A CLASSIC REISSUE

Foundations for Microwave Engineering

SECOND EDITION

Robert E. Collin

The IEEE Press Series on Electromagnetic Wave Theory
Donald G. Dudley, Series Editor

Foundations for Microwave Engineering

SECOND EDITION

Donald G. Dudley, *Series Editor*

Foundations for Microwave Engineering, Second Edition, covers the major topics of microwave engineering. Its presentation defines the accepted standard for both advanced undergraduate and graduate level courses on microwave engineering. An essential reference book for the practicing microwave engineer, it features:

- Planar transmission lines, as well as an appendix that describes in detail conformal mapping methods for their analysis and attenuation characteristics.
- Small aperture coupling and its application in practical components such as directional couplers and cavity coupling.
- Printed circuit components with an emphasis on techniques such as even and odd mode analysis and the use of symmetry properties.
- Microwave linear amplifier and oscillator design using solid-state circuits such as varactor devices and transistors.

Foundations for Microwave Engineering, Second Edition, has extensive coverage of transmission lines, waveguides, microwave circuit theory, impedance matching, and cavity resonators. It devotes an entire chapter to fundamental microwave tubes, as well as other chapters on periodic structures, microwave filters, small signal solid-state microwave amplifier and oscillator design, and negative resistance devices and circuits. Completely updated in 1992, it is being reissued by the IEEE Press in response to requests from our many members, who found it an invaluable textbook and an enduring reference for practicing microwave engineers.

About the Author

Robert E. Collin is the author or coauthor of more than 150 technical papers and five books on electromagnetic theory and applications. His classic text, *Field Theory of Guided Waves*, is also a volume in the series. Professor Collin has had a long and distinguished academic career at Case Western Reserve University. In addition to his professorial duties, he has served as chairman of the Department of Electrical Engineering and as interim dean of engineering.

Professor Collin is a life fellow of the IEEE and a member of the Microwave Theory and Techniques Society and the Antennas and Propagation Society (APS). He is a member of the U.S. Commission B of URSI and a member of the Geophysical Society. Other honors include the Diekman Award from Case Western Reserve University for distinguished graduate teaching, the IEEE APS Distinguished Career Award (1992), the IEEE Schelkunoff Prize Paper Award (1992), the IEEE Electromagnetics Award (1998), and an IEEE Third Millennium Medal in 2000. In 1990 Professor Collin was elected to the National Academy of Engineering.

The IEEE Press Series on Electromagnetic Wave Theory offers outstanding coverage of the field. It consists of new titles of contemporary interest, as well as reissues and revisions of recognized classics by established authors and researchers. The series emphasizes works of long-term archival significance in electromagnetic waves and applications. Designed specifically for graduate students, researchers and practicing engineers, the series provides affordable volumes that explore and explain electromagnetic waves beyond the undergraduate level.

IEEE
IEEE PRESS

WILEY-
INTERSCIENCE

ISBN 0-7803-6031-1

90000

9 780780 360310

An IEEE Press Classic Reissue

FOUNDATIONS FOR MICROWAVE ENGINEERING

SECOND EDITION

IEEE PRESS SERIES ON ELECTROMAGNETIC WAVE THEORY

The IEEE Press Series on Electromagnetic Wave Theory consists of new titles as well as reprintings and revisions of recognized classics that maintain long-term archival significance in electromagnetic waves and applications.

BOOKS IN THE IEEE PRESS SERIES ON ELECTROMAGNETIC WAVE THEORY

Christopoulos, C., *The Transmission-Line Modeling Methods: TLM*
Clemmow, P. C., *The Plane Wave Spectrum Representation of Electromagnetic Fields*
Collin, R. E., *Field Theory of Guided Waves*, Second Edition
Collin, R. E., *Foundations for Microwave Engineering, Second Edition*
Dudley, D. G., *Mathematical Foundations for Electromagnetic Theory*
Elliot, R. S., *Electromagnetics: History, Theory, and Applications*
Felsen, L. B., and Marcuvitz, N., *Radiation and Scattering of Waves*
Harrington, R. F., *Field Computation by Moment Methods*
Hansen et al., *Plane-Wave Theory of Time-Domain Fields: Near-Field Scanning Applications*
Ishimaru, A., *Wave Propagation and Scattering in Random Media*
Jones, D. S., *Methods in Electromagnetic Wave Propagation*, Second Edition
Lindell, I. V., *Methods for Electromagnetic Field Analysis*
Peterson et al., *Computational Methods for Electromagnetics*
Tai, C. T., *Generalized Vector and Dyadic Analysis: Applied Mathematics in Field Theory*
Tai, C. T., *Dyadic Green Functions in Electromagnetic Theory*, Second Edition
Van Bladel, J., *Singular Electromagnetic Fields and Sources*
Volakis et al., *Finite Element Method for Electromagnetics: Antennas, Microwave Circuits, and Scattering Applications*
Wait, J., *Electromagnetic Waves in Stratified Media*

An IEEE Press Classic Reissue

FOUNDATIONS FOR MICROWAVE ENGINEERING

SECOND EDITION

IEEE Press Series on
Electromagnetic Wave Theory

Robert E. Collin
Professor of Electrical Engineering
Case Western Reserve University
Cleveland, OH

IEEE Antennas & Propagation Society, *Sponsor*

IEEE Microwave Theory and Techniques Society, *Sponsor*

IEEE
The Institute of Electrical and Electronics Engineers, Inc., New York

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

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

Printed in the United States of America
10 9 8 7 6 5 4 3 2 **ISBN 0-7803-6031-1**

Library of Congress Cataloging-in-Publication Data

Collin, Robert E.
Foundations for microwave engineering / Robert E. Collin.-- 2nd ed.
p. cm. – (IEEE Press series on electromagnetic wave theory)
Originally published : New York : McGraw Hill, c1992.
"An IEEE Press classic reissue."
Includes bibliographical references and index.
ISBN 0-7803-6031-1
1. Microwave devices. I. Title. II. Series.

TK7876 .C645 2000
621.381'3--dc21

00-053874

FOREWORD TO THE REISSUED EDITION

The purpose of the IEEE Press Series on Electromagnetic Wave Theory is to publish books of long-term archival significance in electromagnetics. Included are new titles as well as reprints and revisions of recognized classics. The book *Foundations for Microwave Engineering*, by Robert E. Collin, is by any measure such a classic. The original edition of the book appeared in 1966 and remained in print until the appearance of the second edition in 1992, a span of 26 years.

In the second edition, Professor Collin completely updated and modernized his book to include the many advances that had occurred in microwave engineering since the appearance of the original edition. That the second edition has gone out of print has caused concern among many of my colleagues in the IEEE Antennas and Propagation Society (APS) and the IEEE Microwave Theory and Techniques Society (MTT). We at the IEEE Press are delighted to be able to overcome this difficulty by introducing a reprint of the second edition into our Series on Electromagnetic Wave Theory. The book is a thorough and in-depth exposition on microwave engineering. Furthermore, it will make an excellent companion to Professor Collin's book, *Field Theory of Guided Waves*, also included in the series.

Professor Collin has been a valued colleague for many years. He is the author or coauthor of five books and more than 150 technical papers. His contributions to electromagnetics span a wide range of subjects and have brought him international respect and many awards. Among these are election to the National Academy of Engineering, the IEEE Electromagnetics Field Award, the IEEE/APS Distinguished Career Award, an IEEE/APS Schelkunoff Prize Paper Award, and the IEEE Third Millennium Medal.

It is with pleasure that I welcome this book into the series.

Donald G. Dudley
University of Arizona
Series Editor
IEEE Press Series on Electromagnetic Wave Theory

IEEE Press
445 Hoes Lane, P.O. Box 1331
Piscataway, NJ 08855-1331

CONTENTS

CHAPTER 1

INTRODUCTION

The purpose of this introductory chapter is to provide a short, and admittedly incomplete, survey of what the microwave engineering field encompasses. Section 1.2 presents a brief discussion of many of the varied and sometimes unique applications of microwaves. This is followed by a third section in which an attempt is made to show in what ways microwave engineering differs from the engineering of communication systems at lower frequencies. In addition, a number of microwave devices are introduced to provide examples of the types of devices and circuit elements that are examined in greater detail later on in the text.

1.1 MICROWAVE FREQUENCIES

The descriptive term *microwaves* is used to describe electromagnetic waves with wavelengths ranging from 1 cm to 1 m. The corresponding frequency range is 300 MHz up to 30 GHz for 1-cm-wavelength waves. Electromagnetic waves with wavelengths ranging from 1 to 10 mm are called millimeter waves. The infrared radiation spectrum comprises electromagnetic waves with wavelengths in the range 1 μm (10^{-6} m) up to 1 mm. Beyond the infrared range is the visible optical spectrum, the ultraviolet spectrum, and finally x-rays. Several different classification schemes are in use to designate frequency bands in the electromagnetic spectrum. These classification schemes are summarized in Tables 1.1 and 1.2. The radar band classification came into use during World War II and is still in common use today even though the new military band classification is the recommended one.

In the UHF band up to around a frequency of 1 GHz, most communications circuits are constructed using lumped-parameter circuit compo-

Appendixes

PREFACE

The first edition of *Foundations for Microwave Engineering* was published in 1966. The text has remained continuously in use since that time, but it has become clear that it no longer gives an adequate account of modern microwave engineering practice. Since the publication of the first edition there has been a dramatic advance in the microwave field brought about by the development of solid state transistors that can provide amplification and signal generation well into the millimeter wavelength region. Along with the widespread use of solid state devices, compatible transmission line structures and passive components were developed that could be integrated with the solid state devices into compact miniaturized microwave systems. These developments made it mandatory that the text be thoroughly revised if it were to continue serving the needs of the student and the practicing microwave engineer.

In the revised addition I have adhered to the same general philosophy that governed the preparation of the first edition. Fundamental principles are stressed and complete derivations are provided for all significant formulas and relationships. All important fundamental concepts and principles are covered to the extent possible within a text of reasonable size. The applications of basic theory and principles are illustrated through detailed analysis of a large number of important components that find widespread use in practical microwave systems.

Chapter 1 is an updated introductory chapter. Chapter 2 is essentially the same as in the original edition and provides a comprehensive summary of basic electromagnetic theory that is needed as background for proper understanding of the rest of the text. Many students will already have knowledge of this material before they pursue a course in microwave engineering. For these students, Chapter 2 will serve as a concise reference or review of familiar material.

Chapter 3 is very different from that in the first edition. The first part of this chapter provides a more basic introduction to waves on transmission

lines using distributed circuit models. The propagation of pulse signals is also covered. The second part of this chapter is a long section covering the characteristics of planar transmission lines, such as microstrip lines, coupled microstrip lines, strip lines, and coplanar lines or waveguides. The treatment is considerably broader than what is available in any other current text. Most of the formulas for the quasi-TEM mode parameters are derived using conformal mapping methods in a new Appendix III and are not just quoted from the literature. Several new formulas for attenuation have been derived as well as suitable modifications of existing formulas to account for anisotropic substrates. The last part of the chapter covers the basic properties of rectangular and circular waveguides, as in the original edition.

Chapter 4 develops the basic microwave circuit theory and includes detailed discussions of the impedance, admittance and scattering matrix descriptions of microwave junctions. New material has been added on signal flow graphs and the generalized scattering matrix for power waves. The material on small aperture coupling has been updated to include radiation reaction that will account for power transmission through an aperture and thereby lead to physically meaningful equivalent circuits for small apertures.

Chapter 5 treats a number of topics related to impedance matching and transformations. The old topic of impedance matching with lumped reactive elements has been revived because this is now frequently used in microwave integrated circuits. The design of complex load terminations has also been included because this is required for microwave solid state amplifier design. The available power at any point in a lossless reciprocal network is an invariant quantity. This concept is explained in terms of the impedance mismatch factor. The invariance of the impedance mismatch factor places an important constraint on the design of interstage matching networks in a microwave amplifier and is used in Chapter 10 in the design of microwave amplifiers. The last part of Chapter 5 discusses multisection quarter-wave transformers and tapered transmission lines. A new example of a microstrip half-wave filter design based on the quarter-wave transformer as a prototype circuit has been included.

A variety of passive components are described along with detailed analysis in Chapter 6. In addition to those components described in the original edition, new material has been added on coupled-microstrip-line directional couplers, the branch-line coupler, hybrid junctions, and the Wilkinson power divider. New material on electronic controlled attenuators and phase shifters has also been added.

Chapter 7 on resonators has been expanded to include new material on microstrip resonators and dielectric resonators. The old material on Fabry–Perot resonators has been deleted in order to make room for a short section on cavity perturbation theory.

Chapter 8 on periodic structures and filters now includes a detailed treatment of gap-coupled and edge-coupled microstrip filters. The treatment

of admittance and impedance inverters was rewritten in order to more fully explain the use of inverters in filter design.

Apart from a brief discussion of gyratron tubes, Chapter 9 on microwave tubes remains essentially the same as in the first edition.

The old Chapter 10 on masers has been replaced by a new chapter on microwave solid state amplifier design. This chapter gives a complete discussion of the scattering matrix approach to small signal narrow band amplifier design. The treatment is self-contained and all important relations for gain, stability, and low noise design are derived. A design strategy for low noise single stage and double stage amplifiers is developed along with considerations for the necessary tradeoffs that must be made between input and output VSWR, gain, low noise figure, and stability.

The original Chapter 11 on parametric amplifiers has been retained without any change.

A new Chapter 12 on oscillators and mixers has been added. This chapter is of limited scope because of the need to keep the overall length of the text within reasonable bounds. Solid state oscillators using Gunn devices and IMPATT diodes are described in a qualitative way only. An introduction to transistor oscillator design based on small signal scattering matrix parameters is provided. Included in this discussion is the relationship between the two-port and three-port scattering matrix description of a transistor because this is needed in order to efficiently analyze the effect of an impedance inserted in series with one of the transistor leads for feedback purposes.

Many textbooks provide introductory treatments of diode mixers without any consideration of the embedding network. Such treatments do not provide a good understanding of diode mixers because it is the impedance properties of the embedding network that determine the diode voltages at the various harmonic frequencies. The introductory treatment of diode mixers in Chapter 12 does include the embedding network and this should provide the student with a more complete understanding of mixer analysis and design. The last part of the chapter describes the harmonic balancing method for the analysis of mixers.

I have tried to provide a broad, comprehensive, and self-contained treatment of the fundamental theory and principles, and the methods of analysis and design that are the foundations for microwave engineering. There are, of course, limitations because all books must have a finite length. Many references have been included for the benefit of the reader who wishes to pursue a given topic in greater depth or refer to the original papers that a lot of the material has been based on. This text, in many respects, is a compilation of the work of a great many people. Unfortunately, it has not been possible to always give proper credit to those who were the originators of new concepts and the inventors of new devices.

It is my belief that the revised edition will prove to be useful for both senior elective as well as beginning graduate level courses in microwave engineering, and will also serve as a useful reference source on fundamental

TABLE 1.1
Frequency band designation

Frequency band	Designation	Typical service
3–30 kHz	Very low frequency (VLF)	Navigation, sonar
30–300 kHz	Low frequency (LF)	Radio beacons, navigational aids
300–3,000 kHz	Medium frequency (MF)	AM broadcasting, maritime radio, Coast Guard communication, direction finding
3–30 MHz	High frequency (HF)	Telephone, telegraph, and facsimile; shortwave international broadcasting; amateur radio; citizen's band; ship-to-coast and ship-to-aircraft communication
30–300 MHz	Very high frequency (VHF)	Television, FM broadcast, air-traffic control, police, taxicab mobile radio, navigational aids
300–3,000 MHz	Ultrahigh frequency (UHF)	Television, satellite communication, radiosonde, surveillance radar, navigational aids
3–30 GHz	Superhigh frequency (SHF)	Airborne radar, microwave links, common-carrier land mobile communication, satellite communication
30–300 GHz	Extreme high frequency (EHF)	Radar, experimental

TABLE 1.2
Microwave frequency band designation

	Microwave band designation	
Frequency	Old	New
500–1,000 MHz	VHF	C
1–2 GHz	L	D
2–3 GHz	S	E
3–4 GHz	S	F
4–6 GHz	C	G
6–8 GHz	C	H
8–10 GHz	X	I
10–12.4 GHz	X	J
12.4–18 GHz	Ku	J
18–20 GHz	K	J
20–26.5 GHz	K	K
26.5–40 GHz	Ka	K

principles for the practicing microwave engineer. There is clearly much more material in the revised edition than can be covered in a one semester course. The last four chapters alone would provide sufficient material for a one semester course on active microwave circuits.

As an instructor I have always believed that it was very important to fully understand where formulas came from and how they are derived in order to present the material to students in a meaningful way. It is for this reason that I have attempted to make the text self-contained. In presenting many of the topics to undergraduate students I will only outline the basic approach used and will omit the details. It is my hope that other instructors will also view the detailed derivations that are provided in the text as a useful source of information in preparing a microwave engineering course and not as material that must always be presented in class. A number of topics that can be omitted in an undergraduate course are identified by a star. The problems based on these sections are also identified by a star.

In recent years the microwave engineering course that I have taught to seniors at Case Western Reserve University has drawn heavily on the material in Chapters 3 through 5, which is very basic core material. In addition, topics have been selected from Chapters 6 and 7 on components and resonators in order to illustrate the application of basic microwave circuit theory. The last quarter of the semester has been generally devoted to microwave solid state amplifier design along with a brief coverage of oscillators and mixers.

A better selection of problems and a solutions manual has been prepared for the revised edition. Over the past several years I have also prepared a number of short stand alone computer programs that provide useful tools to remove the drudgery of solving many of the homework problems. These programs are included on a floppy disk along with user instructions as part of the solutions manual. The programs cover the calculation of the characteristics of various planar transmission lines, including attenuation; the cutoff frequency, propagation constant, and attenuation of the dominant mode in rectangular and circular waveguides; impedance transformation along a transmission line; input and output impedances, admittances, and reflection coefficients for a linear two-port, which can be described in terms of impedance, admittance, or scattering matrix parameters; double-stub and lumped element impedance matching with frequency scans; two-port and three-port scattering matrix parameters for a transistor; and a rather long program that implements a design strategy for low noise one- and two-stage microwave amplifiers with various imposed constraints. Students have generally found these programs to be of significant help in problem solving. They have enjoyed working with the microwave amplifier design program. Without a computer program, the design of a microwave amplifier using potentially unstable devices and subject to various constraints on gain, noise figure, and input and output VSWR, is not feasible for students to carry out. The scope of each program

has been purposefully limited in order to ensure that the student will be fully aware of the solution strategy involved.

Many users of the first edition have provided me with helpful comments on the original material. In addition, I have received many helpful comments and suggestions from the following reviewers of the material for the revised edition. They are Chin-Lin Chen, Purdue University; M. Yousif El-Ibiary, University of Oklahoma; Irving Kaufman, Arizona State University; Stuart Long, University of Houston; Glenn S. Smith, Georgia Institute of Technology; and Robert J. Weber, Iowa State University. For the most part their suggestions and recommendations have been incorporated.

The new material for the revised edition was typed by Sue Sava. I would like to acknowledge the professional skill with which she prepared this material as well as her willingness to rearrange her schedule so as to meet various deadlines.

The last acknowledgment is to my wife Kathleen, who was willing to give up many other activities so that the revision could be carried out. Her encouragement and support of the project never faltered, and without it the revision could not have been undertaken.

Robert E. Collin

nents. In the frequency range from 1 up to 100 GHz, lumped circuit elements are usually replaced by transmission-line and waveguide components. Thus by the term *microwave engineering* we shall mean generally the engineering and design of information-handling systems in the frequency range from 1 to 100 GHz corresponding to wavelengths as long as 30 cm and as short as 3 mm. At shorter wavelengths we have what can be called optical engineering since many of the techniques used are derived from classical optical techniques. The characteristic feature of microwave engineering is the short wavelengths involved, these being of the same order of magnitude as the circuit elements and devices employed.

The short wavelengths involved in turn mean that the propagation time for electrical effects from one point in a circuit to another point is comparable with the period of the oscillating currents and charges in the system. As a result, conventional low-frequency circuit analysis based on Kirchhoff's laws and voltage-current concepts no longer suffices for an adequate description of the electrical phenomena taking place. It is necessary instead to carry out the analysis in terms of a description of the electric and magnetic fields associated with the device. In essence, it might be said, microwave engineering is applied electromagnetic fields engineering. For this reason the successful engineer in this area must have a good working knowledge of electromagnetic field theory.

There is no distinct frequency boundary at which lumped-parameter circuit elements must be replaced by distributed circuit elements. With modern technological processes it is possible to construct printed-circuit inductors that are so small that they retain their lumped-parameter characteristics at frequencies as high as 10 GHz or even higher. Likewise, optical components, such as parabolic reflectors and lenses, are used to focus microwaves with wavelengths as long as 1 m or more. Consequently, the *microwave engineer will frequently employ low-frequency lumped-parameter* circuit elements, such as miniaturized inductors and capacitors, as well as optical devices in the design of a microwave system.

1.2 MICROWAVE APPLICATIONS

The great interest in microwave frequencies arises for a variety of reasons. Basic among these is the ever-increasing need for more radio-frequency-spectrum space and the rather unique uses to which microwave frequencies can be applied. When it is noted that the frequency range 10^9 to 10^{12} Hz contains a thousand sections like the frequency spectrum from 0 to 10^9 Hz, the value of developing the microwave band as a means of increasing the available usable frequency spectrum may be readily appreciated.

At one time (during World War II and shortly afterward), microwave engineering was almost synonymous with radar (*RA*dio *D*etection *A*nd *R*anging) engineering because of the great stimulus given to the development of microwave systems by the need for high-resolution radar capable of

detecting and locating enemy planes and ships. Even today radar, in its many varied forms, such as missile-tracking radar, fire-control radar, weather-detecting radar, missile-guidance radar, airport traffic-control radar, etc., represents a major use of microwave frequencies. This use arises predominantly from the need to have antennas that will radiate essentially all the transmitter power into a narrow pencil-like beam similar to that produced by an optical searchlight. The ability of an antenna to concentrate radiation into a narrow beam is limited by diffraction effects, which in turn are governed by the relative size of the radiating aperture in terms of wavelengths. For example, a parabolic reflector-type antenna produces a pencil beam of radiated energy having an angular beam width of $140^\circ/(D/\lambda_0)$, where D is the diameter of the parabola and λ_0 is the wavelength. A 90-cm (about 3 ft) parabola can thus produce a 4.7° beam at a frequency of 10^{10} Hz, i.e., at a wavelength of 3 cm. A beam of this type can give reasonably accurate position data for a target being observed by the radar. To achieve comparable performance at a frequency of 100 MHz would require a 300-ft parabola, a size much too large to be carried aboard an airplane.

In more recent years microwave frequencies have also come into widespread use in communication links, generally referred to as microwave links. Since the propagation of microwaves is effectively along line-of-sight paths, these links employ high towers with reflector or lens-type antennas as repeater stations spaced along the communication path. Such links are a familiar sight to the motorist traveling across the country because of their frequent use by highway authorities, utility companies, and television networks. A further interesting means of communication by microwaves is the use of satellites as microwave relay stations. The first of these, the Telstar, launched in July 1962, provided the first transmission of live television programs from the United States to Europe.

Since that time a large number of satellites have been deployed for communication purposes, as well as for surveillance and collecting data on atmospheric and weather conditions. For direct television broadcasting the most heavily used band is the C band. The up-link frequency used is in the 5.9- to 6.4-GHz band and the receive or down-link frequency band is between 3.7 and 4.2 GHz. For home reception an 8-ft-diameter parabolic reflector antenna is commonly used. A second frequency band has also been allocated for direct television broadcasting. For this second band the up-link frequency is in the 14- to 14.5-GHz range and the down-link frequencies are between 10.95 and 11.2 GHz and 11.45 and 11.7 GHz. In this band a receiving parabolic antenna with a 3-ft diameter is adequate. At the present time this frequency band is not being used to any great extent in the United States. It is more widely used in Europe and Japan.

Terrestrial microwave links have been used for many years. The TD-2 system was put into service in 1948 as part of the Bell Network. It operated in the 3.7- to 4.2-GHZ band and had 480 voice circuits, each occupying a

3.1-kHz bandwidth. In 1974, the TN-1 system operating in the 10.7- to 11.7-GHz band was put into operation. This system had a capacity of 1,800 voice circuits or one video channel with a 4.5-MHz bandwidth. Since that time the use of terrestrial microwave links has grown rapidly.

At the present time most communication systems are shifting to the use of digital transmission, i.e., analog signals are digitized before transmission. Microwave digital communication system development is progressing rapidly. In the early systems simple modulation schemes were used and resulted in inefficient use of the available frequency spectrum. The development of 64-state quadrature amplitude modulation (64-QAM) has made it possible to transmit 2,016 voice channels within a single 30-MHz RF channel. This is competitive with FM analog modulation schemes for voice. The next step up is the 256-QAM system which is under development.

For the ready processing and handling of a modulated carrier, modulation sidebands can be only a few percent of the carrier frequency. It is thus seen that the carrier frequency must be in the microwave range for efficient transmission of many television programs over one link. Without the development of microwave systems, our communications facilities would have been severely overloaded and totally inadequate for present operations.

Even though such uses of microwaves are of great importance, the applications of microwaves and microwave technology extend much further, into a variety of areas of basic and applied research, and including a number of diverse practical devices, such as microwave ovens that can cook a small roast in just a few minutes. Some of these specific applications are briefly discussed below.

Waveguides periodically loaded with shunt susceptance elements support slow waves having velocities much less than the velocity of light, and are used in linear accelerators. These produce high-energy beams of charged particles for use in atomic and nuclear research. The slow-traveling electromagnetic waves interact very efficiently with charged-particle beams having the same velocity, and thereby give up energy to the beam. Another possibility is for the energy in an electron beam to be given up to the electromagnetic wave, with resultant amplification. This latter device is the traveling-wave tube, and is examined in detail in a later chapter.

Sensitive microwave receivers are used in radio astronomy to detect and study the electromagnetic radiation from the sun and a number of radio stars that emit radiation in this band. Such receivers are also used to detect the noise radiated from plasmas (an approximately neutral collection of electrons and ions, e.g., a gas discharge). The information obtained enables scientists to analyze and predict the various mechanisms responsible for plasma radiation. Microwave radiometers are also used to map atmospheric temperature profiles, moisture conditions in soils and crops, and for other remote-sensing applications as well.

Molecular, atomic, and nuclear systems exhibit various resonance phenomena under the action of periodic forces arising from an applied

electromagnetic field. Many of these resonances occur in the microwave range; hence microwaves have provided a very powerful experimental probe for the study of basic properties of materials. Out of this research on materials have come many useful devices, such as some of the nonreciprocal devices employing ferrites, several solid-state microwave amplifiers and oscillators, e.g., masers, and even the coherent-light generator and amplifier (laser).

The development of the laser, a generator of essentially monochromatic (single-frequency) coherent-light waves, has stimulated a great interest in the possibilities of developing communication systems at optical wavelengths. This frequency band is sometimes referred to as the *ultramicrowave* band. With some modification, a good deal of the present microwave technology can be exploited in the development of optical systems. For this reason, familiarity with conventional microwave theory and devices provides a good background for work in the new frontiers of the electromagnetic spectrum.

The domestic microwave oven operates at 2,450 MHz and uses a magnetron tube with a power output of 500 to 1000 W. For industrial heating applications, such as drying grain, manufacturing wood and paper products, and material curing, the frequencies of 915 and 2,450 MHz have been assigned. Microwave radiation has also found some application for medical hyperthermia or localized heating of tumors.

It is not possible here to give a complete account of all the applications of microwaves that are being made. The brief look at some of these, as given above, should convince the reader that this portion of the radio spectrum offers many unusual and unique features. Although the microwave engineering field may now be considered a mature and well-developed one, the opportunities for further development of devices, techniques, and applications to communications, industry, and basic research are still excellent.

1.3 MICROWAVE CIRCUIT ELEMENTS AND ANALYSIS

At frequencies where the wavelength is several orders of magnitude larger than the greatest dimensions of the circuit or system being examined, conventional circuit elements such as capacitors, inductors, resistors, electron tubes, and transistors are the basic building blocks for the information transmitting, receiving, and processing circuits used. The description or analysis of such circuits may be adequately carried out in terms of loop currents and node voltages without consideration of propagation effects. The time delay between cause and effect at different points in these circuits is so small compared with the period of the applied signal as to be negligible. It might be noted here that an electromagnetic wave propagates a distance of one wavelength in a time interval equal to one period of a sinusoidally

time-varying applied signal. As a consequence, when the distances involved are short compared with a wavelength λ_0 (λ_0 = velocity of light/frequency), the time delay is not significant. As the frequency is raised to a point where the wavelength is no longer large compared with the circuit dimensions, propagation effects can no longer be ignored. A further effect is the great relative increase in the impedance of connecting leads, terminals, etc., and the effect of distributed (stray) capacitance and inductance. In addition, currents circulating in unshielded circuits comparable in size with a wavelength are very effective in radiating electromagnetic waves. The net effect of all this is to make most conventional low-frequency circuit elements and circuits hopelessly inadequate at microwave frequencies.

If a rather general viewpoint is adopted, one may classify resistors, inductors, and capacitors as elements that dissipate electric energy, store magnetic energy, and store electric energy, respectively. The fact that such elements have the form encountered in practice, e.g., a coil of wire for an inductor, is incidental to the function they perform. The construction used in practical elements may be considered just a convenient way to build these devices so that they will exhibit the desired electrical properties. As is well known, many of these circuit elements do not behave in the desired manner at high frequencies. For example, a coil of wire may be an excellent inductor at 1 MHz, but at 50 MHz it may be an equally good capacitor because of the predominating effect of interturn capacitance. Even though practical low-frequency resistors, inductors, and capacitors do not function in the desired manner at microwave frequencies, this does not mean that such energy-dissipating and storage elements cannot be constructed at microwave frequencies. On the contrary, there are many equivalent inductive and capacitive devices for use at microwave frequencies. Their geometrical form is quite different, but they can be and are used for much the same purposes, such as impedance matching, resonant circuits, etc. Perhaps the most significant electrical difference is the generally much more involved frequency dependence of these equivalent inductors and capacitors at microwave frequencies.

Low-frequency electron tubes are also limited to a maximum useful frequency range bordering on the lower edge of the microwave band. The limitation arises mainly from the finite transit time of the electron beam from the cathode to the control grid. When this transit time becomes comparable with the period of the signal being amplified, the tube ceases to perform in the desired manner. Decreasing the electrode spacing permits these tubes to be used up to frequencies of a few thousand megahertz, but the power output is limited and the noise characteristics are poor. The development of new types of tubes for generation of microwave frequencies was essential to the exploitation of this frequency band. Fortunately, several new principles of operation, such as velocity modulation of the electron beam and beam interaction with slow electromagnetic waves, were discovered that enabled the necessary generation of microwaves to be carried out.

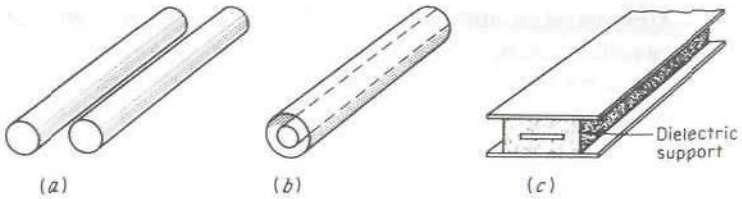


FIGURE 1.1
Some common transmission lines. (*a*) Two-conductor line; (*b*) coaxial line; (*c*) shielded strip line.

These fundamental principles with applications are discussed in a later chapter.

For low-power applications microwave tubes have been largely replaced by solid-state devices, such as transistors and negative resistance diodes. However, for high-power applications microwave tubes are still necessary.

One of the essential requirements in a microwave circuit is the ability to transfer signal power from one point to another without radiation loss. This requires the transport of electomagnetic energy in the form of a propagating wave. A variety of such structures have been developed that can guide electromagnetic waves from one point to another without radiation loss. The simplest guiding structure, from an analysis point of view, is the transmission line. Several of these, such as the open two-conductor line, coaxial line, and shielded strip line, illustrated in Fig. 1.1, are in common use at the lower microwave frequencies.

At the higher microwave frequencies, notably at wavelengths below 10 cm, hollow-pipe waveguides, as illustrated in Fig. 1.2, are often preferred to transmission lines because of better electrical and mechanical properties. The waveguide with rectangular cross section is by far the most common type. The circular guide is not nearly as widely used.

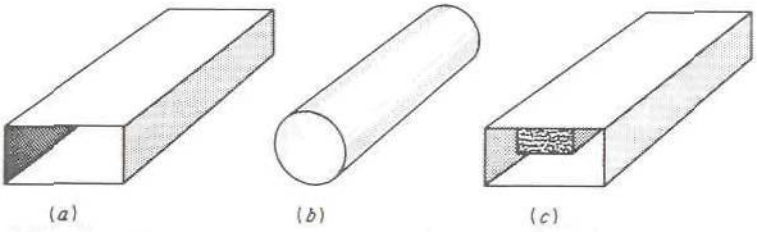


FIGURE 1.2
Some common hollow-pipe waveguides. (*a*) Rectangular guide; (*b*) circular guide; (*c*) ridge guide.

The ridge-loaded rectangular guide illustrated in Fig. 1.2*c* is sometimes used in place of the standard rectangular guide because of better impedance properties and a greater bandwidth of operation. In addition to these standard-type guides, a variety of other cross sections, e.g., elliptical, may also be used.

Another class of waveguides, of more recent origin, is surface waveguides. An example of this type is a conducting wire coated with a thin layer of dielectric. The wire diameter is small compared with the wavelength. Along a structure of this type it is possible to guide an electromagnetic wave. The wave is bound to the surface of the guide, exhibiting an amplitude decay that is exponential in the radial direction away from the surface, and hence is called a surface wave. Applications are mainly in the millimeter-wavelength range since the field does extend a distance of a wavelength or so beyond the wire, and this makes the effective guide diameter somewhat large in the centimeter-wavelength range. A disadvantage of surface waveguides and open-conductor transmission lines is that radiation loss occurs whenever other obstacles are brought into the vicinity of the guide.

The development of solid-state active devices, such as bipolar transistors and, more notably, field-effect transistors (FET), has had a dramatic impact on the microwave engineering field. With the availability of microwave transistors, the focus on waveguides and waveguide components changed to a focus on planar transmission-line structures, such as microstrip lines and coplanar transmission lines. These structures, shown in Fig. 1.3, can be manufactured using printed-circuit techniques. They are compatible with solid-state devices in that it is easy to connect a transistor to a microstrip circuit but difficult to incorporate it as part of a waveguide circuit. By using gallium-arsenide material it has been possible to design field-effect transistors that provide low noise and useful amplification at millimeter wavelengths. At the lower microwave frequencies hybrid integrated microwave circuits are used. In hybrid circuit construction the transmission lines and transmission-line components, such as matching elements, are manufactured first and then the solid-state devices, such as diodes and transistors, are soldered into place. The current trend is toward the use of monolithic microwave integrated circuits (MMIC) in which both the transmission-line circuits and active devices are fabricated on a single chip. A variety of broadband MMIC amplifiers have been designed. The development of MMIC circuits for operation at frequencies up to 100 GHz is well under way.

A unique property of the transmission line is that a satisfactory analysis of its properties may be carried out by treating it as a network with distributed parameters and solving for the voltage and current waves that may propagate along the line. Other waveguides, although they have several properties similar to transmission lines, must be treated as electromagnetic boundary-value problems, and a solution for the electromagnetic fields must be determined. Fortunately, this is readily accomplished for the common

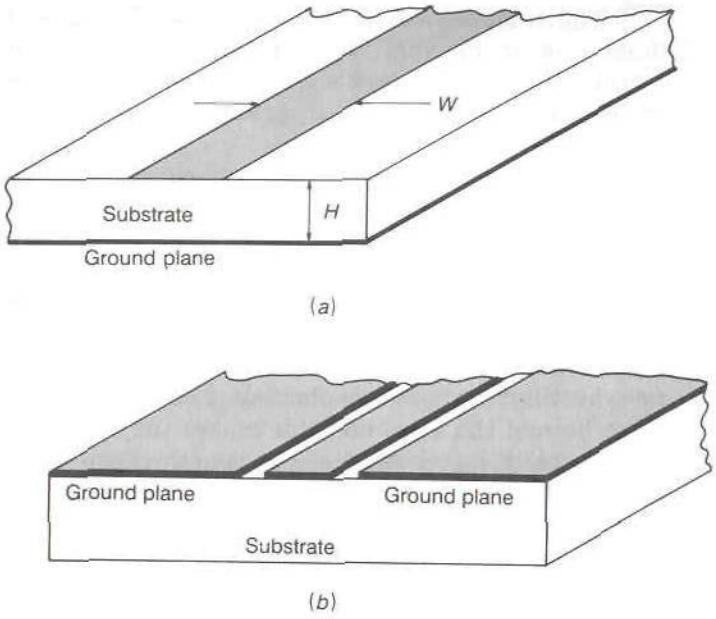


FIGURE 1.3
(*a*) microstrip transmission line; (*b*) coplanar transmission line.

waveguides used in practice. For waveguides it is not possible to define unique voltage and current that have the same significance as for a transmission line. This is one of the reasons why the field point of view is emphasized at microwave frequencies.

Associated with waveguides are a number of interesting problems related to methods of exciting fields in guides and methods of coupling energy out. Three basic coupling methods are used: (1) probe coupling, (2) loop coupling, and (3) aperture coupling between adjacent guides. They are illustrated in Fig. 1.4, and some of them are analyzed later. These coupling

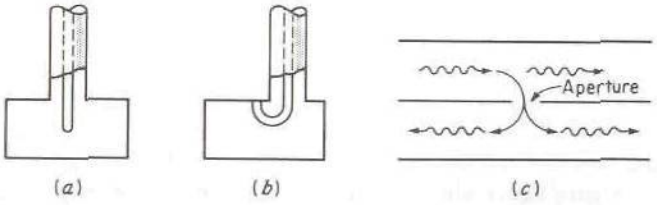


FIGURE 1.4
Basic methods of coupling energy into and out of waveguides. (*a*) Probe coupling; (*b*) loop coupling; (*c*) aperture coupling.

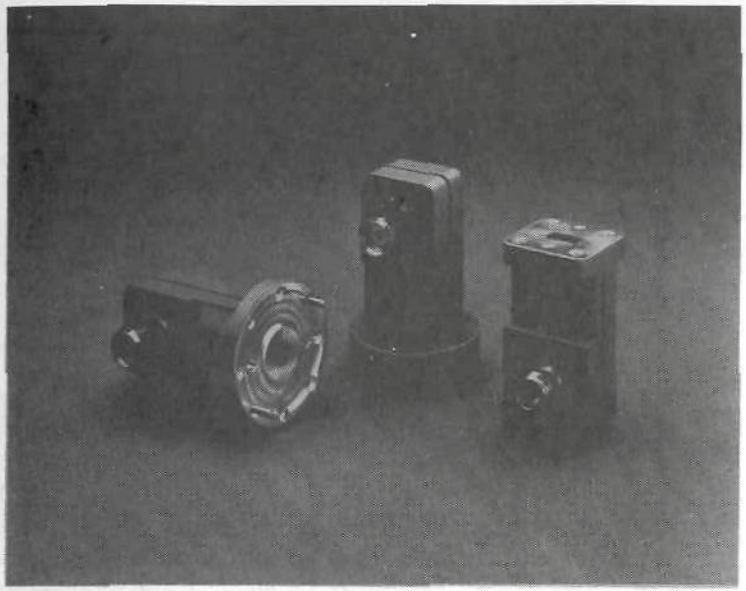


FIGURE 1.5
Waveguide-to-coaxial-line transitions that use probe coupling as shown in Fig. 1.4*a*. (*Photograph courtesy of Ray Moskaluk, Hewlett Packard Company.*)

devices are actually small antennas that radiate into the waveguide. A photograph of a waveguide-to-coaxial-line transition is shown in Fig. 1.5.

Inductive and capacitive elements take a variety of forms at microwave frequencies. Perhaps the simplest are short-circuited sections of transmission line and waveguide. These exhibit a range of susceptance values from minus to plus infinity, depending on the length of the line, and hence may act as either inductive or capactive elements. They may be connected as either series or shunt elements, as illustrated in Fig. 1.6. They are commonly referred to as stubs and are widely used as impedance-matching elements. In a rectangular guide thin conducting windows, or diaphragms, as illustrated in Fig. 1.7, also act as shunt susceptive elements. Their

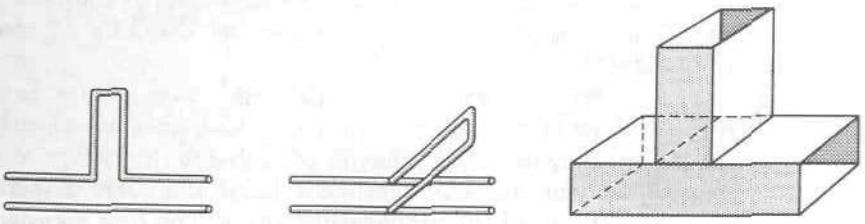


FIGURE 1.6
Stub-type reactive elements. (*a*) Series element; (*b*) shunt element; (*c*) waveguide stub.

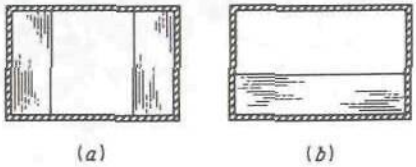


FIGURE 1.7
Shunt susceptive elements in a waveguide. (*a*) Inductive window; (*b*) capacitive window.

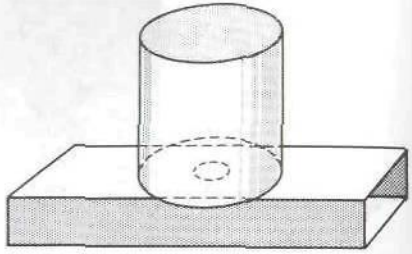


FIGURE 1.8
Cylindrical cavity aperture coupled to a rectangular waveguide.

inductive or capacitive nature depends on whether there is more magnetic energy or electric energy stored in local fringing fields.

Resonant circuits are used both at low frequencies and at microwave frequencies to control the frequency of an oscillator and for frequency filtering. At low frequencies this function is performed by an inductor and capacitor in a series or parallel combination. Resonance occurs when there are equal average amounts of electric and magnetic energy stored. This energy oscillates back and forth between the magnetic field around the inductor and the electric field between the capacitor plates. At microwave frequencies the *LC* circuit may be replaced by a closed conducting enclosure, or cavity. The electric and magnetic energy is stored in the field within the cavity. At an infinite number of specific frequencies, the resonant frequencies, there are equal average amounts of electric and magnetic energy stored in the cavity volume. In the vicinity of any one resonant frequency, the input impedance to the cavity has the same properties as for a conventional *LC* resonant circuit. One significant feature worth noting is the very much larger Q values that may be obtained, these being often in excess of 10^4, as compared with those obtainable from low-frequency *LC* circuits. Figure 1.8 illustrates a cylindrical cavity that is aperture coupled to a rectangular waveguide. Figure 1.9 is a photograph of a family of waveguide low-pass filters. The theory and design of microwave filters is given in Chap. 8. A photograph of a family of waveguide directional couplers is shown in Fig. 1.10. The design of directional couplers is covered in Chap. 6. The photograph in Fig. 1.11 shows a family of coaxial-line GaAs diode detectors.

When a number of microwave devices are connected by means of sections of transmission lines or waveguides, we obtain a microwave circuit. The analysis of the behavior of such circuits is carried out either in terms of equivalent transmission-line voltage and current waves or in terms of the amplitudes of the propagating waves. The first approach leads to an equivalent-impedance description, and the second emphasizes the wave nature of the fields and results in a *scattering-matrix* formulation. Both approaches are used in this book. Since transmission-line circuit analysis forms the basis, either directly or by analogy, for the analysis of all microwave circuits,

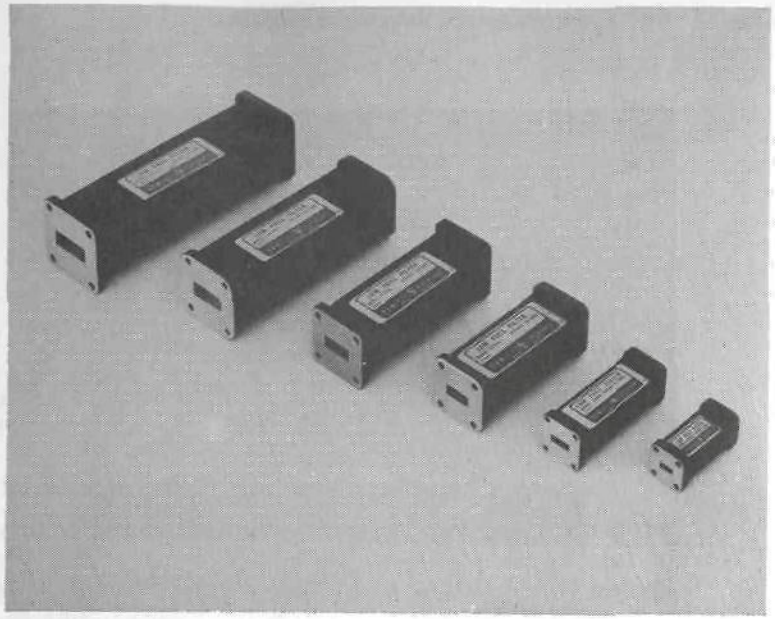


FIGURE 1.9
A family of waveguide low-pass filters for various microwave frequency bands. (*Photographs courtesy of Ray Moskaluk, Hewlett Packard Company.*)

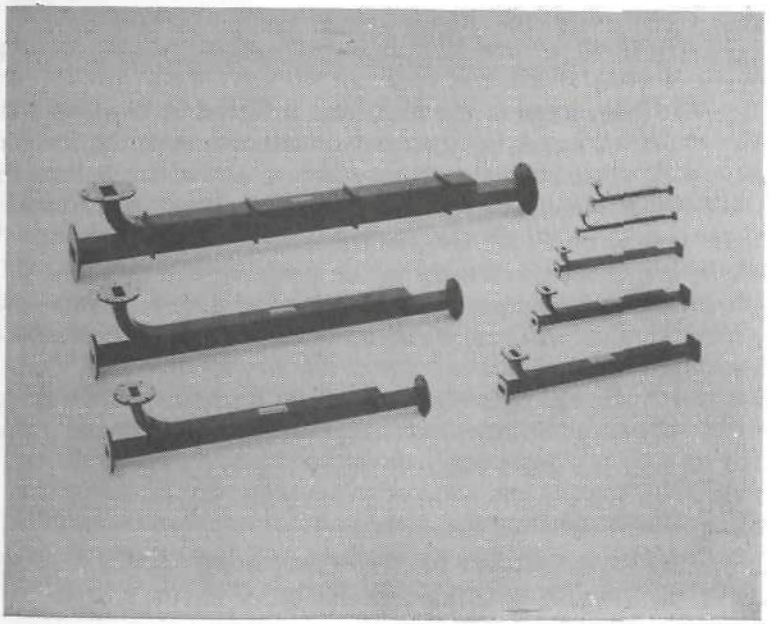


FIGURE 1.10
A family of waveguide directional couplers for various microwave frequency bands. (*Photographs courtesy of Ray Moskaluk, Hewlett Packard Company.*)

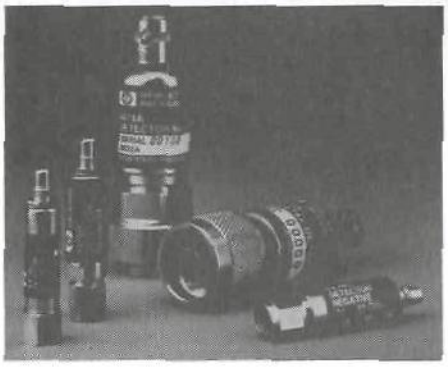


FIGURE 1.11
Coaxial-line GaAs diode detectors for various microwave frequency bands. (*Photographs courtesy of Ray Moskaluk, Hewlett Packard Company.*)

a considerable amount of attention is devoted to a fairly complete treatment of this subject early in the text. This material, together with the field analysis of the waves that may propagate along waveguides and that may exist in cavities, represents a major portion of the theory with which the microwave engineer must be familiar.

The microwave systems engineer must also have some understanding of the principles of operation of various microwave tubes, such as klystrons, magnetrons, and traveling-wave tubes, and of the newer solid-state devices, such as masers, parametric amplifiers, and microwave transistors. This is required in order to make intelligent selection and proper use of these devices. In the text sufficient work is done to provide for this minimum level of knowledge of the principles involved. A treatment that is fully adequate for the device designer is very much outside the scope of this book.

Solid-state oscillators for use as local oscillators in receiver front ends have largely replaced the klystron. Solid-state oscillators for low-power transmitters are also finding widespread use. Thus the future for microwave engineering is clearly in the direction of integrated solid-state circuits and the development of the necessary passive components needed in these circuits, which are also compatible with the fabrication methods that are used.

In the light of the foregoing discussion, it should now be apparent that the study of microwave engineering should include, among other things, at least the following:

1. Electromagnetic theory
2. Wave solutions for transmission lines and waveguides
3. Transmission-line and waveguide circuit analysis
4. Resonators and slow-wave structures
5. Microwave oscillators and amplifiers
6. Antennas
7. Microwave propagation
8. Systems considerations

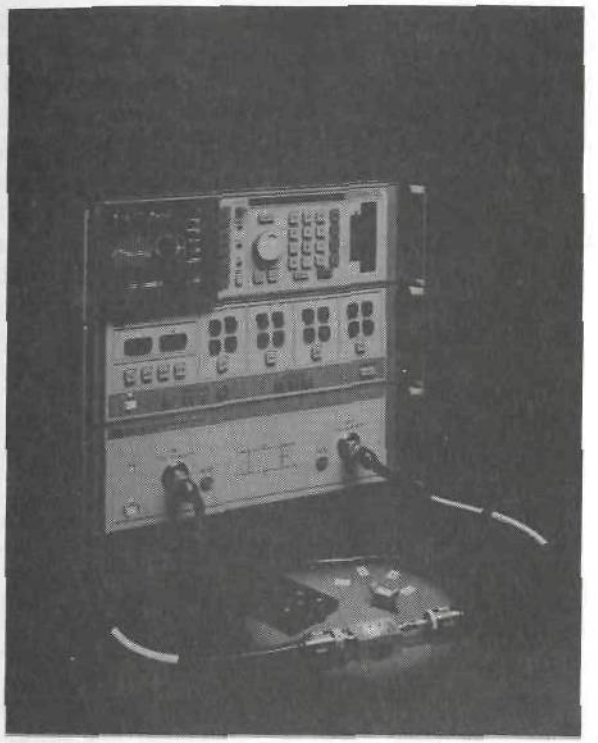


FIGURE 1.12
A microwave network analyzer used to measure scattering matrix parameters. (*Photographs courtesy of Ray Moskaluk, Hewlett Packard Company.*)

Apart from the last three, these are the major topics covered in the text. It is not possible to discuss in any great detail more than a few of the many microwave devices available and in current use. Therefore only a selected number of them are analyzed, to provide illustrative examples for the basic theory being developed. The available technical literature may be, and should be, consulted for information on devices not included here. Appropriate references are given throughout the text.

The number of topics treated in this text represents a good deal more than can be covered in a one-semester course. However, rather than limit the depth of treatment, it was decided to separate some of the more specialized analytical treatments of particular topics from the less analytical discussion. These specialized sections are marked with a star, and can be eliminated in a first reading without significantly interrupting the continuity of the text.† The student or engineer interested in the design of microwave devices, or in a fuller understanding of various aspects of microwave theory, is advised to read these special sections.

As in any engineering field, measurements are of great importance in providing the link between theory and practice at microwave frequencies.

†Problems based on material in these sections are also marked by a star.

Space does not permit inclusion of the subject of microwave measurements in this text. A number of excellent texts devoted entirely to microwave measurements are available, and the reader is referred to them.

There are a variety of commercially available instruments that enable microwave measurements to be carried out automatically with computer control. The photograph in Fig. 1.12 shows a network analyzer equipped to measure the scattering-matrix parameters of a microwave device. The scattering-matrix parameters, as a function of frequency, can be displayed on a Smith chart. The scattering-matrix parameters are commonly used in place of the usual impedance and admittance parameters to characterize a microwave device and are described in Chap. 4.

REFERENCES

1. Historical Perspectives of Microwave Technology, *IEEE Trans.*, vol. MTT-32, September, 1984, Special Centennial Issue.
2. *Kraus, J. D.: "Antennas," 2nd ed., McGraw-Hill Book Company, New York, 1988.*
3. Collin, R. E.: "Antennas and Radiowave Propagation," McGraw-Hill Book Company, New York, 1985.
4. Stutzman, W. L., and G. A. Thiele: "Antenna Theory and Design," John Wiley & Sons, Inc., New York, 1981.
5. Elliott, R. S.: "Antenna Theory and Design," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1981.
6. Balanis, C. A.: "Antenna Theory, Analysis, and Design," Harper & Row Publishers, Inc., New York, 1982.
7. Pratt, T., and C. W. Bostian: "Satellite Communications," John Wiley & Sons, New York, 1986.
8. Ivanek, F. (ed.): "Terrestrial Digital Microwave Communications," Artech House Books, Norwood, Mass., 1989.
9. Skolnik, M. I.: "Introduction to Radar Systems," McGraw-Hill Book Company, New York, 1962.
10. Montgomery, C. G.: "Technique of Microwave Measurements," McGraw-Hill Book Company, New York, 1947.
11. Ginzton, E. L.: "Microwave Measurements," McGraw-Hill Book Company, New York, 1957.
12. Bailey, A. E. (ed.): "Microwave Measurement," Peter Peregrinus, London, 1985.
13. Okress, E. C.: "Microwave Power Engineering," Academic Press, New York, 1968.
14. Ulaby, F. T., R. K. Moore, and A. K. Fung: "Microwave Remote Sensing: Active and Passive. Microwave Remote Sensing, Fundamentals and Radiometry," vol. 1, Addison-Wesley, Reading, Mass., 1981.

CHAPTER 2

ELECTROMAGNETIC THEORY

2.1 MAXWELL'S EQUATIONS

Electric and magnetic fields that vary with time are governed by physical laws described by a set of equations known collectively as Maxwell's equations. For the most part these equations were arrived at from experiments carried out by several investigators. It is not our purpose here to justify the basis for these equations, but rather to gain some understanding of their physical significance and to learn how to obtain solutions of these equations in practical situations of interest in the microwave engineering field. The electric field $\mathscr{E}$ and magnetic field $\mathscr{B}$ are vector fields and in general have amplitudes and directions that vary with the three spatial coordinates x, y, z and the time coordinate t.† In mks units, which are used throughout, the electric field is measured in volts per meter and the magnetic field in webers per square meter. Since these fields are vector fields, the equations governing their behavior are most conveniently written in vector form.‡

The electric field $\mathscr{E}$ and magnetic field $\mathscr{B}$ are regarded as fundamental in that they give the force on a charge q moving with velocity $\mathbf{v}$; that is,

$$\mathbf{F} = q(\mathscr{E} + \mathbf{v} \times \mathscr{B}) \tag{2.1}$$

†Boldface script type is used to represent vector fields having arbitrary time dependence. Boldface roman type is used later for the phasor representation of fields having sinusoidal time dependence.

‡It is assumed that the reader is familiar with vector analysis. However, for convenient reference, a number of vector formulas and relations are summarized in App. I.

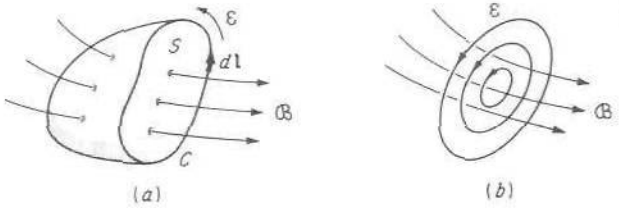


FIGURE 2.1
Illustration of Faraday's law.

where $\mathbf{F}$ is the force in newtons, q is the charge measured in coulombs, and $\mathbf{v}$ is the velocity in meters per second. This force law is called the Lorentz force equation. In addition to the $\mathscr{E}$ and $\mathscr{B}$ fields, it is convenient to introduce two auxiliary field vectors, namely, the electric displacement $\mathscr{D}$ and the magnetic intensity $\mathscr{H}$. These are related to $\mathscr{E}$ and $\mathscr{B}$ through the electric and magnetic polarization of material media, a topic covered in the next section. In this section we consider fields in vacuum, or *free space*, only. In this case the following simple relationships hold:

$$\mathscr{H} = \frac{1}{\mu_0}\mathscr{B} \tag{2.2a}$$

$$\mathscr{D} = \epsilon_0 \mathscr{E} \tag{2.2b}$$

where $\mu_0 = 4\pi \times 10^{-7}$ H/m and is called the permeability of vacuum, and $\epsilon_0 = 10^{-9}/36\pi = 8.854 \times 10^{-12}$ F/m and is known as the permittivity of vacuum.

One of the basic laws of electromagnetic phenomena is Faraday's law, which states that a time-varying magnetic field generates an electric field. With reference to Fig. 2.1, let C denote an arbitrary closed curve that forms the boundary of a nonmoving surface S. The time rate of change of total magnetic flux through the surface S is $\partial(\int_S \mathscr{B} \cdot d\mathbf{S})/\partial t$. According to Faraday's law, *this time rate of change of total magnetic flux is equal to the* negative value of the total voltage measured around C. The later quantity is given by $-\oint_C \mathscr{E} \cdot d\mathbf{l}$. Hence the mathematical statement of Faraday's law is

$$\oint_C \mathscr{E} \cdot d\mathbf{l} = -\frac{\partial}{\partial t}\int_S \mathscr{B} \cdot d\mathbf{S} \tag{2.3}$$

The line integral of $\mathscr{E}$ around C is a measure of the circulation, or "curling up," of the electric field in space. The time-varying magnetic field may be properly regarded as a vortex source that produces an electric field having nonzero curl, or circulation. Although (2.3) is in a form that is readily interpreted physically, it is not in a form suitable for the analysis of a *physical problem. What is required is a differential equation that is equiva*lent to (2.3). This equation may be obtained by using Stokes' theorem from vector analysis, which states that the line integral of a vector around a closed contour C is equal to the integral of the normal component of the

curl of this vector over any surface having C as its boundary. The curl of a vector is written $\nabla \times \mathscr{E}$ (App. 1), and hence (2.1) becomes

$$\oint_C \mathscr{E} \cdot d\mathbf{l} = \int_S \nabla \times \mathscr{E} \cdot d\mathbf{S} = -\frac{\partial}{\partial t} \int_S \mathscr{B} \cdot d\mathbf{S}$$

Since S is completely arbitrary, the latter two integrals are equal only if

$$\nabla \times \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t} \tag{2.4}$$

which is the desired differential equation describing Faraday's law. The curl is a measure of the circulation of a vector field at a point.

Helmholtz's theorem from vector analysis states that a vector field is completely defined only when the curl, or circulation, of the field, and also its divergence, are given at every point in space. Now the divergence (or convergence) of field lines arises only if a proper source (or sink) is available. The electric field, in addition to having a curl produced by the vortex source $-\partial \mathscr{B}/\partial t$, has a divergence produced by electric charge. Gauss' law states that the total flux of $\mathscr{D} = \epsilon_0 \mathscr{E}$ from a volume V is equal to the net charge contained within V. If ρ represents the charge density in coulombs per cubic meter, Gauss' law may be written as

$$\oint_S \epsilon_0 \mathscr{E} \cdot d\mathbf{S} = \int_V \rho \, dV \tag{2.5}$$

This equation may be converted to a differential equation by using the divergence theorem to give

$$\oint_S \epsilon_0 \mathscr{E} \cdot d\mathbf{S} = \int_V \nabla \cdot \epsilon_0 \mathscr{E} \, dV = \int_V \rho \, dV$$

Since V is arbitrary, it follows that

$$\nabla \cdot \epsilon_0 \mathscr{E} = \nabla \cdot \mathscr{D} = \rho \tag{2.6}$$

where $\nabla \cdot \mathscr{D}$ is the divergence of $\mathscr{D}$, that is, a measure of the total outward flux of $\mathscr{D}$ from a volume element, divided by the volume of the element, as this volume shrinks to zero. Since both the curl and divergence of the electric field are now specified, this field is completely determined in terms of the two sources, $\partial \mathscr{B}/\partial t$ and ρ.

To complete the formulation of electromagnetic phenomena, we must now relate the curl and divergence of the magnetic field to their sources. The vortex source that creates the circulation, or curl, of the magnetic field $\mathscr{H}$ is the current. By current is meant the total current density, the conduction current density $\mathscr{J}$ measured in amperes per square meter, the displacement current density $\partial \mathscr{D}/\partial t$, and the convection current $\rho \mathbf{v}$ consisting of charge in motion if present. Convection current is not included in this chapter. However, in the chapter dealing with microwave tubes, convection current plays a central role and is discussed in detail there. The displacement current density $\partial \mathscr{D}/\partial t$ was first introduced by Maxwell, and leads to

the possibility of wave motion, as will be seen. Mathematically, the circulation of $\mathscr{H}$ around a closed contour C bounding a surface S as in Fig. 2.1 is given by

$$\oint_C \mathscr{H} \cdot d\mathbf{l} = \int_S \frac{\partial \mathscr{D}}{\partial t} \cdot d\mathbf{S} + \int_S \mathscr{J} \cdot d\mathbf{S} \tag{2.7}$$

Application of Stokes' law to the left-hand side yields

$$\int_S \nabla \times \mathscr{H} \cdot d\mathbf{S} = \int_S \frac{\partial \mathscr{D}}{\partial t} \cdot d\mathbf{S} + \int_S \mathscr{J} \cdot d\mathbf{S}$$

from which it may be concluded that

$$\nabla \times \mathscr{H} = \frac{\partial \mathscr{D}}{\partial t} + \mathscr{J} \tag{2.8}$$

Since magnetic charge, as the dual of electric charge, does not exist in nature, it may be concluded that the divergence of $\mathscr{B}$ is always zero; i.e., the flux lines of $\mathscr{B}$ are always closed since there are no charges for them to terminate on. Thus the net flux of $\mathscr{B}$ through any closed surface S is always zero; i.e., just as much flux enters through the surface as leaves it. Corresponding to (2.5) and (2.6), we thus have

$$\oint_S \mathscr{B} \cdot d\mathbf{S} = 0 \tag{2.9}$$

$$\nabla \cdot \mathscr{B} = 0 \tag{2.10}$$

Conduction current, of density $\mathscr{J}$, is the net flow of electric charge. Since charge is conserved, the total rate of flow of charge out of a volume V is equal to the time rate of decrease of total charge within V, as expressed by the equation

$$\oint_S \mathscr{J} \cdot d\mathbf{S} = -\frac{\partial}{\partial t} \int_V \rho \, dV \tag{2.11}$$

This is the continuity equation, and it may be converted to a differential equation by using the divergence theorem in the same manner as was done to derive (2.6) from (2.5). It is readily found that

$$\nabla \cdot \mathscr{J} + \frac{\partial \rho}{\partial t} = 0 \tag{2.12}$$

This equation may also be derived from (2.8) and (2.6). Since the divergence of the curl of any vector is identically zero, the divergence of (2.8) yields

$$0 = \frac{\partial \nabla \cdot \mathscr{D}}{\partial t} + \nabla \cdot \mathscr{J}$$

Using (2.6) converts this immediately into the continuity equation (2.12). If the displacement current density $\partial \mathscr{D}/\partial t$ had not been included as part of the total current density on the right-hand side of (2.8), that equation would

have led to the conclusion that $\nabla \cdot \mathcal{J} = 0$, a result inconsistent with the continuity equation unless the charge density was independent of time.

In summary, the four equations, known as Maxwell's equations, that describe electromagnetic phenomena in vacuum are

$$\nabla \times \mathcal{E} = -\frac{\partial \mathcal{B}}{\partial t} \tag{2.13a}$$

$$\nabla \times \mathcal{H} = \frac{\partial \mathcal{D}}{\partial t} + \mathcal{J} \tag{2.13b}$$

$$\nabla \cdot \mathcal{D} = \rho \tag{2.13c}$$

$$\nabla \cdot \mathcal{B} = 0 \tag{2.13d}$$

where in (2.13b) the convection current $\rho\mathbf{v}$ has not been included. The continuity equation may be derived from (2.13b) and (2.13c), and hence contains no additional information. Although $-\partial\mathcal{B}/\partial t$ may be regarded as a source for $\mathcal{E}$, and $\partial\mathcal{D}/\partial t$ as a source of $\mathcal{H}$, the ultimate sources of an electromagnetic field are the current $\mathcal{J}$ and charge ρ. For time-varying fields, that charge density ρ which varies with time is not independent of $\mathcal{J}$ since it is related to the latter by the continuity equation. As a consequence, it is possible to derive the time-varying electromagnetic field from a knowledge of the current density $\mathcal{J}$ alone.

It is not difficult to show in a qualitative way that (2.13a) and (2.13b) lead to wave propagation, i.e., to the propagation of an electromagnetic disturbance through space. Consider a loop of wire in which a current varying with time flows as in Fig. 2.2. The conduction current causes a circulation, or curling, of the magnetic field around the current loop as in Fig. 2.2a (for clarity only a few flux lines are shown). The changing magnetic field in turn creates a circulating, or curling, electric field, with field lines that encircle the magnetic field lines as in Fig. 2.2b. This changing electric field creates further curling magnetic field lines as in Fig. 2.2c, and so forth. The net result is the continual growth and spreading of the electromagnetic field into all space surrounding the current loop. The

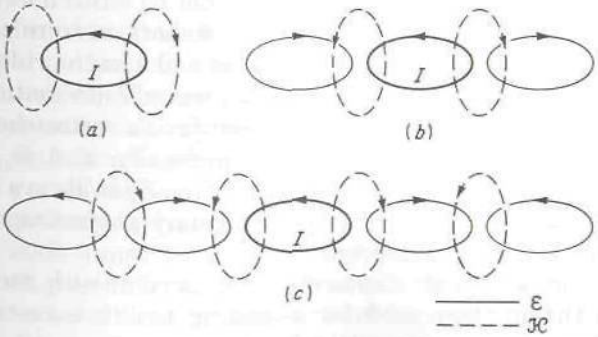


FIGURE 2.2
The growth or generation of an electromagnetic wave from a current loop.

disturbance moves outward with the velocity of light. A little thought will show that the same characteristic mutual effect between two quantities must always exist for wave motion. That is, quantity A must be generated by quantity B, and vice versa. For example, in an acoustical wave the excess pressure creates a motion of the adjacent air mass. The motion of the air mass by virtue of its inertia in turn creates a condensation, or excess pressure, farther along. The repetition of this process generates the acoustical wave.

For the most part, as at lower frequencies, it is sufficient to consider only the steady-state solution for the electromagnetic field as produced by currents having sinusoidal time dependence. The time derivative may then be eliminated by denoting the time dependence of all quantities as $e^{j\omega t}$ and representing all field vectors as complex-phasor space vectors independent of time. Boldface roman type is used to represent these complex-phasor space vectors. For example, the mathematical representation for the electric field $\mathscr{E}(x, y, z, t)$ will be $\mathbf{E}(x, y, z)e^{j\omega t}$. Each component of $\mathbf{E}$ is in general complex, with a real and imaginary part; thus

$$\mathbf{E} = \mathbf{a}_x(E_{xr} + jE_{xi}) + \mathbf{a}_y(E_{yr} + jE_{yi}) + \mathbf{a}_z(E_{zr} + jE_{zi}) \qquad (2.14)$$

where the subscript r refers to the real part and the subscript i refers to the imaginary part. Each component is allowed to be complex in order to provide for an arbitrary time phase for each component. This may be seen by recalling the usual method of obtaining $\mathscr{E}$ from its phasor representation. That is, by definition,

$$\mathscr{E}(x, y, z, t) = \mathrm{Re}\left[\mathbf{E}(x, y, z)e^{j\omega t}\right] \qquad (2.15)$$

Thus

$$\begin{aligned} E_x &= \mathrm{Re}\left[(E_{xr} + jE_{xi})e^{j\omega t}\right] \\ &= \mathrm{Re}\left(\sqrt{E_{xr}^2 + E_{xi}^2}\, e^{j\omega t + j\phi}\right) \\ &= \sqrt{E_{xr}^2 + E_{xi}^2}\, \cos(\omega t + \phi) \end{aligned}$$

where $\phi = \tan^{-1}(E_{xi}/E_{xr})$. Unless E_x had both an imaginary part jE_{xi} and a real part E_{xr}, the arbitrary phase angle ϕ would not be present. As a general rule, the time factor $e^{j\omega t}$ will not be written down when the phasor representation is used. However, it is important to remember both the fact that such a time dependence is implied and also the rule (2.15) for obtaining the physical field vector from its phasor representation. The real and imaginary parts of the space components of a vector should not be confused with the space components; for example, E_{xr} and E_{xi} are not two space components of E_x since the component $\mathbf{a}_x E_x$ is always directed along the x axis in space, with the real and imaginary parts simply accounting for an arbitrary time phase or origin.

A further point of interest in connection with the phasor representation is the method used for obtaining the time-average value of a field quantity.

For example, if

$$\mathscr{E} = \mathbf{a}_x E_1 \cos(\omega t + \phi_1) + \mathbf{a}_y E_2 \cos(\omega t + \phi_2) + \mathbf{a}_z E_3 \cos(\omega t + \phi_3)$$

the time-average value of $|\mathscr{E}|^2$ is

$$\begin{aligned} |\mathscr{E}|^2_{\text{av}} &= \frac{1}{T}\int_0^T \mathscr{E} \cdot \mathscr{E}\, dt \\ &= \frac{1}{T}\int_0^T \big[E_1^2 \cos^2(\omega t + \phi_1) + E_2^2 \cos^2(\omega t + \phi_2) \\ &\qquad + E_3^2 \cos(\omega t + \phi_3)\big]\, dt \\ &= \frac{1}{2}\left(E_1^2 + E_2^2 + E_3^2\right) \end{aligned} \tag{2.16}$$

where T is the period, equal to $2\pi/\omega$. The same result is obtained by simply taking one-half of the scalar, or dot, product of $\mathbf{E}$ with the complex conjugate $\mathbf{E}^*$; thus

$$|\mathscr{E}|^2_{\text{av}} = \tfrac{1}{2}\mathbf{E} \cdot \mathbf{E}^* = \tfrac{1}{2}\left[\left(E_{xr}^2 + E_{xi}^2\right) + \left(E_{yr}^2 + E_{yi}^2\right) + \left(E_{zr}^2 + E_{zi}^2\right)\right] \tag{2.17}$$

since $E_x E_x^* = (E_{xr} + jE_{xi})(E_{xr} - jE_{xi}) = E_{xr}^2 + E_{xi}^2$, etc. This is equal to (2.16), since $E_1^2 = E_{xr}^2 + E_{xi}^2$, etc.

By using the phasor representation, the time derivative $\partial/\partial t$ may be replaced by the factor $j\omega$ since $\partial e^{j\omega t}/\partial t = j\omega e^{j\omega t}$. Hence Maxwell's equations, with steady-state sinusoidal time dependence, become

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B} \tag{2.18a}$$

$$\nabla \times \mathbf{H} = j\omega\mathbf{D} + \mathbf{J} \tag{2.18b}$$

$$\nabla \cdot \mathbf{D} = \rho \tag{2.18c}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.18d}$$

2.2 CONSTITUTIVE RELATIONS

In material media the auxiliary field vectors $\mathscr{H}$ and $\mathscr{D}$ are defined in terms of the polarization of the material and the fundamental field quantities $\mathscr{B}$ and $\mathscr{E}$. The relationships of $\mathscr{H}$ to $\mathscr{B}$ and of $\mathscr{D}$ to $\mathscr{E}$ are known as constitutive relations, and must be known before solutions for Maxwell's equations can be found.

Consider first the electric case. If an electric field $\mathscr{E}$ is applied to a material body, this force results in a distortion of the atoms or molecules in such a manner as to create effective electric dipoles with a dipole moment $\mathscr{P}$ per unit volume. The total displacement current is the sum of the vacuum displacement current $\partial\epsilon_0\mathscr{E}/\partial t$ and the polarization current $\partial\mathscr{P}/\partial t$. To avoid accounting for the polarization current $\partial\mathscr{P}/\partial t$ explicitly, the

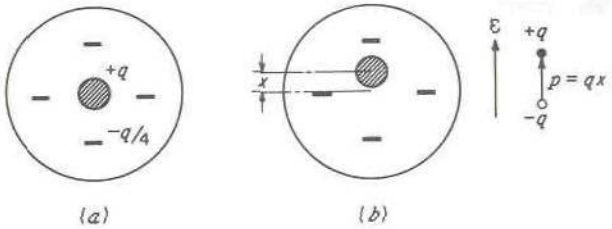


FIGURE 2.3
Model for determining the polarization of an atom.

displacement vector $\mathscr{D}$ is defined as

$$\mathscr{D} = \epsilon_0 \mathscr{E} + \mathscr{P} \tag{2.19}$$

whence the total displacement current density can be written as $\partial \mathscr{D}/\partial t$.

For a great many materials the polarization $\mathscr{P}$ is in the direction of the electric field $\mathscr{E}$, although rarely will $\mathscr{P}$ have the same time phase as $\mathscr{E}$. A simple classical model will serve to illustrate this point. Figure 2.3*a* shows a model of an atom consisting of a nucleus with charge q surrounded by a spherically symmetrical electron cloud of total charge $-q$. The application of a field $\mathscr{E}$ displaces the electron cloud an effective distance x as in Fig. 2.3*b*. This displacement is resisted by a restoring force kx proportional to the displacement (Prob. 2.1). In addition, dissipation, or damping, effects are present and result in an additional force, which we shall assume to be proportional to the velocity. If m is the effective mass of the electron cloud, the dynamical equation of motion is obtained by equating the sum of the inertial force $m\,d^2x/dt^2$, viscous force $m\nu\,dx/dt$, and restoring force kx to the applied force $-q\mathscr{E}$; thus

$$m\frac{d^2x}{dt^2} + m\nu\frac{dx}{dt} + kx = -q\mathscr{E} \tag{2.20}$$

When $\mathscr{E} = E_x \cos \omega t$, the solution for x is of the form $x = -A\cos(\omega t + \phi)$.

If $E_x \cos \omega t$ is represented by the phasor E_x, and x by the phasor X, the solution for X is readily found to be

$$X = \frac{-qE_x}{-\omega^2 m + j\omega\nu m + k}$$

and hence

$$x = \mathrm{Re}(Xe^{j\omega t}) = A\cos(\omega t + \phi)$$

where

$$A = \frac{(q/m)E_x}{\left[(\omega^2 - \omega_0^2)^2 + \omega^2\nu^2\right]^{1/2}}$$

$$\phi = \tan^{-1}\frac{\omega\nu}{\omega^2 - \omega_0^2}$$

and we have replaced k/m by ω_0^2.

The dipole moment is p_x, where

$$p_x = -qx = \frac{q^2 E_x}{m\left[\left(\omega^2 - \omega_0^2\right)^2 + \omega^2\nu^2\right]^{1/2}} \cos(\omega t + \phi) \tag{2.21}$$

For N such atoms per unit volume the polarization per unit volume is $\mathscr{P}_x = Np_x$ and the displacement $\mathscr{D}_x$ is given by

$$\mathscr{D}_x = \epsilon_0 E_x \cos \omega t + \frac{Nq^2 E_x}{m\left[\left(\omega^2 - \omega_0^2\right)^2 + \omega^2\nu^2\right]^{1/2}} \cos(\omega t + \phi)$$

This equation may also be put into the following form:

$$\mathscr{D}_x = E_x \left\{ \frac{\left[\epsilon_0\left(\omega_0^2 - \omega^2\right) + Nq^2/m\right]^2 + \left(\omega\nu\epsilon_0\right)^2}{\left(\omega_0^2 - \omega^2\right)^2 + \left(\omega\nu\right)^2} \right\}^{1/2} \cos(\omega t - \theta) \tag{2.22}$$

where $$\theta = \tan^{-1} \frac{\omega\nu}{\omega_0^2 - \omega^2} - \tan^{-1} \frac{\omega\nu}{\omega_0^2 - \omega^2 + Nq^2/\epsilon_0 m}$$

Two points are of interest in connection with (2.22). One is the linear relationship between $\mathscr{P}$ and $\mathscr{E}$, and hence between $\mathscr{D}$ and $\mathscr{E}$. The second is the phase lag in $\mathscr{D}$ relative to $\mathscr{E}$ whenever damping forces are present.

The phase difference between $\mathscr{P}$, $\mathscr{E}$, and $\mathscr{D}$ makes it awkward to handle the relations between these quantities unless phasor representation is used. In phasor representation (2.21) and (2.22) become

$$P_x = \frac{q^2 E_x}{\left(\omega_0^2 - \omega^2 + j\omega\nu\right)m} \tag{2.23}$$

$$D_x = \frac{\epsilon_0\left(\omega_0^2 - \omega^2 + j\omega\nu\right) + Nq^2/m}{\omega_0^2 - \omega^2 + j\omega\nu} E_x \tag{2.24}$$

In general, for linear media, we may write

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} \tag{2.25}$$

where χ_e is a complex constant of proportionality called the electric susceptibility. The equation for $\mathbf{D}$ becomes

$$\begin{aligned} \mathbf{D} &= \epsilon_0 \mathbf{E} + \mathbf{P} = \epsilon_0(1 + \chi_e)\mathbf{E} \\ &= \epsilon \mathbf{E} = \epsilon_r \epsilon_0 \mathbf{E} = (\epsilon' - j\epsilon'')\mathbf{E} \end{aligned} \tag{2.26}$$

where $\epsilon = \epsilon_0(1 + \chi_e)$ is called the permittivity, and $\epsilon_r = \epsilon/\epsilon_0$, the dielectric constant of the medium. Note that ϵ is complex whenever damping effects are present and that the imaginary part is always negative. A positive imaginary part would imply energy creation instead of energy loss. [The reader may verify from (2.22) that θ is always positive.]

Loss in a dielectric material may also occur because of a finite conductivity σ. The two mechanisms are indistinguishable as far as external effects

related to power dissipation are concerned. The curl equation for **H** may be written as

$$\nabla \times \mathbf{H} = j\omega(\epsilon' - j\epsilon'')\mathbf{E} + \sigma\mathbf{E}$$

where $\mathbf{J} = \sigma\mathbf{E}$ is the conduction current density in the material. We may also write

$$\nabla \times \mathbf{H} = j\omega\left[\epsilon' - j\left(\epsilon'' + \frac{\sigma}{\omega}\right)\right]\mathbf{E} = j\omega\epsilon'\mathbf{E} + (\omega\epsilon'' + \sigma)\mathbf{E} \qquad (2.27)$$

where by $\epsilon'' + \sigma/\omega$ may be considered as the effective imaginary part of the permittivity, or $\omega\epsilon'' + \sigma$ as the total effective conductivity.

The loss tangent of a dielectric medium is defined by

$$\tan \delta_l = \frac{\omega\epsilon'' + \sigma}{\omega\epsilon'} \qquad (2.28)$$

Any measurement of $\tan \delta_l$ always includes the effects of finite conductivity σ. At microwave frequencies, however, $\omega\epsilon''$ is usually much larger than σ because of the large value of ω.

Materials for which **P** is linearly related to **E** and in the same direction as **E** are called linear isotropic materials. Nonlinear effects generally occur only for very large applied fields, and as a consequence are rarely encountered in microwave work. However, nonisotropic material is of some importance. If the crystal structure lacks spherical symmetry such as that in a cubic crystal, it may be anticipated that the polarization per unit volume will depend on the direction of the applied field. In Fig. 2.4 a two-dimensional sketch of a crystal lacking cubic symmetry is given. The polarization produced when the field is applied along the x axis may be greater than that when the field is applied along the y or z axis because of the greater ease of polarization along the x axis. In this case we must write

$$D_x = \epsilon_{xx}E_x \qquad D_y = \epsilon_{yy}E_y \qquad D_z = \epsilon_{zz}E_z \qquad (2.29)$$

where ϵ_{xx}, ϵ_{yy}, and ϵ_{zz} are, in general, all different. The dielectric constants $\epsilon_{rx} = \epsilon_{xx}/\epsilon_0$, $\epsilon_{ry} = \epsilon_{yy}/\epsilon_0$, $\epsilon_{rz} = \epsilon_{zz}/\epsilon_0$ are known as the principal dielectric constants, and the material is said to be anisotropic. If the coordinate system used had a different orientation with respect to the crystal structure,

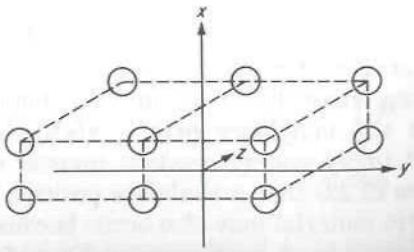


FIGURE 2.4
A noncubic crystal exhibiting anisotropic effects.

the relation between **D** and **E** would become

$$D_x = \epsilon_{xx}E_x + \epsilon_{xy}E_y + \epsilon_{xz}E_z$$
$$D_y = \epsilon_{yx}E_x + \epsilon_{yy}E_y + \epsilon_{yz}E_z$$
$$D_z = \epsilon_{zx}E_x + \epsilon_{zy}E_y + \epsilon_{zz}E_z$$

or in matrix form,

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{2.30}$$

Only for a particular orientation of the coordinate system does (2.30) reduce to (2.29). This particular orientation defines the principal axis of the medium. For anisotropic media the permittivity is referred to as a tensor permittivity (a tensor of rank 2 may be represented by a matrix). For the most part the materials dealt with in this text are isotropic. Nevertheless, an awareness of the existence of anisotropic media and of the nature of the constitutive relations for such media is important.

For the magnetic case, **H** is defined by the constitutive relation

$$\mu_0\mathbf{H} = \mathbf{B} - \mu_0\mathbf{M} \tag{2.31}$$

where **M** is the magnetic dipole polarization per unit volume. For most materials (ferromagnetic materials excluded), **M** is linearly related to **B** and hence to **H**. By convention this is expressed by the equation

$$\mathbf{M} = \chi_m\mathbf{H} \tag{2.32}$$

where χ_m is called the magnetic susceptibility. Substituting (2.32) into (2.31) gives

$$\mathbf{B} = \mu_0(\mathbf{M} + \mathbf{H}) = \mu_0(1 + \chi_m)\mathbf{H} = \mu\mathbf{H} \tag{2.33}$$

where $\mu = \mu_0(1 + \chi_m)$ is called the permeability.

As in the electric case, damping forces cause μ to be a complex parameter with a negative imaginary part; that is, $\mu = \mu' - j\mu''$. Also, there are magnetic materials that are anisotropic; in particular, ferrites are anisotropic magnetic materials of great usefulness at microwave frequencies. These exhibit a tensor permeability of the following form:

$$[\mu] = \begin{bmatrix} \mu_1 & j\mu_2 & 0 \\ -j\mu_2 & \mu_1 & 0 \\ 0 & 0 & \mu_3 \end{bmatrix} \tag{2.34}$$

when a static magnetic field is applied along the axis for which the permeability is μ_3. A discussion of ferrites and their uses is presented later; so further comments on their anisotropic properties is deferred until then.

In Sec. 2.1 care was taken to write Maxwell's equations in a form valid not only in vacuum but also in material media. Thus (2.13) and (2.18) are

valid in general, but with the constitutive relations of this section replacing the free-space relations (2.2). Note, however, that it is not possible to write, in general, constitutive relations of the form $\mathscr{D} = \epsilon\mathscr{E}$, $\mathscr{B} = \mu\mathscr{H}$, when $\mathscr{D}$ and $\mathscr{E}$, and likewise $\mathscr{B}$ and $\mathscr{H}$, are not in time phase. For arbitrary time dependence we must write instead $\mathscr{D} = \epsilon_0\mathscr{E} + \mathscr{P}$, $\mathscr{B} = \mu_0(\mathscr{H} + \mathscr{M})$ and relate $\mathscr{P}$ and $\mathscr{M}$ to $\mathscr{E}$ and $\mathscr{H}$ through the dynamical equation of motion governing the polarization mechanism. This difficulty may be circumvented by using the phasor representation for which relations such as $\mathbf{D} = \epsilon\mathbf{E}$ are perfectly valid because the complex nature of ϵ accounts for the difference in time phase.† It should be pointed out, however, that for many materials used at frequencies up to and including microwaves, the losses are so small that $\mathscr{D}$ and $\mathscr{E}$, and also $\mathscr{H}$ and $\mathscr{B}$, are very nearly in time phase. In such cases constitutive relations such as $\mathscr{D} = \epsilon\mathscr{E}$, $\mathscr{B} = \mu\mathscr{H}$ apply with negligible error. Significant departure in time phase between $\mathscr{D}$ and $\mathscr{E}$ or $\mathscr{B}$ and $\mathscr{H}$ occurs only in the vicinity of a natural resonance frequency of the equation of motion for the polarization.

2.3 STATIC FIELDS

For electric and magnetic fields that are independent of time, the electric and magnetic fields are not coupled, and likewise the current and charge are not coupled. Putting all time derivatives equal to zero in (2.13) yields‡

$$\nabla \times \mathbf{E} = 0 \tag{2.35a}$$

$$\nabla \cdot \epsilon\mathbf{E} = \rho \tag{2.35b}$$

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{2.36a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.36b}$$

$$\nabla \cdot \mathbf{J} = 0 \tag{2.36c}$$

The last equation is the continuity equation for the special case $\partial\rho/\partial t = 0$.

The static electric field has zero curl, or circulation, and this means that the line integral of $\mathbf{E}$ around any arbitrary closed contour is zero. This property is just the condition that permits $\mathbf{E}$ to be derived from the gradient of a scalar potential function Φ; that is, since $\nabla \times \nabla\Phi$ is identically zero, we may put

$$\mathbf{E} = -\nabla\Phi \tag{2.37}$$

---

†The situation here is like that encountered in ac circuit analysis, where in phasor notation the voltage V equals the current I multiplied by the impedance Z; that is, $V = IZ$. An Ohm's law of this sort cannot be written for the physical voltage and current, for if $\mathscr{V} = \mathrm{Re}(Ve^{j\omega t}) = V\cos\omega t$, then $\mathscr{I} = \mathrm{Re}(Ie^{j\omega t}) = [V/(R^2 + X^2)^{1/2}]\cos(\omega t - \phi)$, where $\phi = \tan^{-1}(X/R)$. Clearly, $\mathscr{V}$ cannot be equated to $\mathscr{I}$ multiplied by a constant because of the difference in phase.

‡For static fields we are using boldface roman type to represent the physically real vector fields.

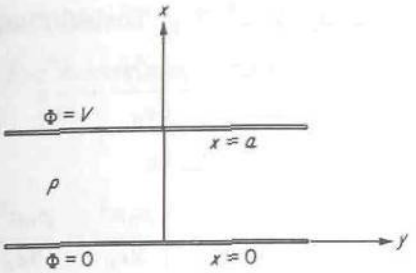


FIGURE 2.5
A simple potential problem.

Substituting (2.37) into (2.35*b*) and assuming that ϵ is a constant independent of the coordinates give

$$-\nabla \cdot \mathbf{E} = \nabla^2 \Phi = -\frac{\rho}{\epsilon} \tag{2.38}$$

This equation is known as Poisson's equation. When $\rho = 0$, Laplace's equation

$$\nabla^2 \Phi = 0 \tag{2.39}$$

is obtained. The basic field problem in electrostatics is to solve Poisson's or Laplace's equation for a potential function Φ that satisfies specified boundary conditions.

As a simple example consider two infinite conducting planes at $x = 0, a$, as in Fig. 2.5. Let charge be distributed with a density $\rho = \rho_0 x$ between the two plates.† It is required to find a Φ which is a solution of Poisson's equation and which equals zero on the plane $x = 0$ and V on the plane $x = a$. The potential will depend on x only; so (2.38) becomes

$$\frac{d^2\Phi}{dx^2} = -\rho_0 \frac{x}{\epsilon_0}$$

Integrating this equation twice gives $\Phi = -\rho_0 x^3/6\epsilon_0 + C_1 x + C_2$. Imposing the boundary conditions at $x = 0, a$ yields $0 = C_2$,

$$V = -\frac{\rho_0 a^3}{6\epsilon_0} + C_1 a + C_2$$

†The example is somewhat artificial since the assumed charge distribution is not a stable one; i.e., the electric field it produces would cause the charge distribution to change.

and hence $C_2 = 0$, $C_1 = V/a + \rho_0 a^2/6\epsilon_0$. The solution for Φ is thus

$$\Phi = -\frac{\rho_0 x^3}{6\epsilon_0} + \frac{\rho_0 a^2 x}{6\epsilon_0} + \frac{V}{a}x$$

The electric field between the two plates is

$$\mathbf{E} = -\nabla\Phi = -\mathbf{a}_x \frac{\partial \Phi}{\partial x} = \mathbf{a}_x \left(\frac{\rho_0 x^2}{2\epsilon_0} - \frac{\rho_0 a^2}{6\epsilon_0} - \frac{V}{a} \right)$$

The solution for the electrostatic field is greatly facilitated by introduction of the scalar potential Φ. For the same reason it is advantageous to introduce a potential function for the solution of magnetostatic problems. Since $\mathbf{B}$ always has zero divergence, it may be derived from the curl of a vector potential $\mathbf{A}$; that is,

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{2.40}$$

This makes the divergence of $\mathbf{B}$ vanish identically because the divergence of the curl of a vector is identically zero. Using (2.40) in (2.36a) and assuming that μ is constant yields the equation

$$\nabla \times \mu \boldsymbol{H} = \nabla \times \mathbf{B} = \nabla \times \nabla \times \mathbf{A} = \mu \boldsymbol{J}$$

A vector identity of use here is $\nabla \times \nabla \times \mathbf{A} = \nabla\nabla \cdot \mathbf{A} - \nabla^2\mathbf{A}$. The divergence of $\mathbf{A}$ may be placed equal to zero without affecting the value of $\mathbf{B}$ derived from the curl of $\mathbf{A}$, and hence the equation for $\mathbf{A}$ is

$$\nabla^2 \mathbf{A} = -\mu \mathbf{J} \tag{2.41}$$

This equation is a vector Poisson's equation. In rectangular coordinates, (2.41) represents three scalar Poisson's equations, the first being

$$\nabla^2 A_x = -\mu J_x \tag{2.42}$$

In a curvilinear coordinate system, such as a cylindrical coordinate system, (2.41) cannot be written in such a simple component form. The reason is that, for example, $\nabla^2 \mathbf{a}_r A_r$ does not equal $\mathbf{a}_r \nabla^2 A_r$ because, even though the unit vector $\mathbf{a}_r$ is of constant length, its orientation varies from point to point since it is always directed along the radius vector from the origin to the point under consideration. The evaluation of $\nabla^2\mathbf{A}$ in curvilinear coordinates is made by using the vector identity quoted above to give $\nabla^2\mathbf{A} = \nabla\nabla \cdot \mathbf{A} - \nabla \times \nabla \times \mathbf{A}$. These latter operations are readily carried out.

The interest in static field solutions at microwave frequencies arises because the field distribution over a cross-sectional plane of a transmission line is a static field distribution and because static field solutions are good approximate solutions to the actual fields in the vicinity of obstacles that are small compared with the wavelength. The potential theory introduced above may be extended to the time-varying case also, and this is done in a following section.

2.4 WAVE EQUATION

For convenience, the two curl equations are repeated here:

$$\nabla \times \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t} \tag{2.43a}$$

$$\nabla \times \mathscr{H} = \frac{\partial \mathscr{D}}{\partial t} \tag{2.43b}$$

where it is assumed for the present that the current density $\mathscr{J}$ is zero in the region of interest. These equations, together with the assumed constitutive relations $\mathscr{D} = \epsilon\mathscr{E}$, $\mathscr{B} = \mu\mathscr{H}$, may be combined to obtain a separate equation for each field. The curl of (2.43a) is

$$\nabla \times \nabla \times \mathscr{E} = -\frac{\partial(\nabla \times \mathscr{B})}{\partial t} = -\mu\frac{\partial(\nabla \times \mathscr{H})}{\partial t}$$

Using (2.43b) and expanding $\nabla \times \nabla \times \mathscr{E}$ now yields

$$\nabla\nabla \cdot \mathscr{E} - \nabla^2\mathscr{E} = -\mu\epsilon\frac{\partial^2\mathscr{E}}{\partial t^2}$$

Since ρ is assumed zero and ϵ is taken as a constant, $\nabla \cdot \mathscr{E} = 0$, and we obtain

$$\nabla^2\mathscr{E} - \mu\epsilon\frac{\partial^2\mathscr{E}}{\partial t^2} = 0 \tag{2.44}$$

which is a three-dimensional wave equation. The velocity of propagation v is equal to $(\mu\epsilon)^{-1/2}$. In free space v is equal to the velocity of light c. To illustrate the nature of the solutions of (2.44), consider a case where $\mathscr{E}$ has only an x component and depends only on the z coordinate. In this instance

$$\frac{\partial^2\mathscr{E}_x}{\partial z^2} - \mu\epsilon\frac{\partial^2\mathscr{E}_x}{\partial t^2} = 0$$

Any function of the form $f(z - vt)$ is a solution of this equation since

$$\frac{\partial^2 f}{\partial z^2} = f'' \qquad \frac{\partial^2 f}{\partial t^2} = v^2\frac{\partial^2 f}{\partial (vt)^2} = v^2 f''$$

and hence

$$\frac{\partial^2 f}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} = 0$$

This solution is illustrated in Fig. 2.6 and clearly represents a disturbance propagating in the positive z direction with velocity v. An equally valid solution is $f(z + vt)$ and represents a disturbance propagating in the negative z direction.

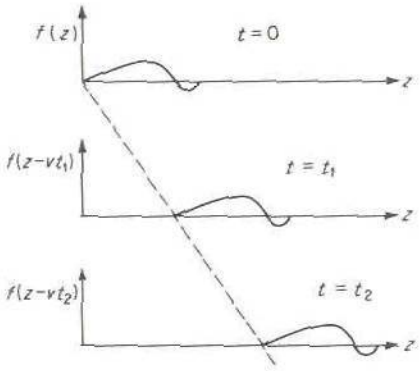


FIGURE 2.6
Propagation of a disturbance $f(z - vt)$.

By eliminating the electric field, it is readily found that the magnetic field $\mathscr{H}$ also satisfies the wave equation (2.44). In practice, however, we solve the wave equation for either $\mathscr{E}$ or $\mathscr{H}$ and then derive the other field by using the appropriate curl equation. When constitutive relations such as $\mathscr{D} = \epsilon\mathscr{E}$ and $\mathscr{B} = \mu\mathscr{H}$ cannot be written, the polarization vectors $\mathscr{P}$ and $\mathscr{M}$ must be exhibited explicitly in Maxwell's equations. Wave equations for $\mathscr{E}$ and $\mathscr{H}$ may still be derived, but $\mathscr{P}$ and $\mathscr{M}$ will now enter as equivalent sources for the field (which they actually are). The derivation is left as a problem at the end of this chapter.

For harmonic time dependence, the equation obtained in place of (2.44) is

$$\nabla^2\mathbf{E} + k^2\mathbf{E} = 0 \tag{2.45}$$

where $k^2 = \omega^2\mu\epsilon$. This equation is referred to as the Helmholtz equation, or reduced wave equation. The constant k is called the wave number and may be expressed in the form

$$k = \omega\sqrt{\mu\epsilon} = \frac{\omega}{v} = 2\pi\frac{f}{v} = \frac{2\pi}{\lambda} \tag{2.46}$$

where the wavelength λ is equal to v/f. In free space the wave number will be written as k_0, and is equal to $\omega\sqrt{\mu_0\epsilon_0} = 2\pi/\lambda_0$. The magnetic field $\mathbf{H}$, as may be surmised, satisfies the same reduced wave equation.

In a medium with finite conductivity σ, a conduction current $\mathscr{J} = \sigma\mathscr{E}$ will exist, and this results in energy loss because of Joule heating. The wave equation in media of this type has a damping term proportional to σ and the first time derivative of the field. In metals, excluding ferromagnetic materials, the permittivity and permeability are essentially equal to their free-space values, at least for frequencies up to and including the microwave range. Thus Maxwell's curl equations become

$$\nabla \times \mathscr{E} = -\mu_0\frac{\partial\mathscr{H}}{\partial t} \qquad \nabla \times \mathscr{H} = \epsilon_0\frac{\partial\mathscr{E}}{\partial t} + \sigma\mathscr{E}$$

Elimination of $\mathscr{H}$ in the same manner as before leads to the following wave equation for $\mathscr{E}$:

$$\nabla^2\mathscr{E} - \mu_0\sigma\frac{\partial\mathscr{E}}{\partial t} - \mu_0\epsilon_0\frac{\partial^2\mathscr{E}}{\partial t^2} = 0 \tag{2.47}$$

The magnetic field $\mathscr{H}$ also satisfies this equation. For the time-harmonic case damping effects enter in through the complex nature of ϵ and μ, and hence the wave number k. It should be recalled here that, as shown by (2.27), a finite conductivity σ is equivalent to an imaginary term in the permittivity ϵ. In the present case the equivalent permittivity is $\epsilon = \epsilon_0 - j\sigma/\omega$ and the Helmholtz equation is

$$\nabla^2\mathbf{E} + \omega^2\mu_0\epsilon_0\left(1 - j\frac{\sigma}{\omega\epsilon_0}\right)\mathbf{E} = 0 \tag{2.48}$$

In metals the conduction current $\sigma\mathbf{E}$ is generally very much larger than the displacement current $\omega\epsilon_0\mathbf{E}$, so that the latter may be neglected. For example, σ is equal to 5.8×10^7 S/m for copper, and at a frequency of 10^{10} Hz, $\omega\epsilon_0 = 0.55$, which is much smaller than σ. Only for frequencies in the optical range will the two become comparable. Thus (2.47) may be simplified to

$$\nabla^2\mathscr{E} - \mu_0\sigma\frac{\partial\mathscr{E}}{\partial t} = 0 \tag{2.49}$$

and (2.48) reduces to

$$\nabla^2\mathbf{E} - j\omega\mu_0\sigma\mathbf{E} = 0 \tag{2.50}$$

Equation (2.49) is a diffusion equation similar to that which governs the flow of heat in a thermal conductor.

