottky barrier diodes are not described here, since they are mostly used out of necessity for low-noise high-microwave frequency applications.

Consider a *pn* junction nonlinearity that is excited by two signals (plus a dc term):

$$V(t) = V_{\mathrm{dc}} + V_{\mathrm{p}} \cos \omega_{\mathrm{p}} t + V_1 \cos \omega_1 t \tag{11.1}$$

The device current would then be of the form

$$I(t) = I_{\mathrm{s}} e^{V_{\mathrm{dc}}/V_{\mathrm{T}}} [e^{V_{\mathrm{p}} \cos \omega_{\mathrm{p}} t} \cdot e^{V_1 \cos \omega_1 t}] \tag{11.2}$$

where the thermal voltage, $V_{\mathrm{T}}$, is defined as $kT/q$, $k$ is Boltzmann's constant, $T$ is the absolute temperature, and $q$ is the magnitude of the electronic charge. It is known, however, that this can be simplified by expressing it in terms of modified Bessel functions because

$$e^{z \cos \theta} = I_0(z) + 2 \sum_{n=1}^{\infty} I_n(z) \cos n\theta \tag{11.3}$$

where $I_n(z)$ is the modified Bessel function of order $n$ and argument $z$ [1]. The Bessel function has the property that as $n$ increases and $z$ decreases the function itself decreases. The two exponentials in Eq. (11.2) indicate there are two infinite

series of the form shown in Eq. (11.3): the first with summation index $n$ and the second with index $m$. The current given in Eq. (11.2) can be found by the appropriate substitution:

$$\begin{aligned} I(t) &= I_s e^{V_{dc}/V_T}\left[I_0(V_p) + 2\sum_{n=1}^{\infty} I_n(V_p)\cos n\omega_p t\right] \\ &\quad \times \left[I_0(V_1) + 2\sum_{m=1}^{\infty} I_n(V_1)\cos m\omega_1 t\right] \\ &= I_{dc}e^{V_{dc}/V_T} I_0(V_p) I_0(V_1) \\ &\quad + 2I_{dc}e^{V_{dc}/V_T}\left(I_0(V_1)\sum_{n=1}^{\infty} I_n(V_p)\cos n\omega_p t + I_0(V_p)\sum_{m=1}^{\infty} I_m(V_1)\cos m\omega_1 t\right) \\ &\quad + 4I_{dc}e^{V_{dc}/V_T}\left[\sum_{n=1}^{\infty} I_n(V_p)\cos n\omega_p t\right]\cdot\left[\sum_{m=1}^{\infty} I_m(V_1)\cos m\omega_1 t\right] \end{aligned} \tag{11.4}$$

The basic result is a set of frequencies $n\omega_p + m\omega_1$ where $n$ and $m$ can take on any integer. The actual current values at any given frequency would be greatly modified by circuit impedances at these frequencies as well as by variations in the device itself. In the usual mixer application shown in Fig. 11.1, the amplitude of the local oscillator voltage, $V_p$, is typically 40 dB greater than the RF signal voltage $V_1$. Consequently the number of frequencies drops to

$$\omega_n = n\omega_p + \omega_0 \tag{11.5}$$

The higher-order mixing products are reduced in amplitude by approximately $1/n$. The usual desired output for a receiver is the intermediate frequency (IF), $\omega_0$. The frequencies of primary interest are given the following names:

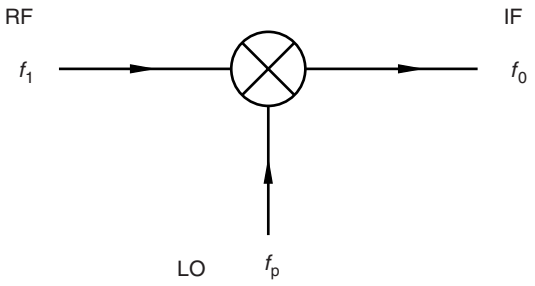



**FIGURE 11.1** Schematic diagram for a mixer.

| | |
|---|---|
| $\omega_p$ | Local oscillator (pump) frequency |
| $\omega_0 = \omega_1 - \omega_p$ | Intermediate frequency |
| $\omega_1$ | RF signal frequency |
| $\omega_{-1} = -\omega_p + \omega_0$ | Image frequency |
| $\omega_2 = 2\omega_p + \omega_0$ | Sum frequency |

In the FET type of nonlinearity, the current as a result of excitation given by Eq. (11.1) is

$$\frac{I(t)}{I_{DSS}} = \left(1 - \frac{V_p}{V_T}\cos\omega_p t - \frac{V_1}{V_T}\cos\omega_1 t\right)^2 \tag{11.6}$$

$$= 1 - 2\left(\frac{V_p}{V_T}\cos\omega_p t + \frac{V_1}{V_T}\cos\omega_1 t\right) + \frac{V_p^2}{2V_T^2}\left(1 + \cos 2\omega_p t\right) + \frac{V_1^2}{2V_T^2}\left(1 + \cos 2\omega_1 t\right) + \frac{V_p V_1}{V_T^2}\left(\cos(\omega_p + \omega_1)t + \cos(\omega_p - \omega_1)t\right) \tag{11.7}$$

While it may appear that the FET is "less nonlinear" than the *pn* junction type of nonlinearity, it should be remembered that the circuit into which the device is embedded will reflect back into the nonlinear device and create multiple mixing products. However, ultimately the frequencies will follow, at least potentially, the values shown in Fig. 11.2.

Readily apparent from the foregoing, a measure of patience is necessary to unravel all the frequency terms and their relative amplitudes. The nonlinear device is sometimes modeled as a power series of the applied voltages:

$$I(t) = I_{dc} + aV(t) + bV^2(t) + cV^3(t)\ldots \tag{11.8}$$

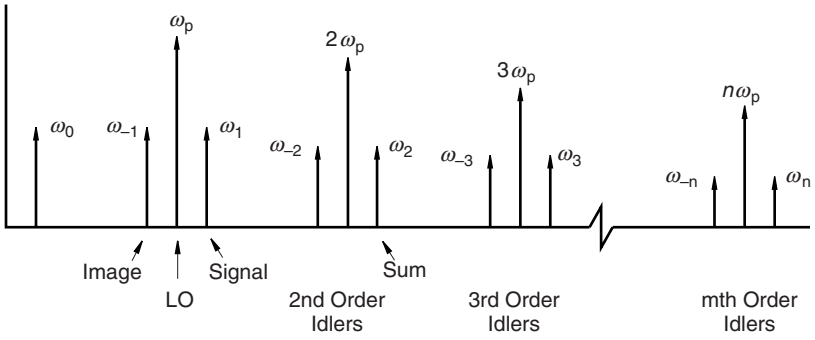



**FIGURE 11.2** Frequency components in a mixer.

Rather than directly determining the mixing products by multiplication, it is more convenient to determine these in the frequency domain by employing the Fourier transform [2]. The most convenient way of writing the Fourier transform pair is symmetrically where $f$ is used rather than $\omega$:

$$g(t) = \int_{-\infty}^{\infty} G(f)e^{-j2\pi ft}df \tag{11.9}$$

$$G(f) = \int_{-\infty}^{\infty} g(t)e^{j2\pi ft}dt \tag{11.10}$$

The Fourier transform of an exponential function is a Dirac delta function:

$$\mathcal{F}(e^{-j\omega_a}) = 2\pi\delta(\omega - \omega_a) \tag{11.11}$$

If two voltages are represented as $V_a = \cos(\omega_a t + \theta_a)$ and $V_b = \cos(\omega_b t + \theta_b)$. These are to be multiplied together. Before doing this, they can each be converted into the frequency domain by Eq. (11.10) and substituted into the convolution theorem:

$$G_o(f) = \int_{-\infty}^{\infty} G_a(\lambda)G_b(f - \lambda)d\lambda \tag{11.12}$$

For sinusoidal voltages the integral is simply a series of $\delta$ functions that are trivial to integrate. The process proceeds most easily graphically. The function $G_a(\lambda)$ is fixed, and the $G_b(f - \lambda)$ is allowed to slide from right to left. The nonzero parts of the integration occurs when $\delta$ functions coincide.

As an example, consider the frequencies that would result from the product of two voltages:

$$V_a(t) \cdot V_b(t) \tag{11.13}$$

where

$$V_a(t) = \cos(2\pi f_1 t + \theta_1)$$

and

$$V_b(t) = \cos(2\pi f_2 t + \theta_2)$$

The Fourier transform of $V_a(t)$ is

$$G_a(f) = \int_{-\infty}^{\infty} V_a(t)e^{j2\pi ft}dt \tag{11.14}$$

$$= \int_{-\infty}^{\infty} \frac{1}{2}\left[e^{j(2\pi f_1 t+\theta_1)} + e^{-j(2\pi f_1 t+\theta_1)}\right] e^{j2\pi ft}\frac{2\pi dt}{2\pi}$$

$$= \frac{1}{2}\left[\delta(f + f_1)e^{j\theta_1} + \delta(f - f_1)e^{-j\theta_1}\right] \tag{11.15}$$

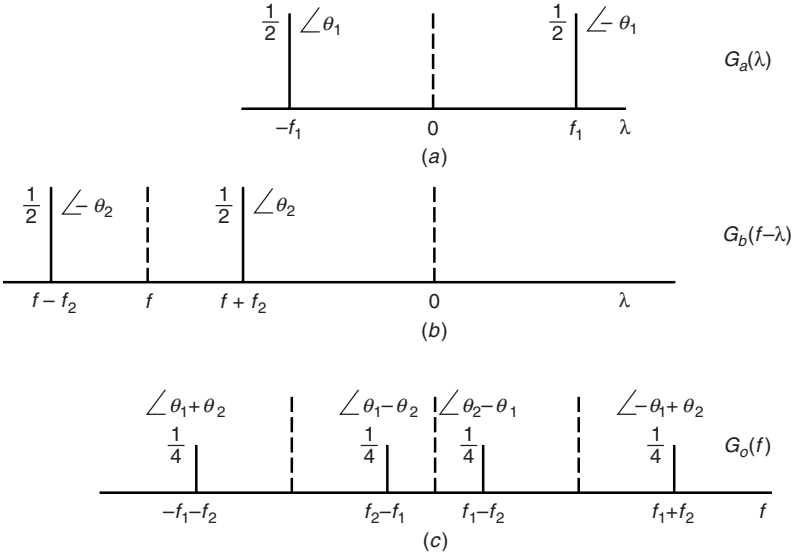



**FIGURE 11.3** Graphical integration of the convolution integral where (*a*) is $G_a(\lambda)$, (*b*) is $G_b(f-\lambda)$, and (*c*) is the result of the integration.

A similar function results from $V_b(t)$. The Fourier transform, $G_a(\lambda)$, is shown in Fig. 11.3*a*, which displays both the magnitude and the phase of terms. The term $G_b(-\lambda)$, which is found in similar fashion, is offset by $f$ as indicated in Eq. (11.12) and is seen in Fig. 11.3*b*. As $f$ increases, $G_b(f-\lambda)$ moves from left to right. No contribution to the convolution integral occurs until $f + f_2 = f_1$ or $f = -f_1 - f_2$. This is the leftmost line shown in Fig. 11.3*c*. As $f$ continues to increase, all four intercepts between $G_a(\lambda)$ and $G_b(f-\lambda)$ are found. While the amount of effort in using the frequency domain approach described here and the time domain approach of multiplying sines and cosines in this example is about the same, adding a third frequency quickly tilts the ease of calculation toward the frequency domain approach.

## 11.2 FIGURES OF MERIT FOR MIXERS

The quality of a mixer rests on a number of different mixer parameters which of course must fit the application under consideration. The first of these is *conversion loss*, $L$. This is the ratio of the delivered output power to the input available power.

$$L = \frac{\text{output IF power delivered to the load, } \mathrm{P_0}}{\text{available RF input signal power, } \mathrm{P_1}}$$

Clearly, the conversion loss is dependent on the load of the input RF circuit as well as the output impedance of the mixer at the IF port. The conversion loss for a typical diode mixer is between 6 and 7 dB.

The *noise figure* is a measure of the noise added by the mixer itself to the RF input signal as it gets converted to the output IF. It specifically excludes the noise figure of the following IF amplifier and neglects the $1/f$ flicker noise. In practice, the mixer noise figure is very nearly the same as the conversion loss.

The *isolation* is the amount of local oscillator power that leaks into either the IF or the RF ports. For double-balanced mixers this value typically lies in the 15 to 20 dB range.

A single- or double-balanced mixer will convert energy in the upper or lower sidebands with equal efficiency. Consequently noise in the sideband with no signal will be added to the IF output, which of course will increase the noise figure by 3 dB in the IF port. *Image rejection* mixers will block this unwanted noise from the IF port.

The *conversion compression* is the RF input power, above which the RF input in terms of the IF output deviates from linearity by a given amount. For example, the 1 dB compression point occurs when the conversion loss increases by 1 dB above the conversion loss in the low-power linear range. A typical value of 1.0 dB compression occurs when the RF power is $+7$ dBm and the LO is $+13$ dBm.

The *LO drive power* is the required LO power level needed to make the mixer operate in optimal fashion. For a double-balanced mixer, this is typically $+6$ dBm to $+20$ dBm.

The *dynamic range* is the maximum RF input power range for the mixer. The maximum amplitude is limited by the conversion compression, and the minimum amplitude is limited by the noise figure.

The *input intercept point* is the RF input power at which the output power levels of the undesired intermodulation products and the desired IF output would be equal. In defining the input intercept point, it is assumed that the IF output power does not compress. It is therefore a theoretical value and is obtained by extrapolating from low-power levels. The higher this power level, the better is the mixer. Sometimes an output intercept point is used. This is the input intercept point minus the conversion loss. The idea of intercept points is described in greater detail in Section 11.6.

The *two-tone third-order intermodulation point* is a measure of how the mixer reacts when two equal amplitude RF frequencies excite the RF input port of the mixer.

## 11.3 SINGLE-ENDED MIXERS

Mixers are usually classed as single ended, single balanced, or double balanced. The technical advantages of the double-balanced mixer over the other two usually precludes using the slightly lower cost of the single-ended or single-balanced types in RF circuits. The are used, though, in millimeter wave circuits where geometrical constraints and other complexities favor using the simpler single-ended mixer.

The single-ended mixer in Fig. 11.4 shows that the RF input signal and the local oscillator signal enter the mixer at the same point. Some degree of isolation between the two is achieved by using a directional coupler in which the RF

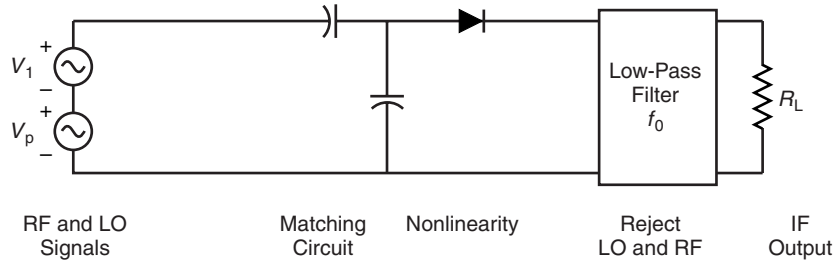



**FIGURE 11.4** Single-ended mixer.

signal enters the direct port and the local oscillator enters through the coupled port. The amplitude of the local oscillator, even after passing through the coupler, is large enough to turn the diode on and off during each cycle. Indeed, the LO power is so large as to cause clipping of the LO voltage, thereby approximating a square wave. The small RF signal is then presented with alternately a short or open circuit at the LO rate. It is this turning on and off of the RF frequency that produces the $|n f_p \pm f_1|$ set of frequencies. The one of most interest in the standard receiver is $f_0 = f_p - f_1$. Among the disadvantages of the single-ended mixer are a high-noise figure, a large number of frequencies generated because of the nonlinear diode, a lack of isolation between the RF and LO signals, and large LO currents in the IF circuit. The RF to LO isolation problem can be very important, since the LO can leak back out of the RF port and be radiated through the receiver antenna. The LO currents in the IF circuit would have to be filtered out with a low-pass filter that has sufficient attenuation at the LO frequency to meet system specifications. It does have the advantage of requiring lower LO power than the other types of mixers.

Rather than using a switching diode, a FET can be switched at the LO rate. One such design is when the LO and RF signal both enter the FET gate and the output IF signal is developed in the drain circuit. The nonlinearity of the FET implies that fewer spurious signals are generated than the "more" nonlinear diode. Furthermore it is possible to achieve conversion gain between the input RF and output IF signals. A second alternative would be to excite the gate with the RF signal and the source with the LO; then the output IF is developed in the drain. This circuit offers improved isolation between the RF and LO signals but at the cost of higher LO power requirements. The dual gate FET is often used in which one gate is excited with the LO and the other with the RF. The IF is again developed in the drain circuit. This circuit offers even better isolation between RF and LO but its gain is somewhat lower.

## 11.4 SINGLE-BALANCED MIXERS

The single-balanced (or simply balanced) mixer has either two or four diodes as shown in the examples of Fig. 11.5. In all of these cases, when the LO voltage

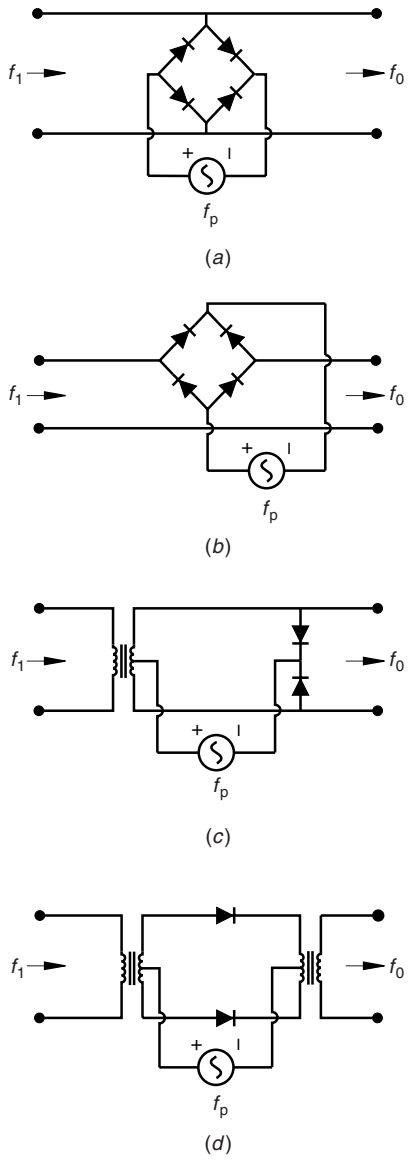



**FIGURE 11.5** Four possible single-balanced mixers.

has a large positive value, all the diodes are shorted. When the LO voltage has a large negative value, all the diodes are open. In either case, the LO power cannot reach the IF load nor the RF load because of circuit symmetry. However, the incoming RF voltage sees alternately a path to the IF load and a blockage to the IF load. The block may either be an open circuit to the IF load or a short circuit to ground.

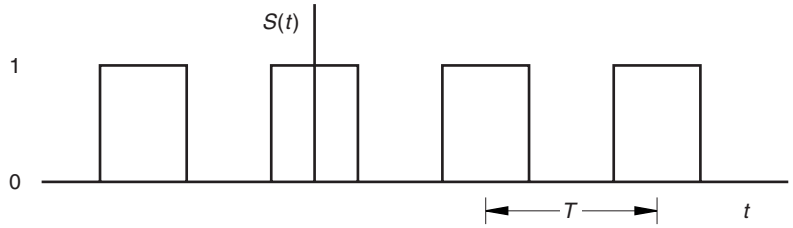



**FIGURE 11.6** Single-balanced mixer waveform.

As before, it is assumed that the LO voltage is much greater than the RF voltage, so $V_p \gg V_1$. The LO voltage can be approximated as a square wave with period $T = 1/f_p$ that modulates the incoming RF signal (Fig. 11.6). A Fourier analysis of the square wave results in a switching function designated by $S(t)$:

$$S(t) = \frac{1}{2} + \sum_{n=1}^{\infty} \frac{\sin(n\pi/2)}{n\pi/2} \cos n\omega_p t \tag{11.16}$$

If the input RF signal is expressed as $V_1 \cos \omega_1 t$, then the output voltage is this multiplied by the switching function:

$$V_0 = V_1 \cos \omega_1 t \cdot S(t) \tag{11.17}$$

$$= V_1 \cos \omega_1 t \left( \frac{1}{2} + \sum_{n=1}^{\infty} \frac{\sin(n\pi/2)}{n\pi/2} \cos n\omega_p t \right) \tag{11.18}$$

Clearly, the RF input signal voltage will be present in the IF circuit. However, only the odd harmonics of the local oscillator voltage will effect the IF load. Thus the spurious voltages appearing in the IF circuit are

$$f_1,\ f_p + f_1,\ 3f_p \pm f_1,\ 5f_p \pm f_1, \ldots$$

and all even harmonics of $f_p$ are suppressed (or balanced out).

## 11.5 DOUBLE-BALANCED MIXERS

The double-balanced mixer is capable of isolating both the RF input voltage and the LO voltage from the IF load. The slight additional cost of some extra diodes and a balun is usually outweighed by the improved intermodulation suppression, improved dynamic range, low conversion loss, and low noise. The two most widely used double balanced mixers for the RF and microwave band are the "ring" mixer and the "star" mixer depicted in Fig. 11.7. In the single-balanced mixer all the diodes were either turned on or turned off, depending on the instantaneous polarity of the local oscillator voltage. The distinguishing feature of the

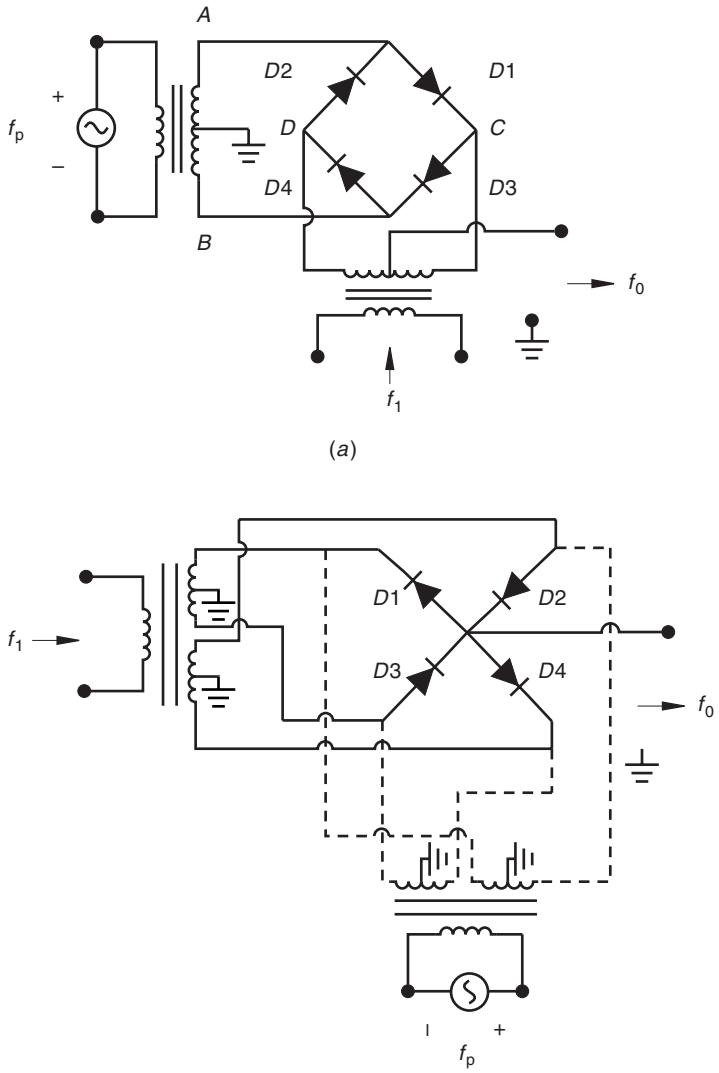



**FIGURE 11.7** Double-balanced mixers using (*a*) ring diode design and (*b*) diode star design.

double-balanced mixer is that half the diodes are on and half off at any given time. The diode pairs are switched on or off according to the local oscillator polarity. Thus the path from the signal port with frequency $f_1$ to the intermediate frequency port, $f_0$, reverses polarity at the rate of $1/f_p$.

In Fig 11.7*a*, when the LO is positive at the upper terminal, diodes *D*1 and *D*2 are shorted while diodes *D*3 and *D*4 are open. Current from the RF port

flows out of node $C$. As far as the RF frequency is concerned, nodes $A$ and $B$ lie midway between the positive and negative RF signal voltage. Therefore at the signal frequency, $f_1$, nodes $A$ and $B$ are at zero potential. During this instant, current is drawn from nodes $A$ and $B$ by way of the LO transformer secondary. The RF signal current is induced into the RF transformer secondary and on out to the IF load. When the LO switches to the negative polarity, diodes $D3$ and $D4$ are shorted and diodes $D1$ and $D2$ are open. The RF signal current will then flow into node $D$ and on to nodes $A$ and $B$ as before. Now, however, the RF current at $f_1$ flows in the opposite direction in the RF signal transformer secondary and thus out of the IF load. The switching of the polarity at the LO frequency, $f_p$, of the current in the IF circuit produces the difference frequency, $f_0$. Symmetry would suggest that the IF power could be extracted from the center tap of the LO secondary rather than the RF signal secondary. However, the LO power, being so much higher than the RF signal power, the isolation between the LO and IF would be poorer.

An analysis of this mixer can be done in SPICE in which the diodes are replaced by ideal voltage switches. An example of this is illustrated in Fig. 11.8 in which the local oscillator is set at 900 MHz and the RF signal is at 800 MHz. The resulting time domain output shown in Fig. 11.9 is not easily interpreted. The Fourier transform in Fig. 11.10 clearly shows the resulting IF output frequency at 100 MHz along with other frequencies generated by the mixer.

The star circuit shown in Fig. 11.7*b* also acts as a double-balanced mixer. An advantage over the ring mixer is that the central node of the four diodes allows direct connection to the IF circuit. On the other hand, the star mixer requires a more complicated transformer in the RF signal and LO ports. When the LO voltage is positive, diodes $D1$ and $D2$ are shorted and diodes $D3$ and $D4$ are open. The RF signal current from the upper terminals of the secondary winding flows to the IF port. When the LO voltage is negative, diodes $D3$ and $D4$ are shorted and diodes $D1$ and $D2$ open. The current then flows from the lower terminals of the RF signal transformer secondary. The RF signal current in the IF circuit has switched polarity. The switching rate produces an output at the difference frequency, $f_0$. In both these cases the switching function is shown in Fig. 11.11. Fourier analysis provides the following time domain representation of the switching function, which differs from Eq. (11.18) by a lack of a dc term:

$$S(t) = 2\sum_{n=1}^{\infty} \frac{\sin(n\pi/2)}{n\pi/2} \cos n\omega_p t \tag{11.19}$$

The IF voltage is found as before for the single-balanced mixer:

$$\begin{aligned} V_0 &= V_1 \cos \omega_1 t \cdot S(t) \\ &= 2V_1 \cos \omega_1 t \left( \sum_{n=1}^{\infty} \frac{\sin(n\pi/2)}{n\pi/2} \cos n\omega_p t \right) \end{aligned} \tag{11.20}$$

```
Double Balanced Diode Mixer
* Local Oscillator
VP      10   0   SIN(0   2.   900E6)
RP      10  1   .01
RRF     20  4   .01
* RF signal
VRf     20   0   SIN(0   .2    800E6)
LP      1   0    1uH
LPA     2   0   .5uH
LPB     0   3   .5uH
KP1     LP  LPA  LPB  1
LR      4   0    1uH
LRA     5   6   .5uH
LRB     6   7   .5uH
KRF1   LR   LRA  LRB  1
*  Ideal voltage switches represent diodes.
SD1      2   7   2   7   SWMOD
SD2      5   2   5   2   SWMOD
SD3      7   3   7   3   SWMOD
SD4      3   5   3   5   SWMOD
RLIF    6   0    50
.MODEL  SWMOD  VSWITCH (RON=.2, ROFF=1.E5 VON=.7
  VOFF=.6)
.PROBE
.OP
*       Start    Final   Begin Prt   ceiling
.TRAN      1nS      50nS      0
*.TRAN    .05nS     20nS      0        10pS
* IF output is V(6)
.PRINT TRAN V(6)
.END
```

**FIGURE 11.8** SPICE net list for diode ring mixer.

Clearly, there is no RF signal nor LO voltage seen in the IF circuit, nor any even harmonics of the LO voltage.

The description above of mixers has assumed the use of ideal diodes. The diodes are in fact either *pn* or Schottky barrier (metal–semiconductor) junctions with a nonzero forward voltage drop and nonzero leakage current in the reverse bias condition. The Schottky barrier devices are particularly useful when low noise is required at high microwave frequencies. The device and package parasitic elements limit mixer frequency response, although designs based on the above analysis have been made to work at frequencies exceeding 26 GHz.