2.5 ENERGY AND POWER

When currents exist in conductors as a result of the application of a suitable potential source, energy is expended by the source in maintaining the currents. The energy supplied by the source is stored in the electric and magnetic fields set up by the currents or propagated (radiated) away in the form of an electromagnetic wave. Under steady-state sinusoidal time-varying conditions, the time-average energy stored in the electric field is

$$W_e = \mathrm{Re}\,\frac{1}{4}\int_V \mathbf{E}\cdot\mathbf{D}^*\,dV = \frac{1}{4}\int_V \epsilon'\mathbf{E}\cdot\mathbf{E}^*\,dV \tag{2.51a}$$

If ϵ is a constant and real, (2.51a) becomes

$$W_e = \frac{\epsilon}{4}\int_V \mathbf{E}\cdot\mathbf{E}^*\,dV \tag{2.51b}$$

The time-average energy stored in the magnetic field is given by

$$W_m = \text{Re}\,\frac{1}{4}\int_V \mathbf{H}^* \cdot \mathbf{B}\, dV = \frac{1}{4}\int_V \mu' \mathbf{H} \cdot \mathbf{H}^*\, dV \qquad (2.52a)$$

which, for μ real and constant, becomes

$$W_m = \frac{\mu}{4}\int_V \mathbf{H} \cdot \mathbf{H}^*\, dV \qquad (2.52b)$$

These expressions for W_e and W_m are valid only for nondispersive media, i.e., media for which ϵ and μ can be considered independent of ω in the vicinity of the angular frequency ω with which the fields vary. In general, when the losses are small, so that $\epsilon'' \ll \epsilon'$ and $\mu'' \ll \mu'$, we have

$$W_e = \frac{1}{4}\int_V \mathbf{E} \cdot \mathbf{E}^* \frac{\partial \omega \epsilon'}{\partial \omega}\, dV \qquad (2.53a)$$

$$W_m = \frac{1}{4}\int_V \mathbf{H} \cdot \mathbf{H}^* \frac{\partial \omega \mu'}{\partial \omega}\, dV \qquad (2.53b)$$

for the time-average stored electric and magnetic energy.

The above equations for the time-average energy in a dispersive medium may be established by considering a classical model of the polarization mechanism similar to that discussed in Sec. 2.2. In a unit volume let the effective oscillating charge of the dipole distribution be $-q$ with an effective mass m. Let the damping force be equal to $m\nu$ times the velocity of the charge. This damping force takes account of collision effects and loss of energy by radiation from the oscillating charge. The equation of motion for the polarization charge displacement u is

$$m\frac{d^2u}{dt^2} + m\nu\frac{du}{dt} + ku = -q\mathscr{E}$$

where u is parallel to the direction of the field $\mathscr{E}$. In this equation k is the elastic constant giving rise to the restoring force. This constant arises from the Coulomb forces acting on the displaced charge, and hence is of electrical origin. The dipole polarization $\mathscr{P}$ is $-qu$, and the polarization current $\mathscr{J}_p = d\mathscr{P}/dt$. Introducing the polarization current into the equation of motion gives

$$\frac{m}{q^2}\frac{d\mathscr{J}_p}{dt} + \frac{m\nu}{q^2}\mathscr{J}_p + \frac{k}{q^2}\int^t \mathscr{J}_p\, dt = \mathscr{E}$$

This equation is formally the same as that which describes the current in a series LCR circuit with an applied voltage $\mathscr{V}$ equal to $\mathscr{E}$ and with

$$L = \frac{m}{q^2} \qquad R = \frac{m\nu}{q^2} \qquad C = \frac{q^2}{k}$$

An equivalent circuit describing the polarization is illustrated in Fig. 2.7. If

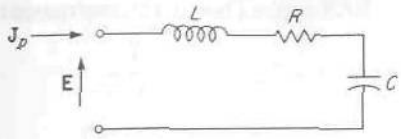


FIGURE 2.7
Equivalent circuit for polarization current.

a time dependence $e^{j\omega t}$ is assumed and phasor notation is used,

$$J_p = EY = E\frac{R - jX}{R^2 + X^2}$$

where Y is the input admittance and $X = \omega L - 1/\omega C$. Since $P = \epsilon_0 \chi_e E$ and $J_p = j\omega P$, we see that

$$\omega\epsilon_0\chi_e = \omega\epsilon_0(\chi_e' - j\chi_e'') = -jY = \frac{-X - jR}{R^2 + X^2}$$

and hence

$$\omega\epsilon_0\chi_e' = \frac{-X}{R^2 + X^2} \tag{2.54a}$$

$$\omega\epsilon_0\chi_e'' = \frac{R}{R^2 + X^2} \tag{2.54b}$$

The time-average power loss associated with the polarization is the same as the power loss in R in the equivalent circuit. This is given by

$$P_l = \frac{1}{2}EE^*\frac{R}{R^2 + X^2} = \frac{1}{2}EE^*\omega\epsilon_0\chi_e'' \tag{2.55}$$

per unit volume. This equation shows that $\omega\epsilon_0\chi_e'' = \omega\epsilon''$ is an equivalent conductance. The time-average energy stored in the system is of two forms. First there is the kinetic energy of motion, that is, $\frac{1}{2}m(du/dt)^2$ averaged over a cycle, and this is equal to the magnetic energy stored in the inductor in the equivalent circuit. This time-average kinetic energy per unit volume is given by

$$U_m = \frac{1}{4}LJ_pJ_p^* = \frac{1}{4}EE^*\frac{L}{R^2 + X^2} \tag{2.56a}$$

The second form of stored energy is the potential energy associated with the charge displacement. The time-average value of this energy is equal to the time-average electric energy stored in the capacitor C in the equivalent circuit, and is given by

$$U_e = \frac{1}{4}EE^*\frac{1}{\omega^2 C(R^2 + X^2)} \tag{2.56b}$$

The total time-average energy stored per unit volume is $U = U_m + U_e$. Note

that U is not given by $\frac{1}{4}EE^*\epsilon_0\chi_e'$. The latter expression gives

$$\frac{1}{4}EE^*\epsilon_0\chi_e' = \frac{1}{4}EE^*\frac{-X}{\omega(R^2+X^2)}$$

$$= \frac{1}{4}EE^*\frac{1/\omega^2C - L}{R^2+X^2} = U_e - U_m$$

or the difference between the potential and kinetic energy stored.

To obtain an expression for the total stored energy, note that

$$\frac{d}{d\omega}\left(\frac{X}{R^2+X^2}\right) = \frac{L+1/\omega^2C}{R^2+X^2}\left(1 - \frac{2X^2}{R^2+X^2}\right)$$

For a low-loss system, $R^2 \ll X^2$, and we then have $1 - 2X^2/(R^2+X^2) \approx -1$; so

$$\frac{d}{d\omega}\left(\frac{-X}{R^2+X^2}\right) = \frac{d}{d\omega}(\omega\epsilon_0\chi_e') \approx \frac{L+1/\omega^2C}{R^2+X^2}$$

Multiplying this expression by $\frac{1}{4}EE^*$ now gives the total time-average energy stored, as comparison with (2.56*a*) and (2.56*b*) shows. Thus the final expression for the time-average electric energy stored in a volume V is given by the volume integral of $U = U_e + U_m$ plus the free-space energy density $\epsilon_0(\mathbf{E}\cdot\mathbf{E}^*)/4$ and is

$$W_e = \int_V\left(U + \frac{\epsilon_0}{4}\mathbf{E}\cdot\mathbf{E}^*\right)dV$$

$$= \int_V \frac{\mathbf{E}\cdot\mathbf{E}}{4}\left(\epsilon_0 + \frac{\partial\omega\epsilon_0\chi_e'}{\partial\omega}\right)dV$$

$$= \frac{1}{4}\int_V \mathbf{E}\cdot\mathbf{E}^*\frac{\partial\omega\epsilon'}{\partial\omega}\,dV$$

since $\epsilon' = \epsilon_0(1+\chi_e')$. This equation is the result given earlier by (2.53*a*).

A similar type of model may be used to establish (2.53*b*) for the average stored magnetic energy. It should be pointed out that under time-varying conditions the average stored energy associated with either electric or magnetic polarization includes a kinetic-energy term. This term is negligible at low frequencies and also when ϵ' and μ' are essentially independent of ω for the range of ω of interest. When this energy is not negligible, the modified expressions for stored energy must be used.

Although (2.53) is more general than (2.51) and (2.52), we shall, in the majority of instances, use the latter equations for the stored energy. We thereby tacitly assume that we are dealing with material that is nondispersive or very nearly so.

The time-average power transmitted across a closed surface S is given by the integral of the real part of one-half of the normal component of the

complex Poynting vector $\mathbf{E} \times \mathbf{H}^*$; that is,

$$P = \text{Re}\,\frac{1}{2}\oint_S \mathbf{E} \times \mathbf{H}^* \cdot d\mathbf{S} \tag{2.57}$$

The above results are obtained from the interpretation of the complex Poynting vector theorem, which may be derived from Maxwell's equations as follows: If the divergence of $\mathbf{E} \times \mathbf{H}^*$, that is, $\nabla \cdot \mathbf{E} \times \mathbf{H}^*$, is expanded, we obtain

$$\nabla \cdot \mathbf{E} \times \mathbf{H}^* = (\nabla \times \mathbf{E}) \cdot \mathbf{H}^* - (\nabla \times \mathbf{H}^*) \cdot \mathbf{E}$$

From Maxwell's equations $\nabla \times \mathbf{E} = -j\omega\mathbf{B}$ and $\nabla \times \mathbf{H}^* = -j\omega\mathbf{D}^* + \mathbf{J}^*$, and hence

$$\nabla \cdot \mathbf{E} \times \mathbf{H}^* = -j\omega\mathbf{B} \cdot \mathbf{H}^* + j\omega\mathbf{D}^* \cdot \mathbf{E} - \mathbf{E} \cdot \mathbf{J}^*$$

The integration of this equation throughout a volume V bounded by a closed surface S gives the complex Poynting vector theorem; i.e.,

$$\begin{aligned}\frac{1}{2}\int_V \nabla \cdot \mathbf{E} \times \mathbf{H}^*\, dV &= \frac{1}{2}\oint_S \mathbf{E} \times \mathbf{H}^* \cdot d\mathbf{S} \\ &= -j\frac{\omega}{2}\int_V (\mathbf{B} \cdot \mathbf{H}^* - \mathbf{E} \cdot \mathbf{D}^*)\, dV - \frac{1}{2}\int_V \mathbf{E} \cdot \mathbf{J}^*\, dV\end{aligned} \tag{2.58a}$$

where the divergence theorem has been used on the left-hand side integral. The above result may be rewritten as

$$\begin{aligned}\frac{1}{2}\oint_S \mathbf{E} \times \mathbf{H}^* \cdot (-d\mathbf{S}) &= 2j\omega\int_V \left(\frac{\mathbf{B} \cdot \mathbf{H}^*}{4} - \frac{\mathbf{E} \cdot \mathbf{D}^*}{4}\right) dV \\ &\quad + \frac{1}{2}\int_V \mathbf{E} \cdot \mathbf{J}^*\, dV\end{aligned} \tag{2.58b}$$

where $-d\mathbf{S}$ is a vector element of area directed into the volume V. If the medium in V is characterized by parameters $\epsilon = \epsilon' - j\epsilon''$, $\mu = \mu' - j\mu''$, and conductivity σ, the real and imaginary parts of (2.58) may be equated to give

$$\begin{aligned}\text{Re}\,\frac{1}{2}\oint_S \mathbf{E} \times \mathbf{H}^* \cdot (-d\mathbf{S}) &= \frac{\omega}{2}\int_V (\mu''\mathbf{H} \cdot \mathbf{H}^* + \epsilon''\mathbf{E} \cdot \mathbf{E}^*)\, dV \\ &\quad + \frac{1}{2}\int_V \sigma\mathbf{E} \cdot \mathbf{E}^*\, dV\end{aligned} \tag{2.59a}$$

$$\text{Im}\,\frac{1}{2}\oint_S \mathbf{E} \times \mathbf{H}^* \cdot (-d\mathbf{S}) = 2\omega\int_V \left(\mu'\frac{\mathbf{H} \cdot \mathbf{H}^*}{4} - \epsilon'\frac{\mathbf{E} \cdot \mathbf{E}^*}{4}\right) dV \tag{2.59b}$$

Equation (2.59a) is interpreted to state that the real electromagnetic power transmitted through the closed surface S into V is equal to the power loss produced by conduction current $\sigma\mathbf{E}$, resulting in Joule heating plus the

power loss resulting from polarization damping forces. Note that $\omega\epsilon''$ could be interpreted as an equivalent conductance, as pointed out in Sec. 2.2. This equation also shows that μ'' and ϵ'' must be positive in order to represent energy loss, and hence the imaginary parts of ϵ and μ must be negative. Equation (2.59*b*) states that the imaginary part of the complex rate of energy flow into V is equal to 2ω times the net reactive energy $W_m - W_e$ stored in the magnetic and electric fields in V. The complex Poynting vector theorem is essentially an energy-balance equation.

A result analogous to the above may be derived for a conventional network, and serves to demonstrate the validity of the interpretation of (2.58). Consider a simple series RLC circuit as in Fig. 2.7. If the current in the circuit is I and the applied voltage is V, the complex input power is given by

$$\frac{1}{2}VI^* = \frac{1}{2}ZII^* = \frac{1}{2}II^*\left(R + j\omega L - \frac{j}{\omega C}\right)$$

The time-average power loss in R, magnetic energy stored in the field around L, and electric energy stored in the field associated with C are given, respectively, by

$$P_l = \frac{1}{2}RII^* \qquad W_m = \frac{1}{4}LII^* \qquad W_e = \frac{1}{4}\frac{II^*}{\omega^2 C}$$

since the voltage across C is $I/\omega C$. Hence

$$\frac{1}{2}VI^* = \frac{1}{2}ZII^* = P_l + 2j\omega(W_m - W_e)$$

which has the same interpretation as (2.58). This equation may also be solved for the impedance Z to give

$$Z = \frac{P_l + 2j\omega(W_m - W_e)}{\frac{1}{2}II^*} \tag{2.60}$$

and provides a general definition of the impedance of a network in terms of the associated power loss and stored reactive energy. The factor $\frac{1}{2}II^*$ in the denominator serves as a normalization factor, and is required in order to make Z independent of the magnitude of the current at the input to the network.

In the case of a general time-varying field, an expansion of $\nabla \cdot \mathscr{E} \times \mathscr{H}$ and substitution from Maxwell's equations (2.13) lead to the following Poynting vector theorem for general time-varying fields:

$$\oint_S \mathscr{E} \times \mathscr{H} \cdot (-d\mathbf{S}) = \int_V \left(\mu_0 \mathscr{H} \cdot \frac{\partial \mathscr{H}}{\partial t} + \mu_0 \mathscr{H} \cdot \frac{\partial \mathscr{M}}{\partial t} + \epsilon_0 \mathscr{E} \cdot \frac{\partial \mathscr{E}}{\partial t} + \mathscr{E} \cdot \frac{\partial \mathscr{P}}{\partial t} + \mathscr{E} \cdot \mathscr{J} \right) dV$$

Since $\mathscr{H} \cdot \partial\mathscr{H}/\partial t = \frac{1}{2}\partial(\mathscr{H} \cdot \mathscr{H})/\partial t$, etc., and the electric and magnetic polarization currents are $\mathscr{J}_p = \partial\mathscr{P}/\partial t$, $\mathscr{J}_m = \mu_0(\partial\mathscr{M}/\partial t)$, we have

$$\oint_S \mathscr{E} \times \mathscr{H} \cdot (-d\mathbf{S}) = \frac{\partial}{\partial t}\int_V \left(\frac{\mu_0 \mathscr{H} \cdot \mathscr{H}}{2} + \frac{\epsilon_0 \mathscr{E} \cdot \mathscr{E}}{2}\right) dV + \int_V \left[\mathscr{E} \cdot (\mathscr{J} + \mathscr{J}_p) + \mathscr{H} \cdot \mathscr{J}_m\right] dV \qquad (2.61)$$

where $-d\mathbf{S}$ is an element of surface area directed into V. This equation states that the rate of energy flow into V is equal to the time rate of change of the free-space field energy stored in V plus the rate of energy dissipation in Joule heating arising from the conduction current $\mathscr{J}$ and, in addition, the instantaneous rate of energy supplied in maintaining the polarization currents. If $\mathscr{M}$ and $\mathscr{H}$, and also $\mathscr{P}$ and $\mathscr{E}$, are in phase, there is no energy loss associated with the polarization currents. If these quantities are not in phase, some energy dissipation takes place, leading to increased heating of the material.

If the susceptibilities χ_e and χ_m can be considered as constants, so that $\partial\mathscr{P}/\partial t = \epsilon_0\chi_e(\partial\mathscr{E}/\partial t)$ and $\partial\mathscr{M}/\partial t = \chi_m(\partial\mathscr{H}/\partial t)$, then (2.61) becomes

$$\oint_S \mathscr{E} \times \mathscr{H} \cdot (-d\mathbf{S}) = \frac{\partial}{\partial t}\int_V \left(\frac{\mathscr{H} \cdot \mathscr{B}}{2} + \frac{\mathscr{E} \cdot \mathscr{D}}{2}\right) dV + \int_V \mathscr{E} \cdot \mathscr{J}\, dV \qquad (2.62)$$

which is the usual form of the Poynting vector theorem. The first term on the right is now interpreted as the instantaneous rate of change of the total electric and magnetic energy stored in the volume V.

The susceptibilities can usually be considered as true constants whenever the inertial and damping forces are small compared with the elastic restoring force in the dynamical equation describing the polarization. For example, with reference to (2.54a), this is the case when k is much greater than $\omega m\nu$ or $\omega^2 m$, that is, when $1/\omega C$ is large compared with ωL and R, so that

$$\epsilon_0\chi_e' \approx C = \frac{q^2}{k}$$

2.6 BOUNDARY CONDITIONS

In order to find the proper and unique solutions to Maxwell's equations for situations of practical interest (these always involve material bodies with boundaries), a knowledge of the behavior of the electromagnetic field at the boundary separating material bodies with different electrical properties is required. From a mathematical point of view, the solution of a partial differential equation, such as a wave equation, in a region V is not unique unless boundary conditions are specified, i.e., the behavior of the field on the boundary of V. Boundary conditions play the same role in the solution of

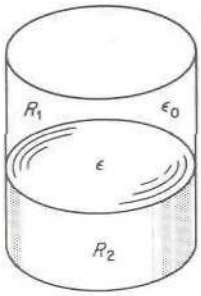


FIGURE 2.8
A cylindrical cavity partially filled with a dielectric medium.

partial differential equations that initial conditions play in the solution of the differential equations that govern the behavior of electric circuits.

As an example, consider the problem of finding a solution to Maxwell's equations inside a cylindrical cavity partially filled with a dielectric medium of permittivity ϵ, as in Fig. 2.8. In practice, the solution is obtained by finding general solutions valid in the two regions labeled R_1 and R_2. These general solutions must satisfy prescribed conditions on the metallic boundaries and in addition contain arbitrary amplitude constants that can be determined only from a knowledge of the boundary conditions to be applied at the air-dielectric boundary separating regions R_1 and R_2.

The integral form of Maxwell's equations provides the most convenient formulation in order to deduce the required boundary conditions. Consider two media with parameters ϵ_1, μ_1 and ϵ_2, μ_2, as in Fig. 2.9*a*. If there is no surface charge on the boundary, which is the usual case for nonconducting media, the integral of the displacement flux over the surface of the small "coin-shaped" volume centered on the boundary as in Fig. 2.9*b* gives, in the limit as h tends to zero,

$$\lim_{h \to 0} \oint_S \mathbf{D} \cdot d\mathbf{S} = D_{2n}\,\Delta S - D_{1n}\,\Delta S = 0$$

or

$$D_{2n} = D_{1n} = \mathbf{n} \cdot \mathbf{D}_2 = \mathbf{n} \cdot \mathbf{D}_1 \tag{2.63}$$

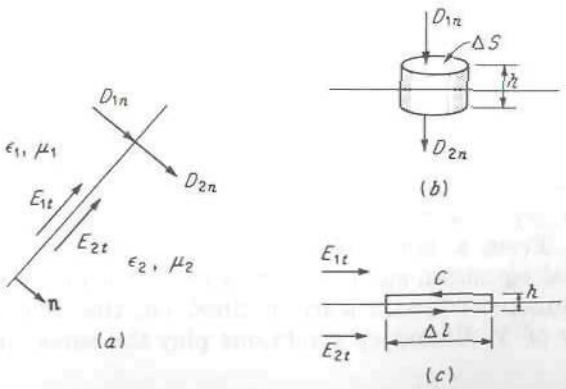


FIGURE 2.9
Boundary between two different media.

where n denotes the normal component. The limit $h \to 0$ is taken so that the flux through the sides of the coin-shaped region vanishes. Equation (2.63) simply states that the displacement flux lines are continuous in the direction normal to the boundary. A similar result clearly must hold also for the magnetic flux lines since $\nabla \cdot \mathbf{B} = 0$, and hence, by analogy,

$$\mathbf{n} \cdot \mathbf{B}_2 = \mathbf{n} \cdot \mathbf{B}_1 \tag{2.64}$$

To obtain boundary conditions on the tangential components of the electric field $\mathbf{E}$ and magnetic field $\mathbf{H}$, the circulation integrals for $\mathbf{E}$ and $\mathbf{H}$ are used. If for the contour C in Fig. 2.9*c*, the width h is made to approach zero, the magnetic flux flowing through this contour vanishes and

$$\lim_{h \to 0} \oint_C \mathbf{E} \cdot d\mathbf{l} = \lim_{h \to 0} -j\omega \int_S \mathbf{B} \cdot d\mathbf{S} = 0$$

$$= E_{2t}\,\Delta l - E_{1t}\,\Delta l$$

or

$$E_{1t} = E_{2t} \tag{2.65}$$

For the same contour C the total displacement current directed through the contour vanishes as $h \to 0$, so that

$$\lim_{h \to 0} \oint_C \mathbf{H} \cdot d\mathbf{l} = \lim_{h \to 0} \left(j\omega \int_S \mathbf{D} \cdot d\mathbf{S} \right) = 0$$

$$= (H_{2t} - H_{1t})\,\Delta l$$

or

$$H_{2t} = H_{1t} \tag{2.66}$$

where t denotes the components tangential to the boundary surface. These latter relations state that the components of $\mathbf{E}$ and $\mathbf{H}$ tangent to the boundary are continuous across the boundary; i.e., the tangential components on adjacent sides of the boundary are equal at the boundary surface.

For the boundary conditions at the surface separating a good conductor (any metal) and free space or air, some simplification is possible. As shown in a later section, the electromagnetic field can penetrate into a conductor only a minute distance at microwave frequencies. The field amplitude decays exponentially from its surface value according to e^{-u/δ_s}, where u is the normal distance into the conductor measured from the surface, and δ_s is called the skin depth. The skin depth is given by

$$\delta_s = \left(\frac{2}{\omega \mu \sigma} \right)^{1/2} \tag{2.67}$$

For copper ($\sigma = 5.8 \times 10^7$ S/m) at a frequency of 10^{10} Hz, the skin depth is 6.6×10^{-5} cm, truly a very small distance. Likewise, the current $\mathbf{J} = \sigma \mathbf{E}$ is concentrated near the surface. As the conductivity is made to approach infinity, δ_s approaches zero and the current is squeezed into a narrower and narrower region and in the limit $\sigma \to \infty$ becomes a true surface current. *Since the skin depth is so small at microwave frequencies for metals, the* approximation of infinite conductivity may be made with negligible error (an

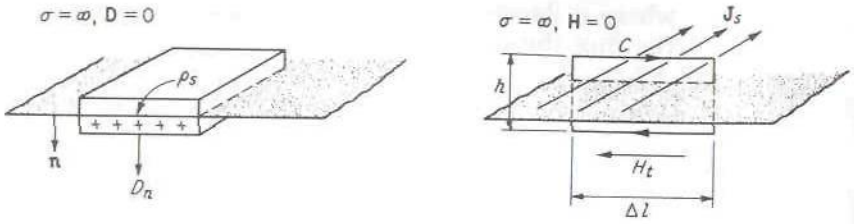


FIGURE 2.10
Boundary of a perfect conductor.

exception is when attenuation is to be calculated, since then infinite conductivity implies no loss). For infinite conductivity the field in the conductor must be zero. Since the flux lines of **B** are continuous and likewise since the tangential component of **E** is continuous across the boundary, it is necessary that

$$\mathbf{n} \cdot \mathbf{B} = 0 \tag{2.68a}$$

$$\mathbf{E}_t = \mathbf{n} \times \mathbf{E} = 0 \tag{2.68b}$$

at the surface of a perfect conductor. This same argument cannot be applied to the normal component of **D** and the tangential component of **H** because, as noted above, a surface current $\mathbf{J}_s$ will exist on the surface in the limit $\sigma \to \infty$. Applying Maxwell's equation

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = j\omega \int \mathbf{D} \cdot d\mathbf{S} + \int \mathbf{J} \cdot d\mathbf{S}$$

to the contour C illustrated in Fig. 2.10 gives

$$\lim_{h \to 0} \oint_C \mathbf{H} \cdot d\mathbf{l} = H_t \, \Delta l = \lim_{h \to 0} \int j\omega \mathbf{D} \cdot d\mathbf{S} + \lim_{h \to 0} \int \mathbf{J} \cdot d\mathbf{S}$$

$$= \lim_{h \to 0} hJ \, \Delta l = J_s \, \Delta l$$

or in vector form,

$$\mathbf{n} \times \mathbf{H} = \mathbf{J}_s \tag{2.68c}$$

Note that the field in the conductor goes to zero, that the total displacement current through C vanishes as $h \to 0$, but that hJ tends to the limiting value J_s as the conductivity is made infinite and h is made to approach zero. Associated with the surface current is a charge of density ρ_s on which the normal displacement flux lines terminate. Hence, at the surface of a perfect conductor,

$$\mathbf{n} \cdot \mathbf{D} = D_n = \rho_s \tag{2.68d}$$

When it is desired to take into account the large but finite conductivity (as would be the case in attenuation calculations), an impedance boundary

condition may be used with little error. The metallic surface exhibits a surface impedance Z_m, with equal resistive and inductive parts, given by

$$Z_m = \frac{1+j}{\sigma\delta_s} \tag{2.69}$$

At the surface a surface current exists, and the relation between this and the electric field tangent to the surface is

$$\mathbf{E}_t = Z_m \mathbf{J}_s \tag{2.70}$$

Note that the tangential electric field cannot be zero for finite conductivity, although it may be very small. Now $\mathbf{n} \times \mathbf{H} = \mathbf{J}_s$, so that

$$\mathbf{E}_t = Z_m \mathbf{J}_s = Z_m \mathbf{n} \times \mathbf{H} \tag{2.71}$$

From (2.69) it is seen that the resistive part of the surface impedance is equal to the dc resistance per square of a unit square of metal of thickness δ_s. In a later section the above results are verified; so further comments are reserved until then.

In practice, it suffices to make the tangential components of the fields satisfy the proper boundary conditions since, when they do, the normal components of the fields automatically satisfy their appropriate boundary conditions. The reason is that when the fields are a solution of Maxwell's equations, not all the components of the field are independent. For example, when the tangential part of the electric field is continuous across a boundary, the derivatives of the tangential component of electric field with respect to coordinates on the boundary surface are also continuous. Thus the curl of the electric field normal to the surface is continuous, and this implies continuity of the normal component of $\mathbf{B}$. More specifically, if the xy plane is the boundary surface and E_x, E_y are continuous, then $\partial E_x/\partial x$, $\partial E_x/\partial y$, $\partial E_y/\partial x$, and $\partial E_y/\partial y$ are also continuous. Hence $-j\omega B_z = \partial E_y/\partial x - \partial E_x/\partial y$ is continuous. For the same reasons continuity of the tangential components of $\mathbf{H}$ ensures the continuity of the normal component of $\mathbf{D}$ across a boundary.

In addition to the boundary conditions given above, a boundary condition must be imposed on the field solutions at the edge of a conducting body such as a wedge. The edge condition requires that the energy stored in the field in the vicinity of an edge of a conducting body be finite. This limits the maximum rate at which the field intensities can increase as the edge is approached.† A detailed analysis shows that at the edge of a two-dimensional perfectly conducting wedge with an internal angle ϕ, the field components normal to the edge must not increase any faster than r^α, where r is

†J. Meixner, The Behavior of Electromagnetic Fields at Edges, *N.Y. Univ. Inst. Math. Sci. Res. Rept.*, vol. EM-72, December, 1954. The theory is also discussed in R. E. Collin, "Field Theory of Guided Waves," chap. 1, IEEE Press, Piscataway, N.J., 1991, revised edition.

the perpendicular radial distance away from the edge and

$$\alpha = \frac{n\pi}{2\pi - \phi} - 1$$

where the integer n must be chosen so that α is greater than or equal to $-\frac{1}{2}$ at least.

When solving for fields in an infinite region of space, the behavior of the field at infinity must also be specified. This boundary condition is called a radiation condition, and requires that the field at infinity be a wave propagating a finite amount of energy outward, or else that the field vanish so fast that the energy stored in the field and the energy flow at infinity are zero.

2.7 PLANE WAVES

In this section and the two following ones we shall introduce wave solutions *by considering plane waves propagating in free space and reflection of a* plane wave from a boundary separating free space and a dielectric, or conducting, medium. The latter problem will serve to derive the boundary conditions given by (2.68) to (2.71) in the preceding section.

Plane Waves in Free Space

The electric field is a solution of the Helmholtz equation

$$\nabla^2\mathbf{E} + k_0^2\mathbf{E} = \frac{\partial^2\mathbf{E}}{\partial x^2} + \frac{\partial^2\mathbf{E}}{\partial y^2} + \frac{\partial^2\mathbf{E}}{\partial z^2} + k_0^2\mathbf{E} = 0$$

This vector equation holds for each component of $\mathbf{E}$, so that

$$\frac{\partial^2 E_i}{\partial x^2} + \frac{\partial^2 E_i}{\partial y^2} + \frac{\partial^2 E_i}{\partial z^2} + k_0^2 E_i = 0 \qquad i = x, y, z \tag{2.72}$$

The standard procedure for solving a partial differential equation is the method of *separation of variables*. However, this method does not work for all types of partial differential equations in all various coordinate systems, and when it does not work, a solution is very difficult, if not impossible, to obtain. For the Helmholtz equation the method of separation of variables does work in such common coordinate systems as rectangular, cylindrical, and spherical. Hence this method suffices for the class of problems discussed in this text. The basic procedure is to assume for the solution a product of functions each of which is a function of one coordinate variable only. Substitution of this solution into the partial differential equation then separates the partial differential equation into three ordinary differential equations which may be solved by standard means.

In the present case let $E_x = f(x)g(y)h(z)$. Substituting this expression into (2.72) gives

$$ghf'' + fhg'' + fgh'' + k_0^2 fgh = 0$$

where the double prime denotes the second derivative. Dividing this equation by fgh gives

$$\frac{f''}{f} + \frac{g''}{g} + \frac{h''}{h} + k_0^2 = 0 \tag{2.73}$$

Each of the first three terms in (2.73), such as f''/f, is a function of a single independent variable only, and hence the sum of these terms can equal a constant $-k_0^2$ only if each term is constant. Thus (2.73) separates into three equations:

$$\frac{f''}{f} = -k_x^2 \qquad \frac{g''}{g} = -k_y^2 \qquad \frac{h''}{h} = -k_z^2$$

or

$$\frac{d^2f}{dx^2} + k_x^2 f = 0 \qquad \frac{d^2g}{dy^2} + k_y^2 g = 0 \qquad \frac{d^2h}{dz^2} + k_z^2 h = 0 \tag{2.74}$$

where k_x^2, k_y^2, k_z^2 are called separation constants. The only restriction so far on k_x^2, k_y^2, k_z^2 is that their sum must equal k_0^2, that is,

$$k_x^2 + k_y^2 + k_z^2 = k_0^2 \tag{2.75}$$

so that (2.73) will be satisfied.

Equations (2.74) are simple-harmonic differential equations with exponential solutions of the form $e^{\pm jk_x x}$, $e^{\pm jk_y y}$, $e^{\pm jk_z z}$. As one suitable solution for E_x we may therefore choose

$$E_x = Ae^{-jk_x x - jk_y y - jk_z z} \tag{2.76}$$

where A is an amplitude factor. This solution is interpreted as the x component of a wave propagating in the direction specified by the propagation vector

$$\mathbf{k} = \mathbf{a}_x k_x + \mathbf{a}_y k_y + \mathbf{a}_z k_z \tag{2.77}$$

because the scalar product of $\mathbf{k}$ with the position vector

$$\mathbf{r} = \mathbf{a}_x x + \mathbf{a}_y y + \mathbf{a}_z z$$

equals $k_x x + k_y y + k_z z$ and is k_0 times the perpendicular distance from the origin to a plane normal to the vector $\mathbf{k}$, as illustrated in Fig. 2.11. The $\mathbf{k}$ vector may also be written as $\mathbf{k} = \mathbf{n}k_0$, where $\mathbf{n}$ is a unit vector in the direction of $\mathbf{k}$ and k_0 is the magnitude of $\mathbf{k}$ by virtue of (2.75).

Although (2.76) gives a possible solution for E_x, this is not the complete solution for the electric field. Similar solutions for E_y and E_z may be found. The three components of $\mathbf{E}$ are not independent since the divergence relation $\nabla \cdot \mathbf{E} = 0$ must hold in free space. This constraint means

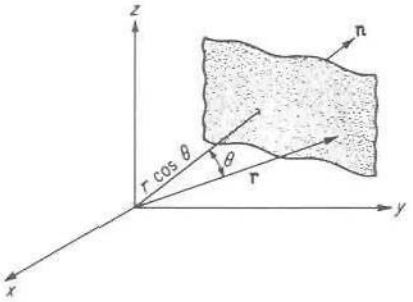


FIGURE 2.11
Illustration of plane normal to vector **k**.

that only two components of **E** can have arbitrary amplitudes. However, for $\nabla \cdot \mathbf{E}$ to vanish everywhere, all components of **E** must have the same spatial dependence, and hence appropriate solutions for E_y and E_z are

$$E_y = Be^{-j\mathbf{k}\cdot\mathbf{r}} \qquad E_z = Ce^{-j\mathbf{k}\cdot\mathbf{r}}$$

with B and C amplitude coefficients. Let $\mathbf{E}_0$ be the vector $\mathbf{a}_x A + \mathbf{a}_y B + \mathbf{a}_z C$; then the total solution for **E** may be written in vector form as

$$\mathbf{E} = \mathbf{E}_0 e^{-j\mathbf{k}\cdot\mathbf{r}} \tag{2.78}$$

The divergence condition gives

$$\nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{E}_0 e^{-j\mathbf{k}\cdot\mathbf{r}} = \mathbf{E}_0 \cdot \nabla e^{-j\mathbf{k}\cdot\mathbf{r}} = -j\mathbf{k} \cdot \mathbf{E}_0 e^{-j\mathbf{k}\cdot\mathbf{r}} = 0$$

or

$$\mathbf{k} \cdot \mathbf{E}_0 = 0 \tag{2.79}$$

since $\nabla e^{-j\mathbf{k}\cdot\mathbf{r}} = -j\mathbf{k}e^{-j\mathbf{k}\cdot\mathbf{r}}$, as may be verified by expansion in rectangular coordinates. The divergence condition is seen to constrain the amplitudes A, B, C so that the vector $\mathbf{E}_0$ is perpendicular to the direction of propagation as specified by **k**. The solution (2.78) is called a uniform plane wave since the constant-phase surfaces given by $\mathbf{k} \cdot \mathbf{r} = \text{const}$ are planes and the field **E** does not vary on a constant-phase plane.

The solution for **H** is obtained from Maxwell's equation

$$\nabla \times \mathbf{E} = -j\omega\mu_0 \mathbf{H}$$

which gives

$$\begin{aligned}\mathbf{H} &= -\frac{1}{j\omega\mu_0} \nabla \times \mathbf{E}_0 e^{-j\mathbf{k}\cdot\mathbf{r}} = \frac{1}{j\omega\mu_0} \mathbf{E}_0 \times \nabla e^{-j\mathbf{k}\cdot\mathbf{r}} \\ &= \frac{1}{\omega\mu_0} \mathbf{k} \times \mathbf{E}_0 e^{-j\mathbf{k}\cdot\mathbf{r}} = \frac{k_0}{\omega\mu_0} \mathbf{n} \times \mathbf{E} \\ &= \sqrt{\frac{\epsilon_0}{\mu_0}}\, \mathbf{n} \times \mathbf{E} = Y_0 \mathbf{n} \times \mathbf{E}\end{aligned} \tag{2.80}$$

where $Y_0 = \sqrt{\epsilon_0/\mu_0}$ has the dimensions of an admittance and is called the intrinsic admittance of free space. The reciprocal $Z_0 = 1/Y_0$ is called the

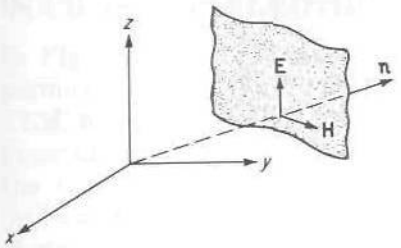


FIGURE 2.12
Space relationship between **E**, **H**, and **n** in a TEM wave.

intrinsic impedance of free space. Note that **H** is perpendicular to **E** and to **n**, and hence both **E** and **H** lie in the constant-phase planes. For this reason this type of wave is called a transverse electromagnetic wave (TEM wave). The spatial relationship between **E**, **H**, and **n** is illustrated in Fig. 2.12.

The physical electric field corresponding to the phasor representation (2.78) is

$$\mathbf{E} = \mathrm{Re}(\mathbf{E}_0 e^{-j\mathbf{k}\cdot\mathbf{r}+j\omega t}) = \mathbf{E}_0 \cos(\mathbf{k}\cdot\mathbf{r} - \omega t) \tag{2.81}$$

where, for simplicity, $\mathbf{E}_0$ has been assumed to be real. The wavelength is the distance the wave must propagate to undergo a phase change of 2π. If we let λ_0 denote the wavelength in free space, it follows that

$$|\mathbf{k}|\lambda_0 = k_0\lambda_0 = 2\pi$$

so that

$$k_0 = \omega\sqrt{\mu_0\epsilon_0} = \frac{\omega}{c} = \frac{2\pi}{\lambda_0} \tag{2.82}$$

This result is the familiar relationship between wavelength λ_0, frequency $f = \omega/2\pi$, and velocity c in free space. A wavelength in a direction other than that along the direction of propagation **n** may also be defined. For example, along the direction of the x axis the wavelength is

$$\lambda_x = \frac{2\pi}{k_x} \tag{2.83}$$

and since k_x is less than k_0, λ_x is greater than λ_0. The phase velocity is the velocity with which an observer would have to move in order to see a constant phase. From (2.81) it is seen that the phase of **E** is constant as long as $\mathbf{k}\cdot\mathbf{r} - \omega t$ is constant. If the angle between **k** and **r** is θ, then $\mathbf{k}\cdot\mathbf{r} - \omega t = k_0 r\cos\theta - \omega t$. Differentiating the relation

$$k_0 r\cos\theta - \omega t = \text{const}$$

gives

$$\frac{dr}{dt} = v_p = \frac{\omega}{k_0\cos\theta} \tag{2.84}$$

for the phase velocity v_p in the direction **r**. Along the direction of propagation, $\cos\theta = 1$ and $v_p = \omega/k_0 = c$. In other directions, the phase velocity is

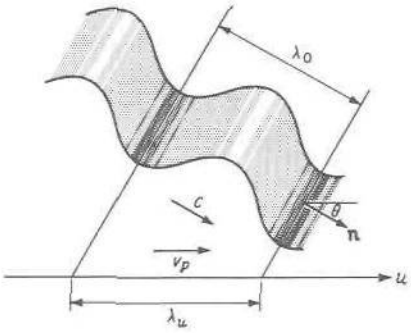


FIGURE 2.13
A wave propagating obliquely to the u axis.

greater than c. These results may be understood by reference to Fig. 2.13. When the wave has moved a distance λ_0 along the direction $\mathbf{n}$, the constant-phase-plane intersection with the u axis has moved a distance $\lambda_u = \lambda_0 \sec\theta$ along the direction u. For this reason the wavelength and phase velocity along u are greater by a factor $\sec\theta$ than the corresponding quantities measured along the direction of propagation $\mathbf{n}$.

The time-average rate of energy flow per unit area in the direction $\mathbf{n}$ is given by

$$P = \tfrac{1}{2}\,\mathrm{Re}\,\mathbf{E}\times\mathbf{H}^*\cdot\mathbf{n} = \tfrac{1}{2}\,\mathrm{Re}\,Y_0\mathbf{E}\times(\mathbf{n}\times\mathbf{E}^*)\cdot\mathbf{n} = \tfrac{1}{2}Y_0\mathbf{E}_0\cdot\mathbf{E}_0^* \quad (2.85)$$

The time-average energy densities in the electric and magnetic fields of a TEM wave are, respectively,

$$U_e = \frac{\epsilon_0}{4}\mathbf{E}\cdot\mathbf{E}^* = \frac{\epsilon_0}{4}\mathbf{E}_0\cdot\mathbf{E}_0^*$$

$$U_m = \frac{\mu_0}{4}\mathbf{H}\cdot\mathbf{H}^* = \frac{\mu_0}{4}Y_0^2(\mathbf{n}\times\mathbf{E})\cdot(\mathbf{n}\times\mathbf{E}^*) = \frac{\epsilon_0}{4}\mathbf{E}_0\cdot\mathbf{E}_0^* = U_e$$

and are seen to be equal. Since power is a flow of energy, the velocity v_g of energy propagation is such that

$$(U_e + U_m)v_g = P$$

or

$$v_g = \frac{P}{U_e + U_m} = \frac{\frac{1}{2}Y_0\mathbf{E}_0\cdot\mathbf{E}_0^*}{\frac{1}{2}\epsilon_0\mathbf{E}_0\cdot\mathbf{E}_0^*} = \frac{Y_0}{\epsilon_0} = c \quad (2.86)$$

Thus, for a TEM wave in free space, the energy in the field is transported with a velocity $c = 3\times10^8$ m/s, which is also the phase velocity. Since the phase velocity is independent of frequency, a modulated carrier or signal will have all its frequency components propagated with the same velocity c. Hence the signal velocity is also the velocity of light c. Later on, in the study of waveguides, situations arise where the phase velocity is dependent on frequency and consequently is not equal to the velocity of energy propagation or the signal velocity.

2.8 REFLECTION FROM A DIELECTRIC INTERFACE

In Fig. 2.14 the half-space $z \geq 0$ is filled with a dielectric medium with permittivity ϵ (dielectric constant $\epsilon_r = \epsilon/\epsilon_0$; index of refraction $\eta = \sqrt{\epsilon_r}$). A TEM wave is assumed incident from the region $z < 0$. Without loss in generality, the xy axis may be oriented so that the unit vector $\mathbf{n}_1$ specifying the direction of incidence lies in the xz plane. It is convenient to solve this problem as two special cases, namely (1) parallel polarization, where the electric field of the incident wave is coplanar with $\mathbf{n}_1$ and the interface normal, i.e., lies in the xz plane, and (2) perpendicular polarization, where the electric field of the incident wave is perpendicular to the plane of incidence as defined by $\mathbf{n}_1$ and the interface normal, i.e., along the y axis. An incident TEM wave with arbitrary polarization can always be decomposed into a linear sum of perpendicular and parallel polarized waves. The reason for treating the two polarizations separately is that the reflection and transmission coefficients, to be defined, are different for the two cases.

1 Parallel Polarization

Let the incident TEM wave be

$$\mathbf{E}_i = \mathbf{E}_1 e^{-jk_0 \mathbf{n}_1 \cdot \mathbf{r}} \qquad \mathbf{H}_i = Y_0 \mathbf{n}_1 \times \mathbf{E}_i \tag{2.87}$$

where $\mathbf{E}_1$ lies in the xz plane. Part of the incident power will be reflected, and the remainder will be transmitted into the dielectric medium. Let the reflected TEM wave be

$$\mathbf{E}_r = \mathbf{E}_2 e^{-jk_0 \mathbf{n}_2 \cdot \mathbf{r}} \qquad \mathbf{H}_r = Y_0 \mathbf{n}_2 \times \mathbf{E}_r \tag{2.88}$$

where $\mathbf{n}_2$ and $\mathbf{E}_2$ are to be determined. In the dielectric medium the solution for a TEM wave is the same as that in free space, but with ϵ_0 replaced by ϵ. Thus, in place of $k_0 = \omega\sqrt{\mu_0\epsilon_0}$ and $Y_0 = \sqrt{\epsilon_0/\mu_0}$, the parameters $k = \omega\sqrt{\mu_0\epsilon} = \eta k_0$ and $Y = \sqrt{\epsilon/\mu_0} = \eta Y_0$ are used, where $\eta = \sqrt{\epsilon_r}$ is the index of refraction. The transmitted wave in the dielectric may be

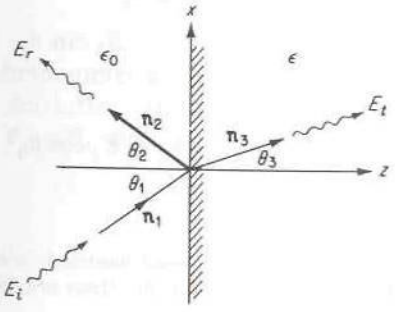


FIGURE 2.14
Plane wave incident on a dielectric interface.

expressed by

$$\mathbf{E}_t = \mathbf{E}_3 e^{-jk\mathbf{n}_3 \cdot \mathbf{r}} \qquad \mathbf{H}_t = Y\mathbf{n}_3 \times \mathbf{E}_t \tag{2.89}$$

with $\mathbf{E}_3$ *and* $\mathbf{n}_3$ *as yet unknown.*

The boundary conditions to be applied are the continuity of the tangential components of the electric and magnetic fields at the interface plane $z = 0$. These components must be continuous for all values of x and y on the $z = 0$ plane, and this is possible only if the fields on adjacent sides of the boundary have the same variation with x and y. Hence we must have

$$k_0 n_{1x} = k_0 n_{2x} = k n_{3x} = \eta k_0 n_{3x} \tag{2.90}$$

i.e., the propagation phase constant along x must be the same for all waves. Since n_{1y} was chosen as zero, it follows that $n_{2y} = n_{3y} = 0$ also. The unit vectors $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$ may be expressed as

$$\mathbf{n}_1 = \mathbf{a}_x \sin\theta_1 + \mathbf{a}_z \cos\theta_1$$

$$\mathbf{n}_2 = \mathbf{a}_x \sin\theta_2 + \mathbf{a}_z \cos\theta_2$$

$$\mathbf{n}_3 = \mathbf{a}_x \sin\theta_3 + \mathbf{a}_z \cos\theta_3$$

Equation (2.90) gives

$$\sin\theta_1 = \sin\theta_2$$

or

$$\theta_1 = \theta_2 \tag{2.91}$$

which is the well-known Snell's law of reflection; in addition, (2.90) gives

$$\sin\theta_1 = \eta \sin\theta_3 \tag{2.92}$$

which is also a well-known result specifying the angle of refraction θ_3 in terms of the angle of incidence θ_1 and the index of refraction η.

The incident electric field $\mathbf{E}_1$ has components $\mathbf{E}_{1x} = E_1 \cos\theta_1$,

$$E_{1z} = -E_1 \sin\theta_1$$

since $\mathbf{n}_1 \cdot \mathbf{E}_1$ *must equal zero. Note that* E_1 *is used to denote the magnitude of the vector* $\mathbf{E}_1$. Since the incident electric field has no y component, the reflected and transmitted electric fields also have zero y components.† Expressing all fields in component form, i.e.,

$$E_{2x} = E_2 \cos\theta_2$$

$E_{2z} = E_2 \sin\theta_2$, $E_{3x} = E_3 \cos\theta_3$, $E_{3z} = -E_3 \sin\theta_3$, and imposing the boundary condition of continuity of the x component at $z = 0$ yields the relation

$$E_1 \cos\theta_1 + E_2 \cos\theta_2 = E_3 \cos\theta_3$$

†If the reflected and transmitted electric fields were assumed to have a y component, the boundary conditions which must apply would show that these are, indeed, zero.

or
$$(E_1 + E_2)\cos\theta_1 = E_3\sqrt{1 - \sin^2\theta_3} = E_3\frac{\sqrt{\epsilon_r - \sin^2\theta_1}}{\eta} \tag{2.93}$$

by using (2.91) and (2.92). Apart from the propagation factor, the magnetic field is given by

$$\mathbf{H}_1 = Y_0\mathbf{n}_1 \times \mathbf{E}_1 = Y_0\mathbf{a}_y(-n_{1x}E_{1z} + n_{1z}E_{1x}) = Y_0\mathbf{a}_y E_1$$
$$\mathbf{H}_2 = -Y_0\mathbf{a}_y E_2$$
$$\mathbf{H}_3 = Y\mathbf{a}_y E_3$$

and has only a y component. Continuity of this magnetic field at the boundary requires that

$$Y_0(E_1 - E_2) = YE_3 = \eta Y_0 E_3 \tag{2.94}$$

If a reflection coefficient Γ_1 and a transmission coefficient T_1 are introduced according to the following relations:

$$\Gamma_1 = \frac{\text{amplitude of reflected electric field}}{\text{amplitude of incident electric field}} = \frac{E_2}{E_1} \tag{2.95a}$$

$$T_1 = \frac{\text{amplitude of transmitted electric field}}{\text{amplitude of incident electric field}} = \frac{E_3}{E_1} \tag{2.95b}$$

then the boundary conditions (2.93) and (2.94) may be expressed as

$$1 + \Gamma_1 = T_1\frac{(\epsilon_r - \sin^2\theta_1)^{1/2}}{\eta\cos\theta_1} \tag{2.96a}$$

$$1 - \Gamma_1 = \eta T_1 \tag{2.96b}$$

These equations may be solved to give the Fresnel reflection and transmission coefficients for the case of parallel polarization, namely,

$$\Gamma_1 = \frac{(\epsilon_r - \sin^2\theta_1)^{1/2} - \epsilon_r\cos\theta_1}{(\epsilon_r - \sin^2\theta_1)^{1/2} + \epsilon_r\cos\theta_1} \tag{2.97a}$$

$$T_1 = \frac{2\eta\cos\theta_1}{(\epsilon_r - \sin^2\theta_1)^{1/2} + \epsilon_r\cos\theta_1} \tag{2.97b}$$

An interesting feature of Γ_1 is that it vanishes for an angle of incidence $\theta_1 = \theta_b$, called the Brewster angle, where, from (2.97a),

$$\epsilon_r - \sin^2\theta_b = \epsilon_r^2\cos^2\theta_b$$

or
$$\sin\theta_b = \left(\frac{\epsilon_r}{\epsilon_r + 1}\right)^{1/2} \tag{2.98}$$

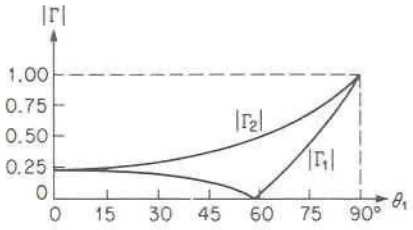


FIGURE 2.15
Modulus of reflection coefficient at a dielectric interface for $\epsilon_r = 2.56$, $|\Gamma_1|$ parallel polarization, $|\Gamma_2|$ perpendicular polarization.

At this particular angle of incidence all the incident power is transmitted into the dielectric medium. In Fig. 2.15 the reflection coefficient Γ_1 is plotted as a function of θ_1 for polystyrene, for which $\epsilon_r = 2.56$.

2 Perpendicular Polarization

For perpendicular polarization the roles of electric and magnetic fields are interchanged so that the electric field has only a y component. The fields may, however, still be expressed in the form given by (2.87) to (2.89), but with $\mathbf{E}_1$, $\mathbf{E}_2$, and $\mathbf{E}_3$ having y components only. As in the previous case, the boundary conditions must hold for all values of x and y on the $z = 0$ plane. Therefore Snell's laws of reflection and refraction again result; i.e., (2.91) and (2.92) must be satisfied. In place of the boundary conditions (2.93) and (2.94), we have

$$E_1 + E_2 = E_3 \tag{2.99a}$$

$$Y_0(E_1 - E_2)\cos\theta_1 = YE_3\cos\theta_3 \tag{2.99b}$$

Introducing the following reflection and transmission coefficients:

$$\Gamma_2 = \frac{E_2}{E_1} \qquad T_2 = \frac{E_3}{E_1}$$

into (2.99) yields

$$1 + \Gamma_2 = T_2 \tag{2.100a}$$

$$1 - \Gamma_2 = T_2\frac{(\epsilon_r - \sin^2\theta_1)^{1/2}}{\cos\theta_1} \tag{2.100b}$$

The Fresnel reflection and transmission coefficients for the case of perpendicular polarization thus are

$$\Gamma_2 = \frac{\cos\theta_1 - (\epsilon_r - \sin^2\theta_1)^{1/2}}{(\epsilon_r - \sin^2\theta_1)^{1/2} + \cos\theta_1} \tag{2.101a}$$

$$T_2 = \frac{2\cos\theta_1}{(\epsilon_r - \sin^2\theta_1)^{1/2} + \cos\theta_1} \tag{2.101b}$$

A notable difference for this case is the nonexistence of a Brewster angle for which Γ_2 vanishes. For comparison with the case of parallel polarization, Γ_2 is plotted in Fig. 2.15 for $\epsilon_r = 2.56$.

2.9 REFLECTION FROM A CONDUCTING PLANE

The essential features of the behavior of the electromagnetic field at the surface of a good conductor may be deduced from an analysis of the simple problem of a TEM wave incident normally onto a conducting plane. The problem is illustrated in Fig. 2.16, which shows a medium with parameters ϵ, μ, σ filling the half-space $z \geq 0$. Let the electric field be polarized along the x axis so that the incident and reflected fields may be expressed as

$$\begin{aligned} \mathbf{E}_i &= E_1 \mathbf{a}_x e^{-jk_0 z} \\ \mathbf{H}_i &= Y_0 E_1 \mathbf{a}_y e^{-jk_0 z} \\ \mathbf{E}_r &= \Gamma E_1 \mathbf{a}_x e^{+jk_0 z} \\ \mathbf{H}_r &= -Y_0 \Gamma E_1 \mathbf{a}_y e^{+jk_0 z} \end{aligned} \tag{2.102b}$$

where Γ is the reflection coefficient.

In the conducting medium the conduction current $\sigma \mathbf{E}$ is much greater than the displacement current $j\omega\epsilon\mathbf{E}$, so that Helmholtz's equation reduces to (2.50); i.e.,

$$\nabla^2 \mathbf{E} - j\omega\mu\sigma\mathbf{E} = 0$$

The transmitted field is a solution of

$$\left(\frac{\partial^2}{\partial z^2} - j\omega\mu\sigma\right)\mathbf{E}_t = 0$$

since no x or y variation is assumed. The solution for a wave with an x component only and propagating in the z direction is

$$\mathbf{E}_t = E_3 \mathbf{a}_x e^{-\gamma z} \tag{2.103a}$$

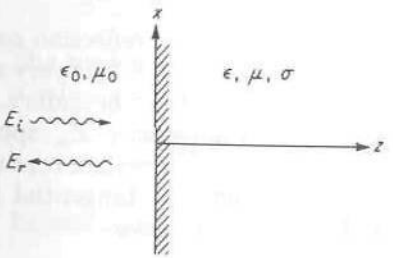


FIGURE 2.16
A TEM wave incident normally on a conducting plane.

with a corresponding magnetic field

$$\mathbf{H}_t = -\frac{1}{j\omega\mu}\nabla \times \mathbf{E}_t = \frac{\gamma}{j\omega\mu}\mathbf{a}_y E_3 e^{-\gamma z} \tag{2.103b}$$

where

$$\gamma = (j\omega\mu\sigma)^{1/2} = \frac{1+j}{\delta_s} \tag{2.104}$$

and the skin depth $\delta_s = (\omega\mu\sigma/2)^{-1/2}$. The propagation constant $\gamma = \alpha + j\beta$ has equal phase and attenuation constants. In the conductor the fields decay by an amount e^{-1} in a distance of one skin depth δ_s, which is a very small distance for metals at microwave frequencies (about 10^{-5} cm). The intrinsic impedance of the metal is Z_m, where

$$Z_m = \frac{j\omega\mu}{\gamma} = \frac{j\omega\mu}{(j\omega\mu\sigma)^{1/2}} = \frac{1+j}{\sigma\delta_s} \tag{2.105}$$

and is very small compared with the intrinsic impedance $Z_0 = (\mu_0/\epsilon_0)^{1/2}$ of free space. For example, for copper at 10^4 MHz, $Z_m = 0.026(1+j)\ \Omega$ as compared with 377 Ω for Z_0. Note that (2.103*b*) may be written as

$$\mathbf{H}_t = \frac{1}{Z_m}\mathbf{a}_y E_3 e^{-\gamma z} = Y_m \mathbf{a}_y E_3 e^{-\gamma z}$$

which shows that the ratio of the magnitudes of the electric field to the magnetic field for a TEM wave in a conductor is the intrinsic impedance Z_m.

Returning to the boundary-value problem and imposing the boundary conditions of continuity of tangential fields at the boundary plane $z = 0$ give

$$(1+\Gamma)E_1 = E_3 = TE_1 \tag{2.106a}$$

$$(1-\Gamma)Y_0 E_1 = H_3 = Y_m E_3 = Y_m T E_1 \tag{2.106b}$$

where $E_3/E_1 = T$, the transmission coefficient. Solving (2.106) for the reflection coefficient Γ and T yields

$$\Gamma = \frac{Z_m - Z_0}{Z_m + Z_0} \tag{2.107a}$$

$$T = 1 + \Gamma = \frac{2Z_m}{Z_m + Z_0} \tag{2.107b}$$

Since $|Z_m|$ is small compared with Z_0, the reflection coefficient Γ is almost equal to -1 and the transmission coefficient T is very small. Almost all the incident power is reflected from the metallic boundary. As the conductivity σ is made to approach infinity, the impedance Z_m approaches zero and in the limit $\Gamma = -1$ and $T = 0$. Hence, for a perfect conductor, the tangential electric field at the surface is zero and the tangential magnetic field has a value equal to twice that of the incident wave.

The current density in the conductor is $\mathbf{J} = \sigma \mathbf{E}_t = \sigma T E_1 \mathbf{a}_x e^{-\gamma z}$. The total current per unit width of conductor along y is

$$\mathbf{J}_s = \int_0^\infty \mathbf{J}\, dz = \sigma T E_1 \mathbf{a}_x \int_0^\infty e^{-\gamma z}\, dz = \frac{\sigma T E_1 \mathbf{a}_x}{\gamma} \text{ A/m}$$

This result may also be expressed in the following form:

$$\mathbf{J}_s = \frac{2\sigma Z_m^2 E_1}{(Z_m + Z_0) j\omega\mu} \mathbf{a}_x \tag{2.108}$$

by substituting for T from (2.107b) and replacing γ by $j\omega\mu/Z_m$ from (2.105). As $\sigma \to \infty$, the limiting value of $\mathbf{J}_s$ becomes

$$\mathbf{J}_s = \frac{2E_1}{Z_0} \mathbf{a}_x = 2Y_0 E_1 \mathbf{a}_x \tag{2.109}$$

since $Z_m \to 0$ and $\sigma Z_m^2 \to j\omega\mu$. This current exists only on the surface of the conductor since, as $\sigma \to \infty$, the skin depth $\delta_s \to 0$; that is, the field decays infinitely fast with distance into the conductor. When σ is infinite, $\Gamma = -1$ and the total tangential magnetic field at the surface is $2Y_0 E_1 \mathbf{a}_y$ and equal in magnitude to $\mathbf{J}_s$. In vector form the boundary conditions at the surface of a perfect conductor are thus seen to be

$$\mathbf{n} \times \mathbf{E} = 0 \tag{2.110a}$$

$$\mathbf{n} \times \mathbf{H} = \mathbf{J}_s \tag{2.110b}$$

where $\mathbf{n}$ is a unit outward normal at the conductor surface.

For finite conductivity the current density at the surface is $\sigma T E_1$ and the magnetic field at the surface is $Y_m T E_1$. In terms of these quantities the total current per unit width may be expressed as

$$\mathbf{J}_s = \frac{\sigma T E_1}{\gamma} \mathbf{a}_x = \frac{\sigma Z_m}{\gamma} (Y_m T E_1) \mathbf{a}_x = Y_m T E_1 \mathbf{a}_x$$

In other words, the total current per unit width is equal to the tangential magnetic field at the surface.

The time-average power transmitted into the conductor per unit area is given by the real part of one-half of the complex Poynting vector at the surface, and is

$$P_t = \tfrac{1}{2} \operatorname{Re} \mathbf{E} \times \mathbf{H}^* \cdot \mathbf{a}_z = \tfrac{1}{2} T T^* E_1 E_1^* \operatorname{Re} Y_m = \tfrac{1}{4} T T^* E_1 E_1^* \sigma \delta_s \tag{2.111}$$

The reader may readily verify that this is equal to the result obtained from a volume integral of $\mathbf{J} \cdot \mathbf{J}^*$; that is,

$$P_t = \frac{1}{2\sigma} \int_0^\infty \mathbf{J} \cdot \mathbf{J}^*\, dz$$

Equation (2.111) may be simplified with little error by making the following

approximation:

$$\sigma TT^* = \frac{4\sigma Z_m Z_m^*}{(Z_m + Z_0)(Z_m + Z_0)^*}$$
$$\approx \frac{4\sigma Z_m Z_m^*}{Z_0^2} = \frac{8}{\sigma \delta_s^2 Z_0^2}$$

whence (2.111) becomes

$$P_t \approx \frac{1}{2} \frac{(2Y_0 E_1)(2Y_0 E_1^*)}{\sigma \delta_s} \tag{2.112}$$

Note that $2Y_0 E_1$ is the value of the magnetic field, tangent to the surface, that would exist if σ were infinite. Hence an excellent approximate technique for evaluating power loss in a conductor is to find the tangential magnetic field, say H_t, that would exist for a perfect conductor, and then compute the power loss according to the relation

$$P_t = \tfrac{1}{2} \operatorname{Re}(H_t H_t^* Z_m) = \tfrac{1}{2} \operatorname{Re}(J_s J_s^* Z_m) \tag{2.113}$$

This procedure is equivalent to assuming that the metal exhibits a surface impedance Z_m and the current is essentially the same as that which would exist for infinite conductivity.

The procedure outlined above for power-loss calculations is widely used in microwave work. Although the derivation was based on a consideration of a very special boundary-value problem, the same conclusions result for more complex structures such as conducting spheres and cylinders. In general, the technique of characterizing the metal by a surface impedance Z_m and assuming that the current $\mathbf{J}_s$ is the same as that for infinite conductivity is valid as long as the conductor surface has a radius of curvature at least a few skin depths in magnitude.

2.10 POTENTIAL THEORY

The wave solutions presented in the previous sections have all been source-free solutions; i.e., the nature of the sources giving rise to the field was not considered. When it is necessary to consider the specific field generated by a given source, as in antenna problems, waveguide and cavity coupling, etc., this is greatly facilitated by introducing an auxiliary vector potential function $\mathbf{A}$. As will be seen, the vector potential $\mathbf{A}$ is determined by the current source, and the total electromagnetic field may be derived from $\mathbf{A}$.

Since $\nabla \cdot \mathbf{B} = 0$ always, this condition will hold identically if $\mathbf{B}$ is expressed as the curl of a vector potential $\mathbf{A}$ since the divergence of the curl of a vector is identically zero. Thus let

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{2.114}$$

The assumed time dependence $e^{j\omega t}$ is not written out explicitly in (2.114)

since this is a phasor representation. The curl equation for $\mathbf{E}$ gives

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B} = -j\omega\nabla \times \mathbf{A}$$

or

$$\nabla \times (\mathbf{E} + j\omega\mathbf{A}) = 0$$

Now the curl of the gradient of a scalar function Φ is identically zero; so the general integral of the above equation is

$$\mathbf{E} + j\omega\mathbf{A} = -\nabla\Phi$$

or

$$\mathbf{E} = -j\omega\mathbf{A} - \nabla\Phi \tag{2.115}$$

Substituting this expression into the $\nabla \times \mathbf{H}$ equation gives

$$\nabla \times \mathbf{H} = \frac{1}{\mu}\nabla \times \nabla \times \mathbf{A} = j\omega\epsilon\mathbf{E} + \mathbf{J} = \omega^2\epsilon\mathbf{A} - j\omega\epsilon\nabla\Phi + \mathbf{J} \tag{2.116}$$

Up to this point the divergences of $\mathbf{A}$ and $\nabla\Phi$ have not been specified [note that (2.114) specifies the curl of $\mathbf{A}$ only]. Therefore a relation between $\nabla \cdot \mathbf{A}$ and Φ may be chosen so as to simplify (2.116). Expanding $\nabla \times \nabla \times \mathbf{A}$ to give $\nabla\nabla \cdot \mathbf{A} - \nabla^2\mathbf{A}$ enables us to write (2.116) as

$$\nabla\nabla \cdot \mathbf{A} - \nabla^2\mathbf{A} = k^2\mathbf{A} - j\omega\epsilon\mu\nabla\Phi + \mu\mathbf{J}$$

where $k^2 = \omega^2\mu\epsilon$. If now the following condition is specified:

$$\nabla\nabla \cdot \mathbf{A} = -j\omega\epsilon\mu\nabla\Phi$$

or

$$\nabla \cdot \mathbf{A} = -j\omega\mu\epsilon\Phi \tag{2.117}$$

this equation simplifies to

$$\nabla^2\mathbf{A} + k^2\mathbf{A} = -\mu\mathbf{J} \tag{2.118}$$

Thus $\mathbf{A}$ is a solution of the inhomogeneous Helmholtz equation, the current $\mathbf{J}$ being the source term. The condition imposed on $\nabla \cdot \mathbf{A}$ and Φ in (2.117) is called the Lorentz condition in honor of the man first to propose its use.

In the preceding derivation three of Maxwell's equations have been used and are therefore satisfied. The fourth equation, $\nabla \cdot \mathbf{D} = \rho$, must also hold, and this will be shown to be the case provided the Lorentz condition is obeyed. Hence the three equations (2.114), (2.115), and (2.118), together with the Lorentz condition (2.117), are fully equivalent to Maxwell's equations. To verify the equation $\nabla \cdot \mathbf{D} = \rho$, take the divergence of (2.115) to obtain

$$\nabla \cdot \epsilon\mathbf{E} = -j\omega\epsilon\nabla \cdot \mathbf{A} - \epsilon\nabla^2\Phi \tag{2.119}$$

where ϵ is a constant. Using the Lorentz condition yields

$$\nabla \cdot \mathbf{D} = -j\omega\epsilon\nabla \cdot \mathbf{A} - \nabla^2\frac{\nabla \cdot \mathbf{A}}{-j\omega\mu} = \frac{1}{j\omega\mu}\nabla \cdot (\nabla^2\mathbf{A} + k^2\mathbf{A}) = -\frac{1}{j\omega}\nabla \cdot \mathbf{J}$$

by using (2.118) and noting that $\nabla^2\nabla \cdot \mathbf{A} = \nabla \cdot \nabla^2\mathbf{A}$; that is, these differential operators commute. Now $\nabla \cdot \mathbf{J} = -j\omega\rho$ from the continuity equation;

so we obtain

$$\nabla \cdot \mathbf{D} = -\frac{1}{j\omega}(-j\omega\rho) = \rho$$

If, instead of eliminating Φ in (2.119), $\nabla \cdot \mathbf{A}$ is eliminated by use of the Lorentz condition, we get

$$\nabla \cdot \mathbf{D} = \rho = -j\omega\epsilon(-j\omega\mu\epsilon\Phi) - \epsilon\nabla^2\Phi$$

or

$$\nabla^2\Phi + k^2\Phi = -\frac{\rho}{\epsilon} \tag{2.120}$$

Hence the scalar potential Φ is a solution of the inhomogeneous scalar Helmholtz equation, with the charge density ρ as a source term.

For the time-varying field, $\mathbf{J}$ and ρ are not independent, and hence the field can be determined in terms of $\mathbf{A}$ and $\mathbf{J}$ alone. The scalar potential can always be found from the Lorentz relation, and ρ from the continuity equation, but explicit knowledge of these is not required in order to solve radiation problems. For convenience, the pertinent equations are summarized here:

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{2.121a}$$

$$\mathbf{E} = -j\omega\mathbf{A} - \nabla\Phi = -j\omega\mathbf{A} + \frac{\nabla\nabla \cdot \mathbf{A}}{j\omega\mu\epsilon} = \frac{k^2\mathbf{A} + \nabla\nabla \cdot \mathbf{A}}{j\omega\mu\epsilon} \tag{2.121b}$$

$$\nabla^2\mathbf{A} + k^2\mathbf{A} = -\mu\mathbf{J} \tag{2.121c}$$

where the Lorentz condition was used to eliminate $\nabla\Phi$ in (2.121*b*). Note that, in rectangular coordinates, (2.121*c*) is three scalar equations of the form

$$\nabla^2 A_x + k^2 A_x = -\mu J_x$$

but that, in other coordinate systems, $\nabla^2\mathbf{A}$ must be expanded according to the relation $\nabla^2\mathbf{A} = \nabla\nabla \cdot \mathbf{A} - \nabla \times \nabla \times \mathbf{A}$.

The simplest solution to (2.121*c*) is that for an infinitesimal current element $\mathbf{J}(x', y', z') = \mathbf{J}(\mathbf{r}')$ located at the point x', y', z', as specified by the position vector $\mathbf{r}' = \mathbf{a}_x x' + \mathbf{a}_y y' + \mathbf{a}_z z'$, as in Fig. 2.17. This solution is

$$\mathbf{A}(x, y, z) = \mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi}\mathbf{J}(\mathbf{r}')\frac{e^{-jkR}}{R}\,dV' \tag{2.122}$$

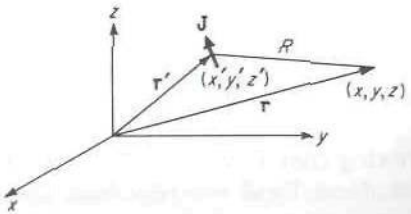


FIGURE 2.17
Coordinates used to describe vector potential from a current sheet.

where $R = |\mathbf{r} - \mathbf{r}'|$ is the magnitude of the distance from the source point to the field point at which $\mathbf{A}$ is evaluated; i.e.,

$$R = \left[(x - x')^2 + (y - y')^2 + (z - z')^2\right]^{1/2}$$

and $\mathbf{J}(\mathbf{r}')\,dV'$ is the total source strength. In terms of this fundamental solution, the vector potential from a general current distribution may be obtained by superposition. Thus, adding up all the contributions from each infinitesimal current element gives

$$\mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi}\int_V \mathbf{J}(x', y', z')\frac{e^{-jkR}}{R}\,dx'\,dy'\,dz' = \frac{\mu}{4\pi}\int_V \mathbf{J}(\mathbf{r}')\frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\,dV' \tag{2.123}$$

where the integration is over the total volume occupied by the current. Note that the solution for $\mathbf{A}$ as given by (2.122) is a spherical wave propagating radially outward from $\mathbf{J}$ and with an amplitude falling off as $1/R$. The solution (2.123) is a superposition of such elementary spherical waves.

*2.11 DERIVATION OF SOLUTION FOR VECTOR POTENTIAL

In this section a detailed derivation of the solution to the inhomogeneous Helmholtz equation for a unit current source is given. A unit source is a source of unit strength, localized at a point in space (a familiar example is a point charge). Such a unit source in a three-dimensional space is a generalization of a unit current impulse localized at a time t' along the time coordinate. A current pulse is represented by the Dirac delta function $\delta(t - t')$ in circuit theory, where $\delta(t - t')$ has the property

$$\delta(t - t') = 0 \qquad t \neq t' \tag{2.124a}$$

and at $t = t'$ it becomes infinite but is integrable to give

$$\int_{t'-\tau}^{t'+\tau} \delta(t - t')\,dt = 1 \tag{2.124b}$$

A further property is that, for any function $f(t)$ which is continuous at t',

$$\int_{t'-\tau}^{t'+\tau} f(t)\delta(t - t')\,dt = f(t') \tag{2.124c}$$

This result follows since τ can be chosen so small that, in the interval $t' - \tau < t < t' + \tau$, the function $f(t)$ differs by a vanishing amount from $f(t')$ since $f(t)$ is continuous at t'. Hence (2.124c) may be written as

$$f(t')\int_{t'-\tau}^{t'+\tau} \delta(t - t')\,dt = f(t')$$

by virtue of (2.124b).

As the preceding discussion has shown, the delta function is a convenient mathematical way to represent a source of unit strength localized at a point along a coordinate axis, in the above example along the time axis. In an N-dimensional space a product of N delta functions, one for each coordinate, may be used to represent a unit source. Thus, in three dimensions, a unit source is represented by

$$\delta(x - x')\delta(y - y')\delta(z - z') = \delta(\mathbf{r} - \mathbf{r}') \tag{2.125}$$

where $\delta(\mathbf{r} - \mathbf{r}')$ is an abbreviated notation for the product of the three one-dimensional delta functions. The source function $\delta(\mathbf{r} - \mathbf{r}')$ has the following properties:

$$\delta(\mathbf{r} - \mathbf{r}') = 0 \qquad \mathbf{r} \neq \mathbf{r}' \tag{2.126a}$$

$$\int_V \delta(\mathbf{r} - \mathbf{r}')\, dV = \begin{cases} 1 & \mathbf{r}' \text{ in } V \\ 0 & \mathbf{r}' \text{ not in } V \end{cases} \tag{2.126b}$$

$$\int_V \mathbf{F}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}')\, dV = \begin{cases} \mathbf{F}(\mathbf{r}') & \mathbf{r}' \text{ in } V \\ 0 & \mathbf{r}' \text{ not in } V \end{cases} \tag{2.126c}$$

where $\mathbf{F}$ is an arbitrary vector (or scalar) function that is continuous at $\mathbf{r}'$, that is, at x', y', z'. These properties follow from the properties of the one-dimensional delta functions that make up $\delta(\mathbf{r} - \mathbf{r}')$.

A unit current source directed along the unit vector $\mathbf{a}$ at $\mathbf{r}'$ may be expressed as $\mathbf{J} = \mathbf{a}\delta(\mathbf{r} - \mathbf{r}')$. The vector potential is a solution of

$$\nabla^2\mathbf{A} + k^2\mathbf{A} = -\mu\mathbf{a}\delta(\mathbf{r} - \mathbf{r}') \tag{2.127}$$

Since the current is in the direction $\mathbf{a}$, the vector potential must also be in this direction, and hence $\mathbf{A} = A\mathbf{a}$. Equation (2.127) may therefore be written as a scalar equation:

$$\nabla^2 A + k^2 A = -\mu\delta(\mathbf{r} - \mathbf{r}') \tag{2.128}$$

At all points $\mathbf{r} \neq \mathbf{r}'$, A is a solution of

$$\nabla^2 A + k^2 A = 0 \tag{2.129}$$

If the source point $\mathbf{r}'$ is considered as the origin in a spherical coordinate system, then, since no angle variables occur in the source term in (2.128), the solution for A must have spherical symmetry about the source point $\mathbf{r}'$. Thus, in terms of the spherical radial coordinate $R = |\mathbf{r} - \mathbf{r}'|$, which is the radial distance from the origin at $\mathbf{r}'$, (2.129) is a function of R only and may be written as

$$\frac{1}{R^2}\frac{\partial}{\partial R}\left(R^2\frac{\partial A}{\partial R}\right) + k^2 A = 0$$

or

$$\frac{d^2A}{dR^2} + \frac{2}{R}\frac{dA}{dR} + k^2 A = 0 \tag{2.130}$$

after expressing the R-dependent part of ∇^2 in spherical coordinates. In

anticipation of a spherical-wave solution, let $A = f(R)e^{-jkR}$. Substitution in (2.130) leads to the following equation for $f(R)$:

$$\frac{d^2f}{dR} + \left(\frac{2}{R} - 2jk\right)\frac{df}{dR} - \frac{2jk}{R}f = 0$$

which is readily verified to have the solution $f = C/R$, where C is an arbitrary constant. Consequently, the solution to (2.129) is $A = Ce^{-jkR}/R$. This solution is singular at $R = 0$, and the singularity must correspond to that of the source term at this point.

To determine the constant C, integrate (2.128) throughout a small sphere of radius r_0 centered on $\mathbf{r}'$ and use the delta-function property (2.126*b*) to obtain

$$\int_0^{2\pi}\int_0^{\pi}\int_0^{r_0}(\nabla^2 A + k^2 A)\,R^2 \sin\theta\, d\theta\, d\phi\, dR$$

$$= \int_V (\nabla^2 A + k^2 A)\, dV = -\mu \int_V \delta(\mathbf{r} - \mathbf{r}')\, dV = -\mu$$

Now the integral of the term k^2R^2A, which is proportional to R^2, will vanish as r_0 tends to zero. Hence, for sufficiently small r_0,

$$\int_V \nabla^2 A\, dV = -\mu$$

Since $\nabla^2 A = \nabla \cdot \nabla \mathbf{A}$, the divergence theorem may be used to give

$$\int_V \nabla^2 A\, dV = \oint_S \nabla A \cdot d\mathbf{S} = \oint_S \nabla A \cdot \mathbf{a}_r r_0^2\, d\Omega$$

since $d\mathbf{S} = \mathbf{a}_r r_0^2\, d\Omega$, where $d\Omega$ is an element of solid angle. Since A is a function of R only, $\nabla A = \mathbf{a}_r(\partial A/\partial R)$, and hence

$$r_0^2 \oint_S \nabla A \cdot \mathbf{a}_r\, d\Omega = r_0^2 \oint_S \frac{\partial A}{\partial R}\, d\Omega = 4\pi r_0^2 \frac{\partial A}{\partial R} = -\mu$$

Evaluating $\partial A/\partial R$ for $R = r_0$ shows that

$$4\pi r_0^2 \frac{\partial A}{\partial R} = -4\pi C r_0^2 \left(\frac{jk}{r_0} e^{-jkr_0} + \frac{e^{-jkr_0}}{r_0^2}\right) = -4\pi C$$

in the limit as r_0 tends to zero. Hence $4\pi C = \mu$, or $C = \mu/4\pi$, in order for the singularity in the solution for A to correspond to that for a unit source.

The above solution for the vector potential from a unit source, namely,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi} \frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|}\mathbf{a} \tag{2.131}$$

is clearly a function of both the source point and field point. Since (2.131) is the solution for a unit source, it is often called a Green's function and

denoted by the symbol $\mathbf{G}$ as

$$\mathbf{G}(\mathbf{r}|\mathbf{r}') = G(\mathbf{r}|\mathbf{r}')\mathbf{a} = G(x, y, z|x', y', z')\mathbf{a} = \frac{\mu}{4\pi}\frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\mathbf{a} \quad (2.132)$$

because, by definition, a Green's function is the solution of a differential equation for a unit source.

The vector potential from a general current distribution may now be expressed in the form

$$\mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi}\int_V \mathbf{J}(\mathbf{r}')\frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\,dV' = \int_V \mathbf{J}(\mathbf{r}')G(\mathbf{r}|\mathbf{r}')\,dV' \quad (2.133)$$

since any current distribution $\mathbf{J}$ may be considered as a sum of weighted unit sources.

2.12 LORENTZ RECIPROCITY THEOREM

The Lorentz reciprocity theorem is one of the most useful theorems in the solution of electromagnetic problems, since it may be used to deduce a number of fundamental properties of practical devices. It provides the basis for demonstrating the reciprocal properties of microwave circuits and for showing that the receiving and transmitting characteristics of antennas are the same. It also may be used to establish the orthogonality properties of the modes that may exist in waveguides and cavities.† Another important use is in deriving suitable field expansions (analogous to a Fourier series expansion) for the fields radiated or coupled into waveguides and cavities by probes, loops, or coupling apertures.

To derive the theorem, consider a volume V bounded by a closed surface S as in Fig. 2.18. Let a current source $\mathbf{J}_1$ in V produce a field $\mathbf{E}_1, \mathbf{H}_1$, while a second source $\mathbf{J}_2$ produces a field $\mathbf{E}_2, \mathbf{H}_2$. These fields satisfy Maxwell's equations; so

$$\nabla \times \mathbf{E}_1 = -j\omega\mu\mathbf{H}_1 \qquad \nabla \times \mathbf{H}_1 = j\omega\epsilon\mathbf{E}_1 + \mathbf{J}_1$$
$$\nabla \times \mathbf{E}_2 = -j\omega\mu\mathbf{H}_2 \qquad \nabla \times \mathbf{H}_2 = j\omega\epsilon\mathbf{E}_2 + \mathbf{J}_2$$

Expanding the relation $\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1)$ and using Maxwell's equation show that

$$\begin{aligned}\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) &= (\nabla \times \mathbf{E}_1) \cdot \mathbf{H}_2 - (\nabla \times \mathbf{H}_2) \cdot \mathbf{E}_1 - (\nabla \times \mathbf{E}_2) \cdot \mathbf{H}_1 + (\nabla \times \mathbf{H}_1) \cdot \mathbf{E}_2 \\ &= -\mathbf{J}_2 \cdot \mathbf{E}_1 + \mathbf{J}_1 \cdot \mathbf{E}_2 \end{aligned} \quad (2.134)$$

†In any waveguide or cavity an infinite number of field solutions are possible. Any one solution is called a mode for the same reason that the various solutions for vibrating strings and membranes are called modes. Orthogonality of modes is discussed in Sec. 3.14.

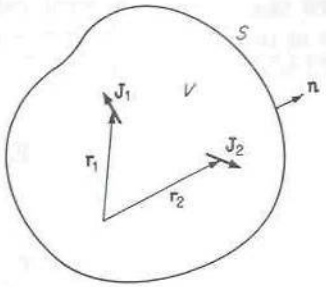


FIGURE 2.18
Illustration of the Lorentz reciprocity theorem.

Integrating both sides over the volume V and using the divergence theorem give

$$\int_V \nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1)\, dV = \oint_S (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot \mathbf{n}\, dS$$

$$= \int_V (\mathbf{E}_2 \cdot \mathbf{J}_1 - \mathbf{E}_1 \cdot \mathbf{J}_2)\, dV \qquad (2.135)$$

where $\mathbf{n}$ is the unit outward normal to S.

Equation (2.135) is the basic form of the Lorentz reciprocity theorem.† For a number of typical situations that occur, the surface integral vanishes. If S is a perfectly conducting surface, then $\mathbf{n} \times \mathbf{E}_1 = \mathbf{n} \times \mathbf{E}_2 = 0$ on S. Since $\mathbf{E}_1 \times \mathbf{H}_2 \cdot \mathbf{n} = (\mathbf{n} \times \mathbf{E}_1) \cdot \mathbf{H}_2$, etc., the surface integral vanishes in this case. If the surface S is characterized by a surface impedance Z_m, then, according to (2.71),

$$\mathbf{E}_t = -Z_m \mathbf{n} \times \mathbf{H} \qquad \text{or} \qquad \mathbf{n} \times \mathbf{E} = -Z_m \mathbf{n} \times (\mathbf{n} \times \mathbf{H})$$

[note that in (2.71) $\mathbf{n}$ points into the region occupied by the field, and hence the minus sign is used here, since $\mathbf{n}$ is directed out of V]. Consequently,

$$(\mathbf{n} \times \mathbf{E}_1) \cdot \mathbf{H}_2 - (\mathbf{n} \times \mathbf{E}_2) \cdot \mathbf{H}_1$$

$$= -Z_m[\mathbf{n} \times (\mathbf{n} \times \mathbf{H}_1)] \cdot \mathbf{H}_2 + Z_m[\mathbf{n} \times (\mathbf{n} \times \mathbf{H}_2)] \cdot \mathbf{H}_1$$

$$= Z_m(\mathbf{n} \times \mathbf{H}_2) \cdot (\mathbf{n} \times \mathbf{H}_1) - Z_m(\mathbf{n} \times \mathbf{H}_1) \cdot (\mathbf{n} \times \mathbf{H}_2) = 0$$

and the surface integral vanishes again.

†In anisotropic media with nonsymmetric permittivity or permeability tensors, a modified form must be used. See, for example, R. F. Harrington and A. T. Villeneuve, Reciprocity Relations for Gyrotropic Media, *IRE Trans.*, vol. MTT-6, pp. 308–310, July, 1958.

Another example where the surface integral vanishes is when S is chosen as a spherical surface at infinity for which $\mathbf{n} = \mathbf{a}_r$. The radiated field at infinity is a spherical TEM wave for which

$$\mathbf{H} = Y\mathbf{a}_r \times \mathbf{E} = \left(\frac{\epsilon}{\mu}\right)^{1/2} \mathbf{a}_r \times \mathbf{E}$$

Therefore

$$(\mathbf{n} \times \mathbf{E}_1) \cdot \mathbf{H}_2 - (\mathbf{n} \times \mathbf{E}_2) \cdot \mathbf{H}_1 = Y(\mathbf{a}_r \times \mathbf{E}_1) \cdot (\mathbf{a}_r \times \mathbf{E}_2)$$
$$-Y(\mathbf{a}_r \times \mathbf{E}_2) \cdot (\mathbf{a}_r \times \mathbf{E}_1) = 0$$

and the surface integral vanishes.

Actually, for any surface S which encloses all the sources for the field, the surface integral will vanish. This result may be seen by applying (2.135) to the volume V bounded by S and the surface of a sphere of infinite radius. There are no sources in this volume, and since the surface integral over the surface of the sphere with infinite radius is zero, we must have, from (2.135),

$$\oint_S (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot (-\mathbf{n})\, dS = 0$$
$$= \oint_S (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot \mathbf{n}\, dS$$

Hence the surface integral taken over any closed surface S surrounding all the sources vanishes.

When the surface integral vanishes, (2.135) reduces to

$$\int_V \mathbf{E}_1 \cdot \mathbf{J}_2\, dV = \int_V \mathbf{E}_2 \cdot \mathbf{J}_1\, dV \tag{2.136}$$

If $\mathbf{J}_1$ and $\mathbf{J}_2$ are infinitesimal current elements, then

$$\mathbf{E}_1(\mathbf{r}_2) \cdot \mathbf{J}_2(\mathbf{r}_2) = \mathbf{E}_2(\mathbf{r}_1) \cdot \mathbf{J}_1(\mathbf{r}_1) \tag{2.137}$$

which states that the field $\mathbf{E}_1$ produced by $\mathbf{J}_1$ has a component along $\mathbf{J}_2$ that is equal to the component along $\mathbf{J}_1$ of the field radiated by $\mathbf{J}_2$ when $\mathbf{J}_1$ and $\mathbf{J}_2$ have unit magnitude. The form (2.137) is essentially the reciprocity principle used in circuit analysis except that $\mathbf{E}$ and $\mathbf{J}$ are replaced by the voltage V and current I. The applications of the reciprocity theorem are illustrated at various points throughout the text and hence are not discussed further at this time.

PROBLEMS

2.1. An atom of atomic number Z has a nuclear charge Ze and Z electrons revolving around it. As a model of this atom, consider the nucleus as a point charge and treat the electron cloud as a total charge $-Ze$ distributed uniformly throughout a sphere of radius r_0. When an external field E is applied, the nucleus is displaced an amount x. Show that a restoring force $x(Ze)^2/4\pi r_0^3\epsilon_0$ is produced and must be equal to ZeE. Thus show that the induced dipole moment is $p = 4\pi\epsilon_0 r_0^3 E$ and is linearly related to E.

2.2. In a certain material the equation of motion for the polarization is

$$\frac{d^2\mathscr{P}}{dt^2} + \nu\frac{d\mathscr{P}}{dt} + \omega_0^2\mathscr{P} = 2\epsilon_0\omega_0^2\mathscr{E}$$

where $\mathscr{E}$ is the total field in the dielectric. Find the relation between $\mathscr{P}$ and $\mathscr{E}$ when $\mathscr{E} = \mathrm{Re}(Ee^{j\omega t})$ and E is real. If $\omega_0 = 10^{11}$ and $\nu = 10^{10}$, over what frequency range can a relationship such as $\mathscr{D} = \epsilon\mathscr{E} = \epsilon_0\mathscr{E} + \mathscr{P}$ be written if it is assumed that the criterion to be used is that the phase difference between $\mathscr{D}$ and $\mathscr{E}$ should not exceed 5°? Plot the magnitude and phase angle of the dielectric constant $\epsilon_r = \epsilon/\epsilon_0 = (\epsilon' - j\epsilon'')/\epsilon_0$ as a function of ω.

2.3. A dielectric material is characterized by a matrix (tensor) permittivity

$$\begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{xy} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{xz} & \epsilon_{yz} & \epsilon_{zz} \end{bmatrix} = \frac{\epsilon_0}{4}\begin{bmatrix} 7 & 3 & -2\sqrt{0.5} \\ 3 & 7 & -2\sqrt{0.5} \\ -2\sqrt{0.5} & -2\sqrt{0.5} & 10 \end{bmatrix}$$

when referred to the xyz coordinate frame. If the coordinate axis is rotated into the principal axis u, v, w, the permittivity is exhibited in diagonal form:

$$[\epsilon] = \begin{bmatrix} \epsilon_{uu} & 0 & 0 \\ 0 & \epsilon_{vv} & 0 \\ 0 & 0 & \epsilon_{ww} \end{bmatrix}$$

Find the principal axis and the values of the principal dielectric constants ϵ_{uu}/ϵ_0, etc.

Hint: By definition, along a principal axis a scalar equation such as $D_u = \epsilon_{uu}E_u$ holds. In general, if $\mathbf{D}$ is directed along a principal axis, then

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \frac{\epsilon_0}{4}\begin{bmatrix} 7 & 3 & -2\sqrt{0.5} \\ 3 & 7 & -2\sqrt{0.5} \\ -2\sqrt{0.5} & -2\sqrt{0.5} & 10 \end{bmatrix}\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \lambda\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}$$

or in words, when $\mathbf{D}$ is directed along a principal axis, it is related to $\mathbf{E}$ by a scalar constant λ. The above constitutes a set of three homogeneous equations, of which the first is

$$\left(\frac{7\epsilon_0}{4} - \lambda\right)E_x + \frac{3\epsilon_0}{4}E_y - \frac{2\epsilon_0\sqrt{0.5}}{4}E_z = 0$$

Verify that, for a solution, the following determinant must vanish:

$$\begin{vmatrix} 7 - 4\lambda/\epsilon_0 & 3 & -2\sqrt{0.5} \\ 3 & 7 - 4\lambda/\epsilon_0 & -2\sqrt{0.5} \\ -2\sqrt{0.5} & -2\sqrt{0.5} & 10 - 4\lambda/\epsilon_0 \end{vmatrix} = 0$$

This cubic equation gives three roots for λ, which may be identified as $\epsilon_{uu}, \epsilon_{vv}, \epsilon_{ww}$. For any one root, say ϵ_{uu}, the components of a vector directed along the corresponding principal axis are proportional to the cofactors of the above determinant. This type of problem is called a matrix eigenvalue problem. The λ's are the eigenvalues.

Answer: $\epsilon_{uu} = 3\epsilon_0$, $\epsilon_{vv} = 2\epsilon_0$, $\epsilon_{ww} = \epsilon_0$.

Unit vectors along the principal axis are

$$\mathbf{a}_u = 0.5\mathbf{a}_x + 0.5\mathbf{a}_y - \sqrt{0.5}\,\mathbf{a}_z$$

$$\mathbf{a}_v = 0.5\mathbf{a}_x + 0.5\mathbf{a}_y + \sqrt{0.5}\,\mathbf{a}_z$$

$$\mathbf{a}_w = \sqrt{0.5}\,\mathbf{a}_x - \sqrt{0.5}\,\mathbf{a}_y$$

2.4. In the interior of a medium with conductivity σ and permittivity ϵ, free charge is distributed with a density $\rho_0(x, y, z)$ at time $t = 0$. Show that the charge decays according to

$$\rho = \rho_0 e^{-t/\tau} \qquad \tau = \frac{\epsilon}{\sigma}$$

Evaluate the relaxation time τ for copper for which $\sigma = 5.8 \times 10^7$ S/m, $\epsilon = \epsilon_0$. Find τ for sea water also for which $\sigma = 4$ S/m and $\epsilon = 80\epsilon_0$. If the relaxation time is short compared with the period of an applied time-harmonic field, there will be negligible accumulation of free charge and $\nabla \cdot \mathbf{D}$ may be assumed to be zero. What is the upper frequency limit for which this is true in the case of copper and sea water, i.e., the frequency for which τ is equal to the period?

Hint: Use the continuity equation, Ohm's law, and the divergence equation for $\mathbf{D}$.

2.5. Show that, when the relaxation time for a material is small compared with the period of the time-harmonic field, the displacement current may be neglected in comparison with the conduction current.

2.6. Consider two concentric spheres of radii a and b. The outer sphere is kept at a potential V, and the inner sphere at zero potential. Solve Laplace's equation in spherical coordinates to find the potential and electric field between the spheres. Take $b > a$.

2.7. Solve Laplace's equation to find the potential and electric field between two coaxial cylinders of radii a and b if the center cylinder is kept at a potential V and the outer cylinder at zero potential. Take $b > a$.

2.8. Derive (2.45) from (2.18).

2.9. Derive (2.47).

2.10. Express the scalar Helmholtz equation $\nabla^2\psi + k^2\psi = 0$ in cylindrical coordinates. If $\psi = f(\phi)g(r)h(z)$, find the differential equations satisfied by f, g, and h.

2.11. When material polarization $\mathscr{P}$ and $\mathscr{M}$ are explicitly taken into account, show that the wave equations satisfied by $\mathscr{E}$ and $\mathscr{H}$ are

$$\nabla^2\mathscr{H} - \mu_0\epsilon_0\frac{\partial^2\mathscr{H}}{\partial t^2} = -\nabla\nabla\cdot\mathscr{M} + \mu_0\epsilon_0\frac{\partial^2\mathscr{M}}{\partial t^2} - \frac{\partial}{\partial t}\nabla\times\mathscr{P} - \nabla\times\mathscr{J}$$

$$\nabla^2\mathscr{E} - \mu_0\epsilon_0\frac{\partial^2\mathscr{E}}{\partial t^2} = \mu_0\frac{\partial^2\mathscr{P}}{\partial t^2} + \mu_0\frac{\partial\mathscr{J}}{\partial t} + \mu_0\frac{\partial}{\partial t}\nabla\times\mathscr{M} + \frac{1}{\epsilon_0}\nabla\rho - \frac{\nabla\nabla\cdot\mathscr{P}}{\epsilon_0}$$

Note that $\nabla\cdot\mathscr{B} = 0$; so $\nabla\cdot\mathscr{H} = -\nabla\cdot\mathscr{M}$ and $\nabla\cdot\mathscr{D} = \rho$; so $\nabla\cdot\epsilon_0\mathscr{E} = \rho - \nabla\cdot\mathscr{P}$. Examination of the source terms in the above equations shows that $\partial\mathscr{P}/\partial t$ is a polarization current analogous to conduction current $\mathscr{J}$.

2.12. Derive (2.62).

2.13. Between two perfectly conducting coaxial cylinders of radii a and b, $b > a$, the electromagnetic field is given by

$$\mathbf{E} = \mathbf{a}_r E_0 r^{-1} e^{-jk_0 z} \qquad H = \mathbf{a}_\phi Y_0 E_0 r^{-1} e^{-jk_0 z}$$

where $k_0 = \omega(\mu_0\epsilon_0)^{1/2}$, $Y_0 = (\epsilon_0/\mu_0)^{1/2}$. Find the potential difference between the cylinders and the total current on the inner and outer cylinders. Express the power in terms of the voltage and current, and show that it is equal to that computed from an integration of the complex Poynting vector over the coaxial-line cross section. Show that the characteristic impedance of the line is $V/I = (Z_0/2\pi)\ln(b/a) = 60\ln(b/a)$, where V is the voltage and I is the total current on one cylinder.

2.14. A round wire of radius r_0 much greater than the skin depth δ_s has a uniform electric field E applied in the axial direction at its surface. Use the surface-impedance concept to find the total current on the wire. Show that the ratio of the ac impedance of the wire to the dc resistance is

$$\frac{Z_{ac}}{R_{dc}} = \frac{r_0\sigma}{2}Z_m$$

Evaluate this ratio for copper at $f = 10^6$ Hz for $\sigma = 5.8 \times 10^7$ S/m, $r_0 = 0.1$ cm, $\mu = \mu_0$.

2.15. The half-space $z \geq 0$ is filled with a material with permittivity ϵ_0 and permeability $\mu \neq \mu_0$. A parallel polarized plane TEM wave is incident at an angle θ_1, as in Fig. 2.14. Find the reflection and transmission coefficients for the electric field. Does a Brewster angle exist for which the reflection coefficient vanishes?

2.16. Repeat Prob. 2.15 for the case of a perpendicular polarized incident wave. Does a Brewster angle exist? If so, obtain an expression for it.

2.17. The half-space $z \geq 0$ is filled with a material with permeability μ and permittivity ϵ. When a plane wave is incident normally on this material, show that the reflection and transmission coefficients are

$$\Gamma = \frac{Z - Z_0}{Z + Z_0} \qquad T = 1 + \Gamma = \frac{2Z}{Z + Z_0}$$

where $Z = (\mu/\epsilon)^{1/2}$, $Z_0 = (\mu_0/\epsilon_0)^{1/2}$. Choose an electric field with an x component only.

2.18. The half-space $z \geq 0$ is filled with a material of permittivity ϵ_2 and with $\mu = \mu_0$. A second sheet with permittivity ϵ_1 is placed in front. A plane wave is incident normally on the structure from the left, as illustrated. Verify that the reflection coefficient at the first interface vanishes if $\epsilon_1 = (\epsilon_2\epsilon_0)^{1/2}$ and the thickness $d = \frac{1}{4}\lambda_0(\epsilon_0/\epsilon_1)^{1/2}$. The electric field may be assumed to have an x component only. The matching layer is known as a quarter-wave transformer (actually an impedance transformer). This matching technique is used to reduce reflections from optical lenses and is called *lens blooming*, or coated lenses.

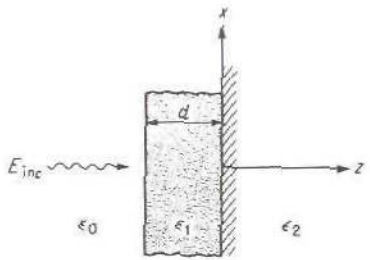


FIGURE P2.18

2.19. In terms of the vector potential $\mathbf{A}$ from a short current element $\Delta z I_0 \mathbf{a}_z$ located at the origin, show that the radiated electric and magnetic fields are

$$\mathbf{H} = \frac{I_0 \Delta z}{4\pi}\left(\frac{jk_0}{r} + \frac{1}{r^2}\right)\mathbf{a}_\phi \sin\theta\, e^{-jk_0 r}$$

$$\mathbf{E} = -\frac{I_0 \Delta z}{2\pi}\frac{jZ_0}{k_0}\left(\frac{jk_0}{r^2} + \frac{1}{r^3}\right)\mathbf{a}_r \cos\theta\, e^{-jk_0 r}$$

$$-\frac{I_0 \Delta z}{4\pi}\frac{jZ_0}{k_0}\left(\frac{-k_0^2}{r} + \frac{jk_0}{r^2} + \frac{1}{r^3}\right)\mathbf{a}_\theta \sin\theta\, e^{-jk_0 r}$$

Hint: Use (2.122) and (2.121), and express $\mathbf{A}$ as components in a spherical coordinate system r, θ, ϕ. Note that $\mathbf{a}_z = \mathbf{a}_r \cos\theta - \mathbf{a}_\theta \sin\theta$.

2.20. A dielectric may be characterized by its dipole polarization $\mathbf{P}$ per unit volume. If $\rho = \mathbf{J} = 0$ and $\mathbf{P}$ is taken into account explicitly, show that, if a vector potential $\mathbf{A}$ is introduced according to $\mathbf{B} = \nabla \times \mathbf{A}$, then $\mathbf{A}$ is a solution of

$$\nabla^2\mathbf{A} + k_0^2\mathbf{A} = -j\omega\mu_0\mathbf{P}$$

and that the fields are given by

$$\mathbf{B} = \nabla \times \mathbf{A} \qquad \mathbf{E} = \frac{\nabla\nabla \cdot \mathbf{A} + k_0^2\mathbf{A}}{j\omega\mu_0\epsilon_0}$$

Note that a Lorentz condition is used. Thus an electric dipole $\mathbf{P}$ is equivalent to a current element $j\omega\mathbf{P}$, or alternatively, a current element $\mathbf{J}$ may be considered as an electric dipole $\mathbf{P} = \mathbf{J}/j\omega$.

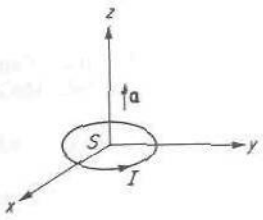


FIGURE P2.21

2.21. A small current loop constitutes a magnetic dipole $\mathbf{M} = IS\mathbf{a}$, where I is the current, S the area of the loop, and $\mathbf{a}$ a vector normal to the plane of the loop and pointing in the direction that a right-hand screw, rotating in the direction of the current, would advance. The field radiated by such a current loop, with linear dimensions much smaller than a wavelength, may be obtained by a potential theory analogous to that given in Prob. 2.20 by treating the loop as a *magnetic dipole* $\mathbf{M}$. *Thus replace* $\mathbf{B}$ *by* $\mu_0\mathbf{H} + \mu_0\mathbf{M}$ *in Maxwell's equation and* treat $\mathbf{M}$ as a source term. Since ρ is zero, $\nabla \cdot \mathbf{D} = 0$, and this permits $\mathbf{D}$ to be expressed as $\mathbf{D} = -\nabla \times \mathbf{A}_m$, where $\mathbf{A}_m$ is a magnetic-type vector potential. By paralleling the development in the text for the potential $\mathbf{A}$, show that the following relations are obtained:

$$\nabla^2\mathbf{A}_m + k_0^2\mathbf{A}_m = -j\omega\mu_0\epsilon_0\mathbf{M}$$

$$\mathbf{D} = -\nabla \times \mathbf{A}_m$$

$$\mathbf{H} = \frac{k_0^2 A_m + \nabla\nabla \cdot \mathbf{A}_m}{j\omega\mu_0\epsilon_0}$$

Hence, for a z-directed magnetic dipole at the origin,

$$\mathbf{A}_m = \frac{j\omega\mu_0\epsilon_0\mathbf{M}}{4\pi r}e^{-jk_0r}$$

from which the fields are readily found. Note that in this problem $\mathbf{M}$ represents the magnetic dipole source density in Maxwell's equations, but in the solution for the vector potential it represents the total magnetic dipole strength. It would have been more consistent to use $\mathbf{M}\,\delta(\mathbf{r} - \mathbf{r}')$ to represent the source density, where $\delta(\mathbf{r} - \mathbf{r}')$ is the three-dimensional Dirac delta function which has the property

$$\int_V \delta(\mathbf{r} - \mathbf{r}')\,dV' = 1 \qquad \mathbf{r} \text{ in } V$$

2.22. *Consider* an arbitrary current element $\mathbf{J}_1$ in front of a *perfectly conducting* plane. This current radiates a field $\mathbf{E}_1$ having zero tangential components on the *conducting plane. Use the Lorentz reciprocity theorem* to show that a current $\mathbf{J}_2$ parallel to the conducting plane and an infinitesimal distance in front of it does not radiate.

REFERENCES

1. Kraus, J. D.: "Electromagnetics," 3rd ed., McGraw-Hill Book Company, New York, 1984.
2. Hayt, W. H., Jr.: "Engineering Electromagnetics," 5th ed., McGraw-Hill Book Company, New York, 1989.
3. Schwarz, S. E.: "Electromagnetics for Engineers," Holt, Rinehart, and Winston, Inc., Philadelphia, 1990.
4. Johnk, C. T. A.: "Engineering Electromagnetic Fields and Waves," 2nd ed., John Wiley & Sons, Inc., New York, 1988.
5. Wait, J. R.: "Electromagnetic Wave Theory," Harper & Row Publishers, Inc., New York, 1985.
6. Stratton, J. A. "Electromagnetic Theory," McGraw-Hill Book Company, New York, 1941.
7. Shen, L. C., and J. A. Kong: "Applied Electromagnetism," Brooks-Cole, Calif., 1983.

CHAPTER 3

TRANSMISSION LINES AND WAVEGUIDES

This chapter is a long one and for this reason has been divided into three parts, namely:

Part 1—Waves on transmission lines
Part 2—Field analysis of transmission lines
Part 3—Rectangular and circular waveguides

The three parts are closely related but independent with the exception of Sec. 3.7, which is needed as an introduction to both Parts 2 and 3. With the exception of this section, the three parts can be studied independently and in any order.

In Part 1 we give an introduction to waves on transmission lines using a distributed-circuit model of the transmission line. By using the distributed-circuit model, we are able to study the excitation and propagation of current and voltage waves on a transmission line without the need to invoke Maxwell's equations.

The electrical characteristics of a transmission line such as the propagation constant, attenuation constant, characteristic impedance, and the distributed-circuit parameters can only be determined from a knowledge of the fields surrounding the transmission line. Thus in Part 2 we carry out a detailed field analysis of transmission lines. This part also includes an extensive discussion of planar transmission-line structures such as the microstrip line.

Part 3 presents the theory for waves in hollow rectangular and circular waveguides (pipes). In the beginning section of Part 2, we show that Maxwell's equations can be separated into equations that describe three types of waves. These are transverse electromagnetic waves (TEM), transverse electric (TE), and transverse magnetic (TM) waves. The TEM wave is the principal wave that can exist on a transmission line. The TE and TM waves are characterized by having no axial component of electric and magnetic field respectively. The TE and TM waves are the fundamental wave types that can exist in hollow-pipe waveguides. Hollow-pipe waveguides do not support TEM waves. The ability to reduce Maxwell's equations into three set of equations, one set for each wave type, facilitates the analysis of transmission lines and waveguides. Thus this decomposition of Maxwell's equations is carried out in the first section of Part 2.

PART 1
WAVES ON TRANSMISSION LINES

In this section we introduce the topic of voltage and current waves on a two-conductor transmission line by using a distributed-circuit model of the transmission line. This allows us to explore a number of fundamental properties of one-dimensional waves without having to consider the electromagnetic fields in detail. The distributed-circuit-model approach has limitations and in general must be replaced by a detailed solution for the electromagnetic field associated with the guiding structure if we want to determine the distributed-circuit parameters. The field analysis of transmission lines is presented in Part 2.

3.1 WAVES ON AN IDEAL TRANSMISSION LINE

In Fig. 3.1*a* we show a two-conductor transmission line consisting of two parallel round conductors (wires). The conductors will be assumed to be perfect, i.e., have infinite conductivity. The conductors extend from $z = 0$ to infinity, thus forming a semiinfinite transmission line. At $z = 0$ a voltage generator with internal resistance R_g is connected to the transmission line. The generator produces a voltage $\mathcal{V}_g(t)$ that is impressed across the transmission line. If the generator is switched on at time $t = 0$, a current $\mathcal{I}(t)$ will begin to flow into the upper conductor. A return current $-\mathcal{I}(t)$ must then flow on the lower conductor since current flow through the generator must be continuous. The return current is produced by the action of the electric field established between the two conductors. Since the transmission line is semiinfinite in length, there is no direct conducting path between the upper and lower conductors. However, there is a distributed capacitance C

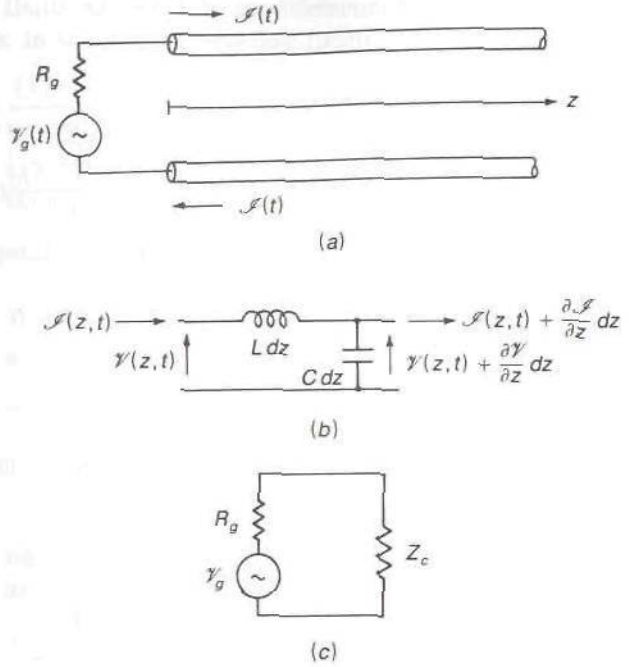


FIGURE 3.1
(*a*) An ideal two-conductor transmission line connected to a voltage generator; (*b*) equivalent circuit of a differential section of the transmission line with no loss; (*c*) equivalent circuit seen by the generator.

per meter between the two conductors; so we have a capacitive or displacement current flowing from the upper conductor to the lower conductor.

The electric current results in a magnetic field around the conductors and consequently the transmission line will also have a distributed series inductance L per meter. We can model a differential section dz of this transmission line as a series inductance $L\,dz$ and a shunt capacitance $C\,dz$ as shown in Fig. 3.1*b*. If the conductors had finite conductivity, we would also need to include a series resistance in the equivalent circuit of a differential section. However, we are assuming that the conductors are perfect; so the series distributed resistance R per meter is zero.

Since electrical effects propagate with a finite velocity v (the speed of light in vacuum), it should be clear that the voltage $\mathscr{V}(z,t)$ and current $\mathscr{I}(z,t)$ at some arbitrary point z on the transmission line will be zero until a time z/v has elapsed after switching the generator on. We will show that the generator launches voltage and current waves on the transmission line that propagate with a finite velocity. The equations that describe these waves are established by applying Kirchhoff's circuit laws to the equivalent circuit of a differential section of the transmission line, along with a specification of the terminal relationships (boundary conditions) that must hold at the generator end.

At some arbitrary point z on the transmission line, let the voltage and current be given by $\mathscr{V}(z,t)$, $\mathscr{I}(z,t)$. At a differential distance dz further

along, the voltage and current have changed by small amounts $(\partial\mathscr{V}/\partial z)\,dz$ and $(\partial\mathscr{I}/\partial z)\,dz$; so the output voltage and current at $z + dz$ will be

$$\mathscr{V}(z + dz, t) = \mathscr{V}(z, t) + \frac{\partial\mathscr{V}(z, t)}{\partial z}\,dz$$

$$\mathscr{I}(z + dz, t) = \mathscr{I}(z, t) + \frac{\partial\mathscr{I}(z, t)}{\partial z}\,dz$$

The sum of all potential drops around the circuit must be zero; so we have

$$-\mathscr{V} + L\,dz\frac{\partial\mathscr{I}}{\partial t} + \mathscr{V} + \frac{\partial\mathscr{V}}{\partial z}\,dz = 0$$

or

$$\frac{\partial\mathscr{V}(z, t)}{\partial z} = -L\frac{\partial\mathscr{I}(z, t)}{\partial t} \tag{3.1a}$$

The sum of currents flowing into the output node must also be zero; so we can write

$$\mathscr{I} - C\,dz\frac{\partial\mathscr{V}}{\partial t} - \mathscr{I} - \frac{\partial\mathscr{I}}{\partial z}\,dz = 0$$

or

$$\frac{\partial\mathscr{I}(z, t)}{\partial z} = -C\frac{\partial\mathscr{V}(z, t)}{\partial t} \tag{3.1b}$$

These two partial differential equations describe the relationship between the voltage and current waves on the transmission line.

We can obtain an equation for the voltage $\mathscr{V}(z, t)$ by differentiating (3.1a) with respect to z and using (3.1b) to eliminate the current; thus

$$\frac{\partial^2\mathscr{V}(z, t)}{\partial z^2} = -L\frac{\partial^2\mathscr{I}}{\partial z\,\partial t} = -L\left(-C\frac{\partial^2\mathscr{V}}{\partial t^2}\right)$$

or

$$\frac{\partial^2\mathscr{V}(z, t)}{\partial z^2} - LC\frac{\partial^2\mathscr{V}(z, t)}{\partial t^2} = 0 \tag{3.2a}$$

In a similar way we obtain

$$\frac{\partial^2\mathscr{I}(z, t)}{\partial z^2} - LC\frac{\partial^2\mathscr{I}(z, t)}{\partial t^2} = 0 \tag{3.2b}$$

The product LC has the dimensions of one over velocity squared. These two equations are one-dimensional wave equations and describe waves propagating with a velocity†

$$v = \frac{1}{\sqrt{LC}} \tag{3.3}$$

†For an ideal transmission line in air, $v = c = 3 \times 10^8$ m/s, the velocity of light.

Consider the equation

$$\frac{\partial^2 \mathscr{V}}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 \mathscr{V}}{\partial t^2} = 0$$

We can readily show that any two arbitrary functions of the form $f^+(t - z/v)$ and $f^-(t + z/v)$ are solutions of this equation. If we let $w = t - z/v$ then we have

$$\frac{\partial f^+(t - z/v)}{\partial z} = \frac{\partial f^+(w)}{\partial z} = \frac{\partial f^+(w)}{\partial w}\frac{\partial w}{\partial z} = -\frac{1}{v}\frac{\partial f^+(w)}{\partial w}$$

and

$$\frac{\partial^2 f^+(w)}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 f^+(w)}{\partial w^2}$$

For $\partial^2 f^+/\partial t^2$ we get $\partial^2 f^+/\partial w^2$. Consequently,

$$\frac{\partial^2 f^+}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 f^+}{\partial t^2} = \frac{\partial^2 f^+}{\partial w^2}\left(\frac{1}{v^2} - \frac{1}{v^2}\right) = 0$$

so $f^+(t - z/v)$ is clearly a solution of the one-dimensional wave equation, as is $f^-(t + z/v)$.

The function $f^+(t - z/v)$ is the same as the function $f^+(t)$ but delayed in time by an amount z/v which equals the distance z divided by the velocity of propagation v. We interpret this solution as a wave propagating in the positive z direction and identify this solution with a superscript $+$ sign. The other solution represents a wave propagating in the $-z$ direction and is identified by the $-$ sign as a superscript.

The general solution for the voltage waves on the transmission line is

$$\mathscr{V}(z,t) = V^+ f^+\left(t - \frac{z}{v}\right) + V^- f^-\left(t + \frac{z}{v}\right) \tag{3.4}$$

where V^+ and V^- are amplitude constants. By using (3.1b) we see that

$$\frac{\partial \mathscr{I}}{\partial z} = -C\left(V^+\frac{\partial f^+}{\partial t} + V^-\frac{\partial f^-}{\partial t}\right)$$

If we assume that $\mathscr{I}$ is of the form

$$\mathscr{I}(z) = I^+ f^+\left(t - \frac{z}{v}\right) - I^- f^-\left(t + \frac{z}{v}\right)$$

then

$$\frac{\partial \mathscr{I}}{\partial z} = -\frac{1}{v}\left(I^+\frac{\partial f^+}{\partial t} + I^-\frac{\partial f^-}{\partial t}\right)$$

upon using

$$\frac{\partial f^\pm}{\partial z} = \frac{\partial f^\pm}{\partial (t \mp z/v)}\frac{\partial (t \mp z/v)}{\partial z} = \mp\frac{1}{v}\frac{\partial f^\pm}{\partial t}$$

An examination of these equations shows that the assumed solution for

$\mathcal{I}(z,t)$ is compatible with that for the voltage $\mathcal{V}(z,t)$ if we choose

$$I^+ = vCV^+ \qquad I^- = vCV^-$$

The parameter vC has the dimensions of an admittance and is also equal to $C/\sqrt{LC} = \sqrt{C/L}$. The characteristic admittance Y_c of the transmission line is defined by this parameter. The reciprocal parameter is called the characteristic or surge impedance of the transmission line. It is given by

$$Z_c = \sqrt{\frac{L}{C}} = Y_c^{-1} \tag{3.5}$$

By using this parameter the solution for the current waves on the transmission line can be expressed in the form

$$\mathcal{I}(z,t) = \frac{V^+}{Z_c} f^+\left(t - \frac{z}{v}\right) - \frac{V^-}{Z_c} f^-\left(t + \frac{z}{v}\right) \tag{3.6}$$

The negative sign preceding the wave with amplitude V^- indicates a reversal in the direction of current flow for the wave propagating in the $-z$ direction.

For the transmission-line circuit shown in Fig. 3.1, the generator will launch voltage and current waves propagating in the $+z$ direction. Since the transmission line extends to infinity, no waves propagating in the $-z$ direction will be present. Later on we will consider a transmission line that is terminated at $z = l$ with either a resistance, capacitance, or a combination of these elements. Waves propagating in both the $+z$ and $-z$ directions will then exist. For the present case the voltage and current waves on the transmission line are assumed to be

$$\mathcal{V}(z,t) = V^+ f^+\left(t - \frac{z}{v}\right)$$

$$\mathcal{I}(z,t) = I^+ f^+\left(t - \frac{z}{v}\right)$$

with $V^+ = I^+ Z_c$. At the generator end $z = 0$ the terminal conditions require that

$$\mathcal{V}_g(t) = \mathcal{I}_g R_g + \mathcal{V}(0,t)$$

$$\mathcal{I}(0,t) = \mathcal{I}_g$$

where $\mathcal{I}_g$ is the current supplied by the generator. These terminal conditions can be expressed in the form

$$\mathcal{V}_g(t) = R_g \frac{V^+}{Z_c} f^+(t) + V^+ f^+(t)$$

$$\frac{V^+}{Z_c} f^+(t) = \mathcal{I}_g$$

from which we find that

$$V^+f^+(t) = \frac{Z_c}{Z_c + R_g}\mathscr{V}_g(t) \tag{3.7}$$

The voltage wave launched on the transmission line is thus given by

$$\mathscr{V}(z,t) = \frac{Z_c}{Z_c + R_g}\mathscr{V}_g\left(t - \frac{z}{v}\right) \tag{3.8a}$$

with a corresponding current wave

$$\mathscr{I}(z,t) = \frac{1}{Z_c + R_g}\mathscr{V}_g\left(t - \frac{z}{v}\right) \tag{3.8b}$$

At any point on the transmission line, the voltage waveform is the same as that produced by the generator but delayed in time and reduced in amplitude by the factor $Z_c/(R_g + Z_c)$. The voltage reduction is the usual voltage

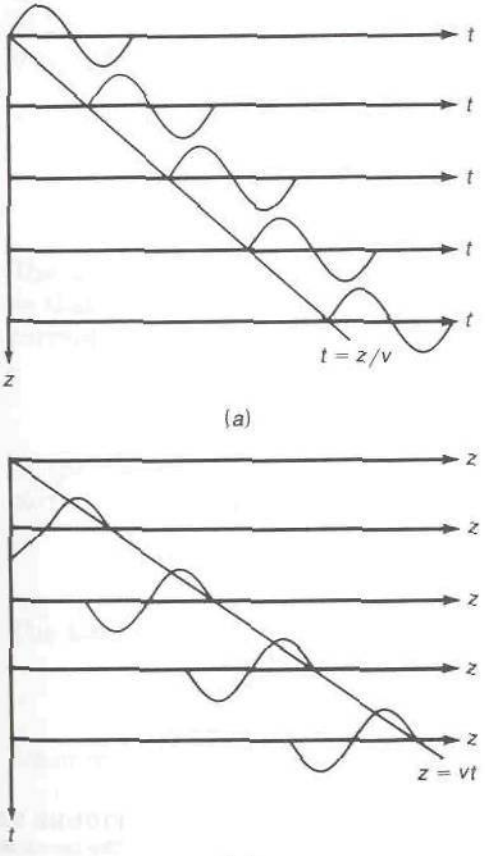


FIGURE 3.2
Time-distance and distance-time plots of voltage waveform $\mathscr{V}(z,t)$ on a transmission line for a single-cycle sinusoidal generator voltage pulse.

division factor associated with the equivalent circuit shown in Fig. 3.1*c*. For the infinite line the generator sees only an equivalent impedance Z_c equal to the transmission-line characteristic impedance.

In Fig. 3.2*a* we show a time-distance plot of the voltage waveform on a transmission line for the case when the generator produces a single cycle of a sinusoidal waveform, i.e.,

$$\mathscr{V}_g(t) = V_0 \sin t \qquad 0 \le t \le 2\pi$$

Figure 3.2*a* shows $\mathscr{V}(z,t)$ as a function of t at various distances z, while Fig. 3.2*b* shows $\mathscr{V}(z,t)$ as a function of z for various values of t. In the latter plot note that the leading edge of the waveform is the initial voltage produced by the generator at time $t = 0$ and hence the waveform appears reversed when plotted as a function of z.

3.2 TERMINATED TRANSMISSION LINE: RESISTIVE LOAD

In Fig. 3.3 we show a transmission line terminated at a distance l from the generator in a load resistance R_L. At the load end the terminal conditions are

$$\mathscr{V}(l,t) = \mathscr{V}_L = \mathscr{I}_L R_L \tag{3.9a}$$

$$\mathscr{I}(l,t) = \mathscr{I}_L \tag{3.9b}$$

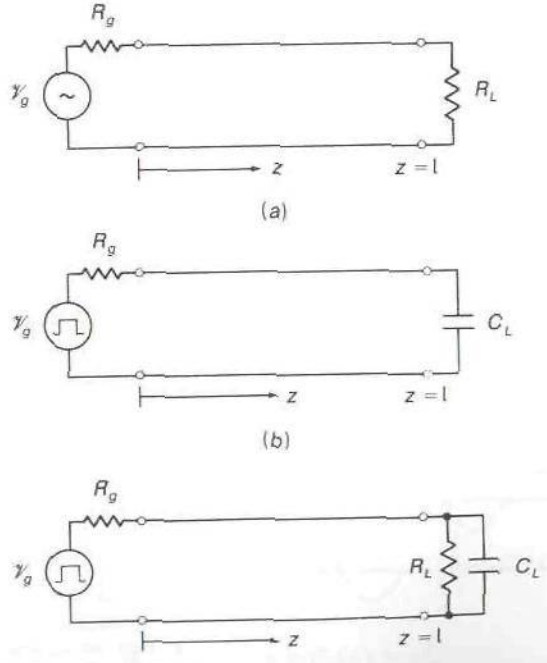


FIGURE 3.3
The terminated transmission line.

If we choose R_L equal to the characteristic impedance Z_c, then $\mathscr{V}_L = \mathscr{I}_L R_L = \mathscr{I}_L Z_c$. For a wave propagating in the $+z$ direction, $\mathscr{V}(z,t) = Z_c \mathscr{I}(z,t)$, so that at $z = l$, $\mathscr{V}(l,t) = Z_c \mathscr{I}(l,t)$, which satisfies the load terminal condition. Thus by choosing $R_L = Z_c$ the forward propagating wave will be completely absorbed by the load resistor and no reflected wave will be generated at the load end. Thus, in order to avoid a reflected wave, such as a reflected pulse in a digital circuit application, the transmission line should be terminated in its characteristic impedance.

When $R_L \neq Z_c$ the terminal conditions at the load end cannot be satisfied without introducing a reflected wave. The incident wave at $z = l$ is given by

$$\mathscr{V}_i(l,t) = \frac{Z_c}{Z_c + R_g}\mathscr{V}_g\left(t - \frac{l}{v}\right) = V^+\mathscr{V}_g\left(t - \frac{l}{v}\right)$$

$$\mathscr{I}_i(l,t) = \frac{1}{Z_c}\mathscr{V}_i(l,t)$$

where V^+ is the amplitude of $\mathscr{V}_i$ relative to $\mathscr{V}_g$. In order for a reflected wave to combine with the incident wave so as to satisfy the terminal conditions (3.9), the reflected wave must have the same time dependence as that of the incident wave. Hence the form of the reflected wave will be

$$\mathscr{V}_r(z,t) = V^-\mathscr{V}_g\left(t - \frac{l}{v} + \frac{(z-l)}{v}\right)$$

$$= V^-\mathscr{V}_g\left(t + \frac{z}{v} - \frac{2l}{v}\right)$$

The argument must contain the factor $t + z/v$ plus additional delay factors, so that at $z = l$ the reflected wave has the form $\mathscr{V}_g(t - l/v)$. The reflected current wave is given by

$$\mathscr{I}_r(z,t) = -\frac{1}{Z_c}\mathscr{V}_r(z,t)$$

At the load end the total current on the transmission line must equal the current $\mathscr{I}_L$ flowing through R_L; thus

$$\frac{1}{Z_c}(V^+ - V^-)\mathscr{V}_g\left(t - \frac{l}{v}\right) = \mathscr{I}_L$$

The total voltage on the transmission line must equal the load voltage; so

$$(V^+ + V^-)\mathscr{V}_g\left(t - \frac{l}{v}\right) = \mathrm{V}_L = \mathscr{I}_L R_L$$

When we divide this equation by the first one, we obtain

$$\frac{V^+ + V^-}{V^+ - V^-} = \frac{R_L}{Z_c}$$

which gives

$$\frac{V^-}{V^+} = \frac{R_L - Z_c}{R_L + Z_c} = \Gamma_L \tag{3.10}$$

The parameter Γ_L is called the load voltage reflection coefficient. The amplitude V^- is that of the reflected voltage wave and V^+ is the amplitude of the incident voltage wave. The ratio is determined by the *conditions at the load end only.*

Once a reflected wave has been launched from the load termination, the total voltage on the transmission line will consist of the incident voltage wave plus the reflected voltage wave until the time at which the reflected wave reaches the generator end. If the generator internal impedance R_g equals the characteristic impedance Z_c, then the reflected wave is absorbed at the generator end. If $R_g \neq Z_c$ then the reflected wave is reflected at the generator end to produce another forward propagating wave. For the reflected wave at the generator end, the terminal conditions are obtained by short-circuiting the voltage generator. Thus the reflected wave sees a termination R_g and will be reflected with a reflection coefficient Γ_g given by

$$\Gamma_g = \frac{R_g - Z_c}{R_g + Z_c} \tag{3.11}$$

As long as the generator continues to produce a voltage $\mathscr{V}_g(t)$, it continues to launch a first forward propagating wave with voltage $\mathscr{V}_i(0,t) = Z_c\mathscr{V}_g(t)/(R_g + Z_c)$ and with current $\mathscr{I}_i(0,t) = \mathscr{I}_g$. Thus the superposition of a reflected wave at the generator end requires the launching of a second forward propagating wave with a voltage amplitude that cancels that of the reflected wave at the generator terminals, i.e., the generator is treated as being short-circuited. The second forward propagating wave will also undergo reflection at the load termination, so that as time proceeds we will end up with a multitude of forward propagating and reflected waves on the transmission line. This collection of waves can be described as follows:

$$\begin{aligned}\mathscr{V}(z,t) = {} & V^+\mathscr{V}_g\left(t - \frac{z}{v}\right) + \Gamma_L V^+\mathscr{V}_g\left(t + \frac{z-2l}{v}\right)U\left(t - \frac{l}{v}\right) \\ & + \Gamma_g\Gamma_L V^+\mathscr{V}_g\left(t - \frac{z+2l}{v}\right)U\left(t - \frac{2l}{v}\right) \\ & + \Gamma_g\Gamma_L^2 V^+\mathscr{V}_g\left(t + \frac{z-4l}{v}\right)U\left(t - \frac{3l}{v}\right) \\ & + \Gamma_g^2\Gamma_L^2 V^+\mathscr{V}_g\left(t - \frac{z+4l}{v}\right)U\left(t - \frac{4l}{v}\right) + \cdots\end{aligned} \tag{3.12}$$

where $V^+ = Z_c/(R_g + Z_c)$ and $U(t-\alpha)$ is the unit step function which equals zero for $t < \alpha$ and equals unity for $t \geq \alpha$. The unit step function is a

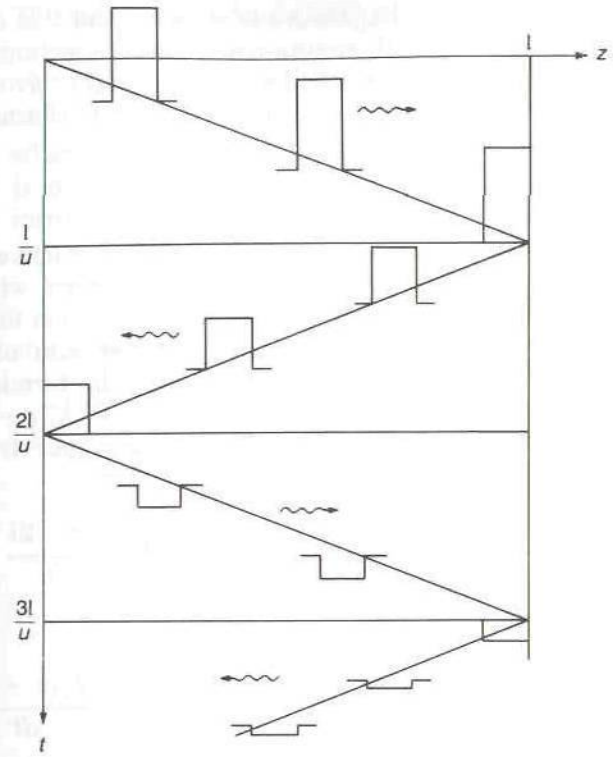


FIGURE 3.4
Distance-time plot of a pulse undergoing multiple reflection on a transmission line when $\Gamma_L = -\Gamma_g = 0.5$. Reflection at the generator end causes a reversal in the polarity of the pulse.

convenient function to use to specify when a waveform begins. In the case of multiple reflected waves on a transmission line, each reflected wave begins after time delays of l/v, $2l/v$, $3l/v$, etc., corresponding to the time delay to propagate a certain number of times back and forth between the generator end and load end.† The current wave can be obtained by multiplying the forward propagating waves by Y_c and the backward propagating waves by $-Y_c$. When the generator voltage $\mathscr{V}_g(t)$ exists for only a finite time interval, the total voltage wave on the line will decay toward zero since each successive reflected wave is multiplied by a reflection coefficient, either Γ_L or Γ_g, which is less than one in magnitude, and hence successive waves are of diminishing amplitude.

The sequence of multiple reflected waves can be illustrated in a distance-time plot. In Fig. 3.4 we show this type of plot for a generator producing a rectangular pulse. We have chosen $R_L = 3Z_c$ and $R_g = Z_c/3$ so that $\Gamma_L = 0.5$, $\Gamma_g = -0.5$. When the reflection coefficient is negative, the

†The unit step functions were introduced for clarity in describing the physical process but are actually not required in (3.12) since $\mathscr{V}_g(t - \tau) = 0$ for $t < \tau$.

sign of the reflected voltage wave is reversed and this is illustrated in Fig. 3.4. In high-speed digital circuits using interconnecting transmission lines, multiple reflected pulses are undesirable and can be avoided by terminating each transmission line in a resistance equal to its characteristic impedance.

3.3 CAPACITIVE TERMINATION

When the transmission line is terminated in a reactive element such as a capacitor as shown in Fig. 3.3*b*, the reflected wave will have a waveform different from that of the incident wave. The solution for the reflected wave is readily found from the condition that the sum of the incident plus reflected voltage wave at $z = l$ must satisfy the terminal conditions. The incident wave is again chosen to be $\mathscr{V}_i(l,t) = V^+\mathscr{V}_g(t - l/v)$. The reflected wave is initiated at time $t = l/v$ and will propagate from the point $z = l$ toward the generator. Therefore it is of the form

$$\mathscr{V}_r\left(t - \frac{l}{v} + \frac{z-l}{v}\right) = \mathscr{V}_r\left(t + \frac{z-2l}{v}\right)$$

For a capacitor we require

$$\mathscr{I}_L = C_L\frac{d\mathscr{V}_L}{dt} = C_L\left[\frac{d\mathscr{V}_i(l,t)}{dt} + \frac{d\mathscr{V}_r(t-l/v)}{dt}\right]$$

where $\mathscr{V}_r(t - l/v)$ is the reflected voltage wave at $z = l$. In addition, we use the condition

$$\mathscr{I}_L = Y_c\left[\mathscr{V}_i(l,t) - \mathscr{V}_r\left(t - \frac{l}{v}\right)\right]$$

From these two equations we obtain

$$\begin{aligned}\frac{d\mathscr{V}_r(t-l/v)}{dt} &+ \frac{1}{C_L Z_c}\mathscr{V}_r\left(t - \frac{l}{v}\right)\\ &= -\frac{d\mathscr{V}_i}{dt} + \frac{1}{C_L Z_c}\mathscr{V}_i\\ &= -V^+\frac{d\mathscr{V}_g(t-l/v)}{dt} + \frac{V^+}{C_L Z_c}\mathscr{V}_g\left(t - \frac{l}{v}\right)\end{aligned} \tag{3.13}$$

For a specific example we will consider the case when the generator produces a rectangular pulse given by

$$\mathscr{V}_g(t) = 1 \qquad 0 \le t \le T$$

The right-hand side of (3.13) will now become the source function

$$V^+\left\{\frac{1}{C_L Z_c}\left[U\left(t-\frac{l}{v}\right)-U\left(t-\frac{l}{v}-T\right)\right]-\delta\left(t-\frac{l}{v}\right)+\delta\left(t-\frac{l}{v}-T\right)\right\}$$

where $\delta(t-\alpha)$ is the Dirac delta function or impulse function that arises from the derivative of the rectangular pulse. We can integrate (3.13) by introducing the integrating factor $e^{t/\tau}$ where $\tau = C_L Z_c$. We note that

$$\frac{d}{dt}\left(\mathscr{V}_r e^{t/\tau}\right) = e^{t/\tau}\left(\frac{d\mathscr{V}_r}{dt}+\frac{1}{\tau}\mathscr{V}_r\right)$$

so consequently

$$\begin{aligned}\int_{l/v}^{t}\frac{d}{dt}\mathscr{V}_r e^{t/\tau}\,dt &= \mathscr{V}_r\left(t-\frac{l}{v}\right)e^{t/\tau}-\mathscr{V}_r(0)e^{l/v\tau}\\ &= \frac{V^+}{\tau}\int_{l/v}^{t}\left[U\left(t-\frac{l}{v}\right)-U\left(t-\frac{l}{v}-T\right)\right]e^{t/\tau}\,dt\\ &\quad - V^+\int_{l/v}^{t}\left[\delta\left(t-\frac{l}{v}\right)-\delta\left(t-\frac{l}{v}-T\right)\right]e^{t/\tau}\,dt\\ &= \begin{cases} V^+(e^{t/\tau}-2e^{l/v\tau}) & \dfrac{l}{v}\le t\le\dfrac{l}{v}+T\\[2ex] V^+(-2e^{l/v\tau}+2e^{(l+vT)/v\tau}) & t>\dfrac{l}{v}+T\end{cases}\end{aligned}$$

Since we have included the impulse functions as derivatives of the applied rectangular pulse, the lower limit of integration is regarded to be just before $t = l/v$ and thus $\mathscr{V}_r(0)$ is equal to zero since it corresponds to $\mathscr{V}_r(0-)$. Hence we obtain

$$\mathscr{V}_r\left(t-\frac{l}{v}\right) = \begin{cases} V^+(1-2e^{-(vt-l)/v\tau}) & \dfrac{l}{v}<t\le\dfrac{l}{v}+T\\[2ex] V^+(-2e^{l/v\tau}+2e^{(l+vT)/v\tau})e^{-t/\tau} & t>\dfrac{l}{v}+T\end{cases} \tag{3.14}$$

At $t = l/v$ the reflected wave has an amplitude equal to $-V^+$ which cancels that of the incident pulse. This is consistent with the requirement that initially the capacitor C_L is uncharged and must have a zero voltage across it. The capacitor charges to a final voltage level

$$\begin{aligned}V_c &= \mathscr{V}_i\left(l, t=\frac{l}{v}+T\right)+\mathscr{V}_r(T)\\ &= 2V^+(1-e^{-T/\tau})\end{aligned}$$

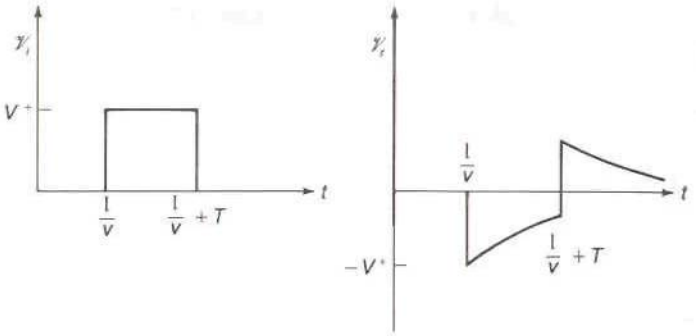


FIGURE 3.5
Incident and reflected voltage waveforms at $z = l$.

at $t = l/v + T$ and then discharges toward zero. The apparent discontinuous change that occurs in the reflected voltage wave at $t = l/v + T$ is caused by the sudden drop to zero volts for the incident pulse, and in order to match the voltage across the capacitor, the reflected voltage wave must have a positive jump of value V^+. The incident and reflected voltage waveforms at $z = l$ are shown in Fig. 3.5. The reflected voltage waveform will propagate toward the generator and will begin to initiate a new forward propagating wave at time $t = 2l/v$. Clearly the capacitor has made a significant change in the waveform of the reflected wave.

The analysis for the case of a capacitor-resistor termination as shown in Fig. 3.3*c* is similar. The terminal conditions are

$$\mathscr{I} = \mathscr{V}_L G_L + C_L \frac{d\mathscr{V}_L}{dt} = Y_c \left[\mathscr{V}_i(l,t) - \mathscr{V}_r\left(t - \frac{l}{v}\right) \right]$$

$$\mathscr{V}_L = \mathscr{V}_i(l,t) + \mathscr{V}_r\left(t - \frac{l}{v}\right)$$

so in place of (3.13) we have

$$\frac{d\mathscr{V}_r}{dt} + \left(\frac{1}{C_L Z_c} + \frac{1}{C_L R_L}\right)\mathscr{V}_r = -\frac{d\mathscr{V}_i}{dt} + \left(\frac{1}{C_L Z_c} - \frac{1}{C_L R_L}\right)\mathscr{V}_i \quad (3.15)$$

The solution is similar to that for (3.13) except that the charging time constant is now $\tau_1 = C_L R_L Z_c/(R_L + Z_c)$. Initially, $\mathscr{V}_r$ has a value equal to $-V^+$ as before. In this case the capacitor charges toward a final voltage equal to

$$V^+\left(1 + \frac{R_L - Z_c}{R_L + Z_c}\right) = \frac{2R_L}{R_L + Z_c} V^+$$

determined by the steady-state voltage across R_L if C_L was absent. At

$t = l/v + T$ the capacitor voltage will be

$$V_c = V^+ \frac{2R_L}{R_L + Z_c}(1 - e^{-T/\tau_1})$$

The reflected-wave voltage at this time will be $V_c - V^+$. When t becomes greater than $l/v + T$ the incident voltage wave pulse drops to zero volts so the reflected-wave voltage jumps to a value equal to V_c and will then decay toward zero with a time constant τ_1.

3.4 STEADY-STATE SINUSOIDAL WAVES

When the generator produces a sinusoidal voltage $\mathscr{V}_g(t) = V_g \cos \omega t$, the steady-state voltage waves on the transmission line will be of the form $\cos \omega(t - z/v)$ and $\cos \omega(t + z/v)$. The steady state is achieved, for all practical purposes, after a few multiple reflections have occurred, since the amplitude of the successive reflected wave decreases quite rapidly because it is multiplied by Γ_g or Γ_L upon each reflection. The solution for steady-state sinusoidal waves is most conveniently obtained using phasor analysis. The generator voltage is represented by $V_g e^{j\omega t}$. The voltage and current waves on the transmission line will then also have an $e^{j\omega t}$ time dependence. The differential equations (3.1a) and (3.1b) now become (the common time factor $e^{j\omega t}$ is dropped)

$$\frac{\partial V(z)}{\partial z} = -j\omega L I(z) \tag{3.16a}$$

$$\frac{\partial I(z)}{\partial z} = -j\omega C V(z) \tag{3.16b}$$

where $V(z)$ and $I(z)$ are complex phasor amplitudes. By eliminating the current we find that $V(z)$ satisfies the equation

$$\frac{d^2V(z)}{dz^2} + \frac{\omega^2}{v^2} V(z) = 0 \tag{3.16c}$$

The solution for $V(z)$ is of the form

$$V(z) = V^+ e^{-j\beta z} + V^- e^{j\beta z} \tag{3.17a}$$

with a corresponding solution

$$I(z) = I^+ e^{-j\beta z} - I^- e^{j\beta z} \tag{3.17b}$$

for the current waves. The constant $\beta = \omega/v$ is the propagation phase constant. As before the current amplitudes are related to the voltage amplitudes through the characteristic impedance of the line, i.e.,

$$I^+ = Y_c V^+ \qquad I^- = Y_c V^-$$

When the time factor is restored, it is readily seen that $e^{-j\beta z + j\omega t} = e^{j\omega(t - z/v)}$

corresponds to a wave propagating in the $+z$ direction, while $e^{j\beta z+j\omega t}$ is a wave propagating in the $-z$ direction. In the next section we will show that for a transmission line with finite conducting wires and possibly also surrounded with lossy dielectric materials, the waves attenuate in amplitude as they propagate. For this case the wave solutions are of the form

$$V = V^+e^{-j\beta z-\alpha z} + V^-e^{j\beta z+\alpha z} \tag{3.18a}$$

$$I = I^+e^{-j\beta z-\alpha z} - I^-e^{j\beta z+\alpha z} \tag{3.18b}$$

where α is the attenuation constant.

3.5 WAVES ON A LOSSY TRANSMISSION LINE

Conductors used in a transmission line will always have a finite conductivity and will therefore exhibit some series resistance. Furthermore, because of the skin effect the current flows in a thin layer at the surface of the conductor, the effective thickness of the layer being equal to the skin depth δ_s given by $(2/\omega\mu\sigma)^{1/2}$ [see (2.104)]. Consequently, the series resistance increases with an increase in the frequency of operation. In order to account for this resistance, a distributed series resistance R per meter must be included in the distributed circuit used to model the transmission line.

The two conductors in a transmission line are usually maintained parallel to each other by supporting them in a dielectric structure. For example, a coaxial transmission line is filled with a dielectric medium in order to keep the center conductor coaxial with the outer shield. Dielectric materials usually have a negligible conductance but do have a small amount of dielectric loss due to polarization loss in the dielectric. Consequently, a shunt conductance G per meter is added to the distributed circuit to account for this loss. Thus, for a lossy transmission line, the equivalent circuit of a differential length dz is chosen to be that shown in Fig. 3.6.

If the voltage and current at the input are $\mathscr{V}(z,t)$, $\mathscr{I}(z,t)$ and if the voltage and current at the output are

$$\mathscr{V} + \frac{\partial \mathscr{V}}{\partial z}\,dz \qquad \mathscr{I} + \frac{\partial \mathscr{I}}{\partial z}\,dz$$

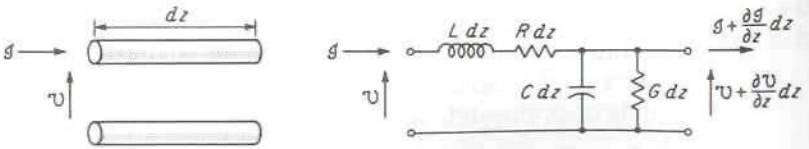


FIGURE 3.6
Equivalent circuit of a differential length of transmission line.

then Kirchhoff's laws give

$$\mathscr{V} - \left(\mathscr{V} + \frac{\partial \mathscr{V}}{\partial z}\,dz\right) = \mathscr{I} R\,dz + L\,dz\,\frac{\partial \mathscr{I}}{\partial t}$$

or

$$\frac{\partial \mathscr{V}}{\partial z} = -\mathscr{I} R - L\frac{\partial \mathscr{I}}{\partial t} \tag{3.19a}$$

Similarly,

$$\mathscr{I} - \left(\mathscr{I} + \frac{\partial \mathscr{I}}{\partial z}\,dz\right) = \mathscr{V} G\,dz + C\,dz\frac{\partial \mathscr{V}}{\partial t}$$

or

$$\frac{\partial \mathscr{I}}{\partial z} = -\mathscr{V} G - C\frac{\partial \mathscr{V}}{\partial t} \tag{3.19b}$$

The first equation states that the potential difference between the input and output is equal to the potential drop across R and L. The second equation states that the output current is less than the input current by an amount equal to the shunt current flowing through C and G. Differentiating (3.19a) with respect to z and (3.19b) with respect to time t gives

$$\frac{\partial^2 \mathscr{V}}{\partial z^2} = -R\frac{\partial \mathscr{I}}{\partial z} - L\frac{\partial^2 \mathscr{I}}{\partial t\,\partial z} \tag{3.20a}$$

$$\frac{\partial^2 \mathscr{I}}{\partial t\,\partial z} = -G\frac{\partial \mathscr{V}}{\partial t} - C\frac{\partial^2 \mathscr{V}}{\partial t^2} \tag{3.20b}$$

Using (3.19b) and (3.20b) in (3.20a) now gives the following equation for the line voltage $\mathscr{V}$:

$$\frac{\partial^2 \mathscr{V}}{\partial z^2} = R\left(G\mathscr{V} + C\frac{\partial \mathscr{V}}{\partial t}\right) + L\left(G\frac{\partial \mathscr{V}}{\partial t} + C\frac{\partial^2 \mathscr{V}}{\partial t^2}\right)$$

or

$$\frac{\partial^2 \mathscr{V}}{\partial z^2} - (RC + LG)\frac{\partial \mathscr{V}}{\partial t} - LC\frac{\partial^2 \mathscr{V}}{\partial t^2} - RG\mathscr{V} = 0 \tag{3.21}$$

The current $\mathscr{I}$ satisfies this one-dimensional wave equation also. If a solution in the form of a propagating wave

$$\mathscr{V} = \mathrm{Re}(Ve^{-\gamma z + j\omega t})$$

is assumed, substitution into (3.21) shows that the propagation constant γ must be a solution of

$$\gamma^2 - j\omega(RC + LG) + \omega^2 LC - RG = 0 \tag{3.22}$$

If only the steady-state sinusoidally time-varying solution is desired, phasor notation may be used. If we let V and I represent the voltage and current without the time dependence $e^{j\omega t}$, the basic equations (3.19) may be

written as

$$\frac{\partial V}{\partial z} = -(R + j\omega L)I \tag{3.23a}$$

$$\frac{\partial I}{\partial z} = -(G + j\omega C)V \tag{3.23b}$$

The wave equation (3.21) becomes

$$\frac{\partial^2 V}{\partial z^2} - (RG - \omega^2 LC)V - j\omega(RC + LG)V = 0 \tag{3.24}$$

The general solution to (3.24) is

$$V = V^+ e^{-\gamma z} + V^- e^{\gamma z} \tag{3.25}$$

where $\gamma = \alpha + j\beta$ is given by

$$\gamma = \left[-\omega^2 LC + RG + j\omega(RC + LG)\right]^{1/2} \tag{3.26}$$

from (3.22). The constants V^+ and V^- are arbitrary amplitude constants for waves propagating in the $+z$ and $-z$ directions, respectively. The solution for the current I may be found from (3.23a), that is

$$I = I^+ e^{-\gamma z} - I^- e^{+\gamma z} = \frac{\gamma}{R + j\omega L}(V^+ e^{-\gamma z} - V^- e^{\gamma z}) \tag{3.27}$$

The parameter

$$Z_c = \frac{R + j\omega L}{\gamma} = \left(\frac{R + j\omega L}{G + j\omega C}\right)^{1/2} \tag{3.28}$$

is the characteristic impedance of the line since it is equal to the ratio V^+/I^+ and V^-/I^-. Note that $\gamma = [(R + j\omega L)(G + j\omega C)]^{1/2}$.

Loss-Free Transmission Line

For a line without loss, i.e., for which $R = G = 0$, the propagation constant is

$$\gamma = j\beta = j\omega\sqrt{LC} \tag{3.29}$$

and the characteristic impedance is pure real and given by

$$Z_c = \sqrt{\frac{L}{C}} \tag{3.30}$$

According to the field analysis, β is also equal to $\omega(\mu\epsilon)^{1/2}$, and hence

$$LC = \mu\epsilon \tag{3.31}$$

for a transmission line. This result may also be verified from the solutions for L and C, as shown later in the section on transmission-line parameters.

Using (3.31) in (3.30) shows that the characteristic impedance is also given by

$$Z_c = \sqrt{\frac{L}{C}} = \sqrt{\frac{\mu\epsilon}{C^2}} = \frac{\epsilon}{C}\sqrt{\frac{\mu}{\epsilon}} = Z\frac{\epsilon}{C} \tag{3.32}$$

where Z is the intrinsic impedance of the medium. The characteristic impedance differs from the intrinsic impedance Z by a factor ϵ/C, which is a function of the line configuration only.

Low-Loss Transmission Line

For most microwave transmission lines the losses are very small; that is, $R \ll \omega L$ and $G \ll \omega C$. When this is the case, the term RG in the expression (3.26) for γ may be neglected. A binomial expansion then gives

$$\gamma \approx j\omega\sqrt{LC} + \frac{1}{2}\sqrt{LC}\left(\frac{R}{L} + \frac{G}{C}\right) = \alpha + j\beta \tag{3.33}$$

To first order the characteristic impedance is still given by (3.30) or (3.32). Thus the phase constant for a low-loss line is

$$\beta = \omega\sqrt{LC} \tag{3.34a}$$

and the attenuation constant α is

$$\alpha = \frac{1}{2}\sqrt{LC}\left(\frac{R}{L} + \frac{G}{C}\right) = \frac{1}{2}(RY_c + GZ_c) \tag{3.34b}$$

where $Y_c = Z_c^{-1} = \sqrt{C/L}$ is the characteristic admittance of the transmission line.

3.6 TERMINATED TRANSMISSION LINE: SINUSOIDAL WAVES

In this section the properties of a transmission line terminated in an arbitrary load impedance Z_L are examined. This will serve to illustrate how the forward and backward propagating waves can be combined to satisfy the boundary conditions at a termination. Figure 3.7 illustrates schematically a transmission line terminated in a load impedance Z_L. The line is assumed lossless and with a characteristic impedance Z_c and a propagation constant

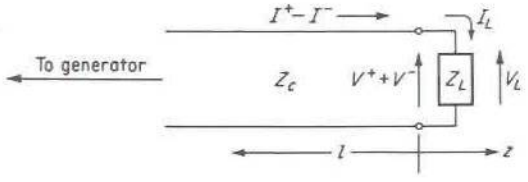


FIGURE 3.7
Terminated transmission line.

$\gamma = j\beta$. It should be noted that at microwave frequencies conventional low-frequency resistors, inductors, or capacitors, when connected across the two conductors of a transmission line, may behave as impedance elements with quite different characteristics from the low-frequency behavior.

If a voltage wave $V^+e^{-j\beta z}$ with an associated current $I^+e^{-j\beta z}$ is incident on the termination, a reflected voltage wave $V^-e^{j\beta z}$ with a current $-I^-e^{j\beta z}$ will, in general, be created. The ratio of the reflected and incident wave amplitudes is determined by the load impedance only. At the load the total line voltage must equal the impressed voltage across the load and the line current must be continuous through the load. Hence, if Z_L is located at $z = 0$,

$$V = V^+ + V^- = V_L \tag{3.35a}$$

$$I = I^+ - I^- = I_L \tag{3.35b}$$

But $I^+ = Y_c V^+$, $I^- = Y_c V^-$, and $V_L/I_L = Z_L$ by definition of load impedance. Therefore

$$V^+ + V^- = V_L \tag{3.36a}$$

$$V^+ - V^- = \frac{Z_c}{Z_L} V_L \tag{3.36b}$$

The ratio of V^- to V^+ is usually described by a voltage reflection coefficient Γ defined as

$$\Gamma_L = \frac{V^-}{V^+} \tag{3.37}$$

In place of (3.36) we may write

$$V^+(1 + \Gamma_L) = V_L$$

$$V^+(1 - \Gamma_L) = \frac{Z_c}{Z_L} V_L$$

Dividing one equation by the other yields

$$\frac{1 + \Gamma_L}{1 - \Gamma_L} = \frac{Z_L}{Z_c} \tag{3.38}$$

The quantity Z_L/Z_c is called the normalized load impedance (load impedance measured in units of Z_c), and $(1 + \Gamma_L)/(1 - \Gamma_L)$ is then the normalized input impedance seen looking toward the load at $z = 0$. The normalized load impedance will be expressed as $\bar{Z}_L$, with the bar on top signifying a normalized impedance in general. Solving for the voltage reflection coefficient Γ gives

$$\Gamma_L = \frac{Z_L - Z_c}{Z_L + Z_c} = \frac{Z_L/Z_c - 1}{Z_L/Z_c + 1} = \frac{\bar{Z}_L - 1}{\bar{Z}_L + 1} \tag{3.39}$$

Analogous to a voltage reflection coefficient, a current reflection coefficient Γ_I could also be introduced. In the present case

$$\Gamma_I = \frac{-I^-}{I^+} = -\frac{Y_c V^-}{Y_c V^+} = -\Gamma_L$$

In this text, however, only the voltage reflection coefficient will be used; so the adjective "voltage" can be dropped without confusion.

The incident voltage wave can be considered as transmitting a voltage V_L across the load, and a voltage transmission coefficient T can be defined as giving V_L in terms of V^+; thus

$$V_L = TV^+ = (1 + \Gamma_L)V^+$$

So

$$T = 1 + \Gamma_L \tag{3.40}$$

A corresponding current transmission coefficient is not used in this book.

Returning to (3.39), it is seen that if $Z_L = Z_c$, the reflection coefficient is zero. In this case all the power in the incident wave is transmitted to the load and none of it is reflected back toward the generator. The power delivered to the load in this case is

$$P = \tfrac{1}{2}\,\mathrm{Re}(VI^*) = \tfrac{1}{2}|V^+|^2 Y_c = \tfrac{1}{2}|V_L|^2 Y_L \tag{3.41}$$

The load is said to be matched to the transmission line when $\Gamma_L = 0$.

If Z_L does not equal Z_c, the load is mismatched to the line and a reflected wave is produced. The power delivered to the load is now given by

$$\begin{aligned} P &= \tfrac{1}{2}\,\mathrm{Re}(V_L I_L^*) = \tfrac{1}{2}\,\mathrm{Re}\left[(V^+ + V^-)(I^+ - I^-)^*\right] \\ &= \tfrac{1}{2}\,\mathrm{Re}\left[Y_c(V^+ + V^-)(V^+ - V^-)^*\right] \\ &= \tfrac{1}{2}\,\mathrm{Re}\left[Y_c|V^+|^2(1 + \Gamma_L)(1 - \Gamma_L)^*\right] \\ &= \tfrac{1}{2}Y_c|V^+|^2\left(1 - |\Gamma_L|^2\right) \end{aligned} \tag{3.42}$$

The final result states the physically obvious result that the power delivered to the load is the incident power minus that reflected from the load.

In the absence of reflection, the magnitude of the voltage along the line is a constant equal to $|V^+|$. When a reflected wave also exists, the incident and reflected waves interfere to produce a standing-wave pattern along the line. The voltage at any point on the line ($z < 0$) is given by

$$V = V^+ e^{-j\beta z} + \Gamma_L V^+ e^{j\beta z}$$

and has a magnitude given by

$$|V| = |V^+|\,|1 + \Gamma_L e^{2j\beta z}| = |V^+|\,|1 + \Gamma_L e^{-2j\beta l}|$$

where $l = -z$ is the positive distance measured from the load toward the

generator, as in Fig. 3.7. Let Γ_L be equal to $\rho e^{j\theta}$, where $\rho = |\Gamma_L|$; then†

$$\begin{aligned} |V| &= |V^+|\,|1 + \rho e^{j(\theta - 2\beta l)}| \\ &= |V^+|\left\{[1 + \rho\cos(\theta - 2\beta l)]^2 + \rho^2 \sin^2(\theta - 2\beta l)\right\}^{1/2} \\ &= |V^+|\left\{(1+\rho)^2 - 2\rho[1 - \cos(\theta - 2\beta l)]\right\}^{1/2} \\ &= |V^+|\left[(1+\rho)^2 - 4\rho \sin^2\left(\beta l - \frac{\theta}{2}\right)\right]^{1/2} \end{aligned} \tag{3.43}$$

This result shows that $|V|$ oscillates back and forth between maximum values of $|V^+|(1 + \rho)$ when $\beta l - \theta/2 = n\pi$ and minimum values $|V^+|(1 - \rho)$ when $\beta l - \theta/2 = n\pi + \pi/2$, where n is an integer. These results also agree with physical intuition since they state that voltage maxima occur when the incident and reflected waves add in phase and that voltage minima occur when they add 180° out of phase. Successive maxima and minima are spaced a distance $d = \pi/\beta = \lambda\pi/2\pi = \lambda/2$ apart, where λ is the wavelength for TEM waves in the medium surrounding the conductors. The distance between a maximum and the nearest minimum is $\lambda/4$.

Since the current reflection coefficient is equal to $-\Gamma_L$ the current waves subtract whenever the voltage waves add up in phase. Hence current maxima and minima are displaced $\lambda/4$ from the corresponding voltage maxima and minima. Figure 3.8 illustrates the voltage and current standing-wave patterns that result when Z_L is a pure resistance equal to $3Z_c$.

The ratio of the maximum line voltage to the minimum line voltage is called the voltage standing-wave ratio S; thus

$$S = \frac{|V^+|(1+\rho)}{|V^+|(1-\rho)} = \frac{1+\rho}{1-\rho} \tag{3.44}$$

This is a parameter of considerable importance in practice for the following reasons: At microwave frequencies instruments for the direct absolute measurement of voltage or current are difficult to construct and use. On the other hand, devices to measure relative voltage or current (or electric or magnetic field) amplitudes are easy to construct. A typical device is a small probe inserted into the region of the electric field around a line. The output of the probe is connected to a crystal rectifier, and produces an output current which is a measure of the relative electric field or voltage at the probe position. By moving the probe along the line, the standing-wave ratio can be measured directly in terms of the maximum and minimum probe

†The symbol ρ denotes both charge density and the modulus of the reflection coefficient. The context makes it clear which quantity is under discussion; so confusion should not occur.

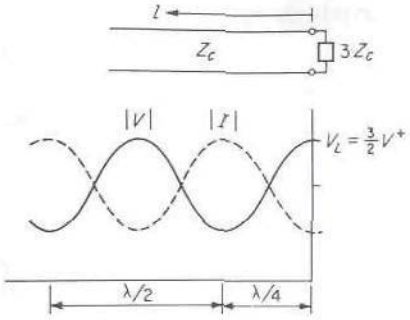


FIGURE 3.8
Voltage and current standing-wave patterns on a line terminated in a load impedance equal to $3Z_c$.

currents. The location of a voltage minimum can also be measured, and this permits the phase angle θ of Γ_L to be calculated. Since ρ is known from the measured value of S, Γ_L is specified, and the normalized load impedance may be calculated from (3.38).

Although the reflection coefficient was introduced as a measure of the ratio of reflected- to incident-wave amplitudes at the load, the definition may be extended to give the corresponding voltage ratio at any point on the line. Thus, at $z = -l$, the reflection coefficient is

$$\Gamma(l) = \frac{V^- e^{-j\beta l}}{V^+ e^{j\beta l}} = \frac{V^-}{V^+} e^{-2j\beta l} = \Gamma_L e^{-2j\beta l} \tag{3.45}$$

where $\Gamma_L = V^-/V^+$ denotes the reflection coefficient of the load. The normalized impedance, seen looking toward the load, at $z = -l$, is

$$\begin{aligned} \bar{Z}_{\text{in}} = \frac{Z_{\text{in}}}{Z_c} = \frac{V}{IZ_c} &= \frac{V^+ e^{j\beta l} + V^- e^{-j\beta l}}{V^+ e^{j\beta l} - V^- e^{-j\beta l}} \\ &= \frac{1 + \Gamma(l)}{1 - \Gamma(l)} = \frac{1 + \Gamma_L e^{-2j\beta l}}{1 - \Gamma_L e^{-2j\beta l}} \end{aligned} \tag{3.46}$$

By replacing Γ_L by $(Z_L - Z_c)/(Z_L + Z_c)$ and $e^{\pm j\beta l}$ by $\cos \beta l \pm j \sin \beta l$, this result may be expressed as

$$\bar{Z}_{\text{in}} = \frac{Z_{\text{in}}}{Z_c} = \frac{Z_L + jZ_c \tan \beta l}{Z_c + jZ_L \tan \beta l} \tag{3.47}$$

A similar result holds for the normalized input admittance; so

$$\bar{Y}_{\text{in}} = \frac{Y_{\text{in}}}{Y_c} = \frac{Y_L + jY_c \tan \beta l}{Y_c + jY_L \tan \beta l} = \frac{\bar{Y}_L + j \tan \beta l}{1 + j\bar{Y}_L \tan \beta l} \tag{3.48}$$

Of particular interest are two special cases, namely, $\beta l = \pi$ or $l = \lambda/2$ and

$\beta l = \pi/2$ or $l = \lambda/4$, for which

$$Z_{\text{in}}\left(l = \frac{\lambda}{2}\right) = Z_L \tag{3.49a}$$

$$Z_{\text{in}}\left(l = \frac{\lambda}{4}\right) = \frac{Z_c^2}{Z_L} \tag{3.49b}$$

The first is equivalent to an ideal one-to-one impedance transformer, whereas in the second case the impedance has been inverted with respect to Z_c^2.

The terminal conditions at the generator end are readily established by using (3.47) to evaluate the input impedance Z_{in} seen looking toward the load at the generator end. If the generator with open-circuit voltage V_g has an internal impedance Z_g, then by the usual voltage division formula the total transmission-line voltage V at $z = -l$ will be given by

$$V = \frac{Z_{\text{in}}}{Z_{\text{in}} + Z_g} V_g$$

But V is the sum of the forward-propagating-wave and reflected-wave voltages at $z = -l$, that is,

$$\begin{aligned} V &= V^+ e^{j\beta l} + V^- e^{-j\beta l} \\ &= V^+ e^{j\beta l}\left(1 + \Gamma_L e^{-2j\beta l}\right) \end{aligned}$$

By using this expression we can solve for V^+ which is found to be

$$\begin{aligned} V^+ &= \frac{Z_{\text{in}} V_g}{(Z_{\text{in}} + Z_g)\left(e^{j\beta l} + \Gamma_L e^{-j\beta l}\right)} \\ &= \frac{Z_{\text{in}}(Z_L + Z_c)V_g}{2(Z_{\text{in}} + Z_g)(Z_L \cos \beta l + jZ_c \sin \beta l)} \end{aligned} \tag{3.50}$$

Terminated Lossy Line

In the case of a lossy line with propagation constant $\gamma = j\beta + \alpha$, the previous equations hold except that $j\beta$ must be replaced by $j\beta + \alpha$, where α is usually so small that, for the short lengths of line used in most experimental setups, the neglect of α is justified. Nevertheless, it is of some interest to examine the behavior of a lossy transmission line terminated in a load Z_L. One simplifying assumption will be made, and this is that the characteristic impedance Z_c can still be considered real. This assumption is certainly valid for low-loss lines of the type used at microwave frequencies.

A detailed calculation justifying this assumption for a typical case is called for in Prob. 3.18.

Clearly, the presence of an attenuation constant α does not affect the definition of the voltage reflection coefficient Γ_L for the load. However, at any other point a distance l toward the generator, the reflection coefficient is now given by

$$\Gamma(l) = \Gamma_L e^{-2j\beta l - 2\alpha l} \tag{3.51}$$

As l is increased, Γ decreases exponentially until, for large l, it essentially vanishes. Thus, whenever a load Z_L is viewed through a long section of lossy line, it appears to be matched to the line since Γ is negligible at the point considered. This effect may also be seen from the expression for the input impedance, namely,

$$Z_{\text{in}} = Z_c \frac{1 + \Gamma_L e^{-2j\beta l - 2\alpha l}}{1 - \Gamma_L e^{-2j\beta l - 2\alpha l}} = Z_c \frac{Z_L + Z_c \tanh(j\beta l + \alpha l)}{Z_c + Z_L \tanh(j\beta l + \alpha l)} \tag{3.52}$$

which approaches Z_c for l large since $\tanh x$ approaches 1 for x large and not a pure imaginary quantity.

The losses also have the effect of reducing the standing-wave ratio S toward unity as the point of observation is moved away from the load toward the generator. As the generator is approached, the incident-wave amplitude increases exponentially whereas the reflected-wave amplitude decreases exponentially. The result is a standing-wave pattern of the type illustrated in Fig. 3.9. For illustrative purposes a relatively large value of α has been assumed here.

The power delivered to the load is given by

$$P_L = \frac{1}{2} \operatorname{Re}(V_L I_L^*) = \frac{1}{2} |V_L|^2 G_L = \frac{Y_c}{2} |V^+|^2 (1 - |\Gamma_L|^2) \tag{3.53}$$

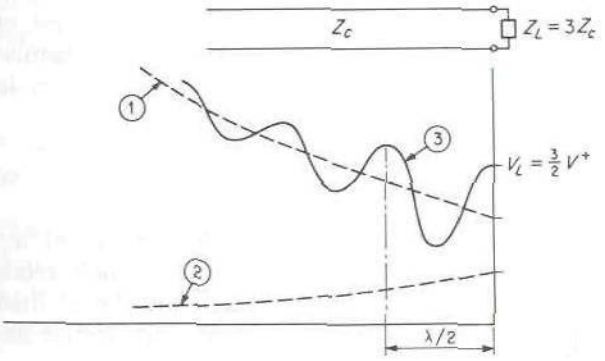


FIGURE 3.9
Voltage standing-wave pattern on a lossy transmission line. (1) Envelope of incident-wave amplitude; (2) envelope of reflected-wave amplitude; (3) standing-wave pattern.

as before. At some point $z = -l$, the power directed toward the load is

$$P(l) = \frac{1}{2}\,\mathrm{Re}(VI^*) = \frac{1}{2}|V|^2 Y_c = \frac{Y_c}{2}|V^+ e^{\alpha l}|^2\left[1 - |\Gamma(l)|^2\right]$$

$$= \frac{Y_c}{2}|V^+|^2\left(e^{2\alpha l} - |\Gamma_L|^2\right) \tag{3.54}$$

where $|\Gamma|e^{\alpha l}$ has been replaced by $|\Gamma_L|$. Of the power given by (3.54) only that portion corresponding to P_L as given by (3.53) is delivered to the load. The remainder is dissipated in the lossy line, this remainder being given by

$$P(l) - P_L = \frac{Y_c}{2}|V^+|^2(e^{2\alpha l} - 1) \tag{3.55}$$

PART 2
FIELD ANALYSIS OF TRANSMISSION LINES

The first section of this part will show that Maxwell's equations can be reformulated so as to describe three classes of waves, TEM, TE, and TM waves. The TEM wave is the principal wave on transmission lines. From the solution for the electric and magnetic fields for the TEM wave, we will be able to establish that there are unique voltage and current waves associated with the TEM wave. We will also be able to evaluate the distributed-circuit parameters R, L, C, and G for a transmission line. The field analysis thus provides a theoretical basis for treating the transmission line as a distributed circuit.

After the basic equations for TEM, TE, and TM waves have been derived, we present the field analysis for transmission lines that support TEM waves. This is followed by several sections dealing with planar transmission lines. Many of the planar lines that we examine support only quasi-TEM waves but can be analyzed as transmission lines once their equivalent distributed-circuit parameters have been determined.

3.7 CLASSIFICATION OF WAVE SOLUTIONS

The transmission lines and waveguides analyzed in this chapter are all characterized by having axial uniformity. Their cross-sectional shape and electrical properties do not vary along the axis, which is chosen as the z axis. Since sources are not considered, the electric and magnetic fields are

solutions of the homogeneous vector Helmholtz equation, i.e.,

$$\nabla^2\mathbf{E} + k_0^2\mathbf{E} = 0 \qquad \nabla^2\mathbf{H} + k_0^2\mathbf{H} = 0$$

The type of solution sought is that corresponding to a wave that propagates along the z axis. Since the Helmholtz equation is separable, it is possible to find solutions of the form $f(z)g(x, y)$, where f is a function of z only and g is a function of x and y or other suitable transverse coordinates only. The second derivative with respect to z enters into the wave equation in a manner similar to the second derivative with respect to time. By analogy with $e^{j\omega t}$ as the time dependence, the z dependence can be assumed to be $e^{\pm j\beta z}$. This assumption will lead to wave solutions of the form $\cos(\omega t \pm \beta z)$ and $\sin(\omega t \pm \beta z)$, which are appropriate for describing wave propagation along the z axis. A wave propagating in the positive z direction is represented by $e^{-j\beta z}$, and $e^{j\beta z}$ corresponds to a wave propagating in the negative z direction. With an assumed z dependence $e^{-j\beta z}$, the del operator becomes $\nabla = \nabla_t + \nabla_z = \nabla_t - j\beta\mathbf{a}_z$, since $\nabla_z = \mathbf{a}_z\,\partial/\partial z$. Note that ∇_t is the transverse part and equals $\mathbf{a}_x\,\partial/\partial x + \mathbf{a}_y\,\partial/\partial y$ in rectangular coordinates. The propagation phase constant β will turn out to depend on the waveguide configuration.

Considerable simplification of Maxwell's equations is obtained by decomposing all fields into transverse and axial components and separating out the z dependence. Thus let (the time dependence $e^{j\omega t}$ is suppressed)

$$\begin{aligned}\mathbf{E}(x, y, z) &= \mathbf{E}_t(x, y, z) + \mathbf{E}_z(x, y, z)\\ &= \mathbf{e}(x, y)e^{-j\beta z} + \mathbf{e}_z(x, y)e^{-j\beta z} \qquad (3.56)\\ \mathbf{H}(x, y, z) &= \mathbf{H}_t(x, y, z) + \mathbf{H}_z(x, y, z)\\ &= \mathbf{h}(x, y)e^{-j\beta z} + \mathbf{h}_z(x, y)e^{-j\beta z} \qquad (3.57)\end{aligned}$$

where $\mathbf{E}_t, \mathbf{H}_t$ are the transverse (x and y) components, and $\mathbf{E}_z, \mathbf{H}_z$ are the axial components. Note also that $\mathbf{e}(x, y), \mathbf{h}(x, y)$ are transverse vector functions of the transverse coordinates only, and $\mathbf{e}_z(x, y), \mathbf{h}_z(x, y)$ are axial vector functions of the transverse coordinates.

Consider the $\nabla \times \mathbf{E}$ equation, which may be expanded to give

$$\nabla \times \mathbf{E} = (\nabla_t - j\beta\mathbf{a}_z) \times (\mathbf{e} + \mathbf{e}_z)e^{-j\beta z} = -j\omega\mu_0(\mathbf{h} + \mathbf{h}_z)e^{-j\beta z}$$

or $$\nabla_t \times \mathbf{e} - j\beta\mathbf{a}_z \times \mathbf{e} + \nabla_t \times \mathbf{e}_z - j\beta\mathbf{a}_z \times \mathbf{e}_z = -j\omega\mu_0\mathbf{h} - j\omega\mu_0\mathbf{h}_z$$

The term $\mathbf{a}_z \times \mathbf{e}_z = 0$, and $\nabla_t \times \mathbf{e}_z = \nabla_t \times \mathbf{a}_z e_z = -\mathbf{a}_z \times \nabla_t e_z$. Note also that $\nabla_t \times \mathbf{e}$ is directed along the z axis only, since it involves factors such as $\mathbf{a}_x \times \mathbf{a}_y$, $\mathbf{a}_x \times \mathbf{a}_x$, $\mathbf{a}_y \times \mathbf{a}_x$, and $\mathbf{a}_y \times \mathbf{a}_y$, whereas $\mathbf{a}_z \times \mathbf{e}$ and $\nabla_t \times \mathbf{e}_z$ have transverse components only. Consequently, when the transverse and axial

components of the above equation are equated, there results

$$\nabla_t \times \mathbf{e} = -j\omega\mu_0 \mathbf{h}_z \tag{3.58a}$$

$$\nabla_t \times \mathbf{e}_z - j\beta \mathbf{a}_z \times \mathbf{e} = -\mathbf{a}_z \times \nabla_t e_z - j\beta \mathbf{a}_z \times \mathbf{e} = -j\omega\mu_0 \mathbf{h} \tag{3.58b}$$

In a similar manner the $\nabla \times \mathbf{H}$ equation yields

$$\nabla_t \times \mathbf{h} = j\omega\epsilon_0 \mathbf{e}_z \tag{3.58c}$$

$$\mathbf{a}_z \times \nabla_t h_z + j\beta \mathbf{a}_z \times \mathbf{h} = -j\omega\epsilon_0 \mathbf{e} \tag{3.58d}$$

The divergence equation $\nabla \cdot \mathbf{B} = 0$ becomes

$$\nabla \cdot \mathbf{B} = \nabla \cdot \mu_0 \mathbf{H} = (\nabla_t - j\beta \mathbf{a}_z) \cdot (\mathbf{h} + \mathbf{h}_z)\mu_0 e^{-j\beta z}$$

$$= (\nabla_t \cdot \mathbf{h} - j\beta \mathbf{a}_z \cdot \mathbf{h}_z)\mu_0 e^{-j\beta z} = 0$$

or

$$\nabla_t \cdot \mathbf{h} = j\beta h_z \tag{3.58e}$$

Similarly, $\nabla \cdot \mathbf{D} = 0$ gives

$$\nabla_t \cdot \mathbf{e} = j\beta e_z \tag{3.58f}$$

This reduced form of Maxwell's equations will prove to be very useful in formulating the solutions for waveguiding systems.

For a large variety of waveguides of practical interest it turns out that all the boundary conditions can be satisfied by fields that do not have all components present. Specifically, for transmission lines, the solution of interest is a transverse electromagnetic wave with transverse components only, that is, $E_z = H_z = 0$, whereas for waveguides, solutions with $E_z = 0$ or $H_z = 0$ are possible. Because of the widespread occurrence of such field solutions, the following classification of solutions is of particular interest.

1. Transverse electromagnetic (TEM) waves. For TEM waves, $E_z = H_z = 0$. The electric field may be found from the transverse gradient of a scalar function $\Phi(x, y)$, which is a function of the transverse coordinates only and is a solution of the two-dimensional Laplace equation.
2. Transverse electric (TE), or H, modes. These solutions have $E_z = 0$, but $H_z \neq 0$. All the field components may be derived from the axial component H_z of magnetic field.
3. Transverse magnetic (TM), or E, modes. These solutions have $H_z = 0$, but $E_z \neq 0$. The field components may be derived from E_z.

In some cases it will be found that a TE or TM mode by itself will not satisfy all the boundary conditions. However, in such cases linear combinations of TE and TM modes may be used, since such linear combinations always provide a complete and general solution. Although other possible types of wave solutions may be constructed, the above three types are the most useful in practice and by far the most commonly used ones.

The appropriate equations to be solved to obtain TEM, TE, or TM modes will be derived below by placing E_z and H_z, E_z, and H_z, respectively, equal to zero in Maxwell's equations.

TEM Waves

For TEM waves $e_z = h_z = 0$; so (3.3) reduces to

$$\nabla_t \times \mathbf{e} = 0 \tag{3.59a}$$

$$\beta \mathbf{a}_z \times \mathbf{e} = \omega \mu_0 \mathbf{h} \tag{3.59b}$$

$$\nabla_t \times \mathbf{h} = 0 \tag{3.59c}$$

$$\beta \mathbf{a}_z \times \mathbf{h} = -\omega \epsilon_0 \mathbf{e} \tag{3.59d}$$

$$\nabla_t \cdot \mathbf{h} = 0 \tag{3.59e}$$

$$\nabla_t \cdot \mathbf{e} = 0 \tag{3.59f}$$

The vanishing of the transverse curl of $\mathbf{e}$ means that the line integral of $\mathbf{e}$ around any closed path in the xy plane is zero. This must clearly be so since *there is no axial magnetic flux passing through such a contour.* Although $\nabla_t \times \mathbf{h} = 0$ when there are no volume currents present, the line integral of $\mathbf{h}$ will not vanish for a transmission line with conductors on which axial currents may exist. This point will be considered again later when transmission lines are analyzed. Equation (3.59a) is just the condition that permits $\mathbf{e}$ to be expressed as the gradient of a scalar potential. Hence let

$$\mathbf{e}(x, y) = -\nabla_t \Phi(x, y) \tag{3.60}$$

Using (3.59f) shows that Φ is a solution of the two-dimensional Laplace equation

$$\nabla_t^2 \Phi(x, y) = 0 \tag{3.61}$$

The electric field is thus given by

$$\mathbf{E}_t(x, y, z) = -\nabla_t \Phi(x, y) e^{-j\beta z}$$

But this field must also satisfy the Helmholtz equation

$$\nabla^2 \mathbf{E}_t + k_0^2 \mathbf{E}_t = 0$$

Since $\nabla = \nabla_t - j\beta \mathbf{a}_z$, $\nabla^2 = \nabla_t^2 - \beta^2$, that is, the second derivative with *respect to z gives a factor $-\beta^2$, this reduces to*

$$\nabla_t^2 \mathbf{E}_t + (k_0^2 - \beta^2)\mathbf{E}_t = 0$$

or

$$\nabla_t \left[\nabla_t^2 \Phi + (k_0^2 - \beta^2)\Phi \right] = 0$$

This shows that $\beta = \pm k_0$ for TEM waves, a result to be anticipated from the wave solutions discussed in Chap. 2. The magnetic field may be found

from the $\nabla \times \mathbf{E}$ equation, i.e., from (3.59b); thus

$$\frac{\omega\mu_0}{k_0}\mathbf{h} = \mathbf{a}_z \times \mathbf{e} = Z_0\mathbf{h} \tag{3.62}$$

In summary, for TEM waves, first find a scalar potential Φ which is a solution of

$$\nabla_t^2\Phi(x,y) = 0 \tag{3.63a}$$

and satisfies the proper boundary conditions. The fields are then given by

$$\mathbf{E} = \mathbf{E}_t = \mathbf{e}e^{\mp jk_0z} = -\nabla_t\Phi e^{\mp jk_0z} \tag{3.63b}$$

$$\mathbf{H} = \mathbf{H}_t = \pm\mathbf{h}e^{\mp jk_0z} = \pm Y_0\mathbf{a}_z \times \mathbf{e}e^{\mp jk_0z} \tag{3.63c}$$

where $k_0 = \omega(\mu_0\epsilon_0)^{1/2}$, $Y_0 = (\epsilon_0/\mu_0)^{1/2}$, and e^{-jk_0z} represents a wave propagating in the $+z$ direction and e^{jk_0z} corresponds to wave propagation in the $-z$ direction. For TEM waves, Z_0 is the wave impedance, and from (3.63c) it is seen that, for wave propagation in the $+z$ direction,

$$\frac{E_x}{H_y} = -\frac{E_y}{H_x} = Z_0 \tag{3.64a}$$

whereas for propagation in the $-z$ direction,

$$\frac{E_x}{H_y} = -\frac{E_y}{H_x} = -Z_0 \tag{3.64b}$$

E Waves

For transverse electric (TE) waves, h_z plays the role of a potential function from which the rest of the field components may be obtained. The magnetic field $\mathbf{H}$ is a solution of

$$\nabla^2\mathbf{H} + k_0^2\mathbf{H} = 0$$

Separating the above equation into transverse and axial parts and replacing ∇^2 by $\nabla_t^2 - \beta^2$ yield

$$\nabla_t^2 h_z(x,y) + k_c^2 h_z(x,y) = 0 \tag{3.65a}$$

$$\nabla_t^2\mathbf{h} + k_c^2\mathbf{h} = 0 \tag{3.65b}$$

where $k_c^2 = k_0^2 - \beta^2$ and a z dependence $e^{-j\beta z}$ is assumed. Unlike the case of TEM waves, β^2 will not equal k_0^2 for TE waves. Instead, β is determined by the parameter k_c^2 in (3.65a). When this equation is solved, subject to appropriate boundary conditions, the eigenvalue k_c^2 will be found to be a function of the waveguide configuration.

The Maxwell equations (3.58) with $\mathbf{e}_z = 0$ become

$$\nabla_t \times \mathbf{e} = -j\omega\mu_0 \mathbf{h}_z \tag{3.66a}$$

$$\beta \mathbf{a}_z \times \mathbf{e} = \omega\mu_0 \mathbf{h} \tag{3.66b}$$

$$\nabla_t \times \mathbf{h} = 0 \tag{3.66c}$$

$$\mathbf{a}_z \times \nabla_t h_z + j\beta \mathbf{a}_z \times \mathbf{h} = -j\omega\epsilon_0 \mathbf{e} \tag{3.66d}$$

$$\nabla_t \cdot \mathbf{h} = j\beta h_z \tag{3.66e}$$

$$\nabla_t \cdot \mathbf{e} = 0 \tag{3.66f}$$

The transverse curl of (3.66c) gives

$$\nabla_t \times (\nabla_t \times \mathbf{h}) = \nabla_t \nabla_t \cdot \mathbf{h} - \nabla_t^2 \mathbf{h} = 0$$

Replacing $\nabla_t \cdot \mathbf{h}$ by $j\beta h_z$ from (3.58e) and $\nabla_t^2 \mathbf{h}$ by $-k_c^2 \mathbf{h}$ from (3.65b) leads to the solution for $\mathbf{h}$ in terms of h_z; namely,

$$\mathbf{h} = -\frac{j\beta}{k_c^2} \nabla_t h_z \tag{3.67}$$

To find $\mathbf{e}$ in terms of $\mathbf{h}$, take the vector product of (3.66b) with $\mathbf{a}_z$ to obtain

$$\beta \mathbf{a}_z \times (\mathbf{a}_z \times \mathbf{e}) = \beta[(\mathbf{a}_z \cdot \mathbf{e})\mathbf{a}_z - (\mathbf{a}_z \cdot \mathbf{a}_z)\mathbf{e}] = -\beta \mathbf{e} = \omega\mu_0 \mathbf{a}_z \times \mathbf{h}$$

or

$$\mathbf{e} = -\frac{\omega\mu_0}{\beta} \mathbf{a}_z \times \mathbf{h} = -\frac{k_0}{\beta} Z_0 \mathbf{a}_z \times \mathbf{h} \tag{3.68}$$

The factor $k_0 Z_0/\beta$ has the dimensions of an impedance, and is called the wave impedance of TE, or H, modes. It will be designated by the symbol Z_h, so that

$$Z_h = \frac{k_0}{\beta} Z_0 \tag{3.69}$$

Thus, in component form, (3.68) gives

$$\frac{e_x}{h_y} = -\frac{e_y}{h_x} = Z_h \tag{3.70}$$

for a wave with z dependence $e^{-j\beta z}$.

The remaining equations in the set (3.66) do not yield any new results; so the solution for TE waves may be summarized as follows: First find a solution for h_z, where

$$\nabla_t^2 h_z + k_c^2 h_z = 0 \tag{3.71a}$$

Then

$$\mathbf{h} = -\frac{j\beta}{k_c^2} \nabla_t h_z \tag{3.71b}$$

and

$$\mathbf{e} = -Z_h \mathbf{a}_z \times \mathbf{h} \tag{3.71c}$$

where $\beta = (k_0^2 - k_c^2)^{1/2}$ and $Z_h = \dfrac{k_0 Z_0}{\beta}$

Complete expressions for the fields are

$$\mathbf{H} = \pm \mathbf{h} e^{\mp j\beta z} + \mathbf{h}_z e^{\mp j\beta z} \tag{3.71d}$$

$$\mathbf{E} = \mathbf{E}_t = \mathbf{e} e^{\mp j\beta z} \tag{3.71e}$$

Note that in (3.71*d*) the sign in front of $\mathbf{h}$ is reversed for a wave propagating in the $-z$ direction since $\mathbf{h}$ will be defined by (3.71*b*), with β positive regardless of whether propagation is in the $+z$ or $-z$ direction. The sign in front of $\mathbf{e}$ does not change since it involves the factor β twice, once in the expression for $\mathbf{h}$ and again in Z_h. Only the sign of one of $\mathbf{e}$ or $\mathbf{h}$ can change if a reversal in the direction of energy flow is to occur. That is, the solution for a wave propagating in the $-z$ direction can be chosen as $\mathbf{E} = -\mathbf{e}e^{j\beta z}$, $\mathbf{H} = (\mathbf{h} - \mathbf{h}_z)e^{j\beta z}$ or as $\mathbf{E} = \mathbf{e}e^{j\beta z}$, $\mathbf{H} = (-\mathbf{h} + \mathbf{h}_z)e^{j\beta z}$. One solution is the negative of the other. The latter solution is arbitrarily chosen as the standard in this text.

M Waves

The TM, or E, waves have $\mathbf{h}_z = 0$, but the axial electric field $\mathbf{e}_z$ is not zero. These modes may be considered the dual of the TE modes in that the roles of electric and magnetic fields are interchanged. The derivation of the equations to be solved parallels that for TE waves, and hence only the final results will be given.

First obtain a solution for e_z, where

$$\nabla_t^2 e_z + k_c^2 e_z = 0 \tag{3.72a}$$

subject to the boundary conditions imposed. This will serve to determine the eigenvalue k_c^2. The transverse fields are then given by

$$\mathbf{E}_t = \mathbf{e} e^{\mp j\beta z} = -\frac{j\beta}{k_c^2} \nabla_t e_z e^{\mp j\beta z} \tag{3.72b}$$

$$\mathbf{H}_t = \pm \mathbf{h} e^{\mp j\beta z} = \pm Y_e \mathbf{a}_z \times \mathbf{e} e^{\mp j\beta z} \tag{3.72c}$$

where $\beta = (k_0^2 - k_c^2)^{1/2}$ and the wave admittance Y_e for TM waves is given by

$$Y_e = Z_e^{-1} = \frac{k_0}{\beta} Y_0 \tag{3.72d}$$

The dual nature of TE and TM waves is exhibited by the relation

$$Z_e Z_h = Z_0^2 \tag{3.73}$$

which holds when both types of waves have the same value of β and is derivable from (3.69) and (3.72d). The complete expression for the electric field is

$$\mathbf{E} = \mathbf{E}_t + \mathbf{E}_z = \mathbf{e}e^{\mp j\beta z} \pm \mathbf{e}_z e^{\mp j\beta z}$$

$$= \left(-\frac{j\beta}{k_c^2} \nabla_t e_z \pm \mathbf{e}_z \right) e^{\mp j\beta z} \tag{3.74}$$

It is convenient to keep the sign of $\mathbf{e}$ *the same for propagation in either* the $+z$ or $-z$ direction. Since $\nabla \cdot \mathbf{E} = 0$, that is, $\nabla_t \cdot \mathbf{E}_t + \partial E_z/\partial z = 0$, this requires that the z component of electric field be $-\mathbf{e}_z e^{j\beta z}$ for a wave propagating in the $-z$ direction, because $\nabla_t \cdot \mathbf{E}_t$ does not change sign, whereas $\partial E_z/\partial z$ does, in view of the change in sign in front of β. The transverse magnetic field must also change sign upon reversal of the direction of propagation in order to obtain a change in the direction of energy flow. For reference, this sign convention is summarized below. The transverse variations of the fields are represented by the functions $\mathbf{e}$, $\mathbf{h}$, $\mathbf{e}_z$, *and* $\mathbf{h}_z$, *independent of the direction of propagation. Waves propagating in* the $+z$ direction are then given by

$$\mathbf{E} = \mathbf{E}^+ = (\mathbf{e} + \mathbf{e}_z)e^{-j\beta z} \tag{3.75a}$$

$$\mathbf{H} = \mathbf{H}^+ = (\mathbf{h} + \mathbf{h}_z)e^{-j\beta z} \tag{3.75b}$$

For propagation in the $-z$ direction the fields are

$$\mathbf{E} = \mathbf{E}^- = (\mathbf{e} - \mathbf{e}_z)e^{j\beta z} \tag{3.76a}$$

$$\mathbf{H} = \mathbf{H}^- = (-\mathbf{h} + \mathbf{h}_z)e^{j\beta z} \tag{3.76b}$$

Additional superscripts $(+)$ or $(-)$ will be used when it is necessary to indicate the direction of propagation. The previously derived equations for TEM, TE, and TM modes are valid in a medium with electrical constants ϵ, μ, provided these are used to replace ϵ_0, μ_0. A finite conductivity can also be taken into account by making ϵ complex, i.e., replacing ϵ by $\epsilon - j\sigma/\omega$.

The wave impedance introduced in the solutions is an extremely useful concept in practice. The wave impedance is always chosen to relate the transverse components of the electric and magnetic fields. The sign is always such that if i, j, k is a cyclic labeling of the coordinates and propagation is along the positive direction of coordinate k, the ratio $E_i/H_j = (Z_w)_k$ is positive. Here $(Z_w)_k$ is the wave impedance referred to the k axis as the direction of propagation. If i, j, k form an odd permutation of the coordinates, then E_i/H_j is negative. The usefulness of the wave-impedance concept stems from the fact that the power is given in terms of the

transverse fields alone. For example, for TE waves,

$$P = \frac{1}{2}\,\mathrm{Re}\int_S \mathbf{E}\times\mathbf{H}^*\cdot\mathbf{a}_z\,dx\,dy$$
$$= \frac{1}{2}\,\mathrm{Re}\int_S \mathbf{e}\times\mathbf{h}^*\cdot\mathbf{a}_z\,dx\,dy$$
$$= -\frac{1}{2}\,\mathrm{Re}\int_S Z_h(\mathbf{a}_z\times\mathbf{h})\times\mathbf{h}^*\cdot\mathbf{a}_z\,dx\,dy$$
$$= \frac{Z_h}{2}\int_S \mathbf{h}\cdot\mathbf{h}^*\,dx\,dy = \frac{Y_h}{2}\int_S \mathbf{e}\cdot\mathbf{e}^*\,dx\,dy$$

upon expanding the integrand. Thus the wave impedance enables the power transmitted to be expressed in terms of one of the transverse fields alone. A further property of the wave impedance, which will be dealt with later, is that it provides a basis for an analogy between conventional multiconductor transmission lines and waveguides.

3.8 TRANSMISSION LINES (FIELD ANALYSIS)

Lossless Transmission Line

A transmission line consists of two or more parallel conductors. Typical examples are the two-conductor line, shielded two-conductor line, and coaxial line with cross sections, as illustrated in Fig. 3.10. Initially, it will be assumed that the conductors are perfectly conducting and that the medium surrounding the conductors is air, with $\epsilon \approx \epsilon_0$, $\mu \approx \mu_0$. The effect of small losses will be considered later.

When the conductors are completely surrounded by a uniform dielectric medium, the principal wave that can exist on the transmission line is a TEM wave. The electric field for this wave can be found from the scalar potential which is a solution of Laplace's equation in the transverse plane. Microstrip lines and other planar transmission lines do not have the dielectric medium completely surrounding the conductors and therefore do not support a pure TEM wave. In this case it is found that only in the low-frequency limit does the dominant mode of propagation approach that of a TEM wave. We refer to the principal wave on these lines as a

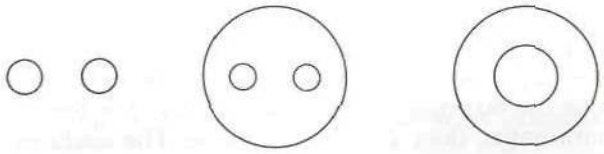


FIGURE 3.10
Cross sections of typical transmission lines. (*a*) Two-conductor line; (*b*) shielded two-conductor line; (*c*) coaxial line.

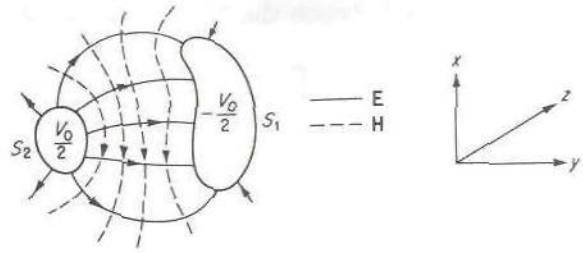


FIGURE 3.11
Cross section of a general two-conductor line showing transverse field patterns.

quasi-TEM wave. The solution for the electric and magnetic fields of the quasi-TEM wave requires a separate solution for both the electric and magnetic fields in order to determine the distributed-circuit line parameters R, L, C, and G. This is because the electric and magnetic fields are no longer related in the simple way that they are for the TEM wave. The solution for the magnetic field can be found by solving for the vector potential function as will be shown later. In this section and the following one, we consider only transmission lines that support a TEM wave.

With reference to Fig. 3.11, let the one conductor be at a potential $V_0/2$ and the other conductor at $-V_0/2$. To determine the field of a TEM wave, a suitable potential $\Phi(x, y)$ must be found first. It is necessary that Φ be a solution of

$$\nabla_t^2 \Phi = 0$$

and satisfy the boundary conditions

$$\Phi = \begin{cases} \dfrac{V_0}{2} & \text{on } S_2 \\ -\dfrac{V_0}{2} & \text{on } S_1 \end{cases}$$

Since Φ is unique only to within an additive constant, we could equally well choose $\Phi = V_0$ on S_2 and $\Phi = 0$ on S_1. If a solution for Φ is possible, a TEM mode or field solution is also possible. When two or more conductors are present, this is always the case. The solution for Φ is an electrostatic problem that can be solved when the line configuration is simple enough, as exemplified in Fig. 3.10.

The fields are given by (3.63), and for propagation in the $+z$ direction are

$$\mathbf{E} = \mathbf{E}_t = \mathbf{e} e^{-jk_0 z} = -\nabla_t \Phi e^{-jk_0 z} \tag{3.77a}$$

$$\mathbf{H} = \mathbf{H}_t = Y_0 \mathbf{a}_z \times \mathbf{e} e^{-jk_0 z} \tag{3.77b}$$

The line integral of **e** between the two conductors is

$$\int_{S_1}^{S_2} \mathbf{e} \cdot d\mathbf{l} = \int_{S_1}^{S_2} -\nabla_t \Phi \cdot d\mathbf{l}$$

$$= -\int_{S_1}^{S_2} \frac{d\Phi}{dl}\, dl = -[\Phi(S_2) - \Phi(S_1)] = -V_0$$

Associated with the electric field is a unique voltage wave

$$V = V_0 e^{-jk_0 z} \tag{3.78}$$

since the line integral of **e** between S_1 and S_2 is independent of the path chosen because **e** is the gradient of a scalar potential.

The line integral of **h** around one conductor, say S_2, gives

$$\oint_{S_2} \mathbf{h} \cdot d\mathbf{l} = \oint_{S_2} J_s\, dl = I_0$$

by application of Ampère's law, $\nabla \times \mathbf{H} = j\omega\mathbf{D} + \mathbf{J}$, and noting that there is no axial displacement flux D_z for a TEM mode. On the conductors the boundary conditions require $\mathbf{n} \times \mathbf{e} = 0$ and $\mathbf{n} \times \mathbf{h} = \mathbf{J}_s$, where **n** is a unit outward normal and $\mathbf{J}_s$ is the surface current density. Since **n** and **h** lie in a transverse plane, the current $\mathbf{J}_s$ is in the axial direction. In the region remote from the conductors, $\nabla_t \times \mathbf{h} = 0$, but the line integral around a conductor is not zero because of the current that exists. The current on the two conductors is oppositely directed, as may be verified from the expression $\mathbf{n} \times \mathbf{h} = \mathbf{J}_s$. Associated with the magnetic field there is a unique current wave

$$I = I_0 e^{-jk_0 z} \tag{3.79}$$

Since the potential Φ is independent of frequency, it follows that the transverse fields **e** and **h** are also independent of frequency and are, in actual fact, the static field distributions which exist between the conductors if the potential difference is V_0 and currents $I_0, -I_0$ exist on S_2, S_1, respectively. The magnetic lines of flux coincide with the equipotential lines, since **e** and **h** are orthogonal, as seen from (3.77*b*).