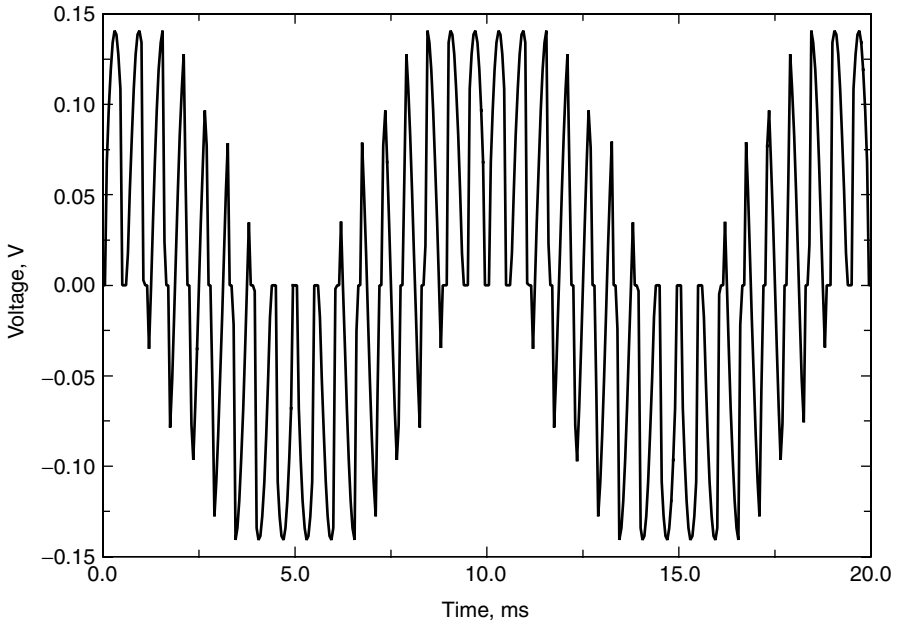



**FIGURE 11.9** Time domain response of a double-balanced mixer using ideal switches.

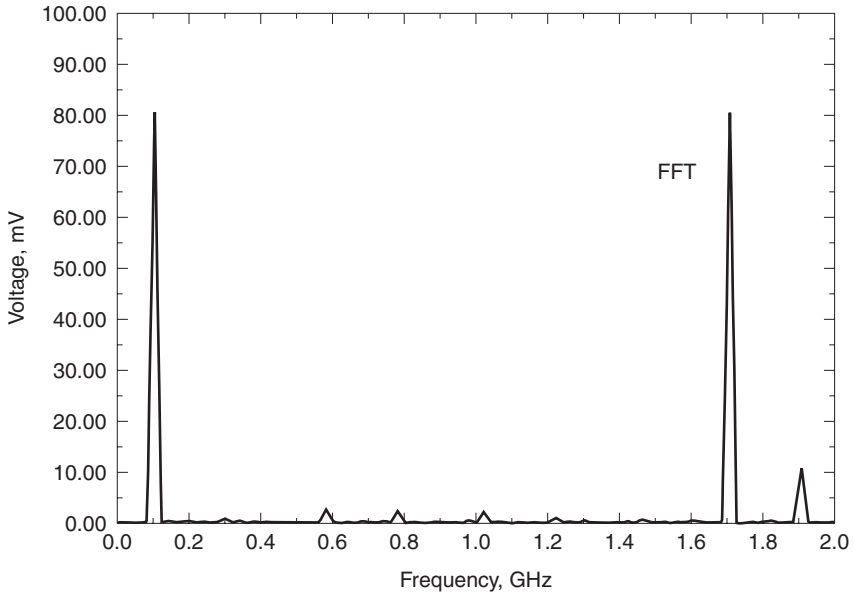



**FIGURE 11.10** Fast Fourier transform of the time function showing the frequency components off the double-balanced mixer.

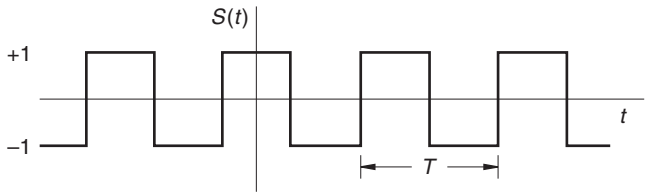



**FIGURE 11.11** Double-balanced mixer waveform.

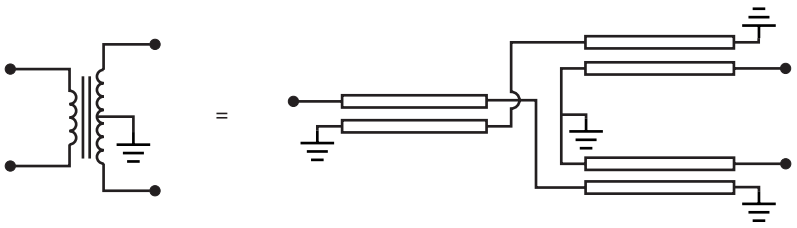



**FIGURE 11.12** Transmission line transformer equivalent to the center-tapped transformer.

This analysis was also based on the availability of ideal center-tapped transformers. At RF frequencies, these can be realized using transmission line transformers, as shown in Fig. 11.12.

The double-balanced ring mixer described above used a single diode in each arm of the ring. Such a mixer is termed a class 1 mixer. Class 2 mixers are obtained by replacing the single diode in each arm of the ring with two diodes in series or with a diode or resistor in series (Fig. 11.13). The precision resistor in the later case can be adjusted to improve the ring balance and thus the intermodulation distortion. More complex ring elements can be used to further improve intermodulation distortion with the added cost of increasing the amount of LO power required to drive the diodes. More detailed information on design of RF and microwave mixers is available in [3,4].

## 11.6 DOUBLE-BALANCED TRANSISTOR MIXERS

Transistors can also be used as the mixing element in all three types of mixers described above, though only the double-balanced configuration is described here. These are called *active mixers* because they provide the possibility of conversion gain that the diode mixers are not capable of doing. They produce approximately the same values of port isolation and suppression of even harmonic distortion as the diode mixers. One example of such a circuit is a transistor ring of enhancement mode $n$-channel MOSFETs in which the gate voltage must exceed zero in order for the transistor to turn on (Fig. 11.14). When the LO voltage is positive as indicated, the pair of transistors on the right-hand side is turned on, and the

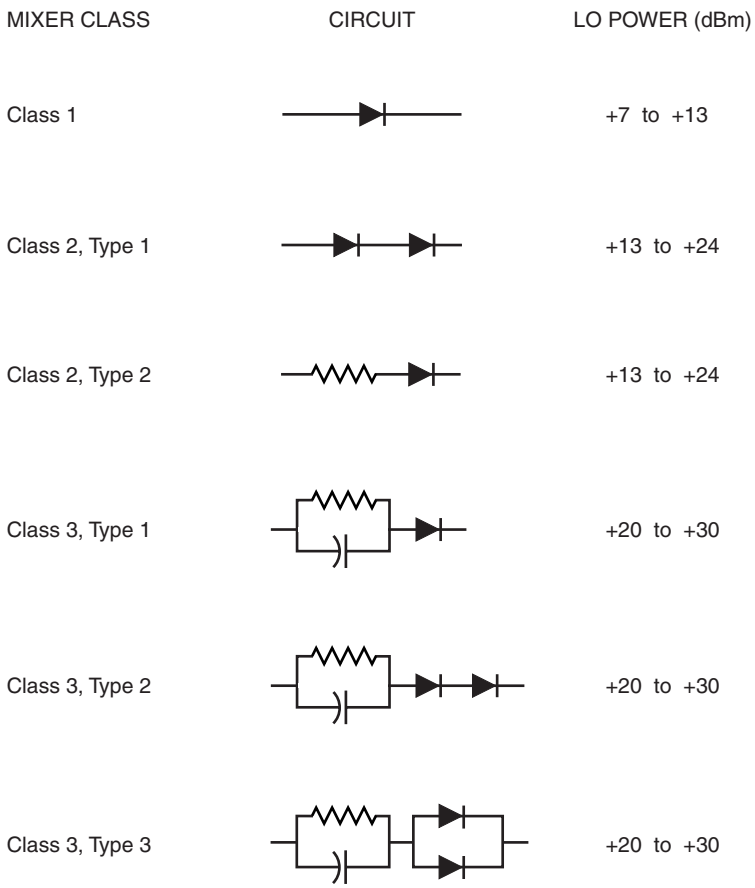



**FIGURE 11.13** Double-balanced mixer classes is based on the elements in each branch. Required LO power levels increases with circuit complexity. (From [5].)

left-hand pair is turned off. When the LO voltage is negative, the two pairs of transistors switch roles. In this process the path from the RF signal switches back and forth between the positive and negative IF ports at the LO switching rate. While the balance of the polarity of the RF signal voltage precludes it from being seen at the IF port, the difference frequency generated by the switching action does appear across the IF terminals.

An alternative design is based on the Gilbert cell multiplier [6]. An analysis of the elementary Gilbert cell in Fig. 11.15 is most easily accomplished by assuming that the base and reverse bias saturation currents are negligible, that the output resistances of the transistors are infinite, and that the bias source is ideal. Considering, for the moment, transistors $Q1$, $Q2$, and $Q5$ current continuity demands,

$$I_{C5} = I_{C1} + I_{C2} \tag{11.21}$$

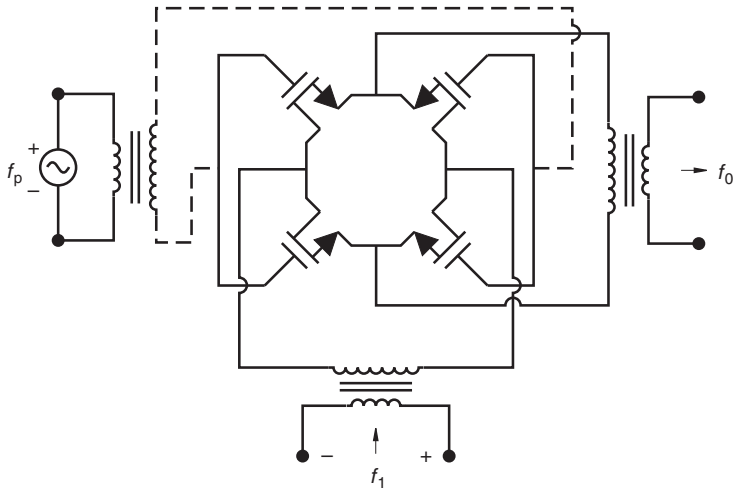



**FIGURE 11.14** Double-balanced mixer using MOSFETs.

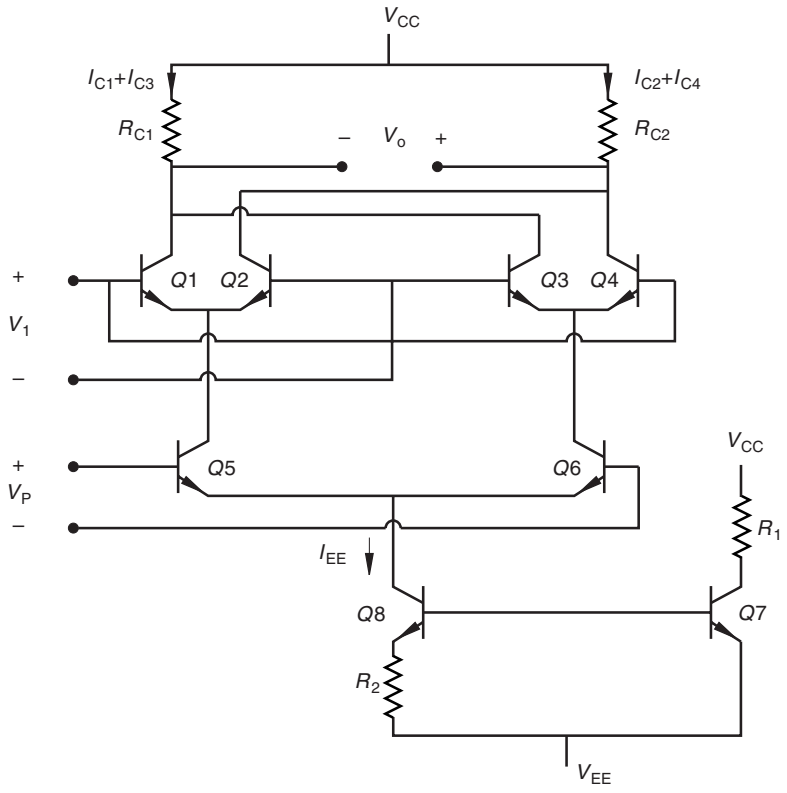



**FIGURE 11.15** Gilbert cell used as a modulator.

The ratio of the Schottky diode equations with negligible saturation current gives a second relationship:

$$\frac{I_{C1}}{I_{C2}} = \frac{e^{V_{BE1}/V_T}}{e^{V_{BE2}/V_T}} = e^{V_1/V_T} \tag{11.22}$$

Combining of these two equations gives an expression for $I_{C1}$. In like manner the currents for $Q2$, $Q3$, and $Q4$ are found:

$$I_{C1} = \frac{I_{C5}}{1 + e^{-V_1/V_T}} \tag{11.23}$$

$$I_{C2} = \frac{I_{C5}}{1 + e^{V_1/V_T}} \tag{11.24}$$

$$I_{C3} = \frac{I_{C6}}{1 + e^{V_1/V_T}} \tag{11.25}$$

$$I_{C4} = \frac{I_{C6}}{1 + e^{-V_1/V_T}} \tag{11.26}$$

For $Q5$ and $Q6$ the collector currents are

$$I_{C5} = \frac{I_{EE}}{1 + e^{-V_2/V_T}} \tag{11.27}$$

$$I_{C6} = \frac{I_{EE}}{1 + e^{V_2/V_T}} \tag{11.28}$$

The output voltage is proportional to the difference of the currents through the collector resistors:

$$\begin{aligned} V_O &= [(I_{C1} + I_{C3}) - (I_{C2} + I_{C4})]R \qquad (11.29)\\ &= [(I_{C1} - I_{C4}) - (I_{C2} - I_{C3})]R \\ &= \frac{R(I_{C5} - I_{C6})}{1 + e^{-V_1/V_T}} - \frac{R(I_{C5} - I_{C6})}{1 + e^{V_1/V_T}} \\ &= \frac{I_{EE}R}{1 + e^{-V_1/V_T}}\left(\frac{1}{1 + e^{-V_2/V_T}} - \frac{1}{1 + e^{V_2/V_T}}\right) \\ &\quad - \frac{I_{EE}R}{1 + e^{V_1/V_T}}\left(\frac{1}{1 + e^{-V_2/V_T}} - \frac{1}{1 + e^{V_2/V_T}}\right) \\ &= \frac{I_{EE}R}{1 + e^{-V_1/V_T}}\left(\frac{e^{V_2/2V_T}}{e^{V_2/2V_T} + e^{-V_2/2V_T}} - \frac{e^{-V_2/V_T}}{e^{-V_2/2V_T} + e^{V_2/2V_T}}\right) \\ &\quad - \frac{I_{EE}R}{1 + e^{V_1/V_T}}\left(\frac{e^{V_2/2V_T}}{e^{V_2/2V_T} + e^{-V_2/2V_T}} - \frac{e^{-V_2/2V_T}}{e^{-V_2/2V_T} + e^{V_2/2V_T}}\right) \\ &= I_{EE}R \tanh\left(\frac{V_2}{2V_T}\right)\tanh\left(\frac{V_1}{2V_T}\right) \qquad (11.30) \end{aligned}$$

```
Gilbert Cell
VRF       1    4    SIN (0 .2  800MEG ) DC   0
VP        8    9    SIN (0  2  900MEG ) DC   0
VCC       7    0    DC   15
VEE       0    12   DC   15
Q1        2    1    3    DEVICE
Q2        6    4    3    DEVICE
Q3        2    4    5    DEVICE
Q4        6    1    5    DEVICE
Q5        3    8    10   DEVICE
Q6        5    9    10   DEVICE
Q7        11   11   12   DEVICE
Q8        10   11   13   DEVICE
R1        7    11   15
R2        13   12   100
RC1       7    2    30k
RC2       7    6    30k
.MODEL    DEVICE   NPN
.PROBE
.DC VRF -100m  100m 10m VP -100m 100m 20m
*   Print step, Final time, Print start, Step ceiling
.TRAN  1nS     100nS      0
* IF output is V(2,6)
* DC analysis
.TF    V(6) VRF
*.TF   V(6) VP
.END
```

**FIGURE 11.16** SPICE list for the Gilbert multiplier.

Since $\tanh x \approx x$ for $x \ll 1$, the monomial type of multiplication between the two input voltages will occur as long as $V_i \ll 2V_T$, where $i = 1, 2$. At the other extreme, when $x \gg 1$, $\tanh x \approx 1$.

The modulator application typically has one large input voltage (LO) and one small one (RF signal). A positive value of the LO voltage, shown as $V_1$ in Fig. 11.15, will then cause $Q1$ and $Q4$ to be turned on, while $Q2$ and $Q3$ are turned off. As in the previous double-balanced mixers, the LO switches the RF signal voltage path to the IF port at the frequency, $f_p$, so that the difference frequency is generated. A SPICE analysis of the Gilbert cell (Fig. 11.16) again demonstrates the production of an IF output between the collectors of $Q1$ and $Q2$.

This same circuit can be realized using field effect transistors. In either case a large RF signal input can cause the mixer to operate outside of its linear region. The mixer dynamic range can be improved by adding emitter (source) degeneracy. This is a small resistor (usually in the 100's of Ohms) in the emitter circuit. Another scheme is to introduce a filter between the lower two transistors

and the upper ones [7]. Distortion products produced in $Q5$ and $Q6$ are thus filtered out before the RF signal reaches the transistors being switched by the LO. A 20 dB improvement in dynamic range over the conventional Gilbert cell is reported using this filtering technique.

## 11.7 SPURIOUS RESPONSE

The previous sections considered some representative mixer circuits. Here some of the primary mixer performance criteria for mixers are described. The first of these are the spurious frequencies generated when the mixer is excited by a single tone RF signal. A second measurement of mixer performance results from exciting it with two tones near to each other that produces two IF terms. The latter is termed *two-tone intermodulation distortion*.

Single-tone intermodulation is an effect of the imbalance in the transformers or the diodes used in the mixer. A distinction is made between the inherent nonlinear current–voltage curve of a diode and the nonlinearity associated with the switching action of the diode [8]. Fitting a polynomial function to an ideal diode characteristic whose current is zero when off, and whose $i-V$ slope is a straight line when the diode is on, would yield a polynomial fitting function with many powers of the independent variable. Indeed, the switching of the diodes appears to be the predominant effect in a mixer. Analytical estimates of intermodulation distortion suppression can be made solely on the basis of the switching action of the diodes in the mixer rather than on any curvature of individual diode curves. Such an expression is presented in Appendix H. That equation has also been coded in the program IMSUP as described in Appendix H. Basically the intermodulation suppression in dBc (dB below the carrier) is $S_{nm}$ for a set of frequencies $nf_p \pm mf_1$.

Two-tone intermodulation distortion is best explained by following a simple experimental procedure. Normally one RF signal excites the RF port of the mixer, which then produces the IF output frequency along with various higher-order terms that can be easily filtered out of the IF circuit. Now consider exciting the RF port of the mixer with two RF signals, $f_{1a}$ and $f_{1b}$, spaced close together, which thus lie within the pass band of the mixer input. The nonlinear mixer circuit will then produce the following frequencies:

$$(\pm m_1 f_{1a} \pm m_2 f_{1b}) \pm n f_p \tag{11.31}$$

The order of the mixing product is $m_1 + m_2$. It would be nice if the IF output were only $|f_{1a} - f_p|$ and $|f_{1b} - f_p|$, since that would represent the down-converted signal to the IF output. Those terms containing harmonics of $f_p$ would be far outside the band of interest and could be filtered out. There are essentially two possibilities for the second-order intermodulation products:

$$(\pm 1 f_{1a} \pm 1 f_{1b}) \pm f_p$$

$$(\pm 1 f_{1a} \mp 1 f_{1b}) \pm f_p$$

In the first case, the output is near $3f_p$, and therefore well outside the IF pass band. The second case presents an output frequency slightly above or below the local oscillator frequency, $f_p$, which again is well outside the IF pass band. However, the third-order intermodulation products do present a special problem:

$$(\pm 2f_{1a} \mp 1f_{1b}) \pm f_p$$

$$(\pm 1f_{1a} \mp 2f_{1b}) \pm f_p$$

A numerical example illustrates what occurs with the third-order intermodulation products. If $f_p = 500$ MHz, the desired RF input signal is $f_{1a} = 410$ MHz, and a second signal of the same amplitude is at $f_{1b} = 400$ MHz. The first-order products would give the desired output IF frequencies and a high frequency that could be easily filtered out:

$$|f_{1a} \pm f_p| = 90, 910 \quad \text{MHz}$$

$$|f_{1b} \pm f_p| = 100, 900 \quad \text{MHz}$$

The third-order intermodulation products would be

$$|2f_{1a} - f_{1b} \pm f_p| = |820 - 400 \pm 500| = 80,\ 920 \quad \text{MHz}$$

$$|2f_{1b} - f_{1a} \pm f_p| = |800 - 410 \pm 500| = 110,\ 890 \quad \text{MHz}$$

As shown if Fig. 11.17, the undesired 80 and 110 MHz third-order intermodulation products could lie inside the IF pass band and thus distort the signal. The surest defense against this is to keep the amplitudes of the third-order intermodulation products small.

The measure of the size of the third-order intermodulation product is the intersection of third-order term with the desired first-order term, $f_0 = f_p - f_1$, (Fig. 11.17). The second-order intermodulation product is a result of having two

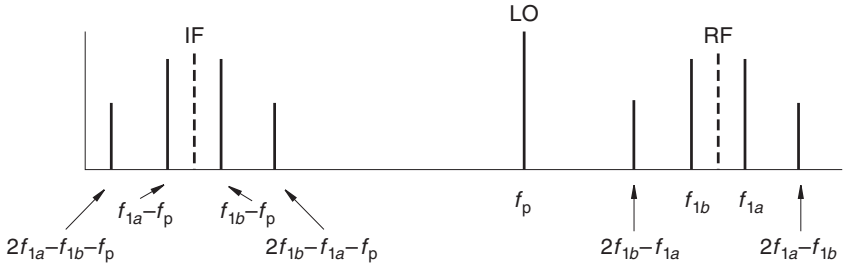



**FIGURE 11.17** Third-order intermodulation distortion.

RF signals that are multiplied together because of a quadratic nonlinearity:

$$[A \cos \omega_{1a} t \cdot B \cos \omega_{1b} t] \cos \omega_{\mathrm{p}} t$$

The resulting amplitude proportional to $AB$ will increase 2 dB when $A$ and $B$ each increase by 1 dB. The third-order intermodulation product is a result of a cubic nonlinearity:

$$[A^2 \cos^2 \omega_{1a} t \cdot B \cos \omega_{1b} t] \cos \omega_{\mathrm{p}} t$$

The resulting amplitude proportional to $A^2B$ will increase by 3 dB for every 1 dB rise in $A$ and $B$. Thus, when the RF signal rises by 1 dB, the desired IF term will rise by 1 dB, but the undesired third-order intermodulation term rises by 3 dB (Fig. 11.18). The interception of the extrapolation of these two lines in the output power relative to the input power coordinates is called the *third-order intercept point*. The input power level where this intersection occurs is called the *input intercept point*. The actual third-order intermodulation point cannot be directly measured, since that point must be found by extrapolation from lower-power levels. It nevertheless can give a single-valued criterion for determining the upper end of the dynamic range of a mixer (or power amplifier). The conversion compression on the desired output curve is the point where the desired IF output drops by 1 dB below the linear extrapolation of the low level values.

The range of mixer LO frequencies and RF signal frequencies should be chosen so as to reduce to a minimum the possibility of producing intermodulation products that will end up in the IF bandwidth. When dealing with multiple bands of frequencies, keeping track of all the possibilities that may cause problems is often done with the aid of computer software. Such programs are available free of charge off the internet, and other programs that are not so free.

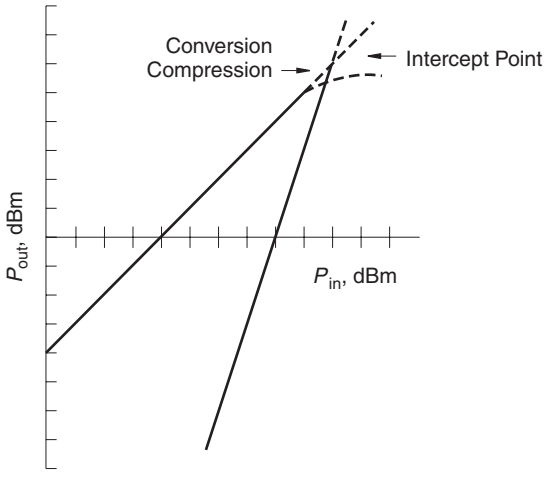



**FIGURE 11.18** Two-tone third-order intermodulation intercept point.

## 11.8 SINGLE-SIDEBAND NOISE FIGURE AND NOISE TEMPERATURE

The frequency independent noise power from a resistor is to a good approximation $kT$ where $k$ is Boltzmann's constant, and $T$ is the absolute temperature. In the two-port circuit shown in Fig. 11.19, a generator resistance, $R_G$, produces noise with an equivalent noise temperature of $T_G$. The network itself is characterized as having a certain transducer power gain, $G_T$, and noise temperature. When describing the noise temperature of a two-port, it must be decided if the noise is measured at the input or the output. The noise power at the output is presumably

$$T_{out} = G_T T_{in} \tag{11.32}$$

where $T_{in}$ is the noise temperature referred to the input port and $G_T$ is the transducer power gain. For mixers, this is the conversion gain between the signal and IF ports. In the land where amplifiers are broadband, linear and have wide dynamic range, Eq. (11.32) is accurate. However, low-level random noise voltages may not necessarily be amplified the same way a clean sinusoid would. But to wander from this idealistic world would complicate things beyond their basic usefulness for the present discussion. So the noise power delivered to the load, $Z_L$, is

$$N_L = k(G_T T_G + T_{out}) \tag{11.33}$$

or

$$\begin{aligned} T_L &= G_T T_G + T_{out} \\ &= G_T(T_G + T_{in}) \end{aligned} \tag{11.34}$$

While the load will generate its own noise, this is defined out of the equation. What is described here is the noise *delivered* to the load.

The noise figure is sometimes defined in terms of the signal-to-noise ratio at the input to the signal-to-noise ratio at the output of a two-port:

$$\begin{aligned} F &= \frac{S_i/N_i}{S_o/N_o} = \frac{S_i}{S_o} \cdot \frac{N_o}{N_i} \\ &= \frac{1}{G_T} \cdot \frac{G_T(T_G + T_n)}{T_G} \\ &= \left(1 + \frac{T_n}{T_G}\right) \end{aligned} \tag{11.35}$$

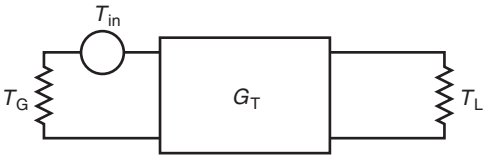



**FIGURE 11.19** Noise within the circuit is referred to the input side.

The noise figure depends on the temperature of the generator. This ambiguity in noise figure is removed by choosing by convention that the generator is at room temperature, $T_G = 290^\circ\text{K} \triangleq T_0$. Thus the noise characteristics of a two-port such as a mixer (the LO port being conceptually ignored) can be characterized with either noise figure or noise temperature. Because of the greater expansion of the temperature scale over that of noise figure in dB, noise temperature is preferred when describing very low noise systems and noise figure for higher-noise systems. However, the concept of noise temperature becomes increasingly convenient when describing mixers with their multiple frequency bands.

The noise figure of a mixer can be described in terms of single-sideband (SSB) noise figure or double-sideband (DSB) noise figure. If the IF term, $\omega_0$ in Fig. 11.2 comes solely from the signal $\omega_1$ and the image frequency $\omega_{-1}$ is entirely noise free, then the system is described in terms of its single-sideband noise figure, $F_{SSB}$ (Fig. 11.20*a*). Double-sideband noise figure comes from considering both the noise contributions of the signal and the image frequencies (Fig. 11.20*b*). In general, the output noise of the mixer will be the sum of the noise generated within the mixer itself and the noise power coming into the mixer multiplied by the mixer conversion gain. The noise power from inside the mixer itself can be referred to either the output port or the input port as described by Eq. (11.32). If all the internal mixer noise is referred back to the input RF signal port, then this will designated as $N_{SSB}$. The total noise power delivered to the load is found by multiplying $N_{SSB}$ by the RF port conversion gain, $G_{rf}$, and adding to this the power entering from the signal source, $N_G$, at both the RF signal and image frequencies:

$$N_L = (N_{SSB} + N_G)G_{rf} + N_G G_{im} \tag{11.36}$$

The gains at the RF signal and image frequencies, $G_{rf}$ and $G_{im}$, are typically very close to being the same since these two frequencies are close together. The terms in this definition are readily measurable, but Eq. (11.36) is at variance with the

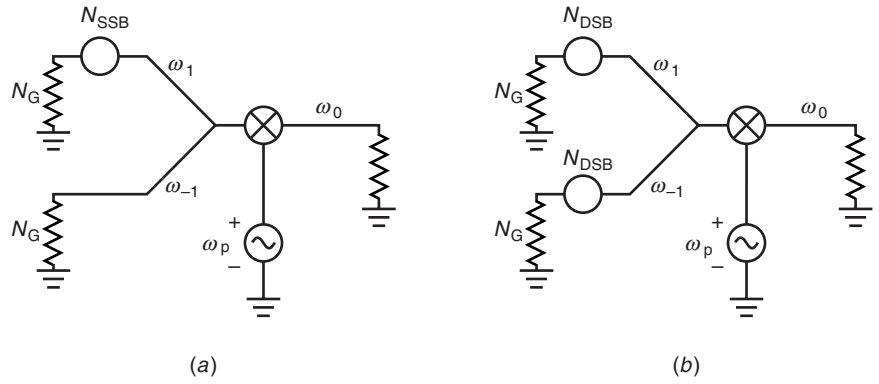



**FIGURE 11.20** Mixer noise specification using (*a*) single-sideband noise, and (*b*) double-sideband noise.

way the IEEE standards define single-sideband noise figure. For further discussion on this point, see [3]. The single-sideband noise figure is conventionally defined as the ratio of the total noise power delivered to the load to the noise power entering at the RF signal frequency from a generator whose temperature is $T_0$ and when the mixer itself is considered to be noise free:

$$F_{\mathrm{SSB}} = \frac{N_{\mathrm{L}}}{N_{\mathrm{G}} G_{\mathrm{rf}}} \tag{11.37}$$

Making the assumption $G_{\mathrm{rf}} = G_{\mathrm{im}}$,

$$\begin{aligned} F_{\mathrm{SSB}} &= \frac{N_{\mathrm{SSB}} G_{\mathrm{rf}} + 2 G_{\mathrm{rf}} N_{\mathrm{G}}}{G_{\mathrm{rf}} N_{\mathrm{G}}} \\ &= \frac{T_{\mathrm{SSB}}}{T_0} + 2 \end{aligned} \tag{11.38}$$

Since $N_{\mathrm{SSB}}$ is referred to the mixer input, so its associated noise temperature, $T_{\mathrm{SSB}}$, is also referred to the input side.