Example 3.1 Coaxial line. Figure 3.12 illustrates a coaxial transmission line for which the solution for a TEM mode will be constructed. In cylindrical coordinates r, ϕ, z, the two-dimensional Laplace equation is

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \Phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \Phi}{\partial \phi^2} = 0$$

or for a potential function independent of the angular coordinate ϕ,

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \Phi}{\partial r}\right) = 0$$

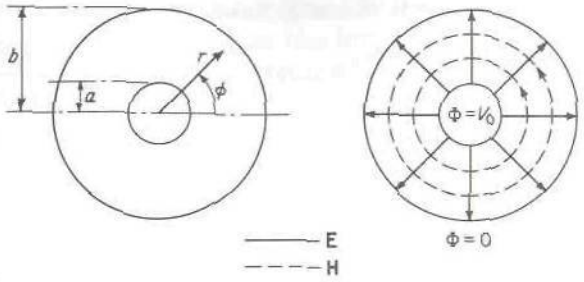


FIGURE 3.12
Coaxial transmission line.

Integrating this equation twice gives

$$\Phi = C_1 \ln r + C_2$$

Imposing the boundary conditions $\Phi = V_0$ at $r = a$, $\Phi = 0$ at $r = b$, gives

$$V_0 = C_1 \ln a + C_2 \qquad 0 = C_1 \ln b + C_2$$

and hence $C_2 = -C_1 \ln b$, $C_1 = V_0/[\ln(a/b)]$,

$$\Phi = V_0 \frac{\ln(r/b)}{\ln(a/b)} \tag{3.80}$$

The electric and magnetic fields of a TEM mode propagating in the $+z$ direction are given by (3.77) and are

$$\mathbf{E} = -\mathbf{a}_r \frac{\partial \Phi}{\partial r} e^{-jk_0 z} = -\frac{V_0}{\ln(a/b)} \frac{\mathbf{a}_r}{r} e^{-jk_0 z}$$

$$= \frac{V_0}{\ln(b/a)} \frac{\mathbf{a}_r}{r} e^{-jk_0 z} \tag{3.81a}$$

$$\mathbf{H} = Y_0 \mathbf{a}_z \times \mathbf{e} e^{-jk_0 z} = \frac{Y_0 V_0}{\ln(b/a)} \frac{\mathbf{a}_\phi}{r} e^{-jk_0 z} \tag{3.81b}$$

The potential difference between the two conductors is obviously V_0; so the voltage wave associated with the electric field is

$$V = V_0 e^{-jk_0 z} \tag{3.82}$$

The current density on the inner conductor is

$$\mathbf{J}_s = \mathbf{n} \times \mathbf{H} = \mathbf{a}_r \times \mathbf{H} = \frac{Y_0 V_0}{\ln(b/a)} \frac{\mathbf{a}_z}{a} e^{-jk_0 z}$$

The total current, apart from the factor $e^{-jk_0 z}$, is

$$I_0 = \frac{Y_0 V_0}{a \ln(b/a)} \int_0^{2\pi} a \, d\phi = \frac{Y_0 V_0 2\pi}{\ln(b/a)} \tag{3.83}$$

The current on the inner surface of the outer conductor is readily shown to be equal to I_0 also, but directed in the $-z$ direction. The current wave associated with the magnetic field is therefore

$$I = I_0 e^{-jk_0 z} \tag{3.84}$$

The power, or rate of energy flow, along the line is given by

$$P = \frac{1}{2}\,\mathrm{Re}\int_a^b\int_0^{2\pi}\mathbf{E}\times\mathbf{H}^*\cdot\mathbf{a}_z r\,dr\,d\phi = \frac{1}{2}\frac{Y_0V_0^2}{[\ln(b/a)]^2}\int_a^b\int_0^{2\pi}\frac{d\phi\,dr}{r}$$

$$= \frac{\pi Y_0 V_0^2}{\ln(b/a)} \tag{3.85}$$

The power transmitted is seen to be also given, as anticipated, by the expression

$$\frac{1}{2}\,\mathrm{Re}(VI^*) = \frac{1}{2}V_0I_0 = \frac{1}{2}V_0^2\frac{2\pi Y_0}{\ln(b/a)}$$

The characteristic impedance of the line is defined by the ratio

$$\frac{V_0}{I_0} = Z_c \tag{3.86}$$

in terms of which the power may be expressed as $P = \frac{1}{2}Z_cI_0^2 = \frac{1}{2}Y_cV_0^2$, where Y_c is the characteristic admittance of the line and equal to Z_c^{-1}. The characteristic impedance is a function of the cross-sectional shape of the transmission line.

Transmission Line with Small Losses

Practical transmission lines always have some loss caused by the finite conductivity of the conductors and also loss that may be present in the dielectric material surrounding the conductors. Consider first the case when the conductors are surrounded by a dielectric with permittivity $\epsilon = \epsilon' - j\epsilon''$ but the conductors are still considered to be perfect. The presence of a lossy dielectric does not affect the solution for the scalar potential Φ. Consequently, the field solution is formally the same as for the ideal line, except that k_0 and Y_0 are replaced by $k = k_0(\epsilon_r' - j\epsilon_r'')^{1/2}$ and $Y = Y_0(\epsilon_r' - j\epsilon_r'')^{1/2}$, where the dielectric constant $\epsilon_r = \epsilon_r' - j\epsilon_r'' = \epsilon/\epsilon_0$. For small losses such that $\epsilon_r'' \ll \epsilon_r'$, the propagation constant is

$$jk = \alpha + j\beta = j(\epsilon_r')^{1/2}k_0\left(1 - j\frac{\epsilon_r''}{\epsilon_r'}\right)^{1/2} \approx j(\epsilon_r')^{1/2}k_0 + \frac{\epsilon_r''k_0}{2(\epsilon_r')^{1/2}}$$

Thus

$$\alpha = \frac{\epsilon_r''k_0}{2(\epsilon_r')^{1/2}} \tag{3.87a}$$

$$\beta = (\epsilon_r')^{1/2}k_0 \tag{3.87b}$$

where α is the attenuation constant and β is the phase constant. The wave consequently attenuates according to $e^{-\alpha z}$ as it propagates in the $+z$ direction.

It will be instructive to derive the above expression for α by means of a perturbation method that is widely used in the evaluation of the attenua-

tion, or damping, factor for a low-loss physical system. This method is based on the assumption that the introduction of a small loss does not substantially perturb the field from its loss-free value. The known field distribution for the loss-free case is then used to evaluate the loss in the system, and from this the attenuation constant can be calculated. In the present case, if $\epsilon_r'' = 0$, the loss-free solution is

$$\mathbf{E} = -\nabla_t \Phi e^{-jkz} \qquad \mathbf{H} = Y\mathbf{a}_z \times \mathbf{E}$$

where $k = (\epsilon_r')^{1/2} k_0$ and $Y = (\epsilon_r')^{1/2} Y_0$. When ϵ_r'' is small but not zero, the imaginary part of ϵ, that is, ϵ'', is equivalent to a conductivity

$$\sigma = \omega\epsilon'' = \omega\epsilon_0\epsilon_r''$$

A conductivity σ results in a shunt current $\mathbf{J} = \sigma\mathbf{E}$ between the two conductors. The power loss per unit length of line is

$$P_l = \frac{1}{2\sigma}\int_S \mathbf{J}\cdot\mathbf{J}^*\,dS = \frac{\omega\epsilon''}{2}\int_S \mathbf{E}\cdot\mathbf{E}^*\,dS \tag{3.88}$$

where the integration is over the cross section of the line, and the loss-free solution for $\mathbf{E}$ is used to carry out the evaluation of P_l. Since loss is present, the power propagated along the line must decrease according to a factor $e^{-2\alpha z}$. The rate of decrease of power propagated along the line equals the power loss. If the power at $z = 0$ is P_0, then at z it is $P = P_0 e^{-2\alpha z}$. Consequently,

$$-\frac{\partial P}{\partial z} = P_l = 2\alpha P_0 e^{-2\alpha z} = 2\alpha P \tag{3.89}$$

which states that the power loss at any plane z is directly proportional to the total power P present at this plane. The power propagated along the line is given by

$$\begin{aligned} P &= \frac{1}{2}\,\mathrm{Re}\int_S \mathbf{E}\times\mathbf{H}^*\cdot\mathbf{a}_z\,dS \\ &= \frac{Y}{2}\,\mathrm{Re}\int_S \mathbf{E}\times(\mathbf{a}_z\times\mathbf{E}^*)\cdot\mathbf{a}_z\,dS = \frac{Y}{2}\int_S \mathbf{E}\cdot\mathbf{E}^*\,dS \end{aligned}$$

Hence the attenuation α is given by

$$\alpha = \frac{P_l}{2P} = \frac{\sigma}{2Y} = \frac{\omega\epsilon''}{2Y_0(\epsilon_r')^{1/2}} = k_0\frac{\epsilon_r''}{2(\epsilon_r')^{1/2}}$$

which is the same as the expression (3.87a). For this example the perturbation method does not offer any advantage. However, often the field solution for the lossy case is very difficult to find, in which case the perturbation method is extremely useful and simple to carry out by comparison with other methods. The case of transmission lines with conductors having finite conductivity is an important example of this, and is discussed below.

If the conductors of a transmission line have a finite conductivity, they exhibit a surface impedance

$$Z_m = \frac{1+j}{\sigma\delta_s} \tag{3.90}$$

where $\delta_s = (2/\omega\mu\sigma)^{1/2}$ is the skin depth (Sec. 2.9). At the surface the electric field must have a tangential component equal to $Z_m\mathbf{J}_s$, where $\mathbf{J}_s$ is the surface current density. Therefore it is apparent that an axial component of electric field must be present, and consequently the field is no longer that of a TEM wave. The axial component of electric field gives rise to a component of the Poynting vector directed into the conductor, and this accounts for the power loss in the conductor. Generally, it is very difficult to find the exact solution for the fields when the conductors have finite conductivity. However, since $|Z_m|$ is very small compared with Z_0, the axial component of electric field is also very small relative to the transverse components. Thus the field is very nearly that of the TEM mode in the loss-free case. The perturbation method outlined earlier may be used to evaluate the attenuation caused by finite conductivity.

The current density $\mathbf{J}_s$ is taken equal to $\mathbf{n} \times \mathbf{H}$, where $\mathbf{n}$ is the unit outward normal to the conductor surface and $\mathbf{H}$ is the *loss-free* magnetic field. The power loss in the surface impedance per unit length of line is

$$\begin{aligned} P_l &= \frac{1}{2}\,\mathrm{Re}\, Z_m \oint_{S_1+S_2} \mathbf{J}_s \cdot \mathbf{J}_s^*\, dl \\ &= \frac{R_m}{2}\oint_{S_1+S_2} (\mathbf{n}\times\mathbf{H})\cdot(\mathbf{n}\times\mathbf{H}^*)\, dl \\ &= \frac{R_m}{2}\oint_{S_1+S_2} \mathbf{H}\cdot\mathbf{H}^*\, dl \end{aligned} \tag{3.91}$$

where $R_m = 1/\sigma\delta_s$ is the high-frequency surface resistance, and

$$\begin{aligned} (\mathbf{n}\times\mathbf{H})\cdot(\mathbf{n}\times\mathbf{H}^*) &= \mathbf{n}\cdot\mathbf{H}\times(\mathbf{n}\times\mathbf{H}^*) \\ &= \mathbf{n}\cdot[(\mathbf{H}\cdot\mathbf{H}^*)\mathbf{n} - (\mathbf{H}\cdot\mathbf{n})\mathbf{H}^*] = \mathbf{H}\cdot\mathbf{H}^* \end{aligned}$$

since $\mathbf{n}\cdot\mathbf{H} = 0$ for the infinite-conductivity case. The integration is taken around the periphery $S_1 + S_2$ of the two conductors. The attenuation constant arising from conductor loss is thus

$$\alpha = \frac{P_l}{2P} = \frac{R_m \oint_{S_1+S_2} \mathbf{H}\cdot\mathbf{H}^*\, dl}{2Z\int \mathbf{H}\cdot\mathbf{H}^*\, dS} \tag{3.92}$$

where the power propagated along the line is given by

$$\mathrm{Re}\,\frac{1}{2}\int \mathbf{E}\times\mathbf{H}^*\cdot\mathbf{a}_z\, dS = \frac{1}{2}Z\int \mathbf{H}\cdot\mathbf{H}^*\, dS$$

and Z is the intrinsic impedance of the medium; that is, $Z = (\mu/\epsilon)^{1/2}$.

When both dielectric and conductor losses are present, the attenuation constant is the sum of the attenuation constants arising from each cause, provided both attenuation constants are small.

Example 3.2 Lossy coaxial line. Let the coaxial line in Fig. 3.12 be filled with a lossy dielectric ($\epsilon = \epsilon' - j\epsilon''$), and let the conductors have finite conductivity σ. For the loss-free case ($\epsilon'' = 0$, $\sigma = \infty$) the fields are given by (3.81), with k_0 and Y_0 replaced by $k = (\epsilon'/\epsilon_0)^{1/2}k_0$,

$$Y = \left(\frac{\epsilon'}{\epsilon_0}\right)^{1/2} Y_0$$

Thus

$$\mathbf{E} = \frac{V_0}{\ln(b/a)} \frac{\mathbf{a}_r}{r} e^{-jkz} \tag{3.93a}$$

$$\mathbf{H} = \frac{YV_0}{\ln(b/a)} \frac{\mathbf{a}_\phi}{r} e^{-jkz} \tag{3.93b}$$

The power propagated along the line is

$$P = \frac{1}{2} \operatorname{Re} \int_0^{2\pi} \int_a^b \mathbf{E} \times \mathbf{H}^* \cdot \mathbf{a}_z r \, d\phi \, dr = \frac{\pi Y V_0^2}{\ln(b/a)} \tag{3.94}$$

The power loss P_{l1} from the lossy dielectric is, from (3.88),

$$P_{l1} = \frac{\omega \epsilon''}{2} \int_a^b \mathbf{E} \cdot \mathbf{E}^* r \, d\phi \, dr = \frac{\omega \epsilon'' V_0^2 \pi}{\ln(b/a)} \tag{3.95a}$$

The power loss from finite conductivity is given by (3.91), and is

$$\begin{aligned} P_{l2} &= \frac{R_m}{2} \frac{Y^2 V_0^2}{[\ln(b/a)]^2} \int_0^{2\pi} \left(\frac{1}{a} + \frac{1}{b}\right) d\phi \\ &= \frac{R_m \pi Y^2 V_0^2}{[\ln(b/a)]^2} \frac{b+a}{ab} \end{aligned} \tag{3.95b}$$

Hence the attenuation constant α for the coaxial line is given by

$$\begin{aligned} \alpha &= \frac{P_{l1} + P_{l2}}{2P} = \frac{\omega \epsilon''}{2Y} + \frac{R_m Y}{2 \ln(b/a)} \frac{b+a}{ab} \\ &= \frac{k_0 \epsilon_r''}{2(\epsilon_r')^{1/2}} + \frac{R_m}{2Z \ln(b/a)} \frac{b+a}{ab} \end{aligned} \tag{3.96}$$

For the lossy case the propagation constant is consequently taken as

$$\alpha + j\beta = \alpha + jk$$

with α given by (3.96).

3.9 TRANSMISSION-LINE PARAMETERS

In this section the field analysis to determine the circuit parameters L, R, C, and G for a transmission line is examined in greater detail. This will serve further to correlate the field analysis and circuit analysis of transmission lines.

Consider first the case of a loss-free line such as that illustrated in Fig. 3.11. When the scalar potential Φ has been determined, the charge density on the conductors may be found from the normal component of electric field at the surface; that is, $\rho_s = \epsilon \mathbf{n} \cdot \mathbf{e} = -\epsilon \mathbf{n} \cdot \nabla\Phi = -\epsilon\, \partial\Phi/\partial n$, where ϵ is the permittivity of the medium surrounding the conductors. The total charge Q per unit length on conductor S_2 is

$$Q = \oint_{S_2} \epsilon \mathbf{n} \cdot \mathbf{e}\, dl$$

The total charge on the conductor S_1 is $-Q$ per meter. The potential of S_2 is V_0, and hence the capacitance C per unit length is

$$C = \frac{Q}{V_0} = \frac{\epsilon \int_{S_2} \mathbf{n} \cdot \mathbf{e}\, dl}{\int_{S_2}^{S_1} \mathbf{e} \cdot d\mathbf{l}} \tag{3.97}$$

The total current on S_2 is

$$I_0 = \oint_{S_2} \mathbf{h} \cdot d\mathbf{l} = \oint_{S_2} Y\mathbf{n} \cdot \mathbf{e}\, dl = \frac{YQ}{\epsilon}$$

since $|\mathbf{h}| = Y|\mathbf{e}| = Y\mathbf{n} \cdot \mathbf{e}$ at the surface of S_2 because the normal component of $\mathbf{h}$ and the tangential component of $\mathbf{e}$ are zero at the perfectly conducting surface S_2. The characteristic impedance of the line is given by

$$Z_c = \frac{V_0}{I_0} = \frac{V_0 \epsilon}{YQ} = \frac{\epsilon Z}{C} \tag{3.98}$$

A knowledge of the capacitance per unit length suffices to determine the *characteristic impedance.*

To determine the inductance L per unit length, refer to Fig. 3.13, which illustrates the magnetic flux lines around the conductors. Since $\mathbf{h}$ is

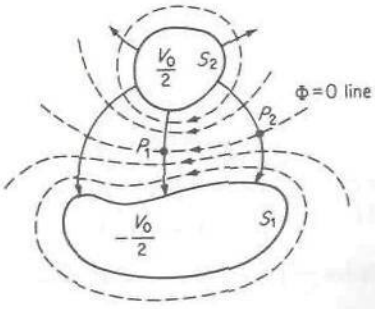


FIGURE 3.13
Magnetic flux lines in a transmission line.

orthogonal to **e**, these coincide with the equipotential lines. All the flux lines from the $\Phi = 0$ to the $\Phi = V_0/2$ line link the current on S_2. The flux linkage is the total flux cutting any path joining the $\Phi = 0$ line to the surface S_2. If a path such as P_1S_2 or P_2S_2 is chosen, which is orthogonal to the flux lines, this path coincides with a line of electric force. The flux cutting such a path is

$$\psi = \int_{P_1}^{S_2} \mu h \, dl = \mu Y \int_{P_1}^{S_2} - \mathbf{e} \cdot d\mathbf{l} = \mu Y \frac{V_0}{2}$$

since $|\mathbf{h}| = Y|\mathbf{e}|$ for a TEM wave. The inductance of one conductor of the line is

$$L_1 = \frac{\psi}{I_0} = \mu Y \frac{V_0}{2I_0}$$

The inductance of both conductors per unit length is twice this value; so

$$L = \mu Y \frac{V_0}{I_0} = \mu Y Z_c \tag{3.99}$$

From this relation and (3.98) it is seen that $Z = \mu Z_c/L = CZ_c/\epsilon$, and hence

$$Z^2 = \frac{\mu}{\epsilon} = \frac{\mu Z_c}{L} \frac{CZ_c}{\epsilon}$$

which gives

$$Z_c = \sqrt{\frac{L}{C}} \tag{3.100}$$

Equations (3.98) and (3.99) also show that

$$\mu\epsilon = LC \tag{3.101}$$

for a transmission line. The above expressions for C and L can also be obtained from the definitions based on stored energy. The derivation is left as a problem.

If the dielectric has a complex permittivity $\epsilon = \epsilon' - j\epsilon''$, *where* ϵ'' includes the conductivity of the dielectric if it is not zero, the total shunt current consists of a displacement current I_D and a conduction current I_S. The current leaving conductor S_2 per unit length is

$$I = I_D + I_S = j\omega\epsilon \oint_{S_2} \mathbf{e} \cdot \mathbf{n} \, dl = j\omega\epsilon' \oint_{S_2} \mathbf{e} \cdot \mathbf{n} \, dl + \omega\epsilon'' \oint_{S_2} \mathbf{e} \cdot \mathbf{n} \, dl$$

where the first integral on the right gives the displacement current and the second integral gives the conduction current. The total shunt admittance is given by $Y = j\omega C + G = (I_S + I_D)/V_0$, and hence it is seen that

$$G = \frac{I_S}{V_0} = \frac{I_S}{I_D} \frac{I_D}{V_0} = \frac{\omega\epsilon''}{\epsilon'} C \tag{3.102}$$

since $j\omega C = I_D/V_0$ and $j\omega C/j\omega\epsilon' = C/\epsilon'$. This relation shows that G differs from C by the factor $\omega\epsilon''/\epsilon'$ only.

The transmission-line loss from finite conductivity may be accounted for by a series resistance R per unit length provided R is chosen so that

$$\frac{1}{2}RI_0^2 = \frac{R_m}{2}\oint_{S_1+S_2} |\mathbf{h}|^2\, dl \tag{3.103}$$

The right-hand side gives the total power loss per unit length arising from the high-frequency resistance of the conductors. In terms of this quantity, the resistance R is thus defined as

$$R = R_m \frac{\oint_{S_1+S_2} |\mathbf{h}|^2\, dl}{\left(\oint_{S_2} |\mathbf{h}|\, dl\right)^2} \tag{3.104}$$

where $R_m = 1/\sigma\delta_s$ and δ_s is the skin depth.

A further effect of the finite conductivity is to increase the series inductance of the line by a small amount because of the penetration of the magnetic field into the conductor. This skin-effect inductance L_s is readily evaluated on an energy basis. The surface impedance Z_m has an inductive part $jX_m = j/\sigma\delta_s$ equal in magnitude to R_m. The magnetic energy stored in X_m is (note that X_m is equivalent to a surface inductance $X_m/\omega = L_m$)

$$W_m = \frac{X_m}{4\omega}\oint_{S_1+S_2} |\mathbf{J}_s|^2\, dl$$

$$= \frac{X_m}{4\omega}\oint_{S_1+S_2} |\mathbf{h}|^2\, dl = \frac{X_m}{4\omega}\frac{RI_0^2}{R_m} = \frac{RI_0^2}{4\omega}$$

by using (3.103) to replace the integral. Defining L_s by the relation

$$\tfrac{1}{4}L_s I_0^2 = W_m$$

gives

$$\omega L_s = R \tag{3.105}$$

The series inductive reactance of the line is increased by an amount equal to the series resistance. However, for low-loss lines, $R \ll \omega L$, so that $L_s \ll L$, and the correction is not significant for most practical lines. The inductance L_s is called the internal inductance since it arises from flux linkage internal to the conductor surfaces.

It should not come as a surprise to find that $\omega L_s = R$ since both the inductive reactance and resistance arise from the penetration of the current and fields into the conductor. The effect of this penetration into the conductor by an effective distance equal to the skin depth δ_s is correctly accounted for in a simplified manner by introduction of the surface impedance $Z_m = (1 + j)/\sigma\delta_s$.

Example 3.3 Coaxial-line parameters. For the coaxial line of Fig. 3.12 the potential Φ is given by

$$\Phi = V_0 \frac{\ln(r/b)}{\ln(a/b)}$$

The charge on the inner conductor is

$$Q = \epsilon \int_0^{2\pi} \mathbf{a}_r \cdot \mathbf{e} a \, d\phi = \epsilon \int_0^{2\pi} -\frac{\partial \Phi}{\partial r} a \, d\phi$$

$$= \frac{-\epsilon V_0}{\ln(a/b)} \int_0^{2\pi} d\phi = \frac{2\pi\epsilon V_0}{\ln(b/a)}$$

Hence the capacitance per unit length is

$$C = \frac{\epsilon'}{\epsilon} \frac{Q}{V_0} = \frac{2\pi\epsilon'}{\ln(b/a)} \tag{3.106}$$

since the capacitance arises only from that part of the charge associated with ϵ' whereas ϵ'' gives rise to the shunt conductance.

The magnetic field is given by (3.93*b*) as

$$\mathbf{H} = \mathbf{h} e^{-jk_0 z} = \frac{YV_0}{\ln(b/a)} \frac{\mathbf{a}_\phi}{r} e^{-jk_0 z}$$

The current I_0 is

$$I_0 = \int_0^{2\pi} \mathbf{h} \cdot \mathbf{a}_\phi a \, d\phi = \frac{2\pi Y V_0}{\ln(b/a)}$$

Thus the characteristic impedance is

$$Z_c = \frac{V_0}{I_0} = \frac{Z \ln(b/a)}{2\pi} \tag{3.107}$$

The flux linking the center conductor is

$$\psi = \mu \int_a^b \mathbf{h} \cdot \mathbf{a}_\phi \, dr = \frac{\mu Y V_0}{\ln(b/a)} \int_a^b \frac{dr}{r} = \mu Y V_0$$

Consequently, the inductance per unit length is

$$L = \frac{\psi}{I_0} = \frac{\mu Y V_0}{2\pi Y V_0} \ln \frac{b}{a} = \frac{\mu}{2\pi} \ln \frac{b}{a} \tag{3.108}$$

from which it is seen that $LC = \mu\epsilon'$ and $Z_c = (L/C)^{1/2}$.

The shunt conductance G is given by $\omega\epsilon'' C/\epsilon'$, and is

$$G = \frac{\omega\epsilon''}{\epsilon'} \frac{2\pi\epsilon'}{\ln(b/a)} = \frac{2\pi\omega\epsilon''}{\ln(b/a)} \tag{3.109}$$

To find the series resistance, the power loss in the inner and outer conductors must be evaluated. This was done in Example 3.2, with the result

[Eq. (3.95*b*)]

$$\frac{1}{2}RI_0^2 = P_{l2} = \frac{R_m \pi Y^2 V_0^2}{[\ln(b/a)]^2}\frac{b+a}{ab}$$

Solving for R gives

$$R = \frac{R_m}{2\pi}\frac{b+a}{ab} \tag{3.110}$$

The internal inductance L_s is equal to R/ω; so the total series line inductance per unit length is

$$L + L_s = \frac{\mu}{2\pi}\ln\frac{b}{a} + \frac{b+a}{2\pi\omega ab\delta_s\sigma} \tag{3.111}$$

The distributed-circuit parameters R, L, C, and G for a transmission line can also be determined from an evaluation of the stored electric and magnetic field energy and the power loss per unit length. Energy storage in the magnetic field is accounted for by the series inductance L per unit length, whereas energy storage in the electric field is accounted for by the distributed shunt capacitance C per unit length. Power loss in the conductors is taken into account by a series resistance R per unit length. Finally, the power loss in the dielectric may be included by introducing a shunt conductance G per unit length. Suitable definitions for the parameters L, C, R, and G based on the above concepts are

$$L = \frac{\mu}{I_0 I_0^*}\int_S \mathbf{H}\cdot\mathbf{H}^*\,dS \tag{3.112a}$$

$$C = \frac{\epsilon'}{V_0 V_0^*}\int_S \mathbf{E}\cdot\mathbf{E}^*\,dS \tag{3.112b}$$

$$R = \frac{R_m}{I_0 I_0^*}\oint_{S_1+S_2} \mathbf{H}\cdot\mathbf{H}^*\,dl \tag{3.112c}$$

$$G = \frac{\omega\epsilon''}{V_0 V_0^*}\int \mathbf{E}\cdot\mathbf{E}^*\,dS \tag{3.112d}$$

where I_0 is the total current on the line, and V_0 the potential difference. These expressions are obtained, for example, by equating the magnetic energy $\frac{1}{4}I_0 I_0^* L = W_m$ stored in the equivalent series inductance L to the expression for W_m in terms of the field. The above definitions are readily shown to be equivalent to the other commonly used definitions such as

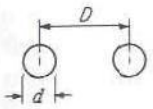
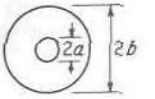
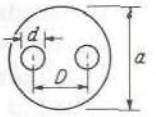
$$L = \frac{\text{magnetic flux linkage}}{\text{total current}} \tag{3.113a}$$

$$C = \frac{\text{total charge per unit length}}{\text{voltage difference between conductors}} \tag{3.113b}$$

$$G = \frac{\text{total shunt current}}{\text{voltage difference between conductors}} \tag{3.113c}$$

Parameters of some common transmission lines are given in Table 3.1.

TABLE 3.1
Parameters of common transmission lines†

	Z_c	R
	$\frac{1}{\pi}\left(\frac{\mu_0}{\epsilon'}\right)^{1/2}\cosh^{-1}\frac{D}{d}$	$\frac{2R_m}{\pi d}\frac{D/d}{\left[(D/d)^2-1\right]^{1/2}}$
	$\frac{1}{2\pi}\left(\frac{\mu_0}{\epsilon'}\right)^{1/2}\ln\frac{b}{a}$	$\frac{R_m}{2\pi}\left(\frac{1}{a}+\frac{1}{b}\right)$
	$\frac{1}{\pi}\left(\frac{\mu_0}{\epsilon'}\right)^{1/2}\left[\ln\left(2p\frac{1-q^2}{1+q^2}\right)-\frac{1+4p^2}{16p^4}(1-4q^2)\right]$	$\frac{2R_m}{\pi d}\left[1+\frac{1+2p^2}{4p^4}(1-4q^2)\right]+\frac{8R_m}{\pi a}q^2\left[(1+q^2)-\frac{1+4p^2}{8p^4}\right]$
$p=\frac{D}{d}$ $q=\frac{D}{a}$		

†For all TEM transmission lines

$$C=\frac{(\mu_0\epsilon')^{1/2}}{Z_c} \qquad L=(\mu_0\epsilon')^{1/2}Z_c \qquad G=\frac{\omega\epsilon''C}{\epsilon'}$$

$$\alpha_d=\frac{GZ_c}{2} \qquad \alpha_c=\frac{RY_c}{2} \qquad R_m=\frac{1}{\sigma\delta_s}=\left(\frac{\omega\mu}{2\sigma}\right)^{1/2}$$

3.10 INHOMOGENEOUSLY FILLED PARALLEL-PLATE TRANSMISSION LINE

In Fig. 3.14*a* we show a parallel-plate transmission line (waveguide) partially filled with dielectric material having a permittivity $\epsilon=\epsilon_r\epsilon_0$, where ϵ_r is the dielectric constant. The plates are infinitely wide and spaced a distance b apart. The dielectric sheet has a thickness a and rests on the bottom plate.

The purpose for studying this particular waveguide is that it exhibits a number of characteristics that are similar to those of the microstrip transmission lines examined in the following section. We will show that the dominant mode of propagation in the waveguide under consideration is an E mode and that as the frequency approaches zero this mode becomes a TEM mode. Furthermore, in the low-frequency limit, the propagation con-

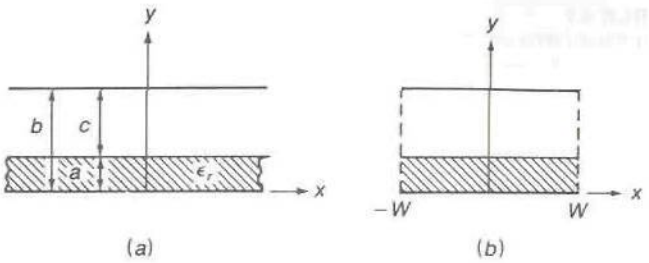


FIGURE 3.14
(*a*) Partially loaded parallel-plate waveguide; (*b*) parallel-plate waveguide with magnetic walls at $x = \pm W$.

stant can be found in terms of the distributed capacitance and inductance per meter by the usual transmission-line formula $\beta = \omega\sqrt{LC}$. As the frequency increases β increases faster than ω, in which case we say that the transmission line exhibits dispersion.

Another feature that can be easily described for this waveguide is the existence of a surface-wave mode of propagation that consists of a field concentrated near the air-dielectric interface.

Since the analytic solution for the partially filled parallel-plate waveguide is readily constructed, this waveguide serves as a useful example to provide some physical insight into the properties of microstrip transmission lines.

An electric wall is a surface on which the tangential electric field must be zero. A good conductor such as copper provides a surface with a very small skin-effect surface impedance (see Sec. 2.9). When we let the conductivity σ become infinite, we obtain an electric wall on which the boundary condition $\mathbf{n} \times \mathbf{E} = 0$ holds. The dual of an electric wall is a magnetic wall on which the tangential magnetic field is zero, i.e., the boundary condition $\mathbf{n} \times \mathbf{H} = 0$ holds. The magnetic wall does not have a physical realization but is, nevertheless, a useful theoretical concept. In practice, a magnetic wall can be inserted into a field region, without disturbing the field, along any surface on which the tangential magnetic field is zero. Such surfaces usually correspond to certain symmetry planes in a given problem. In addition to the above boundary conditions, Maxwell's equations show that on an electric wall the normal component of $\mathbf{H}$ is zero, that is, $\mathbf{n} \cdot \mathbf{H} = 0$. The dual boundary condition $\mathbf{n} \cdot \mathbf{E} = 0$ holds on a magnetic wall.

For the E mode that we will consider in the partially filled parallel-plate waveguide, we will assume that the fields do not depend on the x coordinate but are functions of y and z only. A consequence of this assumption is that only the field components E_y, E_z and H_x are present. Thus we can place a magnetic wall along any x = constant surface without disturbing the field. We will now assume that magnetic walls are inserted at $x = \pm W$ as shown

in Fig. 3.14*b*. By means of this artifice, we are able to talk about a closed waveguide structure, closed by electric walls at $y = 0, b$ and by magnetic walls at $x = \pm W$.

In order to find the solutions for E modes having the z dependence $e^{-j\beta z}$, we must find solutions for the axial electric field component $e_z(y)$ first. In an ideal transmission line the propagation constant equals that for plane TEM waves in the surrounding medium. In the structure under investigation we have a nonuniform medium, namely dielectric in the region $0 \le y \le a$ and air in the region $a \le y \le b$. Consequently, we can anticipate that the propagation phase constant β for the dominant mode will take on an intermediate value, i.e.,

$$k_0 = \omega\sqrt{\mu_0\epsilon_0} < \beta < \sqrt{\epsilon_r}k_0 = k$$

The equation satisfied by $e_z(y)$ is (3.72*a*) which is repeated below

$$\nabla_t^2 e_z + k_c^2 e_z = 0$$

Since we assume no variation with x, the transverse laplacian operator becomes simply $\partial^2/\partial y^2$. In this equation $k_c^2 = k_0^2 - \beta^2$ in the air region and equals $k^2 - \beta^2$ in the dielectric region. The propagation constant β *must be the same* in both regions because the tangential electric and magnetic fields must match at the air-dielectric interface for all values of z. For convenience, we will let $k_c = l$ in the dielectric region and let it equal p in the air region. We thus require that $p^2 - k_0^2 = l^2 - k^2$ or

$$l^2 - p^2 = (\epsilon_r - 1)k_0^2 \tag{3.114}$$

In the two regions the axial electric field is thus a solution of

$$\frac{d^2e_z}{dy^2} + l^2 e_z = 0 \qquad 0 \le y \le a \tag{3.115a}$$

$$\frac{d^2e_z}{dy^2} + p^2 e_z = 0 \qquad a \le y \le b \tag{3.115b}$$

along with the boundary conditions

$$e_z(y) = 0 \qquad y = 0, b \tag{3.116a}$$

$$e_z(y) \text{ continuous at } y = a \tag{3.116b}$$

$$\frac{\epsilon_r}{l^2}\frac{\partial e_z}{\partial y}\bigg|_a = \frac{1}{p^2}\frac{\partial e_z}{\partial y}\bigg|_a \tag{3.116c}$$

The third boundary condition comes from the requirement that H_x be continuous across the air-dielectric interface. The transverse fields are given

by (3.72*b*) and (3.72*c*). The generic form of the equations is

$$E_y = -\frac{j\beta}{k_c^2}\frac{\partial e_z}{\partial y}e^{-j\beta z}$$

$$H_x = Y_e\frac{j\beta}{k_c^2}\frac{\partial e_z}{\partial y}e^{-j\beta z}$$

upon using $\mathbf{a}_z \times \mathbf{a}_y = -\mathbf{a}_x$. In the equation for H_x the wave admittance Y_e is given by kY/β in the dielectric and by $(k_0 Y_0)/\beta$ in the air region where $kY = \omega\sqrt{\mu_0\epsilon}\sqrt{\epsilon/\mu_0} = \epsilon_r k_0 Y_0$. Thus we have

$$e_y(y) = \begin{cases} -\dfrac{j\beta}{l^2}\dfrac{\partial e_z}{\partial y} & \text{dielectric region} \\ -\dfrac{j\beta}{p^2}\dfrac{\partial e_z}{\partial y} & \text{air region} \end{cases} \tag{3.117a}$$

$$h_x(y) = \begin{cases} \dfrac{j\epsilon_r k_0 Y_0}{l^2}\dfrac{\partial e_z}{\partial y} & \text{dielectric region} \\ \dfrac{jk_0 Y_0}{p^2}\dfrac{\partial e_z}{\partial y} & \text{air region} \end{cases} \tag{3.117b}$$

An examination of the expression for $h_x(y)$ shows that continuity of h_x at $y = a$ gives the boundary condition specified by (3.116*c*). We also find that in an inhomogeneously filled waveguide, the wave impedance, defined by the ratio $-E_y/H_x$, is not constant since it has a different value in the air region from that in the dielectric region.

The reader can readily verify that the solutions of (3.115*a*) and (3.115*b*) that satisfy the boundary conditions at $y = 0$ and b are

$$e_z(y) = C_1 \sin ly \qquad 0 \le y \le a$$

$$e_z(y) = C_2 \sin p(b - y) \qquad a \le y \le b$$

where C_1 and C_2 are unknown amplitude constants. The boundary condition (3.116*b*) requires that

$$C_1 \sin la = C_2 \sin pc$$

where $c = b - a$. The last boundary condition (3.116*c*) requires

$$\frac{\epsilon_r}{l}C_1 \cos la = -\frac{1}{p}C_2 \cos pc$$

When we divide the first equation by the second one, we obtain

$$l \tan la = -\epsilon_r p \tan pc \tag{3.118}$$

This transcendental equation must be solved simultaneously with (3.114) to determine the allowed values of l and p. There will be an infinite number of solutions; consequently an infinite number of E modes are possible. Since β is given by

$$\beta = \sqrt{k_0^2 - p^2} = \sqrt{k^2 - l^2} \tag{3.119}$$

most of the modes will be nonpropagating since increasing values of p and l give $p > k_0$ which makes β imaginary. When β is imaginary the z dependence is of the form $e^{-|\beta|z}$ and the field decays exponentially from the point at which it is excited. These nonpropagating modes are called evanescent modes.

We note from (3.119) that a value of β between k_0 and k can occur only if p is imaginary. Thus we must consider the possibility that an imaginary p, say $p = jp_0$, is a possible solution to (3.118). If we let l_0 be the corresponding value of l, then our relevant equations become

$$l_0 \tan l_0 a = \epsilon_r p_0 \tanh p_0 c \tag{3.120a}$$

$$l_0^2 + p_0^2 = (\epsilon_r - 1) k_0^2 \tag{3.120b}$$

We consider solutions of these equations and the corresponding fields in the low- and high-frequency limits in the next two subsections.

Low-Frequency Solution

When the frequency is very low, k_0^2 is a very small number (at 1 MHz, k_0 equals 0.02094 rad/m); hence l_0 and p_0 are then also small. We will assume that b is at most a few centimeters, then $l_0 a$ and $p_0 c$ are also small and we can replace the tangent function and the hyperbolic tangent function by their arguments. Thus (3.120a) becomes

$$l_0^2 a = \epsilon_r p_0^2 c$$

Upon using (3.120b) we readily find that

$$(\epsilon_r - 1) k_0^2 - p_0^2 = \frac{\epsilon_r p_0^2 c}{a}$$

or

$$p_0^2 = \frac{(\epsilon_r - 1) k_0^2 a}{a + \epsilon_r c}$$

The solution for β in the low-frequency limit is thus

$$\beta = \sqrt{k_0^2 + p_0^2} = \sqrt{\frac{\epsilon_r b}{a + \epsilon_r c}}\, k_0 = \sqrt{\epsilon_e}\, k_0 \tag{3.121}$$

where ϵ_e, given by this equation, is called the effective dielectric constant.

We will now show that this equals $\omega\sqrt{LC}$, where L and C are the static distributed inductance and capacitance per meter for the given structure.

If we have a uniform current density J_z on the inner surface of the upper plate and $-J_z$ on the inner surface of the lower plate, the magnetic field between the plates will be given by $H_x = J_z$. The time-average stored magnetic energy per unit length is given by

$$W_m = \frac{\mu_0}{4}\int_0^b \int_{-W}^{W} H_x^2 \, dx \, dy = \frac{\mu_0}{2} WbJ_x^2$$

We equate this to $\frac{1}{4}LI_z^2$ where the total current $I_z = 2WJ_z$ and then find that

$$L = \frac{\mu_0 b}{2W} \tag{3.122}$$

The distributed capacitance C per meter is found by considering the capacitance of the dielectric and air regions as represented by two equivalent parallel-plate capacitances C_d and C_a in series where

$$C_d = \frac{\epsilon_r \epsilon_0 2W}{a} \qquad C_a = \frac{\epsilon_0 2W}{c}$$

The capacitance C_d is that of a parallel-plate capacitor of width $2W$, unit length, plate spacing a, and filled with dielectric. C_a is the capacitance of the air-filled section which has a spacing c.

The series capacitance is given by

$$C = \frac{C_a C_d}{C_a + C_d} = \frac{2W\epsilon_r \epsilon_0}{\epsilon_r c + a} \tag{3.123}$$

The product $LC = \epsilon_r \epsilon_0 \mu_0 b/(\epsilon_r c + a)$ which gives the solution for $\beta = \omega\sqrt{LC}$ equal to that in (3.121).

The expressions for the fields can be written down in simplified form using the small argument approximations and the relationship $C_2 = C_1 \sin l_0 a / j \sinh p_0 c \approx -jC_1 l_0 a / p_0 c$ obtained from the boundary condition requiring continuity of e_z at $y = a$. We readily find that in the region $0 \le y \le a$,

$$e_z = C_1 l_0 y \tag{3.124a}$$

$$e_y = -\frac{j\beta}{l_0} C_1 = -jC_1 \sqrt{\frac{b}{(\epsilon_r - 1)c}} \tag{3.124b}$$

$$h_x = \frac{j\epsilon_r k_0 Y_0}{l_0} C_1 = jY_0 C_1 \sqrt{\frac{\epsilon_r(\epsilon_r c + a)}{(\epsilon_r - 1)c}} \tag{3.124c}$$

and in the air region

$$e_z = C_1 \frac{l_0 a}{c}(b - y) \tag{3.124d}$$

$$e_y = -\frac{j\beta l_0 a}{p_0^2 c} C_1 = -\frac{j\beta}{l_0}\epsilon_r C_1 = -jC_1\epsilon_r\sqrt{\frac{b}{(\epsilon_r - 1)c}} \tag{3.124e}$$

$$h_x = \frac{jk_0 Y_0}{l_0} C_1 = jY_0 C_1\sqrt{\frac{\epsilon_r(\epsilon_r c + a)}{(\epsilon_r - 1)c}} \tag{3.124f}$$

We note that in the low-frequency limit e_z *vanishes as* k_0, *and hence* l_0, approach zero, while e_y and h_x remain constant. If we define the voltage V between the upper and lower plates by the line integral of e_y, then

$$V = -\int_0^b e_y\, dy = jC_1\sqrt{\frac{b}{(\epsilon_r - 1)c}}\,(a + \epsilon_r c)$$

The total z-directed current on the upper plate is $I_z = 2WJ_z = 2WH_x$, and hence the characteristic impedance is given by

$$Z_c = \frac{V}{I_z} = \frac{Z_0}{2W}\sqrt{\frac{(\epsilon_r c + a)b}{\epsilon_r}} = \sqrt{\frac{L}{C}} \tag{3.125}$$

Thus we find that in the low-frequency limit the dominant mode of propagation in the partially filled parallel-plate waveguide becomes a TEM mode and the waveguide may be analyzed as a transmission line. The propagation constant and characteristic impedance are determined by the static distributed inductance and capacitance. In general, at low frequencies the mode of propagation would be called a quasi-TEM mode since the axial electric field e_z, even though it is small, is not zero. At high frequencies the mode of propagation is an E mode and departs significantly from a TEM mode in its field distribution.

High-Frequency Solution

At high frequencies k_0, and hence l_0 and p_0, are large. In this case $p_0 c$ is large so we can replace $\tanh p_0 c$ by unity and (3.120a) gives

$$l_0 \tan l_0 a = \epsilon_r p_0 = \epsilon_r\sqrt{(\epsilon_r - 1)k_0^2 - l_0^2} \tag{3.126}$$

upon using (3.120b) to eliminate p_0. This equation is independent of the

plate separation b. The solution for e_z can be approximated as follows:

$$e_z(y) = C_1 \sin l_0 y \qquad 0 \le y \le a \tag{3.127a}$$

$$\begin{aligned} e_z(y) &= C_2 j \sinh p_0(b-y) \\ &= C_1 \sin l_0 a \frac{\sinh p_0(b-y)}{\sinh p_0(b-a)} \\ &\approx C_1 \sin l_0 a \frac{e^{p_0(b-y)}}{e^{p_0(b-a)}} \\ &= C_1 \sin(l_0 a) e^{-p_0(y-a)} \qquad a \le y \le b \end{aligned} \tag{3.127b}$$

This is a field that decays exponentially away from the air-dielectric surface and does not depend on b as long as $p_0 c = p_0(b-a)$ is large. This field is guided by the dielectric sheet on the ground plane (lower conductor) even if the upper plate is removed to infinity. This type of mode is called a surface-wave mode because its field is confined close to the guiding surface. The axial electric field for this surface-wave mode is illustrated in Fig. 3.15.

The first root for l_0 in the eigenvalue equation (3.126) occurs for $l_0 a < \pi/2$ or $l_0 < \pi/2a$. Thus, as k_0 approaches infinity, l_0 remains bounded but p_0 will become large because $p_0^2 = (\epsilon_r - 1)k_0^2 - l_0^2$. Consequently, for large enough k_0 we will have $l_0^2 \ll k^2$ and then $\beta \approx k$. As we go from zero frequency to very high frequencies, the propagation constant varies from a low value of $\sqrt{\epsilon_e} k_0$ given by (3.121) to an asymptotic value of $\sqrt{\epsilon_r} k_0$. We see that β is not a linear function of ω or k_0 and for this reason is said to exhibit dispersion. The term *dispersion* arises from a consideration of signal propagation. In our discussion of waveguides later on in this chapter, we will show that a signal consisting of a band of frequencies will have its frequency components dispersed whenever β is not a linear function of ω. This is caused by the phase velocity v_p, given by the relation $\beta = \omega/v_p$ and thus $v_p = \omega/\beta = 1/\sqrt{\epsilon_e \mu_0 \epsilon_0}$, being a function of ω. The ratio β^2/k_0^2 gives the effective dielectric constant at any frequency. In Fig. 3.16 we show a plot of ϵ_e versus frequency for the case when $\epsilon_r = 10$, $a = 0.4$ cm, and $b = 1$ cm. This curve is derived by solving the pair of equations (3.120). Microstrip transmission lines exhibit similar dispersion characteristics.

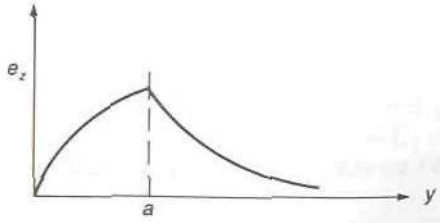


FIGURE 3.15
Axial electric field for surface-wave mode.

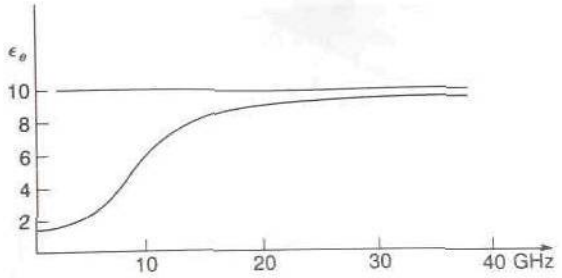


FIGURE 3.16
Effective dielectric constant as a function of frequency for $\epsilon_r = 10$, $a = 0.4$ cm, and $b = 1$ cm.

A second surface-wave mode solution can be found from (3.120*a*) with $l_0 a$ in the range $\pi < l_0 a < 3\pi/2$ provided $\sqrt{\epsilon_r - 1}\, k_0$ is larger than π so that (3.120*b*) can also be satisfied. For large $k_0 a$ many surface-wave modes can propagate. In addition to the surface-wave modes, there are also an infinite number of solutions to (3.118) for real values of p. The higher-order solutions have values of p on the order of $n\pi/b$ in value, where n is an integer. Provided π/b is greater than k_0, these values of p will give imaginary values of β and hence nonpropagating modes. The cutoff occurs when $p = k_0$ giving $\beta = 0$. Thus at cutoff

$$l \tan la = -\epsilon_r k_0 \tan k_0 c = \sqrt{\epsilon_r}\, k_0 \tan \sqrt{\epsilon_r}\, k_0 a$$

since $l = k$ for $\beta = 0$. This equation reduces to

$$\tan \sqrt{\epsilon_r}\, k_0 a = -\sqrt{\epsilon_r} \tan k_0 c$$

which can be solved for the values of k_0 at which the various modes cease to propagate.

We will not consider the partially filled parallel-plate waveguide any further even though a good deal more could be said about its mode spectrum. The purpose of our discussion is to highlight those features that will be displayed by microstrip transmission lines, which is the next topic taken up.

3.11 PLANAR TRANSMISSION LINES

A planar transmission line is a transmission line with conducting metal strips that lie entirely in parallel planes. The most common structure is one or more parallel metal strips placed on a dielectric substrate material adjacent to a conducting ground plane. A planar transmission line that is widely used is the microstrip line shown in Fig. 3.17. It consists of a single conducting strip of width W placed on a dielectric substrate of thickness H and located on a ground plane. By image theory this transmission line is equivalent to a line consisting of two parallel conducting strips placed opposite each other on a dielectric sheet of thickness $2H$ as also shown in

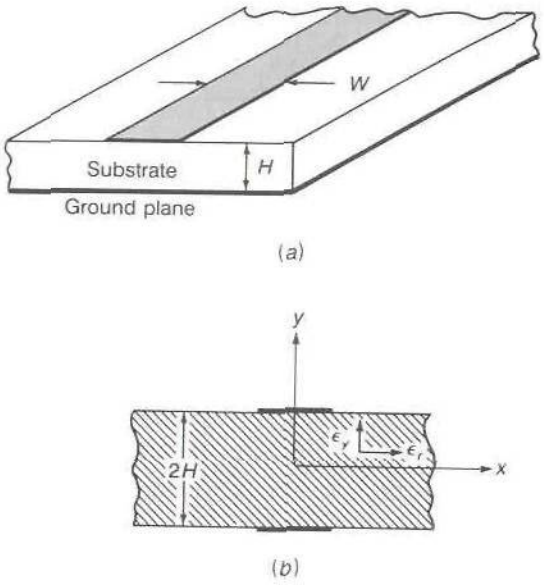


FIGURE 3.17
(*a*) The microstrip transmission line; (*b*) equivalent parallel strip line obtained by using image theory.

Fig. 3.17. Typical dimensions for a microstrip line are substrate thickness of 0.25 to 1 mm and strip widths of 0.1 to 5 mm.

The microstrip transmission line can be fabricated using conventional printed-circuit-board techniques which result in good mechanical tolerances and a low cost.

In addition to the microstrip line, there are many other planar-transmission-line structures that are used for various purposes. A number of these other transmission-line configurations are shown in Figs. 3.18 to 3.20. The coupled microstrip line shown in Fig. 3.18*a* is used in directional couplers. The coupled microstrip line supports two modes of propagation. The even mode of propagation has the same voltage and current on the two strips, while the odd mode of propagation has opposite voltages and currents on the two strips.

The coplanar transmission line shown in Fig. 3.18*b* consists of a single strip mounted between two ground planes on the same side of the dielectric substrate. The coplanar line has an advantage over the microstrip line in that shunt connection of components to the ground plane can be made on the same side of the substrate. In addition, it allows the series connection of components to be made with equal facility to that for microstrip lines. The coplanar strip line shown in Fig. 3.18*c* is similar to the coplanar line in that all conducting strips are in the same plane (coplanar). It is less desirable than the coplanar line because it is not balanced relative to a ground plane and thus wave propagation on this line is more strongly influenced by nearby conductors such as a shielding enclosure. In practice, a shielding

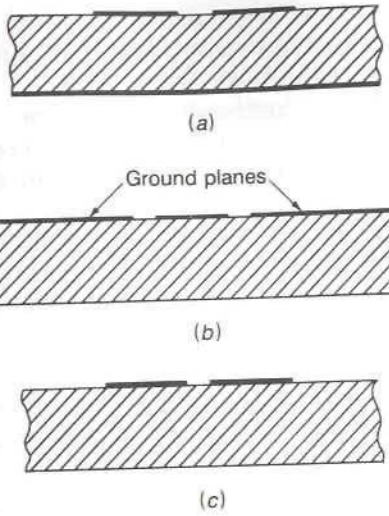


FIGURE 3.18
(*a*) Coupled microstrip lines; (*b*) coplanar transmission line; (*c*) coplanar strip transmission line.

FIGURE 3.19
Suspended and inverted suspended microstrip line.

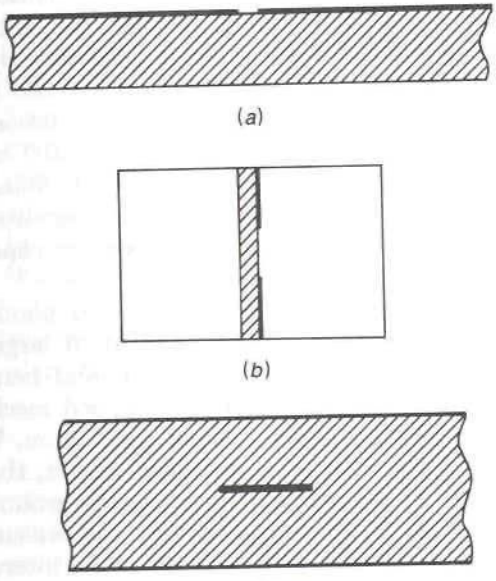


FIGURE 3.20
(*a*) Slot line; (*b*) shielded slot line or fin line; (*c*) strip line.

enclosure for a microwave circuit is needed to reduce spurious radiation from the circuit, eliminate electromagnetic coupling with nearby circuits, and for environmental protection.

In Fig. 3.19 are illustrated the suspended and inverted suspended microstrip lines which are quite similar to the conventional microstrip line but involve less dielectric substrate material. Figure 3.20*a* shows a slot line. The open slot line is not as widely used as the microstrip and coplanar lines are. The shielded slot line, which is typically a slot line placed inside a rectangular waveguide as shown in Fig. 3.20*b*, is called a fin line and has found to be useful for many circuit applications. Only shunt-connected components can be used with a slot line.

The final transmission-line structure illustrated here is the strip line shown in Fig. 3.20*c*. This line consists of a strip placed between two parallel plates that function as ground planes. The strip may be rigid enough to be suspended in air or it may be sandwiched between two dielectric sheets as shown in the figure. The strip line was often used in microwave filters and couplers before the other forms of planar transmission lines became popular. The strip line is somewhat more difficult to fabricate but has certain advantages for special applications to filters and couplers. Many directional coupler, power divider, and filter designs using strip lines were developed in the period from 1955 to 1975. An excellent reference source for design data for strip-line circuits is the book by Howe listed in the references at the end of this chapter.

The methods used to fabricate planar-transmission-line structures and related circuit elements are compatible with integrated circuit fabrication and have allowed the development of microwave integrated circuits (MIC circuits). In integrated microwave circuits the active devices and all interconnecting transmission lines, impedance-matching elements, needed capacitors and resistors, etc., are fabricated on the same chip. In these applications the microstrip and coplanar transmission lines are the ones most easily adapted for on-chip fabrication. In MIC circuits the substrate thickness and line widths are generally much smaller than in hybrid circuits. The term *hybrid* is used to describe integrated microwave circuits where the discrete components such as transistors, capacitors, and resistors are soldered in place.

The dielectric substrate material used in a planar transmission line must have low loss, i.e., a small loss tangent. A large dielectric constant results in a shorter propagation wavelength and hence a more compact circuit. The substrate material should have good mechanical strength, be easy to machine, and have good thermal conduction. When active devices are mounted into a planar-transmission-line circuit, the heat generated by the active device is in part conducted away to the ground plane through the substrate material. It is difficult to use metal heat sinks in a microwave circuit because these large metal structures would interact with the electro-

magnetic field in an undesirable, and often unpredictable, manner. Consequently, in power amplifier circuits substrate materials with good thermal conductivity are required. Dielectric materials used in low-frequency circuit boards are generally too lossy to be used for microwave transmission lines. The dielectric constant and thickness must be maintained to a high level of uniformity in the manufacturing of substrates because otherwise the fabricated transmission lines will not perform according to the specified design since the propagation phase constant and characteristic impedance both depend on these parameters. Uniform thickness and dielectric constant is particularly important in the design of filters and impedance-matching elements whose dimensions are critical. Once a printed microwave structure such as a filter has been constructed, it is not very easy to add external tuning elements to bring the constructed filter performance into specifications.

A commonly used substrate material is polytetrafluoroethylene (PTFE)† which has a dielectric constant of 2.1 and a loss tangent of 0.0002 at 1 MHz and around 0.0005 at microwave frequencies. This material has excellent resistance to chemicals used in the photoetching process. In order to increase the mechanical strength, it can be loaded with woven fiberglass mat or glass microparticles. This increases the dielectric constant to the range 2.2 to 3. The use of glass fiber results in some anisotropy in the dielectric constant. In the manufacturing process the glass fibers are generally aligned parallel with the substrate so the dielectric constant along the substrate is typically 5 to 10 percent larger than that normal to the substrate. By using ceramic powders as fillers, notably titanium oxide, much larger dielectric constants can be obtained. Typical values are in the range 5 to 15.

Ceramic materials such as aluminum oxide (alumina) and boron nitride, as well as the glasslike material sapphire, are also used for substrates. These materials are very difficult to machine. Alumina is perhaps the most commonly used material. It has excellent thermal conductivity. For integrated microwave circuits the usual semiconductor materials germanium, silicon, and gallium arsenide are used. These substrate materials have a high dielectric constant and may exhibit some conductivity depending on the doping level.

In Table 3.2 we have summarized the important properties of a number of substrate materials. In this table ϵ_r is the dielectric constant along the substrate and ϵ_y is the dielectric constant normal to the substrate.‡

†This material is commonly known as Teflon, which is a registered trade name of Du Pont.

‡In order to keep the notation as simple as possible, we use ϵ_y instead of ϵ_{ry} for the relative permittivity (dielectric constant) in the y direction.

TABLE 3.2
Properties of substrate materials

Material	ϵ_r	ϵ_y	Loss tangent	Thermal conductivity	Machinability
PTFE/woven glass	2.84	2.45	0.001–0.002	Fair	Good
PTFE/microfiberglass	2.26	2.2	0.0005–0.001	Fair	Good
*CuFlon	2.1	2.1	0.0004	Fair	Good
*RT/Duroid 5880	2.26	2.2	0.001	Fair	Good
*RT/Duroid 6006	6.36	6		Medium	Good
*Epsilam 10	13	10.3		Medium	Good
Boron nitride	5.12	3.4		Good	Poor
Silicon	11.7–12.9	11.7–12.9	0.001–0.003	Medium	Poor
Germanium	16	16		Medium	Poor
Gallium arsenide	12.9	12.9	0.0005–0.001	Medium	Poor
Alumina	9.6–10.1	9.6–10.1	0.0005–0.002	Good	Poor
Sapphire	9.4	11.6	0.0002	Good	Poor
Beryllium oxide	6.7	6.7	0.001–0.002	Good	Poor

*CuFlon is a registered trademark of Polyflon Company. It is a Teflon material electroplated with copper. RT/Duroid is a registered trademark of Rogers Corporation. Rogers Corporation also manufactures substrates with dielectric constants around 10. Epsilam 10 is a registered trademark of the 3M Company. It is a ceramic-filled Teflon material.

The data in Table 3.2 have been compiled from a variety of sources.† Since the dielectric constant and loss tangent are frequency dependent and also influenced by the material processing, the listed data should be viewed as representative values at microwave frequencies.

Substrate materials are usually plated with copper in 0.5-, 1-, or 2-oz weights (amount of copper per square foot). The use of 1-oz copper weight gives a plating thickness of 0.0014 in. Gold plating is sometimes used on top of the copper to prevent oxidation of the metal. In integrated microwave circuit construction a metalization thickness of a few microns is typical. One-half oz copper-clad board has a metalization thickness of 18 μm.

3.12 MICROSTRIP TRANSMISSION LINE

In a microstrip transmission line the dielectric material does not completely surround the conducting strip and consequently the fundamental mode of propagation is not a pure TEM mode. At low frequencies, typically below a

†H. Howe, "Stripline Circuit Design," Artech. House Books, Dedham, Mass., 1974.

T. Laverghetta, Microwave Materials: The Choice is Critical, *Microwave J.*, vol. 28, p. 163, 1985.

M. N. Afsar and K. J. Button, Precise Millimeter-Wave Measurements of Complex Refractive Index, Complex Dielectric Permittivity and Loss Tangent of GaAs, Si, SiO_2, Al_2O_3, BeO, Macor, and Glass, *IEEE Trans.*, vol. MTT-31, pp. 217–223, 1983.

Some data were also obtained from manufacturers' literature.

few gigahertz for practical microstrip lines, the mode is a quasi-TEM mode. In the frequency range up to a gigahertz or somewhat higher, the microstrip transmission line can be characterized in terms of its distributed capacitance and inductance per meter in a manner similar to what was found for the partially loaded parallel-plate transmission line in the previous section. Unfortunately, there are no simple closed-form analytic expressions that can be derived for describing the field distribution or the characteristics of planar transmission lines. Formal solutions can be derived and evaluated on a computer and have been used to compile data on the characteristics of these transmission-line structures. Static field analysis has also been extensively used to obtain the low-frequency characteristics. However, even the static field analysis is quite complex.

The analysis of planar transmission lines can be based directly on a solution for the electric and magnetic fields in the structure. An alternative approach is to first solve for the scalar and vector potential functions and from these find the corresponding electromagnetic field. In actual fact the propagation constant and characteristic impedance can be found from the potentials without a detailed consideration of the fields. The advantage of using the scalar and vector potentials in the analysis is that this approach provides a direct link to the quasistatic solutions in terms of more familiar low-frequency concepts.

In this section we will develop the essential equations to be satisfied by the scalar and vector potentials for a microstrip transmission line. From these equations we then obtain simplified ones that will describe the quasi-TEM mode of propagation at low frequencies. The term low frequency is a relative one. It is the ratio of line dimensions to wavelength that determines whether a microstrip line can be adequately described in terms of the quasi-TEM mode of propagation. In MIC circuits with line widths as small as 100 μm, the low-frequency region can extend as high as 20 to 30 GHz. Even though space does not permit a full development of analytic methods suitable for solving planar-transmission-line problems, some insight into the properties of these structures is obtained from the basic formulation of the relevant equations.

After we have presented the theoretical foundations, typical dispersion curves and graphical results for characteristic impedances are given for a number of important substrate materials and a range of microstrip conductor widths.

The vector and scalar potential functions are solutions of Helmholtz equations as described in Sec. 2.10 when the sources are located in vacuum (air). For the microstrip line shown in Fig. 3.17, two added complications enter due to the presence of the dielectric in only part of the region of interest and the anisotropic nature of some substrate materials. For this reason we need to derive new equations to be satisfied by the potential functions. The substrate material will be characterized by a dielectric constant ϵ_y in the y direction which is normal to the interface and by a

dielectric constant ϵ_r in the x and z directions. The unknown charge and current densities on the conducting microstrip will be denoted by ρ and $\mathbf{J}$. These source densities are concentrated along y since they exist only on the microstrip which is assumed to have negligible thickness. The source concentration can be described by introducing the delta function $\delta(y - H)$ to localize the sources at $y = H$. Thus we can write

$$\rho(x, y, z) = \rho_s(x, z)\delta(y - H) \tag{3.128a}$$

$$\mathbf{J}(x, y, z) = \mathbf{J}_s(x, z)\delta(y - H) \tag{3.128b}$$

where $\mathbf{J}_s$ and ρ_s now describe surface densities rather than volume densities.

We will assume that the dielectric constants $\epsilon_y(y)$ and $\epsilon_r(y)$ are functions of y that are constant in the substrate and undergo a rapid change in value to unity as the interface is crossed into the air region. The reason for doing this is that the equations we then obtain for the potentials will automatically give us the boundary conditions needed to properly join the solutions for the potentials in the substrate region to those in the air region.

We begin the derivation by letting

$$\mathbf{B} = \nabla \times \mathbf{A}$$

From Maxwell's equation

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B} = -j\omega\nabla \times \mathbf{A}$$

we get $\nabla \times (\mathbf{E} + j\omega\mathbf{A}) = 0$ which has the general solution

$$\mathbf{E} = -j\omega\mathbf{A} - \nabla\Phi$$

where Φ is a scalar potential function. Up to this point we have followed the same steps as in Sec. 2.10. Maxwell's curl equation for the magnetic field is

$$\nabla \times \mathbf{H} = j\omega\mathbf{D} + \mathbf{J}$$

For an anisotropic dielectric we have

$$\mathbf{D} = \epsilon_0\epsilon_r(E_x\mathbf{a}_x + E_z\mathbf{a}_z) + \epsilon_0\epsilon_y E_y\mathbf{a}_y$$

We can replace $\mathbf{H}$ by $\mu_0^{-1}\nabla \times \mathbf{A}$ to obtain

$$\nabla \times \nabla \times \mathbf{A} = \nabla\nabla \cdot \mathbf{A} - \nabla^2\mathbf{A} = j\omega\mu_0\mathbf{D} + \mu_0\mathbf{J}$$

and express $\mathbf{D}$ in terms of the potentials as follows:

$$\mathbf{D} = -\epsilon_0\epsilon_r\left[j\omega(A_x\mathbf{a}_x + A_z\mathbf{a}_z) + \mathbf{a}_x\frac{\partial\Phi}{\partial x} + \mathbf{a}_z\frac{\partial\Phi}{\partial z}\right]$$

$$- \epsilon_0\epsilon_y\left(jA_y\mathbf{a}_y + \mathbf{a}_y\frac{\partial\Phi}{\partial y}\right)$$

By adding and subtracting a term to the y component that includes the factor ϵ_r, we can reexpress $\mathbf{D}$ in the form

$$\mathbf{D} = -\epsilon_r\epsilon_0(j\omega\mathbf{A} + \nabla\Phi) - \epsilon_0(\epsilon_y - \epsilon_r)\left(j\omega A_y\mathbf{a}_y + \mathbf{a}_y\frac{\partial\Phi}{\partial y}\right)$$

We wish to eliminate the $\nabla\nabla\cdot\mathbf{A}$ term in the equation for $\mathbf{A}$ by setting it equal to the gradient of another function. For this purpose we now express $\epsilon_r\nabla\Phi$ in the form $\nabla(\epsilon_r\Phi) - \Phi\nabla\epsilon_r$, where $\nabla\epsilon_r$ has only a y component since ϵ_r is a function of y only. We can set $\nabla\nabla\cdot\mathbf{A}$ equal to $-j\omega\epsilon_0\mu_0\nabla(\epsilon_r\Phi)$ which gives the Lorentz condition

$$\nabla\cdot\mathbf{A} = -j\omega\epsilon_0\epsilon_r\mu_0\Phi \tag{3.129}$$

The equation for the vector potential now becomes

$$-\nabla^2\mathbf{A} = j\omega\mu_0\left[-j\omega\epsilon_0\epsilon_r\mathbf{A} + \epsilon_0\Phi\nabla\epsilon_r - \epsilon_0(\epsilon_y - \epsilon_r)\left(j\omega\mathbf{a}_y A_y + \mathbf{a}_y\frac{\partial\Phi}{\partial y}\right)\right] + \mu_0\mathbf{J}$$

The current $\mathbf{J}$ does not have a y component so the x and z components of this equation are

$$\nabla^2 A_x + \epsilon_r(y)k_0^2 A_x = -\mu_0 J_x \tag{3.130a}$$

$$\nabla^2 A_z + \epsilon_r(y)k_0^2 A_z = -\mu_0 J_z \tag{3.130b}$$

while the y component becomes

$$\nabla^2 A_y + \epsilon_y(y)k_0^2 A_y = -j\omega\mu_0\epsilon_0\left[\Phi\frac{\partial\epsilon_r}{\partial y} - (\epsilon_y - \epsilon_r)\frac{\partial\Phi}{\partial y}\right]$$

$$= j\omega\mu_0\epsilon_0\left[(\epsilon_y - \epsilon_r)\frac{\partial\Phi}{\partial y} + \Phi(H)(\epsilon_r - 1)\delta(y - H)\right] \tag{3.131}$$

where $-(\epsilon_r - 1)\delta(y - H)$ expresses the derivative of the step change that occurs in ϵ_r as y crosses the interface at H, that is, $\epsilon_r(y)$ changes from ϵ_r to unity.

The equations for A_x and A_z are of the same form as derived in Sec. 2.10, but the equation for A_y is new. The equation for A_y is coupled to the scalar potential $\Phi(H)$ at the boundary even if we have an isotropic substrate. Thus boundary conditions require the presence of an A_y component *even though there is no y component of current.*

A separate equation for the scalar potential is obtained by using Gauss' law $\nabla \cdot \mathbf{D} = \rho$ and the Lorentz condition (3.129). Thus we find that

$$\nabla \cdot \mathbf{D} = \nabla \cdot \left[-\epsilon_r \epsilon_0 (j\omega \mathbf{A} + \nabla \Phi) - \epsilon_0 (\epsilon_y - \epsilon_r) \left(j\omega \mathbf{a}_y A_y + \mathbf{a}_y \frac{\partial \Phi}{\partial y} \right) \right]$$

$$= -\epsilon_0 \left[j\omega \nabla \cdot (\epsilon_r \mathbf{A}) + \nabla \cdot (\epsilon_r \nabla \Phi) + j\omega \frac{\partial}{\partial y} (\epsilon_y - \epsilon_r) A_y + \frac{\partial}{\partial y} (\epsilon_y - \epsilon_r) \frac{\partial \Phi}{\partial y} \right]$$

$$= -\epsilon_0 \left[j\omega (\epsilon_r \nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla \epsilon_r) + \nabla \cdot (\epsilon_r \nabla \Phi) + j\omega \frac{\partial}{\partial y} (\epsilon_y - \epsilon_r) A_y + \frac{\partial}{\partial y} (\epsilon_y - \epsilon_r) \frac{\partial \Phi}{\partial y} \right] = \rho$$

By replacing $\nabla \cdot \mathbf{A}$ with $-j\omega \epsilon_0 \mu_0 \epsilon_r \Phi$ from the Lorentz condition, we obtain

$$\nabla \cdot (\epsilon_r \nabla \Phi) + \frac{\partial}{\partial y} (\epsilon_y - \epsilon_r) \frac{\partial \Phi}{\partial y} + \epsilon_r^2 k_0^2 \Phi = -j\omega A_y \frac{\partial \epsilon_r}{\partial y} - j\omega \frac{\partial}{\partial y} (\epsilon_y - \epsilon_r) A_y - \frac{\rho}{\epsilon_0}$$

The last step is to simplify this equation using

$$\frac{\partial}{\partial y} (\epsilon_y - \epsilon_r) A_y = (\epsilon_y - \epsilon_r) \frac{\partial A_y}{\partial y} + A_y \left(\frac{\partial \epsilon_y}{\partial y} - \frac{\partial \epsilon_r}{\partial y} \right)$$

and

$$\nabla \cdot \epsilon_r \nabla \Phi = \epsilon_r \nabla^2 \Phi + \frac{\partial \epsilon_r}{\partial y} \frac{\partial \Phi}{\partial y}$$

$$\frac{\partial}{\partial y} (\epsilon_y - \epsilon_r) \frac{\partial \Phi}{\partial y} = (\epsilon_y - \epsilon_r) \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial \Phi}{\partial y} \left(\frac{\partial \epsilon_y}{\partial y} - \frac{\partial \epsilon_r}{\partial y} \right)$$

By using these expressions a number of terms cancel and we obtain the final form

$$\epsilon_r \left(\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial z^2} \right) + \frac{\partial}{\partial y} \left(\epsilon_y \frac{\partial \Phi}{\partial y} \right) + \epsilon_r^2 k_0^2 \Phi = -\frac{\rho}{\epsilon_0} + j\omega (\epsilon_y - 1) A_y(H) \delta(y - H) - j\omega (\epsilon_y - \epsilon_r) \frac{\partial A_y}{\partial y} \quad (3.132)$$

where we have also used

$$A_y \frac{\partial \epsilon_y(y)}{\partial y} = -(\epsilon_y - 1) A_y(H) \delta(y - H)$$

This equation also displays a coupling between the scalar potential and $A_y(H)$ at the interface as well as coupling within the substrate whenever ϵ_y does not equal ϵ_r, that is, for anisotropic substrates. In the air region $y > H$, both ϵ_r and ϵ_y are replaced by unity in (3.130) to (3.132).

After the above lengthy derivation we can now obtain simpler equations to be solved in each separate region along with boundary conditions to use in joining the solutions at the interface $y = H$. The source terms ρ and $\mathbf{J}$ when expressed in the form (3.128) contain the $\delta(y - H)$ factor. In order for the left-hand sides of (3.130) to (3.132) to equal the corresponding right-hand sides, we must obtain a delta function $\delta(y - H)$ from the derivative of the potentials with respect to y. In (3.130a) and (3.130b) this requires that $\partial A_x/\partial y$ and $\partial A_z/\partial y$ have a step change at the interface so that the second derivative with respect to y will produce a delta function. The required step change can be found by integrating both sides of the equation over a vanishingly small interval centered on $y = H$. The integral of terms not involving a derivative with respect to y will vanish since these terms must be continuous at $y = H$ and the interval length vanishes. For example, if A_x were not continuous at $y = H$, the second derivative with respect to y would generate a singular term corresponding to the derivative of the delta function and no such term exists on the right-hand side of the equation. Thus from (3.130a) we obtain

$$\lim_{\tau \to 0} \int_{H-\tau}^{H+\tau} \frac{\partial^2 A_x}{\partial y^2}\, dy = \frac{\partial A_x}{\partial y}\bigg|_{H^-}^{H^+}$$

$$= \lim_{\tau \to 0} \int_{H-\tau}^{H+\tau} -\mu_0 J_{sx}(x, z)\delta(y - H)\, dy = -\mu_0 J_{sx}$$

or

$$\frac{\partial A_x}{\partial y}\bigg|_{H^-}^{H^+} = -\mu_0 J_{sx} \tag{3.133a}$$

In a similar way we obtain

$$\frac{\partial A_z}{\partial y}\bigg|_{H^-}^{H^+} = -\mu_0 J_{sz} \tag{3.133b}$$

The notation H^+ and H^- means evaluating the derivative on adjacent sides of the interface at $y = H$. These two equations state that the tangential components of the magnetic field must be discontinuous across the current sheet $\mathbf{J}_s$ since from the equation $\mathbf{B} = \nabla \times \mathbf{A}$:

$$\frac{\partial A_z}{\partial y} = \mu_0 H_x \qquad \frac{\partial A_x}{\partial y} = -\mu_0 H_z$$

In a similar way we obtain the following boundary conditions by integrating

(3.131) and (3.132) about a small interval centered on $y = H$:

$$\left.\frac{\partial A_y}{\partial y}\right|_{H^-}^{H^+} = j\omega\mu_0\epsilon_0(\epsilon_r - 1)\Phi(H) \tag{3.133c}$$

$$\left.\frac{\partial \Phi}{\partial y}\right|_{H^+} - \epsilon_y \left.\frac{\partial \Phi}{\partial y}\right|_{H^-} = -\frac{\rho_s}{\epsilon_0} + j\omega(\epsilon_y - 1)A_y(H) \tag{3.133d}$$

A term such as $(\epsilon_y - \epsilon_r)\partial\Phi/\partial y$ that occurs in (3.131), and a similar term occurring in (3.132), does not contribute because

$$\lim_{\tau\to 0}\int_{H-\tau}^{H+\tau}(\epsilon_y - \epsilon_r)\frac{\partial\Phi}{\partial y}\,dy = \lim_{\tau\to 0}(\epsilon_y - \epsilon_r)\int_{H-\tau}^{H}\frac{\partial\Phi}{\partial y}\,dy = 0$$

since $\partial\Phi/\partial y$ is continuous in the interval $H - \tau \le y < H$ and for $y > H$ we have $\epsilon_y = \epsilon_r = 1$. The boundary conditions on Φ reflect the fact that the total y-directed electric field has a contribution from A_y so that the discontinuity in D_y across the charge layer is given by

$$-\epsilon_0\left(\frac{\partial\Phi}{\partial y} + j\omega A_y\right)\bigg|_{H^+} + \epsilon_y\epsilon_0\left(\frac{\partial\Phi}{\partial y} + j\omega A_y\right)\bigg|_{H^-} = \rho_s$$

which is (3.133d).

By using the above boundary conditions, we can solve (3.130) to (3.132) in each respective region away from the interface. Thus we need only to solve the following homogeneous equations, subject to the specified boundary conditions, in the substrate region:

$$(\nabla^2 + \epsilon_r k_0^2)A_x = 0 \tag{3.134a}$$

$$(\nabla^2 + \epsilon_r k_0^2)A_z = 0 \tag{3.134b}$$

$$\nabla^2 A_y + \epsilon_y k_0^2 A_y = j\omega\mu_0\epsilon_0(\epsilon_y - \epsilon_r)\frac{\partial\Phi}{\partial y} \tag{3.134c}$$

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial z^2} + \frac{\epsilon_y}{\epsilon_r}\frac{\partial^2\Phi}{\partial y^2} + \epsilon_r k_0^2\Phi = -j\omega(\epsilon_y - \epsilon_r)\frac{\partial A_y}{\partial y} \tag{3.134d}$$

In the air region the equations to be solved are obtained by setting $\epsilon_r = \epsilon_y = 1$. There is no volume coupling between Φ and A_y in the air region or in an isotropic substrate region. Since we are interested in wave solutions representing waves propagating in the z direction, we can assume that the z dependence is $e^{-j\beta z}$. The second derivative with respect to z can then be replaced by $-\beta^2$. The common factor $e^{-j\beta z}$ can be deleted from the equations just as $e^{j\omega t}$ was dropped for convenience.

Low-Frequency Solutions

We can obtain the equations to be solved in the zero-frequency limit by assuming that the potentials and the source terms can be expanded as

power series in ω. Thus we let

$$\mathbf{A} = \mathbf{A}^0 + \omega\mathbf{A}^1 + \omega^2\mathbf{A}^2 + \cdots \tag{3.135a}$$

$$\Phi = \Phi^0 + \omega\Phi^1 + \omega^2\Phi^2 + \cdots \tag{3.135b}$$

$$\mathbf{J} = \mathbf{J}^0 + \omega\mathbf{J}^1 + \omega^2\mathbf{J}^2 + \cdots \tag{3.135c}$$

$$\rho = \rho^0 + \omega\rho^1 + \omega^2\rho^2 + \cdots \tag{3.135d}$$

The parameter $k_0^2 = \omega^2\mu_0\epsilon_0$ is of second order in ω. The propagation constant β can be expressed in the form $\beta = \sqrt{\epsilon_e}\,k_0$, where ϵ_e is a frequency-dependent effective dielectric constant. Consequently, β^2 is also of second order in ω.

We now substitute these power-series expansions into (3.130) to (3.132) and equate all zero-order terms to obtain the following lowest-order equations:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) A_x^0 = -\mu_0 J_x^0 \tag{3.136a}$$

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) A_z^0 = -\mu_0 J_z^0 \tag{3.136b}$$

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) A_y^0 = 0 \tag{3.136c}$$

$$\left(\epsilon_r \frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial y}\epsilon_y \frac{\partial}{\partial y}\right) \Phi^0 = -\frac{\rho^0}{\epsilon_0} \tag{3.136d}$$

Further information is obtained from the Lorentz condition (3.129) which gives

$$\frac{\partial A_x^0}{\partial x} + \frac{\partial A_y^0}{\partial y} = 0 \tag{3.137a}$$

$$-j\beta A_z^0 = -j\omega\epsilon_0\epsilon_r\Phi^0 \tag{3.137b}$$

In the air region ϵ_r and ϵ_y are set equal to unity. From the continuity equation relating current and charge, namely,

$$\nabla \cdot \mathbf{J} = -j\omega\rho$$

we obtain

$$\frac{\partial J_x^0}{\partial x} = 0 \tag{3.138a}$$

$$-j\beta J_z^0 = -j\omega\rho^0 \tag{3.138b}$$

In the above equations the $e^{-j\beta z}$ factor is not included but any derivative with respect to z was replaced by $-j\beta$.

Since J_x^0 must be zero at the edges $x = \pm W/2$ of the microstrip, we conclude that J_x^0 is zero because the integral of $\partial J_x^0/\partial x$ is at most a constant. Hence, to lowest order, there is no x-directed current on the microstrip and A_x^0 is zero. The Lorentz condition then requires that $\partial A_y^0/\partial y = 0$ and hence $A_y^0 = 0$ also since a constant A_y^0 is a trivial solution and would not produce any magnetic field contribution. Thus, to lowest order, we only have to solve for an A_z^0 and a scalar potential Φ^0. If we assume the microstrip to be at a potential V, then the boundary condition on Φ^0 is that it equals V on the microstrip and equals zero on the ground plane.

We can integrate the continuity equation (3.138*b*) across the microstrip line to get

$$\beta I_z^0 = \omega Q^0 \tag{3.139}$$

where I_z^0 is the total z-directed current on the microstrip and Q^0 is the total charge. On the microstrip the axial electric field must be zero. To *lowest order this boundary condition is*

$$E_z^1 = -j\omega A_z^0 - \frac{\partial \Phi^0}{\partial z} = -j\omega A_z^0 + j\beta\Phi^0 = 0$$

or

$$\omega A_z^0 = \beta V \tag{3.140}$$

Hence A_z^0 is also constant on the microstrip.

We will show shortly that the inductance L per unit length of the microstrip line is given by the equation

$$I_z^0 L = A_z^0 \tag{3.141}$$

The capacitance per unit length is given by

$$C = \frac{Q^0}{V} \tag{3.142}$$

By using these expressions to eliminate Q^0 in (3.139) and A_z^0 in (3.140), we *obtain the pair of equations*

$$\beta I_z^0 = \omega C V \tag{3.143a}$$

$$\omega L I_z^0 = \beta V \tag{3.143b}$$

from which we find that

$$\beta^2 = \omega^2 LC \tag{3.144a}$$

$$\frac{V}{I_z^0} = \sqrt{\frac{L}{C}} \tag{3.144b}$$

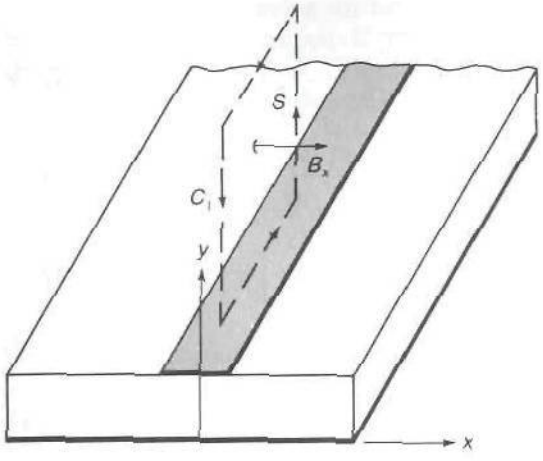


FIGURE 3.21
Surface used to find the magnetic flux linkage in a microstrip line.

We have thus been able to show in a rigorous way that in the *low-frequency limit the microstrip line can be analyzed as a static field* problem and that its propagation constant and characteristic impedance are *determined by the low-frequency distributed capacitance and inductance.* The analysis leading up to (3.144) is quite general and applies to other planar transmission lines as well.

At this point we return to the promised derivation of (3.141) giving the line inductance. With reference to Fig. 3.21 we note that the magnetic flux ψ linking the microstrip per unit length is given by the integral of B_x over the area extending from the microstrip to infinity. Thus

$$\psi = \int_0^1 \int_H^\infty \mathbf{B} \cdot \mathbf{a}_x \, dy \, dz$$

By using $\mathbf{B} = \nabla \times \mathbf{A}$ and Stokes' law, we can write

$$\psi = \int_0^1 \int_H^\infty \nabla \times \mathbf{A} \cdot \mathbf{a}_x \, dy \, dz$$

$$= \oint_{C_l} \mathbf{A} \cdot d\mathbf{l}$$

where C_l is the boundary of the area. Since A_y^0 is zero and A_z^0 is zero at infinity and is constant on the microstrip, we obtain $\psi = A_z^0$ for the flux linkage. The inductance is given by ψ / I_z^0 and this gives (3.141).

We can also derive equations for the next level of approximation. However, the solution of these equations is not much easier than the solution of the original equations; so it is not worthwhile to develop the power-series solutions beyond the lowest order. Thus the equations to be

solved are

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) A_z^0 = 0 \qquad y < H, y > H \tag{3.145a}$$

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\epsilon_y}{\epsilon_r}\frac{\partial^2}{\partial y^2}\right) \Phi^0 = 0 \qquad y < H \tag{3.145b}$$

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) \Phi^0 = 0 \qquad y > H \tag{3.145c}$$

with the boundary conditions

$$\left.\frac{\partial A_z^0}{\partial y}\right|_{H^-}^{H^+} = -\mu_0 J_{sz}^0 \tag{3.145d}$$

$$\left.\frac{\partial \Phi^0}{\partial y}\right|_{H^+} - \epsilon_y \left.\frac{\partial \Phi^0}{\partial y}\right|_{H^-} = -\frac{\rho_s}{\epsilon_0} \tag{3.145e}$$

$$\Phi^0 = V \qquad \text{on microstrip}$$

$$A_z^0 = \text{constant} \qquad \text{on microstrip}$$

Along the interface and away from the microstrip, the right-hand sides of (3.145d) and (3.145e) are zero. In addition, Φ^0 and A_z^0 must be zero on the ground plane in order to make the tangential electric field vanish on this surface.

The equations for A_z^0 do not depend on the dielectric constants of the substrate material. Hence the line inductance is the same as for an air-filled line. But for an air-filled transmission line with distributed capacitance C_a, we have

$$\sqrt{LC_a} = \sqrt{\mu_0 \epsilon_0} \qquad \text{and hence} \qquad L = \frac{\mu_0 \epsilon_0}{C_a}$$

so we can find L by finding the distributed capacitance of an air-filled microstrip line. By introducing C_a in place of L, the solutions for β and Z_c can be expressed in the form

$$\beta = \omega\sqrt{LC} = \sqrt{\frac{C}{C_a}}\, k_0 = \sqrt{\epsilon_e}\, k_0 \tag{3.146a}$$

$$Z_c = \sqrt{\frac{L}{C}} = \frac{\sqrt{\mu_0 \epsilon_0}}{\sqrt{CC_a}} = \sqrt{\frac{L}{C_a}}\sqrt{\frac{C_a}{C}} = \frac{Z_{c0}}{\sqrt{\epsilon_e}} \tag{3.146b}$$

where Z_{c0} is the characteristic impedance of the air-filled line and the ratio C/C_a gives the low-frequency equivalent (effective) dielectric constant ϵ_e.

The effect of having an anisotropic dielectric substrate does not add any additional complication. If we introduce a new variable $u = (\epsilon_r/\epsilon_y)^{1/2}y$, then upon using

$$\frac{\partial}{\partial y} = \frac{\partial}{\partial u}\frac{\partial u}{\partial y} = \sqrt{\frac{\epsilon_r}{\epsilon_y}}\frac{\partial}{\partial u}$$

we find that (3.145*b*) reduces to

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial u^2}\right)\Phi^0 = 0 \tag{3.147}$$

When $y = H$ the corresponding value of u is $(\epsilon_r/\epsilon_y)^{1/2}H$; so the solution of (3.147) is that for a microstrip with an equivalent substrate thickness H_e given by $(\epsilon_r/\epsilon_y)^{1/2}H = H_e$. The boundary condition (3.145*e*) becomes

$$\left.\frac{\partial\Phi^0}{\partial y}\right|_{H^+} - \epsilon_y\sqrt{\frac{\epsilon_r}{\epsilon_y}}\left.\frac{\partial\Phi^0}{\partial u}\right|_{H_e^-} = \left.\frac{\partial\Phi^0}{\partial y}\right|_{H^+} - \sqrt{\epsilon_r\epsilon_y}\left.\frac{\partial\Phi^0}{\partial u}\right|_{H_e^-} = -\frac{\rho_s}{\epsilon_0} \tag{3.148}$$

which shows that the equivalent dielectric constant of the substrate should be taken as the geometric mean $\epsilon_g = \sqrt{\epsilon_r\epsilon_y}$. Thus, by modifying the substrate thickness and introducing the equivalent dielectric constant, the solution for the distributed capacitance C for the case of an anisotropic substrate can be reduced to that for an isotropic substrate. The distributed capacitance C_a is that for the unscaled microstrip line.

The unit of length does not enter directly into the differential equations for the potentials. Thus x and y can be in units of meters, centimeters, inches, or any other convenient unit. What this means is that the distributed capacitance and inductance per unit length is dependent only on the ratio of strip width to substrate thickness, i.e., on W/H. If we have found C and L for a given set of values for W and H on a per-meter basis, then if we change W to sW and H to sH, where s is a scaling factor, both C and L on a per-meter basis do not change. Hence the characteristic impedance, effective dielectric constant, and propagation constant β for any planar transmission line is invariant to a scaling of the cross-sectional dimensions. However, the attenuation caused by conductor loss does not scale since the series resistance is inversely proportional to the conductor widths. The attenuation due to conductor losses will double if the conductor size is reduced by a factor of 2. The scaling law is clearly illustrated for an ideal parallel-plate capacitor with a plate dimension of Wl and separation H and having a capacitance $\epsilon_0 Wl/H$. Clearly keeping the ratio W/H fixed keeps the capacitance unchanged.

A variety of methods exist for solving the two-dimensional Laplace equation (3.147). For planar transmission lines the conformal mapping method is widely used, generally along with some approximations that are necessary because of not having a dielectric medium filling all of the space

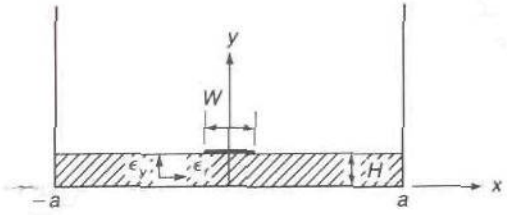


FIGURE 3.22
A microstrip line with perfectly conducting side walls inserted at $x = \pm a$ with $a \gg W$.

around the conductors. A number of useful solutions obtained by conformal mapping methods are described in App. III. We will refer to some of these solutions as needed.

In order to illustrate the general method of solution, we will develop a Fourier series solution to (3.145) which will turn out to provide an efficient method to obtain the parameters of a microstrip transmission line. In order to use the Fourier series method, we place perfectly conducting (electric) walls at $x = \pm a$ as shown in Fig. 3.22. Provided a is chosen equal to $10W$ or $10H$, whichever is larger, the sidewalls have a negligible effect on the field which is concentrated near the microstrip.

We can expand the unknown charge density ρ_s into a Fourier series of the form

$$\rho_s(x) = \sum_{n=1,3,\ldots}^{\infty} \rho_n \cos \frac{n\pi x}{2a} \tag{3.149a}$$

The charge coefficients ρ_n are given by

$$\rho_n = \frac{1}{a} \int_{-W/2}^{W/2} \rho_s(x') \cos \frac{n\pi x'}{2a}\, dx' \tag{3.149b}$$

The charge density is an even function of x because of the symmetry involved; so only a cosine series is needed. The functions are chosen so that they vanish at $x = \pm a$, a required boundary condition for the potential; so only odd integers n are used.

The potential $\Phi(x, y)$ can also be expanded into Fourier series; so we let

$$\begin{aligned} \Phi(x, y) &= \sum_{n=1,3,\ldots}^{\infty} f_n(y) \cos \frac{n\pi x}{2a} \qquad y \geq H \\ &= \sum_{n=1,3,\ldots}^{\infty} g_n(y) \cos \frac{n\pi x}{2a} \qquad 0 \leq y \leq H_e \end{aligned} \tag{3.150}$$

where $f_n(y)$ and $g_n(y)$ are to be found. We are using an effective substrate thickness H_e, so that an anisotropic substrate can be accommodated. We have dropped the superscript 0 since it is understood from the context that we are solving for a static potential field.

Each Fourier term in the expansion of Φ must be a solution of Laplace's equation. Hence we require that

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\begin{Bmatrix} f_n(y) \\ g_n(y) \end{Bmatrix} \cos\frac{n\pi x}{2a} = 0$$

which gives

$$\frac{d^2 f_n(y)}{dy^2} - w_n^2 f_n(y) = 0$$

$$\frac{d^2 g_n(y)}{dy^2} - w_n^2 g_n(y) = 0$$

where $w_n^2 = (n\pi/2a)^2$. In the region $y \leq H_e$, a suitable solution for $g_n(y)$ that vanishes on the ground plane is

$$g_n(y) = C_n \sinh w_n y$$

where C_n is an unknown constant. In the region $y \geq H$, we need a solution that will vanish as y approaches infinity; so we choose

$$f_n(y) = D_n e^{-w_n y}$$

where D_n is another unknown constant. At $y = H_e$, H the two potential functions must match; so we have

$$C_n \sinh w_n H_e = D_n e^{-w_n H}$$

The Fourier series expansion of the charge density ρ_s represents this charge density as sheets of charge $\rho_n \cos n\pi x/2a$ that extend from $x = -a$ to $x = a$. By superimposing an infinite number of such charge sheets, we obtain a charge density ρ_s that is nonzero only on the microstrip $-W/2 \leq x \leq W/2$. The boundary condition (3.148) is applied to each Fourier term to obtain

$$\left(\left.\frac{df_n}{dy}\right|_H - \epsilon_g \left.\frac{dg_n}{dy}\right|_{H_e}\right) = -\frac{\rho_n}{\epsilon_0}$$

or

$$-w_n D_n e^{-w_n H} - \epsilon_g w_n C_n \cosh w_n H_e = -\frac{\rho_n}{\epsilon_0}$$

We now have two equations which we can solve to find C_n and D_n. The solutions are

$$C_n = \frac{\rho_n}{\epsilon_0 w_n(\sinh w_n H_e + \epsilon_g \cosh w_n H_e)}$$

$$D_n = \frac{\rho_n e^{w_n H} \sinh w_n H_e}{\epsilon_0 w_n(\sinh w_n H_e + \epsilon_g \cosh w_n H_e)}$$

where $\epsilon_g = \sqrt{\epsilon_x \epsilon_y}$.

We now substitute our solutions for f_n and g_n into (3.150) and use (3.149*b*) for ρ_n. Thus we obtain

$$\Phi(x,y) = \sum_{n=1,3,\ldots}^{\infty} \int_{-W/2}^{W/2} \frac{\cos w_n x \cos w_n x'}{\epsilon_0 w_n a(\sinh w_n H_e + \epsilon_g \cosh w_n H_e)} \times \begin{Bmatrix} \sinh w_n y \\ \sinh w_n H_e e^{-w_n(y-H)} \end{Bmatrix} \rho_s(x')\,dx' \tag{3.151}$$

where the upper term is for $y \leq H_e$ and the lower one is for $y \geq H$. The factor multiplying $\rho_s(x')$ under the integral sign represents the Green's function for this problem. It is designated by the symbol $G(x, y; x', y')$; so in abbreviated form we express (3.151) as

$$\Phi(x,y) = \int_{-W/2}^{W/2} G(x, y; x', H_e)\rho_s(x')\,dx' \tag{3.152}$$

The last boundary condition to be imposed is the requirement that $\Phi = V$ on the microstrip; thus

$$V = \int_{-W/2}^{W/2} G(x, H_e; x', H_e)\rho_s(x')\,dx' \qquad -\frac{W}{2} \leq x \leq \frac{W}{2} \tag{3.153}$$

This is an integral equation whose solution would determine the unknown charge density $\rho_s(x')$. Once we know the charge density, we can calculate the total charge on the microstrip using

$$Q = \int_{-W/2}^{W/2} \rho_s(x')\,dx'$$

and find $C = Q/V$.

Integral equations are most often not solvable by analytic means. However, various numerical schemes exist for obtaining good approximate solutions. The most popular method is the *Method of Moments*.† In this method the first step is to choose a finite number of basis functions and to expand $\rho_s(x')$ in terms of these in the form

$$\rho_s(x') = \sum_{n=1}^{N} Q_n \psi_n(x') \qquad -\frac{W}{2} \leq x' \leq \frac{W}{2}$$

where Q_n are unknown coefficients. The basis functions could be the unit height pulse functions shown in Fig. 3.23, the cosine functions $\cos 2n\pi x/W$, or any other reasonable set that would give a good approximation to $\rho_s(x')$.

†R. F. Harrington, "Field Computation by Moment Methods," Krieger Publishing Company, Inc., Malabar, Fla., 1968.

FIGURE 3.23
Unit height pulse functions for expanding the charge density.

When this substitution is made in (3.153), we obtain

$$V = \sum_{n=1}^{N} Q_n G_n(x) \qquad -\frac{W}{2} \le x \le \frac{W}{2} \tag{3.154}$$

where $$G_n(x) = \int_{-W/2}^{W/2} G(x, H_e; x', H_e)\psi_n(x')\,dx'$$

The next step is to convert (3.154) into a matrix equation for the unknowns either by making both sides of the equation equal at N points in x along the microstrip, or by using weighting functions to make N weighted averages of both sides equal. We can choose the $\psi_m(x)$ as weighting functions, in which case the method is called Galerkin's method. Other choices for the weighting functions can also be made. If we use Galerkin's method, we obtain

$$\sum_{n=1}^{N} G_{nm} Q_n = V_m \qquad m = 1, 2, \ldots, N \tag{3.155}$$

where the matrix elements are given by

$$G_{nm} = \int_{-W/2}^{W/2} G_n(x)\psi_m(x)\,dx$$

$$= \iint_{-W/2}^{W/2} G(x, x')\psi_n(x)\psi_m(x')\,dx\,dx'$$

and the components V_m of the source vector are given by

$$V_m = \int_{-W/2}^{W/2} V\psi_m(x)\,dx \qquad m = 1, 2, \ldots, N$$

The system of linear equations given in (3.155) can be solved for the unknown charge amplitude coefficients Q_n. If N is chosen sufficiently large, we will obtain a good approximation for the charge density.