If the internal mixer noise power is referred back to both the RF frequency band and the image frequency band, then this power will be designated as the double-sideband power, $N_{\mathrm{DSB}}$. For the double-sideband analysis, both the RF signal and image frequencies are considered as inputs to the mixer. In this case the total power delivered to the load is

$$N_{\mathrm{L}} = (N_{\mathrm{G}} + N_{\mathrm{DSB}})(G_{\mathrm{rf}} + G_{\mathrm{im}}) \tag{11.39}$$

The double-sideband noise figure is determined by taking the ratio of the power delivered to the load and the power from both of these frequency bands if the mixer were considered noise free:

$$F_{\mathrm{DSB}} = \frac{N_{\mathrm{L}}}{(G_{\mathrm{rf}} + G_{\mathrm{im}}) N_{\mathrm{G}}} \tag{11.40}$$

Substituting Eq. (11.39) into Eq. (11.40) and again assuming that $G_{\mathrm{rf}} = G_{\mathrm{im}}$,

$$F_{\mathrm{DSB}} = \frac{T_{\mathrm{DSB}}}{T_0} + 1 \tag{11.41}$$

In the single-sideband case, all mixer noise power is referred to the mixer input at the RF signal frequency. In the double-sideband case, all the mixer noise is referred to the mixer input at both the RF signal and image frequencies. Since the internal mixer power is split between the two frequency bands,

$$T_{\mathrm{SSB}} = 2 T_{\mathrm{DSB}} \tag{11.42}$$

so that

$$F_{\text{SSB}} = \frac{T_{\text{SSB}}}{T_0} + 2 = \frac{2T_{\text{DSB}}}{T_0} + 2 = 2F_{\text{DSB}} \tag{11.43}$$

This illustrates the of-stated difference between single- and double-sideband noise figures. Noise figure specification of a mixer should always state which of these is being used.

## PROBLEMS

**11.1** Using the Fourier transform pair, show that $\mathcal{F}(e^{-j\omega_a}) = 2\pi\delta(\omega - \omega_a)$.

**11.2** Two closely separated frequencies are delivered to the input signal port of a mixer of a receiver. The center frequency of the receiver is 400 MHz, and the two input frequencies are at 399.5 and 400.5 MHz. The mixer has a conversion loss of 6 dB and the local oscillator is at 350 MHz. The power level of these two input frequencies is −14 dBm (dB below a milliwatt). At this input power, the third-order modulation products are at −70 dBm.

**(a)** What are the numerical values for the output frequencies of most concern to the receiver designer?

**(b)** What is the output third-order intercept point?

## REFERENCES

1. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Wiley, eq. 1972, 9.6.33–9.6.35.
2. H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: Wiley, 1980.
3. S. A. Maas, *Microwave Mixers*, 2nd ed., Norwood, MA: Artech House, 1993.
4. G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design*, New York: Wiley , 1990, ch. 7.
5. *Microwave Designer's Handbook*, Watkins-Johnson Co., 1997–98 Catalog.
6. B. Gilbert, "A Precise Four-Quadrant Multiplier with Subnanosecond Response," *IEEE J. Solid State Circuits*, pp. 365–373, 1968.
7. J. M. Moniz and B. Maoz, "Improving the Dynamic Range of Si MMIC Gilbert Cell Mixers for Homodyne Receivers," *IEEE 1994 Microwave and Millimeter-Wave Monolithic Circuits Symp.*, pp. 103–106, 1994.
8. D. G. Tucker, "Intermodulation Distortion in Rectifier Modulators," *Wireless Engineer*, pp. 145–152, 1954.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER TWELVE

# Phase Lock Loops

## 12.1 INTRODUCTION

Phase lock loops (PLL) are not a recent invention. Their use became widespread with the availability of high-quality integrated circuit operational amplifiers (op-amps) in the 1960s. This versatile circuit has found applications across the frequency spectrum in consumer, commercial, deep space, and military projects. Tracking *Voyager* through the solar system and tuning a car radio are made to order uses for a PLL. To understand a PLL, a good working knowledge of RF techniques, oscillator design, closed loop control theory, analog circuit design, and digital circuit design is required. A comprehension of each of the components and its place in the system is essential. Fortunately not all of this knowledge is required at once. The books listed at the end of this chapter can each provide an in-depth insight into areas beyond the present scope [1–5]. This discussion will begin with the basic concepts and rapidly expand these ideas into practical considerations.

## 12.2 PLL DESIGN BACKGROUND

Discussion of the PLL draws heavily on many other areas of analysis, which includes an understanding of the principles of closed loop control theory. From control theory comes the concept of negative feedback to tailor the performance of closed loop systems. Response time, transient performances, bandwidth, damping ratio, and phase margin are used to describe PLL operation. The type and order of a closed loop system define the complexity and response to a stimulus.

In most PLL's, at least two of the components, the voltage-controlled oscillator (VCO) and phase detector, are high-frequency components. There may also be amplifiers, mixers, frequency multipliers, and other oscillators. To use these items, a familiarity with RF design practices and terminology is important.

Frequency multiplication may require digital integrated circuits (ICs) within the PLL. These ICs require digital control words to set the desired frequency.

Many integrated circuits are presently available that combine many of the PLL functions on a single chip. Most of the interface control is digital.

Analog circuit design is perhaps the most demanding of the circuit areas within a PLL. Op-amps are used in many of the filtering circuits used within a loop. Inverting and noninverting circuits are required for loop filters and search circuits. Integrators, dc amplifiers, Schmitt triggers, and offset circuits are used to set the loop operation. Resistor/capacitor circuits provide phase shift for stability. The oscillator is an intrinsic part of a PLL, and its design in itself is a specialized and technically challenging area.

## 12.3 PLL APPLICATIONS

A phase lock loop is a frequency domain device that can be used to multiply, divide, or filter different frequencies. Consider a space probe rapidly moving away from the earth. To recover data from the probe, the transmitter frequency must be known. The signal is very weak because of the distance, and the low signal-to-noise ratio requires a very small receiver filter bandwidth to recover the data. However, because of the relative motion, there is a significant and changing Doppler shift to the transmit frequency. The system requires a filter that may be only a few Hertz wide operating at a varying frequency that is centered at several GHz.

An electronic phase lock loop is one form of a closed loop system. The cruise control is another. A switching power supply, a camera's light meter, a radio's automatic gain control, the temperature control in a building, a car's emission system controls, and a Touch-Tone dialing system are examples of closed loop systems. A broadcast receiver changes frequency with a button push or electronically. Each time the station is accurately centered with no manual adjustment required. Physically these PLLs are all very different working at different jobs and in different environments. However, they all must follow the same rules, and the loops must all be stable.

A clear understanding of the concept of feedback control is illustrated by an everyday situation of the simple action of controlling the speed of a car. If the desired speed is 60 mph, then this becomes the reference. Any deviation from this speed is an error. The accelerator pedal is the control element. On level terrain, a constant pressure on the pedal will maintain constant speed. As the car goes up a hill, it will slow down. The difference between the actual speed and the reference value generates an error. This error generates a command to push the accelerator pedal. Pushing the pedal will increase the speed, but there will continue to be a slight error. As the car crests the hill and starts down, the speed will increase. Releasing pedal pressure will slow the acceleration, but an error will remain until a steady state condition is again reached. For this example, the driver's brain is the feedback path. The driver controls the sense of the feedback by knowing when to push and when to release the pedal. By his reaction time, he controls how close to the reference he maintains the car's speed. He may

decide to rapidly change the correction to tightly match the desired speed, or he may choose to compensate slowly so his speed averages out to the correct value. His actions coupled with the car's controls form a system closely analogous to a phase lock loop. Replace the human with an electrical circuit that senses the speed error, include another circuit that tempers the response time, and couple it to the accelerator controls. This is the typical cruise control system. The elements of understanding the operation of a phase lock loop are all available here. The next step is to apply the concepts of this example to the classical elements that make up a PLL.

## 12.4 PLL BASICS

A PLL is a closed loop system used for frequency control. Several building blocks are common to most PLL designs:

1. The phase detector
2. The loop filter
3. The voltage-controlled oscillator

Figure 12.1 illustrates the connection of these blocks to make a complete phase lock loop. The phase detector has two inputs and one output. This block can be realized by a specialized mixer described in Chapter 11 where the IF port passband goes down to dc. If the two input signals are very close in frequency, then the output will contain a term at twice the input frequency and a term that is almost zero frequency. The loop error signal in the PLL is the near-zero term. This signal contains everything that is needed to control the VCO. This error signal goes to the loop filter for amplification and frequency limiting. The loop filter may be as simple as a capacitor and a resistor, or it can be one or more operational amplifiers with many resistors and capacitors. The loop filter is generally a simple circuit that requires an in depth analysis. It is the "glue" that holds all the other parts together and makes the PLL work the way it is supposed to work. The VCO is the control element of this loop. The input is a control voltage from the loop filter, while the output is the required frequency.

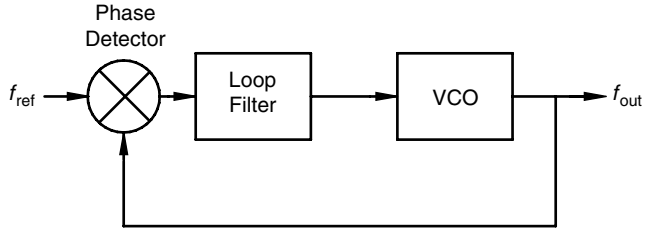



**FIGURE 12.1** Basic phase lock loop.

Frequency is the time rate of change of phase, so phase is the time integral of frequency. Consequently, if the frequency of the VCO is proportional to the tuning voltage, the output phase is proportional to the integral of tuning voltage. The output frequency can range from a few Hertz to many GHz. The VCO output becomes the second input to the phase detector.

When the loop is first turned on, the VCO frequency is not controlled. The loop filter output voltage can be anywhere between the high and low limits set by the power supply. However, the phase detector produces an error voltage that is the difference between the VCO frequency and the reference frequency. Like the cruise control example, this signal tells the loop filter whether the VCO is going too fast or too slow relative to the reference frequency. The loop filter is "smart enough" to know what to do. If the error signal indicates that the VCO frequency is less than the reference, the loop filter adjusts the control voltage to raise the VCO frequency. If the VCO frequency is too high, the loop filter changes the voltage and lowers the VCO frequency. The loop filter design sets how fast this happens. Some loops may be designed for a fast bumpy ride, while others may require a slow smooth ride. When the loop filter has done its job, the VCO frequency will exactly match the reference frequency, and the two inputs will have a constant phase difference. This match in frequency and constant phase difference will be maintained even if the reference frequency changes. With each change, the PLL again goes through the settling out process. If the reference is noisy, the PLL is in a continual state of change, working hard to follow the input. If a PLL were nothing more than a box with the same frequency in and out, it would not be much of an invention. Happily there is a lot more to it than that. Before preceding any farther, the basic idea of the frequency control by a closed loop operation must be clarified.

## 12.5 LOOP DESIGN PRINCIPLES

There are many design rules that go into a successful, stable closed loop design, but many of these details are beyond the scope of his text. However, this discussion will lead to an understanding of how the loop works and how to select parts to customize its operation. From the overall system point of view, the PLL designer must know the overall requirements and be able to translate these requirements into the PLL parameters for the design.

With a PLL, the important top-level parameters are the input and output frequency, the response time, the loop bandwidth and the loop damping ratio. The top-level view concentrates on these system level values. The detailed aspects of the phase detector and VCO must be considered. These are usually high-frequency parts requiring high-frequency design techniques. The design of the loop filter requires an understanding of analog design techniques. Figure 12.2 illustrates the PLL block diagram, which includes a frequency divider in the feedback path. Both frequencies coming into the phase detector must be the locked together at the same frequency. This will force the output frequency,

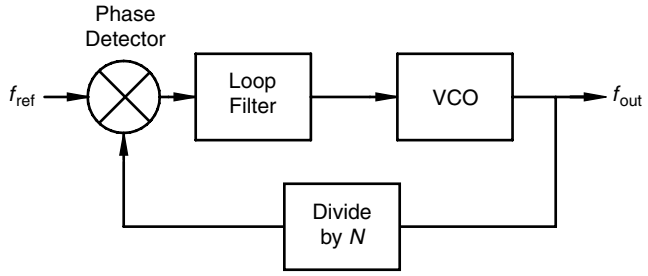



**FIGURE 12.2** Phase lock loop with a frequency divider.

$f_{out}$, to be $N$ times the reference frequency, $f_{ref}$. Familiarity with digital circuit design is required to complete the design. A PLL design is not just a system design nor an RF design nor an analog design, but it is a combination of all these areas. The actual design process can be summarized by three principles.

1. *Know each component.* The components of a PLL, the VCO, phase detector, loop filter, and frequency divider must be thoroughly understood and tested as stand-alone individual components. Testing of the actual parts must show that they each meet the design goals. These tests may extend far beyond the information on the manufacturer's data sheet.
2. *Test the components together.* The individual components must work correctly when connected together in an open loop configuration. The loop filter must put out enough voltage to drive the VCO. The VCO must have enough output to drive the frequency divider. The phase detector output must be large enough for the loop filter to use. An open loop analysis must show the correct phase margin and bandwidth for stability. These issues are resolved with an open loop analysis.
3. *Compare the closed loop configuration to the design goals.* A closed loop analysis should show that the final connection matches the system level design goals. Both test measurement techniques and analysis can be applied for PLL design verification.

## 12.6 PLL COMPONENTS

The basic building blocks, except in exotic applications, are those shown in Fig. 12.2. This section describes in greater depth each of these functions.

### 12.6.1 Phase Detectors

Phase detectors come in many configurations. These include those with logic level inputs, passive and active analog versions, and sampling versions used for high-frequency multiplication. In addition there are phase detectors with automatic

frequency search features to aid in initial frequency acquisition. In its simplest form, a phase detector is a frequency mixer. As described in Chapter 11, when two signals come into the mixer, the output consists primarily in the sum and difference frequencies. The sum frequency is filtered out by the loop filter. The difference frequency, historically called the *beat note*, is typically a few kHz or less in a PLL. If the two input frequencies are exactly the same, the phase detector output is the phase difference between the two inputs. This loop error signal is filtered and used to control the VCO frequency. The two input signals can be represented by sine waves:

$$V_1 = V_a \sin(\omega_1 t + \phi_1) \tag{12.1}$$

$$V_2 = V_b \sin(\omega_2 t + \phi_2) \tag{12.2}$$

The difference frequency term is the error voltage given as

$$V_e = K_m \cdot V_1 \cdot V_2 = \frac{K_m V_a V_b}{2} \cos[(\omega_1 - \omega_2)t + (\phi_1 - \phi_2)] \tag{12.3}$$

where $K_m$ is a constant describing the conversion loss of the mixer. Equation (12.3) gives the time-varying cosine waveform at the beat note frequency. When the two frequencies are identical, the output voltage is a function of the phase difference, $\Delta\phi = \phi_1 - \phi_2$:

$$V_e = \frac{K_m V_a V_b}{2} \cos(\Delta\phi) \tag{12.4}$$

This is maximum when $\Delta\phi = 0^\circ$, a minimum when $\Delta\phi = 180^\circ$, and zero when $\Delta\phi = 90^\circ$ or $270^\circ$ (Fig. 12.3).

When modeling a PLL in the frequency domain, the phase detector can be modeled as

$$K_{pd} \frac{a}{a + s} \tag{12.5}$$

which at low frequencies is simply the slope of the voltage with relation to the phase curve. The units for $K_{pd}$ is volts/rad:

$$K_{pd} = \frac{dV_e}{d\Delta\phi} = -\frac{K_m V_a V_b}{2} \sin(\Delta\phi) \tag{12.6}$$

Thus $K_{pd} = 0$ when $\Delta\phi = 0^\circ$ or $180^\circ$ and a maximum absolute value at $\Delta\phi = 90^\circ$ or $270^\circ$.

### 12.6.2 Voltage-Controlled Oscillator

The voltage-controlled oscillator is the control element for a PLL in which the output frequency changes monotonically with the tuning voltage. A linear

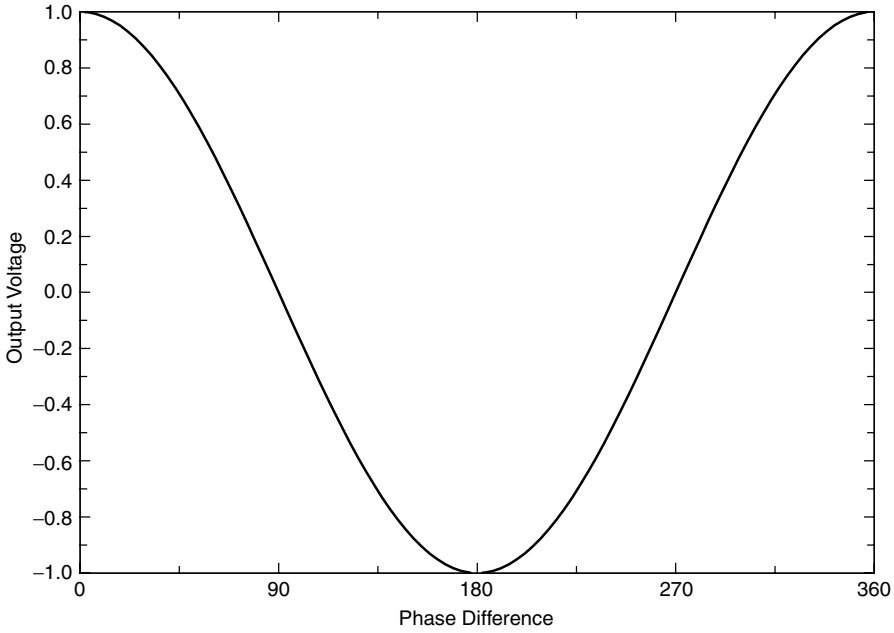



**FIGURE 12.3** Phase detector voltage output as function of phase difference.

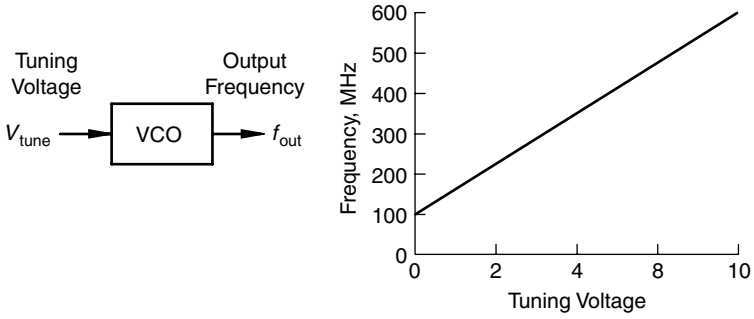



**FIGURE 12.4** Voltage-controlled oscillator tuning.

frequency versus tuning voltage is an adequate model for understanding its operation (Fig. 12.4):

$$\omega_{\text{out}} = K_{\text{vco}} \cdot V_{\text{tune}} + \omega_0 \tag{12.7}$$

In a PLL the ideal VCO output phase may be expressed as

$$\phi(t) = \omega_0 t + \int_0^t K_{\text{vco}} V_{\text{tune}} dt + \phi_0 \tag{12.8}$$

where $\omega_0$ is the free-running VCO frequency when the tuning voltage is zero and $K_{vco}$ is the tuning rate with the dimension of rad/s-volt.

The error voltage from the phase detector first steers the frequency of the VCO to exactly match the reference frequency, and then holds it there with a constant phase difference. It is modeled as having a low-frequency gain $K_{vco}$ and one or more poles of the following form:

$$\frac{K_{vco}}{s(s+a)} \tag{12.9}$$

### 12.6.3 Loop Filters

A loop filter is a low-frequency circuit that filters the phase detector error voltage with which it controls the VCO frequency. While it can be active or passive, it is usually analog and very simple. In extreme cases it might be an entire microprocessor. This discussion will be limited to analog loop filters, such as the representative topologies shown in Fig. 12.5. Figure 12.5*a* shows an op-amp integrator with nearly infinite dc gain. This is the loop filter often associated with the type 2 PLL. The order and type of a PLL is defined in Section 12.9. Figure 12.5*b* shows an op-amp loop filter with a finite gain and is associated with a type 1 PLL. Figure 12.5*c* is a passive filter used with a phase detector whose output is current rather than voltage. This type of detector is frequently found in synthesizer ICs and is associated with a type 2 PLL. While the loop filter is a simple circuit, its characteristic is important in determining the final

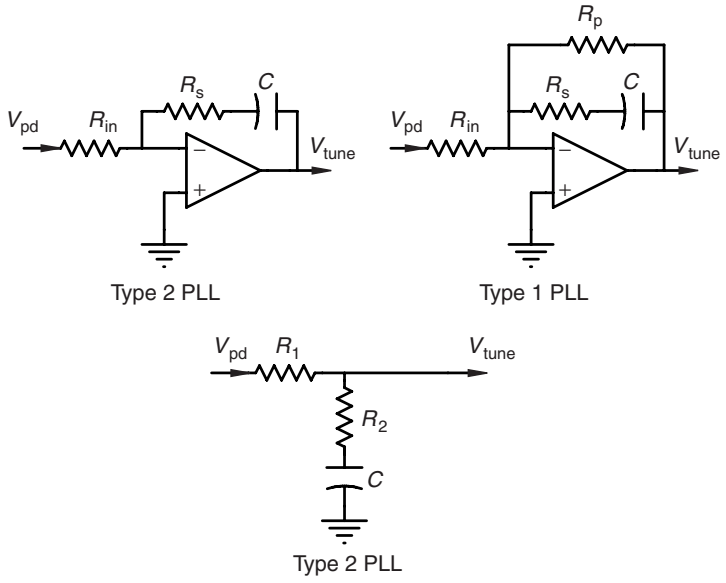



**FIGURE 12.5** Loop filter used in a (*a*) type-2 PLL, (*b*) type-1 PLL, and (*c*) type-2 PLL.

closed loop operation. The wrong design will make the loop unstable causing oscillation or so slow that it is unusable. The loop filters shown in Fig. 12.5 are modeled by

$$F(s) = \frac{s + a}{s^n(s + b)(s + c)} \tag{12.10}$$

where $a$ is a zero and $b$ and $c$ are poles.

### 12.6.4 Frequency Dividers

When the output frequency must be a multiple of the input frequency, frequency dividers may be included in a PLL. Most dividers are a digital circuit, although analog techniques dating from 1939 are available for very high frequency devision. With the availability of complete synthesizers on a single IC, fewer standalone divider circuits are on the market. Most dividers have a division ratio equal to a binary number or switchable from a binary to a binary +1 (e.g., divide by 64 or 65). The upper limit on the input frequency is about 3 GHz, although only a few ICs will go that high. Divide by four circuits has been demonstrated with inputs above 14 GHz, but this is a very specialized device not required by most PLLs. For a linear analysis when the loop bandwidth is much less than the reference frequency, dividers are modeled as a gain element with a value $= 1/N$.

## 12.7 LINEAR ANALYSIS OF THE PLL [1]†

From the perspective of the time domain, the control voltage for the VCO is

$$V_{\text{tune}}(t) = V_{\text{tune}-0}(t) + \int_0^t v_e(t) f(t - \mu) d\mu \tag{12.11}$$

where $f(t)$ is the impulse response of the filter. Now the Laplace transform of $f(t)$ is

$$F(s) = \int_0^\infty f(t) e^{-st} dt, \qquad t > 0 \tag{12.12}$$

and the inverse transform can be obtained in principle by the contour integral shown below:

$$f(t) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} F(s) e^{st} ds, \qquad \Re s > 0 \tag{12.13}$$

† This material is based on A. J. Viterbi, *Principles of Coherent Communication*, 1966, by permission of the McGraw-Hill Company.

Then the VCO frequency is

$$\frac{d\phi_2(t)}{dt} = \omega_0 + \frac{K_m V_a V_b}{2} \int_0^t f(t-\mu)\cos\Delta\phi(\mu) \tag{12.14}$$

where $\Delta\phi(t) = \phi_1(t) - \phi_2(t)$. Consequently, a the general equation describing the phase error is

$$\frac{d\Delta\phi}{dt} \triangleq \frac{d\phi_1}{dt} - \omega_0 - \frac{K_m V_a V_b}{2} \int_0^t f(t-\mu)\cos\Delta\phi(\mu)d\mu \tag{12.15}$$

For a given input phase $\phi_1$, the solution of this equation describes the exact operation of the PLL. However, to avoid carrying along $\omega_0$, a new phase variable may be defined:

$$\psi_1(t) \triangleq \phi_1(t) - \omega_0 t \tag{12.16}$$

$$\psi_2(t) \triangleq \phi_2(t) - \omega_0 t \tag{12.17}$$

The equation for the phase error is now given without $\omega_0$:

$$\frac{d\Delta\phi}{dt} = \frac{d\psi_1}{dt} - \frac{K_m V_a V_b}{2} \int_0^t f(t-\mu)\cos\Delta\phi(\mu)d\mu \tag{12.18}$$

This suggests an alternate representation for the phase lock loop shown in Fig. 12.6. In this representation the multiplier is replaced by a subtracter and a cosinusoidal nonlinearity, while the VCO is replaced by an integrator.

When the phase error $\Delta\phi$ deviates from 90° by a small amount, $\cos(\Delta\phi + 90°) \approx \Delta\phi$. Then Eq. (12.18) becomes

$$\frac{d\psi_1}{dt} = \frac{d\Delta\phi}{dt} + K\int_0^t f(t-\mu)\Delta\phi(\mu)d\mu \tag{12.19}$$

where

$$K = \frac{K_m V_a V_b}{2} \tag{12.20}$$

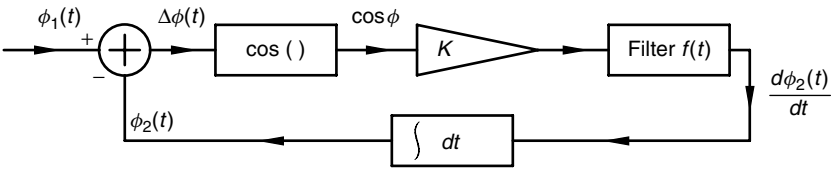



**FIGURE 12.6** Time domain nonlinear phase lock loop.

If the Laplace transform of $\psi_1(t)$ is represented by $\tilde{\psi}(s)$ and the Laplace transform of $\Delta\phi(t)$ is represented by $\Delta\tilde{\phi}(s)$, then the Laplace transform of Eq. (12.19) is

$$s\Delta\tilde{\phi}(s) + KF(s)\Delta\tilde{\phi}(s) = s\tilde{\psi}_1(s) \tag{12.21}$$

This linear frequency domain equation for the PLL can be represented as shown in Fig. 12.7.

The solution for the phase error gives

$$\Delta\tilde{\phi}(s) = \frac{\tilde{\psi}_1(s)}{1 + KF(s)/s} \tag{12.22}$$

so that the phase shift at the output of the PLL is

$$\begin{aligned} \tilde{\psi}_2(s) &= \tilde{\psi}_1(s) - \Delta\tilde{\phi}(s) \\ \frac{\tilde{\psi}_2(s)}{\tilde{\psi}_1(s)} &\triangleq P(s) = \frac{G(s)}{1 + G(s)} \end{aligned} \tag{12.23}$$

where $G(s) = V_a V_b K_m/2F(s)/s$. The phase error can in turn be written in terms of this transfer function:

$$\Delta\tilde{\phi}(s) = \tilde{\psi}_1(s) - \tilde{\psi}_2(s) = [1 - P(s)]\tilde{\psi}_1(s) \tag{12.24}$$

If, for example, the phase of the incoming signal is $\phi_1(t) = \omega t + \phi_0$ and the PLL has no filter, $F(s) = 1$. Readjusting the phase reference as was done in Eqs. (12.16) and (12.17) gives

$$\psi_1 = \phi_1 - \omega_0 t \tag{12.25}$$

$$= (\omega - \omega_0)t + \phi_0 \tag{12.26}$$

In the frequency domain this becomes

$$\tilde{\psi}_1(s) = \frac{\omega - \omega_0}{s^2} + \frac{\phi_0}{s} \tag{12.27}$$

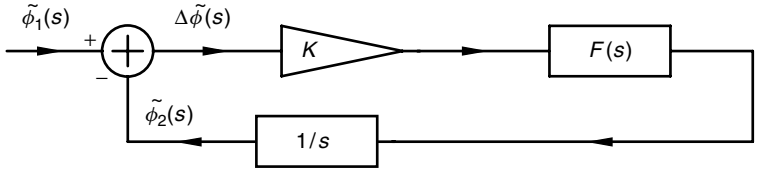



**FIGURE 12.7** Frequency domain linear phase lock loop.

The phase error is found from Eq. (12.22):

$$\Delta\tilde{\phi}(s) = \frac{1}{K/s}\left[\frac{\omega - \omega_0}{s^2} + \frac{\phi_0}{s}\right] \tag{12.28}$$

The inverse transform in this case is straightforward and gives the phase error in the time domain:

$$\Delta\phi(t) = \frac{\omega - \omega_0}{K}(1 - e^{-Kt}) + \phi_0 e^{-Kt} \tag{12.29}$$

The steady state phase error is found by allowing $t \to \infty$:

$$\Delta\phi(t = \infty) = \frac{\omega - \omega_0}{K} \tag{12.30}$$

Clearly, the phase will change when the incoming frequency changes, so that phase lock is not achieved.