If we know a priori how $\rho_s(x')$ is distributed on the strip, or a close approximation to it, we could get an excellent approximate solution using very few basis functions. From conformal mapping solutions we know that on an isolated infinitely long conducting strip of width W and with total

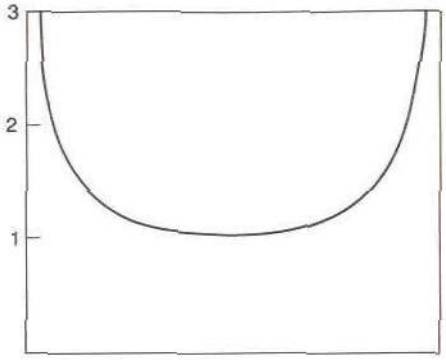


FIGURE 3.24
Charge density on an ideal isolated strip of width W.

charge Q per meter that the charge density is given by

$$\rho_s(x) = \frac{2Q}{\pi W\sqrt{1 - x^2/(W/2)^2}} \tag{3.156}$$

This charge distribution is illustrated in Fig. 3.24. At sharp corners and edges the charge density always exhibits an infinite behavior. However, the density is never so singular that it cannot be integrated. The reader can readily verify that if the substitution $x = (W/2)\sin\theta$, $dx = (W/2)\cos\theta\, d\theta$ is made, then

$$\int_{-W/2}^{W/2} \rho_s(x)\, dx' = \frac{Q}{\pi}\int_{-\pi/2}^{\pi/2} d\theta = Q$$

In a microstrip line the charge density is influenced by the dielectric substrate and the ground plane but surprisingly (3.156) remains a good approximation. By using a two- or three-term polynomial along with (3.156), *i.e., choosing*

$$\rho_s(x) = \frac{Q_0 + Q_1x^2 + Q_2x^4}{\sqrt{1 - x^2/(W/2)^2}} \tag{3.157}$$

an excellent approximate solution for $\rho_s(x)$ will be obtained.

We do not plan to solve the integral equation (3.153) numerically using the Method of Moments. It does require considerable numerical computation to evaluate the matrix elements because we only have the Green's function expressed as an infinite Fourier series. What we are going to do is to express the integral equation in a form that we can interpret as representing the capacitances between a conducting strip *in air* above a ground plane with spacings H_e, $2H_e$, $3H_e$, etc. We can then make use of the conformal mapping solution for a pair of strips in air to evaluate the distributed capacitance C for the microstrip line.

Let us choose $\epsilon_g = 1$ but still retain the strip spacing above the ground plane as H_e. By using (3.151) in the integral equation (3.153) for this case, we obtain (note that $\sinh w_n H_e + \cosh w_n H_e = e^{w_n H_e}$)

$$Q\int_{-W/2}^{W/2} \sum_{n=1,3,\ldots}^{\infty} \frac{\cos w_n x \cos w_n x'}{4\epsilon_0 w_n a}(1 - e^{-2w_n H_e}) \frac{\rho_s(x')}{Q}\, dx' = V \tag{3.158}$$

where we have set $y = H_e$ and multiplied and divided by the total charge Q. The part multiplied by $-e^{-2w_n H_e}$ represents the effect of the ground plane which is equivalent to that of the image strip at $y = -H_e$ and having a charge density $-\rho_s$. A solution of (3.158) provides the solution for the problem of a strip above a ground plane, as well as the solution to the problem of two parallel strips in air and separated by a distance $2H_e$. The capacitance per meter for a strip above a ground plane is given by

$$C_a(H_e) = \frac{Q}{V}$$

while that between two strips separated by a distance $2H_e$ is

$$\frac{Q}{2V} = \frac{C_a(H_e)}{2} = C_a(2H_e)$$

We see that the integral, involving the normalized charge density, represents $1/C_a(H_e)$.

Consider now the factor

$$\frac{\sinh w_n H_e}{\sinh w_n H_e + \epsilon_g \cosh w_n H_e}$$

that occurs in the Green's function in (3.151) at $y = H_e$. We can express this term in the form

$$\frac{e^{w_n H_e} - e^{-w_n H_e}}{(1+\epsilon_g)e^{w_n H_e} - (1-\epsilon_g)e^{-w_n H_e}} = \frac{1 - e^{-2w_n H_e}}{(1+\epsilon_g)\left(1 - \dfrac{1-\epsilon_g}{1+\epsilon_g}e^{-2w_n H_e}\right)}$$

$$= \frac{1}{1+\epsilon_g}(1 - e^{-2w_n H_e})(1 - \eta e^{-2w_n H_e})^{-1}$$

where $\eta = (1 - \epsilon_g)/(1 + \epsilon_g)$. The last factor is now expanded into a power

series to obtain

$$\frac{1}{1+\epsilon_g}(1-e^{-2w_nH_e})\sum_{m=0}^{\infty}\eta^m e^{-2mw_nH_e}$$

$$=\frac{1}{1+\epsilon_g}\left(\sum_{m=0}^{\infty}\eta^m e^{-2mw_nH_e}-\sum_{m=0}^{\infty}\eta^m e^{-2(m+1)w_nH_e}\right)$$

We now add $-\eta^m$ to both series which has no net effect because of the minus sign in front of the second series. This gives

$$\frac{1}{1+\epsilon_g}\left[-\sum_{m=0}^{\infty}\eta^m(1-e^{-2mw_nH_e})+\sum_{m=0}^{\infty}\eta^m(1-e^{-2(m+1)w_nH_e})\right]$$

The $m=0$ term is zero in the first series so we can change m to $m+1$ and still sum this for $m=0,1,2,3,\ldots$ without changing its value; thus we get

$$\frac{1}{1+\epsilon_g}\left[-\sum_{m=0}^{\infty}\eta^{m+1}(1-e^{-2(m+1)w_nH_e})+\sum_{m=0}^{\infty}\eta^m(1-e^{-2(m+1)w_nH_e})\right]$$

$$=-\frac{1}{1+\epsilon_g}(\eta-1)\sum_{m=0}^{\infty}\eta^m(1-e^{-2(m+1)w_nH_e})$$

$$=\frac{2\epsilon_g}{(1+\epsilon_g)^2}\sum_{m=0}^{\infty}\eta^m(1-e^{-2(m+1)w_nH_e}) \tag{3.159}$$

Upon using this expansion in the Green's function, the integral equation (3.153) can be expressed in the form

$$\sum_{m=0}^{\infty}\frac{4\epsilon_g}{(1+\epsilon_g)^2}\eta^m Q\int_{-W/2}^{W/2}\sum_{n=1,3,\ldots}^{\infty}\frac{\cos w_nx\cos w_nx'}{4\epsilon_0 w_n a}$$

$$(1-e^{-2(m+1)w_nH_e})\frac{\rho_s(x')}{Q}\,dx'=V \tag{3.160}$$

We note that the mth term considered by itself is an integral equation of the same form as that in (3.158) apart from the multiplying factor $4\epsilon_g\eta^m/(1+\epsilon_g)^2$. This integral equation would provide a solution for the capacitance of a strip in air spaced a distance $(m+1)H_e$ above the ground plane. If we assumed that the normalized charge density $\rho_s(x')/Q$ was the same for any strip, independent of the spacing above the ground plane, each integral would produce a constant voltage V_m but different from V. With increased strip spacing and constant total charge on each strip, the integral has to give a larger voltage since $Q=C_a[(m+1)H_e]V_m$ and $C_a[(m+1)H_e]$, the capacitance between the strip and the ground plane, decreases with increasing m. The approximation that the charge density is the same independent of strip–ground-plane spacing is a necessary one to make since there is only one charge density expression in the integral equation. The

approximation is a good one and by using it we can express (3.160) in the form

$$\sum_{m=0}^{\infty} \frac{4\epsilon_g}{(1+\epsilon_g)^2} \eta^m \frac{Q}{C_a[(m+1)H_e]} = V \tag{3.161}$$

since by our assumption

$$Q \int_{-W/2}^{W/2} \sum_{n=1,3,\ldots}^{\infty} \frac{\cos w_n x \cos w_n x'}{4\epsilon_0 w_n a} (1 - e^{-2(m+1)H_e})$$

$$\frac{\rho_s(x')}{Q}\, dx' = V_m = \frac{Q}{C_a[(m+1)H_e]}$$

From (3.161) we now obtain the following solution for the distributed capacitance $C = Q/V$:

$$C = \frac{\left[(1+\epsilon_g)^2/4\epsilon_g\right] C_a(H_e)}{1 + C_a(H_e) \sum_{m=1}^{M} \eta^m \frac{1}{C_a[(m+1)H_e]}} \tag{3.162}$$

where M is the number of terms that are kept. Since η is negative the series is an alternating one. Typically, from 10 to 40 terms are needed for good accuracy. The evaluation can be done on a computer very quickly and requires only a simple program to implement. However, we do require an expression for the capacitance between a strip in air as a function of the spacing above the ground plane, which is given below.

The exact solution for the capacitance between a strip of width $2W$ and a distance H above a ground plane is given in App. III, along with tabulated values as a function of $2W/H$. For practical applications it is desirable to have simple formulas that will enable the capacitance to be evaluated with an accuracy of 1 percent or better. A number of investigators have proposed such formulas which are based on approximate analytic solutions, along with empirical adjustment of various numerical constants so as to achieve the desired accuracy.† The following formulas give excellent results for the

†H. A. Wheeler, Transmission Line Properties of a Strip on a Dielectric Sheet on a Plane, *IEEE Trans.*, vol. MTT-25, pp. 631–647, August, 1977.

E. O. Hammerstad, Accurate Models for Microstrip Computer-Aided Design, *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 407–409, 1980.

E. O. Hammerstad, Equations for Microstrip Circuit Design, *Proc. European Micro. Conf., Hamburg, W. Germany*, pp. 268–272, September, 1975. Equations (3.163) and (3.166) come from this publication.

S. Y. Poh, W. C. Chew, and J. A. Kong, Approximate Formulas for Line Capacitance and Characteristic Impedance of Microstrip Line, *IEEE Trans.*, vol. MTT-29, pp. 135–142, February, 1981.

capacitance per meter of a strip of width W at a height H above a ground plane and with air dielectric:

$$C_a = \frac{2\pi\epsilon_0}{\ln\left(\frac{8H}{W} + \frac{W}{4H}\right)} \qquad \frac{W}{H} \leq 1 \quad (3.163a)$$

$$C_a = \epsilon_0\left[\frac{W}{H} + 1.393 + 0.667\ln\left(\frac{W}{H} + 1.444\right)\right] \qquad \frac{W}{H} > 1 \quad (3.163b)$$

These formulas give results that agree with those tabulated in App. III to within 1/4 percent.

The effect of finite thickness T for the microstrip on the distributed capacitance is normally negligible. If necessary the effect of finite thickness can be included by using an effective width W_e, where W_e is given by the following expressions due to Gunston and Weale:†

$$W_e = W + 0.398T\left(1 + \ln\frac{4\pi W}{T}\right) \qquad \frac{W}{H} \leq \frac{1}{2\pi}$$

$$= W + 0.398T\left(1 + \ln\frac{2H}{T}\right) \qquad \frac{W}{H} > \frac{1}{2\pi}$$

The above expressions can be used in (3.162) to evaluate the capacitance C for a microstrip line with an isotropic or anisotropic dielectric substrate. The effective dielectric constant ϵ_e for a microstrip line is given by

$$\epsilon_e = \frac{C}{C_a} \qquad (3.164)$$

where C_a is the capacitance of the unscaled air-filled line. The characteristic impedance is given by

$$Z_c = \sqrt{\frac{L}{C}} = \frac{\sqrt{\epsilon_e}\sqrt{\mu_0\epsilon_0}}{C} = \sqrt{\frac{\mu_0\epsilon_0}{\epsilon_e}}\,\frac{1}{C_a} \qquad (3.165)$$

Even though computations based on (3.162) are straightforward, there is an easier way to find ϵ_e. Schneider has presented a remarkably simple formula for the effective dielectric constant of a microstrip line with an isotropic substrate.‡ This formula was modified by Hammerstad to improve

†M. A. R. Gunston and J. R. Weale, Variation of Microstrip Impedance with Strip Thickness, *Electron. Lett.*, vol. 5, p. 697, 1969.

‡M. V. Schneider, Microstrip Lines for Microwave Integrated Circuits, *Bell System Tech. J.*, vol. 48, pp. 1422–1444, 1969. Schneider's formula has a numerical coefficient of 10 instead of 12 multiplying the H/W term.

the accuracy.† The modified formula, which we will refer to as the S-H formula, is

$$\epsilon_e = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2}\left(1 + \frac{12H}{W}\right)^{-1/2} + F(\epsilon_r, H) - 0.217(\epsilon_r - 1)\frac{T}{\sqrt{WH}} \tag{3.166}$$

where $F(\epsilon_r, H) = 0.02(\epsilon_r - 1)(1 - W/H)^2$ for $W/H < 1$ and equals zero for $W/H > 1$. The last term accounts for the reduction in ϵ_e caused by the finite thickness of the microstrip. We have checked the accuracy of this formula against the results obtained by solving the integral equation and found the agreement to be better than 1 percent for $0.25 \leq W/H \leq 6$ and $1 < \epsilon_r \leq 16$. We can also adapt the S-H formula to treat the case of an anisotropic substrate as follows: For an anisotropic substrate we replace the spacing parameter H by the effective spacing H_e given by

$$H_e = \sqrt{\frac{\epsilon_r}{\epsilon_y}}\, H$$

and use the geometric mean $\epsilon_g = \sqrt{\epsilon_r \epsilon_y}$ for the dielectric constant ϵ_r. The S-H formula gives the capacitance of a microstrip line with spacing H_e and dielectric constant ϵ_g relative to an air-filled line with the same spacing H_e; thus

$$\frac{C(\epsilon_g, H_e)}{C_a(H_e)} = \frac{\epsilon_g + 1}{2} + \frac{\epsilon_g - 1}{2}\left(1 + 12\frac{H_e}{W}\right)^{-1/2} + F(\epsilon_g, H_e)$$

The effective dielectric constant is, however, given by the ratio of $C(\epsilon_g, H_e)$ to the capacitance of the air-filled unscaled line according to (3.164). Hence, for the case of an anisotropic substrate, we have

$$\epsilon_e = \left[\frac{\epsilon_g + 1}{2} + \frac{\epsilon_g - 1}{2}\left(1 + 12\frac{H_e}{W}\right)^{-1/2} + F(\epsilon_g, H_e)\right]\frac{C_a(H_e)}{C_a(H)} \tag{3.167}$$

The capacitances for the air-filled lines are readily computed using (3.163).

For comparison purposes, typical results obtained for ϵ_e using (3.162), formulas (3.166) or (3.167), and those obtained by solving the integral equation are given in Table 3.3. Also listed are the number of terms needed in the formula (3.162) to give a numerical convergence of 0.3 percent. Overall, all three methods give values for ϵ_e that are in close agreement. Equation (3.162) gives values that are on the high side for wide strips and substrates with large dielectric constants. This is caused by the variation in charge density with strip spacing above the ground plane, which is more

†E. O. Hammerstad, *loc. cit.* (1975 paper).

TABLE 3.3
Comparison of values for effective dielectric constant using different formulas

	Integral equation	Eq. (3.162)	Eq. (3.166)	Integral equation	Eq. (3.162)	Eq. (3.166)
W/H		$\epsilon_r = 2$			$\epsilon_r = 5.12, \epsilon_y = 3.4$†	
0.25	1.588	1.589 (7)‡	1.583	2.671	2.675 (15)	2.69
0.5	1.61	1.612 (7)	1.605	2.694	2.698 (15)	2.721
1	1.645	1.649 (7)	1.639	2.731	2.734 (15)	2.731
2	1.696	1.699 (7)	1.689	2.797	2.802 (16)	2.794
4	1.762	1.761 (8)	1.75	2.906	2.929 (16)	2.890
6	1.801	1.799 (8)	1.789	2.979	2.944 (16)	2.963
		$\epsilon_r = 6$			$\epsilon_r = 10$	
0.25	3.896	3.896 (22)	3.913	6.195	6.192 (36)	6.244
0.5	4.003	4.004 (22)	4.025	6.387	6.40 (37)	6.445
1	4.173	4.169 (22)	4.193	6.69	6.70 (37)	6.748
2	4.428	4.447 (23)	4.444	7.15	7.16 (38)	7.201
4	4.763	4.837 (23)	4.75	7.757	7.946 (39)	7.750
6	4.966	5.012 (23)	4.943	8.127	8.303 (40)	8.098

†Boron nitride.
‡The numbers in parentheses are the number of terms used in the numerical solution.

pronounced for wide strips. From the tabulated results it can be inferred that the modified Schneider's formula will be acceptable for most applications since even the dielectric constant of the substrate is often not known to an accuracy much better than 1 percent. The following two examples will illustrate the application of the above formulas to microstrip lines.

Example 3.4. A microstrip line uses a substrate with dielectric constant $\epsilon_r = 9.7$ (alumina) and thickness 0.5 mm. The strip width is also 0.5 mm. We wish to find the effective dielectric constant, the characteristic impedance, and the microstrip wavelength at a frequency of 2 GHz.

Since the substrate is isotropic, we use (3.166) to find ϵ_e. Thus since $W/H = 1$,

$$\epsilon_e = \frac{10.7}{2} + \frac{8.7}{2}(1 + 12)^{-1/2} = 6.556$$

In order to evaluate Z_c using (3.165), we first find C_a using (3.163). Thus

$$\frac{C_a}{\epsilon_0} = \frac{2\pi}{\ln(8 + \frac{1}{4})} = 2.978$$

From (3.165) we get

$$Z_c = \frac{Z_0}{2.978\sqrt{\epsilon_e}} = \frac{120\pi}{2.978\sqrt{6.556}} = 49.44\ \Omega$$

At 2 GHz the propagation constant is $\beta = \sqrt{\epsilon_e}\, 2\pi/\lambda_0 = 1.0725$ rad/cm. Hence $\lambda = 2\pi/\beta = 5.858$ cm. The wavelength can also be found using $\lambda = \lambda_0/\sqrt{\epsilon_e}$.

Example 3.5. A microstrip line uses a sapphire substrate 1 mm thick and having $\epsilon_r = 9.4$, $\epsilon_y = 11.6$. We want to find the effective dielectric constant and characteristic impedance for the case when the strip is 0.5 mm wide.

Since this substrate is anisotropic, we first find $\epsilon_g = \sqrt{\epsilon_r \epsilon_y} = 10.44$ and $H_e = \sqrt{\epsilon_r/\epsilon_y}\, H = 0.9$ mm. We now use (3.167) to find ϵ_e. By using (3.163a) we get the ratio

$$\frac{C_a(H_e)}{C_a(H)} = \frac{2.347}{2.26} = 1.0385$$

From (3.167)

$$\epsilon_e = 1.0385\left[\frac{11.44}{2} + \frac{9.44}{2}\left(1 + 12\frac{0.9}{0.5}\right)^{-1/2} + 0.02 \times 9.44 \times \tfrac{1}{4}\right]$$

$$= 7.02$$

From (3.163a) we get $C_a(H) = 2.26\epsilon_0$ and using (3.165) gives

$$Z_c = \frac{120\pi}{2.26\sqrt{7.02}} = 62.96\ \Omega$$

Microstrip Attenuation

Dielectric losses and conductor losses will introduce attenuation. The attenuation caused by the finite conductivity of the conductors is accounted for by the series resistance R, while attenuation caused by dielectric loss is modeled by the shunt conductance G in the distributed circuit model of the microstrip line. The separate attenuation constants are given by

$$\alpha_c = \frac{R}{2Z_c} \tag{3.168a}$$

$$\alpha_d = \frac{G}{2Y_c} \tag{3.168b}$$

and the total attenuation is given by

$$\alpha = \alpha_c + \alpha_d \tag{3.168c}$$

The attenuation in decibels per unit length is obtained by multiplying α by 8.686.

We will first examine the attenuation caused by dielectric loss. The dielectric loss arises when the permittivity ϵ is complex, that is, $\epsilon = \epsilon' - j\epsilon''$. The loss tangent

$$\tan\delta_l = \frac{\epsilon''}{\epsilon'} \approx \delta_l$$

is the usual given parameter for a dielectric material. Maxwell's equation

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E} + \sigma\mathbf{E} = j\omega\epsilon'\mathbf{E} + (\omega\epsilon'' + \sigma)\mathbf{E}$$

shows that $\omega\epsilon''$ can be viewed as the effective conductivity of a lossy dielectric when the ohmic conductivity σ is zero. Normally we can assume that $\sigma = 0$ except for a semiconductor substrate, in which case σ will depend on the doping level.

The electric energy stored in the substrate region of the microstrip line is given by

$$W_{e1} = \frac{\epsilon'}{4}\int_{V_1} \mathbf{E} \cdot \mathbf{E}^* \, dV$$

where V_1 is the volume of the substrate region per unit length of line. The power loss due to dielectric loss is given by

$$P_l = \frac{\omega\epsilon''}{2}\int_{V_1} \mathbf{E} \cdot \mathbf{E}^* \, dV$$

Thus we see that

$$\frac{P_l}{W_{e1}} = \frac{2\omega\epsilon''}{\epsilon'}$$

If the dielectric filled all of the space around the microstrip, we could equate W_{e1} to $CV^2/4$ and P_l to $GV^2/2$ and thereby obtain

$$G = \frac{\omega\epsilon''}{\epsilon'}C$$

However, for a partially filled line some of the electric energy is located in the air region that occupies a volume we will call V_2. Consequently, we have

$$W_{e1} + W_{e2} = \frac{\epsilon'}{4}\int_{V_1} \mathbf{E} \cdot \mathbf{E}^* \, dV + \frac{\epsilon_0}{4}\int_{V_2} \mathbf{E} \cdot \mathbf{E}^* \, dV = \frac{CV^2}{4}$$

If we had an air-filled line, we could write

$$W_{e1} + W_{e2} = \frac{\epsilon_0}{4}\int_{V_1} \mathbf{E} \cdot \mathbf{E}^* \, dV + \frac{\epsilon_0}{4}\int_{V_2} \mathbf{E} \cdot \mathbf{E}^* \, dV = \frac{C_a V^2}{4}$$

with the understanding that the electric field in the two cases will not be the same. There is no simple exact way to determine how the electric energy is split between the two regions. There is, however, an approximate method to find the division of the total energy between the two regions and this is based on the assumption that the volume integrals of $\mathbf{E} \cdot \mathbf{E}^*$ in the two cases are approximately the same. By making this assumption we can write

$$\epsilon' I_1 + \epsilon_0 I_2 = CV^2$$

$$\epsilon_0 I_1 + \epsilon_0 I_2 = C_a V^2$$

where I_1 and I_2 represent the values of the integrals over V_1 and V_2, respectively. The above two equations can be solved for I_1 and I_2 to give

$$I_1 = \frac{C - C_a}{\epsilon' - \epsilon_0} V^2 \qquad I_2 = \frac{\epsilon'_r C_a - C}{\epsilon' - \epsilon_0} V^2$$

where $\epsilon'_r = \epsilon'/\epsilon_0$. The fraction of the total energy in the dielectric region is

$$\frac{\epsilon'_r I_1}{\epsilon'_r I_1 + I_2} = \frac{\epsilon'_r (C - C_a)}{(\epsilon'_r - 1)C} = \frac{\epsilon'_r}{\epsilon_e} \frac{\epsilon_e - 1}{\epsilon'_r - 1} = q \frac{\epsilon'_r}{\epsilon_e} \tag{3.169}$$

where we have used $C = \epsilon_e C_a$. The parameter q is called the filling factor. If q was found independently, then we could solve for ϵ_e to get $\epsilon_e = \epsilon'_r q + (1 - q)$. The parameter q is the ratio of the integral of $\mathbf{E} \cdot \mathbf{E}^*$ over the volume V_1 to the integral over the total volume $V_1 + V_2$, that is, $q = I_1/(I_1 + I_2)$ and clearly represents a filling factor.

With the above assumption regarding no change in the volume integrals for the two cases, we see that G, as given earlier, should be reduced by the same fraction by which the total electric energy was split since there is essentially no loss in the air region. Hence an estimate for G is

$$G = \frac{\epsilon'_r}{\epsilon_e} \frac{\epsilon_e - 1}{\epsilon'_r - 1} \frac{\omega \epsilon''_r}{\epsilon'_r} C \tag{3.170}$$

By using $Z_c = \sqrt{\epsilon_e}\sqrt{\mu_0 \epsilon_0}/C$ we obtain

$$\alpha_d = \frac{GZ_c}{2} = \frac{\pi}{\lambda_0} \frac{\epsilon'_r}{\sqrt{\epsilon_e}} \frac{\epsilon_e - 1}{\epsilon'_r - 1} \tan \delta_l \tag{3.171}$$

for the attenuation constant due to dielectric loss. In the derivation we used $\omega\sqrt{\mu_0 \epsilon_0} = k_0 = 2\pi/\lambda_0$. As an example if $\epsilon'_r = 9.7$, $\epsilon_e = 6.55$, and $\tan \delta_l = 2 \times 10^{-4}$ we get $\alpha_d = 1.52 \times 10^{-3}$ Np/wavelength. In decibel units this equals 0.013 dB/wavelength, which is a relatively small value. The attenuation caused by conductor losses will be significantly larger. Equation (3.171) is valid for isotropic substrates only.

We now turn our attention to evaluating the attenuation caused by finite conductivity of the microstrip and the ground plane. The continuity equation (3.138b) shows that the current density J_z along the conductors varies the same way as the charge density. Thus on a conducting strip of width W the current density will be similar to the charge density given by (3.156). At the edge of an infinitely thin strip, the current density will increase inversely proportional to the square root of the distance from the edge and becomes infinite at the edge. However, the density can be integrated to give a finite value for the total current. But since power-loss calculations require integrating the square of the current density, we would find that for the current density on an infinitely thin strip we would obtain infinite power loss. In practice, the conductors have a finite thickness and the current density is less singular at the edge and the power loss is finite.

Consequently, it is necessary to take into account the finite thickness of the conductors. In addition, it is necessary to determine how the total current divides between the two faces of the microstrip since the presence of a ground plane results in an unequal division of the current for strip widths greater than one-half of the spacing above the ground plane.

The current distribution, current division, and power loss can be evaluated using conformal mapping techniques. In order to obtain useful formulas, some approximations are necessary. The analysis for a microstrip line is carried out in App. III and the results obtained are repeated here (note that in App. III the strip width is $2w$ and the thickness is $2t$, whereas here we use W for the width and T for the thickness). The normalized series distributed resistance for the microstrip is R_1 where

$$\frac{R_1 W}{R_m} = \mathrm{LR}\left(\frac{1}{\pi} + \frac{1}{\pi^2}\ln\frac{4\pi W}{T}\right) \tag{3.172}$$

The loss ratio LR is given by

$$\begin{aligned} \mathrm{LR} &= 1 && \frac{W}{H} \le 0.5 \\ \mathrm{LR} &= 0.94 + 0.132\frac{W}{H} - 0.0062\left(\frac{W}{H}\right)^2 && 0.5 < \frac{W}{H} \le 10 \end{aligned} \tag{3.173}$$

The loss ratio gives the increase in resistance that results from an unequal division of the current. The normalized series resistance R_2 of the ground plane is given by

$$W\frac{R_2}{R_m} = \frac{W/H}{W/H + 5.8 + 0.03H/W} \qquad 0.1 \le \frac{W}{H} \le 10 \tag{3.174}$$

This formula states that the effective width of the ground plane is $W + 5.8H$ and having uniform current density. The skin-effect resistance R_m is given by $R_m = (\omega\mu/\sigma)^{1/2}$. For copper with a conductivity of 5.8×10^7 S/m, we have $R_m = 8.22 \times 10^{-3}\sqrt{f}\ \Omega$ where f is in gigahertz. The total series resistance is $R_1 + R_2$ and thus upon using (3.34*b*) we get

$$\alpha_c = \frac{R_1 + R_2}{2Z_c} \tag{3.175}$$

for the attenuation caused by conductor losses. For the quasi-TEM mode the magnetic field, and hence the conductor losses, do not depend on the substrate material.

The equations presented above predict somewhat higher theoretical attenuation than that obtained from a formula developed by Pucel, Masse, and Hartwig using Wheeler's incremental inductance rule.† Our formulas

†R. A. Pucel, D. J. Masse, and C. P. Hartwig, Losses in Microstrip, *IEEE Trans.*, vol. MTT-16, pp. 342–350, June, 1966.

appear to be in better agreement with experimental results. For practical microstrip lines surface roughness can increase the attenuation by as much as 50 percent or more depending on the scale of surface roughness relative to the skin depth. The etching process does not produce a perfectly flat end face at the sides of the strip. Some undercutting of the edge occurs along with some roughness, which will also result in an increase of the attenuation above the theoretical values.

Example 3.6. We wish to find the attenuation for a microstrip line using a copper strip of width 0.5 mm, a spacing of 0.5 mm above the ground plane, an alumina substrate with $\epsilon_r = 9.7$, and a loss tangent of 2×10^{-4}. The strip thickness T is 0.02 mm. The frequency of operation is 4 GHz.

The effective dielectric constant and characteristic impedance were found in Example 3.4 and are $\epsilon_e = 6.556$, $Z_c = 49.44\ \Omega$. The wavelength of operation is 7.5 cm. The attenuation caused by dielectric loss was calculated after (3.171) was presented and is 1.52×10^{-3} Np/wavelength or 2.02×10^{-4} Np/cm.

For this microstrip line $W/H = 1$ and $W/T = 25$. From (3.173) we find that the loss ratio is 1.0658 and (3.172) gives

$$R_1 = \frac{1.0658 \times 8.22 \times 10^{-3}\sqrt{4}}{0.05}\left(\frac{1}{\pi} + \frac{1}{\pi^2}\ln 100\pi\right)$$

$$= 3.157 \times 10^{-1}\ \Omega/\text{cm}$$

By using (3.174) we obtain

$$R_2 = \frac{8.22 \times 10^{-3}\sqrt{4}}{0.05}\left[\frac{1}{1 + 5.8 + 0.03}\right] = 4.81 \times 10^{-2}\ \Omega/\text{cm}$$

For this case the loss in the microstrip is a factor of 6.5 greater than that in the ground plane. The reason for this is the high current density near the edges in the microstrip as compared with a more uniform current distribution over a wider area on the ground plane. By using (3.175) we find that the attenuation due to conductor loss is

$$\alpha_c = \frac{3.157 \times 10^{-1} + 4.81 \times 10^{-2}}{2 \times 49.44} = 3.68 \times 10^{-3}\ \text{Np/cm}$$

This attenuation is 18.4 times greater than that caused by dielectric loss. The total attenuation is

$$\alpha = \alpha_c + \alpha_d = 3.88 \times 10^{-3}\ \text{Np/cm}$$

which equals 0.0337 dB/cm. Surface roughness could result in a real attenuation 15 to 25 percent higher than this at 4 GHz. In this example we neglected the correction to ϵ_e and Z_c due to the finite thickness T since it is small.

High-Frequency Properties of Microstrip Lines

The equations describing the quasi-TEM mode can be used with acceptable accuracy for frequencies up to 2 to 4 GHz for a substrate thickness of 1 mm. For a substrate 0.5 mm thick, the upper frequency limit would be 4 to 8 GHz. When these limits are exceeded, it is necessary to take into account the frequency dispersion of the effective dielectric constant and the change in characteristic impedance with frequency. At the higher frequencies the electric field becomes more confined to the region between the microstrip and ground plane. The greater concentration of the field in this region results in an increase in the effective dielectric constant as well as increased

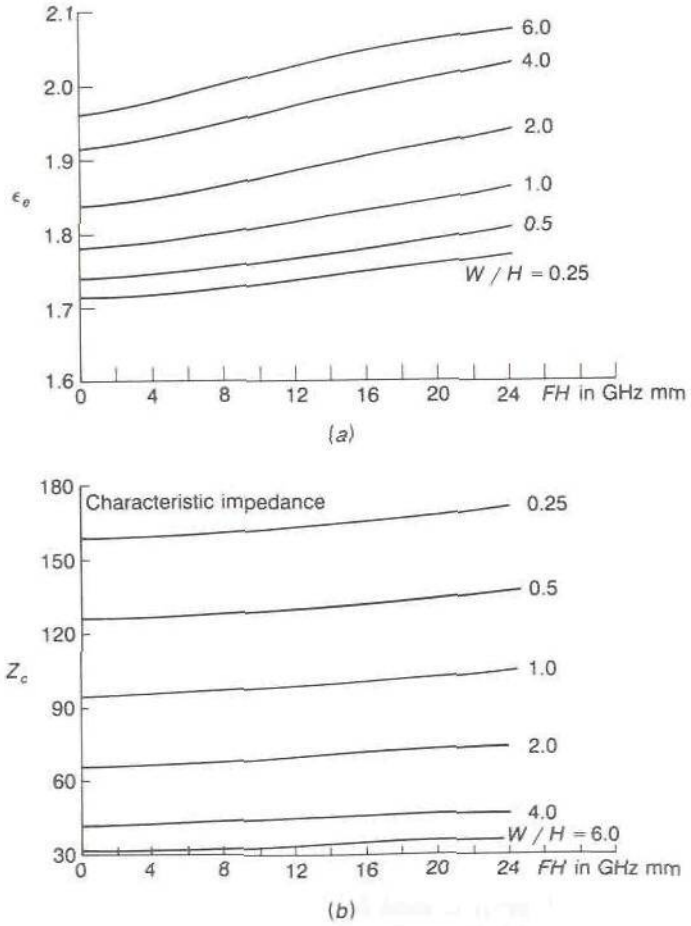


FIGURE 3.25
(*a*) Effective dielectric constant for a PTFE/microfiber glass substrate with $\epsilon_r = 2.26$, $\epsilon_y = 2.2$; (*b*) characteristic impedance.

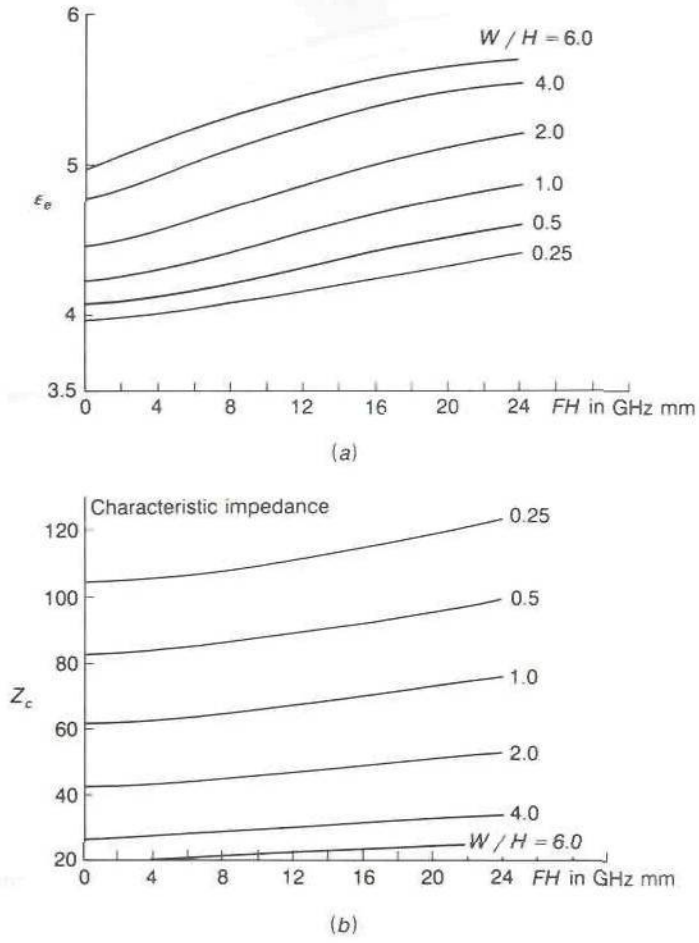
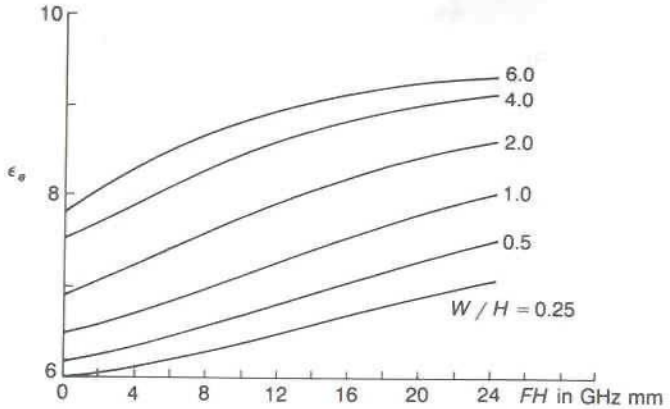


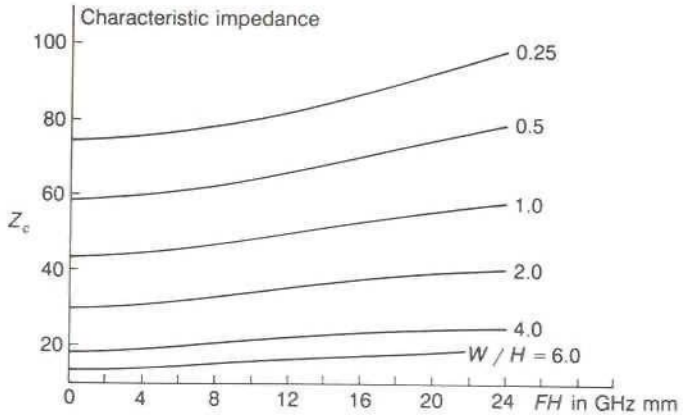
FIGURE 3.26
(*a*) Effective dielectric constant for an RT/Duroid 6006 substrate having $\epsilon_r = 6.36$, $\epsilon_y = 6$; (*b*) characteristic impedance.

attenuation because of a greater concentration of the electric field in the substrate which has some loss. The conductor loss also increases because the skin-effect resistance R_m increases and more of the current flows on the inner face of the microstrip.

In order to determine the effective dielectric constant at high frequencies, it is necessary to carry out a full wave analysis, i.e., the complete set of equations given earlier for the potentials must be solved. In Figs. 3.25 to 3.28 we show the dispersive properties for four common substrate materials, a PTFE/microfiber glass substrate with $\epsilon_r = 2.26$, $\epsilon_y = 2.2$, RT/Duroid 6006 with $\epsilon_r = 6.36$, $\epsilon_y = 6$, alumina with $\epsilon_r = 9.7$, and gallium arsenide



(a)



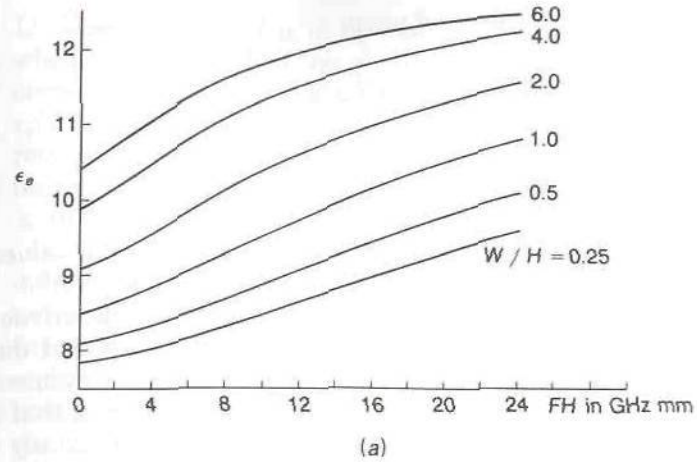
(b)

FIGURE 3.27
(*a*) Effective dielectric constant for an alumina substrate with $\epsilon_r = 9.7$; (*b*) characteristic impedance.

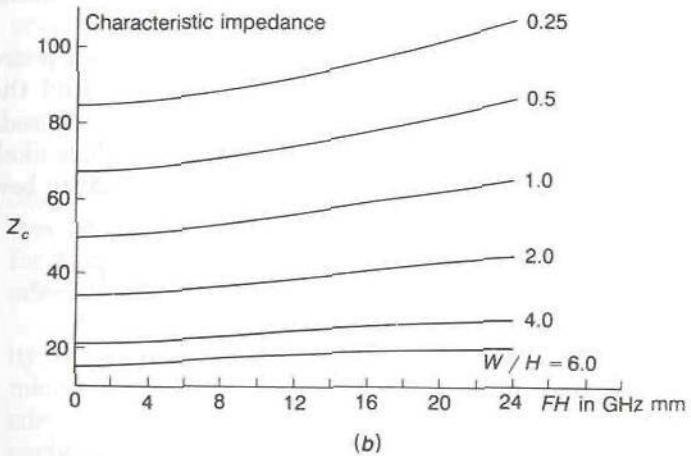
with $\epsilon_r = 12.9$. An examination of these figures shows that dispersion effects are more pronounced for wider strips and large dielectric constants.

Along with each figure giving the effective dielectric constant as a function of normalized frequency is a figure showing the characteristic impedance as a function of normalized frequency. The normalized frequency is the actual frequency in gigahertz multiplied by the substrate thickness in millimeters. Thus, for substrates 0.5 mm thick, the frequency range covered is 0 to 48 GHz.

When the propagating mode is not a TEM or quasi-TEM mode, there is no unique value for the characteristic impedance because it is not possible to define a unique value for the voltage as given by the line integral of the electric field between the ground plane and the microstrip. The characteris-



(a)



(b)

FIGURE 3.28
(*a*) Effective dielectric constant for gallium-arsenide substrate with $\epsilon_r = 12.9$; (*b*) characteristic impedance.

tic impedances given in Figs. 3.25 to 3.28 are based on using the following definition for the equivalent voltage:

$$V = -\int_0^H E_y \, dy = \int_0^H \left(j\omega A_y + \frac{\partial \Phi}{\partial y} \right) dy$$

where the path of integration is along a straight line from the ground plane to the center of the microstrip. The current I is chosen as the total z-directed current on the microstrip and Z_c was calculated from the ratio V/I. In general, the power flow along the microstrip transmission line will not equal $\frac{1}{2}VI$. The characteristic impedance can be defined in terms of the power flow P by choosing either the current I or voltage V according to the

above definitions and using

$$P = \frac{I^2 Z_c}{2}$$

$$= \frac{V^2}{2Z_c}$$

to find Z_c. These two equations will give different values for the characteristic impedance with neither one being equal to V/I.

The lack of a unique value for the characteristic impedance is not a great disadvantage since microstrip junctions and discontinuities can be described by equivalent circuits using any convenient definition for the characteristic impedance. In Chap. 4 we will find that an equivalent transmission-line circuit theory can be formulated for any waveguiding system and does not require that there be a unique characteristic impedance associated with the propagating modes.

For computer-aided design (CAD) of microstrip circuits, it is important to have simple formulas that can be used to find the effective dielectric constant. Many different formulas have been proposed. The most accurate one that covers the full range of parameter values likely to be encountered was developed empirically by Kobayshi and is given below:†

$$\epsilon_e(f) = \epsilon_r - \frac{\epsilon_r - \epsilon_e(0)}{1 + (f/f_a)^m} \tag{3.176}$$

where

$$f_a = \frac{f_b}{0.75 + (0.75 - 0.332\epsilon_r^{-1.73})W/H}$$

$$f_b = \frac{47.746}{H\sqrt{\epsilon_r - \epsilon_e(0)}} \tan^{-1} \epsilon_r \sqrt{\frac{\epsilon_e(0) - 1}{\epsilon_r - \epsilon_e(0)}}$$

$$m = m_0 m_c \leq 2.32$$

$$m_0 = 1 + \frac{1}{1 + \sqrt{W/H}} + 0.32\left(1 + \sqrt{W/H}\right)^{-3}$$

$$m_c = \begin{cases} 1 + \dfrac{1.4}{1 + W/H}(0.15 - 0.235e^{-0.45f/f_a}) & \dfrac{W}{H} \leq 0.7 \\ 1 & \dfrac{W}{H} > 0.7 \end{cases}$$

†M. Kobayashi, A Dispersion Formula Satisfying Recent Requirements in Microstrip CAD, *IEEE Trans.*, vol. MTT-36, pp. 1246–1250, August, 1988.

In these formulas H is in millimeters, the frequency f is in gigahertz, and whenever the product $m_0 m_c$ is greater than 2.32 the parameter m is chosen equal to 2.32. The effective dielectric constant at the frequency f is $\epsilon_e(f)$ and $\epsilon_e(0)$ is the quasistatic value which can be found using (3.166). It requires only a simple computer program to evaluate $\epsilon_e(f)$ using Kobayshi's formula. The accuracy is estimated to be within 0.6 percent for $0.1 \leq W/H \leq 10$, $1 \leq \epsilon_r \leq 128$ and for any value of H/λ_0.†

In microstrip circuit design where junctions of microstrip lines with different widths are involved, it is necessary to characterize the junction in terms of an equivalent circuit. The parameters of this equivalent circuit will depend on frequency. The equivalent characteristic impedances that are assigned to the microstrip lines are arbitrary and often are simply chosen to have normalized values of unity. Any impedance level change that occurs at the junction is incorporated as part of the equivalent circuit of the junction. For these reasons we will not quote any of the formulas that have been proposed for evaluating characteristic impedance as a function of frequency because they are of limited use in practice.

Attenuation

Simple formulas giving the attenuation of microstrip lines at high frequencies do not exist. Equation (3.171) can be expected to give a good estimate for the attenuation due to dielectric loss provided ϵ_e is replaced by the effective dielectric constant that applies at the frequency of interest.

A realistic evaluation of the attenuation caused by the finite conductivity of the conductors requires evaluation of the current density on the microstrip and the ground plane at the frequency of interest. In general, the current tends to be more uniform across the microstrip at high frequencies, particularly for wide strips. In Fig. 3.29 we show several computed current distributions at frequencies of 10 and 20 GHz for an alumina substrate 1 mm thick. The quasistatic current distribution is also shown. The current density has been normalized so that the total current on the microstrip equals W. In view of the tendency for the current density to become nearly uniform at high frequencies for wide strips, the attenuation constant can be estimated with fair accuracy by assuming uniform current density over a width W on both the microstrip and the ground plane. In this limit the attenuation caused by the conductor loss for wide strips that are not too thin is given by

$$\alpha_c = \frac{R_m}{WZ_c} \tag{3.177}$$

†The author has verified the accuracy of Kobayshi's formula by comparison with calculated numerical results for $0.25 \leq W/H \leq 6$ and $2 \leq \epsilon_r \leq 12$.

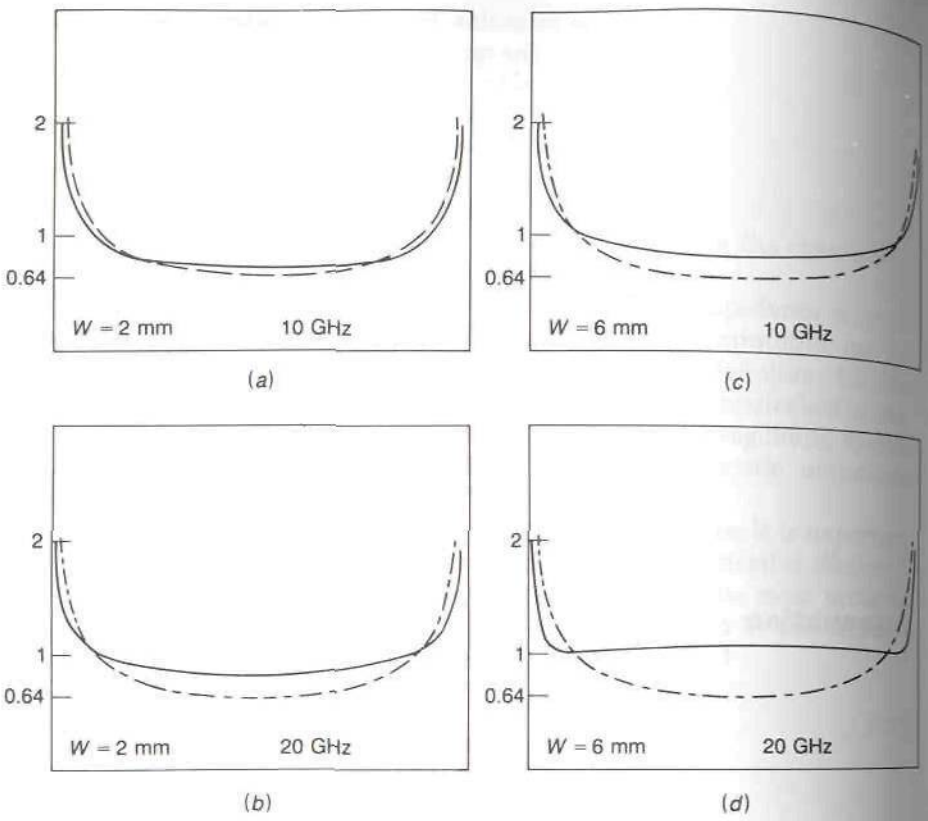


FIGURE 3.29
Current distribution on the microstrip for an alumina substrate with $H = 1$ mm and two different widths. The broken curves give the quasistatic distribution. (a) $W = 2$ mm, $f = 10$ GHz; (b) $W = 2$ mm, $f = 20$ GHz; (c) $W = 6$ mm, $f = 10$ GHz; (d) $W = 6$ mm, $f = 20$ GHz.

For narrow strips and high frequencies, no simple formulas for attenuation appear to be available. For narrow strips, say $W/H < 1$, the quasistatic formula (3.175) is probably a reasonably good estimate since the current density does not depart significantly from the quasistatic distribution for narrow strips.

3.13 COUPLED MICROSTRIP LINES

When two conducting strips of width W are placed side by side on a dielectric substrate above a ground plane as shown in Fig. 3.30a, we obtain a coupled microstrip line. Since this is a three-conductor transmission line, there are two fundamental quasi-TEM modes of propagation. The even mode is the mode corresponding to both strips being at the same potential V

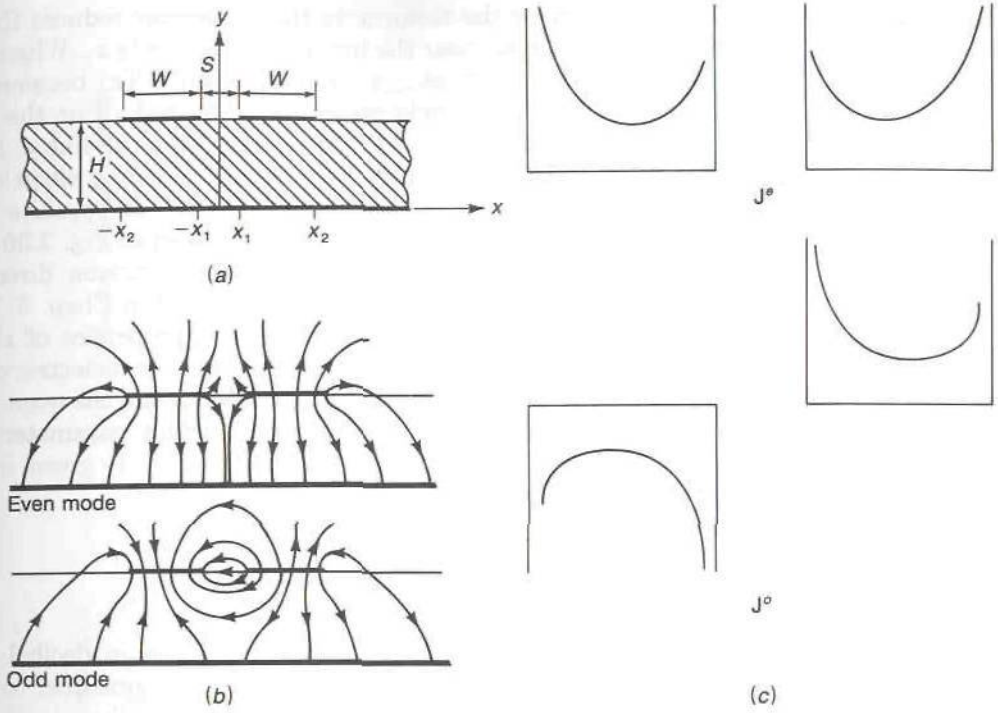


FIGURE 3.30
(*a*) Coupled microstrip line; (*b*) the electric field distribution for the even and odd odes; (*c*) the current distribution for the even and odd modes.

and on which the same currents exist. The odd mode corresponds to the strips being at opposite potentials, $-V$ and V, relative to the ground plane. For the odd mode the currents on the two strips are also equal in amplitude but of opposite sign. A sketch of the electric field lines for the two modes is shown in Fig. 3.30*b*. For isolated strips in air, i.e., with no ground plane and substrate present, the theoretical current distributions for the two modes are:†

$$J^e(x) = \frac{x}{\sqrt{(x^2 - x_1^2)(x_2^2 - x^2)}} \tag{3.178a}$$

$$J^o(x) = \frac{x_2}{\sqrt{(x^2 - x_1^2)(x_2^2 - x^2)}} \tag{3.178b}$$

†R. E. Collin, "Field Theory of Guided Waves," 2nd ed., chap. 4, IEEE Press, Piscataway, N.J., 1991.

For the even mode the factor x in the numerator reduces the amplitude of the singular behavior near the inner edges at $x = \pm x_1$. When x_1 equals zero the current singularity at $\pm x_1$ vanishes and $J^e(x)$ becomes the expected current density on a single strip $2x_2$ units wide. For the odd mode the current singularity at the inner edges $\pm x_1$ is more like a $1/x$ singularity when x_1 is very small. This is caused by the strong electric field across a very narrow slit with the adjacent conductors at opposite potential. The current distribution for the two modes is shown in Fig. 3.30*c*.

The coupled microstrip line is used in various directional coupler designs and these applications will be discussed in Chap. 6. The important parameters describing the quasi-TEM mode properties of the coupled microstrip line are the even- and odd-mode effective dielectric constants $\epsilon_e^e, \epsilon_e^o$ that determine the two propagation constants, and the even- and odd-mode characteristic impedances Z_c^e, Z_c^o. An important parameter in directional coupler design is the coupling coefficient, which is given in terms of the characteristic impedances of the two modes by

$$C = \frac{Z_c^e - Z_c^o}{Z_c^e + Z_c^o} \tag{3.179}$$

The coupling coefficient is commonly expressed in decibel units, that is, $20 \log C$. In a coupled microstrip line it is not practical to achieve much more than a 2.5: 1 impedance ratio; so strong coupling cannot be realized in a simple coupled microstrip directional coupler. However, other designs are available, so this is not a problem for the microwave circuit engineer.

There are no simple formulas giving the quasi-TEM properties of coupled microstrip lines that have an accuracy comparable to that for microstrip lines. Bahl and Bhartia list formulas characterizing coupled microstrip lines that give results which are acceptable for noncritical applications.† The computer program CMST implements these formulas. We have checked the accuracy against numerical results obtained from a full wave solution for an alumina substrate. The effective dielectric constants were found to be accurate to within 3 percent. The characteristic impedance for the even mode was also found to be accurate to within 3 percent. For the odd-mode characteristic impedance, the error was as large as 8 percent for $W/H = 1$, $S/H = 0.25$, but considerably less for larger values of W/H and S/H.

The dispersion properties for coupled microstrip lines on a variety of different substrates have been computed by Morich.‡ In Figs. 3.31 to 3.33

†J. Bahl and P. Bhartia, "Microwave Solid State Circuit Design," p. 28, John Wiley & Sons, Inc., New York, 1988.

‡M. Morich, Broadband Dispersion Analysis of Coupled Microstrip on Anisotropic Substrates by Perturbation-Iteration Theory, M.S. Thesis, Case Western Reserve University, Cleveland, Ohio, May, 1987.

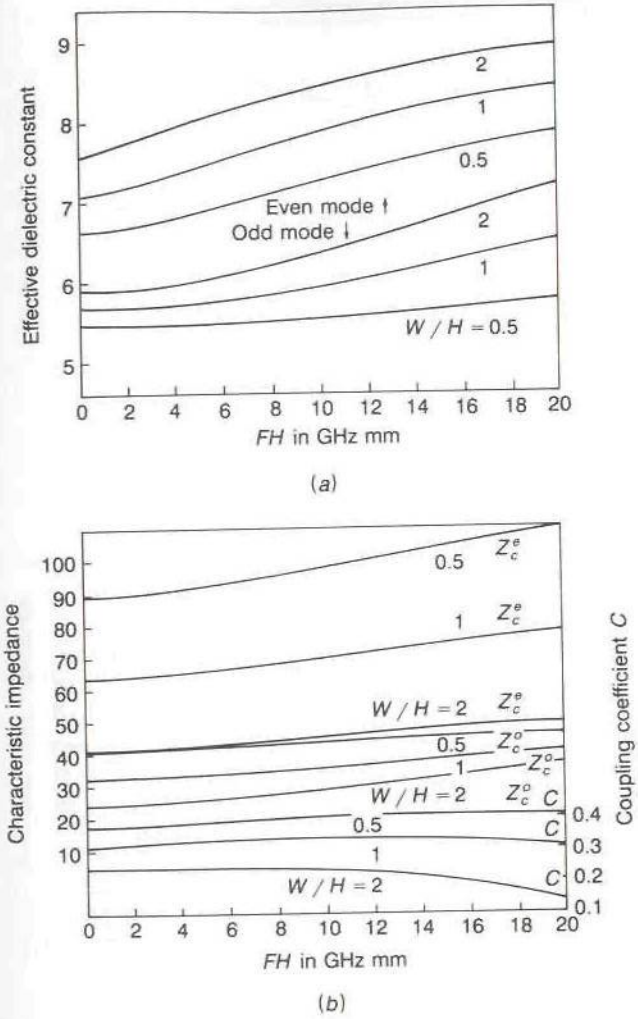


FIGURE 3.31
Dispersion characteristics of a coupled microstrip line on an alumina substrate. $S/H = 0.25$, $\epsilon_r = 9.7$. (*a*) Even- and odd-mode effective dielectric constant; (*b*) even- and odd-mode characteristic impedance and coupling coefficient C.

we show graphical results for coupled microstrip lines on alumina, gallium-arsenide, and RT/Duroid 6006 substrates. The horizontal axis scale is the product of frequency F and substrate thickness H. Thus, for a substrate with $H = 0.5$ mm, the frequency range covered is 0 to 40 GHz. The coupling coefficient C given by (3.179) is also shown. For $W/H = 0.5$ and $S/H = 0.25$, there is little change in the coupling coefficient with frequency. If the spacing is reduced to make $S/H = 0.1$, the coupling coefficient (not shown) for an alumina substrate increases to 0.483 at low frequencies and

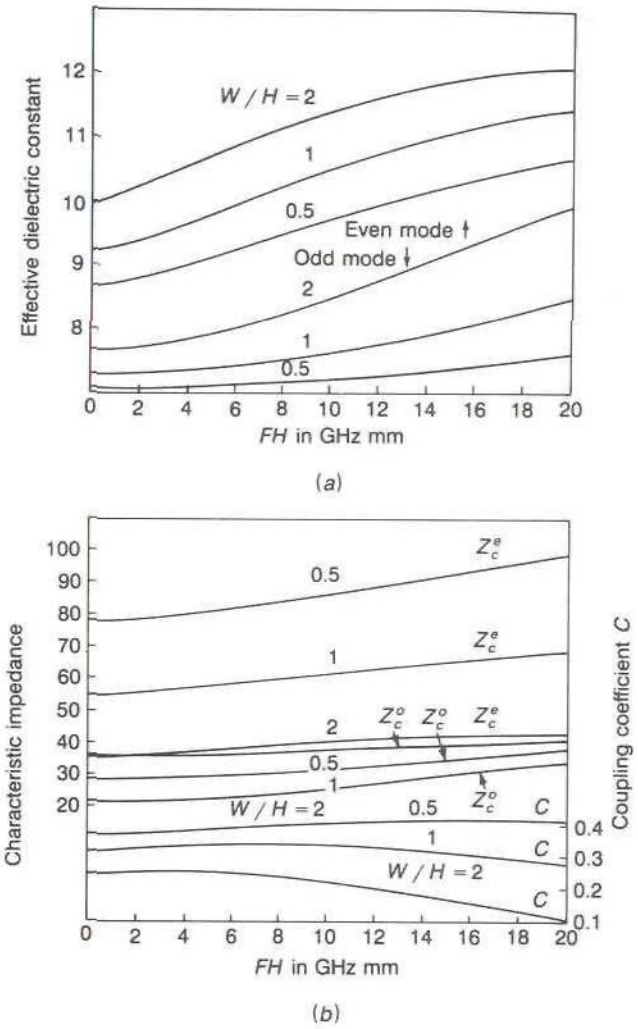


FIGURE 3.32
Dispersion characteristics of coupled microstrip line on a gallium-arsenide substrate. $S/H = 0.25$, $\epsilon_r = 12.9$. (*a*) Even- and odd-mode effective dielectric constant; (*b*) even- and odd-mode characteristic impedance and coupling coefficient C.

rises to a value of 0.524 at $FH = 20$ GHz mm. A coupling coefficient of 0.5 is needed for a 6-dB directional coupler. The effective dielectric constant for the odd mode is smaller than that for the even mode because a larger percentage of the electric field energy is located in the air region. The capacitance between closely spaced parallel strips at opposite potentials is large so the characteristic impedance of the odd mode is smaller than that for the even mode for the normal range of parameters involved. For wide strips with large spacing, there is very little coupling and the two mode impedances will be almost the same.

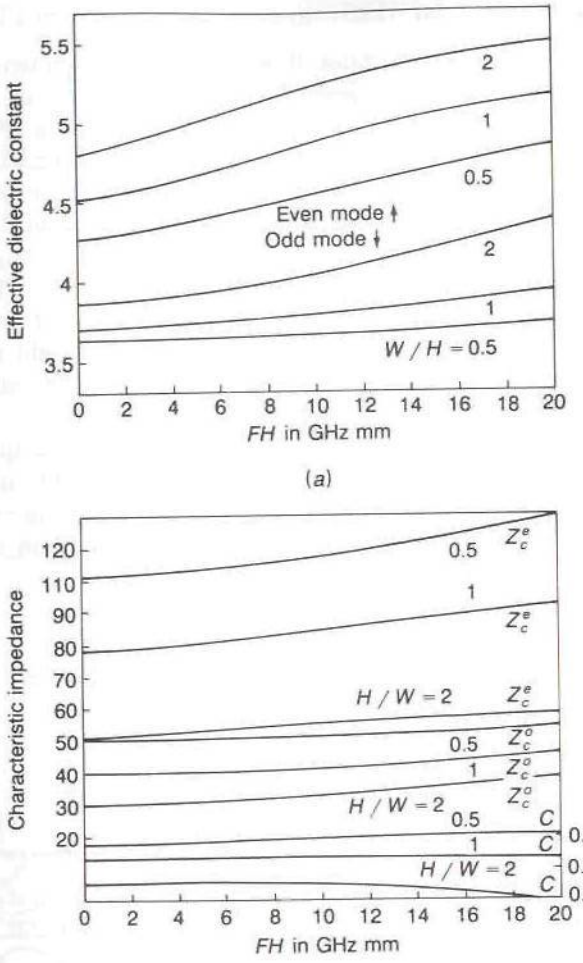


FIGURE 3.33
Dispersion characteristics of coupled microstrip line on an RT/Duroid 6006 substrate. $S/H = 0.25$, $\epsilon_r = 6.36$. (a) Even- and odd-mode effective dielectric constant; (b) even- and odd-mode characteristic impedance and coupling coefficient C.

The attenuation constant for coupled microstrip lines is comparable to that for the microstrip line. For closely spaced strips the increased concentration of the current near the two inner edges for the odd mode, along with the smaller characteristic impedance, increases the attenuation of this mode relative to that for the even mode. For the odd mode the attenuation caused by dielectric loss will be less than that for the even mode since the electric field energy is more evenly distributed between the air region and the substrate region for this mode.

3.14 STRIP TRANSMISSION LINES

The basic strip transmission line consists of a conducting strip embedded in a dielectric medium between two ground planes as shown in Fig. 3.34*a*. Figures 3.34*b* and *c* illustrate coupled strip lines. For broadside coupled strips as shown in Fig. 3.34*c*, the coupling coefficient is significantly greater than for coplanar strips. Thus the strip line with broadside coupled strips is suitable for directional couplers where a coupling of 3 dB is required. Since the dielectric completely surrounds the strips, the strip line and the coupled strip line support pure TEM modes of propagation. There is no frequency dispersion or change in effective dielectric constant with frequency. Consequently, for the coupled strip line the even and odd modes of propagation have the same phase velocity, which is also a desirable feature for directional coupler design.

For the symmetrical strip line and coupled strip line, the TEM-mode characteristic impedance is readily found from the distributed capacitance with the latter determined by conformal mapping techniques. The distributed capacitance for the symmetrical strip line is found in App. III.

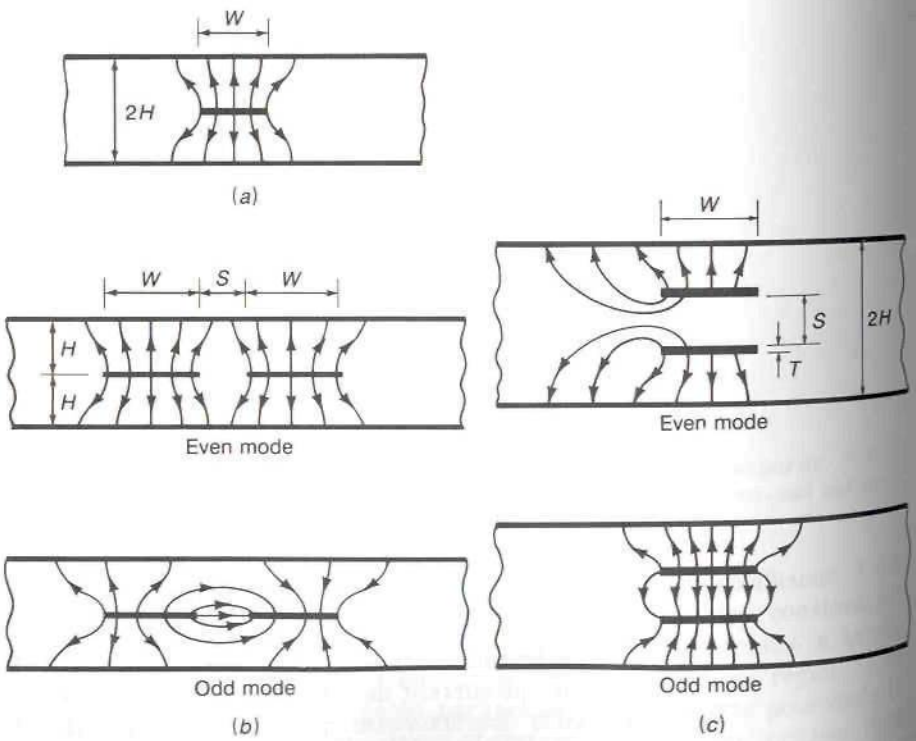


FIGURE 3.34
(*a*) The basic strip-line configuration; (*b*) coupled strip line using coplanar strips; (*c*) coupled strip line using broadside coupled strips. The electric field lines for the TEM modes are also shown.

From the given expression (III.27) and using (III.13c), we have

$$Z_c = \frac{Z_0}{4\sqrt{\epsilon_r'}} \frac{K(k)}{K(k')} = \frac{Z_0 K}{4\sqrt{\epsilon_r'} K'} \tag{3.180a}$$

where

$$k = \frac{1}{\cosh \pi W/4H} \qquad k' = \tanh \pi \frac{W}{4H}$$

$$\frac{K}{K'} = \frac{1}{\pi} \ln\left(2\frac{1+\sqrt{k}}{1-\sqrt{k}}\right) \qquad 0.7 \le k < 1$$

$$\frac{K}{K'} = \left[\frac{1}{\pi} \ln\left(2\frac{1+\sqrt{k'}}{1-\sqrt{k'}}\right)\right]^{-1} \qquad 0 < k \le 0.7$$

and ϵ_r is the dielectric constant of the dielectric material that completely surrounds the center strip. For $W \ge 2H$ the formula for Z_c reduces to the simple form

$$Z_c = \frac{\pi Z_0}{8\sqrt{\epsilon_r'}(\ln 2 + \pi W/4H)} \qquad W \ge 2H \tag{3.180b}$$

For a very narrow strip

$$Z_c = \frac{Z_0}{2\pi\sqrt{\epsilon_r'}} \ln\frac{16H}{\pi W} \qquad W \le 0.4H \tag{3.180c}$$

Attenuation

Since the dielectric material completely fills the strip line, the attenuation due to dielectric loss is given by (3.87a) and is

$$\alpha_d = \frac{\pi\epsilon_r''}{\lambda_0\sqrt{\epsilon_r'}} = \frac{\pi\sqrt{\epsilon_r'}\tan\delta_l}{\lambda_0} \tag{3.181}$$

Formulas for the attenuation in a strip line due to conductor losses and assuming an inner conductor of elliptical cross section have been derived using conformal mapping techniques.† In App. III it is shown that the series resistance of an isolated conductor of elliptical cross section is greater than that for a conductor with a rectangular cross section. By using an equivalent center thickness T_e for the conductor with elliptical cross section, the two series resistances will be equal provided T_e is chosen to be

$$T_e = e^{-\pi/2}\sqrt{\frac{4WT}{\pi}} \tag{3.182}$$

†R. E. Collin, *loc. cit.*, chap. 4, eqs. (73), (74), and (76).

We will assume that this equivalence is also a good approximation for a conductor placed between two ground planes. On this basis the attenuation by the center conductor is

$$\alpha_{c1} = \frac{\pi R_m}{16 Z_c H K'^2 k'} \ln \frac{16Hk'}{k\pi T_e} \tag{3.183a}$$

and that due to the ground planes is

$$\alpha_{c2} = \frac{\pi^2 R_m W}{64 Z_c H^2 K'^2 k'} \tag{3.183b}$$

For wide and narrow strips these formulas reduce to the following simplified forms:

$$\alpha_{c1} = \frac{R_m \sqrt{\epsilon_r'}}{2Z_0 H} \frac{\ln \dfrac{8H}{\pi T_e} + \dfrac{\pi W}{4H}}{\ln 2 + \dfrac{\pi W}{4H}} \qquad W \geq 2H \tag{3.184a}$$

$$\alpha_{c2} = \frac{\pi R_m \sqrt{\epsilon_r'} W}{8 Z_0 H^2 \left(\ln 2 + \dfrac{\pi W}{4H}\right)} \qquad W \geq 2H \tag{3.184b}$$

$$\alpha_{c1} = \frac{2 R_m \sqrt{\epsilon_r'} \ln \dfrac{4W}{T_e}}{\pi Z_0 W \ln \dfrac{16H}{\pi W}} \qquad W \leq 0.4H \tag{3.184c}$$

$$\alpha_{c2} = \frac{R_m \sqrt{\epsilon_r'}}{2 Z_0 H \ln \dfrac{16H}{\pi W}} \qquad W \leq 0.4H$$

Suitable formulas for evaluating K and K' are given by (III.13) in App. III. The total attenuation for a strip line is given by

$$\alpha = \alpha_d + \alpha_{c1} + \alpha_{c2} \tag{3.185}$$

Example 3.7. A strip line has a ground-plane spacing $2H = 1$ cm and uses a centered copper conducting strip of width $W = 1$ cm and thickness $T = 0.002$ cm. The dielectric filling material has a dielectric constant $\epsilon_r' = 2.2$ and a loss tangent equal to 10^{-3}. We want to find the characteristic impedance and attenuation at a frequency of 10 GHz. For this line $W/H = 2$, so (3.180*b*) can be used. Thus

$$Z_c = \frac{120\pi^2}{8\sqrt{2.2}\,(\ln 2 + \pi/2)} = 44.09\ \Omega$$

The wavelength of operation is 3 cm; so by using (3.181) we find the attenuation due to dielectric loss to be

$$\alpha_d = \frac{\pi\sqrt{2.2} \times 10^{-3}}{3}$$

$$= 1.55 \times 10^{-3} \text{ Np/cm} \quad \text{or} \quad 1.35 \times 10^{-2} \text{ dB/cm}$$

From (3.182) we obtain $T_e = 0.00742$ cm for the equivalent thickness. We can use (3.184a) to find

$$\alpha_{c1} = \frac{8.22 \times 10^{-3}\sqrt{10}\sqrt{2.2}}{240\pi \times 0.5} \frac{\ln \dfrac{4}{\pi \times 0.00724} + \dfrac{\pi}{2}}{\ln 2 + \dfrac{\pi}{2}}$$

$$= 3.034 \times 10^{-4} \text{ Np/cm} \quad \text{or} \quad 2.635 \times 10^{-3} \text{ dB/cm}$$

From (3.184b) we obtain

$$\alpha_{c2} = \frac{8.22 \times 10^{-3}\sqrt{10}\sqrt{2.2}\,\pi}{8 \times 120\pi \times 0.5^2(\ln 2 + \pi/2)}$$

$$= 7.096 \times 10^{-5} \text{ Np/cm} \quad \text{or} \quad 6.163 \times 10^{-4} \text{ dB/cm}$$

The total attenuation is 0.0167 dB/cm. For this transmission line the dielectric loss is greater than the conductor loss.

According to a formula for strip-line conductor attenuation developed by Cohn, we would get $\alpha_c = 3.236 \times 10^{-3}$ dB/cm, which agrees very closely with 3.25×10^{-3} dB/cm obtained with our formulas.†

3.15 COUPLED STRIP LINES

For coplanar strips of width W and spacing S, the even- and odd-mode characteristic impedances are given by‡

$$Z_c^e = \frac{Z_0}{4\sqrt{\epsilon_r}} \frac{K(k_e')}{K(k_e)} \tag{3.186a}$$

$$Z_c^o = \frac{Z_0}{4\sqrt{\epsilon_r}} \frac{K(k_o')}{K(k_o)} \tag{3.186b}$$

†See H. Howe, "Strip Line Circuit Design," eq. (1.5), Artech House Books, Dedham, Mass. 1974.

‡S. B. Cohn, Shielded Coupled Strip Transmission Line, *IRE Trans.*, vol. MTT-5, pp. 29–37, October, 1955.

S. B. Cohn, Characteristic Impedances of Broadside-Coupled Strip Transmission Lines, *IRE Trans.*, vol. MTT-8, pp. 633–637, November, 1960.

S. B. Cohn, Thickness Corrections for Capacitive Obstacles and Strip Conductors, *IRE Trans.*, vol. MTT-8, pp. 638–644, November, 1960.

where

$$k_e = \tanh\left(\frac{\pi}{4}\frac{W}{H}\right)\tanh\left(\frac{\pi}{4}\frac{W+S}{H}\right)$$

$$k_o = \tanh\left(\frac{\pi}{4}\frac{W}{H}\right)\coth\left(\frac{\pi}{4}\frac{W+S}{H}\right)$$

and

$$k'_e = \sqrt{1-k_e^2} \qquad k'_o = \sqrt{1-k_o^2}.$$

Note that $K(k') = K'(k)$. For the evaluation of K and K' see (III.13).

For coplanar strips with a thickness T,

$$Z_c^i = \frac{Z_0(2H-T)}{4\sqrt{\epsilon_r}\left(W + (H/\pi)C_f A_i\right)} \qquad i = e, o \qquad W \geq 0.7H \quad (3.187)$$

$$A_e = 1 + \frac{\ln[1+\tanh(\pi S/4H)]}{\ln 2}$$

$$A_o = 1 + \frac{\ln[1+\coth(\pi S/4H)]}{\ln 2}$$

$$C_f = 2\ln\left(\frac{4H-T}{2H-T}\right) - \frac{T}{2H}\ln\left[\frac{T(4H-T)}{(2H-T)^2}\right]$$

For broadside coupled strips as shown in Fig. 3.34*c*, the even- and odd-mode characteristic impedances are given by [valid for $W \geq 0.35S$, $W \geq 0.7H(1 - S/2H)$]

$$Z_c^e = \frac{Z_0}{2\sqrt{\epsilon_r}\left\{\frac{W}{2H-S-2T} + 0.4413 + \frac{1}{\pi}\left[\ln\frac{2H}{2H-S-2T} + \frac{S+2T}{2H-S-2T}\ln\frac{2H}{S+2T}\right]\right\}} \quad (3.188a)$$

$$Z_c^o = \frac{Z_0}{2\sqrt{\epsilon_r}\left\{\frac{W}{2H-S-2T} + \frac{W}{S} + C_f^o + \frac{2}{\pi}\left[\left(1+\frac{T}{S}\right)\ln\left(1+\frac{T}{S}\right) - \frac{T}{S}\ln\frac{T}{S}\right]\right\}} \quad (3.188b)$$

where

$$C_f^o = \frac{2H-2T}{\pi S}\left[\ln\frac{2H-2T}{2H-S-2T} + \frac{S}{2H-S-2T}\ln\frac{2H-2T}{S}\right]$$

All of the characteristic impedances given above refer to the impedance from one strip to the ground planes.

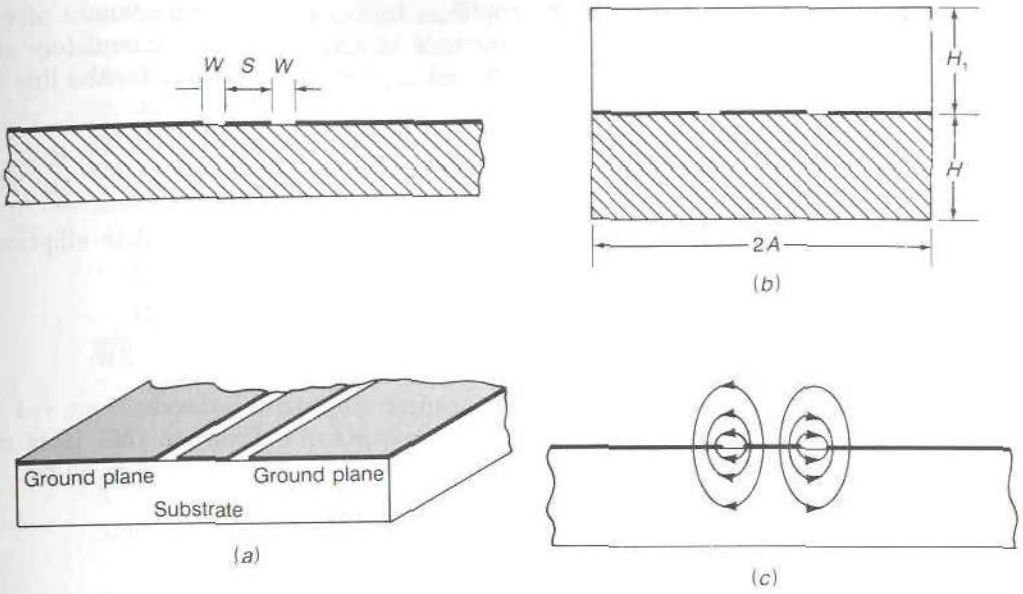


FIGURE 3.35
(*a*) Basic coplanar transmission line; (*b*) a shielded coplanar transmission line; (*c*) electric field distribution.

3.16 COPLANAR TRANSMISSION LINES

An illustration of a coplanar transmission line is shown in Fig. 3.35*a*. A shielded coplanar line is shown in Fig. 3.35*b*. The strip width is S and the strip to ground-plane spacing is W. The coplanar line is often called a coplanar waveguide (CPW). The most significant advantage that a coplanar line has over a microstrip line is the ability to connect active and passive circuit components in shunt from the conducting strip to the ground plane on the same side of the substrate. In a microstrip line a connection to the ground plane requires drilling a hole through the substrate which is somewhat difficult for ceramic materials such as alumina. Figure 3.35*c* shows the electric field distribution in a coplanar line.

The characteristics of a coplanar line at low frequencies can be determined by conformal mapping techniques. A solution for the coplanar line is given in App. III. For the ideal case when the ground planes are very wide relative to the slot width W and the dielectric substrate is very thick, the electric field has the same distribution as for a coplanar line in air. The reason for this is that the mapping of the air-dielectric boundary, shown as the intervals BC and EF in Fig. III.4, is the two parallel sides in the ideal parallel-plate capacitor shown in Fig. III.3. Thus the solution for the potential is not affected by the presence of the dielectric. The distributed

capacitance of the coplanar line is thus the capacitance of one-half of the air-filled line plus one-half of that for a line completely surrounded by dielectric. This capacitance is given by III.14 and for the line with dielectric on one side only we have

$$C = 2\epsilon_0 \frac{K}{K'} + 2\epsilon_r \epsilon_0 \frac{K}{K'} = 4\epsilon_0 \frac{(\epsilon_r + 1)}{2} \frac{K}{K'} \tag{3.189}$$

where $K(k)$ and $K' = K(k')$ are again the complete elliptical integrals of the first kind. The modulus k is given by the ratio

$$k = \frac{1}{1/k} = \frac{S/2}{W + S/2} = \frac{S}{S + 2W} \tag{3.190}$$

since in Fig. III.4 the center conductor extends from -1 to 1 and the ground planes begin at $u = \pm(1/k)$. From (3.189) it is clear that the effective dielectric constant for the coplanar line is given by

$$\epsilon_e = \frac{\epsilon_r + 1}{2} \tag{3.191}$$

Consequently, the characteristic impedance is given by

$$Z_c = \frac{Z_0 K'}{4\sqrt{\epsilon_e} K} \tag{3.192}$$

When the substrate material is anisotropic with a dielectric constant ϵ_y in the direction perpendicular to the air-dielectric interface and ϵ_r in the direction parallel to the conductors, then ϵ_r should be replaced by $\epsilon_g = \sqrt{\epsilon_r \epsilon_y}$ in (3.189) and (3.191).

In most applications it is necessary to provide shielding of a microwave circuit. If the shield dimensions are large, the shield will not produce a significant effect on the line characteristics. In monolithic microwave integrated circuits, the substrate is very thin and fragile; so it is desirable to use another ground plane below the substrate to mechanically strengthen the overall circuit and to also provide a better heat sink to help dissipate the power generated by active devices. Ghione and Naldi have considered a number of different shield arrangements as well as coupled coplanar lines.† We will present results only for the shielded coplanar line shown in Fig. 3.35*b*. For this structure the effect of the widewalls, spaced by amount $2A$, is negligible provided $2A > 10(S + 2W)$. We will assume that this is the case. Ghione and Naldi make the assumption that the air-dielectric interface in the slot regions can be replaced by magnetic walls in order to simplify the

†G. Ghione and C. U. Naldi, Coplanar Waveguides for MMIC Applications: Effect of Upper Shielding, Conductor Backing, Finite Extent Ground Planes, and Line-to-Line Coupling, *IEEE Trans.*, vol. MTT-35, pp. 260–267, March, 1987.

conformal mapping solution. The assumption is a correct one only for the case when the upper and lower shields are spaced the same distance from the coplanar line, that is, $H = H_1$ in Fig. 3.35*b*. For the line dimensions encountered in most practical applications of coplanar lines, the assumption does not introduce significant error. The advantage of this assumption is that it decouples the air and dielectric regions; so it is only necessary to find the capacitance of the air- and dielectric-filled sections separately.

It was found that the effective dielectric constant is given by

$$\epsilon_e = 1 + q(\epsilon_r - 1) \tag{3.193a}$$

where the filling factor q is given by

$$q = \frac{K(k_1)/K(k_1')}{K(k_2)/K(k_2') + K(k)/K(k')} \tag{3.193b}$$

K is the complete elliptic integral of the first kind, k_1' and k_2' are the complementary moduli given by $\sqrt{1 - k_1^2}$ and $\sqrt{1 - k_2^2}$, and

$$k_1 = \frac{\tanh(\pi S/4H)}{\tanh[\pi(S + 2W)/H]} \tag{3.193c}$$

$$k_2 = \frac{\tanh(\pi S/4H_1)}{\tanh[\pi(S + 2W)/4H_1]} \tag{3.193d}$$

The characteristic impedance is given by

$$Z_c = \frac{Z_0}{2\sqrt{\epsilon_e}\,[K(k_1)/K(k_1') + K(k_2)/K(k_2')]} \tag{3.194}$$

The ratios K/K' are easily evaluated using III.13*c*. For an anisotropic substrate ϵ_r should be replaced by $\epsilon_g = \sqrt{\epsilon_r \epsilon_y}$ and H should be replaced by the effective thickness $H_e = \sqrt{\epsilon_r/\epsilon_y}\,H$. When H and H_1 are very large, k_1 and k_2 become equal to k given by (3.190) and the filling factor q equals 0.5.

If the lower shield plate is replaced by a magnetic wall, we obtain an approximate model of a coplanar line with a substrate of finite thickness H and with an upper shield at a spacing H_1 above the coplanar conductors. For this case Ghione and Naldi give

$$q = \frac{K(k_3)/K(k_3')}{K(k_2)/K(k_2') + K(k)/K(k')} \tag{3.195a}$$

$$Z_c = \frac{Z_0}{2\sqrt{\epsilon_e}\,[K(k_2)/K(k_2') + K(k)/K(k')]} \tag{3.195b}$$

where

$$k_3 = \frac{\sinh(\pi S/4H)}{\sinh[\pi(S+2W)/4H]}$$

and k_2 is given by (3.193d), k is given by (3.190), and ϵ_e is given by (3.193a). When H_1 is made infinite k_2 becomes equal to k. The resultant structure is an unshielded coplanar line with a substrate having a finite thickness H.

Example 3.8. A coplanar line with upper and lower shielding and used in an MMIC circuit has the following dimensions: $S = 50\ \mu\text{m} = 0.05$ mm, $W = 50\ \mu\text{m}$, $H = 250\ \mu\text{m}$, $H_1 = 800\ \mu\text{m}$. The substrate material is gallium arsenide with $\epsilon_r = 12.9$. We want to find the effective dielectric constant and characteristic impedance. The first step is to find k_1 and k_2 using (3.193c) and (3.194c). A straightforward evaluation gives $k_1 = 0.3547$, $k_2 = 0.33547$. We now use (III.13c) to find $K(k_1)/K(k_1') = 0.6573$, $K(k_2)/K(k_2') = 0.6414$. By using (3.193a) and (3.193b), we get $q = 0.506$ and $\epsilon_e = 7.0228$. The last calculation is for Z_c using (3.194) and gives $Z_c = 54.77\ \Omega$. This example shows that for the shield spacings used very little effect on the coplanar-line characteristics occurred as is apparent from the fact that $k_1 \approx k_2$ and $q \approx 0.5$, relationships that hold exactly when no shields are used.

Attenuation

The attenuation in a coplanar line caused by dielectric loss is given by the same formula as for a microstrip line, i.e., by (3.171). By using the filling factor q this can be expressed as

$$\alpha_d = \frac{\pi}{\lambda_0}\frac{\epsilon_r'}{\sqrt{\epsilon_e}} q \tan\delta_l \tag{3.196}$$

This formula can be used for shielded and unshielded lines as long as the appropriate filling factor q is used.

For thin conductors with thickness T less than $0.05S$ and with a slot width $W > 0.3S$, the formula for the unshielded coplanar-line attenuation derived in App. III can be used. The center conductor has a series resistance per unit length given by

$$R_1 = \frac{R_m}{4S(1-k^2)K^2(k)}\left[\pi + \ln\frac{4\pi S}{T} - k\ln\frac{1+k}{1-k}\right] \tag{3.197a}$$

The distributed series resistance of the ground planes is given by

$$R_2 = \frac{kR_m}{4S(1-k^2)K^2(k)}\left[\pi + \ln\frac{4\pi(S+2W)}{T} - \frac{1}{k}\ln\frac{1+k}{1-k}\right] \tag{3.197b}$$

where $k = S/(S + 2W)$. The attenuation due to conductor loss is given by

$$\alpha_c = \frac{R_1 + R_2}{2Z_c} \tag{3.197c}$$

In normal situations the additional attenuation introduced by upper and lower shields is small if H and H_1 are greater than $4W$.

Example 3.9. A coplanar line has a copper strip of width 0.6 mm, slot width $W = 0.6$ mm, metal thickness $T = 0.005S = 3$ μm. The dielectric used is alumina with a dielectric constant of 9.7 and a loss tangent equal to 2×10^{-4}. We want to find the attenuation caused by dielectric loss and conductor loss at a frequency of 4 GHz. For this line $k = S/(S + 2W) = 0.333$. The effective dielectric constant, from (3.191), is 5.35. From (3.192) and (III.13c) we obtain

$$Z_c = \frac{120\pi}{4\sqrt{5.35}} \frac{1}{\pi} \ln 2 \frac{1 + \sqrt{k'}}{1 - \sqrt{k'}} = 63.7\ \Omega$$

where we used $k' = \sqrt{1 - k^2} = 0.9428$. From (3.196) we obtain

$$\alpha_d = \frac{9.7\pi}{7.5\sqrt{5.35}} 0.5 \times 2 \times 10^{-4} \times 8.686 = 1.526 \times 10^{-3}\ \text{dB/cm}$$

where $q = (\epsilon_e - 1)/(\epsilon_r' - 1) = 0.5$ was used. To evaluate $K(k) = K(0.333)$, we use (III.13d) to obtain

$$K(0.333) = \frac{2}{1 + k'} K\left(\frac{1 - k'}{1 + k'}\right) = \frac{2}{1.9428} K(0.0294) = \frac{\pi}{1.9428}$$
$$= 1.617$$

From (3.197) we get $R_1 = 19.25R_m$/cm and $R_2 = 5.97R_m$/cm. We now use $R_m = 8.22 \times 10^{-3}\sqrt{f}$ and (3.197c) to get

$$\alpha_c = (2.158 \times 10^{-2} + 6.69 \times 10^{-3}) = 2.827 \times 10^{-2}\ \text{dB/cm}$$

The ground planes contribute 23.7 percent of the conductor loss. The total attenuation is 2.98×10^{-2} dB/cm.

Jackson has made some loss calculations for coplanar and microstrip transmission lines at a frequency of 60 GHz and found that for typical line dimensions and characteristic impedances greater than 50 Ω, the coplanar line has a smaller attenuation.† The substrate considered had a dielectric constant of 12.8 and a loss tangent of 6×10^{-3}. The maximum strip width considered by Jackson was 0.3 mm. For these line dimensions the quasistatic formulas given above can be used. We have verified that indeed the quasistatic formulas give essentially the same attenuation. The lower loss in coplanar lines appears to be due to being able to use a wider center conductor for a given impedance as compared with that for a microstrip line. However, this is not always

†R. W. Jackson, Considerations in the Use of Coplanar Waveguide for Millimeter-Wave Integrated Circuits *IEEE Trans.*, vol. MTT-34, pp. 1450–1456, December, 1986.

necessarily true since the relative attenuation of the two types of transmission lines will depend on other factors such as substrate thickness and dielectric constant also.

High-Frequency Dispersion

Coplanar transmission lines exhibit dispersion effects similar to that for microstrip lines. There are less available computed results for the effective dielectric constant and characteristic impedance for coplanar lines than what is available for the microstrip line. The coplanar line can be viewed as two coupled slot lines and from this point of view it is clear that there are two modes of propagation that are quasi-TEM in character. The reader is referred to the paper by Nakatani and Alexopoulos for typical dispersive properties of a coplanar line.† In many integrated microwave circuit applications, the line dimensions are so small relative to the wavelength that the quasi-TEM formulas can be used even though the frequency may be as high as 50 GHz.

PART 3
RECTANGULAR AND CIRCULAR WAVEGUIDES

Hollow-pipe waveguides do not support a TEM wave. In hollow-pipe waveguides the waves are of the TE and TM variety. The waveguide with a rectangular cross section is the most widely used one. It is available in sizes for use at frequencies from 320 MHz up to 333 GHz. The WR-2300 waveguide for use at 320 MHz has internal dimensions of 58.42 in by 29.1 in and is a very large duct. By contrast, the WR-3 waveguide for use at 333 GHz has internal dimensions of 0.034 in by 0.017 in and is a very miniature structure. The standard WR-90 *X*-band waveguide has internal dimensions of 0.9 in by 0.4 in and is used in the frequency range of 8.2 to 12.5 GHz. The rectangular waveguide is widely used to couple transmitters and receivers to the antenna. For high-power applications the waveguide is filled with an inert gas such as nitrogen and pressurized in order to increase the voltage breakdown rating.

Circular waveguides are not as widely used as rectangular waveguides but are available in diameters of 25.18 in down to 0.239 in to cover the frequency range 800 MHz up to 116 GHz.

†A. Nakatani and N. G. Alexopolous, Toward a Generalized Algorithm for the Modeling of the Dispersive Properties of Integrated Circuit Structures on Anisotropic Substrates, *IEEE Trans.*, vol. MTT-33, pp. 1436–1441, December, 1985.

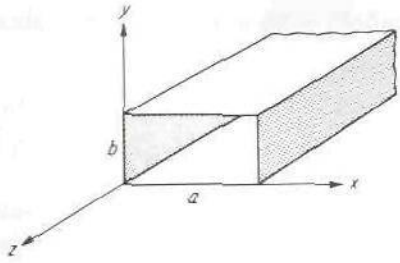


FIGURE 3.36
Rectangular waveguide.

3.17 RECTANGULAR WAVEGUIDE

The rectangular waveguide with a cross section as illustrated in Fig. 3.36 is an example of a waveguiding device that will not support a TEM wave. Consequently, it turns out that unique voltage and current waves do not exist, and the analysis of the waveguide properties has to be carried out as a field problem rather than as a distributed-parameter-circuit problem.

In a hollow cylindrical waveguide a transverse electric field can exist only if a time-varying axial magnetic field is present. Similarly, a transverse magnetic field can exist only if either an axial displacement current or an axial conduction current is present, as Maxwell's equations show. Since a TEM wave does not have any axial field components and there is no center conductor on which a conduction current can exist, a TEM wave cannot be propagated in a cylindrical waveguide.

The types of waves that can be supported (propagated) in a hollow empty waveguide are the TE and TM modes discussed in Sec. 3.7. The essential properties of all hollow cylindrical waveguides are the same, so that an understanding of the rectangular guide provides insight into the behavior of other types as well. As for the case of the transmission line, the effect of losses is initially neglected. The attenuation is computed later by using the perturbation method given earlier, together with the loss-free solution for the currents on the walls.

The essential properties of empty loss-free waveguides, which the detailed analysis to follow will establish, are that there is a double infinity of possible solutions for both TE and TM waves. These waves, or modes, may be labeled by two identifying integer subscripts n and m, for example, TE_{nm}. The integers n and m pertain to the number of standing-wave interference maxima occurring in the field solutions that describe the variation of the fields along the two transverse coordinates. It will be found that each mode has associated with it a characteristic cutoff frequency $f_{c,nm}$ below which the mode does not propagate and above which the mode does propagate. The cutoff frequency is a geometrical parameter dependent on the waveguide cross-sectional configuration. When f_c has been determined, it is found that the propagation factor β is given by

$$\beta = \left(k_0^2 - k_c^2\right)^{1/2} \tag{3.198}$$

where $k_0 = \omega\sqrt{\mu_0\epsilon_0}$ and $k_c = 2\pi f_c\sqrt{\mu_0\epsilon_0}$. The guide wavelength is readily seen to be given by

$$\lambda_g = \frac{2\pi}{\beta} = \frac{\lambda_0}{\left(1 - \lambda_0^2/\lambda_c^2\right)^{1/2}} = \frac{\lambda_0}{\sqrt{1 - f_c^2/f^2}} \tag{3.199}$$

where λ_0 is the free-space wavelength of plane waves at the frequency $f = \omega/2\pi$. Since k_c differs for different modes, there is always a lower band of frequencies for which only one mode propagates (except when k_c may be the same for two or more modes). In practice, waveguides are almost universally restricted to operation over this lower-frequency band for which only the dominant mode propagates, because of the difficulties associated with coupling signal energy into and out of a waveguide when more than one mode propagates. This latter difficulty arises because of the different values of the propagation phase constant β for different modes, since this means that the signal carried by the two or more modes does not remain in phase as the modes propagate along the guide. This necessitates the use of separate coupling probes for each mode at both the input and output and thus leads to increased system complexity and cost.

Another feature common to all empty uniform waveguides is that the phase velocity v_p is greater than the velocity of light c by the factor λ_g/λ_0. On the other hand, the velocity at which energy and a signal are propagated is the group velocity v_g and is smaller than c by the factor λ_0/λ_g. Also, since β, and hence λ_g, v_p, and v_g, are functions of frequency, any signal consisting of several frequencies is dispersed, or spread out, in both time and space as it propagates along the guide. This dispersion results from the different velocities at which the different frequency components propagate. If the guide is very long, considerable signal distortion may take place. Group and signal velocities are discussed in detail in Sec. 3.19.

With some of the general properties of waveguides considered, it is now necessary to consider the detailed analysis that will establish the above properties and that, in addition, will provide the relation between k_c and the guide configuration, the expressions for power and attenuation, etc. The case of TE modes in a loss-free empty rectangular guide is considered first.