The insertion of a low-pass filter into the PLL will produce lock. An active filter such as that shown in Fig. 12.8 is recognized as basically a noninverting amplifier. The inverting amplifier has right half-plane poles and is therefore unstable. For the noninverting case the voltage gain can be found by writing node equations at the input nodes of the operational amplifier:

$$F(s) = \frac{V_o}{V_i} = 1 + \frac{R_2 + 1/sC}{R_1} = \left(1 + \frac{R_2}{R_1}\right) + \frac{1}{sCR_1} \tag{12.31}$$

The phase transfer factor is found from Eq. (12.23):

$$P(s) = \frac{\tilde{\psi}_2(s)}{\tilde{\psi}_1(s)} = \frac{K[(1 + R_2/R_1)R_1Cs + 1]}{s^2CR_1 + CR_1K(1 + R_2/R_1)s + K} \tag{12.32}$$

Thus, using Eq. (12.24), the phase error is easily obtained:

$$\Delta\tilde{\phi}(s) = [1 - P(s)]\psi_1(s) \tag{12.33}$$

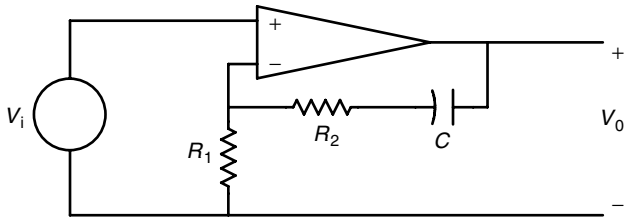



**FIGURE 12.8** Possible active low-pass filter for the PLL.

$$\Delta\tilde{\phi}(s) = \frac{(\omega - \omega_0)CR_1 + \phi_0 CR_1 s}{s^2 CR_1 + CR_1 K(1 + R_2/R_1)s + K} \tag{12.34}$$

Rather than find the inverse transform this time, the final value theorem may be used to find the steady state phase error:

$$\lim_{t\to\infty} \Delta\phi(t) = \lim_{s\to 0} s\Delta\tilde{\phi}(s) = 0 \tag{12.35}$$

In this case the phase error is independent of frequency and in the steady state is zero.

## 12.8 LOCKING A PHASE LOCK LOOP

The previous sections examined each of the elements in a PLL. As an example of how these parts go together, consider a simple loop with no frequency divider. Also assume that initially the loop is not locked and that the reference frequency is 100 MHz. A VCO tuning voltage of 5 volts is required to make the VCO frequency be 100 MHz. The phase detector can produce a cosine wave beat note of 1 volt peak to peak.

To simplify the design, a type 1 loop filter will be used. This is an inverting op-amp circuit with a gain of 100 at low frequency and a gain of 0.1 at high frequency (Fig. 12.9). With the loop unlocked, the VCO frequency could be anywhere within its operating limits. Assume that it is operating at 101 MHz, so that there is a 1 MHz beat note at the phase detector output when the reference frequency is first applied. This beat note frequency is high enough to only be amplified with a gain of 0.1 by the loop filter. The VCO tuning voltage will be modulated by the phase detector output of 0.1 volt peak to peak, but this voltage will not cause any significant change in the VCO frequency.

With the VCO frequency too far away from the reference frequency, there is not enough gain in the loop to bring the loop into lock. However, if the VCO

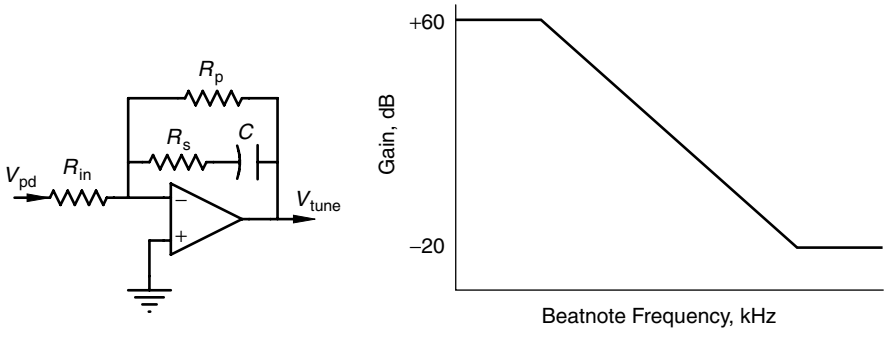



**FIGURE 12.9** (*a*) Type 1 loop filter and (*b*) its frequency response.

frequency is 100.1 MHz when the reference frequency was applied, the beat note frequency would be 100 kHz. That is well within the high-gain frequency range of the loop filter for this design. The amplified beat note voltage modulates the VCO frequency. As the VCO frequency swings closer to the reference frequency, the beat note frequency gets even lower, it enters an even higher-gain region of the loop filter. This action accelerates the VCO frequency change until it crosses the reference frequency. At this point the beat note frequency is zero. The PLL has been designed as a stable closed loop system, and the VCO is at the same frequency as the reference. The transient phase detector output voltage and the VCO tuning voltage are shown in Figs. 12.10 and 12.11, respectively. The input voltage to the VCO is 5 volts when the PLL is at frequency lock. Since the loop filter has a dc inverting gain of 100, the voltage at the phase detector output is

$$V_e = \frac{4}{-100} = -50 \text{ mV} \tag{12.36}$$

The maximum voltage it could reach is 1.0 volt, so from Eq. (12.4) the phase difference is $\Delta\phi = \arccos(V_e/0.5 \cdot 1.0) = 95.7^\circ$. The loop filter will keep the VCO at 100 MHz and maintain a 95.7° phase difference between the two phase detector inputs.

An oscillator accumulates 360° of phase rotation in each cycle. If the frequency increases it will accumulate more phase rotation in a given period of time. If the VCO tries to drift higher in frequency, it will quickly accumulate more phase rotation. The phase detector output voltage will go up, and the loop filter will amplify this change, which will lower the VCO control voltage. The VCO output

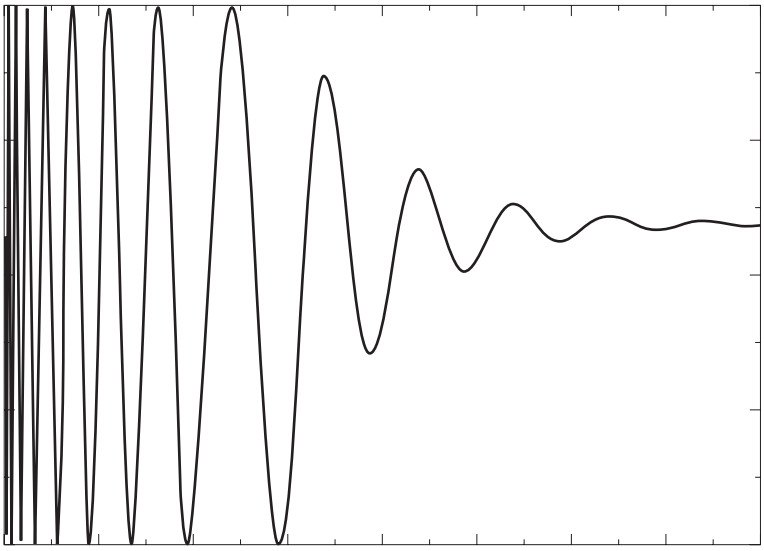


**FIGURE 12.10** Phase detector voltage as the PLL pulls into lock.

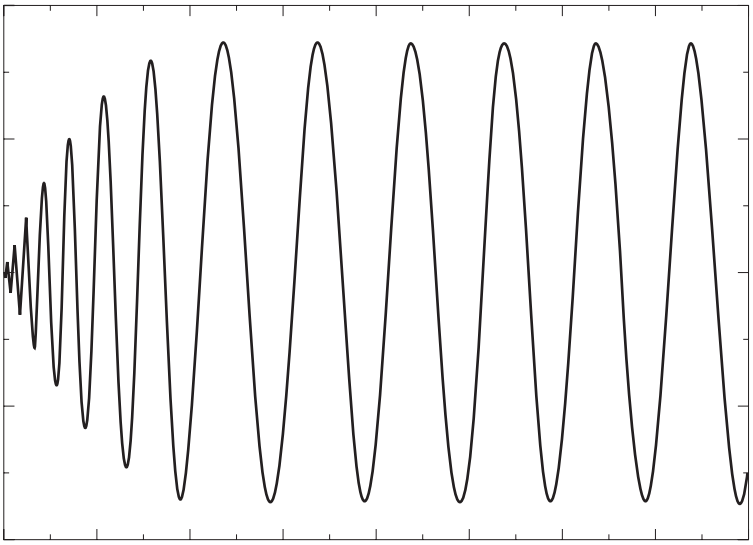


**FIGURE 12.11** VCO tuning voltage as PLL pulls into lock.

frequency will drop and return to 100 MHz. The situation is similar for the VCO trying to move lower in frequency. This is the effect of the negative feedback within the loop. The battle for control goes on continuously. Small changes in the VCO due to temperature, noise, or even gravity cause small frequency changes. The PLL will not tolerate errors due to frequency or phase changes. When an error voltage develops at the phase detector output, the loop filter will amplify it, and the VCO frequency and phase will return to the correct value. The corrective action of the loop will make whatever adjustments are required to hold the phase and frequency constant.

It is important to understand that the initial beat note frequency must be well within the loop filter's bandwidth to achieve lock without frequency aiding. In any loop the first event is to bring the VCO frequency in line with the reference frequency. Once the circuit is in lock, a steady state phase relationship that satisfies the loop feedback and dc requirements is found and that will hold its lock frequency.

## 12.9 LOOP TYPES

The PLL is a closed loop system controlled by negative feedback. The closed loop gain P(s) is described by

$$H(s) = \frac{G(s)}{1 + G(s)/N} \tag{12.37}$$

where $G(s)$ is the open loop or forward gain, and $G(s)/N$ is called the *loop gain*. For this discussion, the forward gain G(s) is the product of the phase detector gain, the loop filter gain, and the VCO gain. The frequency divide ratio is $N$.

The number of pure integrators (or number of poles at the frequency origin) in the denominator of Eq. (12.37) determines the type of the system. This can be produced by an op-amp integrator with near-infinite dc gain. Obviously this cannot be produced with a passive filter where the maximum gain is 1. A VCO is a pure phase integrator that will contribute one pole to the type determination. Therefore a PLL will be at least type 1. A loop filter with a finite dc gain will not increase the type number. A loop filter with an integrator will increase the type to 2.

The order of the PLL is the degree of the denominator polynomial of Eq. (12.37). The loop filter op-amp has at least two significant breakpoints: one at a frequency between 1 and 100 kHz and a second above 10 MHz. The VCO has frequency roll-offs in its modulation performance. A low-pass filter may be included in the phase detector output to further reduce the unwanted high frequencies.

In the previous example using a type 1 loop, the only pure integrator is the VCO, so there is only one pole at dc. The loop filter has a dc inverting gain of 100. If the VCO gain is 1 MHz/volt and the reference frequency is changed to 103 MHz, the VCO tuning voltage will now be 8 volts. With a gain of $-100$, the phase detector voltage must be $V_e = 8/(-100) = -80$ mV.

This represents an angular difference of $\Delta\phi = \arccos(V_e/0.5 \cdot 1.0) = 99.7^\circ$ in contrast to $95.7^\circ$ found earlier when the reference frequency was 100 MHz. The phase difference between the VCO and reference frequency was 95.7 degrees. If the reference frequency continues to change, the VCO frequency will change to match it, which in turn will change the phase detector output voltage. As the reference frequency changes in a type 1 loop, the phase difference changes. This is an important characteristic that is sometimes desirable and other times unacceptable.

If the dc gain of the loop filter is increased to 1000, the phase detector output voltage for a 100 MHz lock is only $-5$ mV. For phase lock at 103 MHz, the phase detector output voltage is $-8$ mV. These values represent phase differences, $\Delta\phi$, of $90.57^\circ$ and $90.92^\circ$, respectively. If the dc gain is further increased, the change of $\Delta\phi$ with frequency will further decrease. If the gain is increased to the limit, the dc feedback resistor, $R_p$, will approach an open circuit, and the loop filter dc gain will increase to infinity. The loop filter in Fig. 12.5*b* is transformed to that shown in Fig. 12.5*a*.

This loop filter is now a pure integrator. The total number of integrators for the PLL with this loop filter is two: one for the VCO and one for the loop filter. This loop filter used in a PLL creates a type-2 loop. Among the features of this loop is the constant phase shift between the VCO and reference frequency that is maintained with a change in frequency.

Type-1 and type-2 loops constitute the majority of applications. Type-3 and higher loops are required to solve frequency change problems in unusual

situations. For example, a ground-launched missile must track an orbiting satellite during its own launch and orbital insertion. During the launch phase, the rocket is consuming fuel and thus reducing its mass. With a constant force, its acceleration will be increasing at an increasing rate. As the satellite comes overhead its transmit frequency is shifted due to the relative motion with the rocket. This shift is constantly changing at an increasing rate. Then the booster separates from the rocket, and the force goes to zero during coast. To track the satellite frequency with no phase error requires a PLL type of at least four. Most high-type loops are used to solve complicated motional problems. This discussion will not cover the design details of loops of a type higher than two.

## 12.10 NEGATIVE FEEDBACK IN A PLL

A frequency change that generates a change in the phase of a stable negative feedback loop generates a correction for the phase error. In the previous example the type 1 loop filter was described as having a dc inverting gain of 100. The VCO requires 5 volts to produce a 100 MHz output. An open loop connection of the PLL components will demonstrate what is called the "sense" of the loop. For open loop testing the VCO is connected to a manually adjustable power supply. With the power supply set at 5 volts, there will be a low-frequency beat note observed at the phase detector output. If the voltage is changed to either 4 or 6 volts, the beat note will be 1 MHz. With a mixer as the loop phase detector, the beat note will be a cosine wave at the difference frequency, 1 MHz. The frequency of the VCO cannot be determined from looking at the beat note. The beat note shows the frequency difference between the two signals, but it does not tell which signal is the higher or lower frequency. A complete description of the difference frequency between the VCO and reference requires both a direction and a magnitude. With the loop out of lock, this type of phase detector can only determine magnitude, $|\Delta\phi|$. The VCO frequency must be forced close enough to the reference frequency for the beat note to be inside the loop bandwidth for a PLL with this type of phase detector to pull into phase lock.

In the previous example the VCO frequency increased as the tuning voltage increased. Many VCOs have the opposite characteristic; that is, the frequency decreases with increased tuning voltage. This difference does not change the stability or operation of the closed loop if a mixer is used as the phase detector. If the loop locked up at 90° difference between the two inputs with the positive slope VCO, it will lock up at 270° with the negative slope VCO. The phase detector output in either case will be correct to speed up or slow down the VCO to match the reference input frequency and phase.

Most synthesizer ICs and PLLs using frequency dividers or logic ICs have a different type of phase detector. Using flip-flops to count the input edges, these phase detectors produce an error voltage that has not only a magnitude but also a sense of the direction between the two inputs. The output is a series of voltage or current pulses. The loop filter averages these pulses to form the control voltage

for the VCO. A pulse duty cycle above 50% indicates that the VCO frequency is higher than the reference frequency, and a duty cycle of less than 50% indicates that the VCO frequency is lower. If the VCO is running faster than the reference frequency, the control voltage will force it toward the correct value. If the VCO is running too low, the error voltage will drive the frequency higher. This type of phase detector can drive a PLL into lock even when the VCO and reference frequencies are a great distance apart, far outside the loop bandwidth. The typical IC synthesizer will have a pin available to reverse the sense of the error voltage to accommodate VCOs of either positive or negative tuning slope.

## 12.11 PLL DESIGN EQUATIONS

A phase lock loop design requires the basic understanding of the locking mechanism as previously discussed. However, the values for the loop filter and other components must be carefully selected to assemble a stable loop. These values can be both analyzed and synthesized using basic closed loop equations and linear algebra.

The normal phase lock loop model includes a phase detector, a loop filter, a VCO and a frequency divider connected as shown in Fig. 12.12. Each block is described by a gain value that may be a constant or a function of frequency. The frequency response of the closed loop is typically displayed as a Bode plot with a minimum frequency of 1 Hz and a maximum frequency between 10 kHz and 10 MHz. This describes the filtering bandwidth and in turn the transient response of the PLL to the input voltage spectrum, $V_{\text{in}}$.

### 12.11.1 Inverting Loop Filter

The analysis proceeds by writing the voltage equations at points $V_{\text{e}}$ and $V_{\text{o}}$. Combining these equations produces the well-known equation for closed loop gain of a system with negative feedback. When $K_{\text{pd}}$ is the phase detector gain, $F(s)$, is the noninverting loop filter function, and $K_{\text{vco}}/s$ is the VCO gain, the

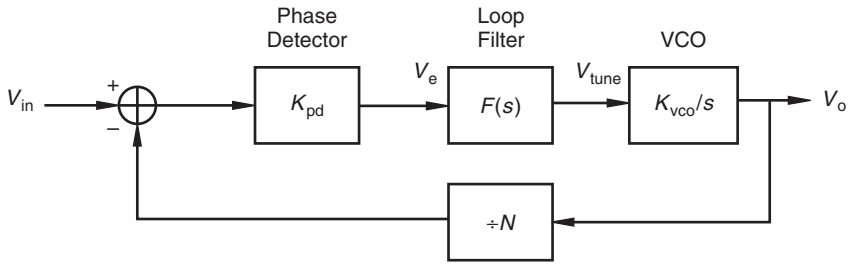



**FIGURE 12.12** Frequency domain closed loop model for a PLL.

error voltage is easily found:

$$V_e = \left(V_{in} - \frac{V_o}{N}\right) K_{pd} \tag{12.38}$$

$$V_o = \left(V_{in} - \frac{V_o}{N}\right) \cdot \left(K_{pd} F(s) \frac{K_{vco}}{s}\right) \tag{12.39}$$

This is solved for the voltage transfer function:

$$H(s) \triangleq \frac{V_o}{V_{in}} = \frac{K_{pd} F(s) K_{vco}/s}{1 + \{[K_{pd} F(s) K_{vco}/s]/N\}}$$

$$= \frac{G(s)}{1 + G(s)/N} \tag{12.40}$$

A similar analysis for a PLL with an inverting loop filter gives a similar equation except for a sign reversal. Positive feedback is used at the summer block since the sign reversal has already occurred in the inverting amplifier:

$$H(s) = \frac{G(s)}{1 - G(s)/N} \tag{12.41}$$

In either case, when the gain, $G(s)$, is large, $|H(s)| \approx N$.

These equations for closed loop gain can now be used to determine the loop filter values required for a desired bandwidth and damping ratio. The procedure initially assumes a second-order type-1 loop, since that is most frequently used. A type 2 loop can then be easily derived. The transfer function for the filter in Fig. 12.5*b* is

$$F(s) = -\frac{R_p \| [R_s + (1/Cs)]}{R_{in}}$$

$$= -\frac{(R_p R_s / R_{in}) sC + (R_p / R_{in})}{(R_p + R_s) sC + 1} \tag{12.42}$$

The open loop gain is

$$G(s) = \frac{F(s) K_{pd} K_{vco}}{s} \tag{12.43}$$

For a type 1 PLL, $R_p \to \infty$ and

$$F(s) = -\frac{R_s sC + 1}{R_{in} sC} \tag{12.44}$$

When Eqs. (12.43) and (12.44) are substituted into the expression for the gain of the closed loop PLL, Eq. (12.41), the result is clearly of second order in

the denominator:

$$H(s) = -\frac{[K_{pd}K_{vco}(R_p + sCR_s)R_p/R_{in}]/C(R_p + R_s)}{s^2 + s[1/C(R_p + R_s) + (K_{pd}K_{vco}/NR_{in})(R_pR_s/R_p + R_s)] + [K_{pd}K_{vco}R_p/NR_{in}C(R_p + R_s)]} \tag{12.45}$$

The denominator can be converted to the familiar form used in control theory, $s^2 + 2\zeta\omega_n s + \omega_n^2$ where $\zeta$ is the damping factor and $\omega_n$ is the natural frequency of the system. In this case,

$$\omega_n = \sqrt{\frac{K_{pd}K_{vco}R_p}{NR_{in}C(R_p + R_s)}} \tag{12.46}$$

$$\zeta = \frac{(1/C) + (K_{pd}K_{vco}R_pR_s/NR_{in})}{2\omega_n(R_p + R_s)} \tag{12.47}$$

For the type-2 PLL when $R_p \to \infty$,

$$\omega_n = \sqrt{\frac{K_{pd}K_{vco}}{NCR_{in}}} \tag{12.48}$$

and

$$\zeta = \frac{K_{pd}K_{vco}R_s}{NR_{in}2\omega_n} \tag{12.49}$$

The design specification for a PLL is typically given in terms of a damping ratio and natural frequency. The design task is to determine circuit values that will meet these specifications. For ease of writing, define

$$K_t \triangleq \frac{K_{pd}K_{vco}}{N} \tag{12.50}$$

Furthermore, the filter response at dc is

$$F_{dc} = -\frac{R_p}{R_{in}} \tag{12.51}$$

Thus Eq. (12.46) can be rearranged to give

$$R_p + R_s = -\frac{K_tF_{dc}}{C\omega_n^2} \tag{12.52}$$

and this substituted into Eq. (12.47) to give

$$R_p + R_s = -\frac{K_tF_{dc}}{C\omega_n^2} = \frac{1}{2C\omega_n\zeta} - \frac{K_tF_{dc}R_s}{2\omega_n\zeta} \tag{12.53}$$

Using Eq. (12.52) to replace $R_s$ above,

$$-\frac{K_t F_{dc}}{C\omega_n^2} = \frac{1}{2C\omega_n\zeta} - \frac{K_t F_{dc}}{2\omega_n\zeta}\left(\frac{-F_{dc}K_t}{C\omega_n^2} - R_p\right) \tag{12.54}$$

If, in addition to the damping ratio and the natural frequency, values for $C$ and the dc gain are chosen, then the required resistance values can be found as summarized below. Solution of Eq. (12.54) gives the value for $R_p$.

$$\begin{aligned} R_p &= \frac{1}{K_{pd}K_{vco}C}\left[\frac{2\zeta K_{pd}K_{vco}}{\omega_n} + \frac{K_{pd}K_{vco}F_{dc}}{N\omega_n^2} - \frac{N}{F_{dc}}\right] \\ &= \frac{1}{K_t C}\left[\frac{2\zeta K_t}{\omega_n} + \frac{K_t F_{dc}}{N\omega_n^2} - \frac{1}{F_{dc}}\right] \end{aligned} \tag{12.55}$$

$$\begin{aligned} R_s &= -\frac{K_{pd}K_{vco}F_{dc}}{NC\omega_n^2} - R_p \\ &= -\frac{K_t F_{dc}}{C\omega_n^2} - R_p \end{aligned} \tag{12.56}$$

$$R_{in} = -\frac{R_p}{F_{dc}} \tag{12.57}$$

The type 2 PLL equations are found by allowing $R_p \to \infty$.

Figure 12.13 illustrates the expected PLL bandwidth verses frequency for several values of damping ratio. These results are calculated for a second-order loop with a natural frequency of 1 Hz. The results can be easily scaled for loops requiring higher natural frequencies. When $\zeta < 1$, the PLL is under damped and peaking occurs. The response of such a loop to a disturbance will be a damped oscillation that finally converges to the final answer. When $\zeta > 1$, the PLL is overdamped. The $-3$ dB gain frequency for a damping of 1.0 is 2.4 Hz. If a $-3$ dB frequency of 50 kHz were required with a damping of 1.0, then a natural frequency of 20.833 kHz would be chosen. The requirements of the PLL design and the available parts will determine the best choice for the natural frequency and damping ratio.

A second-order loop can be built either as type 1 or type 2 with either an inverting or noninverting loop filter. The actual loop order may be several orders higher than 2 when all the extraneous poles are considered. A good design procedure initially ignores these poles and assumes ideal VCOs, phase detectors, op-amps, and so on, with which it determines a set of loop filter values based on the second-order model. Subsequently nonideal parts can then be added to the model and the analysis refined as more values become available. Computer modeling is encouraged for this process.

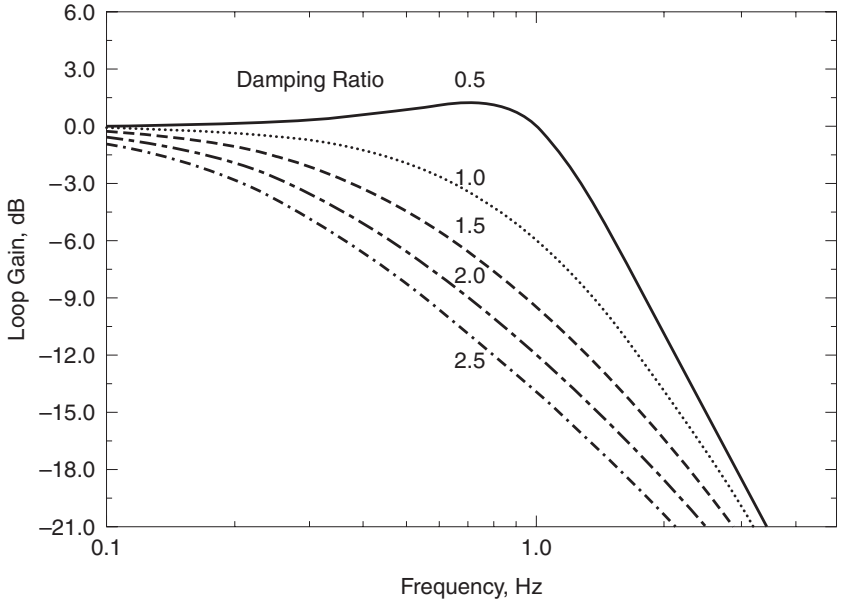



**FIGURE 12.13** PLL response with natural frequency of 1 Hz and various damping ratios.

### 12.11.2 Noninverting Loop Filter

Design equations can be developed for a noninverting loop filter like that shown in Fig. 12.5. The filter transfer function is

$$F(s) = 1 + \frac{R_p \| [R_s + (1/sC)]}{R_{in}}$$
$$= \frac{1 + (R_p/R_{in}) + sC[(R_p R_s/R_{in}) + R_p + R_s]}{1 + sC(R_p + R_s)} \tag{12.58}$$

The closed loop gain is found by substituting Eq. (12.58) into Eq. (12.37) while making use of Eq. (12.50):

$$H(s) = \frac{NK_t\{1 + (R_p/R_s) + sC[(R_p R_s/R_{in}) + R_p + R_s]\}}{s[1 + sC(R_p + R_s)] + \{1 + (R_p/R_{in}) + sC[1 + (R_p R_s/R_{in}) + R_p + R_s]\}K_t} \tag{12.59}$$

$$= \frac{[NK_t/C(R_p + R_s)]\{1 + (R_p/R_s) + sC[(R_p R_s/R_{in}) + R_p + R_s]\}}{s^2 + s\{[1/C(R_p + R_s)] + [R_p R_s K_t/(R_p + R_s)R_{in}] + K_t\} + [(R_{in} + R_p/(R_p + R_s)]K_t/C)} \tag{12.60}$$

From this the loop natural frequency and damping ratio can be identified:

$$\omega_n = \sqrt{\frac{(R_{in} + R_p)K_t}{(R_p + R_s)R_{in}C}} \tag{12.61}$$

$$\zeta = \frac{R_{in} + CR_pR_sK_t + R_{in}(R_p + R_s)CK_t}{2\omega_n C(R_p + R_s)R_{in}} \tag{12.62}$$

The typical synthesis procedure is to design a PLL with a given natural frequency and damping ratio using a specified capacitance, $C$. Solving Eq. (12.61) for $R_p + R_s$ and substituting this into Eq. (12.62) gives an equation in terms of one unknown, $R_p$. First, from Eq. (12.61),

$$R_s = \frac{K_t(R_p + R_{in})}{R_{in}C\omega_n^2} - R_p \tag{12.63}$$

then substitution gives

$$\frac{K_t(R_p + R_{in})}{R_{in}C\omega_n^2} = \frac{1}{2\zeta\omega_n CR_{in}}\left[R_{in} + CR_pK_t\left(\frac{K_t(R_{in} + R_s)}{R_{in}C\omega_n^2} - R_p\right) + CK_tR_{in}\frac{K_t(R_{in} + R_p)}{R_{in}C\omega_n^2}\right] \tag{12.64}$$

This has one unknown, $R_p$, that can be solved by the quadratic formula as follows:

$$0 = R_p^2(K_t^2 - CK_tR_{in}\omega_n^2) + R_p2K_tR_{in}(K_t\zeta - \omega_n) + R_{in}^2(\omega_n^2 + K_t^2 - 2\zeta K_t\omega_n) \tag{12.65}$$

So, if

$$a = K_t^2 - CK_tR_{in}\omega_n^2 \tag{12.66}$$

$$b = 2K_tR_{in}(K_t\zeta - \omega_n) \tag{12.67}$$

$$c = R_{in}^2(\omega_n^2 + K_t^2 - 2\zeta K_t\omega_n) \tag{12.68}$$

then

$$R_p = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{12.69}$$

Again, the value of $R_{in}$ is associated with the voltage gain of the noninverting loop filter:

$$R_{in} = \frac{R_p}{F_{dc} - 1} \tag{12.70}$$

The value for $R_s$ is obtained from Eq. (12.63).

The type 2 PLL parameters with the noninverting loop filter can be found by letting $R_p \to \infty$. Thus the design equations for a given natural frequency, damping ratio, and capacitance are

$$\omega_n^2 = \frac{K_t}{R_{in}C} \tag{12.71}$$

or

$$R_{in} = \frac{K_t}{C\omega_n^2} \tag{12.72}$$

and the damping ratio is

$$2\omega_n\zeta = 0 + \frac{R_s K_t}{R_{in}} + K_t \tag{12.73}$$

or

$$R_s = R_{in}\frac{2\omega_n\zeta - K_t}{K_t} \tag{12.74}$$

The value for $K_t$ is given by Eq. (12.50).

## 12.12 PLL OSCILLATORS

A phase lock loop will have a least two oscillators associated with it: the reference oscillator and the voltage controlled oscillator. The reference is typically a fixed frequency, but in some applications it may change over a wide frequency range. In any event, the VCO must be able to follow it.

The types of oscillators used depend on the design requirements and the frequency range. Low-frequency oscillators use a resistor capacitor combination to set the frequency. A larger charging current in the capacitor results in a higher-oscillation frequency. These oscillators can be used from a few hertz to several megahertz. They can sweep a wide frequency range, but the output is usually noisy and drifts rapidly with temperature. This oscillator is frequently found in simple PLL ICs.

An LC oscillator's frequency is set by a combination of inductors and capacitors in a resonant circuit. This oscillator is useful from about 100 kHz to somewhat above 5 GHz. A high $Q$ network will produce a very clean output with a small tuning range. Frequency control of a voltage-controlled version can be built either by varying a dc voltage in the circuit or, more commonly, by adding a voltage-controlled capacitor, called a *varactor diode*, to the resonant network. A varactor is a low-loss reverse-biased diode whose depletion capacitance has been optimized for large change with changes in bias. Reverse-biased voltages range from about 2 volts to 28 volts. Capacitance values range from fractions of a pF to 500 pF.

Oscillator design at any frequency is a specialized area. The design or purchase of either the reference oscillator or the VCO is an activity that should be completed before the PLL design is attempted. Available power supply voltages will determine the tuning voltage available for the VCO. Simulations can determine an approximate tuning rate for the VCO, from which an appropriate oscillator type can be picked from the various available types and manufacturer's data sheets. The characteristics of both oscillators should be well defined.

## 12.13 PHASE DETECTOR TYPES

Previous sections have introduced both the mixer and flip-flop-based phase detectors. These are the two main configurations that are widely used, although there are many specialized variations of each type. A sampling phase detector is a third type that is frequently used in RF and microwave applications. A careful study of the design requirements will usually point to the correct choice.

### 12.13.1 Mixer Phase Detectors

Mixers with a dc coupled output make an excellent phase detector. At high frequencies, a mixer may be either active (with transistors) or passive (with diodes). The diode versions provide better dc stability and are generally preferred. The diode mixer is the best choice for low-noise designs or when the PLL reference input is a low-level signal buried in noise. When the PLL is locked and running closed loop, both frequencies will be exactly the same. The two inputs are applied to the RF and local oscillator (LO) ports of the mixer, and the beat note output comes from the intermediate frequency (IF) port. The IF port is loaded with a total resistance of 50 to 1000 Ω. Blocking capacitors are used to ac couple the signals into the two input ports, but the IF output port must not be ac coupled. A mixer must be chosen that is appropriate for the frequency range and power level of the inputs. The beat note output is typically 100 mV to 1 volt peak to peak, depending on the mixer type and application. The mixer type phase detector is the best choice when the input signal is pulsed or noncontinuous. The ability of this type of detector to resolve an angular difference is limited to $\pm 90^\circ$. It has no ability to determine which input is the higher frequency. Thus it is not capable of frequency discrimination.

### 12.13.2 Sampling Phase Detectors

Sampling phase detectors (SPD) can be used in a phase lock loop to produce an output frequency that is an integer multiple of the reference frequency. This mixer relies on a device to generate a comb of frequencies at multiples of the reference. The VCO then mixes with the correct spectral line to produce an error signal. A SPD is used in a phase lock loop where the output frequency is an integer multiple of the input frequency. If $f_{\text{out}} = N \cdot f_{\text{in}} + \Delta f$, the SPD output

is a cosine wave of frequency $\Delta f$. If $f_{out}$ is exactly N times $f_{in}$, the SPD output is a dc level proportional to the phase difference between its two inputs. The input frequency, $f_{in}$, is typically between 50 and 200 MHz at a power level of +20 dBm or higher, and $f_{out}$ can be at any harmonic of $f_{in}$ up to 18 GHz. The low-frequency high-power input signal produces a comb of spectral lines inside the SPD where lines up to the 150th multiple are useful.

A sampling phase detector is one of the best choices for a very high frequency PLL where excellent phase noise is a requirement. However, an external circuit is usually required to bring the VCO into lock range. Additionally provisions must be made to ensure that the VCO will be locked to the correct multiple of $f_{in}$.

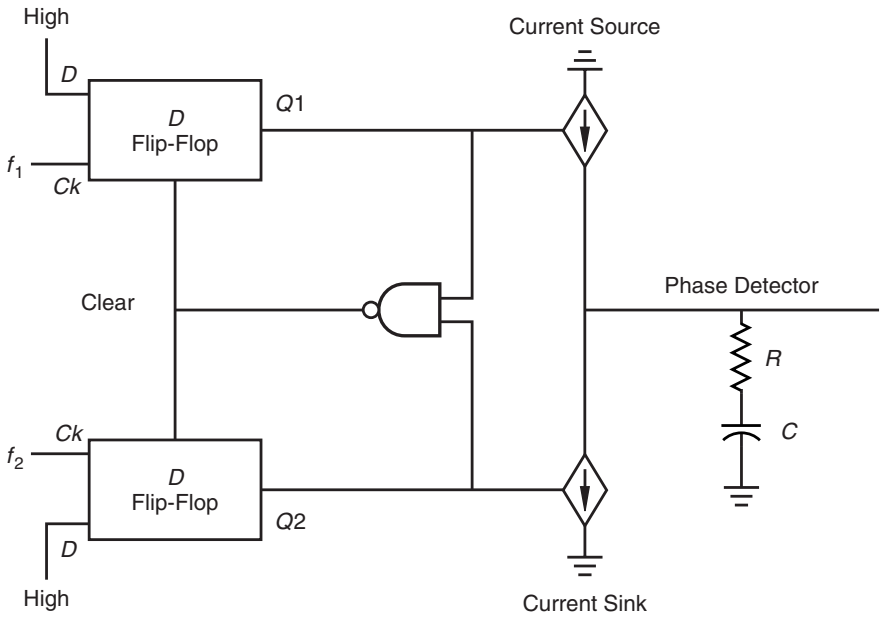
### 12.13.3 Flip-Flop Phase Detector with Frequency Acquisition Aiding

All of the other phase detectors previously discussed have a major drawback. They produce an output equal to the difference between the two input frequencies, either a cosine wave or a triangle wave. However, this output does not have information about whether the VCO is too high in frequency or too low. A PLL using these phase detectors must also include a sweep or search circuit to initially bring the VCO close enough to lock. This can involve a substantial amount of circuitry. The phase detector circuit shown in Fig. 12.14*a* uses positive edge triggered *D* type flip-flops to overcome this problem. The *D* inputs are connected to a logic 1. Figure 12.14*b* illustrates the timing sequence. The signal, $f_1$ positive edge arrives first causing $Q1$ to clock high. Later positive edge of $f_2$ causes $Q2$ to clock high. Two 1's at the NAND gate's input cause its output to go low and clear both $Q1$ and $Q2$. The output at $Q2$ is a pulse whose duty cycle represents the time delay between $f_1$ and $f_2$. The pulse at $Q2$ is very short, being the sum of the propagation times through the flip-flops and gates. Of course, if $f_2$ arrives before $f_1$, then the output pictures are reversed.

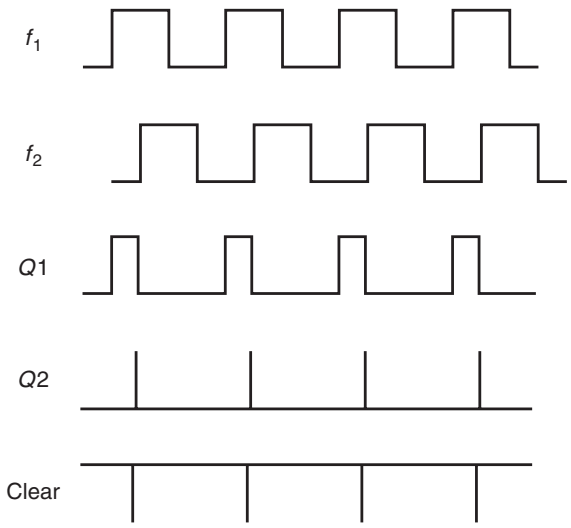
The outputs, $Q1$ and $Q2$, turn on the current sources. These current sources either source or sink current to the capacitor which ramps up or down the phase detector output voltage. The action of this circuit is identical to the op-amp integrator. So the phase detector has added another integrator to the PLL. A PLL using this circuit will be at least a type 2 loop.

The advantage to this circuit is the self-searching capability. If $f_1$ is higher than $f_2$, the output voltage will go to the positive voltage limit. If $f_1$ is lower than $f_2$, the output voltage will go to the negative voltage limit. If $f_1$ equals $f_2$, the output voltage will be proportional to the phase difference. Thus this circuit can sense which input frequency is higher. The output voltage can then be used to drive the VCO in the correct direction to bring the loop into lock. Once the two frequencies are the same, this circuit becomes a phase detector and drives the VCO for no phase error.

This phase detector circuit is used in many present-day frequency synthesizer ICs, where its built-in search capability makes it ideal for a variety of applications. However, there are at least two drawbacks to this circuit that force limits on its usage. The largest problem is the short pulse on one of the flip-flop




(*a*)




(*b*)

**FIGURE 12.14** (*a*) Phase detector using a *D* flip-flop and (*b*) the timing chart.

outputs. With high-speed logic, this pulse is only a few nanoseconds long. If $f_1$ and $f_2$ are high in frequency, this may be a significant part of their period. This pulse dead time due to propagation delays results in a nonlinear phase detector transfer curve. In older versions there were flat spots with zero gain and regions where the gain reversed its slope. The pulse also contributes heavily to the output noise, and it can easily add 20 dB of noise to the PLL output even in the most modern devices. The second problem is that the searching capability can become confused if there is any interruption in either $f_1$ or $f_2$. This circuit should be used in applications where very low phase noise is not required and the inputs are continuous. For this phase detector the gain is $K_{\text{pd}} = 1/4\pi$.

### 12.13.4 Exclusive OR Phase Detector

An exclusive OR gate works as a frequency doubler and phase detector. Figure 12.15 illustrates a typical connection for the phase detector. For correct operation, both inputs $f_1$ and $f_2$ must be at the same frequency, and both must have 50% duty cycles. The XOR output will be a logic level waveform at twice the input frequency. The duty cycle of the output depends on the phase difference between the two inputs. Phase shifts of 90° or 270° produce a 50% output. The RC low-pass filter produces a dc value proportional to the duty cycle. For a 90° or a 270° phase difference, the filter output is one-half the difference between the logic high- and logic low-output voltages.

The XOR gate is the functional equivalent of the balanced mixer. This circuit is useful for PLL applications requiring a high-frequency VCO to be divided down and locked to a low-frequency logic level frequency reference. This phase detector is suitable for low-phase noise applications, but it frequently requires an external search circuit to initially achieve lock. The gain, $K_{\text{pd}}$, of this phase detector is $1/\pi$ volts/rad.

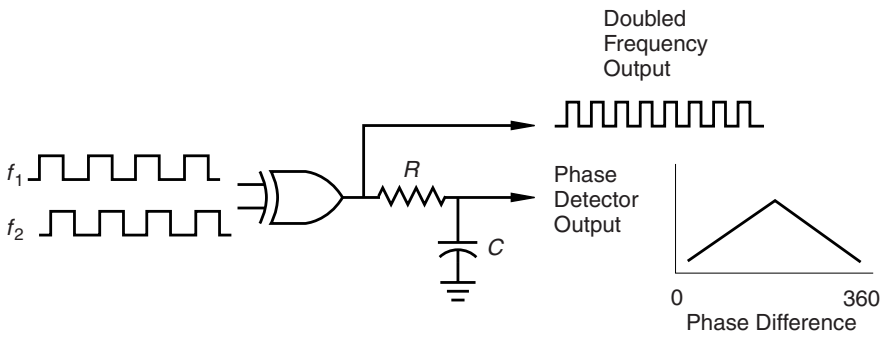



**FIGURE 12.15** Exclusive OR phase detector.

## 12.14 DESIGN EXAMPLES

***Example 1.*** The phase lock loop design shown in Fig. 12.16 requires an output frequency of 1600 MHz and the reference oscillator is 100 MHz. The design approach chosen is to use an inverting type-2 loop filter with a frequency divider and a mixer phase detector. The VCO chosen shows a typical tuning slope of 1 MHz/volt. Measurements of the phase detector output show a cosine wave that is 100 mV peak to peak. A 3 dB bandwidth of 100 kHz is required with a damping ratio of 1.

a. Using a 100 pF capacitor, find the remaining loop filter values.
b. Using a 10 kΩ $R_{in}$, find the remaining loop filter values.

*Solution 1.* From the graph in Fig. 12.13, the 3 dB frequency for a damping ratio of 1 is at 2.45 Hz. The type 2 circuit for this example is shown in Fig. 12.16. If 100 kHz 3 dB frequency is required, the natural frequency is found from scaling the graph. Thus $f_n = 100\ \text{kHz}/2.45 = 41\ \text{kHz}$. The output frequency is 16 times the input frequency, so $N = 16$. The value for $K_{vco}$ is specified at 1 MHz/volt. The phase detector output is a cosine wave. If the loop locks at 90° or 270°, the phase detector output voltage is zero. For a positive $R_{in}$, the slope is the first derivative evaluated at 270°, so $K_{pd} = 50$ mV/rad.

For part a, use the equations derived earlier with $C$ set at 100 pF. Then $K_t = [(1 \cdot 50)/16] \cdot 2\pi \cdot 10^3$, and from Eq. (12.48), $R_{in} = 2.96$ kΩ. Finally Eq. (12.49) gives $R_s = 77.6$ kΩ.

For part b, $R_{in} = 10$ kΩ. With $R_{in}$ set, $C = 29.6$ pF and $R_s = 262.4$ kΩ.

***Example 2.*** The synthesizer design shown in Fig. 12.17 requires an output frequency from 900 to 920 MHz. The output frequency can be changed in 1 kHz steps by changing the divide ratio. Design a PLL using a synthesizer IC and an external VCO. The synthesizer IC data sheet lists the current mode phase detector output as 5 mA/rad. The VCO data sheet lists the tuning rate at 10 MHz/volt.

*Solution 2.* The output frequency must be an integer multiple of the reference frequency, so the reference frequency is 1 kHz. The circuit diagram is shown

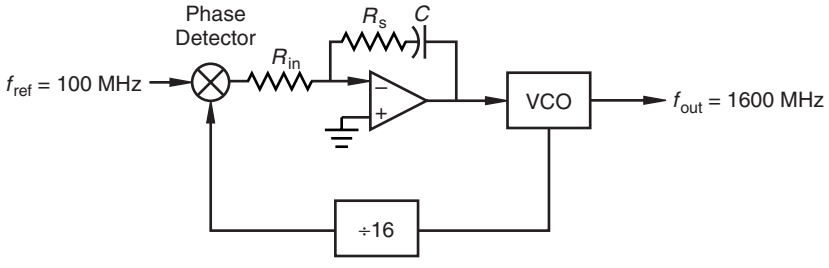



**FIGURE 12.16** PLL for Example 1.

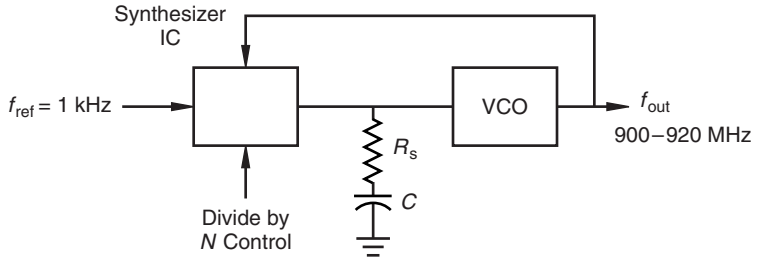



**FIGURE 12.17** Synthesizer design for Example 2.

in Fig. 12.17. The divide ratio must change from 900 MHz/1 kHz or $9 \cdot 10^5$ to 920 MHz/1 kHz or $9.2 \cdot 10^5$. The midpoint value, $9.1 \cdot 10^5$, can be used for the design. A damping ratio of 1 is chosen for a rapid settling time when the divide ratio changes. The loop filter must attenuate the pulses from the phase detector output running at 1 kHz. Figure 12.13 shows that 14 dB of attenuation can be expected at 10 times the natural frequency. With a slope of $-20$ dB/decade, 34 dB attenuation can be expected at 100 times the natural frequency. Choosing $f_n = 10$ Hz will work with a 1 kHz reference frequency. Here the value of $K_t$ is in dimensions of V/A:

$$K_t = \frac{K_{vco}K_{pd}}{N} \cdot \frac{\text{MHz}}{\text{volt}} \cdot \frac{\text{mA}}{\text{Hz}} \cdot \frac{2\pi \text{ rad}}{\text{Hz}}$$

Evaluation of this gives $K_t = 0.345$ A/V. To find $R_s$ and $C$, $K_t$ must be multiplied by $R_{in}$. This will cancel the $R_{in}$ in Eqs. (12.46) and (12.47). Thus from the design equations $R_s = 364\ \Omega$ and $C = 87.45\ \mu$F.

***Example 3.*** A frequency synthesizer contains a phase lock loop circuit. Inspection shows that the loop filter is to be a type 1 noninverting configuration. The data sheets for the phase detector show that the output waveform has a slope of 100 mV/rad. The VCO nominal output frequency is 3 GHz with a tuning rate of 100 MHz/volt. The reference is a 100 MHz crystal oscillator. If $R_{in} = 620\ \Omega$, $R_s = 150\ \Omega$, $R_p = 56\ \text{k}\Omega$, and $C = 1$ nF, what is the expected 3 dB bandwidth and damping ratio for this PLL?

*Solution 3.* With a 3 GHz output and a 100 MHz reference, the frequency divide ratio $N$ must be 30. Also $K_t = 2.094 \cdot 10^6$. Substituting the given circuit values into the analysis equations show that $f_n = 293.7$ kHz and the damping ratio $\zeta = 0.709$. The curve for $\zeta = 0.709$ is not shown in Fig. 12.13, but a value can be found by interpolation. The 3 dB frequency for $\zeta = 0.5$ is 1.8 Hz. The 3 dB frequency for $\zeta = 1$ is 2.45 Hz. A linear approximation for 0.709 is 2.07 Hz. The 3 dB frequency for this PLL is approximately $2.07 \cdot f_n = 608$ kHz.

## PROBLEMS

**12.1** A phase lock loop can be described in the frequency domain in terms of the input and output phase angles shown in Fig. 12.7. The input phase is $\tilde{\phi}_1(s) = a + b/s^2$. The filter transfer function is

$$F(s) = \frac{1 + sCR_2}{1 + sC(R_1 + R_2)}$$

**(a)** What is the steady state phase error?

**(b)** What is the steady state phase error if the capacitance $C = \infty$?

## REFERENCES

1. A. J. Viterbi, *Principles of Coherent Communication*, New York: McGraw-Hill, 1966.
2. U. L. Rhode, *Microwave and Wireless Synthesizers*: Theory and Design, New York: Wiley, 1997.
3. J. A. Crawford, *Frequency Synthesizer Design Handbook*, Norwood, MA: Artech House, 1994.
4. U. L. Rhode, et al., *Communications Receivers, 2nd ed.*, New York: McGraw-Hill, 1996.
5. B. Razavi, *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, New York: IEEE Press, 1996.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

CHAPTER THIRTEEN

# Emerging Technology

## 13.1 INTRODUCTION

The rapid surge in wireless telephones has revolutionized the way people communicate with one another and has brought new impetus to design of radio frequency circuits. The ubiquitous cell phone appears to be only the beginning of possibilities for new forms of communication.

The military introduced the first wireless communication system, and it was limited to voice. The mobile voice service allowed effective deployment of forces in the battle. The equipment was bulky and voice quality was poor. Other agencies such as police, fire, ambulance, marine, and aviation have used mobile communications to facilitate their operations. After World War II the Federal Communications Commission (FCC) established a citizens band at 460 MHz and below allowing private individuals the opportunity to use mobile radio for personal purposes. It is estimated that there were 800,000 authorizations in 1970. Although the "cellular" phone system was conceived by Bell Telephone Laboratories in 1947, the first public system known as Advanced Mobile Phone Service (AMPS), was introduced in 1972 by AT&T (Bell System) in the United States [1,2]. The AMPS phone system was followed by the introduction of Nordic Mobile Telephone (NMT) and Theater Army Communication System (TACS) in Europe, and additional work in Japan. Untethering the telephone and enabling people to conduct communications away from the office and home, and on the move, heralded the wireless era with the goal: "communications-anywhere-anytime." These first generation systems were analog and used frequency modulation. The 1980s witnessed the development of integrated circuits, frequency synthesizers, high-capacity high-speed switch technology, and the like. This resulted in Public Land Mobile Radio (PLMR) networks in the Ultra-High-Frequency bands (UHF), allowing communications for users of Public Switched Telephone Networks (PSTN) or Integrated Service Digital Networks (ISDN) on the move.

Second-generation systems such as the Global System for Mobile (GSM), Digital Cellular Services (DCS), US Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA) digital standards (TIA's IS–54/IS–136, and IS–95), Digital European Cordless Telephone (DECT), and the British Personal Access Communications Systems (PACS) appeared in the late 1980s and early 1990s. These systems in the 900 or 1800 MHz frequency band use complex modulation schemes such as Gaussian minimum shift keying (GMSK), frequency shift keying (FSK), or quadrature phase shift keying (QPSK). The second-generation systems are much more flexible in channel allocation, and they automatically select a base station with strongest signal to communicate. In addition to voice communication, a range of data services is provided by these systems. Presently second-generation systems are commercially deployed in many countries of the world with the subscriber base exceeding 400 million.

Third-generation (3G) wireless systems [3,4] are evolving with the aim of universal access worldwide and a single set of standards meeting a wide range of services and applications. These systems will truly fulfill the dream of "communication-anywhere-anytime." The International Mobile Telecommunications–2000 (IMT–2000) serves as a catalyst and provides the framework for worldwide wireless access by linking the diverse systems of the terrestrial and satellite-based networks. IMT–2000 focuses on such issues as spectrum needs, higher data rate capabilities, Internet Protocol (IP)-based service needs of IMT–2000 and systems beyond IMT–2000. Third-generation systems envision a single universal personal communicator or device capable of roaming globally for voice, data, and video services. There would be provisions for multimedia applications and wide range of video-teleconferencing, high-speed Internet, speech and high-data-rate communication. Systems will use the Broadband Integrated Services Digital Network (B–ISDN) to access the Internet and other information libraries at high speed. The wireless infrastructure will support data rates to the handset of up to 2 Mb/s for stationary use, 384 kb/s at walking pace, and 144 kb/s for vehicles traveling at speeds of up to 60 mph. These systems will enable network operators to radically change the services available on cellular networks. Third-generation systems will require significant investment in technology and resources. The system deployment is expected in the early twentyfirst century with potentially billions of users.

Researchers are beginning to talk about the fourth-generation (4G) systems with powerful attributes such as flexibility and adaptability [5]. Global mobility functions will be supported by software defined radios (SDR), use of digital signal processing (DSP) methods, and low-cost intelligent multimode terminals. Visions of wearable wireless networks with biodegradable sensors embedded in the user's clothes are being proposed. These disposable supersmart systems would be a really low-cost "throw-away price." The multimedia access and system functionality will make the user the "six million dollar man" or the "bionic woman," a rather interesting futuristic concept. The three driving forces for the future global communication remain unchanged as bandwidth, spectrum conservation, and mobility. These areas are currently challenging engineers and research

groups in both large and small countries around the world as well as university laboratories. Additionally the communication network infrastructure has to meet reasonable cost targets to attract investments for enabling technology development. This chapter briefly reviews these three items and presents the efforts to improve efficiency and management.

## 13.2 BANDWIDTH

Today the three transmission modes are copper cable lines, wireless, and optical fiber lines. Telephone lines, as they are currently configured, limit digital data transmission to 56 kb/s. Digital subscriber lines are available that give 1 to 10 Mb/s data rate, but the expense and availability of these still preclude their widespread use. Fiber lines can provide the required bandwidth for most users, but their cost also precludes widespread use. The use of the wireless in the unlicensed wireless bands of 2 to 5 GHz and 20 to 40 GHz can provide data rates in the 5 Mb/s and 10 to 1000 Mb/s, respectively. The economic solution appears to bring fiber lines to a base station, and provide wireless for the last 3 to 5 km individual users. This may ease deployment and lower infrastructure cost. These systems are not significantly affected by multipath because of the high placement of antennas and their narrow beam width. The disadvantages to wireless broadband systems are its restriction to line-of-sight signal transmission resulting in less than 100% coverage, weather effects on radio signals, new technical challenges for commercial application of millimeter wave electronics, and a lack of standards [6].

In 1998 the FCC auctioned the frequency spectrum between 27.5 and 31.3 GHz for local distribution of video, voice, and data for what is called local multipoint distribution service (LMDS). This would provide two-way wireless communications at a high data rate. There are multipoint systems in the 24 to 26 GHz frequency bands as well. Actual frequency plans and bandwidth allocations vary: in the United States (24.25–24.45 GHz and 25.05–25.25 GHz), in Korea (24.25–24.75 GHz and 25.5–26.0 GHz), Germany (24.55–25.05 GHz and 25.56–26.06 GHz), and so on. Bandwidths in the 39 GHz band (38.6–40 GHz) and in the 60 GHz band (59–64 GHz) have also been reserved for future communication systems. Wider bandwidths are naturally available at higher-frequency bands. The need for wider bandwidth is quite clear, since users will demand more and more high-quality multimedia and data applications.

## 13.3 SPECTRUM CONSERVATION

In apparent direct contradiction to the desire for greater bandwidth described in the previous section is the need to conserve spectrum for use by a larger number of people. The three major digital modulation schemes used today, namely frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA), each attempt to bring acceptable voice quality service at a minimum cost in spectrum usage. In addition, use of a smart

antenna that can follow a mobile unit with a narrow width beam also provides a method to conserve (or reuse) spectrum. The cost of a smart antenna is still too high, and this solution still rests in the future. Power levels between two nearly adjacent cells must be kept low enough to avoid interference between the cells. Multiple antennas can sometimes be used along with appropriate digital signal processing to minimize or cancel this interference. All this activity has spurred innovations in diverse areas like voice encoding, antenna design, and linearizing power amplifiers for both amplitude and phase flatness. One example of the later is the feed-forward amplifier.

Since separate channels are spaced close together, and in order to avoid interference in the adjacent channels, a high demand is placed not only on the linearity, amplitude flatness and phase flatness but also on phase noise and spurious emissions of the receiver and transmitter [7]. Nonlinearities can cause "spectral regrowth" in adjacent channels. The degree of linearity can be expressed in terms of (1) a carrier-to-interference ratio, (2) the third-order intercept point, or (3) an adjacent channel power ratio. The third-order intercept point method is the same as that described for mixers in Chapter 11, and as described there, it relates the degree of linearity to a single number. Amplifier linearity is especially crucial when faced with the problem of transmitting digital signals with a short rise time through a narrow channel. Use of more complex modulation schemes such as quadrature phase shift keying (QPSK) and 16 or 64 quadrature amplitude modulation (16QAM, 64QAM) give a special challenge to linear amplifiers because of the shrinking of the distance between symbols (amplitude and phase margin). In many ways the amplifier linearity specification has become more important than the amplifier noise figure.

Noise, though, cannot be neglected, for it eventually will degrade signal quality. The Friis formula described in Chapter 8 will give an overoptimistic view of the noise figure, especially in the millimeter wave wireless systems of the future. Broadband noise from the transmitter as well as the phase noise in the local oscillator of the mixer can add significant noise to the desired intermediate frequency (IF) signal.

## 13.4 MOBILITY

Mobile two-way communications technology has come a long way from the large radio transmitters and receivers to the hand-held cellular telephone. In making this possible, base stations that service areas in the 20 to 30 mile radius must be able to detect a call from anywhere within its boundary and link it eventually with another caller through the telephone system. The problem of making and maintaining contact between people becomes more challenging in areas where there are large obstructions such as buildings and forests. If the mobile caller is traveling in a car at 55 mph and passes from one cell to another, there must be some facility to hand off the call to the new base station. While many of these problems have been solved, there remain many improvements that can be made.

## 13.5 WIRELESS INTERNET ACCESS

Progress in wireless communication has been characterized as three generations of systems (and now maybe even four). The first of these, 1G, has been characterized by the analog mobile cellular telephones acting in concert with a local base station. The second generation, 2G, has been characterized as expansion to a global system for mobile (GSM) communication, using the digital TDMA and CDMA systems. However, this second generation has remained limited in bandwidth and has been limited in carrying data. The third generation, 3G, has been driven by a desire for wireless internet access with high-data-rate capability, as well as the globalization of wireless communication at any time. While the details of how this will work are still being planned, there appears to be a clear desire to put such a system in the hands of consumers shortly. One of the proposed systems is based on what has been termed generic packet radio service (GPRS). This system defines new radio and network interfaces that can handle intermittent bursts of data. It is capable of supporting asymmetric data transmission and the necessary capacity on demand.