TE Waves

For TE, or H, modes, $e_z = 0$ and all the remaining field components can be determined from the axial magnetic field h_z by means of (3.71). The axial field h_z is a solution of

$$\nabla_t^2 h_z + k_c^2 h_z = 0$$

or

$$\frac{\partial^2 h_z}{\partial x^2} + \frac{\partial^2 h_z}{\partial y^2} + k_c^2 h_z = 0 \tag{3.200}$$

If a product solution $h_z = f(x)g(y)$ is assumed, (3.200) becomes

$$\frac{1}{f}\frac{d^2f}{dx^2} + \frac{1}{g}\frac{d^2g}{dy^2} + k_c^2 = 0$$

after substituting fg for h_z and dividing the equation by fg. The term $1/f\,d^2f/dx^2$ is a function of x only, $1/g\,d^2g/dy^2$ is a function of y only, and k_c^2 is a constant, and hence this equation can hold for all values of x and y only if each term is constant. Thus we may write

$$\frac{1}{f}\frac{d^2f}{dx^2} = -k_x^2 \quad \text{or} \quad \frac{d^2f}{dx^2} + k_x^2 f = 0$$

$$\frac{1}{g}\frac{d^2g}{dy^2} = -k_y^2 \quad \text{or} \quad \frac{d^2g}{dy^2} + k_y^2 g = 0$$

where $k_x^2 + k_y^2 = k_c^2$ in order that the sum of the three terms may vanish. The use of the *separation-of-variables* technique has reduced the partial differential equation (3.200) to two ordinary simple-harmonic second-order equations. The solutions for f and g are easily found to be

$$f = A_1 \cos k_x x + A_2 \sin k_x x$$

$$g = B_1 \cos k_y y + B_2 \sin k_y y$$

where A_1, A_2, B_1, B_2 are arbitrary constants. These constants, as well as the separation constants k_x, k_y, can be further specified by considering the boundary conditions that h_z must satisfy. Since the normal component of the transverse magnetic field $\mathbf{h}$ must vanish at the perfectly conducting waveguide wall, (3.71*b*) shows that $\mathbf{n} \cdot \nabla_t h_z = 0$ at the walls, where $\mathbf{n}$ is a unit normal vector at the walls. When this condition holds, tangential $\mathbf{e}$ will also vanish on the guide walls, as (3.71*c*) shows. The requirements on h_z are thus

$$\frac{\partial h_z}{\partial x} = 0 \qquad \text{at } x = 0, a$$

$$\frac{\partial h_z}{\partial y} = 0 \qquad \text{at } y = 0, b$$

where the guide cross section is taken to be that in Fig. 3.36. In the solution for f, the boundary conditions give

$$-k_x A_1 \sin k_x x + k_x A_2 \cos k_x x = 0 \qquad \text{at } x = 0, a$$

Hence, from the condition at $x = 0$, it is found that $A_2 = 0$. At $x = a$, it is

necessary for $\sin k_x a = 0$, and this specifies k_x to have the values

$$k_x = \frac{n\pi}{a} \qquad n = 0, 1, 2, \ldots$$

In a similar manner it is found that $B_2 = 0$ and

$$k_y = \frac{m\pi}{b} \qquad m = 0, 1, 2, \ldots$$

Both n and m equal to zero yields a constant for the solution for h_z and no other field components; so this trivial solution is of no interest.

If we use the above relations and put $A_1 B_1 = A_{nm}$, the solutions for h_z are seen to be

$$h_z = A_{nm} \cos\frac{n\pi x}{a} \cos\frac{m\pi y}{b} \tag{3.201}$$

for $n = 0, 1, 2, \ldots$; $m = 0, 1, 2, \ldots$; $n = m \neq 0$. The constant A_{nm} is an arbitrary amplitude constant associated with the nmth mode. For the nmth mode the cutoff wave number is designated $k_{c,nm}$, given by

$$k_{c,nm} = \left[\left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2\right]^{1/2} \tag{3.202}$$

and is clearly a function of the guide dimensions only. The propagation constant for the nmth mode is given by

$$\begin{aligned}\gamma_{nm} &= j\beta_{nm} = j\left(k_0^2 - k_{c,nm}^2\right)^{1/2} \\ &= j\left[\left(\frac{2\pi}{\lambda_0}\right)^2 - \left(\frac{n\pi}{a}\right)^2 - \left(\frac{m\pi}{b}\right)^2\right]^{1/2}\end{aligned} \tag{3.203}$$

When $k_0 > k_{c,nm}$, β_{nm} is pure real and the mode propagates; when $k_0 < k_{c,nm}$, then γ_{nm} is real but β_{nm} is imaginary and the propagation factor is $e^{-\gamma_{nm}|z|}$, which shows that the mode decays rapidly with distance $|z|$ from the point at which it is excited. This decay is not associated with energy loss, but is a characteristic feature of the solution. Such decaying, or evanescent, modes may be used to represent the local diffraction, or fringing, fields that exist in the vicinity of coupling probes and obstacles in waveguides. The frequency separating the propagation and no-propagation bands is designated the cutoff frequency $f_{c,nm}$. This is given by the solution of $k_0 = k_{c,nm}$; that is,

$$f_{c,nm} = \frac{c}{\lambda_{c,nm}} = \frac{c}{2\pi} k_{c,nm} = \frac{c}{2\pi}\left[\left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2\right]^{1/2} \tag{3.204}$$

where c is the velocity of light. The cutoff wavelength is given by

$$\lambda_{c,nm} = \frac{2ab}{(n^2b^2 + m^2a^2)^{1/2}} \tag{3.205}$$

A typical guide may have $a = 2b$, in which case

$$\lambda_{c,nm} = \frac{2a}{(n^2 + 4m^2)^{1/2}}$$

and $\lambda_{c,10} = 2a$, $\lambda_{c,01} = a$, $\lambda_{c,11} = 2a/\sqrt{5}$, etc. In this example there is a band of wavelengths from a to $2a$, that is, a frequency band

$$\frac{c}{2a} < f < \frac{c}{a}$$

for which only the H_{10} mode propagates. This is the dominant mode in a rectangular guide and the one most commonly used in practice. Above the frequency c/a, other modes may propagate; so the useful frequency band in the present case is a one-octave band from $c/2a$ to c/a.

The remainder of the field components for the TE_{nm}, or H_{nm}, mode are readily found from (3.201) by using (3.71). The results for the complete nmth solution are

$$H_z = A_{nm} \cos\frac{n\pi x}{a} \cos\frac{m\pi y}{b} e^{\mp j\beta_{nm}z} \tag{3.206a}$$

$$H_x = \pm j\frac{\beta_{nm}}{k_{c,nm}^2} A_{nm} \frac{n\pi}{a} \sin\frac{n\pi x}{a} \cos\frac{m\pi y}{b} e^{\mp j\beta_{nm}z} \tag{3.206b}$$

$$H_y = \pm j\frac{\beta_{nm}}{k_{c,nm}^2} A_{nm} \frac{m\pi}{b} \cos\frac{n\pi x}{a} \sin\frac{m\pi y}{b} e^{\mp j\beta_{nm}z} \tag{3.206c}$$

$$E_x = Z_{h,nm} A_{nm} j\frac{\beta_{nm}}{k_{c,nm}^2} \frac{m\pi}{b} \cos\frac{n\pi x}{a} \sin\frac{m\pi y}{b} e^{\mp j\beta_{nm}z} \tag{3.206d}$$

$$E_y = -Z_{h,nm} A_{nm} j\frac{\beta_{nm}}{k_{c,nm}^2} \frac{n\pi}{a} \sin\frac{n\pi x}{a} \cos\frac{m\pi y}{b} e^{\mp j\beta_{nm}z} \tag{3.206e}$$

where the wave impedance for the nmth H mode is given by

$$Z_{h,nm} = \frac{k_0}{\beta_{nm}} Z_0 \tag{3.207}$$

When the mode does not propagate, $Z_{h,nm}$ is imaginary, indicating that there is no net energy flow associated with the evanescent mode. A general field with $E_z = 0$ can be described in a complete manner by a linear superposition of all the H_{nm} modes.

Power

For a propagating H_{nm} mode the power, or rate of energy flow, in the positive z direction is given by

$$\begin{aligned} P_{nm} &= \frac{1}{2}\mathrm{Re}\int_0^a\int_0^b \mathbf{E}\times\mathbf{H}^*\cdot\mathbf{a}_z\,dx\,dy \\ &= \frac{1}{2}\mathrm{Re}\int_0^a\int_0^b (E_xH_y^* - E_yH_x^*)\,dx\,dy \\ &= \frac{1}{2}\mathrm{Re}\,Z_{h,nm}\int_0^a\int_0^b (H_yH_y^* + H_xH_x^*)\,dx\,dy \end{aligned} \tag{3.208}$$

If we substitute from (3.206b) and (3.206c) and note that

$$\int_0^a\int_0^b \sin^2\frac{n\pi x}{a}\cos^2\frac{m\pi y}{b}\,dx\,dy = \int_0^a\int_0^b \cos^2\frac{n\pi x}{a}\sin^2\frac{m\pi y}{b}\,dx\,dy$$

$$= \begin{cases} \dfrac{ab}{4} & n\neq 0,\ m\neq 0 \\ \dfrac{ab}{2} & n \text{ or } m = 0 \end{cases}$$

we find that

$$\begin{aligned} P_{nm} &= |A_{nm}|^2\frac{\frac{1}{2}ab}{\epsilon_{0n}\epsilon_{0m}}\frac{\beta_{nm}^2}{k_{c,nm}^4}Z_{h,nm}\left[\left(\frac{m\pi}{b}\right)^2 + \left(\frac{n\pi}{a}\right)^2\right] \\ &= \frac{|A_{nm}|^2ab}{2\epsilon_{0n}\epsilon_{0m}}\left(\frac{\beta_{nm}}{k_{c,nm}}\right)^2 Z_{h,nm} \end{aligned} \tag{3.209}$$

where ϵ_{0m} is the Neumann factor and equal to 1 for $m = 0$ and equal to 2 for $m > 0$.

If two modes, say the H_{nm} and H_{rs} modes, were present simultaneously, it would be found that the power is the sum of that contributed by each individual mode, that is, $P_{nm} + P_{rs}$. This is a general property of loss-free waveguides. This power orthogonality arises because of the orthogonality of the functions (eigenfunctions) that describe the transverse variation of the fields when integrated over the guide cross section; e.g.,

$$\int_0^a \sin\frac{n\pi x}{a}\sin\frac{r\pi x}{a}\,dx = 0 \qquad n\neq r$$

Even when small losses are present the energy flow may be taken to be that contributed by each individual mode, with negligible error in all cases except when two or more degenerate modes are present. Degenerate modes are modes which have the same propagation constant γ, and for these the presence of even small losses may result in strong coupling between the modes.

Attenuation

If the waveguide walls have finite conductivity, there will be a continuous loss of power to the walls as the modes propagate along the guide. Consequently, the phase constant $j\beta$ is perturbed and becomes $\gamma = \alpha + j\beta$, where α is an attenuation constant that gives the rate at which the mode amplitude must decay as the mode progresses along the guide. For practical waveguides the losses caused by finite conductivity are so small that the attenuation constant may be calculated using the perturbation method outlined in Sec. 3.8 in connection with lossy transmission lines. The method will be illustrated for the dominant H_{10} mode only. For the H_{nm} and also the E_{nm} modes, the calculation differs only in that somewhat greater algebraic manipulation is required.

For the H_{10} mode, the fields are given by (apart from the factor $e^{-j\beta_{10}z}$)

$$h_z = A_{10}\cos\frac{\pi x}{a}$$

$$h_x = j\frac{\beta_{10}}{k_{c,10}^2}A_{10}\frac{\pi}{a}\sin\frac{\pi x}{a}$$

$$e_y = -Z_{h,10}A_{10}\frac{j\beta_{10}}{k_{c,10}^2}\frac{\pi}{a}\sin\frac{\pi x}{a}$$

as reference to (3.206) shows. From (3.209) the rate of energy flow along the guide is

$$P_{10} = |A_{10}|^2\frac{ab}{4}\left(\frac{\beta_{10}}{k_{c,10}}\right)^2 Z_{h,10}$$

The currents on the lossy walls are assumed to be the same as the loss-free currents, and hence are given by

$$\mathbf{J}_s = \mathbf{n} \times \mathbf{H}$$

where $\mathbf{n}$ is a unit inward directed normal at the guide wall. Thus, on the walls at $x = 0, a$, the surface currents are

$$\mathbf{J}_s = \begin{cases} \mathbf{a}_x \times \mathbf{H} = -\mathbf{a}_y A_{10} & x = 0 \\ -\mathbf{a}_x \times \mathbf{H} = -\mathbf{a}_y A_{10} & x = a \end{cases}$$

whereas on the upper and lower walls the currents are

$$\mathbf{J}_s = \begin{cases} \mathbf{a}_y \times \mathbf{H} = -\mathbf{a}_z\dfrac{j\beta_{10}}{k_{c,10}^2}A_{10}\dfrac{\pi}{a}\sin\dfrac{\pi x}{a} + \mathbf{a}_x A_{10}\cos\dfrac{\pi x}{a} & y = 0 \\ -\mathbf{a}_y \times \mathbf{H} = \mathbf{a}_z\dfrac{j\beta_{10}}{k_{c,10}^2}A_{10}\dfrac{\pi}{a}\sin\dfrac{\pi x}{a} - \mathbf{a}_x A_{10}\cos\dfrac{\pi x}{a} & y = b \end{cases}$$

With a finite conductivity σ, the waveguide walls may be characterized as exhibiting a surface impedance given by

$$Z_m = \frac{1+j}{\sigma\delta_s} = (1+j)R_m$$

where δ_s is the skin depth. The power loss in the resistive part R_m of Z_m per unit length of guide is

$$P_l = \frac{R_m}{2}\oint_{\substack{\text{guide}\\ \text{walls}}} \mathbf{J}_s \cdot \mathbf{J}_s^* \, dl$$

$$= \frac{R_m|A_{10}|^2}{2}\left(2\int_0^b dy + 2\int_0^a \frac{\beta_{10}^2}{k_{c,10}^4}\frac{\pi^2}{a^2}\sin^2\frac{\pi x}{a}\,dx + 2\int_0^a \cos^2\frac{\pi x}{a}\,dx\right)$$

Since $k_{c,10} = \pi/a$, the above gives

$$P_l = R_m|A_{10}|^2\left[b + \frac{a}{2}\left(\frac{\beta_{10}}{k_{c,10}}\right)^2 + \frac{a}{2}\right]$$

If P_0 is the power at $z = 0$, then $P_{10} = P_0 e^{-2\alpha z}$ is the power in the guide at any z. The rate of decrease of power propagated is

$$-\frac{dP_{10}}{dz} = 2\alpha P_{10} = P_l$$

and equals the power loss, as indicated in the above equation. The attenuation constant α for the H_{10} mode is thus seen to be

$$\alpha = \frac{P_l}{2P_{10}} = \frac{R_m\left[b + \frac{a}{2}\left(\frac{\beta_{10}}{k_{c,10}}\right)^2 + \frac{a}{2}\right]}{\frac{ab}{2}\left(\frac{\beta_{10}}{k_{c,10}}\right)^2 Z_{h,10}}$$

$$= \frac{R_m}{ab\beta_{10}k_0 Z_0}(2bk_{c,10}^2 + ak_0^2) \text{ Np/m} \qquad (3.210)$$

The attenuation for other TE_{nm} modes is given by the formula in Table 3.4, which summarizes the solutions for TE_{nm} and also TM_{nm} modes. In Fig. 3.37 the attenuation for the TE_{10} mode in a copper rectangular guide is given as a function of frequency. To convert attenuation given in nepers to decibels, multiply by 8.686.

The theoretical formulas for attenuation give results in good agreement with experimental values for frequencies below about 5,000 MHz. For higher frequencies, measured values of α may be considerably higher, depending on the smoothness of the waveguide surface. If surface imperfections of the order of magnitude of the skin depth δ_s are present, it is readily

TABLE 3.4
Properties of modes in a rectangular guide†

	TE modes	TM modes
H_z	$\cos\frac{n\pi x}{a}\cos\frac{m\pi y}{b}e^{-j\beta_{nm}z}$	0
E_z	0	$\sin\frac{n\pi x}{a}\sin\frac{m\pi y}{b}e^{-j\beta_{nm}z}$
E_x	$Z_{h,nm}H_y$	$\frac{-j\beta_{nm}n\pi}{ak_{c,nm}^2}\cos\frac{n\pi x}{a}\sin\frac{m\pi y}{b}e^{-j\beta_{nm}z}$
E_y	$-Z_{n,nm}H_x$	$\frac{-j\beta_{nm}m\pi}{bk_{c,nm}^2}\sin\frac{n\pi x}{a}\cos\frac{m\pi y}{b}e^{-j\beta_{nm}z}$
H_x	$\frac{j\beta_{nm}n\pi}{ak_{c,nm}^2}\sin\frac{n\pi x}{a}\cos\frac{m\pi y}{b}e^{-j\beta_{nm}z}$	$-\frac{E_y}{Z_{e,nm}}$
H_y	$\frac{j\beta_{nm}m\pi}{bk_{c,nm}^2}\cos\frac{n\pi x}{a}\sin\frac{m\pi y}{b}e^{-j\beta_{nm}z}$	$\frac{E_x}{Z_{e,nm}}$
$Z_{h,nm}$	$\frac{k_0}{\beta_{nm}}Z_0$	
$Z_{e,nm}$		$\frac{\beta_{nm}}{k_0}Z_0$
$k_{c,nm}$	$\left[\left(\frac{n\pi}{a}\right)^2+\left(\frac{m\pi}{b}\right)^2\right]^{1/2}$	
β_{nm}	$(k_0^2-k_{e,nm}^2)^{1/2}$	
$\lambda_{c,nm}$	$\frac{2ab}{(n^2b^2+m^2a^2)^{1/2}}$	
α	$\frac{2R_m}{bZ_0(1-k_{c,nm}^2/k_0^2)^{1/2}}\left[\left(1+\frac{b}{a}\right)\frac{k_{c,nm}^2}{k_0^2}+\frac{b}{a}\left(\frac{\epsilon_{0m}}{2}-\frac{k_{c,nm}^2}{k_0^2}\right)\frac{n^2ab+m^2a^2}{n^2b^2+m^2a^2}\right]$	$\frac{2R_m}{bZ_0(1-k_{c,nm}^2/k_0^2)^{1/2}}\frac{n^2b^3+m^2a^3}{n^2b^2a+m^2a^3}$

†$R_m=(\omega\mu_0/2\sigma)^{1/2}$, $\epsilon_{0m}=1$ for $m=0$ and 2 for $m>0$. The expression for α is not valid for degenerate modes.

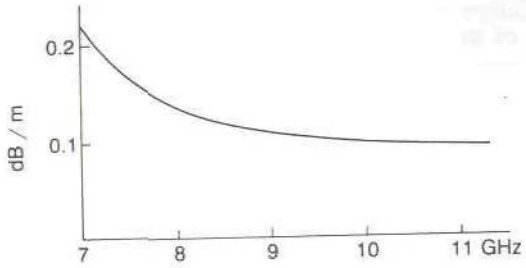


FIGURE 3.37
Attenuation of TE_{10} mode in a copper rectangular waveguide, $a = 2.286$ cm, $b = 1.143$ cm.

appreciated that the effective surface area is much greater, resulting in greater loss. By suitably polishing the surface, the experimental values of attenuation are found to be in substantial agreement with the theoretical values.†

Dominant TE_{10} Mode

Since the TE_{10} mode is the dominant mode in a rectangular guide, and also the most commonly used mode, it seems appropriate to examine this mode in more detail. From the results given earlier, the field components for this mode are described by the following (propagation in the $+z$ direction assumed):

$$H_z = A \cos \frac{\pi x}{a} e^{-j\beta z} \tag{3.211a}$$

$$H_x = \frac{j\beta}{k_c} A \sin \frac{\pi x}{a} e^{-j\beta z} \tag{3.211b}$$

$$E_y = -jAZ_h \frac{\beta}{k_c} \sin \frac{\pi x}{a} e^{-j\beta z} \tag{3.211c}$$

where the subscript 10 has been dropped for convenience since this discussion pertains only to the TE_{10} mode. The parameters β, k_c, and Z_h are given by

$$k_c = \frac{\pi}{a} \tag{3.212a}$$

$$\beta = \left[k_0^2 - \left(\frac{\pi}{a} \right)^2 \right]^{1/2} \tag{3.212b}$$

$$Z_h = -\frac{E_y}{H_x} = \frac{k_0}{\beta} Z_0 \tag{3.212c}$$

†See J. Allison and F. A. Benson, Surface Roughness and Attenuation of Precision Drawn, Chemically Polished, Electropolished, Electroplated and Electroformed Waveguides, *Proc. IEE (London)*, vol. 102, pt. B, pp. 251–259, 1955.

The guide wavelength λ_g is

$$\lambda_g = \frac{2\pi}{\beta} = \frac{\lambda_0}{\left[1 - (\lambda_0/2a)^2\right]^{1/2}} \tag{3.212d}$$

since the cutoff wavelength $\lambda_c = 2a$. The phase and group velocities are

$$v_p = \frac{\lambda_g}{\lambda_0} c \tag{3.212e}$$

$$v_g = \frac{\lambda_0}{\lambda_g} c \tag{3.212f}$$

and are discussed in detail in Sec. 3.19.

In Fig. 3.38 the magnetic and electric field lines associated with the TE_{10} mode are illustrated. Note that the magnetic flux lines encircle the electric field lines; so these can be considered to be the source (displacement current) for the magnetic field. On the other hand, the electric field lines terminate in an electric charge distribution on the inner surface of the upper and lower waveguide walls. This charge oscillates back and forth in the axial and transverse directions and thus constitutes an axial and transverse conduction current that forms the continuation of the displacement current. The total current, displacement plus conduction, forms a

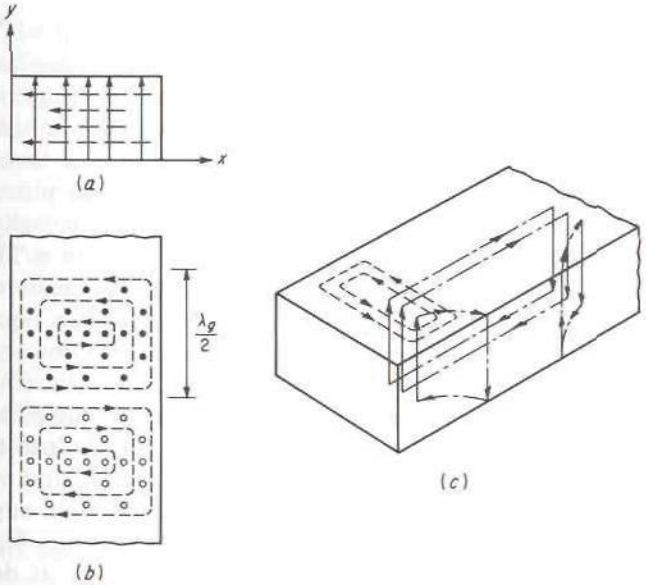


FIGURE 3.38
Magnetic and electric field lines for the TE_{10} mode. (*a*) Transverse plane; (*b*) top view; (*c*) mutual total current and magnetic field linkages.

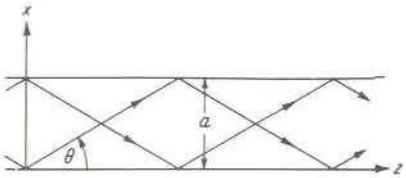


FIGURE 3.39
Decomposition of TE_{10} mode into two plane waves.

closed linkage of the magnetic field lines, and as such may be regarded as being generated by the changing magnetic flux these enclose. This completes the required mutual-support action between the electric and magnetic fields which is required for wave propagation.

The fields for a TE_{10} mode may be decomposed into the sum of two plane TEM waves propagating along zigzag paths between the two waveguide walls at $x = 0$ and $x = a$, as in Fig. 3.39. For the electric field we have

$$E_y = -\frac{Z_h}{2}\frac{\beta}{k_c}(e^{j\pi x/a - j\beta z} - e^{-j\pi x/a - j\beta z})$$

If π/a and β are expressed as

$$\frac{\pi}{a} = k_0 \sin\theta \qquad \beta = k_0 \cos\theta$$

the relation $(\pi/a)^2 + \beta^2 = k_0^2$ still holds. The electric field is now given by

$$E_y = \frac{Z_h}{2}\frac{\beta}{k_c}(e^{-jk_0(x\sin\theta + z\cos\theta)} - e^{-jk_0(-x\sin\theta + z\cos\theta)})$$

which is clearly two plane waves propagating at angles $\pm\theta$ with respect to the z axis, as illustrated. Alternatively, the field may be pictured as a plane wave reflecting back and forth between the two guide walls. As shown in Sec. 2.7, the constant phase planes associated with these obliquely propagating plane waves move in the z direction at the phase velocity $c/\cos\theta = \beta c/k_0$, and this is the reason why the phase velocity of the TE_{10} mode exceeds the velocity of light. Since the energy in a TEM wave propagates with the velocity c in the direction in which the plane wave propagates, this energy will propagate down the guide at a velocity equal to the component of c along the z axis. This component is $v_g = c\cos\theta = (k_0/\beta)c$ and is the group velocity for the TE_{10} mode. When $\theta = \pi/2$, the plane waves reflect back and forth, but do not progress down the guide; so the mode is cutoff.

The above decomposition of the TE_{10} mode into two plane waves may be extended to the TE_{nm} modes also. When n and m are both different from zero, four plane waves result. Although such superpositions of plane waves may be used to construct the field solutions for rectangular guides, this is a rather cumbersome approach. However, it does lend insight into why the phase velocity exceeds that of light, as well as other properties of the modes.

TM Modes

For TM modes, h_z equals zero and e_z plays the role of a potential function from which the remaining field components may be derived. This axial electric field satisfies the reduced Helmholtz equation

$$\nabla_t^2 e_z + k_c^2 e_z = 0 \tag{3.213}$$

of the same type encountered earlier for h_z, that is, (3.200). The solution may be found by using the separation-of-variables method. In the present case the boundary conditions require that e_z vanish at $x = 0, a$ and $y = 0, b$. This condition requires that the solution for e_z be

$$e_z = A_{nm} \sin\frac{n\pi x}{a} \sin\frac{m\pi y}{b} \tag{3.214}$$

instead of a product of cosine functions which was suitable for describing h_z. Again, there are a doubly infinite number of solutions corresponding to various integers n and m. However, unlike the situation for TE modes, $n = 0$ and $m = 0$ are not solutions. The cutoff wave number is given by the same expression as for TE modes; i.e.,

$$k_{c,nm} = \left[\left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2\right]^{1/2} \tag{3.215}$$

and the propagation factor β_{nm} by

$$\beta_{nm} = \left(k_0^2 - k_{c,nm}^2\right)^{1/2} \tag{3.216}$$

The lowest-order propagating mode is the $n = m = 1$ mode, and this has a cutoff wavelength equal to $2ab/(a^2 + b^2)^{1/2}$. Note that the TE_{10} mode can propagate at a lower frequency (longer wavelength), thus verifying that this is the dominant mode.† It should also be noted that for the same values of n and m, the TE_{nm} and TM_{nm} modes are degenerate since they have the same propagation factor. Another degeneracy occurs when $a = b$, for in this case the four modes TE_{nm}, TE_{mn}, TM_{nm}, and TM_{mn} all have the same propagation constant. Still further degeneracies exist if a is an integer multiple of b, or vice versa.

The rest of the solution for TM modes is readily constructed using the general equations (3.72). A summary of this solution is given in Table 3.4. The TM modes are the dual of the TE modes and apart from minor differences have essentially the same properties. For this reason it does not seem necessary to repeat the preceding discussion.

---

†In any hollow waveguide the dominant mode is a TE mode because the boundary conditions $e_z = 0$ for TM modes always require e_z to have a greater spatial variation in the transverse plane than that for h_z for the lowest-order TE mode, and hence the smallest value of k_c occurs for TE modes. Hence a TE mode has the lowest cutoff frequency, i.e., is the dominant mode.

3.18 CIRCULAR WAVEGUIDES

Figure 3.40 illustrates a cylindrical waveguide with a circular cross section of radius a. In view of the cylindrical geometry involved, cylindrical coordinates are most appropriate for the analysis to be carried out. Since the general properties of the modes that may exist are similar to those for the rectangular guide, this section is not as detailed.

TM Modes

For the TM modes a solution of

$$\nabla_t^2 e_z + k_c^2 e_z = 0$$

is required such that e_z will vanish at $r = a$. When we express the transverse laplacian ∇_t^2 in cylindrical coordinates (App. I), this equation becomes

$$\frac{\partial^2 e_z}{\partial r^2} + \frac{1}{r}\frac{\partial e_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 e_z}{\partial \phi^2} + k_c^2 e_z = 0 \tag{3.217}$$

The separation-of-variables method may be used to reduce the above to two ordinary differential equations. Consequently, it is assumed that a product solution $f(r)g(\phi)$ exists for e_z. Substituting for e_z into (3.217) and dividing the equation by fg yield

$$\frac{1}{f}\frac{d^2 f}{dr^2} + \frac{1}{rf}\frac{df}{dr} + \frac{1}{r^2 g}\frac{d^2 g}{d\phi^2} + k_c^2 = 0$$

Multiplying this result by r^2 gives

$$\frac{r^2}{f}\frac{d^2 f}{dr^2} + \frac{r}{f}\frac{df}{dr} + r^2 k_c^2 = -\frac{1}{g}\frac{d^2 g}{d\phi^2}$$

The left-hand side is a function of r only, whereas the right-hand side depends on ϕ only. Therefore this equation can hold for all values of the variables only if both sides are equal to some constant ν^2. As a result, (3.217) is seen to separate into the following two equations:

$$\frac{d^2 f}{dr^2} + \frac{1}{r}\frac{df}{dr} + \left(k_c^2 - \frac{\nu^2}{r^2}\right) f = 0 \tag{3.218a}$$

$$\frac{d^2 g}{d\phi^2} + \nu^2 g = 0 \tag{3.218b}$$

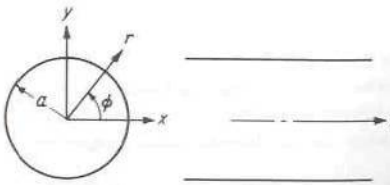


FIGURE 3.40
The circular cylindrical waveguide.

TABLE 3.5
Values of p_{nm} for TM modes

n	p_{n1}	p_{n2}	p_{n3}
0	2.405	5.520	8.654
1	3.832	7.016	10.174
2	5.135	8.417	11.620

In this case the field inside the waveguide must be periodic in ϕ with period 2π, that is, single-valued. It is therefore necessary to choose ν equal to an integer n, in which case the general solution to (3.218*b*) is

$$g(\phi) = A_1 \cos n\phi + A_2 \sin n\phi$$

where A_1 and A_2 are arbitrary constants.

Equation (3.218*a*) is Bessel's differential equation and has two solutions (a second-order differential equation always has two independent solutions) $J_\nu(k_c r)$ and $Y_\nu(k_c r)$, called Bessel functions of the first and second kind, respectively, and of order ν.† For the problem under investigation here, only $J_n(k_c r)$ is a physically acceptable solution since $Y_n(k_c r)$ becomes infinite at $r = 0$. The final solution for e_z may thus be expressed as

$$e_z(r, \phi) = (A_1 \cos n\phi + A_2 \sin n\phi) J_n(k_c r) \tag{3.219}$$

Reference to App. II shows that $J_n(x)$ behaves like a damped sinusoidal function and passes through zero in a quasiperiodic fashion. Since e_z must vanish when $r = a$, it is necessary to choose $k_c a$ in such a manner that $J_n(k_c a) = 0$. If the mth root of the equation $J_n(x) = 0$ is designated p_{nm}, the allowed values (eigenvalues) of k_c are

$$k_{c,nm} = \frac{p_{nm}}{a} \tag{3.220}$$

The values of p_{nm} for the first three modes for $n = 0, 1, 2$ are given in Table 3.5. As in the case of the rectangular guide, there are a doubly infinite number of solutions.

Each choice of n and m specifies a particular TM_{nm} mode (eigenfunction). The integer n is related to the number of circumferential variations in the field, whereas m relates to the number of radial variations. The propagation constant for the nmth mode is given by

$$\beta_{nm} = \left(k_0^2 - \frac{p_{nm}^2}{a^2}\right)^{1/2} \tag{3.221}$$

†Y_ν is also called a Neumann function.

the cutoff wavelength by

$$\lambda_{c,nm} = \frac{2\pi a}{p_{nm}} \tag{3.222}$$

and the wave impedance by

$$Z_{e,nm} = \frac{\beta_{nm}}{k_0} Z_0 \tag{3.223}$$

A cutoff phenomenon similar to that for the rectangular guide exists. For the dominant TM mode, $\lambda_c = 2\pi a/p_{01} = 2.61a$, a value 30 percent greater than the waveguide diameter.

Expressions for the remaining field components may be derived by using the general equations (3.72). Energy flow and attenuation may be determined by methods similar to those used for the rectangular guide. A summary of the results is given in Table 3.6.

TE Modes

The solution for TE modes parallels that for the TM modes with the exception that the boundary conditions require that $\partial h_z/\partial r$ vanish at $r = a$. An appropriate solution for h_z is

$$h_z(r, \phi) = (B_1 \cos n\phi + B_2 \sin n\phi) J_n(k_c r) \tag{3.224}$$

with the requirement that

$$\frac{dJ_n(k_c r)}{dr} = 0 \qquad \text{at } r = a \tag{3.225}$$

The roots of (3.225) will be designated by p'_{nm}; so the eigenvalues $k_{c,nm}$ are given by

$$k_{c,nm} = \frac{p'_{nm}}{a} \tag{3.226}$$

Table 3.7 lists the values of the roots for the first few modes. Note that $p'_{0m} = p_{1m}$ since $dJ_0(x)/dx = -J_1(x)$, and hence the TE_{0m} and TM_{1m} modes are degenerate.

The first TE mode to propagate is the TE_{11} mode, having a cutoff wavelength $\lambda_{c,11} = 3.41a$. This mode is seen to be the dominant mode for the circular waveguide, and is normally the one used. A sketch of the field lines in the transverse plane for this mode is given in Fig. 3.41. The attenuation constant for the dominant TE_{11} mode is given by

$$\alpha = \frac{R_m}{aZ_0}\left(1 - \frac{1.841^2}{k_0^2 a^2}\right)^{-1/2}\left(\frac{1.841^2}{k_0^2 a^2} + 0.4185\right) \tag{3.227}$$

TABLE 3.6
Properties of modes in circular waveguides

	TE modes	TM modes
H_z	$J_n\left(\frac{p'_{nm}r}{a}\right)e^{-j\beta_{nm}z}\begin{Bmatrix}\cos n\phi\\ \sin n\phi\end{Bmatrix}$	0
E_z	0	$J_n\left(\frac{p_{nm}r}{a}\right)e^{-j\beta_{nm}z}\begin{Bmatrix}\cos n\phi\\ \sin n\phi\end{Bmatrix}$
H_r	$-\frac{j\beta_{nm}p'_{nm}}{ak^2_{c,nm}}J'_n\left(\frac{p'_{nm}r}{a}\right)e^{-j\beta_{nm}z}\begin{Bmatrix}\cos n\phi\\ \sin n\phi\end{Bmatrix}$	$-\frac{E_\phi}{Z_{e,nm}}$
H_ϕ	$-\frac{jn\beta_{nm}}{rk^2_{c,nm}}J_n\left(\frac{p'_{nm}r}{a}\right)e^{-j\beta_{nm}z}\begin{Bmatrix}-\sin n\phi\\ \cos n\phi\end{Bmatrix}$	$\frac{E_r}{Z_{e,nm}}$
E_r	$Z_{h,nm}H_\phi$	$-\frac{j\beta_{nm}p_{nm}}{ak^2_{c,nm}}J'_n\left(\frac{p_{nm}r}{a}\right)e^{-j\beta_{nm}z}\begin{Bmatrix}\cos n\phi\\ \sin n\phi\end{Bmatrix}$
E_ϕ	$-Z_{h,nm}H_r$	$-\frac{jn\beta_{nm}}{rk^2_{c,nm}}J_n\left(\frac{p_{nm}r}{a}\right)e^{-j\beta_{nm}z}\begin{Bmatrix}-\sin n\phi\\ \cos n\phi\end{Bmatrix}$
β_{nm}	$\left[k_0^2-\left(\frac{p'_{nm}}{a}\right)^2\right]^{1/2}$	$\left[k_0^2-\left(\frac{p_{nm}}{a}\right)^2\right]^{1/2}$
$Z_{h,nm}$	$\frac{k_0}{\beta_{nm}}Z_0$	
$Z_{e,nm}$		$\frac{\beta_{nm}}{k_0}Z_0$
$k_{c,nm}$	$\frac{p'_{nm}}{a}$	$\frac{p_{nm}}{a}$
$\lambda_{c,nm}$	$\frac{2\pi a}{p'_{nm}}$	$\frac{2\pi a}{p_{nm}}$
Power	$\frac{Z_0k_0\beta_{nm}\varepsilon_n\pi}{2k^4_{c,nm}\epsilon_{0n}}(p'^2_{nm}-n^2)J_n^2(p'_{nm})$	$\frac{Y_0k_0\beta_{nm}\varepsilon_n\pi}{2k^4_{c,nm}\epsilon_{0n}}p^2_{nm}\left[J'_n(k_{c,nm}a)\right]^2$
α	$\frac{R_m}{aZ_0}\left(1-\frac{k^2_{c,nm}}{k_0^2}\right)^{-1/2}\times\left[\frac{k^2_{c,nm}}{k_0^2}+\frac{n^2}{(p'_{nm})^2-n^2}\right]$	$\frac{R_m}{aZ_0}\left(1-\frac{k^2_{c,nm}}{k_0^2}\right)^{-1/2}$

TABLE 3.7
Values of p'_{nm} for TE modes

n	p'_{n1}	p'_{n2}	p'_{n3}
0	3.832	7.016	10.174
1	1.841	5.331	8.536
2	3.054	6.706	9.970

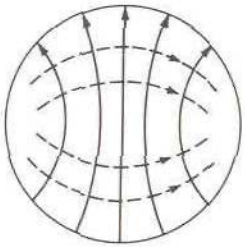


FIGURE 3.41
Field lines for the TE_{11} mode in a circular waveguide.

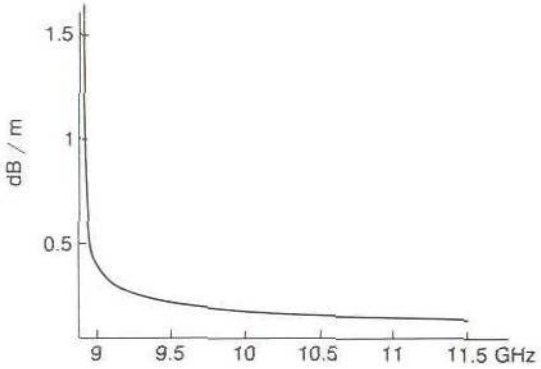


FIGURE 3.42
Attenuation of dominant TE_{11} mode in a circular waveguide. Diameter = 2 cm.

Figure 3.42 shows the attenuation in decibels per meter for a copper waveguide with an internal diameter of 2 cm. For this guide the cutoff frequency is 8.79 GHz. In the normal operating range from 9 to 11 GHz, the attenuation drops from 0.36 dB/m at 9 GHz to 0.11 dB/m at 11 GHz.

3.19 WAVE VELOCITIES

In any system capable of supporting propagating waves, a number of wave velocities occur that pertain to signal propagation, energy propagation, wavefront propagation, etc. These various velocities are examined below in the context of propagation in waveguides.

Phase Velocity

The phase velocity of a wave in a waveguide was introduced earlier and shown to be equal to

$$v_p = \frac{\lambda_g}{\lambda_0} c \tag{3.228}$$

for an air-filled guide. Here λ_g is the guide wavelength, λ_0 the free-space wavelength, and c the velocity of light. The phase velocity is the velocity an observer must move with in order to see a constant phase for the wave propagating along the guide. It is noted that the phase velocity is greater than the velocity of light, and since the principle of relativity states that no signal or energy can be propagated at a velocity exceeding that of light, the physical significance of the phase velocity might very well be questioned.

The clue to the significance of the phase velocity comes from the recognition that this velocity entered into wave solutions that had a steady-state time dependence of $e^{j\omega t}$. A pure monochromatic (single-frequency) wave of this type exists only if the source was turned on at $t = -\infty$ and is kept on for all future time as well. This is clearly not a physically realizable situation. In actual fact the source must be turned on at some finite time, which can be chosen as $t = 0$. The generated signal is then of the form illustrated in Fig. 3.43. Associated with the sudden steplike beginning of the signal is a broad frequency band, as a Fourier analysis shows. If this signal is injected into the guide at $z = 0$, an observer a distance l farther along the guide will, in actual fact, see no signal until a time l/c has elapsed. In other words, the wavefront will propagate along the guide with a velocity c. At the time l/c, the observer will begin to see the arrival of the transient associated with the switching on of the signal. After a suitable period of time has elapsed, the transient will have died out, and the observer will then see the steady-state sinusoidally varying wave. Once steady-state conditions prevail, the phase velocity can be introduced to describe the velocity at which a constant phase point appears to move along the guide. Note, however, that there is no information being transmitted along the guide once steady-state conditions have been established. Thus the phase velocity is not associated with any physical entity such as a signal, wavefront, or energy flow velocity. The term signal is used here to denote a time function that can convey

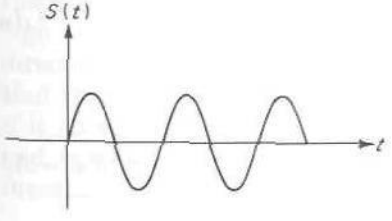


FIGURE 3.43
Sinusoidal signal applied at time $t = 0$.

information to the observer. Thus the step change at $t = 0$ is a signal, but once steady-state conditions are achieved, the observer does not receive any more information. A better understanding of the above features will be obtained after the group velocity, discussed below, has been examined.

Group Velocity

The physical definition of group velocity is the velocity with which a signal consisting of a very narrow band of frequency components propagates. The appropriate tool for the analysis of this situation is the Fourier transform. If a time function is denoted by $f(t)$, this function of time has associated with it a frequency spectrum $F(\omega)$ given by the Fourier transform of $f(t)$,

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt \tag{3.229a}$$

Conversely, if the spectrum $F(\omega)$ is known, the time function may be found from the inverse Fourier transform relation

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\,d\omega \tag{3.229b}$$

From Eq. (3.229b) it is seen that the Fourier transform represents $f(t)$ as a superposition of steady-state sinusoidal functions of infinite duration. These relations are a generalization of the Fourier series relations. If the time function is passed through a device having a response $Z(\omega)$ that is a function of frequency, e.g., filter, the output time function $f_o(t)$ will have a frequency spectrum $F(\omega)Z(\omega)$, and hence, by (3.229b), is given by

$$f_o(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)Z(\omega)e^{j\omega t}\,d\omega$$

In general, $Z(\omega) = |Z(\omega)|e^{-j\psi(\omega)}$; so

$$f_o(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} |Z(\omega)|F(\omega)e^{j(\omega t-\psi)}\,d\omega \tag{3.230}$$

If the output f_o is to be an exact reproduction of the input, then in (3.230) it is necessary for $|Z|$ to equal a constant A, and ψ must be a linear function of ω, say $a\omega + b$. In this case

$$f_o(t) = \frac{A}{2\pi}e^{-jb}\int_{-\infty}^{\infty} F(\omega)e^{j\omega(t-a)}\,d\omega \tag{3.231a}$$

If we put $t - a = t'$, the above becomes

$$f_o(t' + a) = \frac{Ae^{-jb}}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t'}\,d\omega = Ae^{-jb}f(t') \tag{3.231b}$$

as comparison with (3.229*b*) shows. Thus the output time function is

$$f_o(t' + a) = f_o(t) = Ae^{-jb}f(t') = Ae^{-jb}f(t - a) \tag{3.232}$$

i.e., an exact duplicate of the input, apart from a constant multiplier and a time delay a. Thus the conditions given on $|Z|$ and ψ are those sufficient for a distortion-free system.

Now, in a waveguide, the transverse variations of the field are independent of frequency. The only essential frequency-dependent part of the field solution is the propagation factor $e^{-j\beta z}$ since

$$\beta = \left(k_0^2 - k_c^2\right)^{1/2} = \left(\frac{\omega^2}{c^2} - k_c^2\right)^{1/2}$$

is a function of frequency. Thus a waveguide of length l, in which the field has a time dependence $e^{j\omega t}$, $\omega > 0$, can be considered as a frequency filter with a response $e^{-j\beta l}$. Since β is not a linear function of ω, it may be anticipated that some signal distortion will occur for propagation in a waveguide. For an ideal TEM-wave transmission line, $\beta = k_0 = \omega/c$ and distortion-free transmission is possible. However, practical lines have an attenuation which depends on frequency ($R_m \propto \sqrt{f}$), and this will produce distortion. Fortunately, for narrowband signals neither waveguides nor transmission lines produce significant distortion unless very long lines are used.

Consider now a time function $f(t)$ having a band of frequencies between $-f_m$ and f_m. This signal is used to modulate a carrier of frequency $f_c \gg f_m$. The resultant is

$$S(t) = f(t)\cos \omega_c t = \mathrm{Re}\left[f(t)e^{j\omega_c t}\right] \tag{3.233}$$

If $F(\omega)$ is the spectrum of $f(t)$, the spectrum of $S(t)$ is

$$\begin{aligned} F_s(\omega) &= \int_{-\infty}^{\infty} e^{-j\omega t} f(t) \frac{e^{j\omega_c t} + e^{-j\omega_c t}}{2}\, d\omega \\ &= \frac{1}{2}\left[F(\omega - \omega_c) + F(\omega + \omega_c)\right] \end{aligned} \tag{3.234}$$

These spectra are illustrated in Fig. 3.44.

For positive ω the waveguide response is $e^{-j\beta(\omega)l}$. For negative ω the response must be chosen as $e^{j\beta(\omega)l}$ since, if the time variation is $e^{-j\omega t}$, the sign in front of β must be positive for propagation in the positive z direction. In other words, all physical systems will have a response such that $|Z(\omega)|$ is an even function of ω and $\psi(\omega)$ is an odd function of ω. Since β is an even function, the sign must be changed for $\omega < 0$. These even and odd symmetry properties are required simply so that the output time function is real, a physical requirement. The output spectrum for the

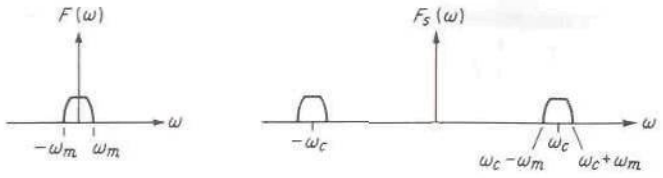


FIGURE 3.44
Spectrum of $f(t)$ and $S(t)$.

waveguide is thus

$$F_o(\omega) = \tfrac{1}{2}\left[F(\omega - \omega_c)e^{-j\beta(\omega)l} + F(\omega + \omega_c)e^{j\beta(\omega)l}\right]$$

and the output signal is

$$S_o(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F_o(\omega)e^{j\omega t}\,d\omega \tag{3.235}$$

The analysis that follows is simplified if the signal is represented in complex form as $f(t)e^{j\omega_c t}$ with a spectrum $F(\omega - \omega_c)$. In this case only the positive half of the spectrum needs to be considered, and the output signal is given by

$$S_o(t) = \mathrm{Re}\frac{1}{2\pi}\int_{\omega_c-\omega_m}^{\omega_c+\omega_m} F(\omega - \omega_c)e^{j\omega t - j\beta(\omega)l}\,d\omega \tag{3.236}$$

since $F(\omega - \omega_c)$ is zero outside the band $\omega_c - \omega_m \le \omega \le \omega_c + \omega_m$. If the band is very narrow, $\omega_m \ll \omega_c$, then $\beta(\omega)$ may be approximated by the first few terms in a Taylor series expansion about ω_c. Thus

$$\beta(\omega) = \beta(\omega_c) + \left.\frac{d\beta}{d\omega}\right|_{\omega_c}(\omega - \omega_c) + \frac{1}{2}\left.\frac{d^2\beta}{d\omega^2}\right|_{\omega_c}(\omega - \omega_c)^2 + \cdots \tag{3.237}$$

Retaining the first two terms only and letting $\beta(\omega_c) = \beta_0$ and $d\beta/d\omega_c = \beta_0'$ at ω_c, (3.236) gives

$$S_o(t) = \mathrm{Re}\frac{1}{2\pi}\int_{\omega_c-\omega_m}^{\omega_c+\omega_m} F(\omega - \omega_c)e^{j\omega(t-\beta_0' l)}e^{-j\beta_0 l + j\beta_0' l\omega_c}\,d\omega$$

If this is compared with (3.231) and (3.232), it is seen that

$$\begin{aligned} S_o(t) &= \mathrm{Re}\left[e^{-j\beta_0 l + j\beta_0' l\omega_c} f(t - \beta_0' l)e^{j\omega_c(t-\beta_0' l)}\right] \\ &= f(t - \beta_0' l)\cos(\omega_c t - \beta_0 l) \end{aligned} \tag{3.238}$$

To the order of approximation used here, the input modulating signal $f(t)$ is reproduced without distortion but with a time delay $\beta_0' l$. This is to be anticipated since β was approximated by a linear function of ω in the band

$\omega_c - \omega_m$ to $\omega_c + \omega_m$ (distortion-free condition). The signal delay defines the group velocity v_g, which is equal to the distance l divided by the delay time; thus

$$v_g = \frac{l}{l\beta_0'} = \left(\frac{d\beta}{d\omega}\right)^{-1} \tag{3.239}$$

This is also the signal velocity. Note, however, that this velocity has significance only if the band, or "group," of frequencies making up the signal is so narrow that β may be approximated by a linear function throughout the frequency band of interest. If this is not the case, more terms in the expansion (3.237) must be retained and signal distortion will occur. In this case the group velocity as given by (3.239) is no longer the signal velocity. In fact, because of signal distortion, no unique signal velocity exists any longer. Different portions of the signal will travel with different velocities, and the resultant signal becomes dispersed in both time and space.

In the case of a waveguide,

$$v_g = \left(\frac{d\beta}{d\omega}\right)^{-1} = c\frac{dk_0}{d\beta} = \left[\frac{d\left(\omega^2/c^2 - k_c^2\right)^{1/2}}{d\omega}\right]^{-1}$$

$$= \frac{\beta c^2}{\omega} = \frac{\beta}{k_0}c = \frac{\lambda_0}{\lambda_g}c \tag{3.240}$$

It is seen that $v_g < c$ and that $v_g v_p = c^2$ for a waveguide.

A typical plot of k_0 versus β for a waveguide is given in Fig. 3.45. From this plot it can be seen that for a narrow band of frequencies a linear approximation for β is good. Also note that for high frequencies (large k_0) β becomes equal to k_0. Thus frequencies well above the cutoff frequency f_c suffer very little dispersion and propagate essentially with the velocity of light c. No frequency components below the cutoff frequency f_c can propagate along the guide.

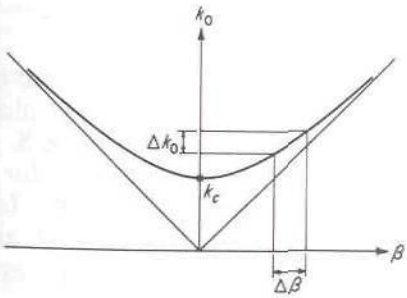


FIGURE 3.45
Plot of k_0 versus β for a waveguide.

The equality of the wavefront velocity and the velocity of light can be readily explained by means of Fig. 3.45. The switching on of a signal results in an initial transient that has a spectrum of frequencies extending out to infinity. Any small group of frequencies at the high end of the spectrum will have a group velocity equal to c since $dk_0/d\beta$ equals unity for k_0 near infinity. Thus the high-frequency part of the transient will propagate along the guide with a velocity c. The lower-frequency components will propagate with smaller group velocities and arrive later.

Energy-Flow Velocity

Power is a flow of energy, and consequently there is a velocity of energy flow such that the average energy density in the guide multiplied by this velocity is equal to the power. In a waveguide it turns out that this velocity of energy flow is equal to the group velocity. A proof for the case of E modes will be given below, that for H modes being very similar.

For E modes the field is given by [see (3.72)]

$$\mathbf{E}_t = -\frac{j\beta}{k_c^2}\nabla_t e_z e^{-j\beta z} \qquad \mathbf{H}_t = \frac{k_0 Y_0}{\beta}\mathbf{a}_z \times \mathbf{E}_t$$

The average rate of energy flow, or power, is given by

$$P = \frac{1}{2}\int_S Y_e |\mathbf{E}_t|^2 \, dS = \frac{1}{2}\frac{k_0 Y_0}{\beta}\int_S |\mathbf{E}_t|^2 \, dS \tag{3.241}$$

where the integration is over the guide cross section.

The energy density in the magnetic field per unit length of guide is

$$U_m = \frac{\mu_0}{4}\int_S |\mathbf{H}_t|^2 \, dS = \frac{\mu_0}{4}\frac{k_0^2 Y_0^2}{\beta^2}\int_S |\mathbf{E}_t|^2 \, dS \tag{3.242}$$

The energy density in the electric field per unit length of guide is equal to that in the magnetic field. This is readily shown to be the case by using the complex Poynting vector theorem, which states that (Sec. 2.5)

$$\frac{1}{2}\int_S \mathbf{E} \times \mathbf{H}^* \cdot \mathbf{a}_z \, dS = P + 2j\omega(W_m - W_e)$$

where the integration is over the guide cross section, and the term on the right gives the power transmitted past the plane S plus $2j\omega$ times the net reactive energy stored in the guide beyond the plane S. Since the integral of the complex Poynting vector over a cross section S for a propagating mode in a loss-free guide is real, it follows that $W_m = W_e$. In addition, since the location of the transverse plane S is arbitrary, it also follows that the energy densities U_m and U_e per unit length of guide are equal.

The velocity of energy flow may now be found from the relation

$$v = \frac{P}{U_e + U_m} = \frac{P}{2U_m} = \frac{k_0 Y_0}{\beta} \frac{\beta^2}{\mu_0 k_0^2 Y_0^2}$$

$$= \frac{\beta}{\mu_0 k_0 Y_0} = \frac{\beta}{k_0 \sqrt{\mu_0 \epsilon_0}} = \frac{\beta}{k_0} c = v_g \qquad (3.243)$$

and comes out equal to the group velocity as stated earlier.

3.20 RIDGE WAVEGUIDE

For a rectangular waveguide with a width a equal to twice the height b, the maximum bandwidth of operation over which only the dominant TE_{10} mode propagates is a 2 : 1 band. For some system applications it is necessary to have a waveguide that operates with only a single mode of propagation over much larger bandwidths. A transmission line supporting only a TEM mode can fulfill this requirement but must then have cross-sectional dimensions that are small relative to the shortest wavelength of interest. A coaxial transmission line will support higher-order TE and TM modes in addition to the TEM mode. Thus, to avoid excitation of a higher-order mode of propagation, the outer radius must be kept small relative to the wavelength. The small cross section implies a relatively large attenuation; so some other form of waveguide is needed. The ridge waveguide illustrated in Fig. 3.46 was developed to fulfill this need for a single-mode waveguide capable of operating over a very broad band. Physically, it is easy to understand why the ridge waveguide has a very large frequency band of operation. The center section of width W and spacing S functions very much like a parallel-plate transmission line and consequently the ridge waveguide has a much lower cutoff frequency for the same width and height as does the

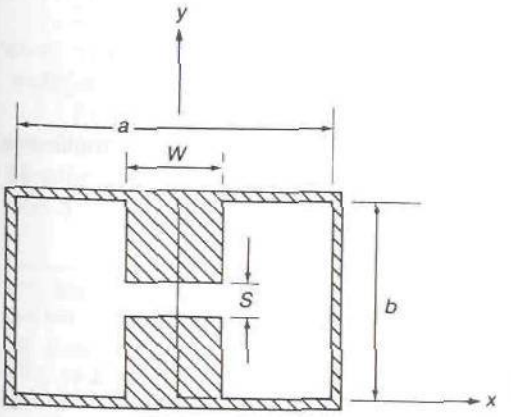


FIGURE 3.46
Ridge waveguide.

conventional rectangular waveguide. Operation over bandwidths of 5 to 1 or more is possible.

The ridge waveguide, when uniformly filled with dielectric, which can be air, has the same general properties as the rectangular and circular waveguides discussed earlier. If we can determine the cutoff wavelength λ_c for the dominant mode, then at any frequency above the cutoff frequency the propagation constant β is given by

$$\beta = \sqrt{k_0^2 - (2\pi/\lambda_c)^2}$$

At the cutoff frequency $\beta = 0$ and the electromagnetic field has no variation with the axial coordinate z. The cutoff wave number $k_c = 2\pi/\lambda_c$ can be found using the transverse-resonance method as described below.

The transverse-resonance method is based on finding the resonant frequency for the transmission-line circuit that provides a model for the waveguide cross section. At cutoff we can view the electromagnetic field as a uniform plane wave with components E_y and H_z that propagates in the x direction and is incident onto a second parallel-plate transmission line of reduced height. The equivalent transmission-line circuit is that of two parallel-plate transmission lines of height b, length $(a - W)/2$, and short-circuited at $x = 0$ and a. These two transmission lines are connected to another parallel-plate line of height S and length W and placed between the first two as shown in Fig. 3.47. At the junction where the height changes, a local fringing electric field occurs and stores electric field energy in the vicinity of the step. The effect of this local fringing electric field is taken into account by a shunt capacitive susceptance jB at each junction as shown in Fig. 3.47.

The standing-wave field pattern along the x direction can exist only at the resonant frequency for the transmission-line circuit shown in Fig. 3.47. For the dominant mode the voltage is a maximum and the current is zero at the midsection. Thus, at $x = a/2$, the impedance looking in the x direction toward the short circuit must be infinite. The corresponding input admittance will be zero. At the step the admittance looking toward the short circuit is

$$Y = -jY_{c1} \cot k_c \frac{a - W}{2} + jB$$

By using the formula for admittance transformation along a transmission

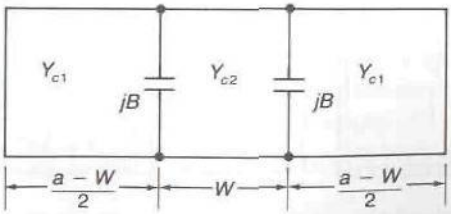


FIGURE 3.47
Equivalent transmission-line circuit of cross section of ridge waveguide.

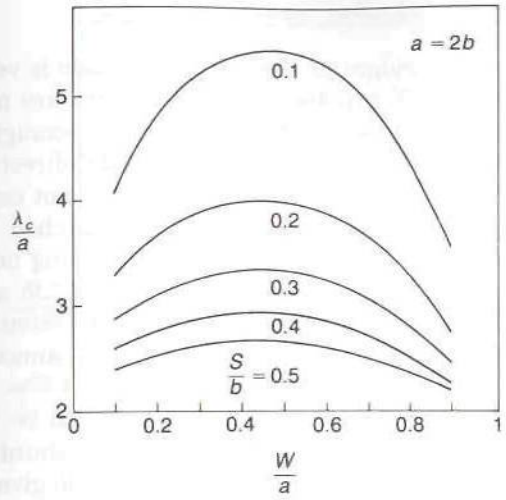


FIGURE 3.48
Normalized cutoff wavelength λ_c/a for a ridge waveguide.

line, we obtain

$$Y_{\text{in}} = Y_{c2}\frac{Y + jY_{c2}\tan k_c W/2}{Y_{c2} + jY\tan k_c W/2}$$

In order for Y_{in} to vanish, we must have

$$jB - jY_{c1}\cot k_c\frac{a-W}{2} + jY_{c2}\tan k_c\frac{W}{2} = 0 \qquad (3.244)$$

which is the transverse-resonance condition. The two characteristic admittances are inversely proportional to the parallel-plate spacing; thus $Y_{c2} = (b/S)Y_{c1}$. An approximate expression for the normalized shunt capacitive susceptance can be found using quasistatic conformal mapping and is†

$$\frac{B}{Y_{c1}} = \frac{2b}{\lambda_c}\left[1 - \ln 4u + \frac{1}{3}u^2 + \frac{1}{2}(1-u^2)^4\frac{b^2}{\lambda_c^2}\right] \qquad u = \frac{S}{b} < 0.5 \quad (3.245)$$

The eigenvalue equation (3.244) is a transcendental equation. The computer program RIDGEWG solves (3.244) for the normalized cutoff wavelength λ_c/a.

Figure 3.48 shows typical results that are obtained. The numerical results obtained from (3.244) agree within 1 percent of the values given by Hopfer and Hoefer and Burton over the commonly used range of parameters.‡

---

†N. Marcuvitz, "Waveguide Handbook," MIT Radiation Lab Series, vol. 10, reprinted by Boston Technical Publishers, Inc., 1964.

‡S. Hopfer, The Design of Ridged Waveguides *IRE Trans.*, vol. MTT-3, pp. 20–29, October, 1955.

W. J. R. Hoefer and M. N. Burton, Closed Form Expression for the Parameters of Finned and Ridged Waveguides, *IEEE Trans.*, vol. MTT-30, pp. 2190–2194, December 1982.

3.21 FIN LINE

If the width W of the ridges in the ridge waveguide is very small, we obtain a fin line as shown in Fig. 3.49*a*. Usually the fins are metal foils on a thin dielectric substrate mounted in the E plane of a rectangular waveguide. For the dominant mode the current flows in the axial direction; so good electrical contact between the fins and the waveguide is not essential. The fin line is a shielded slot line. The fin line can be matched to the rectangular waveguide by means of a tapered section or by using one or more quarter-wave impedance transformers as shown in Figs. 3.49*b* and *c*.

The fin line is suitable for use in microwave circuits that incorporate two-terminal devices such as diodes. Transistors cannot be connected to a fin line since they are three-terminal devices.

The cutoff wavelength for a fin line may also be found by using the transverse-resonance method. The fins produce a shunt normalized capacitive susceptance across the center of the waveguide given by†

$$\frac{B}{Y_c} = \frac{2b}{\lambda_c}\left[\ln\frac{1}{\alpha_2} + \sum_{n=1}^{4}\left(\frac{\pi}{\Gamma_n b} - \frac{1}{n}\right)P_n^2 + \frac{\left(\dfrac{\pi}{\Gamma_1 b} - 1\right)P_1^2}{1 + \left(\dfrac{\pi}{\Gamma_1 b} - 1\right)\alpha_2^2}\right] \tag{3.246}$$

where

$$P_1 = \alpha_1$$

$$P_2 = 2\alpha_1^2 + \alpha_2^2 - 1$$

$$P_3 = 4\alpha_1^3 + 6\alpha_1\alpha_2^2 - 3\alpha_1$$

$$P_4 = 8\alpha_1^4 + 3\alpha_2^4 + 24\alpha_1^2\alpha_2^2 - 8\alpha_1^2 - 4\alpha_2^2 + 1$$

$$\alpha_1 = \cos^2\frac{\pi S}{2b}$$

$$\alpha_2 = \sin^2\frac{\pi S}{2b}$$

$$\Gamma_n = \left[\left(\frac{n\pi}{b}\right)^2 - \left(\frac{2\pi}{\lambda_c}\right)^2\right]^{1/2}$$

The equivalent circuit of the fin-line cross section consists of two short-circuited transmission lines of length $a/2$ with jB connected at the center as

†R. E. Collin, "Field Theory of Guided Waves," 2nd ed., chap. 8, IEEE Press, Piscataway, N.J., 1990.

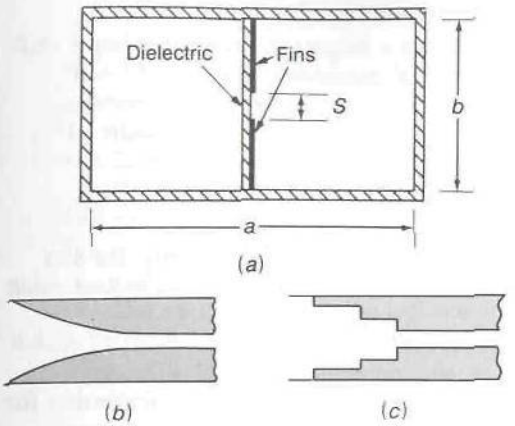


FIGURE 3.49
(*a*) Fin line; (*b*) tapered matching section; (*c*) quarter-wave matching section.

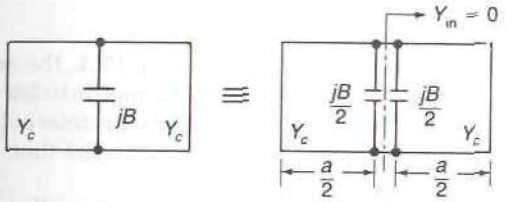


FIGURE 3.50
Equivalent circuit of fin-line cross section.

shown in Fig. 3.50. The resonance condition is

$$\frac{Y_{\text{in}}}{Y_c} = \frac{jB}{2Y_c} - j \cot \frac{\pi a}{\lambda_c} = 0 \tag{3.247}$$

The computer program FINLINE solves this transcendental equation for the normalized cutoff wavelength λ_c/a for $S/b = 0.1$ to 0.9. Typical numerical results are shown in Fig. 3.51 for the case where $a = 2b$. These numerical results agree within 1 percent or better with those given by Hoefer and Burton.†

The propagation constant β is given by the same formula as for a conventional waveguide, i.e.,

$$\beta = \left(k_0^2 - k_c^2\right)^{1/2}$$

When the fins are mounted on a dielectric substrate, a correction is needed for the propagation constant if the dielectric has an appreciable thickness

†W. J. R. Hoefer and M. N. Burton, *loc. cit.*

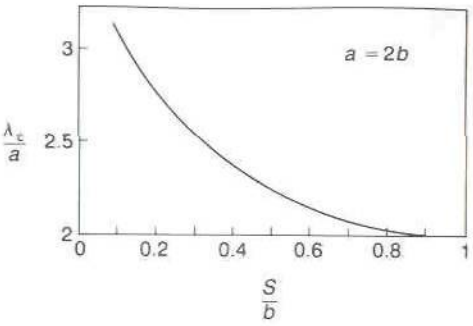


FIGURE 3.51
Normalized cutoff wavelength λ_c/a for a fin line.

and a large dielectric constant. Empirical formulas for this case are available.†

PROBLEMS

3.1. For the ideal transmission line shown in Fig. P3.1, the switch is closed at $t = 0$ and opened 1 μs later. The characteristic impedance of the line is 50 Ω. The load resistance is also 50 Ω. The battery has an internal resistance of 10 Ω.
(*a*) Sketch the voltage across R_L as a function of time for a line 300 m long. The wave velocity $v = 3 \times 10^8$ m/s.
(*b*) Sketch the voltage waveform across R_L when $R_L = 25$ Ω and the line is 900 m long.
(*c*) Sketch the voltage waveform across R_L when $R_L = 100$ Ω and the line is 900 m long.
(*d*) Repeat (*b*) and (*c*) for a line 75 m long.

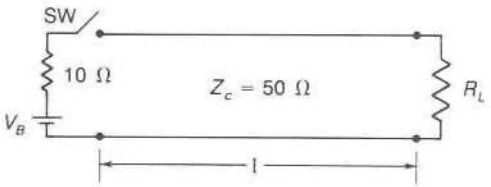


FIGURE P3.1

3.2. Let a generator with internal resistance R_g be connected to a transmission line of length l and having a characteristic impedance Z_c. The line is terminated in a load resistance R_L. Let $\tau = l/v$ be the one-way propagation time delay. The generator produces a pulsed waveform $P(t)$, $0 \le t \le T$. Show that the voltage across R_L is given by

$$V_L = \frac{Z_c}{Z_g + Z_c}(1 + \Gamma_L)\left[P(t - \tau) + \Gamma_L\Gamma_g P(t - 3\tau) + \Gamma_L^2\Gamma_g^2 P(t - 5\tau) + \cdots\right]$$

†K. Chang (ed.), "Handbook of Microwave and Optical Components," vol. 1, pp. 38–39, John Wiley & Sons, Inc., New York, 1989.

Hint: See (3.12) and consider the total line voltage at $z = l$.

3.3. A pulse generator produces a sawtooth pulse $P(t) = 10t/T$, $0 \le t \le T$, where $T = 10^{-8}$ s. The generator has an internal resistance $R_g = 200\ \Omega$ and is connected to a transmission line with $Z_c = 50\ \Omega$. The line is l meters long and terminated in a load resistance R_L. The wave velocity equals 3×10^8 m/s.

(*a*) Find and sketch the load voltage as a function of time when $l = 3$ m and $R_L = 200\ \Omega$.

(*b*) Repeat (*a*) when $R_L = 12.5\ \Omega$.

(*c*) Find an analytic expression for the voltage across R_L when $R_L = 200\ \Omega$ and $l = 12$ m.

(*d*) Make a distance-time plot of the line voltage for (*c*).

3.4. A battery with voltage of 10 V is connected in series with a 50-Ω resistor to the input of a 50-Ω transmission line at time $t = 0$. The transmission line, of length 12 m, is terminated in a 1-μF capacitor. Find and sketch the voltage across the capacitor as a function of time.

Hint: Apply Thévenin's theorem.

Answer:

$$V_c = 10[1 - e^{-(t-0.04)/50}] \qquad t \text{ in microseconds}$$

3.5. In the circuit illustrated in Fig. P3.5, the battery is connected at $t = 0$. Find and sketch the voltage across R_L as a function of time. Assume that $R_L = R_g = Z_c = 50\ \Omega$, $C = 1\ \mu$F, $l = 300$ m, and $v = 3 \times 10^8$ m/s.

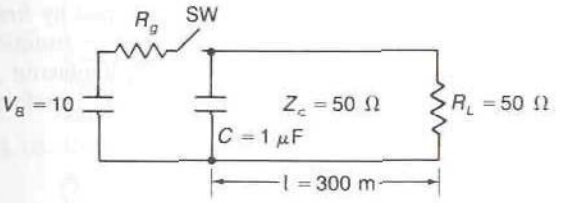


FIGURE P3.5

3.6. The resistor R_L is replaced by a capacitor $C_L = 1\ \mu$F in Fig. P3.5. Find the voltage across C_L during the time interval $1\ \mu s \le t \le 3\ \mu s$.

Answer:

$$V_L = 5[1 + e^{-(t-1)/25}] - 10e^{-(t-1)/50}$$

where t is in microseconds.

3.7. Consider an ideal loss-free transmission line of length l, as shown in Fig. P3.7. The far end is short-circuited. At the input end a battery of voltage V_0 is switched across the line at time $t = 0$. Sketch the voltage wave on the line at the middle over the time interval $0 \le t \le 7l/c$.

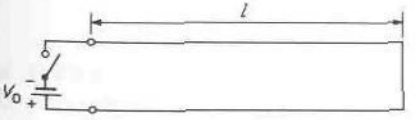


FIGURE P3.7

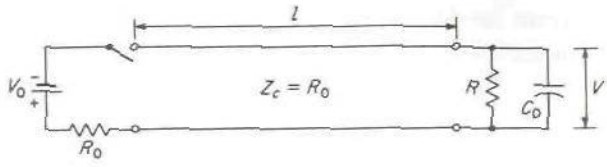


FIGURE P3.8

3.8. Consider the transmission-line circuit illustrated in Fig. P3.8. At time $t = 0$, a battery of voltage V_0 is switched across the line at the input. Determine the output load voltage V as a function of time.

Hint: This transient problem may be solved in a manner similar to that used in low-frequency circuit theory. The governing equations for the transmission line are

$$\frac{\partial V}{\partial z} = -L\frac{\partial \mathcal{I}}{\partial t} \qquad \frac{\partial \mathcal{I}}{\partial z} = -C\frac{\partial V}{\partial t}$$

The time derivative may be eliminated by taking the Laplace transform. The transformed solutions for V and $\mathcal{I}$ are

$$V^+e^{-sz/v} + V^-e^{sz/v} \qquad I^+e^{-sz/v} + I^-e^{sz/v}$$

By transforming the circuit equations for the load termination and the input voltage, the resultant equations may be solved for the Laplace transform of the load voltage. The inverse transform then gives the load voltage as a function of time.

The foregoing procedure may be simplified by first replacing the battery by a source $V_0e^{j\omega t}$ and obtaining the transfer function $V/V_0 = Z_t(j\omega)$ as a function of ω for this steady-state problem. Replacing $j\omega$ by s then gives the transfer function in the s domain. The Laplace transform of the output voltage is then

$$V(s) = \frac{V_0}{s}Z_t(j\omega = s)$$

since the Laplace transform of the input step voltage is V_0/s. The output voltage is obtained from the inverse Laplace transform of $V(s)$.

3.9. Obtain expressions for the voltage and current standing-wave patterns on a lossless open-circuited transmission line. Sketch these patterns. Assume an $e^{j\omega t}$ time dependence.

3.10. A transmission line with $Z_c = 50\ \Omega$ is terminated in an impedance $25 + j25\ \Omega$. Find the reflection coefficient, standing-wave ratio, and fraction of the incident power delivered to the load.

Answer: VSWR = 2.618, 80 percent.

3.11. Verify (3.47) and compute Z_{in} at a distance $\lambda_0/4$ from the termination given in Prob. 3.10.

3.12. On a transmission line with $Z_c = 50\ \Omega$, the voltage at a distance $0.4\lambda_0$ from the load is $4 + j2$. The corresponding current is 0.1 A. Determine the normalized load impedance.

Answer: $0.145 + j0.397$.

3.13. A 50-Ω transmission line is terminated by a 75-Ω load resistor. Find the distance l from the load at which $Y_{in} = 0.02 + jB$. By connecting a shunt

susceptance $-jB$ across the line at this point, the load will be matched to the transmission line. Explain why this is so. The distance l can be expressed in terms of the wavelength λ_0.

3.14. Figure P3.14 illustrates a three-conductor transmission line. Since potential is arbitrary to within an additive constant, the shield S_0 can be chosen to be at zero potential. Show that there are two linearly independent solutions for TEM waves in this transmission line. If S_0 encloses N conductors, how many TEM-wave solutions are possible?

Hint: Note that the potential can be arbitrarily specified on S_1 and S_2.

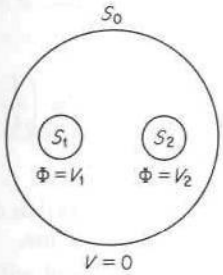


FIGURE P3.14

3.15. Show that power transmitted along a transmission line is given by

$$P = \frac{Y_0}{2} \int |\nabla_t \Phi|^2 \, dx \, dy$$

For Prob. 3.14 show that this equals $\frac{1}{2}(V_1 I_1 + V_2 I_2)$ by using Green's first identity (App. I) to convert the surface integral to a contour integral around the conductor boundaries. I_1 and I_2 are the total currents on S_1 and S_2.

3.16. Consider a three-conductor transmission line as shown in Fig. P3.14 but assume that the cross sections of S_1 and S_2 are not the same. Let Φ_a and Φ_b be two different solutions for the potential field. For Φ_a let V_{a1}, I_{a1} and V_{a2}, I_{a2} be the voltage and currents on S_1 and S_2. Similarly, for Φ_b let V_{b1}, I_{b1} and V_{b2}, I_{b2} be the voltages and currents on S_1 and S_2. For the TEM modes derived from $\Phi_a + \Phi_b$, show that the power flow is given by

$$\tfrac{1}{2}(V_{a1} + V_{b1})(I_{a1} + I_{b1}) + \tfrac{1}{2}(V_{a2} + V_{b2})(I_{a2} + I_{b2})$$

It is convenient to choose the potentials so that the two TEM modes obtained from Φ_a and Φ_b have independent power flow. Show that this will be the case if the interaction term

$$(V_{a1} I_{b1} + V_{b1} I_{a1}) + (V_{a2} I_{b2} + V_{b2} I_{a2})$$

equals zero. Furthermore, show that the interaction term will vanish if the potentials are chosen so that

$$\frac{V_{a1}}{V_{a2}} = -\frac{V_{b1}}{V_{b2}} = \left[\frac{C_{22} + C_{12}}{C_{11} + C_{12}}\right]^{1/2}$$

where C_{11} is the capacitance between S_1 and S_0, C_{22} is the capacitance between S_2 and S_0, and C_{12} is the capacitance between S_1 and S_2. For a

symmetrical line, $C_{11} = C_{22}$ and the two modes correspond to the even and odd modes.

Hint: Use the relations $Q_{a1} = C_{11}V_{a1} + C_{12}(V_{a1} - V_{a2})$, $Q_{a2} = C_{22}V_{a2} + C_{12}(V_{a2} - V_{a1})$ and similar ones for the total charge Q_{b1} and Q_{b2} on S_1, S_2 in terms of V_{b1} and V_{b2}. Also note that $I_{a1} = (Y_0/\epsilon_0)Q_{a1}$, etc.

3.17. Show that, for an air-filled coaxial line, minimum attenuation occurs when $x \ln x = 1 + x$, $x = b/a$. What is the corresponding characteristic impedance?

Hint: Hold the outer radius b constant and find $d\alpha/da$.

3.18. Evaluate Z_c for a lossy coaxial line using (3.28) and computed values of R, G, L, and C. Assume $b = 3a = 1$ cm, $f = 10^9$ Hz, $\sigma = 5.8 \times 10^7$ S/m, and $\epsilon = (2.56 - j0.001)\epsilon_0$. Verify that

$$\text{Im } Z_c \ll \text{Re } Z_c \qquad \text{and} \qquad Z_c \approx \left(\frac{L}{C}\right)^{1/2}$$

See Table 3.1 for coaxial-line parameters.

3.19. Use the energy definitions of L and C [Eqs. (3.112)] to derive the results given by (3.106) and (3.108) for a coaxial transmission line.

3.20. A microstrip line has a substrate 1 mm thick and with a dielectric constant $\epsilon_r = 8$. The strip width $W = 2.5$ mm. Find the low-frequency effective dielectric constant and characteristic impedance.

Answer: $\epsilon_e = 5.953$, $Z_c = 32.13\ \Omega$.

3.21. A microstrip line uses an anisotropic dielectric substrate with $\epsilon_x = 10$ and $\epsilon_y = 8$. The substrate is 0.5 mm thick and the strip width $W = 0.75$ mm. Find the low-frequency effective dielectric constant and characteristic impedance.

Answer: $\epsilon_e = 5.895$, $Z_c = 42.97\ \Omega$.

3.22. A microstrip line has a 1-mm-thick dielectric substrate with a dielectric constant of 6. Use the computer program MSTP to generate data giving the effective dielectric constant and characteristic impedance as a function of strip width W. Use these data to design a microstrip system with an input line having $Z_c = 50\ \Omega$, an output line having $Z_c = 75\ \Omega$, and an intermediate quarter-wave transformer section with $Z_c = \sqrt{50 \times 75}\ \Omega$. Specify the three strip widths W_1, W_2, W_3 and the length of the quarter-wave transformer at a frequency $f = 2$ GHz (see Fig. P3.22).

Answer: Widths are 1.505 mm, 1.0125 mm, 0.649 mm, length = 1.83 cm.

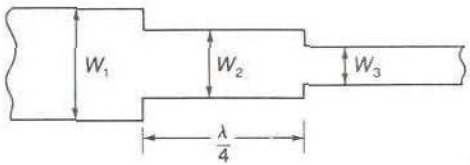


FIGURE P3.22

3.23. Find the effective dielectric constant, characteristic impedance, and attenuation at 2 GHz for a microstrip line with the following parameters: $\epsilon_r = 9.5$, loss tangent $= 2 \times 10^{-3}$, substrate thickness $H = 1$ mm, strip width $= 1.5$ mm, and strip thickness $T = 0.01$ mm. Use the computer program MSTP.

3.24. Use the computer program MSTPD to generate dispersion data for the effective dielectric constant for a microstrip line having the following parameters: dielectric constant = 6, substrate thickness $H = 0.5$ mm, and strip width $W = 2$ mm. At what frequency has the effective dielectric constant increased by 5 percent more than the quasistatic value?

3.25. In a monolithic microwave integrated circuit, gallium arsenide with $\epsilon_r = 12.9$ is used as a substrate material. Find the effective dielectric constant, characteristic impedance, and attenuation at 10 GHz for a microstrip line with the following parameters: substrate thickness = 0.1 mm, strip width = 0.05 mm, strip thickness = 0.002 mm, and loss tangent = 6×10^{-3}. For these dimensions the quasistatic parameters are accurate. The computer program MSTP can be used for the evaluation. What is the attenuation in decibels per wavelength (microstrip) for this microstrip line?

3.26. A microstrip line has the following parameters: strip width $W = 1$ mm, substrate thickness = 1 mm, and anisotropic dielectric with $\epsilon_x = 6.5$, $\epsilon_y = 6$. Find the following: distributed capacitance C and inductance L per centimeter, characteristic impedance, effective dielectric constant, and the quasi-TEM-mode wavelength at 2 GHz.

3.27. Use the computer program CMSTP to find the even- and odd-mode characteristic impedances and the voltage coupling coefficient C for a coupled microstrip line having the following parameters: strip width $W = 1$ mm, strip spacing $S = 0.1$ mm, substrate thickness = 1 mm, and substrate dielectric constant = 9.7.

3.28. A strip line has a ground-plane spacing of 2 mm, a strip width of 1 mm, and is filled with a dielectric medium with dielectric constant 2.3. Find the characteristic impedance.

3.29. Use the computer program STPL to evaluate the characteristic impedance and attenuation of a strip line with the following parameters: ground-plane spacing = 2 mm, strip width $W = 0.5$ mm, strip thickness $T = 0.01$ mm, dielectric constant = 6, loss tangent = 0.006, and frequency of operation = 5 GHz. What is the ratio of the attenuation due to dielectric loss relative to that of conductor loss?

3.30. A broadside coupled strip line is required for a 3-dB directional coupler. The even-mode characteristic impedance is to be 50 Ω. The voltage coupling coefficient is 0.707. From this information determine the required odd-mode characteristic impedance. Find the required strip width W and spacing S for this coupled strip line. The ground-plane spacing is 4 mm and the dielectric constant of the dielectric filling is 5. The strip thickness $T = 0.05$ mm. Use the computer program CSTPL. You will need to follow an iterative procedure to arrive at the required parameters. An accuracy of ± 0.5 percent is adequate.

Hint: Begin with $W = 3.5$ mm, $S = 0.5$ mm.

3.31. In a monolithic microwave integrated circuit, a coplanar transmission line with the following parameters is used: strip width $S = 0.1$ mm, slot width $W = 0.1$ mm, strip thickness = 0.002 mm, substrate thickness = 0.5 mm, dielectric constant = 12.9, loss tangent = 0.0008, and frequency = 10 GHz. Use the computer program CPW to determine the characteristic impedance

and attenuation. If the strip thickness is increased to 0.005 mm, will this significantly reduce the attenuation?

3.32. Figure P3.32 shows a coplanar-transmission-line circuit for use in a MMIC *amplifier circuit. The required input- and output-line characteristic* impedances are 50 Ω and 72 Ω. The impedance of the quarter-wave section is $\sqrt{50 \times 72} = 60$ Ω. The ground-plane spacing $2W + S$ is kept constant at 0.3 mm. The strip thickness is 0.002 mm. The substrate thickness is 0.4 mm and the dielectric constant is 12.9 with a loss tangent of 0.001. Use the computer program CPW to determine the required strip widths S_1, S_2, and S_3. Determine the length l of the quarter-wave matching section at a frequency of 10 GHz. How much attenuation occurs in the quarter-wave section? An accuracy of ±0.25 percent is sufficient.

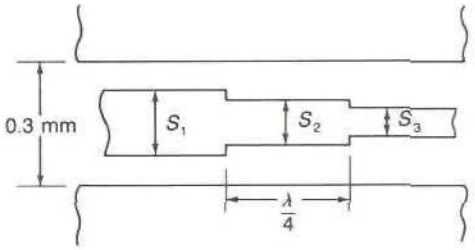


FIGURE P3.32

3.33. In a planar transmission line, the attenuation is 0.25 dB/cm. By what fraction is the wave amplitude reduced in propagating a distance of 1 cm on this transmission line?

3.34. Derive the equations (3.72) for TM waves.

3.35. Find the cutoff frequency for the TE_{10} mode in a rectangular waveguide with dimensions 4 cm by 2 cm. Find the guide wavelength λ_g and phase velocity at a frequency 25 percent higher than the cutoff frequency.

3.36. Derive the solution for a TE_{10} mode in a rectangular guide of wide dimension a and height b when the guide is filled with dielectric of permittivity ϵ. Show that the cutoff frequency is given by $f_c = c/2a\epsilon_r^{1/2}$, where c is the free-space velocity of light and ϵ_r is the dielectric constant. Show that the guide wavelength is smaller for a dielectric-filled guide than for an air-filled guide.

3.37. Obtain an expression for the attenuation of a TE_{10} mode in a dielectric-filled guide when $\epsilon = \epsilon_1 - j\epsilon_2$ but the walls are perfectly conducting. Obtain an exact expression and compare it with the results deduced by an application of the perturbation method.

3.38. Obtain a solution for an H wave in the parallel-plate transmission line with centered dielectric slab as illustrated in Fig. P3.38. Assume that the plates are perfectly conducting and infinitely wide. Can a TEM wave propagate in this structure? Why?

Hint: Assume $h_z = \cos k_d x$ for $|x| \le a/2$ and $h_z = Ae^{-p|x|}$ for $|x| > a/2$. Verify that $k_d^2 + p^2 = (\epsilon_r - 1)k_0^2$. Match the tangential fields at $x = a/2$ to obtain an equation for A and one relating the parameters p and k_d.

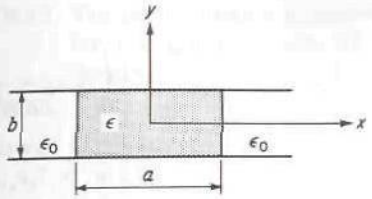


FIGURE P3.38

3.39. Obtain solutions for TE_{n0} modes in the partially filled waveguide illustrated in Fig. P3.39.

Hint: Assume that

$$h_z = \begin{cases} \cos k_d x & 0 < x < t \\ A \cos p(a - x) & t < x < a \end{cases}$$

and match the tangential fields at $x = t$. Thus show that

$$\beta^2 = k_0^2 - p^2 = \epsilon_r k_0^2 - k_d^2$$

and that

$$p \tan k_d t = -k_d \tan pd$$

Note that there are an infinite number of solutions for p and k_d corresponding to various TE_{n0} modes. Obtain numerical values for β, p, and k_d when $k_0 = 2$, $t = 1$ cm, $d = 1.5$ cm, and $\epsilon_r = 4$. Note that there is a lowest-order solution for p pure imaginary.

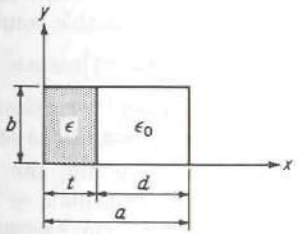


FIGURE P3.39

3.40. Obtain expressions for the surface currents of a TE_{10} mode in a rectangular guide. A narrow slot may be cut in a waveguide along a current flow line without appreciably disturbing the field. Show that, for the TE_{10} mode, narrow centered axial slots may be cut in the broad face of a rectangular guide. This principle is used in standing-wave detectors to provide suitable points of entry for a probe used to sample the interior waveguide field.

3.41. Use the computer program RECTWG to evaluate the parameters of a rectangular waveguide with width $a = 1$ cm, height $b = 0.4$ cm at a frequency of 20 GHz. How much attenuation will occur in a waveguide 5 m long?

3.42. For TE modes in a waveguide, write $\mathbf{H}_t = -I(z)\nabla_t h_z$, $\mathbf{E}_t = V(z)\mathbf{a}_z \times \nabla_t h_z$. Use Maxwell's equations to show that $V(z)$ and $I(z)$ satisfy the transmission-

line equations

$$\frac{dV}{dz} = -j\omega\mu_0 I \qquad \frac{dI}{dz} = -\left(j\omega\epsilon_0 + \frac{k_c^2}{j\omega\mu_0}\right)V$$

Construct an equivalent distributed-parameter circuit for these modes. For TM modes put $\mathbf{E}_t = -V(z)\nabla_t e_z$, $\mathbf{H}_t = -I(z)\mathbf{a}_z \times \nabla_t e_z$, and show that

$$\frac{dV}{dz} = -\left(j\omega\mu_0 + \frac{k_c^2}{j\omega\epsilon_0}\right)I \qquad \frac{dI}{dz} = -j\omega\epsilon_0 V$$

Construct an equivalent distributed-parameter circuit for these modes.†

3.43. Consider an infinitely long rectangular guide. The guide is filled with dielectric for $z \geq 0$, having a dielectric constant ϵ_r. An H_{10} mode is incident from $z < 0$. At $z = 0$, a reflected H_{10} mode and a transmitted H_{10} mode are produced because of the discontinuity. Show that the reflection coefficient is given by $(Z_2 - Z_1)/(Z_2 + Z_1)$, where Z_1 is the wave impedance in the empty guide and Z_2 is the wave impedance in the dielectric-filled guide. Show that the ratio of the wave impedances equals the ratio of the guide wavelengths.

3.44. Find the surface currents for the H_{01} mode in a circular guide.

3.45. Obtain an expression for power in a TE_{11} mode in a circular guide. (See App. II for Bessel-function integrals.)

3.46. Derive an expression for attenuation for TE_{0m} modes in a circular waveguide.

Answer: $\alpha = R_m f_{c,0m}^2/[aZ_0 f(f^2 - f_{c,0m}^2)^{1/2}]$.

3.47. Find the attenuation in decibels per mile for an H_{01} mode in a circular copper guide of 1 in diameter when operated at a frequency of 10 times the cutoff frequency.

3.48. Show that, in a coaxial line with inner radius a and outer radius b, there are solutions for TE_{nm} and TM_{nm} modes. A suitable solution for e_z and h_z is

$$[AJ_n(k_c r) + Y_n(k_c r)]\cos n\phi$$

Obtain equations (transcendental in nature) for determining the cutoff wave number k_c by imposing proper boundary conditions at $r = a, b$.

3.49. Use the computer program RIDGEWG to find the cutoff wavelength and frequency for a ridge waveguide with dimensions $a = 1$ cm, $b = 0.4$ cm, ridge width $= 0.5$ cm, and ridge spacing $S = 0.1$ cm. Compare this with the cutoff frequency of a standard waveguide with $a = 1$ cm and $b = 0.4$ cm.

3.50. Use the transverse-resonance technique to derive the eigenvalue equation for TE_{n0} modes in the partially filled rectangular guide of Prob. 3.39. Verify that the wave impedances in the x direction in the two regions are $k_0 Z_0/p$ and $kZ/k_d = k_0 Z_0/k_d$.

3.51. A rectangular waveguide with internal dimensions $a = 0.9$ in and $b = 0.4$ in (standard X-band waveguide) has a centered fin with a slot spacing $S = 2$ mm. Find the cutoff wavelength and compare it with that for the waveguide without fin loading. Use the computer program FINLINE.

†S. A. Schelkunoff, *Bell System Tech. J.*, vol. 34, p. 995, September, 1955.

3.52. The permittivity ϵ is generally a function of $\epsilon(\omega)$ of ω. Obtain an expression for the group velocity of a coaxial line filled with dielectric. Neglect the frequency dependence of the attenuation due to conductor loss.

3.53. A rectangular guide of dimensions $a = 2b = 2.5$ cm is operated at a frequency of 10^{10} Hz. A pulse-modulated carrier of the above frequency is transmitted through the guide. How much pulse delay time is introduced by a guide 100 m long?

REFERENCES

1. Ramo, S., J. R. Whinnery, and T. Van Duzer: "Fields and Waves in Communication Electronics," 2nd ed., John Wiley & Sons, Inc., New York, 1984.
2. Liboff, R. L., and G. C. Dalman: "Transmission Lines, Waveguides, and Smith Charts," MacMillan Publishing Company, New York, 1985.
3. Pozar, D. M.: "Microwave Engineering," Addison-Wesley Publishing Company, Reading, Mass. 1990.
4. Rizzi, P. A.: "Microwave Engineering—Passive Circuits," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.
5. Edwards, T. C.: "Foundations for Microstrip Circuit Design," John Wiley & Sons, Inc., New York, 1987.
6. Bahl, I., and P. Bhartia: "Microwave Solid State Circuit Design," John Wiley & Sons, Inc., New York, 1988.
7. Chang, K. (ed.): "Handbook of Microwave and Optical Components," vol. 1, John Wiley & Sons, Inc., New York, 1989.
8. Ishii, T. K.: "Microwave Engineering," 2nd ed., Harcourt Brace Jovanovich, New York, 1989.
9. Wolff, E. A., and R. Kaul: "Microwave Engineering and Systems," John Wiley & Sons, Inc., New York, 1988.
10. Baden Fuller, A. J.: "Microwaves," 3rd ed., Pergamon Press, New York, 1990.

CHAPTER 4

CIRCUIT THEORY FOR WAVEGUIDING SYSTEMS†

At low frequencies the interconnection of resistors, capacitors, and inductors results in a circuit. Such circuits are normally linear, so that the superposition principle may be used to find the response when more than one exciting source is present. Kirchhoff's laws form the basis for the analysis, whether in terms of loop currents or node voltages. In these low-frequency circuits the various elements are connected by conducting wires, and generally the length of these connecting wires is not critical or important.

At microwave frequencies equivalent reactive and resistive elements may also be connected to form a microwave circuit. In place of connecting wires, transmission lines and waveguides are used. The length of the connecting link is often several wavelengths, and hence propagation effects become very important. The analysis of microwave circuits is therefore by necessity somewhat more involved than that for the low-frequency case. The circuit theory of transmission-line circuits has been well developed for many decades, and, as will be shown the circuit theory for waveguide systems is formally the same.

†The basic theory of microwave circuits is developed in C. G. Montgomery, R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits," McGraw-Hill Book Company, New York, 1948. Much of the material presented here in Secs. 4.1 to 4.9 must of necessity be similar, in view of its basic nature.

Many of the circuit-analysis techniques and circuit properties that are valid at low frequencies are also valid for microwave circuits. In actual fact, low-frequency circuit analysis is a special case of microwave circuit analysis. As a consequence, a study of microwave circuits provides a deeper physical insight into conventional circuit theory. In this chapter the physical basis for a circuit theory for waveguiding systems is developed. In later chapters we shall utilize this foundation in the study of impedance matching, waveguide devices, resonators, filters, etc.

4.1 EQUIVALENT VOLTAGES AND CURRENTS

At microwave frequencies voltmeters and ammeters for the direct measurement of voltages and currents do not exist. For this reason voltage and current, as a measure of the level of electrical excitation of a circuit, do not play a primary role at microwave frequencies. On the other hand, it is useful to be able to describe the operation of a microwave circuit in terms of voltages, currents, and impedances in order to make optimum use of low-frequency circuit concepts. For the most part this can be done. There is, however, a notable difference, namely, the nonuniqueness of the voltages and currents in most instances. It was noted in the preceding chapter that for the TEM wave on a transmission line there existed a voltage and a current wave uniquely related to the transverse electric and magnetic fields, respectively. In the case of TE and TM modes in a waveguide, no unique voltage or current waves exist that have the same physical significance as those associated with the TEM wave on a transmission line. This result might have been anticipated since the guide boundary is a closed conducting boundary, and one is at a loss as to the two points on the boundary between which the voltage should be measured. Furthermore, if voltage is defined as the line integral of the transverse electric field between two chosen points on the boundary it is found that for TM waves the line integral is zero (Probs. 4.1 and 4.2), whereas for TE waves the value of the line integral depends on the path of integration that is chosen. For these reasons the introduction of voltage and current waves, to be associated with waveguide modes, is done on an equivalent basis and has formal significance only. The basis for the introduction of equivalent voltages and currents is discussed below.

In the previous chapter it was shown that propagating waveguide modes have the following properties:

1. Power transmitted is given by an integral involving the transverse electric and transverse magnetic fields only.
2. In a loss-free guide supporting several modes of propagation, the power transmitted is the sum of that contributed by each mode individually.

3. The transverse fields vary with distance along the guide according to a propagation factor $e^{\pm j\beta z}$ only.
4. The transverse magnetic field is related to the transverse electric field by a simple constant, the wave impedance of the mode; i.e.,

$$Z_w \mathbf{h} = \mathbf{a}_z \times \mathbf{e}$$

for a mode propagating in the $+z$ direction.

These properties suggest letting equivalent voltage and current waves be introduced proportional to the transverse electric and magnetic fields, respectively, since the transverse fields have properties similar to those of the voltage and current waves on a transmission line. That is, in actual fact, what is done.

A propagating waveguide mode may be expressed in general as

$$\mathbf{E} = C^+ \mathbf{e} e^{-j\beta z} + C^+ \mathbf{e}_z e^{-j\beta z} \tag{4.1a}$$

$$\mathbf{H} = C^+ \mathbf{h} e^{-j\beta z} + C^+ \mathbf{h}_z e^{-j\beta z} \tag{4.1b}$$

for propagation in the $+z$ direction, and

$$\mathbf{E} = C^- \mathbf{e} e^{j\beta z} - C^- \mathbf{e}_z e^{j\beta z} \tag{4.2a}$$

$$\mathbf{H} = -C^- \mathbf{h} e^{j\beta z} + C^- \mathbf{h}_z e^{j\beta z} \tag{4.2b}$$

for propagation in the $-z$ direction. In (4.1) and (4.2), C^+ and C^- are arbitrary amplitude constants. Note also that if the mode is a TE or a TM mode, then $\mathbf{e}_z$ and $\mathbf{h}_z$ is zero accordingly. Let the following equivalent voltage and current waves be introduced:

$$V = V^+ e^{-j\beta z} + V^- e^{j\beta z} \tag{4.3a}$$

$$I = I^+ e^{-j\beta z} - I^- e^{j\beta z} \tag{4.3b}$$

where $V^+ = K_1 C^+$, $V^- = K_1 C^-$, and $I^+ = K_2 C^+$, $I^- = K_2 C^-$. K_1 and K_2 are constants of proportionality that will establish the relationship between voltages and the transverse electric field and currents and the transverse magnetic field. In order to conserve power, it is necessary that

$$\frac{1}{2} V^+ (I^+)^* = \frac{|C^+|^2}{2} \int_S \mathbf{e} \times \mathbf{h}^* \cdot \mathbf{a}_z \, dS$$

or

$$K_1 K_2^* = \int \mathbf{e} \times \mathbf{h}^* \cdot \mathbf{a}_z \, dS \tag{4.4}$$

By proper normalization of the functions $\mathbf{e}$ and $\mathbf{h}$, the product $K_1 K_2^*$ can be made equal to unity. Although (4.4) provides one relationship between K_1 and K_2, a second relationship is required before they are determined. This second relationship can be chosen in a variety of ways. For example, the

voltage and current waves given by (4.3) may be thought of as existing on a fictitious transmission line that is equivalent to the waveguide. As such, it may be desirable to choose the characteristic impedance of this transmission line equal to unity, in which case

$$Z_c = \frac{V^+}{I^+} = \frac{V^-}{I^-} = \frac{K_1}{K_2} = 1 \tag{4.5}$$

As an alternative, it might be desirable to choose the characteristic impedance equal to the wave impedance, in which case

$$Z_c = \frac{K_1}{K_2} = Z_w \tag{4.6}$$

Other possibilities are obvious and equally valid. In this text either the definition (4.5) or (4.6) is used. The one that is used will be stated, or else it will be clear from the discussion which definition is being utilized. When the equivalent voltages and currents are chosen so that the equivalent transmission-line characteristic impedance is unity, we shall refer to them as *normalized* voltages and currents. Note that even though equivalent transmission lines may be used to represent a waveguide, the propagation constant of this line must be taken as that for the waveguide.