The technical requirements for designing a mobile internet protocol includes the ability to retain the same permanent address when a mobile unit changes location, security that may be done by means of data encryption, and its independence of particular wireless technology. In analog systems and in systems that look analog to the user, such as normal telephone calls even when an intermediate step is digital, the user can interpolate between data errors or noise to understand the message. However, in strictly digital systems where binary data are being sent, the entire packet of data must be correct, or at least correctable, before the receiver can process it. Wireless data systems must have low bit error rate, and this is compromised by multipath transmission and nonuniform user power levels. The multipath problem can be alleviated by using a multi-carrier modulation system or better, a multi-code system where the data stream is spread over the entire available bandwidth, and by dividing the data stream into a number of parallel low data rate streams [8].

Another application for wireless systems is fulfilling the FCC regulations for location of mobile 911 calls. The basic requirement is that a mobile 911 call be located to within an accuracy of 50 meters 67% of the time and within 150 meters 95% of the time. The typical method is to use triangulation by measuring the time of arrival or angle of arrival of a mobile signal together with receiver signal strength to at least three base stations. Because of buildings, foliage, and other location dependent peculiarities, location data would in many cases need to be supplemented with software mapping of the particular region. Use of the global positioning satellite alone has been ruled out because of the length of time required by the satellite to acquire signals and the weak signal strength from the satellite, especially inside buildings. However, it could be used as a supplement.

Wireless systems are seeking also to make an inroad to reduce copper wire within the home and office by replacing cable to handheld electronic devices.

This project has been called *bluetooth*. At this point, bluetooth is not a recognized standard but is based on using known technology and protocols as much as possible. It operates in the unlicensed 2.4 to 2.487 GHz industrial scientific and medical band, which is almost uniformly accepted around the world. Each receiver must have a sensitivity of −70 dBm and have its own burned-in address. Significant work will be required for bluetooth to be compatible with the Internet, but that is a possibility.

## 13.6 KEY TECHNOLOGIES

Under the 1962 Communications Satellite Act, COMSAT and later in 1964 INTEL−SAT, an International Telecommunications Satellite Organization was created in an effort to make communication available to all nations. Ironically the Bell System, the organization that conceived the satellite idea and launched the first *Telstar*, was also required to cease work on satellite technology. Instead, Bell was asked to develop fiber-optic technology. Satellite-based systems such as *Iridium, Globalstar, ICOG, Teledesic*, and *Skybridge*, among many others, are being designed in the 1.6 GHz, 2.2 GHz, 6 GHz, 12−14 GHz, 30 GHz, and 48 GHz bands. These satellites are planned to be in the low earth orbit to geostationary orbits around the globe. Although these space-based communication systems are expensive, they are emerging. In addition to satellites, innovative solutions of using high-altitude platform (aircraft and tethered balloon)-based systems such as high-altitude long operation (HALO) are in planning and initial testing. For cost and operational purposes, these systems, although feasible today, would primarily service high-density traffic in metropolitan areas. Details of these space based and high altitude systems can be found in published materials in technical journals.

The years 1980 to 2000 have been characterized as the personal computer era, while the years 2000 to 2020 may be characterized as the Internet era [9]. The personal computer era has driven industry to produce digital microprocessors and memory. The next phase of development may well be producing systems on a chip (SOC), which will start sharing the past progress expressed by Moore's law[†] in the digital chip industry. The shrinking of transistors alone will not be sufficient. Rather, an integration of digital signal processing (DSP) and analog functions into a mixed signal (digital and analog) on a single chip will be required. Wireless internet access requires a radio (excluding most likely the power amplifier), a modem, an applications processor, and some form of power management. The incompatibility of analog and digital electronics resides primarily in the voltage requirements for the two technologies. Lower voltage increases logic speed and efficiency, while higher voltage improves analog RF and baseband circuits. Thus, for a true systems on a chip application, fundamental design changes in analog device designs will need to be made.

[†] This was first expressed by Gordon Moore, co-founder of Intel, in 1965. The "law" states that the data storage of a microchip will double every 12 to 18 months.

The beginning premise of this book was that there is a boundary between radio frequency and microwave designs. Implied also is the boundary between analog and digital designs. It is clear that these demarcations are not part of the creation order itself but a separation that has been used for our own convenience. Clearly, future developments will be based on engineers being well versed in all these camps.

## REFERENCES

1. William C. Jakes, *Microwave Mobile Communications*, New York: IEEE Press, Classic Reissue, 1993.
2. W. R. Young, "Advanced Mobile Phone Service Introduction, Background and Objectives," *Bell Sys. Tech. J.*, Vol. 58, pp. 1–14, 1979.
3. Donald C. Cox, "Wireless Personal Communications: What Is It?" *IEEE Personal Communications*, pp. 21–35, April 1995.
4. Shafi, A. Hashimoto, M. Umehira, et al., "Wireless Communications in the Twenty-First Century: A Perspective," *IEEE Proc.*, pp. 1922–1639, 1997.
5. R. Steele, "A Vision of Wireless Communications in the 21st Century," in Shafi, and Ogose, *Wireless Communications in 21st century*, New York: IEEE Press, 2000, ch. 2.1.
6. L. Langston, tutorial on "Broadband Wireless Access: An Overview," 2000 Emerging Technologies Symposium: Broadband, Wireless Internet Access, April 10–11, 2000, Richardson, TX.
7. T. Howard, "Design Challenges in the High Volume Manufacturing of Millimeter-Wave Transceivers," 2000 IEEE Emerging Technologies Symp.: Broadband, Wireless Internet Access, April 10–11, 2000, Richardson, TX.
8. M. F. Madkour and S. C. Gupta, "Performance Analysis of a Wireless Multirate DS–CDMA in Multipath Fading Channels," 2000 IEEE Emerging Technologies Symp.: Broadband, Wireless Internet Access, April 10–11 2000, Richardson, TX.
9. D. D. Buss, et al., "DSP and Analog SOC Integration in the Internet Era," 2000 IEEE Emerging Technologies Symp.: Broadband, Wireless Internet Access, April 10–11 2000, Richardson, TX.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

*APPENDIX A*

# Example of a Solenoid Design

The design of a solenoid inductor was described in Chapter 2. An example for a 100 nH solenoid design at 200 MHz is shown below using the program SOLENOID where the **bold** type indicates user inputs. After initial values for the solenoid are given, the form length and number of integral turns are modified, but the form diameter remains unchanged. The interwire capacitance, C, the self-resonant frequency, Fres, and the Q are calculated.

```
Frequency in Hz for Q calculation = 200.E6
Geometry -> Inductance, type <G>
Inductance -> n, the number of turns, type <L>
L
Desired inductance, L (H), initial form Length,
Form Diameter - inches, Initial value for no. turns
100.E-9, 0.5, 0.1, 20
Number of turns = 13.00000
AWG = 21.0 Wire Diameter = 0.284625E-01
Final Length = 0.576708E + 00 in.


Pitch = 0.44362E-01 Wire Diameter = 0.28462E-01 in.
Turn Diameter = 0.12846E+00 in. Form Diameter
   = 0.10000E+00 in.
L = 0.10002E+00 μH


C = 0.203350E+00 pF Fres = 0.111600E+04 MHz
Q = 0.649760E+02 at F = 0.200000E+03 MHz
```

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

*APPENDIX B*

# Analytical Spiral Inductor Model

Modeling equations for spiral inductors given in [1,2] are collected here for convenience. A straight line of length $l$, width $w$, thickness $t$, resistivity $\rho$, and in a material of permeability $\mu_0$ has the following inductance:

$$L = \frac{\mu_0 l}{2\pi}\left[\ln\left(\frac{2l}{\rho}\right) - 1\right] \tag{B.1}$$

where

$$\ln\left(\frac{\rho}{2c}\right) = -\frac{25}{12} - \frac{1}{6}\left[\left(\frac{w}{t}\right)^2 \ln\sqrt{1+\left(\frac{t}{w}\right)^2} + \left(\frac{w}{t}\right)^2 \ln\sqrt{1+\left(\frac{w}{t}\right)^2}\right] + \frac{2}{3}\left[\frac{w}{t}\arctan\left(\frac{t}{w}\right) + \frac{t}{w}\arctan\left(\frac{w}{t}\right)\right] \tag{B.2}$$

$$2c = \sqrt{w^2 + t^2} \tag{B.3}$$

If inhomogeneous current density across the conductor cross section is to be considered, an additional expression is found in [1].

The single-loop inductor is illustrated in Fig. B.1 where the total angular rotation, $\phi_0$, is somewhat less than 360°. The inductance is found by numerical integration of the following equation:

$$L = \frac{\mu_0}{2\pi w^2}\int_0^{\phi_0/\sqrt{2}} (\phi_0\sqrt{2} - 2\phi)\cos\left(\sqrt{2}\phi\right) F(\phi)\, d\phi \tag{B.4}$$

where

$$F(\phi) = G(r_o, r_o) - G(r_i, r_o) - G(r_o, r_i) + G(r_i, r_i) \tag{B.5}$$

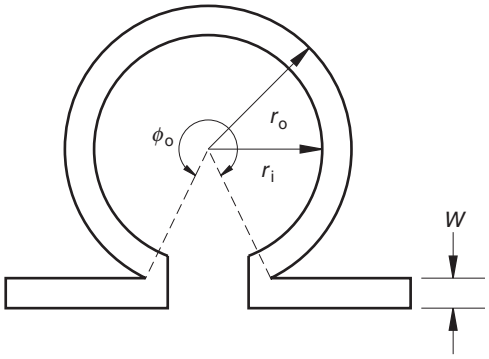



**FIGURE B.1** Single-loop inductor.

$$G(r_1, r_2) = \frac{1}{3}R^3 + \frac{2}{3}r_1 r_2 R \cos\left(\sqrt{2}\phi\right) + \frac{2}{3}r_1^3 \cos\left(\sqrt{2}\phi\right) \arcsin\left(\frac{r_2 - r_1 \cos\left(\sqrt{2}\phi\right)}{r_1 |\sin(\sqrt{2}\phi)|}\right) \tag{B.6}$$

$$R = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\sqrt{2}\phi)}. \tag{B.7}$$

The inductance of a circular spiral with $n$ turns (with air bridge) consists of $n$ static inductances, $L_i$, $i = 1, \ldots, n$, as found above plus mutual inductance terms between the $i$th and $j$th line segments. This mutual inductance is

$$M_{ij} = \mu\sqrt{ab}\left[\left(\frac{2}{k_{ij}} - k_{ij}\right)K(k_{ij}) - \frac{2}{k_{ij}}E(k_{ij})\right] \tag{B.8}$$

where

$$k_{ij} = \frac{4ab}{(a+b)^2} \tag{B.9}$$

$$a = r_i + (i - 0.5)(w + s) \tag{B.10}$$

$$b = r_i + (j - 0.5)(w + s) \tag{B.11}$$

In this expression for the mutual inductance, $r_i$ is the inner radius of the inner most turn of the circular spiral, $w$ is the conductor width, and $s$ is the spacing between turns. The outer most radius of the outer most turn is determined by these parameters together with the number of turns, $n$. The $K(k_{ij})$ and $E(k_{ij})$ are the complete elliptic integrals of the first and second kind, respectively. If there is a ground plane underneath the spiral conductor a distance $h$ away, there is an additional mirrored mutual inductance, $M_{ij}^m$ given by Eq. (B.8) where

$$k_{ij} = \frac{4ab}{4h^2 + (a+b)^2} \tag{B.12}$$

and where $a$ and $b$ are given by Eqs. (B.10) and (B.11). The inductance of the multiple-turn circular spiral is then

$$L = \sum_{i=1}^{n} L_i + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} M_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}^{m} \tag{B.13}$$

The associated capacitances are shown in Fig. B.2, which were given by [3,4,5]:

$$C_{\text{ga}} = \frac{\varepsilon_0}{2}\frac{K(k')}{K(k)} \tag{B.14}$$

The arguments of the elliptic integrals are

$$k = \frac{s}{h}\left(\frac{s}{h} + \frac{2w}{h}\right) \tag{B.15}$$

$$k' = \sqrt{1-k^2} \tag{B.16}$$

The dielectric coupling capacitance is

$$C_{\text{ge}} = \frac{\varepsilon_0\varepsilon_r}{\pi}\ln\left[\coth\left(\frac{\pi}{4}\frac{s}{h}\right)\right] + 0.65C_{\text{f}}\left[\frac{0.02}{s/h}\sqrt{\varepsilon_{\text{r}}} + 1 - \frac{1}{\varepsilon_{\text{r}}^2}\right] \tag{B.17}$$

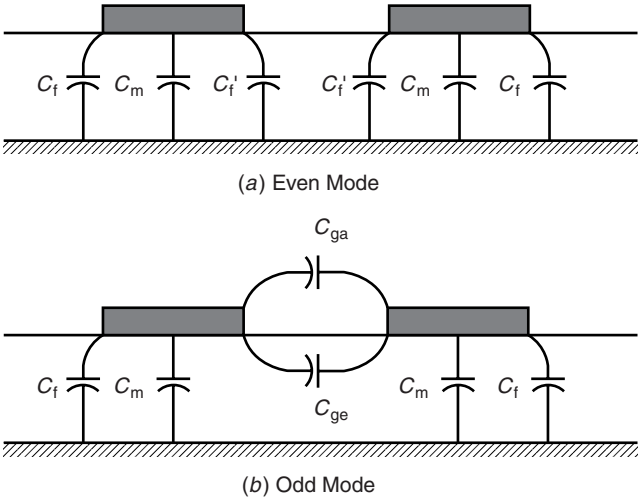



**FIGURE B.2** Capacitances associated with the coupled microstrip.

The main capacitance to ground is

$$C_{\mathrm{m}} = \frac{\varepsilon_0 \varepsilon_{\mathrm{r}} w}{h} \tag{B.18}$$

The fringing capacitances are

$$C_{\mathrm{f}} = \frac{1}{2}\left[\frac{\sqrt{\varepsilon_{\mathrm{eff}}}}{cZ_0} - \frac{\varepsilon_0 \varepsilon_{\mathrm{r}} w}{h}\right] \tag{B.19}$$

$$C_{\mathrm{f}'} = \frac{C_{\mathrm{f}}}{1 + A(h/s)\tanh(8s/h)} \tag{B.20}$$

$$A = \exp\left[-0.1 \exp\left(2.33 - 2.53\frac{w}{h}\right)\right] \tag{B.21}$$

The capacitance, $C_{\mathrm{f}}$, is the fringing capacitance of a single microstrip line of width $w/h$, characteristic impedance, $Z_0$, and effective dielectric constant $\varepsilon_{\mathrm{eff}}$ in which the velocity of light in a vacuum is $c$. The microstrip parameters can be calculated based on Section 4.7.4. Hence the total even-mode capacitance is

$$C_{\mathrm{e}} = C_{\mathrm{m}} + C_{\mathrm{f}} + C_{\mathrm{f}'} \tag{B.22}$$

and the odd-mode capacitance is

$$C_{\mathrm{o}} = C_{\mathrm{m}} + C_{\mathrm{f}} + C_{\mathrm{ga}} + C_{\mathrm{ge}} \tag{B.23}$$

## REFERENCES


1. E. Pettenkpaul, H. Kapusta, A. Weisgerber, H. Mampe, J. Luginsland, and I. Wolff, "CAD Models of Lumped Elements on GaAs up to 18 GHz," *IEEE Trans. Microwave Theory Tech.*, Vol. 36, pp. 294–304, 1988.
2. C. Hentschel, "Die Analyse von Schaltungen mit Dünnfilmschichtspulen," *Arch Elek. Übertragun;g.*, Vol. 26, pp. 319–328, 1972.
3. K. C. Gupta, R. Garg, and I. J. Bahl, *Microstrip Lines and Slotlines*, Norwood, MA: Artech House, 1979, ch. 8.
4. R. Garg and I. J. Bahl, "Characteristics of Coupled Microstrips," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, pp. 700–705, 1979.
5. R. Garg and I. J. Bahl, "Correction to 'Characteristics of Coupled Microstriplines,' " *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, p.272, 1980.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

*APPENDIX C*

# Double-Tuned Matching Circuit Example

Assume that an impedance transformation is required between a 50 Ω source and a 15 Ω load. The matching is to be done using the double-tuned matching circuit described in Chapter 3 for the program DBLTUNE. The center frequency is at 4 MHz, the bandwidth is 100 kHz, and the pass band ripple is 0.5 dB. The capacitances and transformer parameters are to be determined. In the following computer output, the **bold** characters are the responses the program expects from the user. Furthermore, in this example, the verbose mode is chosen by choosing to display the intermediate results. An analysis of this circuit using SPICE is shown in Fig. C.1.

```
Display intermediate results? <Y/N>  Y
Center Freq, Bandwidth (Hz) = ?  4.E6, 100.E3
Fm1 = 0.396480E+07 Fm2 = 0.403551E+07
GTMIN = 0.99992E+00
Passband ripple in dB = ?  0.5
Resistance Ratio r = 0.19841E+01
Q2_m1 = 0.97432E+00 Q2_m2 = 0.10097E+01
Generator and Load resistances values =  50., 15.
L2' = 0.56259E+02 μH C2' = 0.28140E+02pF
RL' = 0.79332E+05 Bm1 = 0.19480E-01 Bm2 = -0.20193E-01
Given terminal resistances: RG = 0.500E+02 RL
   = 0.150E+02
Input Circuit: C1 = 0.446554E+05pF L11 = 0.354637E-01μH
Output Circuit: C2 = 0.148828E+06pF L22
   = 0.106441E-01μH
Transformer coupling coefficient k = 0.250991E-01
```

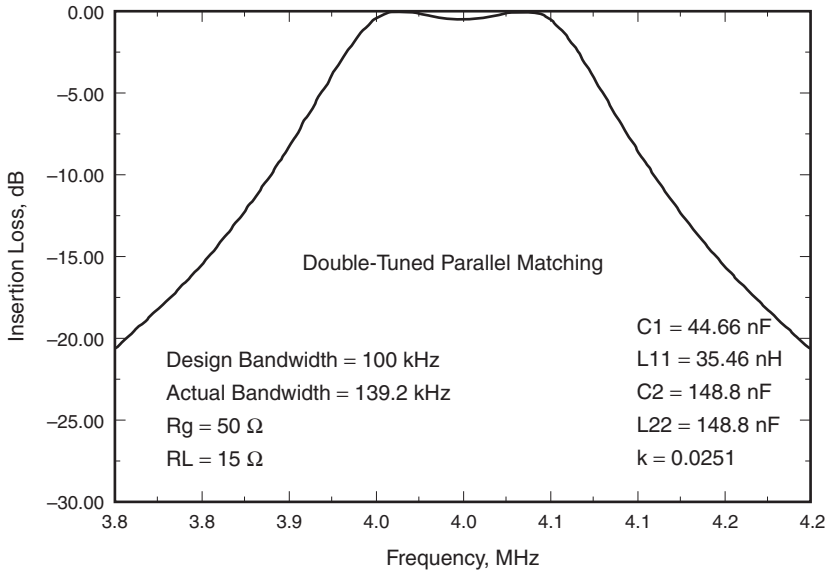



**FIGURE C.1** Double-tuned matching circuit example.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

*APPENDIX D*

# Two-Port Parameter Conversion

## D.1 TWO-PORT VOLTAGE AND WAVE PARAMETERS

Conversion between the $z$, $y$, $h$, and $g$ two-port voltage-current parameters is simply rearrangement of two linear equations relating voltages and currents at the two ports. Converting between these and the $S$ parameters requires relating the voltage waves to voltages and currents. This latter relationship always includes the characteristic impedance, $Z_0$, by which the $S$ parameters are defined. Typically this value is 50 Ω. Table D.1 shows this conversion. The program PARCONV is basically a code of many of the conversions in Table D.1.

The definitions of the various two-port parameters are described below. In each case it is assumed that the current is flowing into the port terminal.

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{D.1}$$

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{D.2}$$

$$\begin{bmatrix} v_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ v_2 \end{bmatrix} \tag{D.3}$$

$$\begin{bmatrix} i_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ i_2 \end{bmatrix} \tag{D.4}$$

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ -i_2 \end{bmatrix} \tag{D.5}$$

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{D.6}$$

**TABLE D.1 *S*-Parameter Conversion Chart**

| | $S$ | $z$ | $y$ | $ABCD$ |
|---|---|---|---|---|
| $S_{11}$ | $S_{11}$ | $\dfrac{(z_{11}-Z_0)(z_{22}+Z_0)-z_{12}z_{21}}{(z_{11}+Z_0)(z_{22}+Z_0)-z_{12}z_{21}}$ | $\dfrac{(Y_0-y_{11})(Y_0+y_{22})+y_{12}y_{21}}{(Y_0+y_{11})(Y_0+y_{22})-y_{12}y_{21}}$ | $\dfrac{A+B/Z_0-CZ_0-D}{A+B/Z_0+CZ_0+D}$ |
| $S_{12}$ | $S_{12}$ | $\dfrac{2z_{12}Z_0}{(z_{11}+Z_0)(z_{22}+Z_0)-z_{12}z_{21}}$ | $\dfrac{-2y_{12}Y_0}{(Y_0+y_{11})(Y_0+y_{22})-y_{12}y_{21}}$ | $\dfrac{2(AD-BC)}{A+B/Z_0+CZ_0+D}$ |
| $S_{21}$ | $S_{21}$ | $\dfrac{2z_{21}Z_0}{(z_{11}+Z_0)(z_{22}+Z_0)-z_{12}z_{21}}$ | $\dfrac{-2y_{21}Y_0}{(Y_0+y_{11})(Y_0+y_{22})-y_{12}y_{21}}$ | $\dfrac{2}{A+B/Z_0+CZ_0+D}$ |
| $S_{22}$ | $S_{22}$ | $\dfrac{(z_{11}+Z_0)(z_{22}-Z_0)-z_{12}z_{21}}{(z_{11}+Z_0)(z_{22}+Z_0)-z_{12}z_{21}}$ | $\dfrac{(Y_0+y_{11})(Y_0-y_{22})+y_{12}y_{21}}{(Y_0+y_{11})(Y_0+y_{22})-y_{12}y_{21}}$ | $\dfrac{-A+B/Z_0-CZ_0+D}{A+B/Z_0+CZ_0+D}$ |
| $z_{11}$ | $Z_0\dfrac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$ | $z_{11}$ | $\dfrac{y_{22}}{y_{11}y_{22}-y_{12}y_{21}}$ | $\dfrac{A}{C}$ |
| $z_{12}$ | $Z_0\dfrac{2S_{12}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$ | $z_{12}$ | $\dfrac{-y_{12}}{y_{11}y_{22}-y_{12}y_{21}}$ | $\dfrac{AD-BC}{C}$ |
| $z_{21}$ | $Z_0\dfrac{2S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$ | $z_{21}$ | $\dfrac{-y_{21}}{y_{11}y_{22}-y_{12}y_{21}}$ | $\dfrac{1}{C}$ |
| $z_{22}$ | $Z_0\dfrac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$ | $z_{22}$ | $\dfrac{y_{11}}{y_{11}y_{22}-y_{12}y_{21}}$ | $\dfrac{D}{C}$ |
| $y_{11}$ | $Y_0\dfrac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}$ | $\dfrac{z_{22}}{z_{11}z_{22}-z_{12}z_{21}}$ | $y_{11}$ | $\dfrac{D}{B}$ |
| $y_{12}$ | $Y_0\dfrac{-2S_{12}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}$ | $\dfrac{-z_{12}}{z_{11}z_{22}-z_{12}z_{21}}$ | $y_{12}$ | $\dfrac{BC-AD}{B}$ |
| $y_{21}$ | $Y_0\dfrac{-2S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}$ | $\dfrac{-z_{21}}{z_{11}z_{22}-z_{12}z_{21}}$ | $y_{21}$ | $\dfrac{-1}{B}$ |
| $y_{22}$ | $Y_0\dfrac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}}$ | $\dfrac{z_{11}}{z_{11}z_{22}-z_{12}z_{21}}$ | $y_{22}$ | $\dfrac{A}{B}$ |
| $A$ | $\dfrac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{2S_{21}}$ | $\dfrac{z_{11}}{z_{21}}$ | $\dfrac{-y_{22}}{y_{21}}$ | $A$ |
| $B$ | $Z_0\dfrac{(1+S_{11})(1+S_{22})-S_{12}S_{21}}{2S_{21}}$ | $\dfrac{z_{11}z_{22}-z_{12}z_{21}}{z_{21}}$ | $\dfrac{-1}{y_{21}}$ | $B$ |
| $C$ | $\dfrac{1}{Z_0}\dfrac{(1-S_{11})(1-S_{22})-S_{12}S_{21}}{2S_{21}}$ | $\dfrac{1}{z_{21}}$ | $-\dfrac{y_{11}y_{22}-y_{12}y_{21}}{y_{21}}$ | $C$ |
| $D$ | $\dfrac{(1-S_{11})(1+S_{22})-S_{12}S_{21}}{2S_{21}}$ | $\dfrac{z_{22}}{z_{21}}$ | $\dfrac{-y_{11}}{y_{21}}$ | $D$ |

For conversion to and from $S$ parameters for circuits with more than two ports, the following formulas may be used [1]. Each variable is understood to be a matrix representing the $S$, $z$, or $y$ parameters. The conversion formulas are

$$S = F(Z - G^*)(Z + G)^{-1}F^{-1} \tag{D.7}$$

$$Z = F^{-1}(I - S)^{-1}(SG + G^*)F \tag{D.8}$$

$$S = F(I - G^*Y)(I + GY)^{-1}F^{-1} \tag{D.9}$$

$$Y = F^{-1}G^{-1}(I + S)^{-1}(I - S)F \tag{D.10}$$

where

$$F = \begin{bmatrix} \dfrac{1}{2\sqrt{Z_{01}}} & 0 & \cdots & 0 \\ 0 & \dfrac{1}{2\sqrt{Z_{02}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dfrac{1}{2\sqrt{Z_{0n}}} \end{bmatrix} \tag{D.11}$$

and

$$G = \begin{bmatrix} Z_{01} & 0 & \cdots & 0 \\ 0 & Z_{02} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_{0n} \end{bmatrix} \tag{D.12}$$

The $I$ in Eqs. (D.8) through (D.10) is the square identity matrix, and the $Z_{0i}$, $i = 1, \ldots, n$, are the characteristic impedances associated with each of the ports. An example of the usage of PARCONV is shown below. In using the program make sure to include the decimals with the input data. The **bold** values represent user inputs to the program. To exit the program, use Ctrl. C.

```
TYPE SOURCE AND LOAD REFERENCE IMPEDANCE ZO1,ZO2 =
 50.,50.
Y --> S = YS OR S --> Y = SY
Z --> S = ZS OR S --> Z = SZ
ABCD --> S = AS OR S --> ABCD = SA
H --> S = HS OR S --> H = SH
H --> Z = HZ OR Z --> H =ZH
 SY
INPUT S11, MAG. AND PHASE (deg)
 .9,-80.
INPUT S21, MAG. AND PHASE (deg)
 1.9,112.
INPUT S12, MAG. AND PHASE (deg)
 .043,48.
```

```
INPUT S22, MAG. AND PHASE (deg)
 .7,-70.
Y(1,1) =   0.162912E-02   J  0.156482E-01
Y(1,2) =   0.304363E-03   J -0.759390E-03
Y(2,1) =   0.360540E-01   J -0.262179E-02
Y(2,2) =   0.483468E-02   J  0.123116E-01
Y --> S = YS OR S --> Y = SY
Z --> S = ZS OR S --> Z = SZ
ABCD --> S = AS OR S --> ABCD = SA
H --> S = HS OR S --> H = SH
H --> Z = HZ OR Z --> H =ZH
```

**TABLE D.2 *S* Parameter to Hybrid Parameter Conversion Chart**

| | $S$ | $h$ |
|---|---|---|
| $S_{11}$ | $S_{11}$ | $\dfrac{(h_{11}-Z_0)(h_{22}Z_0+1)-h_{12}h_{21}Z_0}{(h_{11}+Z_0)(h_{22}Z_0+1)-h_{12}h_{21}Z_0}$ |
| $S_{12}$ | $S_{12}$ | $\dfrac{2h_{12}Z_0}{(h_{11}+Z_0)(h_{22}Z_0+1)-h_{12}h_{21}Z_0}$ |
| $S_{21}$ | $S_{21}$ | $\dfrac{-2h_{21}Z_0}{(h_{11}+Z_0)(h_{22}Z_0+1)-h_{12}h_{21}Z_0}$ |
| $S_{22}$ | $S_{22}$ | $\dfrac{(h_{11}+Z_0)(1-h_{22}Z_0)+h_{12}h_{21}Z_0}{(h_{11}+Z_0)(h_{22}Z_0+1)-h_{12}h_{21}Z_0}$ |
| $h_{11}$ | $Z_0\dfrac{(1+S_{11})(1+S_{22})-S_{12}S_{21}}{(1-S_{11})(1+S_{S22})+S_{12}S_{21}}$ | $h_{11}$ |
| $h_{12}$ | $\dfrac{2S_{12}}{(1-S_{11})(1+S_{S22})+S_{12}S_{21}}$ | $h_{12}$ |
| $h_{21}$ | $\dfrac{-2S_{21}}{(1-S_{11})(1+S_{S22})+S_{12}S_{21}}$ | $h_{21}$ |
| $h_{22}$ | $\dfrac{1}{Z_0}\dfrac{(1-S_{11})(1-S_{22})-S_{12}S_{21}}{(1-S_{11})(1+S_{S22})+S_{12}S_{21}}$ | $h_{22}$ |

## REFERENCES

1. K. Kurokawa, "Power Waves and the Scattering Matrix," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-11, pp. 194–202, 1965.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

*APPENDIX E*

# Termination of a Transistor Port With a Load

In the three-port circuit in Fig. E.1, one of the three ports is terminated with an impedance that has a reflection coefficient relative to the reference impedance $Z_{\text{ref}}$:

$$r_i = \frac{Z_i - Z_{\text{ref}}}{Z_i + Z_{\text{ref}}} \tag{E.1}$$

In this expression the subscript $i$ represents $s$, $g$, or $d$ depending on whether the device connection is common source, gate, or drain terminated with $Z_s$, $Z_g$, or $Z_d$. The incident and scattered waves from the three-port is

$$b_1 = S_{11}a_1 + S_{12}a_2 + S_{13}a_3 \tag{E.2}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 + S_{23}a_3 \tag{E.3}$$

$$b_3 = S_{31}a_1 + S_{32}a_2 + S_{33}a_3 \tag{E.4}$$

When one of the ports is terminated with $r_i$, then the circuit really is a two-port. The scattering parameters for the common source, gate, and drain connection is shown below:

Common source

$$S_{11s} = S_{11} + \frac{S_{12}S_{21}}{1/r_s - S_{22}} \tag{E.5}$$

$$S_{12s} = S_{13} + \frac{S_{12}S_{23}}{1/r_s - S_{22}} \tag{E.6}$$

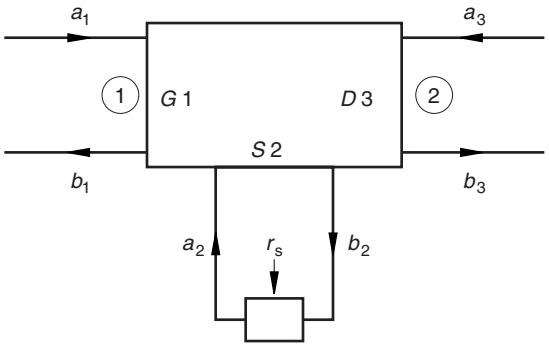



**FIGURE E.1** Three-port with source terminated with $r_s$.

$$S_{21s} = S_{31} + \frac{S_{32}S_{21}}{1/r_s - S_{22}} \tag{E.7}$$

$$S_{22s} = S_{33} + \frac{S_{23}S_{32}}{1/r_s - S_{22}} \tag{E.8}$$

Common gate

$$S_{11g} = S_{22} + \frac{S_{12}S_{21}}{1/r_g - S_{11}} \tag{E.9}$$

$$S_{12g} = S_{23} + \frac{S_{21}S_{13}}{1/r_g - S_{11}} \tag{E.10}$$

$$S_{21g} = S_{32} + \frac{S_{31}S_{12}}{1/r_g - S_{11}} \tag{E.11}$$

$$S_{22g} = S_{33} + \frac{S_{31}S_{13}}{1/r_g - S_{11}} \tag{E.12}$$

Common drain

$$S_{11d} = S_{11} + \frac{S_{13}S_{31}}{1/r_d - S_{33}} \tag{E.13}$$

$$S_{12d} = S_{12} + \frac{S_{13}S_{32}}{1/r_d - S_{33}} \tag{E.14}$$

$$S_{21d} = S_{21} + \frac{S_{23}S_{31}}{1/r_d - S_{33}} \tag{E.15}$$

$$S_{22d} = S_{22} + \frac{S_{23}S_{32}}{1/r_d - S_{33}} \tag{E.16}$$

A numerical example illustrates the process. A given transistor with a set of common source $S$ parameters at 2 GHz is given below:

$$S_{11} = 0.136 \angle 86$$
$$S_{21} = 3.025 \angle 6$$
$$S_{12} = 0.085 \angle -164$$
$$S_{22} = 0.304 \angle -136$$

These are then converted to two-port $y$ parameters. These will be called $y_{11}$, $y_{31}$, $y_{13}$, and $y_{33}$. The indefinite admittance matrix is formed by adding a third row and column such that the sum of each row and the sum of each column is zero. The resulting $3 \times 3$ set of $y$ parameters are obtained:

$$y_{11} = 9.681 \cdot 10^{-3} - 7.695 \cdot 10^{-3}$$
$$y_{12} = -12.77 \cdot 10^{-3} + 6.776 \cdot 10^{-3}$$
$$y_{13} = -3.086 \cdot 10^{-3} + .9194 \cdot 10^{-3}$$
$$y_{21} = 104.2 \cdot 10^{-3} + 20.85 \cdot 10^{-3}$$
$$y_{22} = -82.89 \cdot 10^{-3} + 14.39 \cdot 10^{-3}$$
$$y_{23} = -21.28 \cdot 10^{-3} + 6.452 \cdot 10^{-3}$$
$$y_{31} = -113.8 \cdot 10^{-3} + 13.15 \cdot 10^{-3}$$
$$y_{32} = -95.65 \cdot 10^{-3} + 7.618 \cdot 10^{-3}$$
$$y_{33} = -18.19 \cdot 10^{-3} + 5.533 \cdot 10^{-3}$$

These are then converted to three-port $S$ parameters using Eq. (10.32) [1]:

$$S_{11} = 1.6718 \angle -168.12^\circ$$
$$S_{12} = 1.6573 \angle 3.639^\circ$$
$$S_{13} = 1.0103 \angle 13.684^\circ$$
$$S_{21} = 3.1794 \angle -157.77^\circ$$
$$S_{22} = 2.0959 \angle 14.185^\circ$$
$$S_{23} = 0.7156 \angle 74.511^\circ$$
$$S_{31} = 1.6455 \angle 70.181^\circ$$
$$S_{32} = 2.7564 \angle -167.02^\circ$$
$$S_{33} = 2.1085 \angle -153.87^\circ$$

At this point it is desired to transform these parameters to common gate parameters in which the gate is connected to ground through a short circuit. The resulting common gate two-port $S$ parameters are found from Eqs. (E.9) through (E.12):

$$S_{11g} = 5.317 \angle 170.925^\circ$$

$$S_{21g} = 10.772 \angle -14.852^\circ$$

$$S_{12g} = 2.496 \angle 177.466^\circ$$

$$S_{22g} = 6.250 \angle -7.553^\circ$$

With the transistor now characterized in the orientation that it is to be used in the oscillator, a choice is made for the impedance at the generator side. If this impedance is chosen to be a 5 nH inductor, the output reflection coefficient is

$$\Gamma_o = 1.7775 \angle -30.35^\circ$$

This shows that oscillation is possible under these loading conditions. The expressions given above for the revised $S$ parameters can be found in [2] using slightly different notation.

## REFERENCES

1. K. Kurokawa, "Power Waves and the Scattering Matrix," *IEEE Trans. Microwave Theory Tech.*, pp. 194–202, 1965.
2. R. M. Dougherty, "Feedback Analysis and Design Techniques," *Microwave J.*, Vol. 28, pp. 133–150, 1985.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

*APPENDIX F*

# Transistor and Amplifier Formulas

The following formulas are meant as a reminder of the fundamentals given in most standard electronics textbooks. Notation for the formulas have the traditional meanings. Depletion capacitances are all given with a negative sign in the denominator as in $C = C_0/(1 - V/\phi)^{\gamma}$. Consequently, when the junction is reverse biased, the minus sign turns into a positive sign. Figure F.1 presents the basic FET features and symbols.

Bipolar Transistor Parameters (BJT)

| DESCRIPTION | FORMULA |
|---|---|
| Collector current | $I_C = I_S \exp\left(\frac{qV_{BE}}{kT}\right)$ |
| Transconductance | $g_m = \frac{qI_C}{kT}$ |
| Input resistance | $r_\pi = \frac{\beta_0}{g_m}$ |
| Output resistance | $r_o = \frac{V_A}{I_C}$ |
| Base charging capacitance | $C_D = \tau_F g_m$ |
| Emitter base junction | $C_{je} = A_E \left(\frac{q\varepsilon N_B}{V_j}\right)^{1/3}$ |
| Input capacitance | $C_\pi = C_D + C_{je}$ |
| Collector base | $C_\mu = \frac{C_{\mu o}}{[1 - (V_{BC}/\psi_{oc})]^{1/3}}$ |

Collector substrate $$C_{cs} = \frac{C_{cso}}{[1 - (V_{sc}/\psi_{os})]^{1/2}}$$

Transition frequency $$f_T = \frac{1}{2\pi} \frac{g_m}{(C_\pi + C_\mu)}$$

Thermal voltage $$V_T = \frac{kT}{q} = 0.259 \text{ V}$$

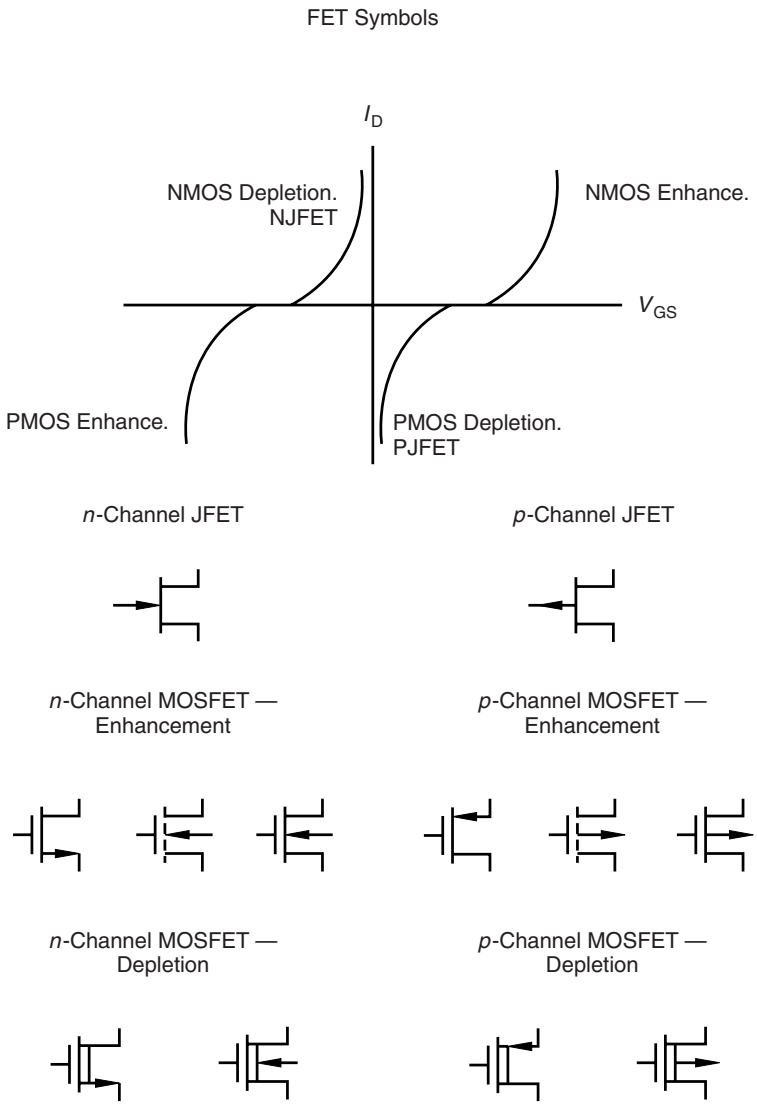



**FIGURE F.1** FET symbols.

## Junction Field Effect Transistor Parameters (JFET)

| DESCRIPTION | FORMULA |
|---|---|
| Saturated drain current | $I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2\left(1 - \frac{V_{DS}}{V_A}\right)$ |
| | $V_{DS} \geq V_{GS} - V_P$ |
| Ohmic region drain current | $I_D = G_o\left[V_{DS} + \frac{3}{2}\frac{(\psi_0 + V_{GS} - V_{DS})^{3/2} - (\psi_0 + V_{GS})^{3/2}}{(\psi_0 + V_P)^{1/2}}\right]$ |
| | $V_{DS} < V_{GS} - V_P$ |
| | $G_0 = \frac{2aW}{L}\sigma_c$ |
| | $I_D \approx K\left[2(V_{GS} - V_P)V_{DS} - V_{DS}^2\right]$ |
| Transconductance | $g_m = \frac{2I_{DSS}}{V_P}\left(1 - \frac{V_{GS}}{V_P}\right)$ |
| Output resistance | $r_o = \frac{V_A}{I_D}$ |
| Gate source capacitance | $C_{gs} = \frac{C_{gs0}}{[1 - (V_{GS}/\psi_0)]^{1/3}}$ |
| Gate drain capacitance | $C_{gd} = \frac{C_{gd0}}{\left(1 - \frac{V_{GD}}{\psi_0}\right)^{1/3}}$ |
| Gate substrate capacitance | $C_{gss} = \frac{C_{gss0}}{[1 - (V_{GSS}/\psi_0)]^{1/2}}$ |
| N Channel JFET | $V_P < 0$ |
| P Channel JFET | $V_P > 0$ |

## Metal Oxide Semiconductor Field Effect Transistor Parameters (MOSFET)

| DESCRIPTION | FORMULA |
|---|---|
| Saturation region drain current | $I_D = \frac{\mu C_{ox} W}{2L}(V_{GS} - V_t)^2\left(1 - \frac{V_{DS}}{V_A}\right)$ |
| | $V_{DS} \geq V_{GS} - V_t$ |
| Ohmic region drain current | $I_D = \frac{\mu C_{ox} W}{2L}\left[2(V_{GS} - V_t)V_{DS} - V_{DS}^2\right] \times \left(1 - \frac{V_{DS}}{V_A}\right)$ |
| | $V_{DS} < V_{GS} - V_t$ |

| | |
|---|---|
| Oxide capacitance | $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$ |
| Transconductance | $g_m = \mu C_{ox} \frac{W}{L}(V_{GS} - V_t)$ |
| Output resistance | $r_o = \frac{\lvert V_A \rvert}{I_{D0}}$ |
| Input capacitance | $C_{in} = C_{GS} + C_{GD} = C_{ox} LW$ |
| Transition frequency | $f_c = \frac{g_m}{2\pi C_{in}} = \frac{\mu_s (V_{GS} - V_t)}{2\pi L^2}$ |
| Surface mobility holes | $\mu_s = 200\ \text{cm}^2/\text{V-s}$ |
| Surface mobility electrons | $\mu_s = 450\ \text{cm}^2/\text{V-s}$ |

| N CHANNEL JFET | P CHANNEL JFET |
|---|---|
| $I_{DSS} > 0$ | $I_{DSS} < 0$ |
| $V_P < 0$ | $V_P > 0$ |
| $g_{mo} \frac{-2I_{DSS}}{V_P}$ | $g_{mo} = \frac{-2I_{DSS}}{V_P}$ |
| $K \triangleq \frac{I_{DSS}}{V_P^2} > 0$ | $K \triangleq \frac{I_{DSS}}{V_P^2} < 0$ |
| $V_P < V_{GS}$ for $\lvert I_{DS} \rvert > 0$ | $V_{GS} < V_P$ for $\lvert I_{DS} \rvert > 0$ |

| NMOS ENHANCEMENT | PMOS ENHANCEMENT |
|---|---|
| $V_t > 0$ | $V_t < 0$ |
| $V_{GS} > V_t$ | $V_{GS} < V_t$ |
| $K \triangleq \frac{\mu_n C_{ox} W}{2L} > 0$ | $K \triangleq \frac{\mu_p C_{ox} W}{2L} < 0$ |

| NMOS DEPLETION | PMOS DEPLETION |
|---|---|
| $V_t < 0$ | $V_t > 0$ |
| $V_{GS} > V_t < 0$ for $\lvert I_{DS} \rvert > 0$ | $V_{GS} < V_t < 0$ for $\lvert I_{DS} \rvert$ |
| $K \triangleq \frac{\mu_n C_{ox} W}{2L} > 0$ | $K \triangleq \frac{\mu_p C_{ox} W}{2L} < 0$ |

## Small Signal Single-Transistor Amplifier Configurations

| MOSFET | BJT |
|---|---|
| Common source | Common emitter |
| $R_{in} = R_G = R_1 \| R_2$<br>$R_{out} = R_D \| r_0$ | $R_{in} = (r_\pi + r_b) \| R_B \approx r_\pi$<br>$R_{out} = R_c \| r_0$ |
| $A_V = -g_m(r_0 \| R_D \| R_L) \propto 1/\sqrt{I_D}$ | $A_V = -g_m(R_c \| r_0 \| R_L)$ |
| Source degeneration | Emitter degeneration |
| $R_{in} = R_G = R_1 \| R_2$ | $R_{in} = r_\pi + R_E(\beta + 1)$<br>$\approx r_\pi(1 + g_m R_E)$ |
| $R_{out} = r_0[1 + (g_m + g_{mb})R_S] + R_S$ | $R_{out} = R_E \| r_\pi + r_0\left[1 + g_m(r_\pi \| R_E)\right]$<br>$\approx r_0(1 + g_m R_E)$ |
| $G_m = \dfrac{g_m}{1 + (g_m + g_{mb})R_S}$ | $G_m = \dfrac{g_m}{1 + g_m R_E}$ |
| Commmon gate | Common base |
| $R_{in} = \dfrac{R_D}{1 + R_D(g_m + g_{mb})} \approx \dfrac{1}{g_m + g_{mb}}$ | $R_{in} = r_e = \dfrac{\alpha_0}{g_m} + \dfrac{r_b}{\beta + 1} \approx \dfrac{\alpha_0}{g_m}$ |
| $R_{out} = R_D$ | $R_{out} = r_0(1 + \beta) \| R_c$ |
| $G_m = g_m + g_{mb}$ | $G_m = g_m\left(\dfrac{1}{1 + r_b/r_\pi}\right) \approx g_m$ |
| | $A_V = g_m(R_c \| R_L)$ |
| | $A_I = \dfrac{g_m R_c}{R_c + R_L}\dfrac{R_E}{1 + g_m R_E} \approx g_m r_e = \alpha_0$ |
| Common drain (source follower) | Commmon collector (emitter follower) |
| $R_{in} = R_1 \| R_2$ | $R_{in} = r_\pi + r_b + (\beta + 1)(r_0 \| R_E)$ |
| $R_{out} = \dfrac{r_0}{1 + r_0(g_m + g_{mb})} \approx \dfrac{1}{g_m + g_{mb}}$ | $R_{out} = \dfrac{r_\pi + R_{BB} + r_b}{1 + \beta} \approx \dfrac{1}{g_m} + \dfrac{R_{BB} + r_b}{1 + \beta}$ |
| $A_V = \dfrac{g_m(R_S \| r_0)}{1 + (R_S \| r_0)(g_m + g_{mb})} \approx \dfrac{1}{1 + g_{mb}/g_m} \approx 1$ | $A_V = \dfrac{1}{1 + [(R_{BB} + r_b + r_\pi)/(R_E \| r_0)(\beta + 1)]} \approx 1$ |

*APPENDIX G*

# Transformed Frequency Domain Measurements Using Spice

## G.1 INTRODUCTION

Time domain measurements taken on an automatic network analyzer can be easily replicated theoretically for dispersionless transmission lines using the SPICE time domain program simulation. A technique is presented here that gives the type, value, and position of the measured discontinuity directly from time domain measurements. Furthermore the effect of multiple discontinuities can be replicated by SPICE.

Microwave impedance measurements have evolved from the slotted line to the computer-controlled automatic network analyzer. One of the major innovations of the network analyzer made possible by its broadband frequency capability was the simulation of time domain measurements. Though fundamentally a frequency domain machine, the network analyzer has been very useful in doing time domain analysis of broadband circuits. This was demonstrated by the pioneering work of Hines and Stinehelfer [1]. Fundamentally time domain equipment such as the time domain reflectometer have lagged because of the requirement for fast switching devices. However, that has not been true of the software side, as the widespread use of SPICE can readily attest. Here it will be demonstrated how time domain measurements on the network analyzer can be simulated using the transient analysis in SPICE.

There have been several papers that have shown how $S$ parameters can be plotted using SPICE [2,3,4]. SPICE can be used to model physical directional couplers by the interconnection of three transmission lines [5]. By this means, a SPICE model was developed for the characteristics of a network analyzer in the frequency domain. However, the above-mentioned applications are not using the time domain capability of SPICE.

First of all, the SPICE *S*-parameter simulation can be expanded to include unequal input and output impedance levels. This is useful in analyzing impedance steps and transformers with two different resistance levels. Second, rather than using ac steady state voltage sources, a time domain pulse can be used. The automatic network analyzer is made to simulate a time domain reflectometer by mathematically producing an "impulse" that can be replicated in SPICE approximately by using the PULSE function or more accurately by the piecewise linear (PWL) transient function. This SPICE replica of the network analyzer "impulse" can be used to show how various circuit elements react to this "impulse." Simple formulas are given that can be used to obtain the value of the inductance, capacitance, or impedance step directly from time domain "impulse" data.

## G.2 FREQUENCY DOMAIN S-PARAMETERS

In [4] the *S* parameters were obtained for a given circuit from SPICE for a two-port circuit in which the input and output resistances were both 50 Ω. However, the conversion circuit used to obtain the *S* parameters can be modified so that the two ports of the circuit are at different impedance levels (Fig. G.1). This is done by the addition of an ideal transformer whose turns ratio is

$$n = \sqrt{\frac{R_{02}}{R_{01}}} \tag{G.1}$$

where $R_{01}$ and $R_{02}$ are the characteristic impedance levels of the input and output sides, respectively. The input impedance, $Z_1$, looking into the transformer from the port-1 side is

$$Z_1 = \frac{Z_L}{n^2} \tag{G.2}$$

which is equal to $R_{01}$ when $Z_L = R_{02}$. The voltage drop caused by the independent current source of value $1/R_{01}$ is

$$V_1 = \frac{1}{R_{01}}(R_{01}||Z_1) = \frac{Z_1}{R_{01} + Z_1} \tag{G.3}$$

The voltage at node 11 in Fig. G.1 is numerically the same as $S_{11}$, since

$$V(11) = 2V_1 - V_{in} = \frac{Z_1 - R_{01}}{Z_1 + R_{01}} \tag{G.4}$$

For the output side, the secondary current in the transformer is $I_2 = I_1/n$. The portion of current going into the transformer primary from the current source side is

$$I_1 = \frac{1}{R_{01}}(R_{01}||Z_1)\frac{1}{Z_1} = \frac{1}{Z_1 + R_{01}} \tag{G.5}$$

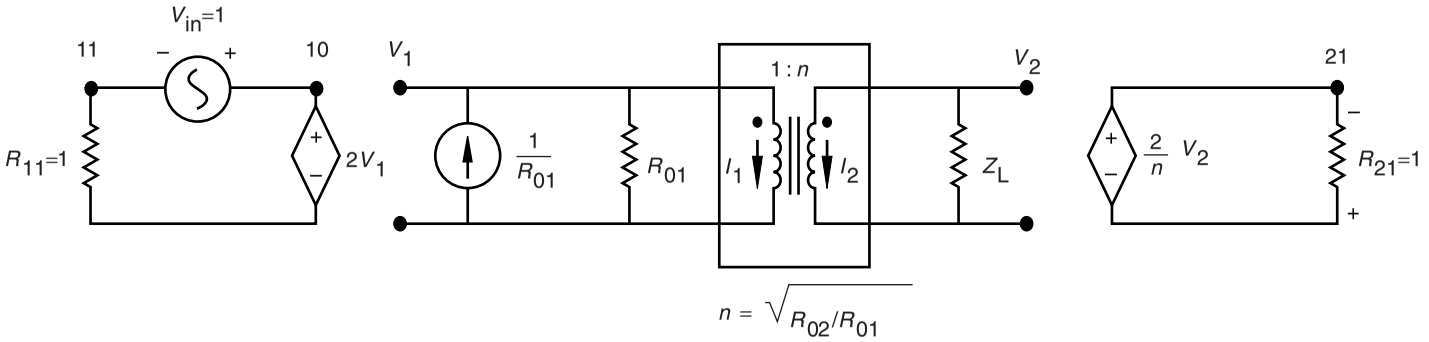



**FIGURE G.1** SPICE circuit for finding *S* parameters with different input and output impedance levels.

The voltage at node 2 is

$$V_2 = -I_2 Z_L = -\frac{I_1}{n} Z_L = \frac{-Z_L}{(Z_1 + R_{01})n} \tag{G.6}$$

The voltage across $R_{21}$ with the marked polarity is $V(21) = -2V_2/n$ which, upon replacing $V_2$ with the value above, gives a value for $V(21)$ that is numerically equal to $S_{21}$:

$$V(21) = \frac{2Z_L}{(Z_1 + R_{01})n^2} = \frac{2Z_L}{Z_L + R_{02}} = S_{21} \tag{G.7}$$

The right-hand side is obtained using Eqs. (G.1) and (G.2). For a matched load when $Z_L = R_{02}$, $S_{21} = 1$ as expected. Finding the $S_{22}$ and $S_{12}$ for a circuit is achieved by direct analogy. A suggested SPICE listing for finding the $S$ parameters with unequal source and load impedance levels is shown in Section G.6.

## G.3 TIME DOMAIN REFLECTOMETRY ANALYSIS

Time domain reflectometer (TDR) measurements from an automatic network analyzer can be directly compared with a theoretical circuit model in the time domain in SPICE. All that is required is to make the two modifications described in Section G.7. Of course, a near-ideal impulse can be implemented in SPICE, but the point of view taken here is to calculate time domain data that would actually be measured using the time domain feature on a network analyzer. Two actual time domain "impulses" were measured on a network analyzer under the conditions in which the maximum frequency was 18 and 26 GHz, respectively. A third "impulse" with a maximum frequency of 50 GHz was calculated from Hewlett Packard's Microwave Development System (MDS) program. This was justified on the basis that (1) the program and the network analyzer use the same algorithm for the chirp-Z transform and (2) the measured and calculated 18 GHz "impulses" are nearly identical. Circuit responses to "impulses" with different maximum frequency content will of course differ, so the time domain response must be always coupled to the frequencies used in producing the "impulse."

Two options for representing the network analyzer "impulse" are illustrated in Fig. G.2. The approximate impulse is modeled in SPICE as a PULSE, while the more accurate approach is achieved by using the piecewise linear (PWL) SPICE function. In the present case, the PWL function uses 77 different $x$, $y$ coordinate pairs to represent the 26 GHz "impulse." The result is indistinguishable from the measured "impulse" within the resolution of the graph in Fig. G.2. A similar fit was made for the 18 and 50 GHz "impulses." The piecewise linear fit of the network analyzer "impulses" for use in SPICE are found in Section G.7. The approximate trapezoidal pulse for a 50 Ω source impedance has the advantage of providing results similar to the PWL approximation with a lot less data entry. However, to closely represent the actual time domain response of the network analyzer, the piecewise linear approach should be used.

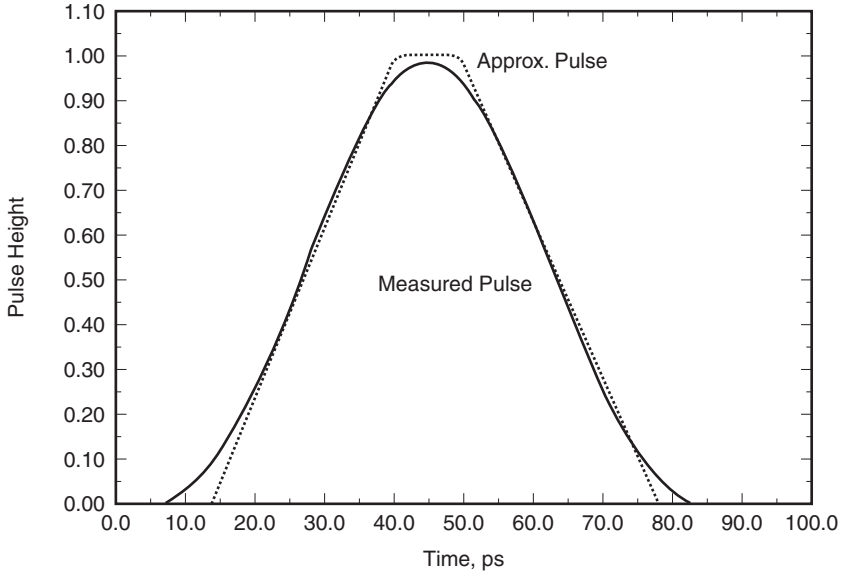



**FIGURE G.2** Measured and approximate trapezoidal 26 GHz pulse used in SPICE analysis.

## G.4 TIME DOMAIN IDENTIFICATION OF CIRCUIT ELEMENTS

Time domain analysis of discontinuities in broadband circuits enables determining the type of circuit element, the position of the circuit element, and the size of the circuit element causing the discontinuity. For example, if a shunt capacitance is causing the discontinuity, the time domain "impulse" reflection first goes below the baseline, then rises above the baseline and then settles back to the baseline, looking roughly like a single-period sine wave. Figure G.3 shows a typical response to a shunt capacitance and series inductance that are a half wave length apart at 10 GHz. The typical response for a series capacitance is shown in Fig. G.4. A shunt inductance response would look like the negative of the series capacitance. The shunt capacitance, series inductance, series capacitance, shunt inductance, and step in characteristic impedance all have their peculiar time domain signatures.