A waveguide supporting N propagating modes may now be formally represented as N fictitious transmission lines supporting equivalent voltage and current waves (from property 2 listed above for waveguide modes). Thus we have

$$V = \sum_{n=1}^{N} \left(V_n^+ e^{-j\beta_n z} + V_n^- e^{j\beta_n z} \right) \tag{4.7a}$$

$$I = \sum_{n=1}^{N} \left(I_n^+ e^{-j\beta_n z} - I_n^- e^{j\beta_n z} \right)$$

$$= \sum_{n=1}^{N} \left(V_n^+ Y_n e^{-j\beta_n z} - V_n^- Y_n e^{j\beta_n z} \right) \tag{4.7b}$$

where the Y_n are arbitrarily chosen characteristic admittances for the equivalent transmission lines. When an obstacle is inserted into a waveguide supporting N modes of propagation, these modes are in general coupled together by the obstacle. This coupling can be described in terms of an equivalent circuit made up of impedance elements. This impedance description of obstacles in waveguides is developed in the next section. Once the equivalent voltage and current amplitudes have been determined, the wave-

guide fields are known from the relations

$$\mathbf{E}_t = \sum_{n=1}^{N} (V_n^+ K_{1n}^{-1} e^{-j\beta_n z} + V_n^- K_{1n}^{-1} e^{j\beta_n z})\mathbf{e}_n \tag{4.8a}$$

$$\mathbf{H}_t = \sum_{n=1}^{N} (I_n^+ K_{2n}^{-1} e^{-j\beta_n z} - I_n^- K_{2n}^{-1} e^{j\beta_n z})\mathbf{h}_n \tag{4.8b}$$

and the specified proportionality constants K_{1n}, K_{2n} for each mode. The axial field components may be found from (4.8) by the use of Maxwell's equations. Note that the equivalent current wave amplitude for propagation in the $-z$ direction is expressed by $-I_n^-$, and hence the corresponding transverse magnetic field is proportional to $-K_{2n}^{-1} I_n^-$. When a waveguide supports several modes of propagation simultaneously at the same frequency, the number of electrical ports will exceed the number of physical ports. That is, power can be fed to a given load by means of any of the propagating modes, and all these modes may be common to a single physical waveguide input port.†

4.2 IMPEDANCE DESCRIPTION OF WAVEGUIDE ELEMENTS AND CIRCUITS

One-Port Circuits

A one-port circuit (equivalent to a two-terminal network) is a circuit for which power can enter or leave through a single waveguide or transmission line. A short-circuited transmission line and a short-circuited waveguide containing a metallic post as illustrated in Fig. 4.1 are examples of one-port circuits.

For one-port devices of the above type, a knowledge of any two of the four quantities $V^+, V^-, V = V^+ + V^-, I = I^+ - I^-$ will serve to describe the effect of the one-port device on an incident wave (it is assumed that the waveguide supports only one propagating mode). These quantities must, of course, be referred to a terminal plane such as t in Fig. 4.1 in order to be unambiguously specified. A terminal plane, or reference plane, is the equivalent of a terminal pair in a low-frequency network. In the present instance an impedance description is desired. If the total voltage and current at the

†The microwave equivalent-circuit theory presented in this chapter may be extended to include nonpropagating modes. However, when nonpropagating modes are included, the impedance and scattering matrices do not have the same properties as when only propagating modes are present at the terminal planes. See H. Haskal, Matrix Description of Waveguide Discontinuities in the Presence of Evanescent Modes, *IEEE Trans.*, vol. MTT-12, pp. 184–188, March, 1964.

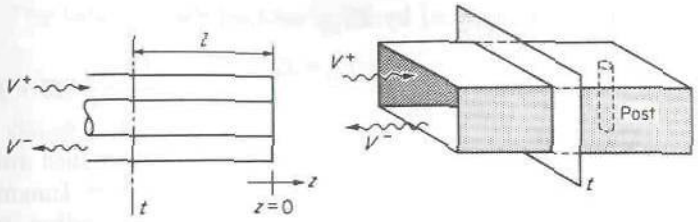


FIGURE 4.1
One-port circuits. (*a*) Short-circuited coaxial line; (*b*) short-circuited waveguide with post.

terminal plane are

$$V = V^+ + V^-$$
$$I = I^+ - I^- = Y_c(V^+ - V^-)$$

where Y_c is the equivalent characteristic admittance (actual characteristic admittance for the transmission line), the input impedance is given by

$$Z_{\text{in}} = \frac{V}{I} = \frac{V^+ + V^-}{V^+ - V^-} Z_c \tag{4.9}$$

The complex Poynting vector may be used to establish the physical nature of one-port impedance functions. From (2.59) and (2.60) we have

$$\frac{1}{2}\oint_S \mathbf{E} \times \mathbf{H}^* \cdot \mathbf{n}\, dS = P_l + 2j\omega(W_m - W_e) \tag{4.10}$$

where $\mathbf{n}$ is a unit inward normal to the closed surface S, P_l is the power dissipated in the volume bounded by S, and $W_m - W_e$ is the net reactive energy stored within S. For the surface S choose the terminal plane, the guide walls, and the short-circuiting plane. For perfectly conducting walls and short circuit, $\mathbf{n} \times \mathbf{E} = 0$; so the integral reduces to that over the terminal plane only. Thus

$$\frac{1}{2}\oint_t \mathbf{E} \times \mathbf{H}^* \cdot \mathbf{a}_z\, dS = P_l + 2j\omega(W_m - W_e) \tag{4.11}$$

Now at the terminal plane the transverse fields are [see (4.8)]

$$\mathbf{E}_t = K_1^{-1}(V^+ + V^-)\mathbf{e} = K_1^{-1}V\mathbf{e} \tag{4.12a}$$
$$\mathbf{H}_t = K_2^{-1}(I^+ - I^-)\mathbf{h} = K_2^{-1}I\mathbf{h} \tag{4.12b}$$

Hence (4.11) becomes

$$\frac{1}{2}(K_1K_2^*)^{-1}VI^*\int_t \mathbf{e} \times \mathbf{h}^* \cdot \mathbf{a}_z\, dS = \frac{1}{2}VI^* = P_l + 2j\omega(W_m - W_e) \tag{4.13}$$

If now V is replaced by IZ_{in}, we find that

$$Z_{\text{in}} = \frac{P_l + 2j\omega(W_m - W_e)}{\frac{1}{2}II^*} = R + jX \tag{4.14}$$

This relates the input impedance to the power loss and net reactive energy stored in the volume beyond the terminal plane. Inasmuch as the current I may be an equivalent current, the corresponding impedance Z_{in} is an equivalent one also. Since P_l, W_m, and W_e are all proportional to $|I^+|^2$, and hence also proportional to $|I|^2$ in view of the linearity of the field equations, the equivalent resistance R and reactance X in (4.14) are independent of the amplitude of the incident wave.

By replacing I^* by $Y_{\text{in}}^* V^*$ in (4.13), we obtain, after taking the complex conjugate,

$$Y_{\text{in}} = \frac{P_l - 2j\omega(W_m - W_e)}{\frac{1}{2}VV^*} = G + jB \tag{4.15}$$

for the input admittance of the one-port device. The susceptance B is positive (capacitive in nature) only if $W_e > W_m$.

The evaluation of an input impedance by means of the general definition (4.14) will be carried out for the simplest case, that of a short-circuited coaxial line. In the short-circuited coaxial line of Fig. 4.1, the fields in the one-port device are given by

$$\mathbf{E} = \frac{V^+}{\ln(b/a)} \frac{\mathbf{a}_r}{r} (e^{-jk_0 z} - e^{jk_0 z})$$

$$\mathbf{H} = \frac{Y_0 V^+}{\ln(b/a)} \frac{\mathbf{a}_\phi}{r} (e^{-jk_0 z} + e^{jk_0 z})$$

since the electric field must vanish at the short-circuited position $z = 0$. If the terminal plane is located at $z = -l$, then

$$\begin{aligned} W_e &= \frac{\epsilon_0}{4} \int_0^{2\pi} \int_a^b \int_{-l}^0 |E|^2 r \, d\phi \, dr \, dz \\ &= \frac{2\pi\epsilon_0 |V^+|^2}{\ln(b/a)} \int_{-l}^0 \sin^2 k_0 z \, dz \\ &= \frac{\pi\epsilon_0 |V^+|^2}{\ln(b/a)} \left(l - \frac{\sin 2k_0 l}{2k_0} \right) \end{aligned}$$

Similarly, it is found that

$$W_m = \frac{\pi\mu_0 Y_0^2 |V^+|^2}{\ln(b/a)} \left(l + \frac{\sin 2k_0 l}{2k_0} \right)$$

The total current at the terminal plane at $z = -l$ is

$$I = Y_c V^+ (e^{jk_0 l} + e^{-jk_0 l}) = 2Y_c V^+ \cos k_0 l$$

Using (4.14) now gives

$$Z_{\text{in}} = \frac{4j\omega\pi\epsilon_0 |V^+|^2}{\ln(b/a)} \frac{\sin 2k_0 l}{k_0 (4Y_c^2 |V^+|^2) \cos^2 k_0 l}$$

$$= \frac{j\omega\pi\epsilon_0}{\ln(b/a)} \frac{\sin 2k_0 l}{k_0 Y_c^2 \cos^2 k_0 l} = jZ_c \tan k_0 l \qquad (4.16)$$

on using the relations $Y_c = 2\pi Y_0/[\ln(b/a)]$, $\sin 2k_0 l = 2 \sin k_0 l \cos k_0 l$, and $k_0 = \omega(\mu_0\epsilon_0)^{1/2}$. This result for the input impedance of a short-circuited coaxial line, *as obtained from the general definition* (4.14), agrees with the simple computation based directly on expressions for the total voltage and current at the terminal plane. However, the purpose of introducing (4.14) was not as a computational tool, but rather for the physical insight it provides into the nature of the impedance function for a one-port circuit.

The second example of a one-port circuit as illustrated in Fig. 4.1 cannot be evaluated in as straightforward a manner as for the coaxial line because it does not consist of a uniform unperturbed waveguide. The presence of a conducting post within the termination results in induced currents on the post that will excite a multitude of waveguide modes. However, since it is assumed that only one mode propagates (the TE_{10} mode), all the other modes decay exponentially in both directions away from the post. By choosing the terminal plane sufficiently far away from the post, the fields at this plane are essentially just those of the incident and reflected dominant modes. The evanescent modes excited by the post will store reactive energy, and this will contribute to the input reactance as viewed from the terminal plane, as reference to (4.14) shows. The presence of the post within the termination modifies the input reactance by changing the amplitude of the reflected dominant wave in just the right amount to account for the additional reactive energy stored.

As seen from the preceding discussion, it is important when dealing with waveguide structures to choose terminal planes sufficiently far away from obstacles that excite evanescent modes, so that only dominant-mode fields have significant amplitudes at these reference planes. This will ensure that all the reactive energy associated with the nonpropagating modes that make up the fringing field around the obstacle is taken into account in the expression for the input reactance. This precaution is particularly important in any experimental setup used to measure the impedance function for a particular obstacle. Once the impedance has been properly determined at a given terminal plane, it may be referred to any other terminal plane by

using the impedance-transformation formula

$$Z(l_2) = Z_c \frac{Z(l_1) + jZ_c \tan \beta(l_2 - l_1)}{Z_c + jZ(l_1) \tan \beta(l_2 - l_1)} \tag{4.17}$$

where l_1 is the location of terminal plane 1 and l_2 specifies the location of the new terminal plane. In particular, shift in the terminal-plane position by a multiple of $\lambda_g/2$ leaves the impedance invariant. Thus an impedance may be referred to terminal planes located in the near vicinity of an obstacle where now it is understood that this impedance describes the effect of the obstacle on the dominant mode only, and does not imply that the total field at this particular terminal plane is that of the dominant mode only. In other words, the impedance description of a waveguide element or obstacle gives information on the effect this element has on the dominant propagating mode but does not give any information on the detailed field structure near the obstacle. Fortunately, the latter information is rarely required.

Lossless One-Port Termination

If there are no losses present in a one-port circuit, the input impedance is a pure reactance given by

$$jX = \frac{4j\omega(W_m - W_e)}{II^*} \tag{4.18}$$

The assumption of a lossless structure is often a very good approximation for microwave circuits. If $W_m = W_e$, the input reactance vanishes and a condition of resonance exists. There are actually two possibilities, namely, $W_m = W_e$ but $I \neq 0$, and $W_m = W_e$ but $V \neq 0$. The first corresponds to a zero in the input reactance (series resonance), whereas the second corresponds to a zero in the input susceptance (parallel resonance) as given by

$$jB = \frac{4j\omega(W_e - W_m)}{VV^*} \tag{4.19}$$

When the input reactance is zero, the input susceptance must be infinite, which implies that $V = 0$ at the terminal plane. This latter condition is possible since, for a pure reactive termination, all the incident power is reflected, so that the total voltage along the waveguide is a standing wave of the form $\sin \beta l$. In the case of a zero for the susceptance function B, the reactance X must be infinite (have a pole) and I must vanish at the terminal plane. It may be anticipated, then, that the reactance and susceptance functions will have a number of zeros and poles, i.e., frequencies at which they vanish or become infinite. This behavior is clearly evident in the expression for the reactance of a short-circuited coaxial line, which is

$$jX = jZ_c \tan k_0 l = jZ_c \tan \frac{\omega l}{c} \tag{4.20}$$

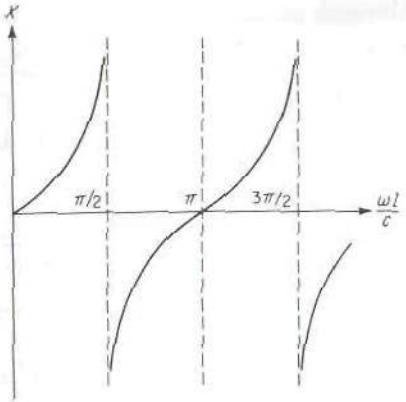


FIGURE 4.2
Input reactance of a short-circuited coaxial line.

A plot of X against frequency is given in Fig. 4.2. In particular, note that *the slope of the reactance is always positive; that is,* $\partial X/\partial\omega > 0$. This positive-slope condition means that the poles and zeros of X must alternate as ω is increased from zero to infinity. We shall show below that this is a general property of any reactive one-port circuit, a result known as Foster's reactance theorem. First, however, it will be instructive to rewrite (4.20) by using the infinite-product representation and also the partial-fraction expansion of the tangent function:†

$$X = Z_c \tan k_0 l = \frac{Z_c \dfrac{\omega l}{c} \displaystyle\prod_{n=1}^{\infty} \left[1 - \left(\frac{\omega l}{n\pi c}\right)^2\right]}{\displaystyle\prod_{n=0}^{\infty} \left[1 - \frac{(\omega l)^2}{\left(n + \frac{1}{2}\right)^2 (\pi c)^2}\right]}$$

$$= Z_c \frac{2\omega c}{l} \sum_{n=1,3,\ldots}^{\infty} \frac{1}{(n\pi c/2l)^2 - \omega^2} \tag{4.21}$$

The first form contains the product of an infinite number of factors in both the numerator and denominator and clearly exhibits both the zeros and poles and their alternating occurrence. The second form exhibits the poles very clearly, but information on the zero locations is lost. In the vicinity of a pole, say that at $\omega \approx \omega_n = n\pi c/2l$, all terms in the partial-fraction expan-

†E. A. Guillemin, "The Mathematics of Circuit Analysis," chap. 6, John Wiley & Sons, Inc., New York, 1949.

E. T. Copson, "Theory of Functions of a Complex Variable," Oxford University Press, Fair Lawn, N.J., 1935.

J. Pierpont, "Functions of a Complex Variable," Dover Publications, Inc., New York, 1959.

sion are small except the nth term, so that

$$X \approx \frac{2\omega c}{l} \frac{Z_c}{(\omega_n - \omega)(\omega_n + \omega)} \approx \frac{cZ_c}{l(\omega_n - \omega)} \tag{4.22}$$

since $\omega \approx \omega_n$. This behavior near a pole is similar to that for a simple LC parallel network for which

$$X = \frac{-\omega L}{\omega^2 LC - 1} \approx \frac{\omega_0}{2}\sqrt{\frac{L}{C}} \frac{1}{\omega_0 - \omega} \tag{4.23}$$

where $\omega_0 = (LC)^{-1/2}$. However, the microwave network is a good deal more complicated, for it has an infinite number of poles and zeros, and not just a double zero and a single pole as a simple parallel LC circuit has [the zeros occur at $\omega = 0$, where ωL vanishes, and at infinity, where $(\omega C)^{-1}$ vanishes]. These similarities and differences are important to note since they are characteristic of microwave networks in general, even though we have demonstrated some of these properties for a short-circuited coaxial line only.

*4.3 FOSTER'S REACTANCE THEOREM

The theorem that will now be proved is that the rate of change of the reactance X and susceptance B with ω is positive. Once this result is established, it follows that the poles and zeros of a reactance function must alternate in position along the ω axis. Figure 4.3 illustrates a general one-port reactive termination. The field within the termination satisfies Maxwell's equations

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \qquad \nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E}$$

The derivative with respect to ω of the complex conjugate of these equations gives

$$\nabla \times \frac{\partial \mathbf{E}^*}{\partial \omega} = j\omega\mu \frac{\partial \mathbf{H}^*}{\partial \omega} + j\mathbf{H}^* \frac{\partial \omega\mu}{\partial \omega} \qquad \nabla \times \frac{\partial \mathbf{H}^*}{\partial \omega} = -j\omega\epsilon \frac{\partial \mathbf{E}^*}{\partial \omega} - j\mathbf{E}^* \frac{\partial \omega\epsilon}{\partial \omega}$$

Consider next the quantity

$$\nabla \cdot \left(\mathbf{E} \times \frac{\partial \mathbf{H}^*}{\partial \omega} + \frac{\partial \mathbf{E}^*}{\partial \omega} \times \mathbf{H}\right) = \nabla \times \mathbf{E} \cdot \frac{\partial \mathbf{H}^*}{\partial \omega} - \mathbf{E} \cdot \nabla \times \frac{\partial \mathbf{H}^*}{\partial \omega} + \nabla \times \frac{\partial \mathbf{E}^*}{\partial \omega} \cdot \mathbf{H} - \frac{\partial \mathbf{E}^*}{\partial \omega} \cdot \nabla \times \mathbf{H}$$

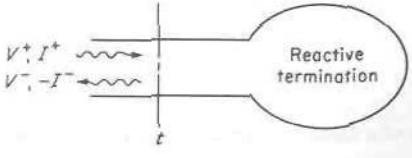


FIGURE 4.3
A one-port reactive termination.

Substituting from above gives

$$\nabla \cdot \left(\mathbf{E} \times \frac{\partial \mathbf{H}^*}{\partial \omega} + \frac{\partial \mathbf{E}^*}{\partial \omega} \times \mathbf{H} \right) = j \left(\mathbf{H} \cdot \mathbf{H}^* \frac{\partial \omega \mu}{\partial \omega} + \mathbf{E} \cdot \mathbf{E}^* \frac{\partial \omega \epsilon}{\partial \omega} \right)$$
$$+ j\omega \left(\mu \mathbf{H} \cdot \frac{\partial \mathbf{H}^*}{\partial \omega} - \mu \mathbf{H} \cdot \frac{\partial \mathbf{H}^*}{\partial \omega} + \epsilon \mathbf{E} \cdot \frac{\partial \mathbf{E}^*}{\partial \omega} - \epsilon \mathbf{E} \cdot \frac{\partial \mathbf{E}^*}{\partial \omega} \right)$$

The second term on the right-hand side vanishes; so we have

$$\nabla \cdot \left(\mathbf{E} \times \frac{\partial \mathbf{H}^*}{\partial \omega} + \frac{\partial \mathbf{E}^*}{\partial \omega} \times \mathbf{H} \right) = j \left(\mathbf{H} \cdot \mathbf{H}^* \frac{\partial \omega \mu}{\partial \omega} + \mathbf{E} \cdot \mathbf{E}^* \frac{\partial \omega \epsilon}{\partial \omega} \right)$$

If we integrate throughout the volume of the termination and use the divergence theorem on the left-hand side, we obtain

$$\oint_S \left(\mathbf{E} \times \frac{\partial \mathbf{H}^*}{\partial \omega} + \frac{\partial \mathbf{E}^*}{\partial \omega} \times \mathbf{H} \right) \cdot d\mathbf{S} = -j \int_V \left(\mathbf{H} \cdot \mathbf{H}^* \frac{\partial \omega \mu}{\partial \omega} + \mathbf{E} \cdot \mathbf{E}^* \frac{\partial \omega \epsilon}{\partial \omega} \right) dV$$
$$= -4j(W_m + W_e) \qquad (4.24a)$$

where $W_m + W_e$ is the total time-average energy stored in the lossless termination, as reference to (2.53) in Sec. 2.5 shows, and $d\mathbf{S}$ is chosen directed into the volume.

Since both $\mathbf{n} \times \mathbf{E}$ and $\mathbf{n} \times \partial \mathbf{E}/\partial \omega$, where $\mathbf{n}$ is a unit inward normal, vanish on the perfectly conducting waveguide walls, the surface integral reduces to an integral over the terminal plane t only. On the terminal plane we have

$$\int_t \left(\mathbf{E} \times \frac{\partial \mathbf{H}^*}{\partial \omega} + \frac{\partial \mathbf{E}^*}{\partial \omega} \times \mathbf{H} \right) \cdot \mathbf{n} \, dS = V \frac{\partial I^*}{\partial \omega} + \frac{\partial V^*}{\partial \omega} I \qquad (4.24b)$$

where V and I are the equivalent terminal voltage and current. Now $V = jIX$ for a lossless reactive termination; so

$$\frac{\partial V^*}{\partial \omega} = -jX \frac{\partial I^*}{\partial \omega} - jI^* \frac{\partial X}{\partial \omega}$$

Thus

$$V \frac{\partial I^*}{\partial \omega} + \frac{\partial V^*}{\partial \omega} I = jXI \frac{\partial I^*}{\partial \omega} - jXI \frac{\partial I^*}{\partial \omega} - jII^* \frac{\partial X}{\partial \omega}$$

and hence we find that (4.24b) yields

$$II^* \frac{\partial X}{\partial \omega} = 4(W_m + W_e)$$

or

$$\frac{\partial X}{\partial \omega} = \frac{4(W_m + W_e)}{II^*} \qquad (4.25)$$

The right-hand side is proportional to the total energy stored in the termination and can never be negative. Consequently, the slope of the reactance function must always be positive. If I is replaced by jBV in

(4.24*b*), it is readily found that

$$\frac{\partial B}{\partial \omega} = \frac{4(W_m + W_e)}{VV^*} \tag{4.26}$$

and hence the susceptance is also an increasing function of frequency. The above relations also show that the frequency sensitivity of the reactance or susceptance is proportional to the total average energy stored. These relations are readily verified in the case of simple LC reactive networks, and a problem calling for this verification is given at the end of this chapter.

*4.4 EVEN AND ODD PROPERTIES OF Z_{in}

Before terminating the discussion dealing with one-port impedance functions, one further general property should be pointed out. This property is that the real part of $Z_{in} = R + jX$ is an even function of ω, whereas the imaginary part is an odd function. The physical necessity of this was pointed out in the previous chapter in the section dealing with group velocity. The property stems from the requirement that the response of a circuit to a real-time-dependent driving function must also be real. That is, if $\mathscr{V}(t)$ is the applied voltage at the terminal plane, the frequency spectrum is given by the Fourier transform:

$$V(\omega) = \int_{-\infty}^{\infty} e^{-j\omega t}\mathscr{V}(t)\,dt \tag{4.27}$$

The frequency spectrum of the current that flows is

$$I(\omega) = \frac{V(\omega)}{Z_{in}(\omega)} = \frac{V}{R + jX} \tag{4.28}$$

The current as a function of time is

$$\mathscr{I}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{V(\omega)}{R(\omega) + jX(\omega)} e^{j\omega t}\,d\omega \tag{4.29}$$

and must be a real function. This will be the case if

$$\frac{V(-\omega)}{R(-\omega) + jX(-\omega)} = \frac{V^*(\omega)}{[R(\omega) + jX(\omega)]^*}$$

for then (4.29) becomes

$$\mathscr{I}(t) = \frac{1}{2\pi}\int_{-\infty}^{0} \frac{V(\omega)e^{j\omega t}}{R(\omega) + jX(\omega)}\,d\omega + \frac{1}{2\pi}\int_{0}^{\infty} \frac{V(\omega)e^{j\omega t}}{R(\omega) + jX(\omega)}$$

which we can show is a real function. In the first integral on the right,

replace ω by $-\omega$ to obtain

$$\mathscr{I}(t) = \frac{1}{2\pi}\int_0^\infty \left[\frac{V(-\omega)e^{-j\omega t}}{R(-\omega)+jX(-\omega)} + \frac{V(\omega)e^{j\omega t}}{R(\omega)+jX(\omega)}\right] d\omega \quad (4.30)$$

The two terms in the integrand are complex conjugates of each other, and hence the sum is real; i.e.,

$$\mathscr{I}(t) = \frac{1}{\pi}\,\mathrm{Re}\int_0^\infty \frac{V(\omega)e^{j\omega t}}{R(\omega)+jX(\omega)}\, d\omega \quad (4.31)$$

The condition specified on $V/(R+jX)$ is satisfied by V alone and by $R+jX$ alone. Clearly, from (4.27), $V(-\omega) = V^*(\omega)$. If

$$R(-\omega)+jX(-\omega) = [R(\omega)+jX(\omega)]^*$$

then $$R(-\omega) = R(\omega) \qquad X(-\omega) = -X(\omega)$$

and R is an even function of ω, whereas X is an odd function of ω, as was to be proved. These even and odd properties are useful to know when approximate expressions for impedance functions are constructed from experimental data. For example, a series such as

$$a_1\omega + a_3\omega^3 + a_5\omega^5 + \cdots$$

could be used to represent X, but not to represent R, since the series is an odd function of ω.

4.5 *N*-PORT CIRCUITS

Figure 4.4 illustrates the junction of N waveguides or transmission lines (or a combination of the two) that terminate in a common region or junction. The region between the N chosen terminal planes may contain any arbitrary collection of passive elements. If each guide can support only one mode of propagation, this circuit constitutes an N-port microwave circuit. If one or more of the guides can support several independent modes of propaga-

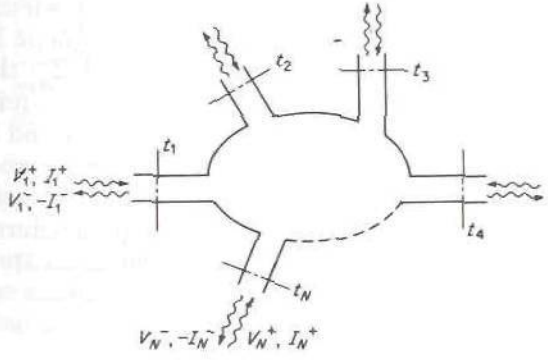


FIGURE 4.4
An N-port microwave circuit.

tion, the number of electrical ports exceeds the number of mechanical ports. Each mode, since it carries power independently of all other modes, corresponds to an electrical port through which power may enter or leave the junction. To simplify the discussion, it will be assumed that each guide supports only a single propagating mode. The extension of the theory to the case when some or all of the guides may support several propagating modes is more or less obvious.

Let the terminal planes be chosen sufficiently far from the junction so that the fields on the terminal planes are essentially just those of the incident and reflected dominant propagating modes. These fields may be uniquely defined at the terminal planes in terms of suitably defined equivalent voltages and currents. Clearly, the amplitudes of all the incident waves may be arbitrarily specified, i.e., chosen independently. The amplitudes of all the reflected waves are then determined by the physical properties of the junction; i.e., all the V_n^- are linear functions of the V_n^+. When the V_n^+ and V_n^- are known, the corresponding currents I_n^+, I_n^- are known from the relations

$$I_n^+ = Y_n V_n^+ \qquad I_n^- = Y_n V_n^-$$

Since Maxwell's equations are linear and the junction is assumed to be linear in its behavior, any N linearly independent combinations of the $4N$ quantities V_n^+, V_n^-, I_n^+, and I_n^- may be chosen as the independent variables to describe the electrical behavior of the junction. For an impedance description the total currents $I_n = I_n^+ - I_n^-$ at the terminal planes are chosen as independent variables. The N total terminal-plane voltages $V_n = V_n^+ + V_n^-$ are then the dependent variables, and are linearly related to the currents as follows:

$$\begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ V_N \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1N} \\ Z_{21} & Z_{22} & \cdots & Z_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ Z_{N1} & Z_{N2} & \cdots & Z_{NN} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \cdots \\ I_N \end{bmatrix} \tag{4.32}$$

The matrix of elements Z_{ij} is the impedance matrix and provides a complete description of the electrical properties of the N-port circuit. Some of the properties of this impedance matrix are discussed below.

If the junction contains a nonreciprocal medium such as a plasma (ionized gas) or a ferrite with an applied dc magnetic biasing field (ferrites are discussed in Chap. 6), then, in general, $Z_{ij} \neq Z_{ji}$; that is, the impedance matrix $[Z]$ is not symmetrical. The junction then requires $2N^2$ parameters to describe it completely since each Z_{ij} is complex and has two independent terms. If the junction does not contain any nonreciprocal media, $Z_{ij} = Z_{ji}$, and the impedance matrix is symmetrical. In this instance a total of only $2N^2 - (N^2 - N) = N(N + 1)$ independent parameters are required to describe the junction since $N^2 - N$ of the parameters are equal. Finally, if the junction is lossless—and many microwave junctions may be approximated as such with negligible error—then all the Z_{ij} must be pure imaginary

since there can be no power loss within the junction. In this case there are only $\frac{1}{2}N(N+1)$ independent parameters required for a complete description. Any network containing the required number of resistive and reactive elements may be used as a representation for the junction at a given frequency. However, it must be kept in mind that when the frequency is changed, the values of the network elements (resistance, capacitance, and inductance) must also be changed. Rarely would any one particular network representation provide a complete description of the junction over a band of frequencies unless the network parameters are changed in value when the frequency is changed.

The foregoing discussion applies also to the admittance matrix $[Y]$, which relates the currents to the total voltages at the terminal planes; i.e.,

$$\begin{bmatrix} I_1 \\ I_2 \\ \cdots \\ I_N \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1N} \\ Y_{21} & Y_{22} & \cdots & Y_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ Y_{N1} & Y_{N2} & \cdots & Y_{NN} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ V_N \end{bmatrix} \tag{4.33}$$

The impedance and admittance matrices are reciprocals of each other; so

$$[Y] = [Z]^{-1} \tag{4.34}$$

*Proof of Symmetry for the Impedance Matrix

The symmetry of the impedance matrix is readily proved when the junction contains media characterized by scalar parameters μ and ϵ. Let incident-wave amplitudes V_n^+ be so chosen that the total voltage V_n equals zero at all terminal planes except the ith plane. Let the corresponding field solution be $\mathbf{E}_i, \mathbf{H}_i$. Similarly, let a second solution $\mathbf{E}_j, \mathbf{H}_j$ correspond to the case when incident-wave amplitudes are chosen so that all V_n equal zero except V_j. The *Lorentz reciprocity theorem [Eq. (2.135)] gives*

$$\oint_S (\mathbf{E}_i \times \mathbf{H}_j - \mathbf{E}_j \times \mathbf{H}_i) \cdot \mathbf{n}\, dS = 0$$

when there are no sources within the closed surface S. Let S consist of the conducting walls bounding the junction and the N terminal planes. The integral over the walls vanishes if they are perfectly conducting or if they exhibit a surface impedance Z_m (Sec. 2.12). Therefore we obtain an integral over the terminal planes only, i.e.,

$$\sum_{n=1}^{N} \int_{t_n} (\mathbf{E}_i \times \mathbf{H}_j - \mathbf{E}_j \times \mathbf{H}_i) \cdot \mathbf{n}\, dS = 0 \tag{4.35}$$

However, for the particular solutions considered here, $\mathbf{n} \times \mathbf{E}_i$ and $\mathbf{n} \times \mathbf{E}_j$, that is, $\mathbf{E}_{ti}, \mathbf{E}_{tj}$ are zero on all terminal planes except t_i and t_j, respectively, since all V_n except V_i and V_j have been chosen equal to zero. Thus (4.35)

becomes

$$\int_{t_i} \mathbf{E}_i \times \mathbf{H}_j \cdot \mathbf{n}\, dS = \int_{t_j} \mathbf{E}_j \times \mathbf{H}_i \cdot \mathbf{n}\, dS$$

or

$$V_i(I_i)_j = V_j(I_j)_i \tag{4.36}$$

where $(I_i)_j$ is the current at the terminal plane i arising from an applied voltage at plane j, and similarly for $(I_j)_i$. From the admittance description (4.33) of the junction, we have

$$I_i = (I_i)_j = Y_{ij}V_j \qquad \text{for } V_n = 0,\ n \neq j$$

$$I_j = (I_j)_i = Y_{ji}V_i \qquad \text{for } V_n = 0,\ n \neq i$$

Hence (4.36) gives

$$V_iV_jY_{ij} = V_jV_iY_{ji}$$

or

$$Y_{ij} = Y_{ji} \tag{4.37}$$

which proves the symmetry of the admittance matrix. Since the reciprocal of a symmetrical matrix is a symmetrical matrix also, it follows that the impedance matrix is also symmetrical. The symmetry of the impedance and admittance matrices is a consequence of reciprocity. For nonreciprocal media, μ or ϵ (or both) are nonsymmetrical matrices, and (2.135) no longer applies. In this case the impedance matrix is no longer symmetrical. Nonreciprocal microwave devices are discussed in Chap. 6; so no further comments on these are made in this section.

*Proof of Imaginary Nature of [Z] for a Lossless Junction

For a lossless junction all the elements in the impedance and admittance matrices are pure imaginary. Let $[V]]$ and $[I]$ be column matrices representing the terminal voltages and currents, respectively. The transposed matrices $[V]_t, [I]_t$ are row matrices of the form

$$[V]_t = [V_1 \quad V_2 \quad \cdots \quad V_N]$$

$$[I]_t = [I_1 \quad I_2 \quad \cdots \quad I_N]$$

The total complex power into the junction is

$$\frac{1}{2}[I^*]_t[V] = \frac{1}{2}[I^*]_t[Z][I]$$

$$= \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} I_n^* Z_{nm} I_m = P_l + 2j\omega(W_m - W_e)$$

For a lossless junction $P_l = 0$ and the double sum must be pure imaginary. Since the I_n can be chosen as independent variables, they may all be chosen

as zero except for the nth one. In this case

$$\mathrm{Re}(I_n^* Z_{nn} I_n) = 0$$

or

$$\mathrm{Re}\, Z_{nn} = 0$$

If all but I_n and I_m are chosen equal to zero, we obtain

$$\mathrm{Re}\left[(I_n^* I_m + I_m^* I_n) Z_{nm} + I_n I_n^* Z_{nn} + I_m I_m^* Z_{mm}\right] = 0$$

But $I_n I_n^*$, $I_m I_m^*$, and $I_n^* I_m + I_n I_m^*$ are all real quantities and Z_{nn}, Z_{mm} are imaginary; so this equation can hold only if

$$\mathrm{Re}\, Z_{nm} = 0$$

Therefore all Z_{nm} are pure imaginary for a lossless junction.

Normalized Impedance and Admittance Matrices

Let us assume that we have chosen equivalent voltages and currents V_n, I_n so that the nth equivalent transmission line has a characteristic impedance Z_n given by

$$Z_n = \frac{V_n^+}{I_n^+} = \frac{V_n^-}{I_n^-}$$

For an N-port circuit the impedance description is given by (4.32). The impedance-matrix elements are Z_{nm}. We now wish to redefine the equivalent voltages and currents so that each transmission line has unity characteristic impedance and to find the new impedance-matrix elements which will be designated by $\bar{Z}_{nm}$ and called normalized elements.

Let the new voltage and current amplitudes be $\bar{V}_n^+$, $\bar{V}_n^-$, $\bar{I}_n^+$, and $\bar{I}_n^-$. In order to have the same power flow, we require

$$\tfrac{1}{2} V_n^+ I_n^+ = \tfrac{1}{2} \bar{V}_n^+ \bar{I}_n^+$$

We can express the power flow in the following two equivalent ways also:

$$\tfrac{1}{2} V_n^+ V_n^+ Y_n = \tfrac{1}{2} I_n^+ I_n^+ Z_n = \tfrac{1}{2} \bar{V}_n^+ \bar{V}_n^+ = \tfrac{1}{2} \bar{I}_n^+ \bar{I}_n^+$$

since for the normalized voltage and current amplitudes $\bar{V}_n^+ / \bar{I}_n^+ = 1$. From the last two relations for power flow, we see that the normalized voltage and current amplitudes are given by

$$\bar{V}_n^+ = \sqrt{Y_n}\, V_n^+$$

$$\bar{I}_n^+ = \sqrt{Z_n}\, I_n^+$$

From these relations it readily follows that

$$\bar{V}_n = \sqrt{Y_n}\, V_n \qquad \bar{I}_n = \sqrt{Z_n}\, I_n$$

and

$$V_n = \sqrt{Z_n}\, \bar{V}_n \qquad I_n = \sqrt{Y_n}\, \bar{I}_n$$

If we substitute from these expressions into (4.32), we obtain

$$\begin{bmatrix} \bar{V}_1 \\ \bar{V}_2 \\ \vdots \\ \bar{V}_N \end{bmatrix} = \begin{bmatrix} \sqrt{Y_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{Y_n} & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & & & \sqrt{Y_N} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{Y_1} & 0 & \cdots & 0 \\ 0 & \sqrt{Y_2} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \sqrt{Y_N} \end{bmatrix} \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1N} \\ Z_{21} & & \cdots & Z_{2N} \\ \vdots & & & \\ Z_{N1} & & \cdots & Z_{NN} \end{bmatrix}$$

$$\times \begin{bmatrix} \sqrt{Y_1} & 0 & \cdots & 0 \\ 0 & \sqrt{Y_2} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \sqrt{Y_N} \end{bmatrix} \begin{bmatrix} \bar{I}_1 \\ \bar{I}_2 \\ \vdots \\ \bar{I}_N \end{bmatrix}$$

$$= \begin{bmatrix} Y_1 Z_{11} & \sqrt{Y_1 Y_2}\, Z_{12} & \cdots & \sqrt{Y_1 Y_N}\, Z_{1N} \\ \sqrt{Y_2 Y_1}\, Z_{21} & Y_2 Z_{22} & \cdots & \sqrt{Y_2 Y_N}\, Z_{2N} \\ \vdots & & & \\ \sqrt{Y_N Y_1}\, Z_{N1} & \cdots & & Y_N Z_{NN} \end{bmatrix} \begin{bmatrix} \bar{I}_1 \\ \bar{I}_2 \\ \vdots \\ \bar{I}_N \end{bmatrix}$$

From this expression, we see that the elements of the normalized impedance matrix are given by

$$\bar{Z}_{nm} = \frac{Z_{nm}}{\sqrt{Z_n Z_m}} \tag{4.38a}$$

A similar analysis shows that the elements of the normalized admittance matrix are given by

$$\bar{Y}_{nm} = \frac{Y_{nm}}{\sqrt{Y_n Y_m}} \tag{4.38b}$$

4.6 TWO-PORT JUNCTIONS

At this point it seems advisable to examine the special case of the two-port junction rather than continue with the general theory of N-port circuits. The derivation of some further properties of the N-port junction is called for in some of the problems at the end of this chapter. Three examples of

two-port junctions are shown in Fig. 4.5. The first is the junction of two rectangular guides of unequal height (called an E-plane step since the E vector of the TE_{10} mode lies in the plane containing the step geometry). The second is a symmetrical junction consisting of two similar rectangular guides joined by an intermediate guide of greater width. The third two-port junction consists of a typical coaxial-line–waveguide junction, where the center conductor of the coaxial line extends into the rectangular guide to provide an antenna radiating energy into or coupling energy out of the rectangular guide. A discussion of these particular junctions will serve to develop the general impedance description of two-port junctions or circuits.

Since evanescent modes are excited at each discontinuity, the terminal planes are chosen far enough away so that these decaying waves have negligible amplitudes at the terminal planes. Equivalent voltages and currents are introduced proportional to the total transverse electric and magnetic fields, respectively, at each terminal plane. For example, for the junction in Fig. 4.5a, let the incident and reflected transverse fields of the TE_{10} mode at each terminal plane be (coordinates x, y, z refer to the left-hand-side guide, and the primed coordinates x', y', z' refer to the right-hand-side guide, as in Fig. 4.6)

$$\mathbf{E}_t = (C_1^+ e^{j\beta l_1} + C_1^- e^{-j\beta l_1})\mathbf{a}_y \sin\frac{\pi x}{a} \qquad \text{at } t_1$$

$$\mathbf{H}_t = -Y_w(C_1^+ e^{j\beta l_1} - C_1^- e^{-j\beta l_1})\mathbf{a}_x \sin\frac{\pi x}{a} \qquad \text{at } t_1$$

$$\mathbf{E}_t = (C_1^+ e^{j\beta l_2} + C_2^- e^{-j\beta l_2})\mathbf{a}_y \sin\frac{\pi x'}{a} \qquad \text{at } t_2$$

$$\mathbf{H}_i = Y_w(C_2^+ e^{j\beta l_2} - C_2^- e^{-j\beta l_2})\mathbf{a}_x \sin\frac{\pi x'}{a} \qquad \text{at } t_2$$

These expressions are obtained by choosing $h_z = (j\pi Y_w/\beta a)\cos(\pi x/a)$ in order to simplify the expressions for the transverse fields. Let the equivalent voltages and currents be chosen as

$$V_1^+ = K_1 C_1^+ e^{j\beta l_1} \qquad V_1^- = K_1 C_1^- e^{-j\beta l_1}$$

$$I_1^+ = Y_w K_1 C_1^+ e^{j\beta l_1} \qquad I_1^- = Y_w K_1 C_1^- e^{-j\beta l_1}$$

$$V_2^+ = K_2 C_2^+ e^{j\beta l_2} \qquad V_2^- = K_2 C_2^- e^{-j\beta l_2}$$

$$I_2^+ = Y_w K_2 C_2^+ e^{j\beta l_2} \qquad I_2^- = Y_w K_2 C_2^- e^{-j\beta l_2}$$

In order to simplify the notation, we have expressed K_{11} as K_1 and K_{21} as K_2 and eliminated the second constants K_{12} and K_{22} by introducing the wave admittance using (4.6). Thus the characteristic impedance of the equivalent transmission line is equal to the wave impedance $Z_w = Y_w^{-1}$ of the TE_{10} mode. To conserve power, it is necessary to choose K_1 and K_2

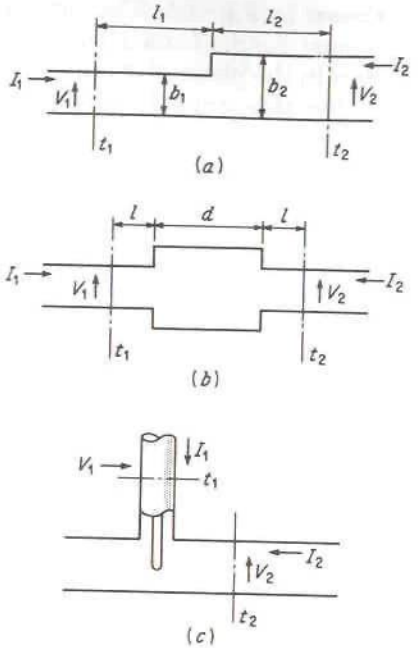


FIGURE 4.5
Examples of two-port junctions.

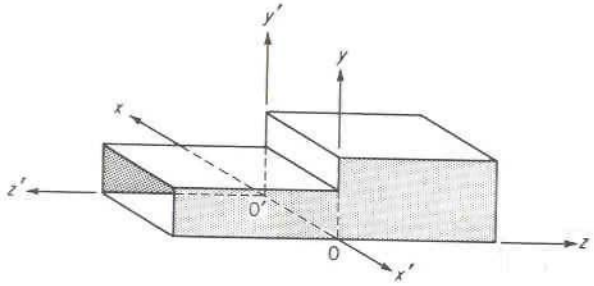


FIGURE 4.6
Coordinates used to describe the junction in Fig. 4.5*a*.

so that

$$V_1^+(I_1^+)^* = Y_w K_1^2 |C_1^+|^2 = Y_w |C_1^+|^2 \int_0^a \int_0^{b_1} \sin^2 \frac{\pi x}{a}\, dx\, dy$$

or $K_1 = \sqrt{ab_1/2}$ and

$$V_2^+(I_2^+)^* = Y_w K_2^2 |C_2^+|^2 = Y_w |C_2^+|^2 \int_0^a \int_0^{b_2} \sin^2 \frac{\pi x'}{a}\, dx'\, dy'$$

or $K_2 = \sqrt{ab_2/2}$.

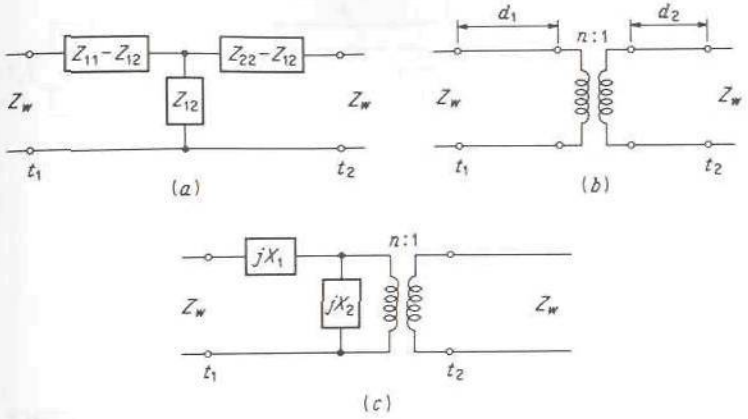


FIGURE 4.7
Equivalent circuits for a lossless two-port circuit.

If we use the above equivalent voltages and currents, we have

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{12} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{4.39}$$

as a suitable description of the E-plane step. An equivalent circuit consisting of a T network joining two transmission lines as in Fig. 4.7*a* provides a convenient equivalent circuit for the junction. Other equivalent circuits are also possible. In particular, if the junction is lossless, any circuit consisting of three independent parameters may be used. Figure 4.7*b* illustrates a circuit consisting of two lengths of transmission lines of length d_1 and d_2 connected by an ideal transformer of turns ratio n:1. Figure 4.7*c* is a variation of this circuit, where the transmission lines are replaced by reactive elements jX_1 and jX_2. The parameters of any one of these circuits may be expressed in terms of the parameters of any of the others. The required derivations may be carried out in the usual manner.

The foregoing discussion applies equally well to the other two-port junctions of Fig. 4.5 provided suitably defined voltages and currents are introduced. Likewise, the equivalent circuits of Fig. 4.7 may be used to describe the behavior of these other junctions. Although general forms for the equivalent networks can be readily specified, the values of the parameters are not so easily found. In some cases the network parameters can be evaluated analytically, whereas in many other cases they must be determined by experimental measurements.†

†For typical analytical solutions and the methods employed, see R. E. Collin, "Field Theory of Guided Waves," 2nd ed., IEEE Press, Piscataway, N.J., 1991.

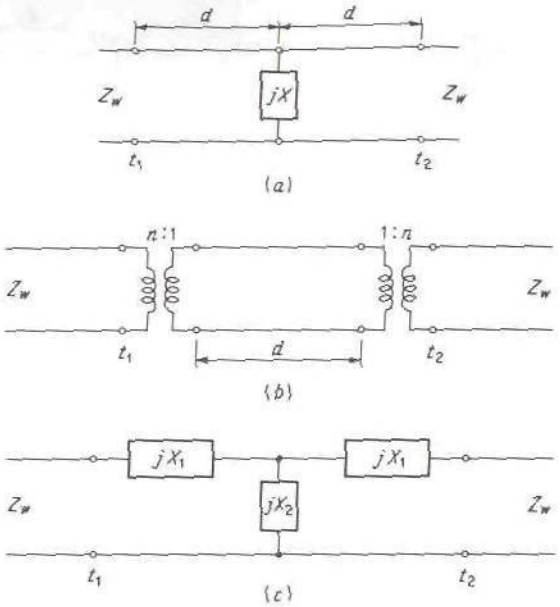


FIGURE 4.8
Equivalent circuits for a symmetrical two-port junction.

For the junction of Fig. 4.5*b*, perfect symmetry exists about the midplane, and hence the equivalent impedance matrix has $Z_{11} = Z_{22}$. For a lossless junction of this type, the equivalent circuit can be any circuit containing two independent parameters in a symmetrical connection. Some typical circuits that may be used are illustrated in Fig. 4.8.

The foregoing discussion could be rephrased so as to apply to the admittance-matrix representation as well; i.e.,

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{12} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{4.40}$$

The basic equivalent circuit described by (4.40) is the Π network illustrated in Fig. 4.9.

Example 4.1. To illustrate the use of equivalent circuits (assuming that their parameters are known at each frequency of interest), consider the coaxial-line–waveguide junction of Fig. 4.5*c*. Let a generator of internal impedance Z_g be connected to the coaxial line a distance l from the terminal plane t_1. Let the output guide be connected to a load that is matched to the guide, i.e., that presents an impedance Z_w at the terminal plane t_2. The overall circuit is that illustrated in Fig. 4.10*a*. We wish to evaluate the power transmitted to the

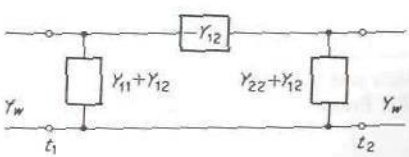


FIGURE 4.9
Equivalent circuit for admittance matrix of a two-port junction.

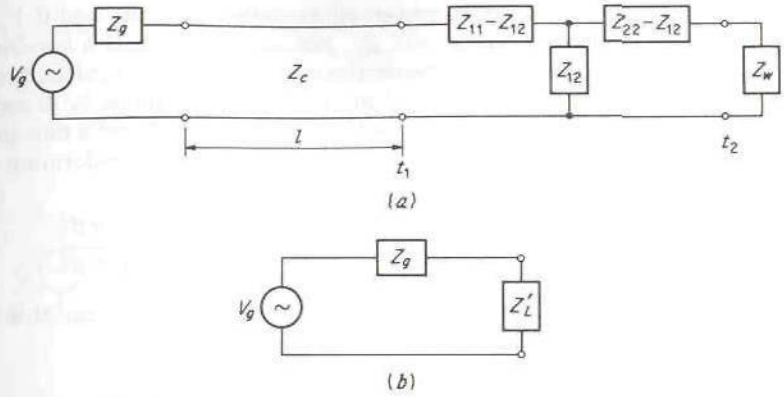


FIGURE 4.10
Equivalent circuit for generator connected to a coaxial-line-fed waveguide.

load and the standing-wave ratio on the input line. The equivalent transformed load impedance at the plane t_1 is found by conventional circuit analysis to be

$$Z_L = Z_{11} - \frac{Z_{12}^2}{Z_{22} + Z_w} \tag{4.41}$$

This impedance is transformed by the length l of coaxial line into an impedance

$$Z_L' = Z_c \frac{Z_L + jZ_c \tan \beta l}{Z_c + jZ_L \tan \beta l} \tag{4.42}$$

at the generator terminals as in Fig. 4.10*b*. This reduced circuit is easily solved. The current supplied by the generator is

$$I_g = \frac{V_g}{Z_g + Z_L'} \tag{4.43}$$

and the power delivered to Z_L' is

$$P = \tfrac{1}{2}|I_g|^2 \operatorname{Re} Z_L' \tag{4.44}$$

If the coaxial line and junction have negligible loss, this is also the power that is delivered to the load.

To compute the standing-wave ratio, note that the effective impedance terminating the coaxial line at t_1 is Z_L. Thus a reflection coefficient Γ given by

$$\Gamma = \frac{Z_L - Z_c}{Z_L + Z_c} \tag{4.45a}$$

is produced. This reflection results in a standing-wave ratio S given by

$$S = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{4.45b}$$

The reflection coefficient and standing-wave ratio depend only on the effective terminating impedance Z_L.

Maximum power will be delivered to the load if Z_g is made equal to the complex conjugate of Z_L', just as in the case of a low-frequency circuit. Other low-frequency network theorems may also be applied. For example, Thévenin's theorem may be applied at the terminal plane t_2 to reduce the circuit to an equivalent generator with a new voltage V_g' and a new internal impedance Z_g'. The new impedance is readily found by transforming Z_g to an equivalent impedance

$$Z_{ge} = Z_c \frac{Z_g + jZ_c \tan \beta l}{Z_c + jZ_g \tan \beta l} \tag{4.46}$$

at the plane t_1. When viewed through the junction, this impedance appears as an impedance

$$Z_g' = Z_{22} - \frac{Z_{12}^2}{Z_{11} + Z_{ge}} \tag{4.47}$$

at the terminal plane t_2.

The evaluation of the open-circuit voltage at t_2 is not quite as straightforward. However, by using Thévenin's theorem twice in succession, the desired result can be deduced with a minimum of labor. We first construct a Thévenin equivalent circuit at plane t_1. The equivalent internal generator impedance to use here is Z_{ge} given above. Let the voltage waves produced by the generator when the coaxial line is open-circuited at t_1 be

$$V^+ e^{-j\beta z} + V^- e^{j\beta z}$$

where z is measured from the generator. At $z = l$ the coaxial line is open-circuited; so the total current

$$I = Y_c(V^+ e^{-j\beta z} - V^- e^{j\beta z})$$

must vanish at $z = l$. Thus $V^- = V^+ e^{-2j\beta l}$. Hence, at the generator end where $z = 0$,

$$V(0) = V^+(1 + e^{-2j\beta l})$$
$$I(0) = Y_c V^+(1 - e^{-2j\beta l})$$

But $I(0) = I_g$ and $V_g = I_g Z_g + V(0)$; so

$$V(0) = V^+(1 + e^{-2j\beta l}) = V_g - I_g Z_g = V_g - I(0)Z_g$$
$$= V_g - Y_c V^+ Z_g(1 - e^{-2j\beta l})$$

When we solve for V^+ we obtain

$$V^+ = \frac{V_g}{(1 + Y_c Z_g) + (1 - Y_c Z_g)e^{-2j\beta l}} \tag{4.48}$$

The open-circuit voltage at t_1 is now readily found to be

$$V_{oc} = V^+ e^{-j\beta l} + V^- e^{j\beta l} = 2V^+ e^{-j\beta l}$$
$$= \frac{2V_g e^{-j\beta l}}{(1 + Y_c Z_g) + (1 - Y_c Z_g)e^{-2j\beta l}} \tag{4.49}$$

The Thévenin equivalent circuit at t_1 is that illustrated in Fig. 4.11.

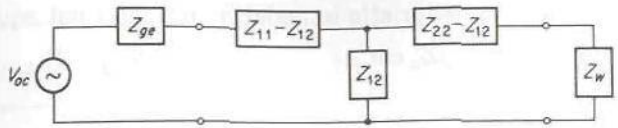


FIGURE 4.11
Thévenin equivalent circuit at terminal plane t_1.

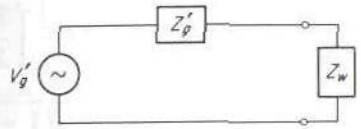


FIGURE 4.12
Thévenin equivalent circuit at terminal plane t_2.

Application of Thévenin's theorem once more in the usual manner leads readily to the circuit of Fig. 4.12, where Z_g' is given by (4.47) and V_g' is given by

$$V_g' = \frac{Z_{12}V_{oc}}{Z_{11} + Z_{ge}} \tag{4.50}$$

The reader may readily verify that the power delivered to Z_w as computed from the circuit of Fig. 4.12 is the same as that given by (4.44) (assuming no circuit losses), i.e.,

$$P = \frac{1}{2}\left|\frac{V_g'}{Z_g' + Z_w}\right|^2 Z_w \tag{4.51}$$

and is equal to (4.44). When there are circuit losses, not all the power delivered to the equivalent load impedance Z_L' of (4.44) is absorbed in the load Z_w. However, (4.51) does give the correct power delivered to the load even if other circuit losses are present. An alternative way of solving the problem is to replace the coaxial line by an equivalent T network also (Prob. 4.9), in which case the circuit is reduced to a conventional lumped-parameter network.

Some Equivalent Two-Port Circuits†

Figure 4.13 illustrates a number of useful equivalent circuits and some of their duals for representing lossless two-port junctions. The impedance parameters Z_{11}, Z_{22}, and Z_{12} are given below in terms of the network parameters, and vice versa. The same equations apply for the admittance parameters for the dual network (replace Z_{11} by Y_{11}, Z_0 by Y_0, etc.). Note that in the dual networks, the turns ratio of the ideal transformers are reversed. In addition, note that Z_0 in these circuits is an independent

†The material in this section has been reproduced in modified form from C. G. Montgomery, R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits," pp. 105–108, McGraw-Hill Book Company, New York, 1948.

parameter or characteristic impedance and does not equal $(\mu_0/\epsilon_0)^{1/2}$.

(*a*)
$$Z_{11} = -jZ_0 \cot \beta l \qquad Z = Z_{22} - Z_{11}$$
$$Z_{22} = Z - jZ_0 \cot \beta l \qquad \cos \beta l = \frac{Z_{11}}{Z_{12}}$$
$$Z_{12} = \pm jZ_0 \csc \beta l \qquad Z_0 = jZ_{12}\left[1 - \left(\frac{Z_{11}}{Z_{12}}\right)^2\right]^{1/2}$$

(*b*)
$$Z_{11} = Z_1 - j \cot \beta l \qquad Z_1 = Z_{11} + \sqrt{1 + Z_{12}^2}$$
$$Z_{22} = Z_2 - j \cot \beta l \qquad Z_2 = Z_{22} + \sqrt{1 + Z_{12}^2}$$
$$Z_{12} = \pm j \csc \beta l \qquad \sin \beta l = \pm \frac{j}{Z_{12}}$$

If $\beta l = \pi/2$, choose
$$Z_{11} = Z_1 \qquad Z_1 = Z_{11}$$
$$Z_{22} = Z_2 \qquad Z_2 = Z_{22}$$
$$Z_{12} = jZ_0 \qquad Z_0 = -jZ_{12} \neq 1$$

(*c*)
$$Z_{11} = -jZ_0 \cot \beta l \qquad \cos \beta l = \frac{\sqrt{Z_{11}Z_{22}}}{Z_{12}}$$
$$Z_{22} = -j\frac{Z_0}{n^2} \cot \beta l \qquad Z_0 = -jZ_{11}\sqrt{\frac{Z_{12}^2}{Z_{11}Z_{22}} - 1}$$
$$Z_{12} = j\frac{Z_0}{n} \csc \beta l \qquad n = \sqrt{\frac{Z_{11}}{Z_{22}}}$$

(*d*)
$$Z_{11} = -j \cot \beta l \qquad \cot \beta l = jZ_{11}$$
$$Z_{22} = Z - \frac{j}{n^2} \cot \beta l \qquad n = \frac{\sqrt{Z_{11}^2 - 1}}{Z_{12}}$$
$$Z_{12} = \frac{j}{n} \csc \beta l \qquad Z = Z_{22} + \frac{Z_{11}Z_{12}^2}{1 - Z_{11}^2}$$

(*e*)
$$Z_{11} = Z_1 + n^2 Z_2 \qquad Z_1 = Z_{11} - \frac{Z_{12}^2}{Z_{22}}$$
$$Z_{22} = Z_2 \qquad Z_2 = Z_{22}$$
$$Z_{12} = \pm nZ_2 \qquad n = \frac{Z_{12}}{Z_{22}}$$

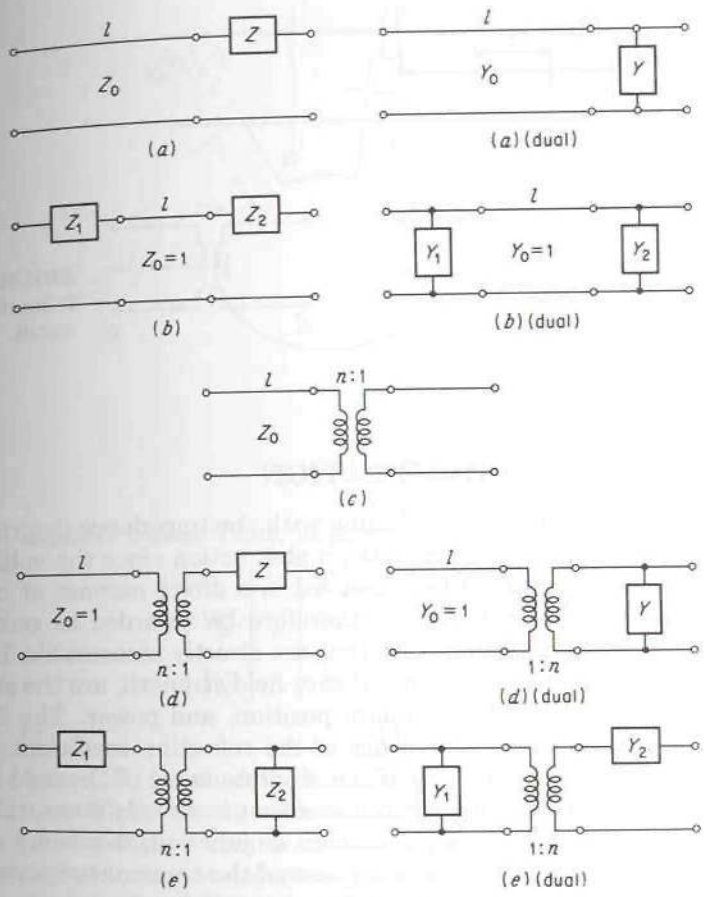


FIGURE 4.13
Some equivalent circuits for lossless two-port junctions.

An equivalent circuit incorporating a length of transmission line is particularly convenient to use since a shift in one (or both) of the terminal planes will reduce the circuit to a very simple form. For example, let the equivalent circuit of Fig. 4.13*c* be used to represent the junction in Fig. 4.14 between the terminal planes t_1 and t_2. If the terminal plane t_1 is shifted a distance l' to the left so that $\beta(l + l') = \pi$, that is, $l + l' = \lambda_g/2$, then the equivalent circuit has a section of transmission line one-half guide wavelength long. But since impedance is invariant to a transformation through a half-wavelength-long section of transmission line, this being equivalent to a 1:1 turns-ratio transformer, the section may be removed and the equivalent circuit reduces to a single ideal transformer. This new circuit represents the junction between the new terminal planes t_1' and t_2.

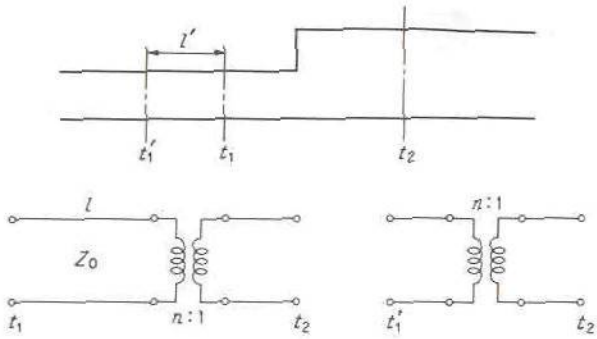


FIGURE 4.14
A junction and its equivalent circuit.

4.7 SCATTERING-MATRIX FORMULATION

The preceding section dealing with the impedance description of microwave circuits is in many respects an abstraction since the voltages, currents, and impedances cannot be measured in a direct manner at microwave frequencies. The quantities must therefore be regarded as secondary, or derived, quantities. The quantities that are directly measurable, by means of a small probe used to sample the relative field strength, are the standing-wave ratio, location of a field minimum position, and power. The first two quantities lead directly to a knowledge of the reflection coefficient. The measurement of power is needed only if the absolute value of the field in the device needs to be known. Another parameter that is directly measurable is the transmission coefficient through a circuit or junction, this being a relative measurement of the amplitude and phase of the transmitted wave as compared with those of the incident wave. In other words, the directly measurable quantities are the amplitudes and phase angles of the waves reflected, or scattered, from a junction relative to the incident-wave amplitudes and phase angles. Again, in view of the linearity of the field equations and most microwave devices, the scattered-wave amplitudes are linearly related to the incident-wave amplitudes. The matrix describing this linear relationship is called the scattering matrix.

Consider the N-port junction of Fig. 4.15. If a wave with an associated equivalent voltage V_1^+ is incident on the junction at terminal plane t_1, a reflected wave $S_{11}V_1^+ = V_1^-$ will be produced in line 1, where S_{11} is the reflection coefficient, or scattering coefficient, for line 1, with a wave incident on line 1. Waves will also be transmitted, or scattered, out of the other junctions and will have amplitudes proportional to V_1^+. These amplitudes can be expressed as $V_n^- = S_{n1}V_1^+$, $n = 2, 3, \ldots, N$, where S_{n1} is a transmission coefficient on line n from line 1. When waves are incident in all lines, the scattered wave in each line has contributions arising from all the

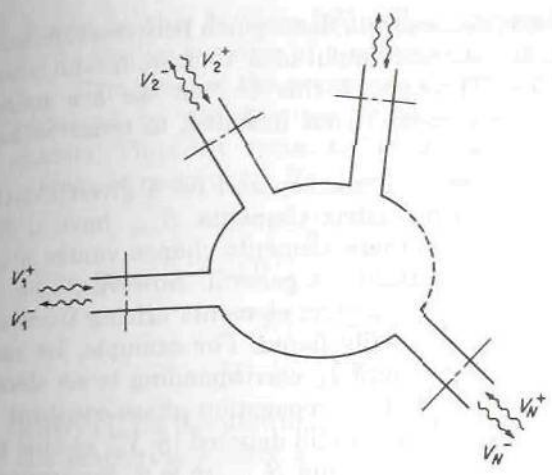


FIGURE 4.15
An N-port junction illustrating scattered waves.

incident waves. Thus, in general, we can write

$$\begin{bmatrix} V_1^- \\ V_2^- \\ \cdots \\ V_N^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} & \cdots & S_{1N} \\ S_{21} & S_{22} & S_{23} & \cdots & S_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_{N1} & S_{N2} & S_{N3} & \cdots & S_{NN} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ \cdots \\ V_N^+ \end{bmatrix} \tag{4.52a}$$

or

$$[V^-] = [S][V^+] \tag{4.52b}$$

where $[S]$ is called the scattering matrix.

When dealing with the scattering-matrix description of a junction, it is convenient to choose all the equivalent voltages (and currents, which, however, do not enter the picture explicitly) so that the power transmitted is given by $\frac{1}{2}|V_n^+|^2$ for all values of n. This corresponds to choosing the equivalent characteristic impedances equal to unity.† The main reason for doing this is to obtain a symmetrical scattering matrix for reciprocal structures. If this normalization is not used, then because of different impedance levels in different lines, the scattering matrix cannot be symmetrical. Note that, with the assumed normalization, $V = V^+ + V^-$ and $I = I^+ - I^- = V^+ - V^-$, and hence $V^+ = \frac{1}{2}(V + I)$ and $V^- = \frac{1}{2}(V - I)$. Thus the new variables V^+ and V^- are linear combinations of the variables V and I used in the impedance description. For this reason the currents do not enter into the scattering-matrix formulation. If desired, they may be calculated from the relation $I = V^+ - V^-$.

†Any value different from unity would also be suitable, the only requirement being that all lines have the same characteristic impedance, so that power will always be equal to a constant times $|V_n^+|^2$.

When it is necessary to distinguish between normalized and unnormalized voltage and current amplitudes, then we use an overbar on the normalized variables. Throughout this section we are using only normalized variables so the overbar is not included, in order to keep the notation as simple as possible.

At any particular frequency and for a given location of the terminal planes, the scattering-matrix elements S_{nm} have definite values. If the frequency is changed, these elements change values also, in a manner not readily deduced analytically in general. However, at a fixed frequency the change in the scattering-matrix elements arising from a shift in the terminal-plane location is readily found. For example, let terminal plane t_n be shifted outward an amount l_n corresponding to an electrical phase shift of $\theta_n = \beta_n l_n$, where β_n is the propagation phase constant for the nth line. If the incident-wave voltage is still denoted by V_n^+ at this new terminal plane, all the transmission coefficients S_{mn}, $m \neq n$, for transmission into line m from line n must be multiplied by $e^{-j\theta_n}$ to account for the additional path length over which the waves must travel. The reflected wave in line n has traveled a distance $2l_n$ more relative to the incident wave at the new terminal plane. Thus the new value of S_{nn} is $e^{-2j\theta_n}S_{nn}$. Likewise, waves traveling from line m to line n must travel a distance l_n farther, and thus S_{nm} is changed to $e^{-j\theta_n}S_{nm}$. These results are readily expressed in the general case by the following transformation of the $[S]$ matrix into the new $[S']$ matrix:

$$[S'] = \begin{bmatrix} e^{-j\theta_1} & & & \\ & e^{-j\theta_2} & & \\ & & \ddots & \\ & & & e^{-j\theta_N} \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & S_{22} & \cdots & S_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ S_{N1} & S_{N2} & \cdots & S_{NN} \end{bmatrix} \times \begin{bmatrix} e^{-j\theta_1} & & & \\ & e^{-j\theta_2} & & \\ & & \ddots & \\ & & & e^{-j\theta_N} \end{bmatrix} \tag{4.53}$$

where $\theta_n = \beta_n l_n$ is the outward electrical phase shift of the nth terminal plane.

Symmetry of Scattering Matrix

For a reciprocal junction the scattering matrix is symmetrical, that is, $S_{nm} = S_{mn}$, provided the equivalent voltages have been chosen so that power is given by $\frac{1}{2}|V_n^+|^2$ for all modes. The latter condition is equivalent to choosing the characteristic impedance of all equivalent transmission lines used to represent the waveguides equal to unity. If the voltages are not

chosen in this fashion, $[S]$ will, in general, not be symmetrical. Problem 4.19 gives an example of a nonsymmetrical scattering matrix.

The proof of the symmetry property of the scattering matrix is readily obtained by utilizing the known symmetry property of the impedance matrix. Thus the symmetry of the scattering matrix is basically a consequence of reciprocity. For the normalization used in the present section,

$$V_n = V_n^+ + V_n^- \qquad I_n = I_n^+ - I_n^- = V_n^+ - V_n^-$$

Thus, since $[V] = [V^+] + [V^-] = [Z][I] = [Z][V^+] - [Z][V^-]$, we have

$$([Z] + [U])[V^-] = ([Z] - [U])[V^+]$$

or

$$[V^-] = ([Z] + [U])^{-1}([Z] - [U])[V^+] \tag{4.54}$$

where $[U]$ is the unit matrix. Comparing this result with (4.52*b*) shows that the scattering matrix is related to the impedance matrix in the following manner:

$$[S] = ([Z] + [U])^{-1}([Z] - [U]) \tag{4.55a}$$

Alternatively, we have

$$[V^+] = \tfrac{1}{2}([V] + [I]) = \tfrac{1}{2}([Z] + [U])[I]$$

and

$$[V^-] = \tfrac{1}{2}([V] - [I]) = \tfrac{1}{2}([Z] - [U])[I]$$

and this gives

$$[V^-] = ([Z] - [U])([Z] + [U])^{-1}[V^+]$$

or

$$[S] = ([Z] - [U])([Z] + [U])^{-1} \tag{4.55b}$$

The transpose of (4.55*a*) is

$$[S]_t = ([Z] - [U])_t([Z] + [U])_t^{-1}$$

But since the matrices in parentheses are symmetrical, they are equal to their transpose; e.g.,

$$([Z] - [U])_t = [Z] - [U]$$

Hence

$$[S]_t = ([Z] - [U])([Z] + [U])^{-1}$$

and using (4.55*b*) now gives

$$[S]_t = [S] \tag{4.56}$$

a result that can hold only if $[S]$ is a symmetrical matrix.

Scattering Matrix for a Lossless Junction

For a lossless junction the total power leaving the N ports must equal the total incident power. The mathematical statement of this power-conserva-

tion condition is

$$\sum_{n=1}^{N} |V_n^-|^2 = \sum_{n=1}^{N} |V_n^+|^2 \tag{4.57}$$

This condition will impose a number of restrictions on the scattering-matrix parameters such as to reduce the total number of independent parameters to $\frac{1}{2}N(N+1)$, the same number of independent parameters as in the impedance matrix for a lossless junction. Replacing V_n^- by

$$V_n^- = \sum_{i=1}^{N} S_{ni} V_i^+$$

the power-conservation condition may be expressed as

$$\sum_{n=1}^{N} \left| \sum_{i=1}^{N} S_{ni} V_i^+ \right|^2 = \sum_{n=1}^{N} |V_n^+|^2 \tag{4.58}$$

The V_n^+ are all independent incident voltages; so if we choose all $V_n^+ = 0$ except V_i^+, we obtain

$$\sum_{n=1}^{N} |S_{ni} V_i^+|^2 = |V_i^+|^2 \tag{4.59}$$

or

$$\sum_{n=1}^{N} |S_{ni}|^2 = \sum_{n=1}^{N} S_{ni} S_{ni}^* = 1 \tag{4.60}$$

The index i is arbitrary; so (4.60) must hold for all values of i. Equation (4.60) states that *for a lossless junction the product of any column* of the scattering matrix with the conjugate of this same column equals unity.

In addition to the above constraint on the S_{nm}, a number of additional constraints may be derived. If we choose all $V_n^+ = 0$ except V_s^+ and V_r^+, (4.58) gives

$$\sum_{n=1}^{N} |S_{ns} V_s^+ + S_{nr} V_r^+|^2 = \sum_{n=1}^{N} (S_{ns} V_s^+ + S_{nr} V_r^+)(S_{ns} V_s^+ + S_{nr} V_r^+)^*$$

$$= |V_s^+|^2 + |V_r^+|^2$$

Expanding the left-hand side gives

$$\sum_{n=1}^{N} |S_{ns} V_s^+|^2 + \sum_{n=1}^{N} |S_{nr} V_r^+|^2 + \sum_{n=1}^{N} S_{ns} S_{nr}^* V_s^+ (V_r^+)^*$$

$$+ \sum_{n=1}^{N} S_{nr} S_{ns}^* V_r^+ (V_s^+)^* = |V_s^+|^2 + |V_r^+|^2$$

Using (4.59) results in a number of terms canceling, and we are left with

$$\sum_{n=1}^{N} [S_{ns} S_{nr}^* V_s^+ (V_r^+)^* + S_{ns}^* S_{nr} V_r^+ (V_s^+)^*] = 0$$

In view of the independent nature of V_s^+ and V_r^+, choose, first of all, $V_s^+ = V_r^+$. We then obtain

$$|V_s^+|^2 \sum_{n=1}^{N} (S_{ns} S_{nr}^* + S_{ns}^* S_{nr}) = 0 \tag{4.61a}$$

If, instead, we choose $V_s^+ = jV_r^+$, with V_r^+ real, we obtain

$$j|V_r^+|^2 \sum_{n=1}^{N} (S_{ns} S_{nr}^* - S_{ns}^* S_{nr}) = 0 \tag{4.61b}$$

Since neither V_s^+ nor V_r^+ is zero, both (4.61a) and (4.61b) can hold only if

$$\sum_{n=1}^{N} S_{ns} S_{nr}^* = 0 \qquad s \neq r \tag{4.62}$$

This equation states that the product of any column of the scattering matrix with the complex conjugate of any other different column is zero.

The conditions (4.60) and (4.62) are sufficient to restrict the number of independent parameters in the scattering matrix to $\frac{1}{2}N(N+1)$. A matrix with elements that satisfy these conditions is called a unitary matrix. To illuminate this unitary property further, it will be instructive to rederive the above results by means of matrix algebra. The power-conservation condition (4.57) can be expressed as

$$\begin{aligned} [V^-]_t[V^-]^* &= [V^+]_t[V^+]^* \\ &= ([S][V^+])_t([S][V^+])^* \\ &= [V^+]_t[S]_t[S]^*[V^+]^* \end{aligned}$$

Upon factoring this equation, we obtain

$$[V^+]_t([U] - [S]_t[S]^*)[V^+]^* = 0$$

This equation can hold only if

$$[S]_t[S]^* = [U] \tag{4.63a}$$

or

$$[S]^* = [S]_t^{-1} \tag{4.63b}$$

since $[V^+]$ is not zero. The result (4.63b) is the definition of a unitary matrix. The conditions (4.60) and (4.62) are obtained by carrying out the matrix multiplication called for in (4.63a).

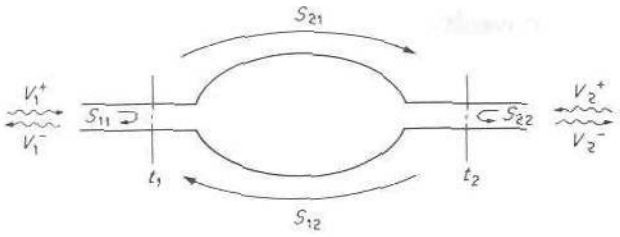


FIGURE 4.16
A two-port junction.

4.8 SCATTERING MATRIX FOR A TWO-PORT JUNCTION

Since many common microwave circuits are two-port junctions, the scattering-matrix description of these is examined in greater detail. With reference to Fig. 4.16, let the scattering-matrix parameters of the junction be S_{11}, S_{21}, S_{12}, and S_{22}. The incident- and scattered-wave amplitudes are related by

$$[V^-] = [S][V^+] \tag{4.64a}$$

or

$$V_1^- = S_{11}V_1^+ + S_{12}V_2^+ \tag{4.64b}$$

$$V_2^- = S_{21}V_1^+ + S_{22}V_2^+ \tag{4.64c}$$

If the output guide is terminated in a matched load, $V_2^+ = 0$. From (4.64*b*) it is seen that S_{11} is the reflection coefficient in the input guide 1, with guide 2 terminated in a matched load. Also, S_{21} is the transmission coefficient into guide 2 from guide 1. Similar remarks, of course, apply to the parameters S_{22} and S_{12}.

If guide 2 is terminated in a normalized impedance $\bar{Z}_2$ at the terminal plane t_2, then V_2^- may be regarded as the incident wave on $\bar{Z}_2$, and V_2^+ is the wave reflected from $\bar{Z}_2$. The ratio must be equal to the reflection coefficient of the load; hence

$$\frac{V_2^+}{V_2^-} = \frac{\bar{Z}_2 - 1}{\bar{Z}_2 + 1} = \Gamma_L \tag{4.65}$$

Substituting into (4.64), we obtain

$$V_1^- - S_{11}V_1^+ = S_{12}V_2^+ = S_{12}\Gamma_L V_2^-$$
$$-S_{21}V_1^+ = S_{22}\Gamma_L V_2^- - V_2^-$$

Solving for V_1^-/V_1^+ gives

$$\frac{V_1^-}{V_1^+} = S_{11} - \frac{S_{12}S_{21}\Gamma_L}{S_{22}\Gamma_L - 1} \tag{4.66}$$

which shows how the input reflection coefficient in guide 1 is modified when the output guide is not terminated in a matched load.

For a reciprocal junction, $S_{12} = S_{21}$, and the scattering matrix contains, at most, six independent parameters, which are the magnitudes and phase angles of S_{11}, S_{12}, and S_{22}. If the junction is lossless, the scattering matrix contains only three parameters, since the S_{nm} are related by conditions (4.60) and (4.62), which in the present case become

$$S_{11}S_{11}^* + S_{12}S_{12}^* = 1 \tag{4.67a}$$

$$S_{22}S_{22}^* + S_{12}S_{12}^* = 1 \tag{4.67b}$$

$$S_{11}S_{12}^* + S_{12}S_{22}^* = 0 \tag{4.67c}$$

The first two equations show that

$$|S_{11}| = |S_{22}| \tag{4.68}$$

and hence the reflection coefficients in the input and output guides are equal in magnitude for a lossless junction. In addition, (4.67a) shows that

$$|S_{12}| = \sqrt{1 - |S_{11}|^2} \tag{4.69}$$

If we let $S_{11} = |S_{11}|e^{j\theta_1}$, $S_{22} = |S_{11}|e^{j\theta_2}$, and $S_{12} = (1 - |S_{11}|^2)^{1/2}e^{j\phi}$, then (4.67$c$) gives

$$|S_{11}|\left(1 - |S_{11}|^2\right)^{1/2}\left(e^{j\theta_1 - j\phi} + e^{j\phi - j\theta_2}\right) = 0$$

or, equivalently,

$$e^{j(\theta_1+\theta_2)} = -e^{2j\phi}$$

Thus

$$\theta_1 + \theta_2 = 2\phi - \pi \pm 2n\pi$$

and

$$\phi = \frac{\theta_1 + \theta_2}{2} + \frac{\pi}{2} \mp n\pi \tag{4.70}$$

The two results (4.69) and (4.70) completely specify the transmission coefficient S_{12} in terms of the reflection coefficients S_{11} and S_{22}. Since S_{11} and S_{22} are readily measured and a knowledge of these suffices for the complete description of a lossless junction, the scattering matrix is a particularly convenient way of describing a lossless microwave two-port circuit.

The direct evaluation of the scattering-matrix parameters is illustrated by considering two simple examples. In Fig. 4.17a a shunt susceptance jB is connected across a transmission line with characteristic impedance Z_c. To find S_{11}, we assume the output line to be matched, so that $V_2^+ = 0$. The reflection coefficient on the input side is

$$S_{11} = \frac{Y_c - Y_{in}}{Y_c + Y_{in}} = \frac{Y_c - Y_c - jB}{2Y_c + jB} = \frac{-jB}{2Y_c + jB}$$

From symmetry considerations it is clear that $S_{22} = S_{11}$. The third parameter S_{12} can be evaluated using (4.69) and (4.70) or by finding the transmitted voltage V_2^- with the output line matched. For a pure shunt element, we

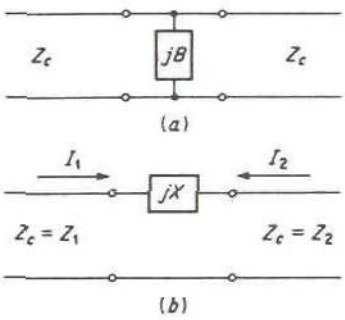


FIGURE 4.17
Shunt and series elements on a transmission line.

must have $V_1^+ + V_1^- = V_2^- = V_1^+(1 + S_{11})$. Since $V_2^- = S_{21}V_1^+$ also, we obtain

$$S_{21} = 1 + S_{11} = S_{12} = \frac{2Y_c}{2Y_c + jB}$$

For the second example we consider a series reactance jX connecting lines with characteristic impedances Z_1 and Z_2, as in Fig. 4.17*b*. In this example the characteristic impedances of the two lines are different; so we must first choose normalized voltages. Let $V_1^+, V_1^-, V_2^+, V_2^-$ be the actual transmission-line voltages for the waves that can exist on the input and output sides. Power flow for a single propagating wave is given by $\frac{1}{2}Y_1|V_1^+|^2$ and $\frac{1}{2}Y_2|V_2^+|^2$. If we choose normalized voltages $\bar{V}_1^+ = Y_1^{1/2}V_1^+$ and $\bar{V}_2^+ = Y_2^{1/2}V_2^+$, then power flow is given by $\frac{1}{2}|\bar{V}_1^+|^2$ and $\frac{1}{2}|\bar{V}_2^+|^2$ and is directly proportional to the voltage wave amplitude squared.

If the output line is matched, we have

$$\frac{V_1^-}{V_1^+} = \frac{\bar{V}_1^-}{\bar{V}_1^+} = S_{11} = \frac{Z_{\text{in}} - Z_1}{Z_{\text{in}} + Z_1} = \frac{Z_2 - Z_1 + jX}{Z_2 + Z_1 + jX}$$

With the input line matched, we find

$$\frac{V_2^-}{V_2^+} = \frac{\bar{V}_2^-}{\bar{V}_2^+} = S_{22} = \frac{Z_1 - Z_2 + jX}{Z_2 + Z_1 + jX}$$

To find S_{21}, again consider the output line matched. On the input line we have $V_1 = V_1^+ + V_1^- = V_1^+(1 + S_{11})$ and

$$I_1 = Y_1(V_1^+ - V_1^-) = Y_1V_1^+(1 - S_{11})$$

The current is continuous through a series element, and hence

$$-I_2 = I_2^- = I_1 = Y_1V_1^+(1 - S_{11})$$

But $I_2^- = Y_2 V_2^-$; so $Y_2 V_2^- = Y_1 V_1^+ (1 - S_{11})$. We now obtain

$$S_{21} = S_{12} = \frac{\overline{V}_2^-}{\overline{V}_1^+} = \left(\frac{Y_2}{Y_1}\right)^{1/2} \frac{V_2^-}{V_1^+} = \left(\frac{Z_2}{Z_1}\right)^{1/2} (1 - S_{11})$$

$$= \left(\frac{Z_2}{Z_1}\right)^{1/2} \frac{2Z_1}{Z_1 + Z_2 + jX} = \frac{2\sqrt{Z_1 Z_2}}{Z_1 + Z_2 + jX}$$

The equality of S_{12} and S_{21} occurs because of the symmetrical manner in which Z_1 and Z_2 enter into this expression. If unnormalized voltages were used, the same expressions would be obtained for S_{11} and S_{22}, but for S_{21} and S_{12} we would obtain instead

$$S_{21} = \frac{2Z_2}{Z_1 + Z_2 + jX} \qquad S_{12} = \frac{2Z_1}{Z_1 + Z_2 + jX}$$

4.9 TRANSMISSION-MATRIX REPRESENTATION

When a number of microwave circuits are connected together in cascade, it is more convenient to represent each junction or circuit by a transmission matrix that gives the output quantities in terms of the input quantities. The reason for this is that, with such a representation, the matrix which describes the complete cascade connection may be obtained simply by multiplying the matrices describing each junction together. The independent variables may be chosen as the input voltages V_n and currents I_n, the incident- and reflected-wave amplitudes V_n^+, V_n^- on the input side, or any other convenient linearly independent quantities. When voltages and currents are chosen, we shall call the corresponding matrix the voltage-current transmission matrix. If incident- and reflected-wave amplitudes are chosen, we shall refer to the matrix as a wave-amplitude transmission matrix. To simplify the discussion, we shall consider the cascade connection of two-port circuits only. However, the general formulation is readily extended to cover the cascade connection of N-port circuits.

The transmission-matrix formulation is of great value in analyzing infinitely long periodic structures such as those used in slow-wave circuits for traveling-wave tubes and linear accelerators. Since examples of these are analyzed in Chap. 8, we shall consider only the basic formulation in this section.

Voltage-Current Transmission Matrix

Figure 4.18*a* illustrates a two-port junction with input total voltage and current V_1, I_1 and output quantities V_2, I_2. Since V_2 and I_2 may be chosen as the independent variables and the junction is linear, the dependent

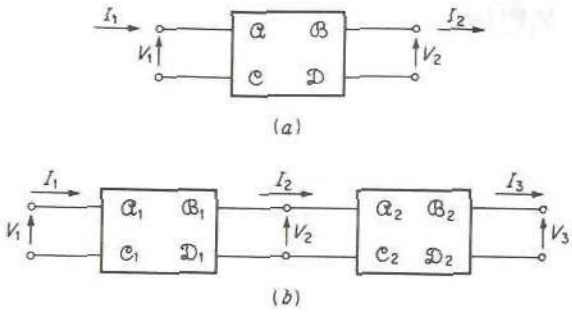


FIGURE 4.18
(*a*) A two-port junction; (*b*) cascade connection of two-port junctions.

variables V_1, I_1 are linearly related to V_2, I_2. Consequently, we may write

$$V_1 = \mathscr{A}V_2 + \mathscr{B}I_2 \tag{4.71a}$$

$$I_1 = \mathscr{C}V_2 + \mathscr{D}I_2 \tag{4.71b}$$

where $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, and $\mathscr{D}$ are suitable constants that characterize the junction. Note that we have chosen the positive direction of current to be to the right at all terminals. This is done so that the output current I_2 becomes the input current to the next junction, etc., in a cascade connection as illustrated in Fig. 4.18*b*.

In matrix form (4.71) becomes

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{C} & \mathscr{D} \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \tag{4.72}$$

The relationship of the voltage-current transmission matrix to the impedance matrix is readily found by rewriting the following equations in the form (4.71):

$$V_1 = I_1 Z_{11} - I_2 Z_{12}$$

$$V_2 = I_1 Z_{12} - I_2 Z_{22}$$

These equations may be solved to give

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} Z_{11}/Z_{12} & (Z_{11}Z_{22} - Z_{12}^2)/Z_{12} \\ 1/Z_{12} & Z_{22}/Z_{12} \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \tag{4.73}$$

The $\mathscr{A}\mathscr{B}\mathscr{C}\mathscr{D}$ parameters of the junction are readily identified in terms of the Z_{nm} from this relation. The determinant of the voltage-current transmission matrix is

$$\mathscr{A}\mathscr{D} - \mathscr{B}\mathscr{C} = 1 \tag{4.74}$$

for a reciprocal junction, as is readily verified from (4.73).

For a cascade connection as illustrated in Fig. 4.18*b*, we may write

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} \mathscr{A}_1 & \mathscr{B}_1 \\ \mathscr{C}_1 & \mathscr{D}_1 \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}$$

$$\begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} \mathscr{A}_2 & \mathscr{B}_2 \\ \mathscr{C}_2 & \mathscr{D}_2 \end{bmatrix} \begin{bmatrix} V_3 \\ I_3 \end{bmatrix}$$

and therefore

$$\begin{aligned} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix} &= \begin{bmatrix} \mathscr{A}_1 & \mathscr{B}_1 \\ \mathscr{C}_1 & \mathscr{D}_1 \end{bmatrix} \begin{bmatrix} \mathscr{A}_2 & \mathscr{B}_2 \\ \mathscr{C}_2 & \mathscr{D}_2 \end{bmatrix} \begin{bmatrix} V_3 \\ I_3 \end{bmatrix} \\ &= \begin{bmatrix} \mathscr{A}_1\mathscr{A}_2 + \mathscr{B}_1\mathscr{C}_2 & \mathscr{A}_1\mathscr{B}_2 + \mathscr{B}_1\mathscr{D}_2 \\ \mathscr{C}_1\mathscr{A}_2 + \mathscr{D}_1\mathscr{C}_2 & \mathscr{C}_1\mathscr{B}_2 + \mathscr{D}_1\mathscr{D}_2 \end{bmatrix} \begin{bmatrix} V_3 \\ I_3 \end{bmatrix} \end{aligned} \tag{4.75}$$

Thus the input quantities are readily found in terms of the output variables simply by multiplying the transmission matrices together. The ratio of the output voltage to current is determined by the load impedance.

Wave-Amplitude Transmission Matrix

The wave-amplitude transmission matrix relates the incident- and reflected-wave amplitudes on the input side of the junction to those on the output side. It bears the same relationship to the scattering matrix as the voltage-current transmission matrix does to the impedance matrix. Just as in the case of the voltage-current transmission-matrix representation, it is convenient to choose the variables in such a fashion that the output variables from one junction become the input variables for the next junction. With reference to Fig. 4.19*a*, we thus choose

$$c_1^+ = V_1^+ \tag{4.76a}$$

$$c_1^- = V_1^- \tag{4.76b}$$

$$c_2^+ = V_2^- \tag{4.76c}$$

$$c_2^- = V_2^+ \tag{4.76d}$$

$$c_3^+ = V_3^- \tag{4.76e}$$

$$c_3^- = V_3^+ \tag{4.76f}$$

The superscript + refers to the amplitude of the wave propagating to the right, and the superscript − refers to the amplitude of the wave propagating to the left. The input and output quantities are linearly related; so we may write

$$\begin{bmatrix} c_1^+ \\ c_1^- \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} c_2^+ \\ c_2^- \end{bmatrix} \tag{4.77}$$

where the A_{nm} are suitable constants that describe the junction. For the

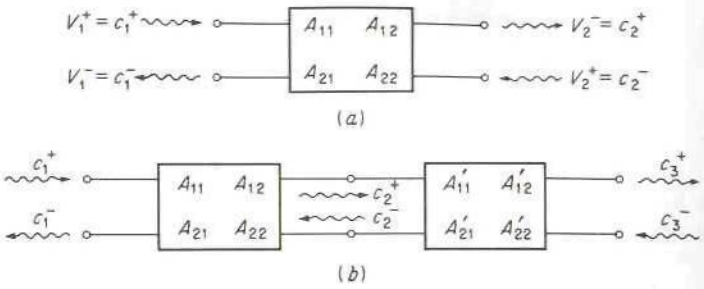


FIGURE 4.19
Wave-amplitude transmission-matrix representation of a junction.

cascade connection of Fig. 4.19*b*, we have

$$\begin{bmatrix} c_1^+ \\ c_1^- \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} A'_{11} & A'_{12} \\ A'_{22} & A'_{22} \end{bmatrix} \begin{bmatrix} c_3^+ \\ c_3^- \end{bmatrix} \tag{4.78}$$

In terms of the scattering matrix for the single junction, we have

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} c_1^- \\ c_2^+ \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix}$$

$$= \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} \begin{bmatrix} c_1^+ \\ c_2^- \end{bmatrix} \tag{4.79}$$

These equations may be solved for c_1^+ and c_1^- to give

$$\begin{bmatrix} c_1^+ \\ c_1^- \end{bmatrix} = \begin{bmatrix} 1/S_{12} & -S_{22}/S_{12} \\ S_{11}/S_{12} & (S_{12}^2 - S_{11}S_{22})/S_{12} \end{bmatrix} \begin{bmatrix} c_2^+ \\ c_2^- \end{bmatrix} \tag{4.80}$$

from which the A_{nm} are readily identified in terms of the S_{nm}. Note that the determinant of the $[A]$ matrix equals unity, i.e.,

$$A_{11}A_{22} - A_{12}A_{21} = 1 \tag{4.81}$$

as is readily verified from (4.80). However, if the wave amplitudes had not been normalized, so that power was given by $\frac{1}{2}|c_n^+|^2$, etc., the determinant of the $[A]$ matrix would be different from unity in general (this follows from the nonsymmetry of the scattering matrix in this case). Some of the problems given at the end of this chapter illustrate these points.

⋆4.10 SIGNAL FLOW GRAPHS

A signal flow graph is a graphical representation of the relationships between a set of independent input variables that are linearly related to a set of dependent output variables. For example, the impedance-matrix

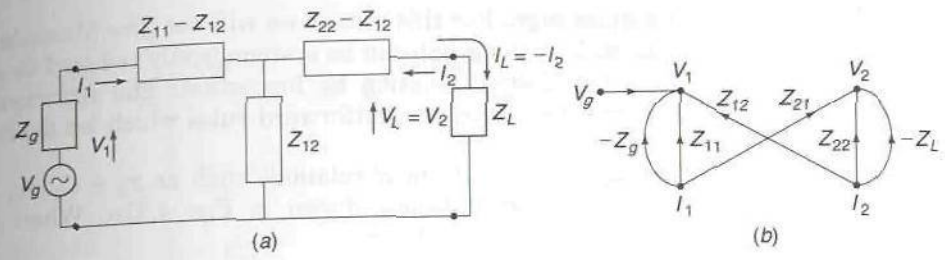


FIGURE 4.20
(*a*) A generator connected to a load through an impedance network; (*b*) signal-flow-graph representation of the linear system in (*a*).

description of a two-port network

$$V_1 = Z_{11}I_1 + Z_{12}I_2$$
$$V_2 = Z_{21}I_1 + Z_{22}I_2$$

may be represented by a signal flow graph by drawing four nodes to represent the variables V_1, V_2, I_1, and I_2 and connecting lines having transmission factors that show how the inputs I_1 and I_2 feed signals to the output nodes labeled V_1 and V_2. The graph is shown in Fig. 4.20. The node labeled V_1 has an input $Z_{11}I_1$ from I_1 and an input $Z_{12}I_2$ from I_2. These inputs are represented by the directed line segments having transmission factors Z_{11} and Z_{12} as shown in Fig. 4.20. Similar connections go from node V_2 to nodes I_1 and I_2. If the output is connected to a load Z_L, we have a further condition, namely, $V_2 = -I_2Z_L$, which can be represented as a directed line segment from the I_2 node to the V_2 node and having a transmission factor $-Z_L$. If we connect a generator with voltage V_g and impedance Z_g to the input, then one additional equation $V_1 = V_g - I_1Z_g$ is imposed on the system. To represent this equation on the graph, a node labeled V_g is added along with a directed line segment from this node to V_1 and with unity transmission factor. Also an additional line segment from node I_1 to node V_1 with transmission factor $-Z_g$ must be added. From this example it should be clear how a signal flow graph is constructed to represent a linear system.

Once the signal flow graph has been constructed, the solution giving any one variable such as V_1, V_2, I_1, or I_2 in terms of the source variable V_g can be determined from the topology of the graph and the application of a set of formal rules known as Mason's rules.† For complicated graphs the use of Mason's rules can be quite intricate and the chances of making a mistake

†S. J. Mason, Feedback Theory—Some Properties of Signal Flow Graphs, *Proc. IRE.*, vol. 41, pp. 1144–1156, 1953; also J. Mason, Feedback Theory—Further Properties of Signal-Flow Graphs, *Proc. IRE.*, vol. 44, pp. 920–926, 1956.

are quite high. For this reason we will not give Mason's rules but will show instead how the graph can be systematically reduced to a simple form which gives the desired solution by inspection. The reduction is carried out by applying five basic straightforward rules which we describe below.