In general, the larger the discontinuity, the larger is the $|\Gamma(t)|$ of the time domain response. For the larger shunt capacitances, the magnitude of the dip below the baseline is not equal to the magnitude of the rise above the baseline. So there is some ambiguity in choosing what part of the curve to use to predict the value of the shunt capacitance. A series of capacitance values were tested using both the piecewise linear representation for the network analyzer time domain "impulse." The results in Fig. G.5 show that if the algebraic maximum of $\Gamma(t)$ (the second extremum) is chosen, there are two possible values of capacitance for $\Gamma(t)$. However, the maximum value of the negative excursion of $\Gamma(t)$ (the first

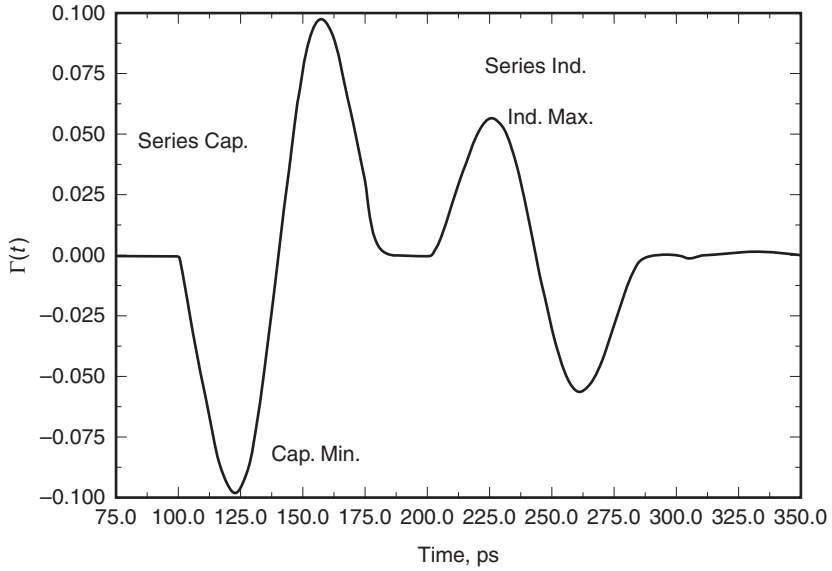



**FIGURE G.3** Time domain response from a shunt capacitor separated from a series inductor by a 100 ps 50 Ω transmission line.

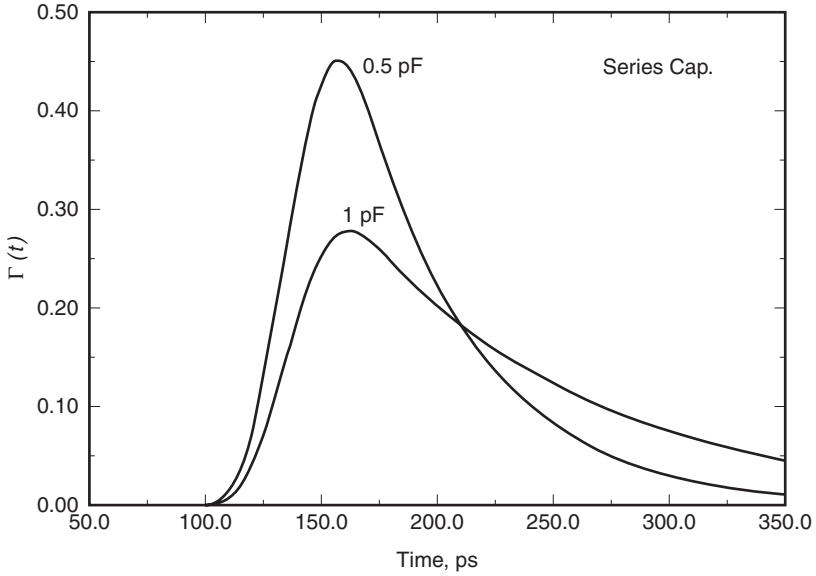



**FIGURE G.4** Typical time domain response for a series capacitor.

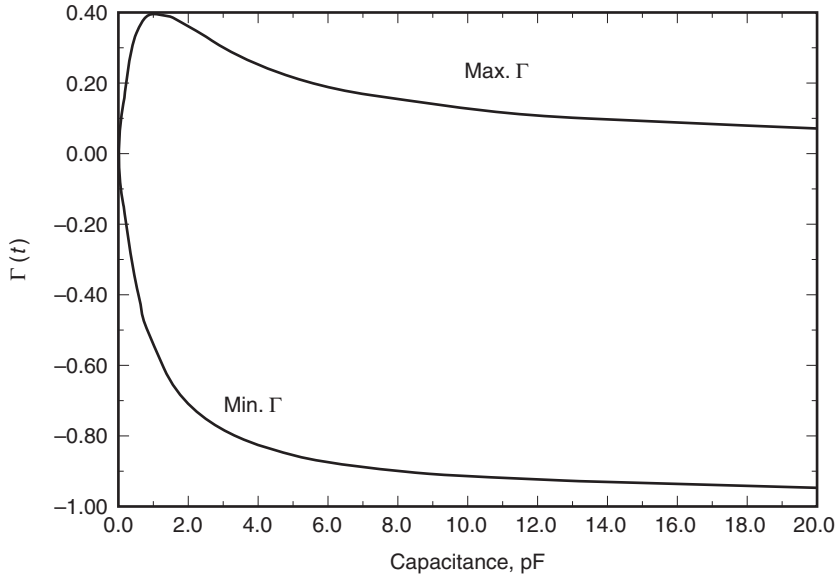



**FIGURE G.5** Maximum and minimum values of the time domain response for a shunt capacitance using the piecewise linear approximation to the measured "impulse."

extremum) is monotonic, and hence it gives a unique value for the capacitance. The series inductance response is a mirror image to the shunt capacitance about the $\Gamma(t) = 0$ line. Thus the first maximum of $\Gamma(t)$ is monotonically related to the value of the series inductance. It is this first maximum of $|\Gamma(t)|$ that can be used to find the value of the two types of discontinuities.

The time domain not only allows determination of the discontinuity type and size but also the discontinuity position in time. Here again there is some ambiguity in the part of the "impulse" response curve that should be used to find the position. For shunt capacitances, a position midway between the lower and upper part of the curve could be used, that is, where $\Gamma(t) = 0$. A comparison of this choice with the theoretical distance is provided in Fig. G.6, where it is seen that this method is accurate if the discontinuity is very small. However, for larger discontinuities a better choice would be at the location of the first extremum of $|\Gamma(t)|$.

The minimum $\Gamma(t)$ of the shunt capacitor, the maximum $\Gamma(t)$ of the series inductor, the maximum $\Gamma(t)$ of the series capacitance, the minimum $\Gamma(t)$ of the shunt inductance, and the peak $\Gamma(t)$ (either negative or positive) of the impedance step can each be described by a simple formula that could be readily stored in a handheld calculator. These formulas are listed below. The parameters for these expressions are in Table G.1 for the 18 GHz, 26 GHz, and 50 GHz "impulses."

Shunt capacitance:

$$C = -\frac{\Gamma A}{1 + \Gamma B} \tag{G.8}$$

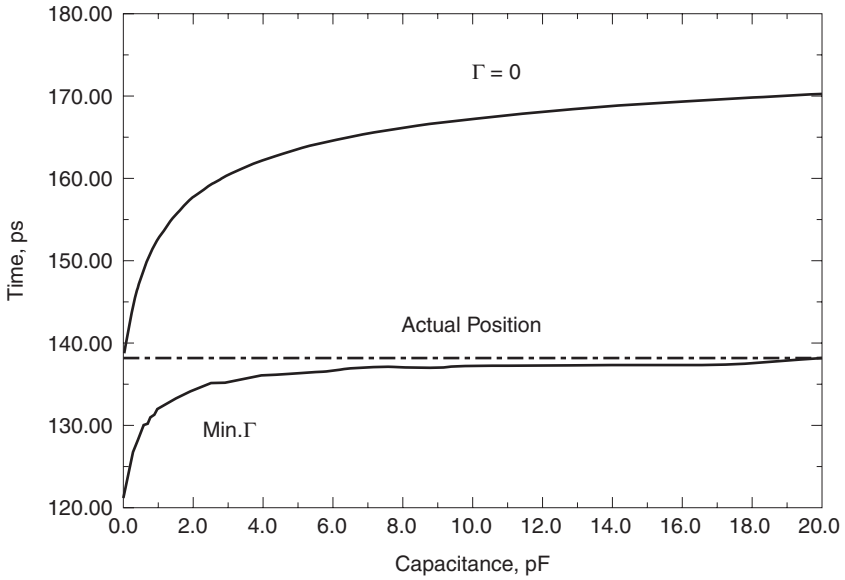



**FIGURE G.6** Estimated position (in time) of a shunt capacitor using the minimum value of $\Gamma(t)$ and the position where $\Gamma(t) = 0$.

**TABLE G.1 Expressions for the Reactance Element Values in Terms of Γ**

| Maximum Frequency | Shunt C, pF $C = -\Gamma A/(1 + \Gamma B)$ | Series L, nH $L = -\Gamma A/(1 + \Gamma B)$ | Series C, pF $C = (1 - \Gamma A)/\Gamma B$ | Shunt L, nH $L = (1 - \Gamma A)/\Gamma B$ |
|---|---|---|---|---|
| 18 GHz | $A = 1.212$ | $A = -2.965$ | $A = 0.9379$ | $A = -0.9239$ |
| pulse | $B = 0.9887$ | $B = -0.9908$ | $B = 1.691$ | $B = -0.6892$ |
| 26 GHz | $A = 0.8216$ | $A = -2.061$ | $A = 0.9068$ | $A = -0.9275$ |
| pulse | $B = 1.004$ | $B = -1.010$ | $B = 2.624$ | $B = -1.036$ |
| 50 GHz | $A = 0.4275$ | $A = -1.091$ | $A = 0.9279$ | $A = -0.8770$ |
| pulse | $B = 0.9875$ | $B = -1.010$ | $B = 4.801$ | $B = -1.9885$ |

Series inductance:

$$L = -\frac{\Gamma A}{1 + \Gamma B} \tag{G.9}$$

Series capacitance:

$$C = \frac{1 - \Gamma A}{B\Gamma} \tag{G.10}$$

Shunt inductance:

$$L = \frac{1 - \Gamma A}{B\Gamma} \tag{G.11}$$

Step in characteristic impedance:

$$Z_0' = Z_0 \frac{1+\Gamma}{1-\Gamma} \tag{G.12}$$

Measurements of a given series-mounted chip capacitor on a LCR meter at, say 1 MHz, would not necessarily give accurate correlation with a time domain model of a simple series capacitance as modeled by Eq. (G.10) or direct measurements on a network analyzer. However, a better fit with the time domain data can be obtained by using a model of a high-frequency capacitor as described in Chapter 2.

## G.5 MULTIPLE DISCONTINUITIES

The formulas above are correct when there is only one significant discontinuity in the circuit that is being measured. When there are multiple discontinuities, the SPICE analysis will also display the results that would be expected in a real time domain reflectometer measurement. The gating error in measuring a discontinuity in the presence of other discontinuities is analyzed in [6]. That analysis uses a rectangular gating function with a depth of 40 dB rather than the chirp-Z transform in the network analyzer software. Four sources of error are identified. The first is out-of-gate attenuation associated with incomplete suppression of reflections outside the gating function. The second is truncation error where the gate is made too narrow to pick up all of the response due to the discontinuity in question. The third is masking error where the transmission coefficients of previous discontinuities reduces the signal getting to the discontinuity under investigation. The fourth is multi-reflection aliasing error that occurs when the circuit has commensurate line lengths, and the residual reflection of one discontinuity adds to or subtracts from the reflection of the discontinuity under investigation.

When two discontinuities are sufficiently separated in time, each can be analyzed separately, though the accuracy of predicting the later discontinuities from the formulas in Table G.1 are less accurate than when there is only one discontinuity. Three impedance steps are set up such that if they were alone, they would have had a $\Gamma(t)$ of 0.3, 0.2, and 0.1, respectively (Fig. G.7), as is done by [6]. However, when the three steps are put in one circuit suitably separated from one another, the SPICE analysis shows that $\Gamma(t) = 0.3$ (0% error), $\Gamma(t) = 0.181$ (9.6% error), and $\Gamma(t) = 0.087$ to 0.076 (12.9% to 24% error depending on whether line lengths are commensurate). The estimated errors in [6] are 0%, 12%, and 20%, respectively, which correlate well with the SPICE results given the approximations that are employed.

In addition the technique described here can be used in a wide variety of circuits to model two elements that are too close together to be resolved separately in time. Adjustment of a theoretical time domain model to make its response match that of the measured time domain measurements gives a method to extract an equivalent circuit in the time domain.

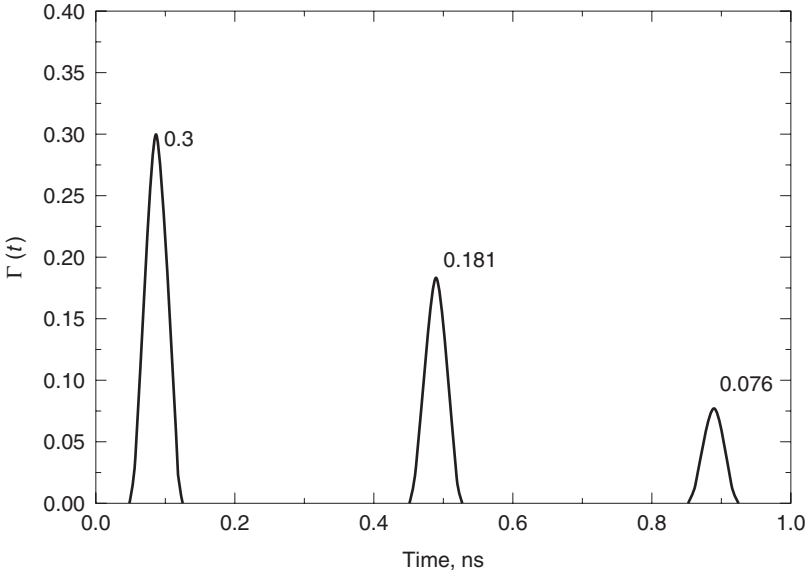



**FIGURE G.7** Predicted response from three discontinuities whose actual reflection coefficients are 0.3, 0.2, and 0.1.

## G.6 SAMPLE SPICE LIST

The SPICE listing below can be used to find the $S_{11}(= V(11)$ and $S_{21}(= V(21))$ parameters when the source and load resistances are 50 Ω, and the circuit is terminated with a 50 Ω load. These may be changed in the PARAM statement. The circuit to be analyzed is entered in the SUBCKT section. The number of frequencies and the frequency range must also be added to the ac statement. The following net list was used in PSPICE. Other versions of SPICE may require some minor modifications as noted in the net list.

```
Analysis of a circuit for S11 and S21
*
* RO1 and RO2 are input and output resistance levels.
* RL is the load resistance. The load may be
   supplemented
* with additional elements.
.PARAM RO1=50, RO2=50. RLOAD=50. IIN=-1/RO1
.FUNC N(RO1,RO2) SQRT(RO2/RO1)
RO1   1    0   RO1
VIN   10   11  AC   1
GI1   1    0    VALUE=-V(10,11)/RO1
*GI1   1    0   10    11   "-1/RO1"
E11   10   0    1   0   2
R11   11   0    1
```

```
Xcircuit   1    2      netname
RL     2    0   RLOAD
E21    21   0   VALUE=V(2)*2/N(R01,R02)
*  n = SQRT(R02/R01)
*E21    21   0   2  0   "2/n"
R21    21   0   1
*
.SUBCKT netname   "first_node"    "last_node"
* Input side
*   .
*   .
*   .
* Output side
.ENDS   netname
* Code for S11 and S21
*.AC DEC "num"  "f1"   "f2"
.PROBE V(11)  V(21)
.END
```

## G.7 IMPULSE RESPONSE SPICE NET LIST MODIFICATION

Time domain analysis with SPICE requires replacing the AC statement with a TRAN statement similar to the following:

```
.TRAN .01ps 250ps 0 .2ps
```

In addition the VIN statement should be replaced with one containing either the PULSE or the PWL transient function. The pulse statement has the form PULSE(initial volt, pulse volt, delay time, rise time, fall time, pulse width, period). The measured "impulse" from the time domain measurements from the network analyzer is approximated by the following PULSE statements. The 18 GHz pulse, with base width of 94.019 ps, is approximated as follows:

```
VIN 10 11 PULSE(0 1 0 39.482p 39.491p 15.05p)
```

The 26 GHz pulse, with a base width of 64.150 ps, is approximated as follows:

```
VIN 10 11 PULSE(0 1 0 26.395p 28.477p 9.278p)
```

Alternately, the more accurate piecewise linear fit could be used. The 18 GHz "impulse" is approximated using 81 points:

```
VIN 10 11 PWL(0ps 0, 1.5ps 0.0037, 3.0ps 0.0089,
   4.5ps 0.0156,
```

```
+ 6.0ps 0.0237, 7.5ps 0.0336, 9.0ps 0.0452,
   10.5ps 0.0589,
+ 12.0ps 0.0745, 13.5ps 0.0922, 15.0ps 0.1121,
   16.5ps 0.1342,
+ 18.0ps 0.1584, 19.5ps 0.1848, 21.0ps 0.2134,
   22.5ps 0.2441,
+ 24.0ps 0.2767, 25.5ps 0.3111, 27.0ps 0.3472,
   28.5ps 0.3848,
+ 30.0ps 0.4236, 31.5ps 0.4634, 33.0ps 0.5038,
   34.5ps 0.5448,
+ 36.0ps 0.5858, 37.5ps 0.6267, 39.0ps 0.6671,
   40.5ps 0.7065,
+ 42.0ps 0.7448, 43.5ps 0.7815, 45.0ps 0.8162,
   46.5ps 0.8489,
+ 48.0ps 0.8789, 49.5ps 0.9062, 51.0ps 0.9304,
   52.5ps 0.9513,
+ 54.0ps 0.9687, 55.5ps 0.9824, 57.0ps 0.9922,
   58.5ps 0.9983,
+ 60.0ps 1.0002, 61.5ps 0.9982, 63.0ps 0.9922,
   64.5ps 0.9823,
+ 66.0ps 0.9686, 67.5ps 0.9512, 69.0ps 0.9303,
   70.5ps 0.9061,
+ 72.0ps 0.8788, 73.5ps 0.8487, 75.0ps 0.8161,
   76.5ps 0.7813,
+ 78.0ps 0.7446, 79.5ps 0.7064, 81.0ps 0.6669,
   82.5ps 0.6266,
+ 84.0ps 0.5857, 85.5ps 0.5446, 87.0ps 0.5037,
   88.5ps 0.4633,
+ 90.0ps 0.4235, 91.5ps 0.3847, 93.0ps 0.3471,
   94.5ps 0.3111,
+ 96.0ps 0.2767, 97.5ps 0.2441, 99.0ps 0.2135,
   100.5ps 0.1849,
+ 102.0ps 0.1585, 103.5ps 0.1342, 105.0ps 0.1122,
   106.5ps 0.0923,
+ 108.0ps 0.0746, 109.5ps 0.0591, 111.0ps 0.0455,
   112.5ps 0.0338,
+ 114.0ps 0.0240, 115.5ps 0.0158, 117.0ps 0.0092,
   118.5ps 0.0040,
+ 120.0ps 0)
```

The piecewise linear fit for the 26 GHz "impulse" is approximated using 77 points:

```
VIN 10 11 PWL(0ps 0,1ps .005, 2ps .015, 3ps .0267,
   4ps .0402,
```

```
+ 5ps .0556, 6ps .0731, 7ps .0925, 8ps .1140,
   9ps .1375,
+ 10ps .1632, 11ps .1909, 12ps .2204, 13ps .2519,
   14ps .2850,
+ 15ps .3198, 16ps .3560, 17ps .3933, 18ps .4318,
   19ps .4709,
+ 20ps .5106, 21ps .5505, 22ps .5904, 23ps .6299,
   24ps .6688,
+ 25ps .7067, 26ps .7433, 27ps .7784, 28ps .8116,
   29ps .8427,
+ 30ps .8713, 31ps .8972, 32ps .9202, 33ps .9401,
   34ps .9566,
+ 35ps .9697, 36ps .9792, 37ps .9850, 38ps .9871,
   39ps .9855,
+ 40ps .9801, 41ps .9710, 42ps .9584, 43ps .9423,
   44ps .9227,
+ 45ps .9001, 46ps .8745, 47ps .8462, 48ps .8155,
   49ps .7825,
+ 50ps .7477, 51ps .7112, 52ps .6734, 53ps .6346,
   54ps .5952,
+ 55ps .5553, 56ps .5154, 57ps .4756, 58ps .4363,
   59ps .3979,
+ 60ps .3604, 61ps .3240, 62ps .2891, 63ps .2557,
   64ps .2240,
+ 65ps .1942, 66ps .1663, 67ps .1404, 68ps .1166,
   69ps .0948,
+ 70ps .0751,71ps .0575, 72ps .0418, 73ps .0279,
   74ps .0160,
+ 75ps .0058, 76ps 0)
```

The piecewise linear fit for the 50 GHz "impulse" is approximated using 46 points:

```
VIN 10 11 PWL( 0ps -4.530E-03, 1ps -611.4E-06,
   2ps 6.941E-03,
+ 3ps 0.019, 4ps 0.037, 5ps 0.061, 6ps 0.091,
   7ps 0.130,
+ 8ps 0.175, 9ps 0.229, 10s 0.289, 11ps 0.355,
   12ps 0.425,
+ 13ps 0.500, 14ps 0.576, 15ps 0.651, 16ps 0.724,
   17ps 0.792,
+ 18ps 0.854, 19ps 0.906, 20ps 0.948, 21ps 0.979,
   22ps 0.996,
+ 23ps 1.000, 24ps 0.990, 25ps 0.967, 26ps 0.931,
   27ps 0.884,
```

```
+ 28ps 0.828, 29ps 0.763, 30ps 0.693, 31ps 0.618,
   32ps 0.542,
+ 33ps 0.467, 34ps 0.394, 35ps 0.325, 36ps 0.262,
   37ps 0.204,
+ 38ps 0.155, 39ps 0.112, 40ps 0.077, 41ps 0.049,
   42ps 0.028,
+ 43ps 0.013, 44ps 3.141E-03, 45ps -2.711E-03)
```

## ACKNOWLEDGMENT

The authors wish to acknowledge the help of Terry Jamison in the analysis of the impulse response data.

## REFERENCES

1. M. E. Hines and H. E. Stinehelfer Sr, "Time-Domain Oscillographic Microwave Network Analysis Using Frequency-Domain Data," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-22, pp. 276–282, 1974.
2. T. B. Mills, "*S*-Parameters in Spice," *RF Design*, pp. 45–48, 1989.
3. K. B. Kumar and T. Wong, "Methods to Obtain *Z*, *Y*, *H*, *G* and *S* Parameters from the SPICE Program," *IEEE Circuits Devices*, Vol. 4, pp. 30–31, 1988.
4. R. Goyal, "*S*-Parameter Output from the SPICE Program,"*IEEE Circuits Devices*, Vol. 4, pp. 28–30, 1988.
5. C. E. Smith, "Frequency Domain Analysis of RF and Microwave Circuits Using SPICE," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-42, pp. 1904–1909, 1994.
6. Ke Lu and T. J. Brazil, "A Systematic Error Analysis of HP8510 Time-Domain Gating Techniques with Experimental Verification," *1993 IEEE MTT-S Microwave Theory Tech. Symp. Digest*, pp. 1259–1262, 1993.

*Radio Frequency Circuit Design*. W. Alan Davis, Krishna Agarwal

Print ISBN 0-471-35052-4 Electronic ISBN 0-471-20068-9

*APPENDIX H*

# Single-Tone Intermodulation Distortion Suppression for Double-Balanced Mixers

An expression is given in [1] for the suppression for single-tone intermodulation distortion in a double-balanced mixer. This is repeated below as well as coded in the program IMSUP. The intermodulation suppression in dBc (dB below the carrier) is $S_{\text{nm}}$ for a set of frequencies $nf_{\text{p}} \pm mf_1$.

$$S_{\text{nm}} = (|m| - 1)\Delta P + 20\log(|A_{\text{nm}}|) \tag{H.1}$$

$\Delta P$ is the difference in dB of RF signal and LO powers.

$$A'_{\text{nm}} = \frac{\Gamma\left(\frac{|n| + |m| - 1}{2}\right)}{\Gamma\left(\frac{|n| - |m| + 3}{2}\right)}\frac{1}{2}\left[\sin\frac{|n|\pi}{2}\sin\frac{|m|\pi}{2}B_{\text{oo}} + \cos\frac{|n|\pi}{2}\cos\frac{|m|\pi}{2}B_{\text{ee}}\right]$$
$$+ \frac{\Gamma\left(\frac{|n| + |m|}{2}\right)}{\Gamma\left(\frac{|n| - |m| + 2}{2}\right)}V_f\left[\sin\frac{|n|\pi}{2}\cos\frac{|m|\pi}{2}B_{\text{oe}} + \cos\frac{|n|\pi}{2}\sin\frac{|m|\pi}{2}B_{\text{eo}}\right] \tag{H.2}$$

$$A_{\text{nm}} = \frac{1}{B_{IF}|m|!}A'_{\text{nm}} \tag{H.3}$$

$$B_{\text{oo}} = 1 + \delta_4 + \alpha(\delta_3 + \delta_2) - |m|[\delta_4 - \delta_2 - \alpha(\delta_3 + \delta_2) - \beta(\delta_3 + \delta_4)]$$

$$B_{\text{ee}} = -1 + \delta_4 - \alpha(\delta_3 - \delta_2) - |m|[\delta_4 - \delta_2 - \alpha(\delta_3 - \delta_2) + \beta(\delta_3 - \delta_4)]$$

$$B_{oe} = |m|[-\delta_4 - \delta_2 + \alpha(\delta_3 + \delta_2) + \beta(\delta_4 - \delta_3)]$$

$$B_{eo} = |m|[\delta_4 + \delta_2 + \alpha(\delta_3 - \delta_2) - \beta(\delta_4 + \delta_3)]$$

$$B_{IF} = B_{oo}(m = 1)$$

The $\alpha$ and $\beta$ represent the isolation in the LO and RF transformers resulting from their imbalance. The imbalance is illustrated in Fig. H.1 which is the same as used in Fig. 11.5 except the diode numbering convention has been made here to conform to that used in [1].

$$\text{LO isolation} = 20\log(1 - \alpha)$$

$$\text{RF isolation} = 20\log(1 - \beta)$$

The values for $\delta$ are a measure of the inequality of the forward voltages across the diodes:

$$\delta_2 = \frac{V_2}{V_1}$$

$$\delta_3 = \frac{V_3}{V_1}$$

$$\delta_4 = \frac{V_4}{V_1}$$

Under ideal conditions

$$\alpha = \beta = \delta_2 = \delta_3 = \delta_4 = 1$$

Typical values for isolation by the transformers are 10 to 15 dB, while the values for $\delta$ range from 0.85 to 1.15.

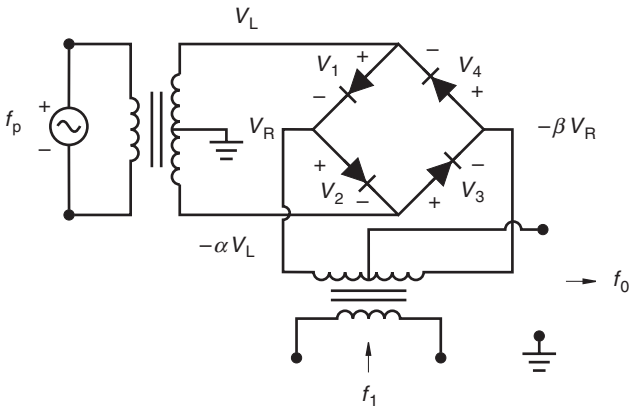



**FIGURE H.1** Double-balanced mixer with transformer and diode imbalance.

The equations have been implemented in a program called IMPSUP, which determines the single-tone intermodulation suppression for a given set of frequency harmonics of the RF signal and LO oscillator, the relative RF signal and LO power levels, the peak value of the LO voltage, imbalances resulting in finite isolation in the transformers, and imbalances in the diode forward voltage drops. A sample run of IMSUP shows the intermodulation suppression for a variety of frequency harmonics:

```
LO and RF Signal Transformer Isolation (typ.
 10 to 15 dB)
 10.,10.
Ring diode voltage ratios: V2/V1, V3/V1, V4/V1 = ?
Typically .85 to 1.15 (ideally =1)
 0.85, 0.90, 1.15
 Difference in LO and RF power in dB (typ. -20.)
 -20.
 Peak LO voltage = ?
 3.
Forward diode saturation voltage (typ. 0.1)
 IM product n x FL +- m x Frf: n,m = ?
 1, 1
 For intermodulation product n x m = 1  1
 IM Suppression =  0.000000E+00 dBc
 New n,m values only? <Y/N>
 y
 IM product n x FL +- m x Frf: n,m = ?
  2, 1
For intermodulation product n x m = 2  1
 IM Suppression = -0.437641E+02 dBc
 New n,m values only? <Y/N>
 y
 IM product n x FL +- m x Frf: n,m = ?
 3, 1
 For intermodulation product n x m = 3  1
 IM Suppression = -0.954243E+01 dBc
 New n,m values only? <Y/N>
 y
 IM product n x FL +- m x Frf: n,m = ?
 3, 2
For intermodulation product n x m = 3  2
 IM Suppression = -0.602423E+02 dBc
 New n,m values only? <Y/N>
 n
Completely new mixer specs? <Y/N>
 n
fin
```

## REFERENCES

1. B. C. Henederson, "Reliably Predict Mixer IM Suppression," *Microwaves RF*, Vol. 22, pp. 63–66, 68–70, 132, 1983. Also reprinted in *RF and Microwave Designer's Handbook*, Watkins-Johnson, 1997–98 Catalog.