RULE 1. A pair of linear relations such as $x_2 = C_{21}x_1$, $x_3 = C_{32}x_2$ has the graphical representation shown in Fig. 4.21*a*. When there are no other inputs to node 2, we have

$$x_3 = C_{21}C_{32}x_1$$

Thus rule 1 states that two series paths are equivalent to a single path from node 1 to node 3 with a transmission factor equal to the product of that from node 1 to node 2 with that from node 2 to node 3 as shown in Fig. 4.21*a*.

RULE 2. If there are two or more parallel paths connecting node 2 to node 1, we have

$$\begin{aligned} x_2 &= A_{21}x_1 + B_{21}x_1 + C_{21}x_1 + \cdots \\ &= (A_{21} + B_{21} + C_{21} + \cdots)x_1 \end{aligned}$$

Rule 2 states the obvious result that several parallel connecting paths are equivalent to a single path with a transmission factor equal to the sum of those of the individual paths. This rule is illustrated in Fig. 4.21*b*.

RULE 3. Consider the linear relationships

$$\begin{aligned} x_2 &= C_{21}x_1 + C_{23}x_3 \\ x_3 &= C_{32}x_2 \end{aligned}$$

which have the graph shown in Fig. 4.21*c*. With no other inputs to nodes 2 and 3, we can eliminate x_3 to obtain

$$\begin{aligned} x_2 &= C_{21}x_1 + C_{23}C_{32}x_2 \\ x_2(1 - C_{23}C_{32}) &= C_{21}x_1 \\ x_2 &= \frac{C_{21}x_1}{1 - C_{23}C_{32}} \\ x_3 = C_{32}x_2 &= \frac{C_{32}C_{21}x_1}{1 - C_{23}C_{32}} \end{aligned}$$

Thus rule 3 states that a feedback loop may be eliminated by dividing the input transmission factor by 1 minus the transmission factor around the loop which is the product $C_{23}C_{32}$ by rule 1. If there are several inputs and outputs from node 2, each input transmission factor is divided by $1 - C_{23}C_{32}$, but all output transmission factors remain unchanged as shown in Fig. 4.21*c*. If node 3 is isolated, then the feedback loop becomes a self-loop with loop gain $C_{23}C_{32} = C_{22}$.

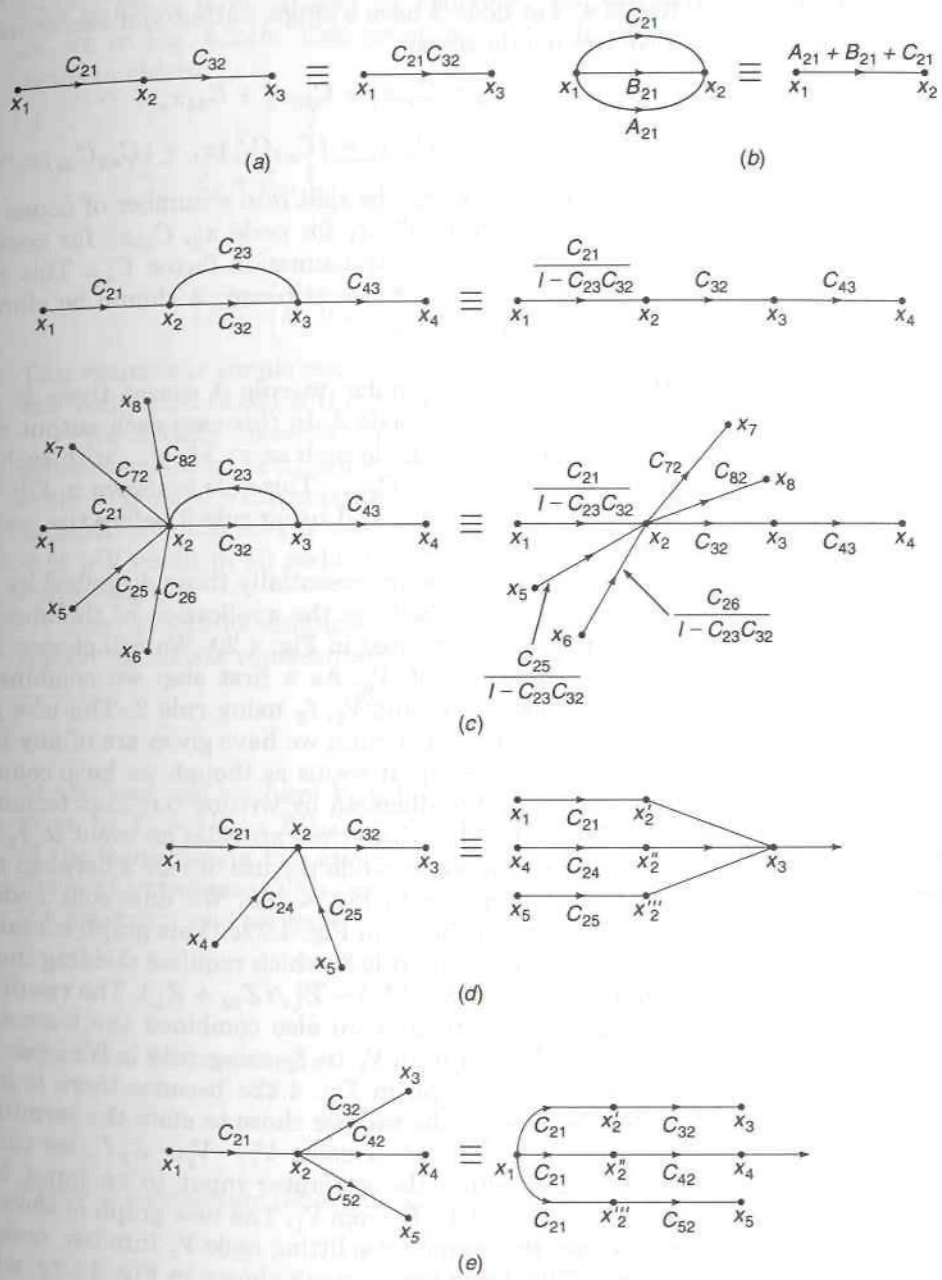


FIGURE 4.21
Illustration of rules for reduction of a signal flow graph to simpler form. (*a*) Rule 1; (*b*) rule 2; (*c*) rule 3; (*d*) rule 4; (*e*) rule 5.

RULE 4. Let node 2 have a single output and two or more inputs. Thus we have the relationships

$$x_2 = C_{21}x_1 + C_{23}x_3 + C_{24}x_4 + \cdots$$

$$x_n = C_{n2}x_2 = (C_{n2}C_{21})x_1 + (C_{n2}C_{23})x_3 + \cdots$$

This shows that x_2 can be split into a number of nodes x_2', x_2'', x_2''', etc., each with separate inputs $C_{21}x_1$ for node x_2', $C_{23}x_3$ for node x_2'', etc., and these all feed node n with a transmission factor C_{n2}. This rule is illustrated in Fig. 4.21*d*. If node 2 has a self-loop, it should be eliminated using rule 3 before the node is split.

RULE 5. Rule 5 is similar to rule 4 except there is only one input but several outputs from node 2. In this case each output can be considered as coming from a single node such as x_2', x_2'', etc., with each of these split nodes having the same input $C_{21}x_1$. This rule is shown in Fig. 4.21*e*. A self-loop at node 2 should be eliminated using rule 3 before the node is split.

The above rules are essentially those described by Kuhn.†

We will now illustrate the application of the above rules to solve the circuit problem illustrated in Fig. 4.20. We will choose I_2 as the variable to be found in terms of V_g. As a first step we combine the parallel paths between nodes V_1, I_1 and V_2, I_2 using rule 2. The new graph is redrawn in Fig. 4.22*a*. None of the rules we have given are of any help in reducing this graph any further and it seems as though we have come to an impasse. We can get around the dilemma by writing our load terminal condition in the form $I_2 = -V_2/Z_L$, which then provides an input to I_2, the variable we are interested in. Hence we undo the use of rule 2 between nodes V_2 and I_2 and redraw the graph as in Fig. 4.22*b*. We now split node 2 using rule 4 to obtain the graph shown in Fig. 4.22*c*. This graph contains a self-loop which we eliminate by using rule 3, which requires dividing the input transmission factors by $1/(1 + Z_{22}/Z_L) = Z_L/(Z_{22} + Z_L)$. The result is the graph shown in Fig. 4.22*e*. Note that we also combined the transmission factors from node I_1 to V_2' and from V_2' to I_2 using rule 1. We again run into a problem in reducing the graph in Fig. 4.22*e* because there is no input to node I_1. This is because of the way we chose to state the terminal conditions at the generator end. Instead of using $V_1 = V_g - Z_g I_1$, we can use $I_1 = V_g/Z_g - V_1/Z_g$ which shifts the generator input to an input to node I_1 and also provides an input to I_1 from V_1. The new graph is shown in Fig. 4.22*f*. We can reduce this graph by splitting node V_1 into two nodes and applying rule 4 again. This leads to the graph shown in Fig. 4.22*g* which has a self-loop.

†N. Kuhn, Simplified Signal Flow Graph Analysis, *Microwave J.*, vol. 6, no. 11, p. 59, November, 1963.

The self-loop is eliminated by applying rule 3 and results in the graph shown in Fig. 4.22*h*. This graph has a feedback loop so we apply rule 3 again to obtain

$$I_2 = \frac{V_g}{Z_g + Z_{11}} \left(\frac{1}{1 - \dfrac{Z_{12}Z_{21}}{(Z_{11} + Z_g)(Z_{22} + Z_L)}} \right) \left(\frac{-Z_{21}}{Z_L + Z_{22}} \right)$$

$$= \frac{-Z_{21}V_g}{(Z_{11} + Z_g)(Z_{22} + Z_L) - Z_{12}Z_{21}} \tag{4.82}$$

This example is simple enough so that we can easily solve for I_2 analytically and verify that (4.82) is the correct answer.

We initially chose to express the terminal conditions at the load and generator ends so as to cause difficulty in reducing the signal flow graph in order to illustrate the importance of choosing the correct way to express the terminal conditions. The terminal conditions should be expressed in a form that will result in all nodes in the signal flow graph having both input and output signals.

As a second example we will solve the same problem using a scattering-matrix representation of the impedance network; thus we use

$$V_1^- = S_{11}V_1^+ + S_{12}V_2^+$$
$$V_2^- = S_{21}V_1^+ + S_{22}V_2^+$$

At the load end we have $V_2^+ = \Gamma_L V_2^-$ where the load reflection coefficient is given by $\Gamma_L = (Z_L - Z_c)/(Z_L + Z_c)$ and Z_c is the characteristic impedance of the transmission line assumed to be connected between the network and the load impedance. This line is considered to have a negligible length. At the generator end we have the terminal conditions

$$V_1^+ + V_1^- = V_g - \frac{(V_1^+ - V_1^-)}{Z_c} Z_g$$

which can be expressed as

$$V_1^+ = \frac{V_g Z_c}{Z_c + Z_g} + \Gamma_g V_1^-$$

where the generator reflection coefficient is given by

$$\Gamma_g = \frac{Z_g - Z_c}{Z_g + Z_c}$$

We again assume that a transmission line of negligible length and with characteristic impedance Z_c is connected between the generator and the network. The signal flow graph for the above system of linear equations is

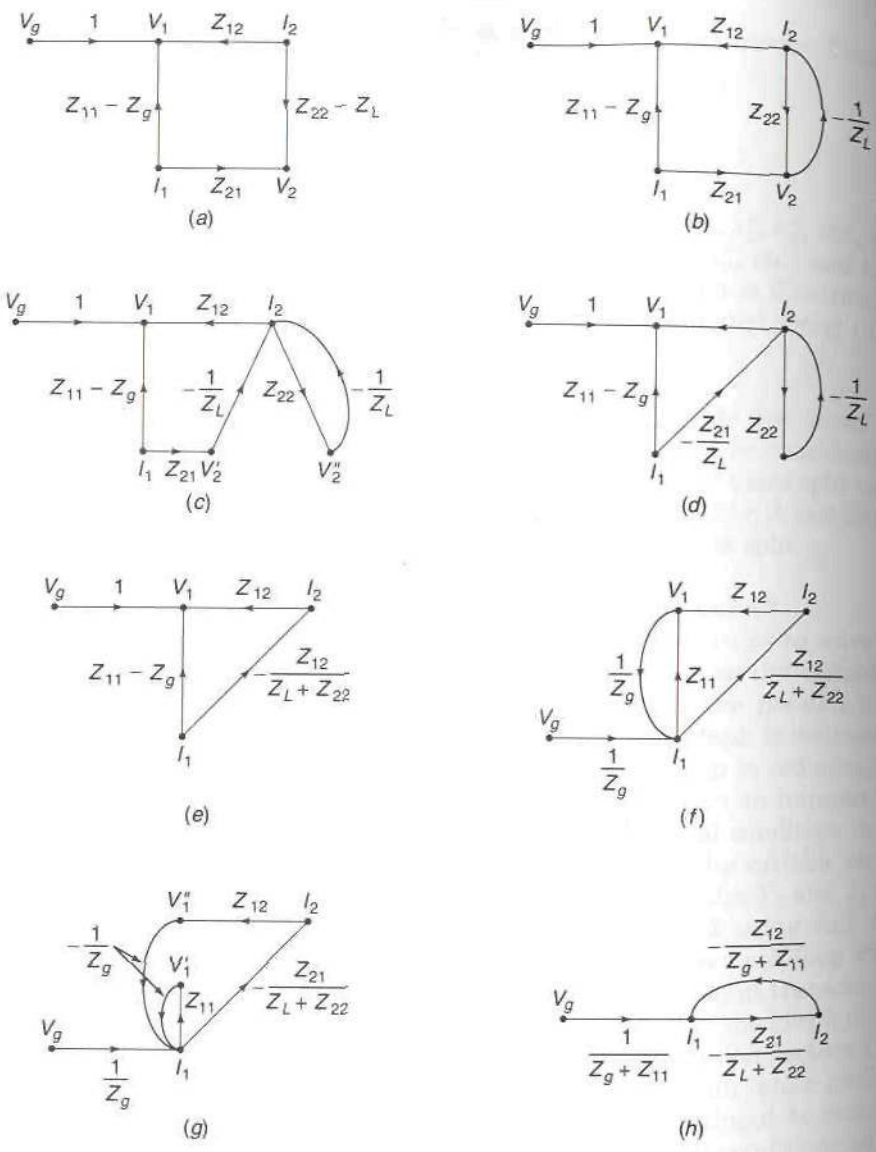


FIGURE 4.22
Illustration of steps used to reduce the signal flow graph in Fig. 4.20*b* to a simple form.

shown in Fig. 4.23*a*. Note that we have chosen to express the terminal conditions in such a form as to provide an input to the node V_2^+ and an output from node V_2^- so as to avoid the problems encountered earlier in reducing the graph to a simple form. The procedure that we will follow parallels that used in the first example. We will choose V_2^- as the variable to be solved since this is the amplitude of the wave that is incident on the load.

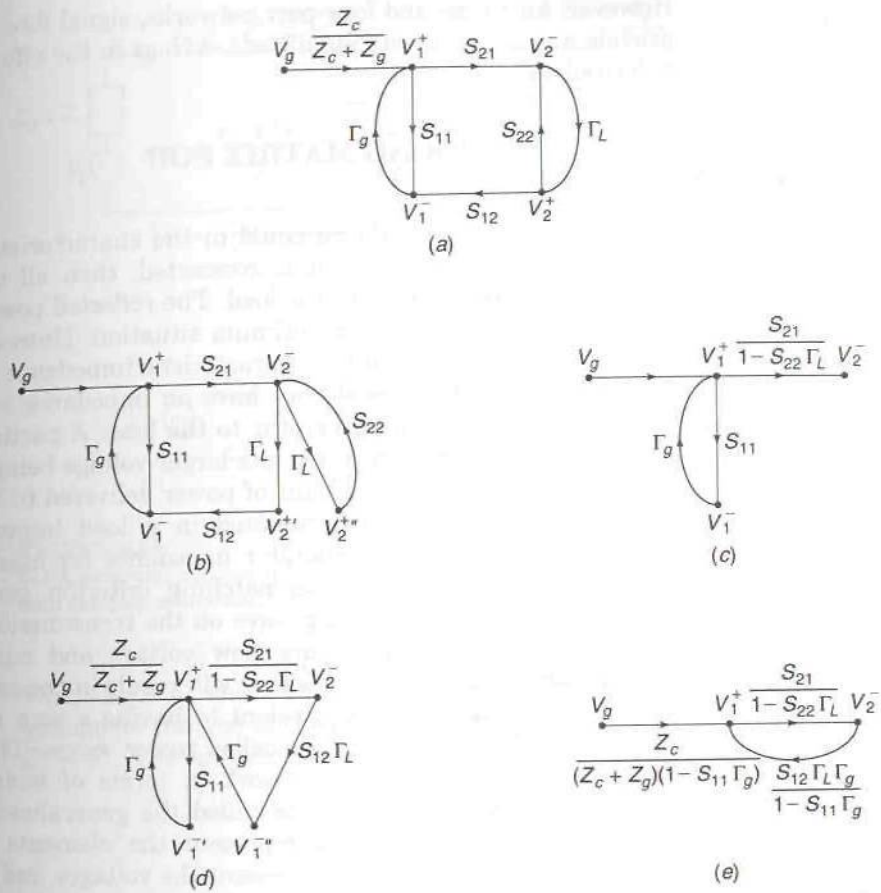


FIGURE 4.23
(*a*) Signal flow graph for the circuit in Fig. 4.20*a* but using a scattering-matrix formulation; (*b*)–(*e*) steps followed in reducing the signal flow graph to a simple form.

The first step is to split node V_2^+ using rule 5 to obtain the graph shown in Fig. 4.23*b*. Rule 3 is now applied to eliminate the self-loop and produces the graph in Fig. 4.23*c*. A similar treatment of node V_1^- but using rule 4 followed by application of rules 2 and 3 results in the graphs shown in Figs. 4.23*d* and *e*. The final graph gives the desired solution upon applying rule 3 once more. Thus we find that

$$V_2^- = \frac{V_g Z_c S_{21}}{(Z_c + Z_g)\left[(1 - S_{11}\Gamma_g)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_L\Gamma_g\right]} \tag{4.83}$$

The load voltage is given by $V_L = V_2^-(1 + \Gamma_L)$.

For simple two-port networks, the use of signal flow graphs does not offer a great advantage over the algebraic method of obtaining a solution.

However, for three- and four-port networks, signal flow graph techniques do provide a useful tool with significant savings in the effort required to obtain a desired solution.

*4.11 GENERALIZED SCATTERING MATRIX FOR POWER WAVES

If we have a load impedance equal to the characteristic impedance of the transmission line to which it is connected, then all of the power in the incident wave is delivered to the load. The reflected power will be zero. This would seem to represent an optimum situation. However, if the generator impedance does not equal the characteristic impedance of the interconnecting transmission line, we do not have an impedance match for maximum power transfer from the generator to the load. A partial standing wave on the transmission line can result in a larger voltage being applied to the load impedance and a greater amount of power delivered to the load. In general, the generator should be terminated in a load impedance equal to the complex conjugate of the generator impedance for maximum power transfer. This conjugate impedance-matching criterion generally means that there will be a partial standing wave on the transmission line.

It is possible to introduce new voltage and current variables and corresponding wave amplitudes that will result in the conjugate impedance-matching condition being equivalent to having a zero reflection coefficient for the load. The new waves are called power waves. The scattering matrix that describes a microwave network in terms of incident, reflected, and transmitted power waves will be called the generalized scattering matrix. We will use script letters to represent the elements of the generalized scattering matrix and also to represent the voltages and currents associated with the power waves. The theory of the generalized scattering matrix and power waves is developed below. This theory provides a useful extension of the conventional theory for analyzing systems in which both the source and load impedances are complex. The theory can be developed by analogy with that of the conventional theory.† We present the key relations from the conventional theory first and then use these to obtain similar relations for the generalized case.

For the circuit shown in Fig. 4.24*a*, let the voltage and current wave amplitudes on the transmission line be V^+, V^- and I^+, I^- at the load. The current amplitudes are related to the voltage amplitudes as follows:

$$I^+ = \frac{V^+}{Z_c} \qquad I^- = -\frac{V^-}{Z_c}$$

†K. Kurokawa, Power Waves and the Scattering Matrix, *IEEE Trans.*, vol. MTT-13, pp. 194–202, March, 1965.

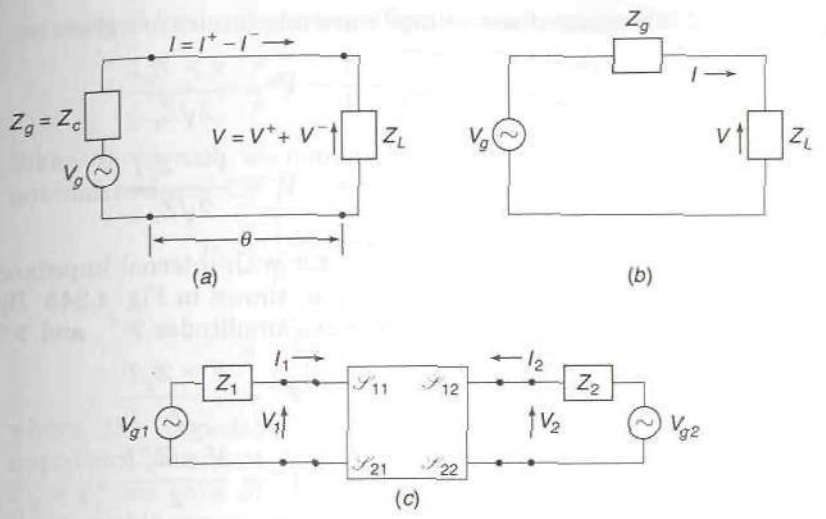


FIGURE 4.24
(*a*) A generator connected to a load impedance by means of a transmission line; (*b*) a circuit with complex source and load impedances; (*c*) a two-port network.

When $Z_L = Z_c$ there is no reflected wave since $\Gamma_L = 0$. Thus $V^- = 0$ and the voltage on the line at the generator end will be $V^+ e^{j\theta}$ and equals $V_g/2$ when $Z_g = Z_c$. The incident power is given by

$$P_{\text{inc}} = P_{\text{ava}} = \frac{|V^+|^2}{2Z_c} = \frac{|V_g|^2}{8Z_c} \tag{4.84}$$

and is equal to the available power P_{ava} from the source. If $Z_L \neq Z_c$ then the power delivered to the load is given by

$$P_L = \left(1 - |\Gamma_L|^2\right) P_{\text{inc}} = \left(1 - |\Gamma_L|^2\right) P_{\text{ava}} \tag{4.85}$$

The total voltage and current at the load are

$$V = V^+ + V^- \qquad I = I^+ - I^- = \frac{V^+ - V^-}{Z_c}$$

We can express the wave amplitudes V^+ and V^- in terms of V and I by the relations

$$V^+ = \frac{V + Z_c I}{2} \tag{4.86a}$$

$$V^- = \frac{V - Z_c I}{2} \tag{4.86b}$$

The normalized voltage wave amplitudes are given by

$$\bar{V}^+ = \frac{V + Z_c I}{2\sqrt{Z_c}} \tag{4.87a}$$

$$\bar{V}^- = \frac{V - Z_c I}{2\sqrt{Z_c}} \tag{4.87b}$$

Consider now a generator with internal impedance Z_g connected to a complex load impedance Z_L as shown in Fig. 4.24*b*. By analogy with (4.86) we will choose the power wave amplitudes $\mathscr{V}^+$ and $\mathscr{V}^-$ to be

$$\mathscr{V}^+ = \frac{V + Z_g I}{2} \tag{4.88a}$$

$$\mathscr{V}^- = \frac{V - Z_g^* I}{2} \tag{4.88b}$$

where V and I are the actual voltage and current in the circuit. When the load impedance Z_L is equal to the complex conjugate Z_g^* of the generator impedance, then

$$V = \frac{V_g Z_L}{Z_g + Z_L} = \frac{V_g Z_g^*}{Z_g + Z_g^*} \qquad I = \frac{V_g}{Z_g + Z_L} = \frac{V_g}{Z_g + Z_g^*}$$

Upon using these expressions in (4.88), we obtain

$$\mathscr{V}^+ = \frac{V_g}{2} \qquad \mathscr{V}^- = \frac{V_g}{2}\frac{Z_g^* - Z_g^*}{Z_g + Z_g^*} = 0$$

Thus the definitions chosen for $\mathscr{V}^+$ and $\mathscr{V}^-$ have the desired property that when the load impedance is conjugate matched to the generator impedance, for maximum power transfer, the amplitude of the reflected power wave is zero. The power delivered to the load is given by

$$P_L = \frac{1}{2}\left|\frac{V_g}{Z_g + Z_g^*}\right|^2 \operatorname{Re} Z_L = \frac{|V_g|^2}{8R_g^2} R_g = \frac{|V_g|^2}{8R_g}$$

where $Z_g = Z_L^* = R_g + jX_g$. This also represents the available power from the source and can be expressed in the following way:

$$P_{\text{ava}} = \frac{|\mathscr{V}^+|^2}{2R_g} = \frac{|\bar{\mathscr{V}}^+|^2}{2}$$

where the normalized voltage wave amplitude has been chosen to be $\mathscr{V}^+/\sqrt{R_g}$.

When the load impedance is not conjugate matched to the generator impedance, we will have a reflected power wave with amplitude $\mathscr{V}^-$. The

generalized load reflection coefficient, which we designate as Γ_L', is given by

$$\Gamma_L' = \frac{\mathscr{V}^-}{\mathscr{V}^+} = \frac{V - IZ_g^*}{V + IZ_g} = \frac{V/I - Z_g^*}{V/I + Z_g} = \frac{Z_L - Z_g^*}{Z_L + Z_g} \tag{4.89}$$

When Z_g is real, we obtain the usual load reflection coefficient. For the nonmatched case the power delivered to the load is given by

$$P_L = \frac{1}{2}\,\mathrm{Re}\,VI^* = \frac{1}{2}\,\mathrm{Re}\,\frac{V_g}{Z_L + Z_g}\,\frac{V_g^*}{Z_L^* + Z_g^*}Z_L$$

$$= \frac{|V_g|^2}{8R_g}\,\frac{4R_L R_g}{|Z_L + Z_g|^2} = MP_{\text{ava}} \tag{4.90}$$

where the impedance mismatch factor $M = 4R_L R_g/|Z_L + Z_g|^2$. The impedance mismatch factor M is always less than or equal to one. For $Z_L = Z_g^*$ we have $M = 1$. The mismatch factor determines the fraction of the available power that is delivered to the load. By using the definition for Γ_L' and some algebra, we can readily show that

$$M = 1 - |\Gamma_L'|^2 \tag{4.91}$$

Consequently, we can write

$$P_L = \left(1 - |\Gamma_L'|^2\right)P_{\text{ava}} \tag{4.92a}$$

which is analogous to (4.85). Upon using $V = Z_L V_g/(Z_g + Z_L)$ and $I = V_g/(Z_g + Z_L)$ in (4.88*a*), we find that $\mathscr{V}^+ = V_g/2$; so we can also write

$$P_L = \frac{1}{2}\left|\frac{\mathscr{V}^+}{\sqrt{R_g}}\right|^2\left(1 - |\Gamma_L'|^2\right) = \left(1 - |\Gamma_L'|^2\right)P_{\text{inc}} \tag{4.92b}$$

where P_{inc} is the power incident from the power wave launched by the generator.

The new set of variables $\mathscr{V}^+$ and $\mathscr{V}^-$, which are linearly related to V and I, form a convenient pair of new variables for analyzing a circuit having complex generator and load impedances. If we have a two-port network as shown in Fig. 4.24*c*, we can introduce incident and scattered power waves on both the input and output sides. The power wave normalized amplitudes are linearly related to the normal voltages and currents on the input and output sides as follows:

$$\mathscr{V}_1^+ = \frac{V_1 + I_1 Z_1}{2\sqrt{R_1}} \qquad \mathscr{V}_1^- = \frac{V_1 - I_1 Z_1^*}{2\sqrt{R_1}}$$

$$\mathscr{V}_2^+ = \frac{V_2 + I_2 Z_2}{2\sqrt{R_2}} \qquad \mathscr{V}_2^- = \frac{V_2 - I_2 Z_2^*}{2\sqrt{R_2}}$$

where Z_1 and Z_2, with real parts R_1 and R_2, are the complex generator

impedances on the input and output sides. The two-port network will force two of the above variables to be linearly related to the remaining two. If we choose the incident power wave amplitudes as the independent variables, then the scattered power wave amplitudes are given in terms of the former ones by the generalized scattering-matrix representation

$$\bar{\mathscr{V}}_1^- = \mathscr{S}_{11}\bar{V}_1^+ + \mathscr{S}_{12}\bar{\mathscr{V}}_2^+ \tag{4.93a}$$

$$\bar{\mathscr{V}}_2^- = \mathscr{S}_{21}\bar{V}_1^+ + \mathscr{S}_{22}\bar{\mathscr{V}}_2^+ \tag{4.93b}$$

The generalized scattering-matrix parameters $\mathscr{S}_{ij}$ cannot be measured directly. However, they can be expressed in terms of the normal two-port scattering-matrix parameters S_{ij}. In practice, we would embed the two-port network into a transmission line with characteristic impedance Z_c (usually 50 Ω) and measure the S_{ij}. We can linearly relate $V_1^+, V_1^-, V_2^+, V_2^-$ to V_1, I_1 and V_2, I_2. The power wave amplitudes are linearly related to the total voltages and currents. By setting up these linear relationships and expressing V_1^- and V_2^- in terms of V_1^+ and V_2^+, we can, after a number of matrix manipulations, show that

$$[\mathscr{S}] = [D^*]^{-1}([S] - [\Gamma^*])([U] - [\Gamma][S])^{-1}[D] \tag{4.94}$$

where $[D]$ is a diagonal matrix with elements

$$D_{ii} = |1 - \Gamma_i^*|^{-1}(1 - \Gamma_i)\sqrt{1 - |\Gamma_i|^2} \qquad i = 1, 2$$

$$\Gamma_i = \frac{Z_i - Z_c}{Z_i + Z_c}$$

and $[\Gamma]$ is also a diagonal matrix with elements $\Gamma_{ii} = \Gamma_i$. The unit matrix is $[U]$.

We will now examine the network shown in Fig. 4.25*a*. The two-port network can be described in terms of the normal scattering-matrix parameters S_{ij} or in terms of the generalized scattering-matrix parameters. In the network shown in Fig. 4.25*a*, we do not have a generator on the output side. The termination is the load impedance Z_L. Also we have labeled the source impedance on the input side as Z_s and the source voltage as V_g. These changes correspond to setting $V_{g2} = 0$, $V_{g1} = V_g$, $Z_1 = Z_s$, $Z_2 = Z_L$ in the previous analysis. We can visualize the two-port network as connected to the source and load by transmission lines of negligible length. Since $V_{g2} = 0$ the output current is $I_2 = -V_2/Z_L$. Consequently,

$$\bar{\mathscr{V}}_2^+ = \frac{V_2 + I_2 Z_L}{2\sqrt{R_L}} = 0$$

so there is no power wave reflected from the load. The power delivered to the load is

$$P_L = \tfrac{1}{2}|\bar{\mathscr{V}}_2^-|^2 = \tfrac{1}{2}|\mathscr{S}_{21}|^2|\bar{\mathscr{V}}_1^+|^2 \tag{4.95}$$

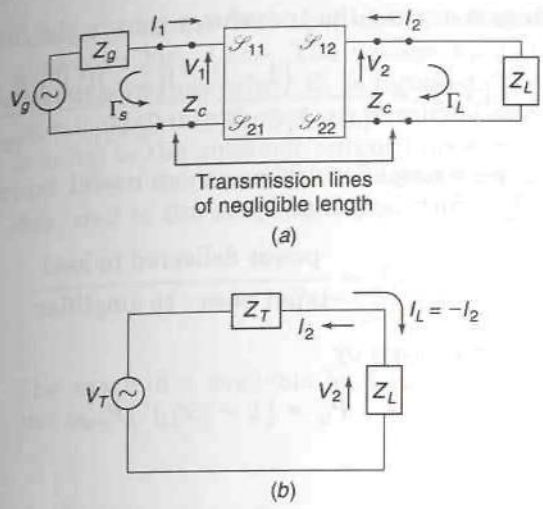


FIGURE 4.25
(*a*) a two-port network terminated in a complex load impedance; (*b*) Thévenin equivalent circuit.

The transducer gain of a two-port network, which could represent a microwave amplifier, is defined by the relation

$$G = \frac{\text{power delivered to load}}{\text{available power from source}} \tag{4.96}$$

Since the available power is simply $|\bar{\mathscr{V}}_1^+|^2/2$, we find that

$$G = |\mathscr{S}_{21}|^2 \tag{4.97}$$

From (4.94) we obtain the following expressions for the generalized scattering-matrix parameters:

$$\mathscr{S}_{11} = \frac{1}{W}\frac{1-\Gamma_s}{1-\Gamma_s^*}\left[(S_{11}-\Gamma_s^*)(1-\Gamma_L S_{22}) + S_{12}S_{21}\Gamma_L\right] \tag{4.98a}$$

$$\mathscr{S}_{12} = \frac{1}{W}\frac{(1-\Gamma_s)(1-\Gamma_L)}{|1-\Gamma_s||1-\Gamma_L|}S_{12}\left[\left(1-|\Gamma_s|^2\right)\left(1-|\Gamma_L|^2\right)\right]^{1/2} \tag{4.98b}$$

$$\mathscr{S}_{21} = \frac{S_{21}}{S_{12}}\mathscr{S}_{12} \tag{4.98c}$$

$$\mathscr{S}_{22} = \frac{1}{W}\frac{1-\Gamma_L}{1-\Gamma_L^*}\left[(S_{22}-\Gamma_L^*)(1-\Gamma_s S_{11}) + S_{12}S_{21}\Gamma_s\right] \tag{4.98d}$$

where

$$W = (1-\Gamma_s S_{11})(1-\Gamma_L S_{22}) - S_{12}S_{21}\Gamma_s\Gamma_L$$

Note that for a reciprocal two-port network with $S_{12} = S_{21}$ that $\mathscr{S}_{12} = \mathscr{S}_{21}$

also. We can now express the transducer gain in the form

$$G = |\mathscr{S}_{21}|^2 = \frac{(1 - |\Gamma_s|^2)(1 - |\Gamma_L|^2)|S_{21}|^2}{|(1 - \Gamma_s S_{11})(1 - \Gamma_L S_{22}) - S_{12}S_{21}\Gamma_s\Gamma_L|^2} \tag{4.99}$$

In microwave amplifier design a more useful expression for gain is the power gain G_p, which is defined by

$$G_p = \frac{\text{power delivered to load}}{\text{input power to amplifier}} \tag{4.100}$$

The input power is given by

$$P_{\text{in}} = (1 - |\mathscr{S}_{11}|^2)P_{\text{ava}} \tag{4.101}$$

so we obtain

$$G_p = \frac{G}{1 - |\mathscr{S}_{11}|^2} = \frac{|\mathscr{S}_{21}|^2}{1 - |\mathscr{S}_{11}|^2} \tag{4.102}$$

Another gain expression is the available power gain, which is given by

$$G_a = \frac{\text{maximum available load power}}{\text{available input power}} \tag{4.103}$$

The maximum power that can be delivered to the load is the load power when the load impedance is made equal to the complex conjugate of the amplifier output impedance Z_{out}. We will show that the actual power delivered to the load is given by

$$P_L = (1 - |\mathscr{S}_{22}|^2)P_{L,\text{ava}} \tag{4.104}$$

where $P_{L,\text{ava}}$ is the available load power. This relationship is of the same form as (4.101) which relates the input power to the power available from the source. When we use this expression in (4.103), we get

$$\begin{aligned} G_a &= \frac{P_L}{P_{\text{ava}}} = \frac{P_L}{P_{\text{in}}}\frac{P_{\text{in}}}{P_{\text{ava}}}\frac{P_{L,\text{ava}}}{P_L} \\ &= G_P\frac{1 - |\mathscr{S}_{11}|^2}{1 - |\mathscr{S}_{22}|^2} = \frac{G}{1 - |\mathscr{S}_{22}|^2} = \frac{|\mathscr{S}_{21}|^2}{1 - |\mathscr{S}_{22}|^2} \end{aligned} \tag{4.105}$$

The above expressions for the various gains associated with an amplifier are all expressed in terms of the parameters of the generalized scattering matrix and are the main reason for introducing the concept of power waves.

We now return to the derivation of (4.104). We can replace the source and two-port network by a Thévenin equivalent network consisting of a

voltage generator with open-circuit voltage V_T and a series impedance Z_T as shown in Fig. 4.25*b*. The voltage V_T is the voltage across the amplifier output terminals when Z_L is removed. The impedance Z_T is the impedance seen looking into the output terminals with the generator short-circuited. It is equal to the amplifier output impedance Z_{out}.

From the Thévenin equivalent circuit, it is readily seen that the power delivered to the load is given by

$$P_L = \frac{1}{2}|I_L|^2 R_L = \frac{|V_T|^2 R_L}{2|Z_T + Z_L|^2}$$

The maximum available load power equals P_L when we choose $Z_L = Z_T^*$ and is

$$P_{L,\text{ava}} = \frac{|V_T|^2}{8R_T}$$

Thus, in general, we can express P_L in the form

$$P_L = \frac{|V_T|^2}{8R_T}\frac{4R_L R_T}{|Z_T + Z_L|^2} = \frac{|V_T|^2}{8R_T}M$$

where $M = 4R_L R_T/|Z_T + Z_L|^2$ is the impedance mismatch factor.

If we can show that $M = 1 - |\mathscr{S}_{22}|^2$, then we can express G_a as

$$G_a = \frac{P_{L,\text{ava}}}{P_{\text{ava}}} = \frac{P_L}{MP_{\text{ava}}} = \frac{G}{M} = \frac{G}{1 - |\mathscr{S}_{22}|^2}$$

which is the expression we are trying to establish. When $V_g = 0$ we have $V_1 + I_1 Z_g = 0$; so $\mathscr{V}_1^+ = 0$. The scattering-matrix relations now give

$$\mathscr{V}_1^- = \mathscr{S}_{12}\mathscr{V}_2^+ \qquad \mathscr{V}_2^- = \mathscr{S}_{22}\mathscr{V}_2^+ = \frac{V_2 - I_2 Z_L^*}{2\sqrt{R_L}}$$

which gives us

$$\frac{\mathscr{V}_2^-}{\mathscr{V}_2^+} = \mathscr{S}_{22} = \frac{V_2 - I_2 Z_L^*}{V_2 + I_2 Z_L} = \frac{V_2/I_2 - Z_L^*}{V_2/I_2 + Z_L} = \frac{Z_T - Z_L^*}{Z_T + Z_L}$$

The Thévenin impedance Z_T is thus found to be

$$Z_T = \frac{Z_L^* + \mathscr{S}_{22} Z_L}{1 - \mathscr{S}_{22}}$$

We now expand the expression $1 - |\mathscr{S}_{22}|^2$ to obtain

$$1 - |\mathscr{S}_{22}|^2 = \frac{|Z_T + Z_L|^2 - (Z_T - Z_L^*)(Z_T^* - Z_L)}{|Z_T + Z_L|^2}$$

$$= \frac{(Z_T + Z_L)(Z_T^* + Z_L^*) - (Z_T - Z_L^*)(Z_T^* - Z_L)}{|Z_T + Z_L|^2}$$

$$= \frac{Z_T(Z_L + Z_L^*) + Z_T^*(Z_L + Z_L^*)}{|Z_T + Z_L|^2}$$

$$= \frac{4R_L R_T}{|Z_T + Z_L|^2} = M$$

which completes the proof.

As a final observation we note that when $Z_L = Z_T^*$ then $M = 1$ and $G_a = G$. Also if $Z_g = Z_{\text{in}}^*$, then $\mathscr{S}_{11} = 0$ and $G_p = G_a = G$. Thus, with conjugate impedance matching at both the input and output, all three gains become equal.

*4.12 EXCITATION OF WAVEGUIDES

The preceding sections have dealt with the circuit aspects of passive microwave junctions. In order to complete the picture, it is necessary to consider also the equivalent-circuit representations for typical sources that are used to excite waves in a waveguide or transmission line. This particular aspect of waveguide theory is somewhat specialized, and it is not possible to give a complete analysis in this text without departing too far from the main theme. However, we shall present certain aspects of the excitation problem that provide a basis for choosing appropriate equivalent circuits and generators for representing typical sources and, in addition, make it possible to solve a number of coupling problems of engineering importance. The theory is, for the most part, developed by considering specific examples.

Probe Coupling in a Rectangular Waveguide

Figure 4.26 illustrates a typical coaxial-line–waveguide probe coupling. The short-circuit position l and probe length d can be adjusted to achieve maximum power transfer from the coaxial line into the waveguide. The center conductor of the coaxial line extends into the waveguide to form an electric probe. Any waveguide mode that has a nonzero electric field along the probe will excite currents on the probe. By reciprocity, when the probe current is produced by a TEM wave incident from the coaxial line, the same waveguide modes will be excited.† It is thus easy to see that, for maximum

†This reciprocity principle is very useful for determining what modes a given probe can excite.

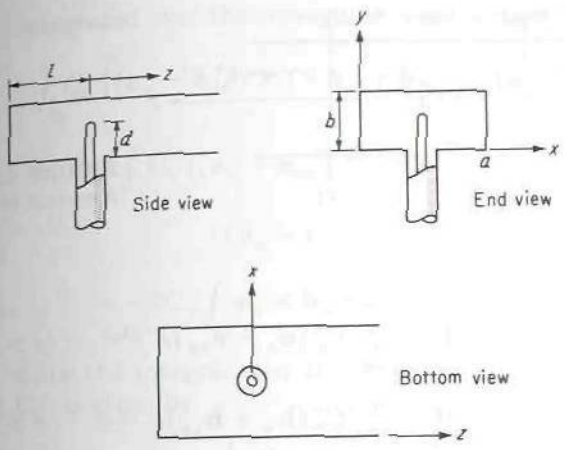


FIGURE 4.26
Coaxial-line probe coupling to a waveguide.

coupling to the dominant TE_{10} mode in a rectangular guide, the probe should extend into the guide through the center of the broad face so as to coincide with the position of maximum electric field for the TE_{10} mode. The evanescent modes that are also excited are localized fields that store reactive energy. These give the junction its reactive properties. The section of short-circuited waveguide provides an adjustable reactance that may be used to tune out the probe reactance. The probe reactance can be evaluated by determining the amplitudes of the evanescent modes that are excited and computing the net reactive energy stored in these nonpropagating modes.† Since the details are rather lengthy, we shall evaluate only the amplitude of the radiated TE_{10} mode.

The current on the probe must be zero at the end of the probe. For a thin probe a sinusoidal standing-wave current distribution is a reasonable approximation to make for the probe current. Thus let the probe current be considered as an infinitely thin filamentary current of the form

$$I = I_0 \sin k_0(d - y) \qquad 0 \le y \le d \qquad x = \frac{a}{2} \qquad z = 0 \qquad (4.106)$$

We wish to determine the amplitude of the TE_{10} mode excited by this current. A general technique for accomplishing this is a mathematical formulation of the reciprocity principle invoked earlier to determine which waveguide modes a given source will excite. The required results are derived below.

Figure 4.27 illustrates an infinitely long waveguide in which a current source $\mathbf{J}$ is located in the region between z_1 and z_2. The field radiated by this source may be expressed as an infinite sum of waveguide modes as

†See, for example, R. E. Collin, "Field Theory of Guided Waves," 2nd. ed., chap. 7, IEEE Press, Piscataway, N.J., 1991.

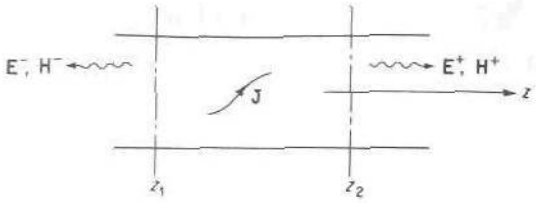


FIGURE 4.27
A current source in a waveguide.

follows:

$$\mathbf{E}^+ = \sum_n C_n^+(\mathbf{e}_n + \mathbf{e}_{zn})e^{-j\beta_n z} \qquad z > z_2 \tag{4.107a}$$

$$\mathbf{H}^+ = \sum_n C_n^+(\mathbf{h}_n + \mathbf{h}_{zn})e^{-j\beta_n z} \qquad z > z_2 \tag{4.107b}$$

$$\mathbf{E}^- = \sum_n C_n^-(\mathbf{e}_n - \mathbf{e}_{zn})e^{j\beta_n z} \qquad z < z_1 \tag{4.107c}$$

$$\mathbf{H}^- = \sum_n C_n^-(-\mathbf{h}_n + \mathbf{h}_{zn})e^{j\beta_n z} \qquad z < z_1 \tag{4.107d}$$

In (4.107) n is a general summation index and implies a summation over all possible TE and TM modes. The unknown amplitudes C_n may be determined by an application of the Lorentz reciprocity formula (2.135). For the volume V, choose that bounded by the waveguide walls and cross-sectional planes located at z_1 and z_2 in Fig. 4.27. Let the field $\mathbf{E}_1, \mathbf{H}_1$, to be used in the Lorentz reciprocity formula, be the field radiated by the current source. This field is given by (4.107). For the field $\mathbf{E}_2, \mathbf{H}_2$, choose the nth waveguide mode $\mathbf{E}_n^-, \mathbf{H}_n^-$; that is,

$$\mathbf{E}_2 = \mathbf{E}_n^- = (\mathbf{e}_n + \mathbf{e}_{zn})e^{j\beta_n z}$$

$$\mathbf{H}_2 = \mathbf{H}_n^- = (-\mathbf{h}_n + \mathbf{h}_{zn})e^{j\beta_n z}$$

Equation (2.135) gives

$$\oint_S (\mathbf{E}_1 \times \mathbf{H}_n^- - \mathbf{E}_n^- \times \mathbf{H}_1) \cdot \mathbf{n}\, dS = \int_V \mathbf{E}_n^- \cdot \mathbf{J}\, dV$$

since the field $\mathbf{E}_2, \mathbf{H}_2$ is a source-free solution ($\mathbf{J}_2 = 0$) within V. The surface integral is zero over the waveguide walls by virtue of the boundary condition $\mathbf{n} \times \mathbf{E}_1 = \mathbf{n} \times \mathbf{E}_n^- = 0$. Since the modes are orthogonal, i.e.,

$$\int_{S_0} \mathbf{E}_m^{\pm} \times \mathbf{H}_n^{\pm} \cdot \mathbf{n}\, dS = 0 \qquad n \neq m$$

all the terms except the nth in the expansion of $\mathbf{E}_1, \mathbf{H}_1$ vanish when

integrated over the waveguide cross section S_0. Thus we have

$$\int_{z_2} C_n^+ [(\mathbf{e}_n + \mathbf{e}_{zn}) \times (-\mathbf{h}_n + \mathbf{h}_{zn}) - (\mathbf{e}_n - \mathbf{e}_{zn}) \times (\mathbf{h}_n + \mathbf{h}_{zn})] \cdot \mathbf{a}_z \, dS$$

$$- \int_{z_1} C_n^- [(\mathbf{e}_n - \mathbf{e}_{zn}) \times (-\mathbf{h}_n + \mathbf{h}_{zn})$$

$$- (\mathbf{e}_n - \mathbf{e}_{zn}) \times (-\mathbf{h}_n + \mathbf{h}_{zn})] \cdot \mathbf{a}_z \, dS$$

$$= -2C_n^+ \int_{z_2} \mathbf{e}_n \times \mathbf{h}_n \cdot \mathbf{a}_z \, dS = \int_V \mathbf{E}_n^- \cdot \mathbf{J} \, dV$$

since the integral over the cross section at z_1 vanishes identically. Hence C_n^+ is given by

$$C_n^+ = -\frac{1}{P_n} \int_V \mathbf{E}_n^- \cdot \mathbf{J} \, dV = -\frac{1}{P_n} \int_V (\mathbf{e}_n - \mathbf{e}_{zn}) \cdot \mathbf{J} e^{j\beta_n z} \, dV \quad (4.108a)$$

If $\mathbf{E}_n^+, \mathbf{H}_n^+$ is chosen for the field $\mathbf{E}_2, \mathbf{H}_2$, we obtain

$$C_n^- = -\frac{1}{P_n} \int_V \mathbf{E}_n^+ \cdot \mathbf{J} \, dV = -\frac{1}{P_n} \int_V (\mathbf{e}_n + \mathbf{e}_{zn}) \cdot \mathbf{J} e^{-j\beta_n z} \, dV \quad (4.108b)$$

where

$$P_n = 2 \int_{S_0} \mathbf{e}_n \times \mathbf{h}_n \cdot \mathbf{a}_z \, dS \quad (4.108c)$$

and S_0 is a cross-sectional surface of the waveguide. The normalization constant P_n depends on the choice of expressions used for $\mathbf{e}_n$ and $\mathbf{h}_n$, the latter being arbitrary.

The above results are now applied to the probe problem introduced earlier. For the TE_{10} mode with fields given by

$$E_y = e_y e^{-j\beta z} = \sin \frac{\pi x}{a} e^{-j\beta z} \quad (4.109a)$$

$$H_x = h_x e^{-j\beta z} = -Y_w \sin \frac{\pi x}{a} e^{-j\beta z} \quad (4.109b)$$

we have

$$P_{10} = 2 \int_0^a \int_0^b Y_w \sin^2 \frac{\pi x}{a} \, dx \, dy = abY_w \quad (4.110)$$

where Y_w is the wave admittance for the TE_{10} mode and β is the propagation constant.

The probe in the short-circuited guide is equivalent to the original probe plus its image at $z = -2l$ placed in an infinite guide, as in Fig. 4.28. If we assume that the field radiated into the $z > 0$ region is

$$E_y^+ = C^+ \sin \frac{\pi x}{a} e^{-j\beta z} \quad (4.111)$$

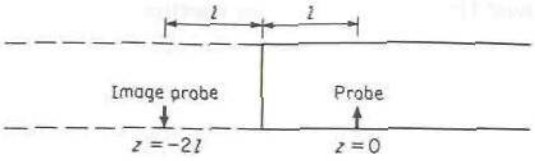


FIGURE 4.28
Probe and its image.

then application of (4.108a) gives

$$C^+ = -\frac{1}{abY_w}\left[\int_0^d I_0 \sin k_0(d-y)\,dy - \int_0^d I_0 \sin k_0(d-y)e^{-j2\beta l}\,dy\right]$$

$$= \frac{I_0 Z_w}{abk_0}(e^{-2j\beta l} - 1)(1 - \cos k_0 d) \tag{4.112}$$

since

$$E_{10}^- = \begin{cases} \sin\dfrac{\pi x}{a}e^{j\beta z} = 1 & \text{at } z = 0,\ x = \dfrac{a}{2} \\ e^{-j2\beta l} & \text{at } x = \dfrac{a}{2},\ z = -2l \end{cases}$$

We have assumed that the current on the probe can be replaced by a line current along the probe axis and with a density given by (4.106). The volume integrals in (4.108) are consequently replaced by line integrals taken along the probe axis. Note also that the direction of the current in the image probe is reversed. This is necessary so that the fields radiated by the probe and its image will give a zero tangential electric field at the short-circuit position.

The total transverse field of the TE_{10} mode radiated by the probe is thus, for $z > 0$,

$$E_y = \frac{I_0 Z_w}{abk_0}(e^{-2j\beta l} - 1)(1 - \cos k_0 d)\sin\frac{\pi x}{a}e^{-j\beta z} \tag{4.113a}$$

$$H_x = -Y_w E_y \tag{4.113b}$$

The total radiated power is given by

$$P = \frac{Y_w}{2}\int_0^a\int_0^b |E_y|^2\,dx\,dy$$

$$= \frac{I_0^2 Z_w}{4abk_0^2}|e^{-2j\beta l} - 1|^2(1 - \cos k_0 d)^2 \tag{4.114}$$

At the base of the probe antenna ($y = 0$), the total coaxial-line current is, from (4.106),

$$I \equiv I_0 \sin k_0 d$$

Let the input impedance seen from the coaxial line, referred to the base, be $Z_{in} = R_0 + jX$. The complex Poynting vector theorem then gives [Eq. (4.14)]

$$Z_{in} = R_0 + jX = \frac{P + 2j\omega(W_m - W_e)}{\frac{1}{2}II^*}$$

where P is the power radiated into the guide and $W_m - W_e$ is the reactive energy stored in the vicinity of the probe owing to the excitation of nonpropagating (evanescent) modes. Since P has been evaluated and is given by (4.114), we can compute the input resistance. We obtain

$$R_0 = \frac{2P}{I_0^2 \sin^2 k_0 d} \approx \frac{Z_w}{2abk_0^2}|1 - e^{-2j\beta l}|^2 \tan^2 \frac{k_0 d}{2} \qquad (4.115)$$

upon using the identities $1 - \cos 2\theta = 2\sin^2\theta$ and $\sin 2\theta = 2\sin\theta\cos\theta$. This input resistance is called the radiation resistance of the probe. Note that its value can be varied by varying the parameters l and d, that is, the short-circuit position and probe length. Varying these parameters thus enables an optimum amount of power transfer to be achieved by adjusting R_0 to equal the characteristic impedance of the coaxial line and introducing a suitable reactance to tune out the reactance jX. Suitable techniques for reactance cancellation are discussed in the next chapter.

Radiation from Linear Current Elements

Figures 2.29*a* and *b* illustrate linear current elements in a waveguide. For the case of the transverse current element, (4.108) shows that

$$C_n^+ = C_n^- = -\frac{1}{P_n}\int \mathbf{e} \cdot \mathbf{J}_l \, dl \qquad (4.116)$$

where J_l is the line current density in amperes. This result may be interpreted to mean that a transverse-current element is equivalent to a shunt voltage source connected across an equivalent transmission line representing the waveguide when only a single mode, say the $n = 1$ mode, propagates. The reason for this is that the transverse current radiates a field with transverse electric field components that are equal on adjacent

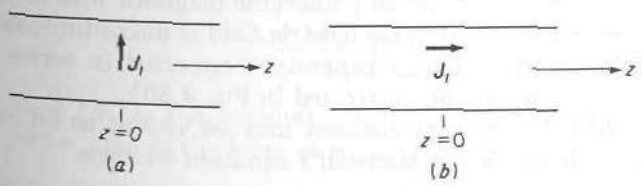


FIGURE 4.29
Linear current elements in a waveguide.

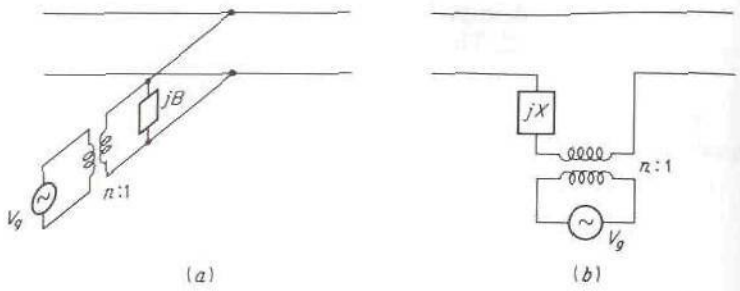


FIGURE 4.30
Equivalent circuits for current sources in a waveguide. (*a*) Transverse current source; (*b*) axial current source.

sides of the current source, and this is equivalent to continuity of the equivalent voltage across the source region. The transverse magnetic field is discontinuous across the source region, and thus the equivalent current is also discontinuous across the equivalent voltage generator. Figure 4.30 illustrates the equivalent circuit for this type of source for the dominant propagating mode. The ideal transformer provides a means of adjusting the coupling between the voltage generator and the transmission line so that the same amount of power is radiated as in the waveguide. The shunt susceptance jB represents the net reactive energy stored in the field of the evanescent modes that are excited.

For an axial current located at $z = 0$, (4.108) gives

$$C_n^+ = \frac{1}{P_n}\int \mathbf{J}_l \cdot \mathbf{e}_{zn} e^{j\beta_n z}\, dl$$

$$C_n^- = -\frac{1}{P_n}\int \mathbf{J}_l \cdot \mathbf{e}_{zn} e^{-j\beta_n z}\, dl$$

If the current is a symmetrical function of z between $-l < z < l$, then, since $\mathbf{e}_{zn}$ is not a function of z, we have

$$C_n^+ = -C_n^- = \frac{1}{P_n}\int_{-l}^{l} \mathbf{J}_l \cdot \mathbf{e}_{zn} \cos \beta_n z\, dz \tag{4.117}$$

For this case the radiated transverse magnetic field is continuous across the source but the transverse electric field is discontinuous. The source is thus equivalent to a voltage generator connected in series with an equivalent transmission line, as illustrated in Fig. 4.30*b*.

A linear current element may be viewed as an equivalent oscillating electric dipole. From Maxwell's equation we have

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E} + \mathbf{J} = j\omega\epsilon_0\mathbf{E} + j\omega\mathbf{P} + \mathbf{J}$$

and hence $\mathbf{J}$ enters into the field equations in the same manner as the

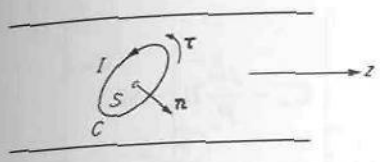


FIGURE 4.31
A current loop in a waveguide.

polarization current $j\omega\mathbf{P}$. Thus $\mathbf{J}$ may be considered equivalent to an electric dipole $\mathbf{P}$ given by

$$\mathbf{P} = \frac{\mathbf{J}}{j\omega}$$

Radiation from Current Loops

Figure 4.31 illustrates a linear current loop in a waveguide. The amplitude of the nth radiated mode is given by

$$C_n^+ = -\frac{1}{P_n}\oint_C \mathbf{E}_n^- \cdot \boldsymbol{\tau} I\, dl$$

where τI is the vector current flowing around the contour C. $\boldsymbol{\tau}$ is a unit vector along C. By Stokes' law we obtain

$$C_n^+ = -\frac{I}{P_n}\oint_C \mathbf{E}_n^- \cdot d\mathbf{l} = -\frac{I}{P_n}\int_S \nabla \times \mathbf{E}_n^- \cdot \mathbf{n}\, dS$$

But $\nabla \times \mathbf{E}_n^- = -j\omega\mathbf{B}_n^- = -j\omega\mu_0\mathbf{H}_n^-$, and hence

$$C_n^+ = \frac{j\omega I}{P_n}\int_S \mathbf{B}_n^- \cdot \mathbf{n}\, dS \qquad (4.118a)$$

Similarly,

$$C_n^- = \frac{j\omega I}{P_n}\int_S \mathbf{B}_n^+ \cdot \mathbf{n}\, dS \qquad (4.118b)$$

It is seen that the excitation amplitude of the nth mode is proportional to the total magnetic flux of this mode passing through the loop.

If the current loop is so small that the field $\mathbf{B}_n$ of the nth mode may be considered constant over the area of the loop, we obtain

$$C_n^+ = \frac{j\omega I}{P_n}\mathbf{B}_n^+ \cdot \int_S \mathbf{n}\, dS = \frac{j\omega I}{P_n}\mathbf{B}_n^+ \cdot \mathbf{S}_0$$

Now $I\mathbf{S}_0$ is the magnetic dipole moment $\mathbf{M}$ of the loop, where $\mathbf{S}_0$ is the vector area of the loop; so we obtain

$$C_n^+ = \frac{j\omega}{P_n}\mathbf{B}_n^- \cdot \mathbf{M} \qquad (4.119a)$$

and similarly

$$C_n^- = \frac{j\omega}{P_n} \mathbf{B}_n^+ \cdot \mathbf{M} \tag{4.119b}$$

Radiation from a small current loop may be considered to be magnetic dipole radiation, as these equations show. For an axial magnetic dipole (transverse current loop), the equivalent source is a shunt-connected voltage source, whereas a transverse magnetic dipole is equivalent to a series-connected voltage source.

*4.13 WAVEGUIDE COUPLING BY APERTURES†

The foregoing formulation of the radiation from currents in a waveguide in terms of radiation from equivalent electric and magnetic dipoles is directly applicable to the coupling of waveguides by small apertures, or holes, in a common wall. To a first approximation a small aperture in a conducting wall is equivalent to an electric dipole normal to the aperture and having a strength proportional to the normal component of the exciting electric field, plus a magnetic dipole in the plane of the aperture and having a strength proportional to the exciting tangential magnetic field. The constants of proportionality are parameters that depend on the aperture size and shape. These constants are called the electric and magnetic polarizabilities of the aperture and characterize the coupling or radiating properties of the aperture.‡ A qualitative argument to demonstrate the physical reasonableness of these properties of an aperture is given below.

Figure 4.32*a* illustrates the normal electric field of strength **E** at a conducting surface without an aperture. When an aperture is cut in the screen, the electric field lines fringe through the aperture in the manner indicated in Fig. 4.32*b*. But this field distribution is essentially that produced by an equivalent electric dipole as shown in Fig. 4.32*c*. Note that the dipole is oriented normal to the aperture.

In a similar manner the tangential magnetic field lines shown in Fig. 4.32*d* will fringe through the aperture as in Fig. 4.32*e*. These fringing field lines are equivalent to those produced by a magnetic dipole located in the plane of the aperture.

In Bethe's original theory the dipole moments are determined by the field in the waveguide in the absence of the aperture. Thus, for a circular aperture of radius $r_0 \ll \lambda_0$, the dipole moments are related to the incident

†The theory was originally developed by H. A. Bethe, Theory of Diffraction by Small Holes, *Phys. Rev.*, vol. 66, pp. 163–182, 1944.

‡For a derivation of these results, see Collin, *loc. cit.*

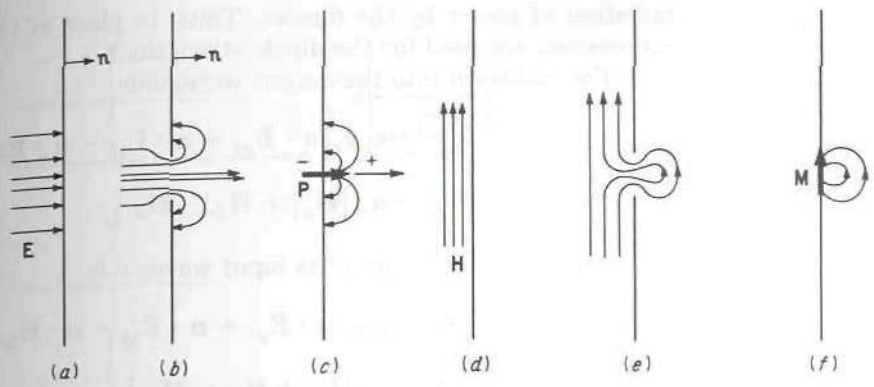


FIGURE 4.32
Aperture in a conducting wall.

fields as follows:

$$\mathbf{P} = -\epsilon_0 \alpha_e (\mathbf{n} \cdot \mathbf{E})\mathbf{n} \tag{4.120a}$$

$$\mathbf{M} = -\alpha_m \mathbf{H}_t \tag{4.120b}$$

where $\mathbf{n} \cdot \mathbf{E}$ is the normal electric field and $\mathbf{H}_t$ is the tangential magnetic field at the center of the aperture. The electric polarizability α_e is given by

$$\alpha_e = -\tfrac{2}{3} r_0^3 \tag{4.121a}$$

and the magnetic polarizability α_m is given by

$$\alpha_m = \tfrac{4}{3} r_0^3 \tag{4.121b}$$

The presence of an aperture also perturbs the field on the incident side of the screen. This perturbed field is that radiated by equivalent dipoles which are the negative of those given by (4.120) and located on the input side of the screen. It is important to note that when the aperture is replaced by equivalent electric and magnetic dipoles, the field radiated by these is computed by assuming that the aperture is now closed by a conducting wall. The equivalent dipoles correctly account for the field coupled through the aperture in the conducting screen.

Bethe's theory does not lead to an equivalent circuit for the aperture that includes a conductance to represent power coupled through the aperture. The reason for this is that the field assumed to excite the dipoles is chosen as the unperturbed incident field in the waveguide. In actual fact one should use the sum of the incident field and excited field as the polarizing field. Since the excited field is small, the correction to the dipole moments is also small. However, by including the excited dominant modes (the propagating modes) as part of the polarizing field, we will obtain the needed correction that results in a conductance element in the equivalent circuit. The dominant-mode fields react back on the dipoles to account for the

radiation of power by the dipoles. Thus, in place of (4.120), the following expressions are used for the dipole strengths:†

For radiation into the output waveguide,

$$\mathbf{P} = -\epsilon_0 \alpha_e \left[\mathbf{n} \cdot \mathbf{E}_{g1} + \mathbf{n} \cdot \mathbf{E}_{1r} - \mathbf{n} \cdot \mathbf{E}_{2r}\right]\mathbf{n} \tag{4.122a}$$

$$\mathbf{M} = -\alpha_m \left[\mathbf{H}_{g1} + \mathbf{H}_{1r} - \mathbf{H}_{2r}\right]_t \tag{4.122b}$$

For the radiation into the input waveguide,

$$\mathbf{P} = \epsilon_0 \alpha_e \left[\mathbf{n} \cdot \mathbf{E}_{g1} + \mathbf{n} \cdot \mathbf{E}_{1r} - \mathbf{n} \cdot \mathbf{E}_{2r}\right]\mathbf{n} \tag{4.122c}$$

$$\mathbf{M} = \alpha_m \left[\mathbf{H}_{g1} + \mathbf{H}_{1r} - \mathbf{H}_{2r}\right]_t \tag{4.122d}$$

where the generator fields $\mathbf{E}_{g1}, \mathbf{H}_{g1}$ are the dominant-mode fields in the input waveguide in the absence of the aperture, $\mathbf{E}_{1r}, \mathbf{H}_{1r}$ are the dominant-mode fields radiated by the dipoles in the input waveguide, and $\mathbf{E}_{2r}, \mathbf{H}_{2r}$ are the dominant-mode fields radiated by the dipoles in the output waveguide. The unit vector $\mathbf{n}$ is normal to the aperture and directed from the input waveguide to the output waveguide. The subscript t denotes the tangential component of the magnetic fields. The field resulting from the aperture is determined by closing the aperture by an electric perfectly conducting surface and calculating the fields radiated by the dipoles given above and located at the center of the circular aperture.

The theory is readily extended to include noncircular apertures. However, the procedure outlined above is restricted to circular apertures in a very thin common wall between two waveguides. There is considerable attenuation in the coupling through an aperture in a thick wall and in many practical applications this attenuation must be taken into account.‡

The examples discussed next will illustrate the application of small-aperture coupling theory to an aperture in a transverse wall and an aperture in the broad wall between two identical rectangular waveguides.

Aperture in a Transverse Wall

Figure 4.33*a* illustrates a small circular aperture in a transverse wall in a rectangular waveguide. To determine the exciting generator field, assume that the aperture is closed. A TE_{10} mode incident from $z < 0$ is reflected by the conducting wall at $z = 0$ to produce a standing-wave field in the region

†R. E. Collin, *loc. cit.*

‡See, for example, N. A. McDonald, Electric and Magnetic Coupling Through Small Apertures in Shield Walls of Any Thickness, *IEEE Trans.*, vol. MTT-20, pp. 689–695, 1972.

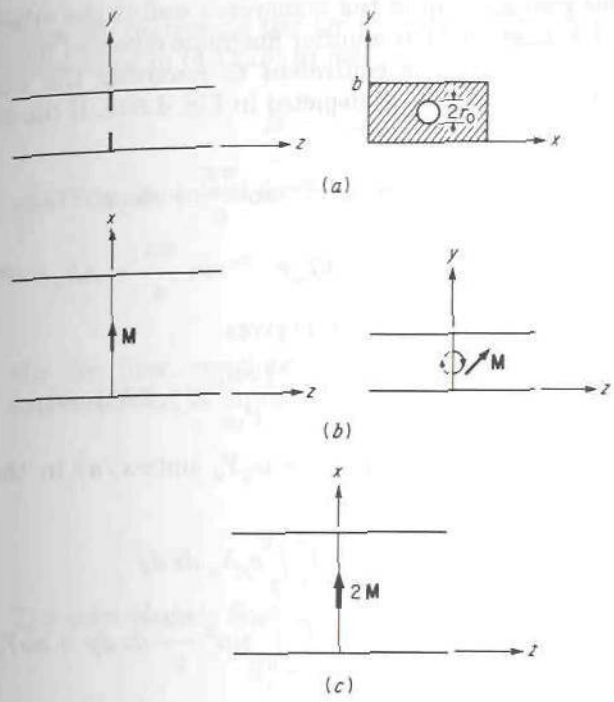


FIGURE 4.33
Aperture in a transverse waveguide wall.

$z < 0$. This field is

$$E_y = C(e^{-j\beta z} - e^{j\beta z})\sin\frac{\pi x}{a} \tag{4.123a}$$

$$H_x = -CY_w(e^{-j\beta z} + e^{j\beta z})\sin\frac{\pi x}{a} \tag{4.123b}$$

plus a z component of magnetic field which is not required to be known for the present problem.

The normal electric field at the aperture is zero; so no induced electric dipole is produced. The tangential magnetic field at the center of the aperture is, from (4.123b),

$$H_x = -2CY_w$$

and hence an induced x-directed magnetic dipole M is produced.

In order to determine the total polarizing field using (4.122b) and (4.122d), we must find the fields $\mathbf{H}_{1r}$ and $\mathbf{H}_{2r}$ radiated into guides 1 and 2 by a magnetic dipole $M\mathbf{a}_x$. The field radiated into the region $z > 0$ is that radiated by the magnetic dipole M, as illustrated in Fig. 4.33b. This dipole is equivalent to a half circular current loop in the yz plane as illustrated. To find the field radiated by this dipole in the presence of the conducting transverse wall, image theory may be used. Since the image of the half

circular current loop in the transverse wall is the other half of the current loop, the image of M is another magnetic dipole of moment M. The effect of the transverse wall is equivalent to removing the wall and doubling the strength of the dipole, as depicted in Fig. 4.33*c*. If the field radiated into the region $z > 0$ is

$$E_y^+ = Ae^{-j\beta z} \sin \frac{\pi x}{a} = Ae_y e^{-j\beta z}$$

$$H_x^+ = -AY_w e^{-j\beta z} \sin \frac{\pi x}{a} = Ah_x e^{-j\beta z}$$

application of formula (4.119*a*) gives

$$A = \frac{j\omega\mu_0}{P_{10}} H_x^-(2M)$$

since the field B_n^- is $-\mu_0 h_x = \mu_0 Y_w \sin(\pi x/a)$ in the present case. The constant P_{10} is given by

$$P_{10} = -2\int_0^a \int_0^b e_y h_x \, dx \, dy$$

$$= 2Y_w \int_0^a \int_0^b \sin^2 \frac{\pi x}{a} \, dx \, dy = abY_w$$

Hence we obtain

$$A = \frac{j\omega\mu_0}{ab} 2M = \frac{jk_0 Z_0}{ab} 2M \tag{4.124}$$

The presence of the aperture causes a field to be scattered into the region $z < 0$ also. For radiation into this region, the effective magnetic dipole moment is the negative of that used to obtain (4.124). Application of (4.119) now gives

$$E_y = A \sin \frac{\pi x}{a} e^{j\beta z} \qquad z < 0$$

for the radiated field in the input waveguide and where A is given by (4.124). As expected, the magnetic dipole $M\mathbf{a}_x$ acts as a shunt source. The x component of the radiation reaction fields $\mathbf{H}_{1r}$ and $\mathbf{H}_{2r}$, at the center of the aperture, are

$$H_{1rx} = AY_w = \frac{jk_0 Z_0}{abZ_w} 2M$$

$$H_{2rx} = -AY_w = -\frac{jk_0 Z_0}{abZ_w} 2M$$

and the generator field is

$$H_{g1x} = -2CY_w$$

Since M represents the dipole strength for radiation into guide 2, we use these fields in (4.122*b*) to obtain

$$M = -\alpha_m \left[-2CY_w + \frac{j4k_0 Z_0}{abZ_w} M \right]$$

which can be solved for M to give

$$M = \frac{2\alpha_m Y_w C}{1 + \dfrac{j4k_0 Z_0}{abZ_w}\alpha_m} \tag{4.125}$$

We can now complete the evaluation of the constant A by using this expression for M in (4.124); thus

$$A = \frac{4\dfrac{jk_0 Z_0}{abZ_w}\alpha_m}{1 + \dfrac{4jk_0 Z_0}{abZ_w}\alpha_m} C \tag{4.126}$$

The total electric field in the input waveguide is

$$E_y = \left[Ce^{-j\beta z} + (A - C)e^{j\beta z} \right] \sin \frac{\pi x}{a}$$

so the input reflection coefficient Γ is given by

$$\Gamma = \frac{A - C}{C}$$

The input normalized admittance is

$$\overline{Y}_{\text{in}} = j\overline{B} + \overline{G} = \frac{1 - \Gamma}{1 + \Gamma}$$

When we substitute for A from (4.126) into the expression for Γ, we find that

$$\overline{Y}_{\text{in}} = \frac{2 - A/C}{A/C} = 1 - j\frac{3ab}{8r_0^3\beta} \tag{4.127}$$

The equivalent circuit of the aperture, as seen from the input waveguide, is a normalized shunt conductance of unit value plus a shunt inductive susceptance. The conductance term is called the radiation conductance and accounts for the power coupled, i.e., radiated, into the output waveguide. The amplitude of the transmitted electric field is A which is given by (4.126). The transmission coefficient is A/C. From the equivalent circuit the transmission coefficient is $1 + \Gamma = 2/(1 + \overline{Y}_{\text{in}})$ which gives the same result. The aperture is equivalent to an inductive susceptance connected across the transmission line. The conductance term represents the output transmission line terminated in a matched load.

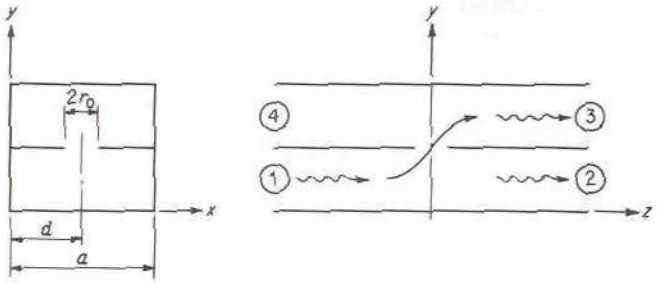


FIGURE 4.34
Aperture in a broad wall separating two waveguides.

Aperture in Broad Wall of a Waveguide

Figure 4.34 illustrates a circular aperture of radius r_0 placed in the broad wall separating two rectangular waveguides. The incident field is a TE_{10} mode in the lower guide, and is given by

$$E_y = C \sin \frac{\pi x}{a} e^{-j\beta z} \tag{4.128a}$$

$$H_x = -CY_w \sin \frac{\pi x}{a} e^{-j\beta z} \tag{4.128b}$$

$$H_z = j\frac{\pi Y_w}{\beta a} C \cos \frac{\pi x}{a} e^{-j\beta z} \tag{4.128c}$$

At the center of the aperture located at $z = 0$, $x = d$, the exciting field is

$$E_y = C \sin \frac{\pi d}{a} \tag{4.129a}$$

$$\mathbf{H} = CY_w \left(-\mathbf{a}_x \sin \frac{\pi d}{a} + j\frac{\pi}{\beta a} \mathbf{a}_z \cos \frac{\pi d}{a} \right) \tag{4.129b}$$

These expressions show that there will be a y-directed electric dipole and x- and z-directed magnetic dipoles excited in the aperture.

Let the fields $\mathbf{E}_{10}^{+}$, $\mathbf{E}_{10}^{-}$, $\mathbf{B}_{10}^{+}$, and $\mathbf{B}_{10}^{-}$ be chosen as

$$\mathbf{E}_{10}^{+} = \mathbf{a}_y \sin \frac{\pi x}{a} e^{-j\beta z}$$

$$\mathbf{E}_{10}^{-} = \mathbf{a}_y \sin \frac{\pi x}{a} e^{j\beta z}$$

$$\mu_0 \mathbf{H}_{10}^{+} = \mathbf{B}_{10}^{+} = -\mu_0 Y_w \left(\mathbf{a}_x \sin \frac{\pi x}{a} - j\frac{\pi}{\beta a} \mathbf{a}_z \cos \frac{\pi x}{a} \right) e^{-j\beta z}$$

$$\mu_0 \mathbf{H}_{10}^{-} = \mathbf{B}_{10}^{-} = \mu_0 Y_w \left(\mathbf{a}_x \sin \frac{\pi x}{a} + j\frac{\pi}{\beta a} \mathbf{a}_z \cos \frac{\pi x}{a} \right) e^{j\beta z}$$

Also let the dominant-mode field radiated by the electric dipole in the upper guide be

$$\mathbf{E} = \begin{cases} A_1\mathbf{E}_{10}^{+} & z > 0 \\ A_2\mathbf{E}_{10}^{-} & z < 0 \end{cases}$$

$$\mathbf{H} = \begin{cases} A_1\mathbf{H}_{10}^{+} & z > 0 \\ A_2\mathbf{H}_{10}^{-} & z < 0 \end{cases}$$

whereas that radiated by the magnetic dipoles is

$$\mathbf{E} = \begin{cases} A_3\mathbf{E}_{10}^{+} & z > 0 \\ A_4\mathbf{E}_{10}^{-} & z < 0 \end{cases}$$

$$\mathbf{H} = \begin{cases} A_3\mathbf{H}_{10}^{+} & z > 0 \\ A_4\mathbf{H}_{10}^{-} & z < 0 \end{cases}$$

The electric dipole $\mathbf{P}$ is equivalent to an electric current given by $j\omega\mathbf{P}$. Since the dipole is oriented in the transverse plane, (4.108) gives

$$A_1 = A_2 = -\frac{1}{P_{10}} j\omega\mathbf{P} \cdot \mathbf{a}_y \sin\frac{\pi d}{a} = -\frac{j\omega P_y}{abY_w} \sin\frac{\pi d}{a} \tag{4.130}$$

Note that no integration is necessary since $\mathbf{P}$ is an infinitesimal dipole of total strength P_y. The constant P_{10} is equal to abY_w.

The field radiated by the magnetic dipoles may be found by using (4.119). Thus

$$A_3 = \frac{j\omega\mu_0 Y_w}{abY_w}\left(\mathbf{a}_x \sin\frac{\pi d}{a} + j\frac{\pi}{\beta a}\cos\frac{\pi d}{a}\mathbf{a}_z\right) \cdot (M_x\mathbf{a}_x + M_z\mathbf{a}_z)$$

$$= \frac{j\omega\mu_0}{ab}\left(M_x \sin\frac{\pi d}{a} + j\frac{\pi M_z}{\beta a}\cos\frac{\pi d}{a}\right) \tag{4.131}$$

Similarly, it is found that

$$A_4 = \frac{j\omega\mu_0}{ab}\left(-M_x \sin\frac{\pi d}{a} + j\frac{\pi M_z}{\beta a}\cos\frac{\pi d}{a}\right) \tag{4.132}$$

With the above expressions for the amplitudes, the total field radiated into the upper waveguide is readily evaluated. It is given by

$$\mathbf{E} = \begin{cases} (A_1 + A_3)\mathbf{E}_{10}^{+} & z > 0 \quad (4.133a) \\ (A_2 + A_4)\mathbf{E}_{10}^{-} & z < 0 \quad (4.133b) \end{cases}$$

$$\mathbf{H} = \begin{cases} (A_1 + A_3)\mathbf{H}_{10}^{+} & z > 0 \quad (4.133c) \\ (A_2 + A_4)\mathbf{H}_{10}^{-} & z < 0 \quad (4.133d) \end{cases}$$

Note that the electric dipole and the magnetic dipole M_z radiate the same in

both directions but the magnetic dipole M_x does not. By correctly choosing the aperture position d, it is possible to obtain zero radiation in one direction; that is, $A_2 + A_4$ can be made to vanish. We will return to this property later.

For radiation into the lower waveguide, the sign of the dipoles are reversed. Since the mode functions are the same, this has the effect of changing the sign of the amplitude constants. By using the expressions for the mode functions given earlier and the derived expressions for the amplitudes, we find that the radiation reaction fields are given by

$$E_{1ry} - E_{2ry} = -2E_{2ry} = -2\left(\frac{A_1 + A_3}{2}\right)\mathbf{E}_{10}^{+} \cdot \mathbf{a}_y - 2\left(\frac{A_2 + A_4}{2}\right)\mathbf{E}_{10}^{-} \cdot \mathbf{a}_y$$

$$= \frac{2j\omega P_y}{abY_w}\sin^2\frac{\pi d}{a} + \frac{2\omega\mu_0\pi M_z}{\beta a^2 b}\cos\frac{\pi d}{a}\sin\frac{\pi d}{a} \qquad (4.134a)$$

$$H_{1rx} - H_{2rx} = -2H_{2rx} = -(A_1 + A_3)\mathbf{H}_{10}^{+} \cdot \mathbf{a}_x - (A_2 + A_4)\mathbf{H}_{10}^{-} \cdot \mathbf{a}_x$$

$$= \frac{2j\beta M_x}{ab}\sin^2\frac{\pi d}{a} \qquad (4.134b)$$

$$H_{1rz} - H_{2rz} = -2H_{2rz} = -(A_1 + A_3)\mathbf{H}_{10}^{+} \cdot \mathbf{a}_z - (A_2 + A_4)\mathbf{H}_{10}^{-} \cdot \mathbf{a}_z$$

$$= -\frac{2\omega\pi P_y}{\beta a^2 b}\sin\frac{\pi d}{a}\cos\frac{\pi d}{a} + \frac{2j\omega\mu_0 Y_w \pi^2 M_z}{\beta^2 a^3 b}\cos^2\frac{\pi d}{a} \qquad (4.134c)$$

Note that in evaluating the reaction fields at the center of the aperture, we take one-half of the field at $z = 0^-$ plus one-half of the field at $z = 0^+$, that is, on adjacent sides of the center of the aperture. The above equations show that there is interaction between the two dipoles P_y and M_z.

The above expressions for the reaction fields along with the generator fields given by (4.129) can now be used in (4.122a) and (4.122b) to write the following set of linear equations which will determine the dipole strengths:

$$P_y = -\alpha_e\epsilon_0\left[C\sin\frac{\pi d}{a} + \frac{2j\omega P_y}{abY_w}\sin^2\frac{\pi d}{a} + \frac{2\omega\mu_0\pi M_z}{\beta a^2 b}\sin\frac{\pi d}{a}\cos\frac{\pi d}{a}\right]$$

$$M_x = -\alpha_m\left[-CY_w\sin\frac{\pi d}{a} + \frac{2j\beta M_x}{ab}\sin^2\frac{\pi d}{a}\right]$$

$$M_z = -\alpha_m\left[\frac{j\pi Y_w}{\beta a}C\cos\frac{\pi d}{a} - \frac{2\omega\pi P_y}{\beta a^2 b}\sin\frac{\pi d}{a}\cos\frac{\pi d}{a} + \frac{2j\omega\mu_0 Y_w\pi^2 M_z}{\beta^2 a^3 b}\cos^2\frac{\pi d}{a}\right]$$

We can solve these equations for the dipole strengths which are found to be

given by

$$P_y = -\frac{\epsilon_0 \alpha_e \sin \dfrac{\pi d}{a}}{\Delta} C \tag{4.135a}$$

$$M_x = \frac{\alpha_m Y_w \sin \dfrac{\pi d}{a}}{1 + 2j \dfrac{\beta \alpha_m}{ab} \sin^2 \dfrac{\pi d}{a}} C \tag{4.135b}$$

$$M_z = -\frac{j\alpha_m \pi Y_0 \cos \dfrac{\pi d}{a}}{k_0 a \Delta} C \tag{4.135c}$$

where

$$\Delta = 1 + 2j\alpha_e \frac{k_0^2 \sin^2 \dfrac{\pi d}{a}}{\beta ab} + 2j\alpha_m \frac{\pi^2 \cos^2 \dfrac{\pi d}{a}}{\beta a^3 b}$$

We can use these expressions to find the amplitudes $A_1, \ldots, A_4$ for the field in the upper waveguide. Thus for the electric field we have

$$\mathbf{E} = \begin{cases} (A_1 + A_3)\mathbf{E}_{10}^{+} & z > 0 \\ (A_2 + A_4)\mathbf{E}_{10}^{-} & z < 0 \end{cases}$$

where

$$A_1 + A_3 = \frac{j\alpha_e k_0^2 \sin^2 \dfrac{\pi d}{a}}{\beta ab\Delta} C + \frac{j\alpha_m \left(\dfrac{\pi}{a}\right)^2 \cos^2 \dfrac{\pi d}{a}}{\beta ab\Delta} C$$

$$+ \frac{j\alpha_m \beta \sin^2 \dfrac{\pi d}{a}}{ab\left(1 + 2j\alpha_m \dfrac{\beta}{ab} \sin^2 \dfrac{\pi d}{a}\right)} C \tag{4.136}$$

The amplitude $A_2 + A_4$ is given by this expression with the sign of the last term changed from positive to negative. If we now set $A_2 + A_4$ equal to zero, we find that this can occur for

$$\sin \frac{\pi d}{a} = \frac{\lambda_0}{\sqrt{6}\, a} \tag{4.137}$$

Thus an aperture position exists such that there is no radiation through port 4. Power entering port 1 in Fig. 4.34 is coupled into ports 2 and 3 only. If the incident field were through port 2, we would find that there is no power coupled into port 3. A four-port network with these properties is a directional coupler, about which more is said in Chap. 6.

In the lower guide the aperture dipoles will radiate dominant-mode fields with amplitudes that are the negative of those for the modes in the upper guide since the sign of the dipoles is reversed. Thus the reflected and transmitted electric fields are

$$\mathbf{E} = (C - A_1 - A_3)\mathbf{E}_{10}^{+} \qquad z > 0 \tag{4.138a}$$

$$\mathbf{E} = -(A_2 + A_4)\mathbf{E}_{10}^{-} \qquad z < 0 \tag{4.138b}$$

This shows that when the aperture is positioned so that there is no power coupled into port 4, then there is also no reflected power in port 1.

We will define the aperture susceptance $j\overline{B}$ and aperture reactance $j\overline{X}$ by the expressions

$$j\overline{B} = j\frac{2\alpha_e k_0^2}{\beta ab}\sin^2\frac{\pi d}{a} + j\frac{2\alpha_m(\pi/a)^2}{\beta ab}\cos^2\frac{\pi d}{a} \tag{4.139a}$$

$$j\overline{X} = j\frac{2\alpha_m \beta}{ab}\sin^2\frac{\pi d}{a} \tag{4.139b}$$

In terms of these parameters, the expressions for the amplitudes $A_1 + A_3$ and $A_2 + A_4$ can be written as

$$A_1 + A_3 = \frac{j\overline{B}/2}{1 + j\overline{B}}C + \frac{j\overline{X}/2}{1 + j\overline{X}}C \tag{4.140a}$$

$$A_2 + A_4 = \frac{j\overline{B}/2}{1 + j\overline{B}}C - \frac{j\overline{X}/2}{1 + j\overline{X}}C \tag{4.140b}$$

The condition $A_2 + A_4 = 0$ is met by setting $\overline{X} = \overline{B}$. Since α_e is negative, we can obtain a negative value of $\overline{B}$. Thus the amplitude $A_1 + A_3$ can also be made small by setting $\overline{X} = -\overline{B}$. However, in this case $A_1 + A_3$ does not vanish exactly. In order for $A_1 + A_3$ to equal zero, we would require $\overline{X} + \overline{B} = 0$ and $\overline{X}\overline{B} = 0$ which has only the trivial solution. By using $\overline{X} = -\overline{B}$ we obtain an imperfect directional coupler.

PROBLEMS

4.1. For TM modes in a waveguide, show that the line integral of the transverse electric field between any two points on the boundary is zero.

Hint: Note that $\nabla_t e_z \cdot d\mathbf{l} = (de_z/dl)\,dl$ = directional derivative of e_z along the path. Integrate this and use the boundary conditions for e_z. As an alternative, note that there is no axial magnetic flux, so that the line integral around a closed path in the transverse plane must vanish.

4.2. For TE modes show that the line integral of the transverse electric field between two points located on the guide boundary depends on the path of integration chosen.

Hint: Note that, because there is an axial magnetic field, the line integral around a closed path does not vanish.

4.3. An obstacle located at $z = 0$ excites evanescent H modes that decay exponentially away from the obstacle in the positive z direction. Integrate the complex

Poynting vector over cross-sectional planes at $z = 0$ and $z = \infty$ and the guide walls for the nmth evanescent H mode, and show that there is no power transmitted into the region $z > 0$. Show also that the reactive energy stored in the nonpropagating H mode in the region $z > 0$ is predominantly magnetic.

Hint: Note from (2.59) that the total inward flux of the complex Poynting vector equals the power loss (which is to be taken equal to zero in this problem) plus $2j\omega(W_m - W_e)$.

4.4. Repeat Prob. 4.3 for the case of an E mode and show that nonpropagating E modes store predominantly electric energy.

4.5. For the circuits illustrated in Fig. P4.5, verify that the slope of the reactance function is given by (4.25).

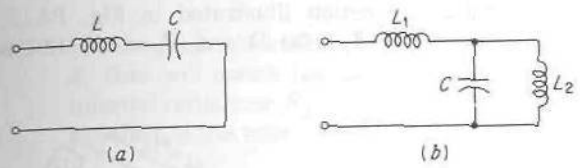


FIGURE P4.5

***4.6.** For the N-port junction choose an excitation such that all $I_n = 0$ except I_j; thus $V_i = Z_{ij}I_j$ for all i. Show that all Z_{ij} must have real parts that are even functions of ω and imaginary parts that are odd functions of ω.

***4.7.** Generalize the result (4.25) to show that for a lossless N-port junction

$$[I^*]_t\left[\frac{\delta Z}{\delta \omega}\right][I] = \sum_{n=1}^{N}\sum_{m=1}^{N} I_n^* \frac{\delta Z_{nm}}{\delta \omega} I_m = 4j(W_e + W_m)$$

4.8. Verify that (4.51) and (4.44) are equal.

4.9. Show that a length l of transmission line of characteristic impedance Z_c is equivalent to a T network with parameters

$$Z_{11} = Z_{22} = -jZ_c \cot \beta l \qquad Z_{12} = -jZ_c \csc \beta l$$

4.10. Let $Z_{sc}^1, Z_{sc}^2, Z_{oc}^1, Z_{oc}^2$ be the input impedance of a T network when terminals 2 are short-circuited, when terminals 1 are short-circuited, when terminals 2 are open-circuited, and when terminals 1 are open-circuited, respectively. In terms of these impedances show that the parameters of the T network are given by

$$Z_{11} = Z_{oc}^1 \qquad Z_{22} = Z_{oc}^2 \qquad Z_{12}^2 = (Z_{oc}^1 - Z_{sc}^1)Z_{oc}^2 = (Z_{oc}^2 - Z_{sc}^2)Z_{oc}^1$$

Use these relations to verify the equations for the circuits of Figs. 4.13*d* and *e*.

4.11. For the mircowave circuit shown in Fig. P4.11, evaluate the power transmitted to the load Z_L. Find the standing-wave ratio in the two transmission-line

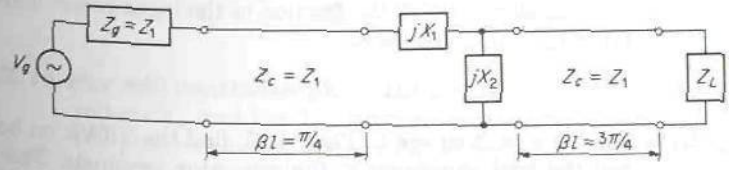


FIGURE P4.11

sections. Assume $Z_L = 2Z_1$, $X_1 = X_2 = Z_1$, $V_g = 5$ V (peak). Check your answers using TLINE.

4.12. For the microwave junction shown in Fig. P4.12, the equivalent-T-network parameters are $Z_{11} = j2$, $Z_{12} = j/\sqrt{2}$, $Z_{22} = -j0.25$. Find the parameters for the alternative equivalent circuit illustrated.

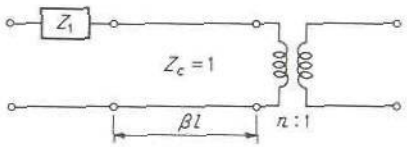


FIGURE P4.12

4.13. For the three-port junction illustrated in Fig. P4.13, compute the power delivered to the loads $Z_1 = 50\ \Omega$ and $Z_2 = 100\ \Omega$. Assume that $V_g = 10$ V peak.

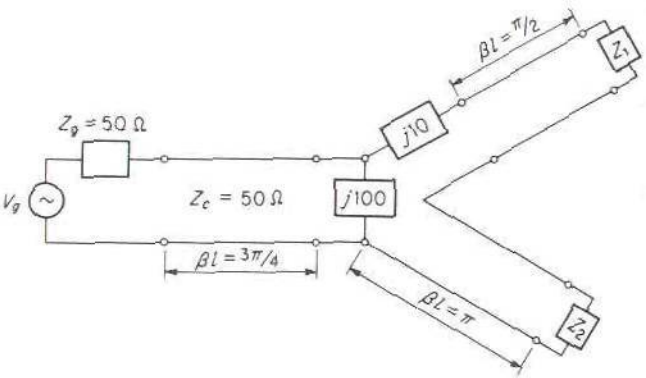


FIGURE P4.13

4.14. For the transmission-line circuit shown in Fig. P4.14, find (*a*) the load reflection coefficient, (*b*) the impedance seen by the generator, (*c*) the VSWR on the

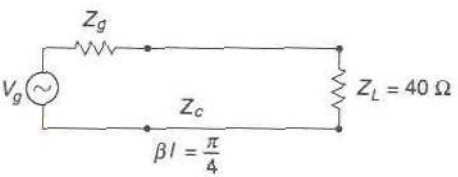


FIGURE P4.14

transmission line, (*d*) the fraction of the input power delivered to the load. The following parameters apply:

$$Z_c = 50\ \Omega \qquad Z_g = 50\ \Omega \qquad \beta l = \pi/4 \qquad Z_L = 40\ \Omega$$

4.15. For the circuit shown in Fig. P4.15, find the VSWR on both transmission lines and the load impedance at the generator terminals. The following parameters

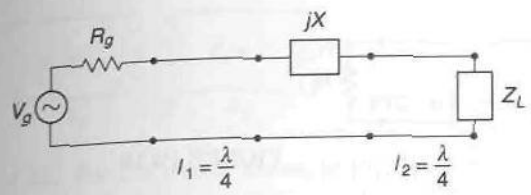


FIGURE P4.15

apply:

$$l_1 = \frac{\lambda}{4} \quad l_2 = \frac{\lambda}{4} \quad Z_c = 50\ \Omega \quad Z_g = 50\ \Omega \quad Z_L = 25 + j25 \quad jX = j25$$

4.16. For the transmission-line circuit shown in Fig. P4.16, find the required value of Z_c that will match the 20-Ω load resistance to the generator. The generator internal resistance $R_g = 60\ \Omega$. Find the VSWR on the transmission line. Is R_L matched to the transmission line?

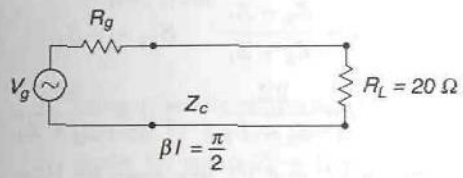


FIGURE P4.16

4.17. For the transmission-line circuit shown in Fig. P4.17, find the VSWR on each

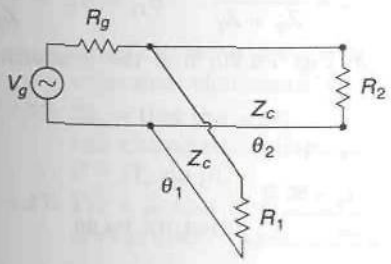


FIGURE P4.17

transmission line and the relative powers delivered to R_1 and R_2. The following parameters apply:

$$\theta_1 = \pi \quad \theta_2 = \frac{\pi}{2} \quad R_1 = 25\ \Omega \quad R_2 = 75\ \Omega \quad Z_c = 50\ \Omega \quad R_g = 50\ \Omega$$

4.18. For the transmission-line circuit shown in Fig. P4.18, find the open-circuit voltage V_{oc} and the Thévenin equivalent impedance. Use these results to find the power delivered to R_L. Assume that $\beta l = \pi/4$, $R_g = 75\ \Omega$, $Z_c = 50\ \Omega$, $R_L = 30\ \Omega$.

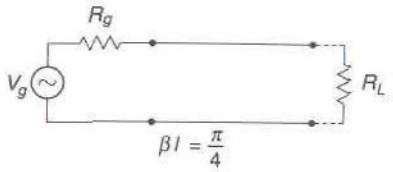


FIGURE P4.18

4.19. Consider the junction of two transmission lines with characteristic impedances Z_1 and Z_2 as illustrated in Fig. P4.19. When the usual transmission-line

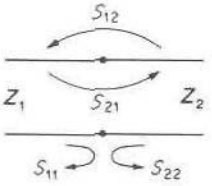


FIGURE P4.19

voltages and currents are used, show that the scattering-matrix parameters are given by

$$S_{11} = \frac{Z_2 - Z_1}{Z_2 + Z_1} \qquad S_{22} = -S_{11}$$

$$S_{12} = \frac{2Z_1}{Z_2 + Z_1} \qquad S_{21} = \frac{2Z_2}{Z_2 + Z_1}$$

The normalized voltages $\bar{V}_1^+$ and $\bar{V}_2^+$ are given by $\bar{V}_1^+ = \sqrt{Y_1}\, V_1^+$, $\bar{V}_2^+ = \sqrt{Y_2}\, V_2^+$, where the unprimed quantities are the usual transmission-line voltages. When normalized voltages are used, show that the scattering-matrix parameters are

$$S_{11} = -S_{22} = \frac{Z_2 - Z_1}{Z_2 + Z_1} \qquad S_{12} = S_{21} = \frac{2\sqrt{Z_1 Z_2}}{Z_2 + Z_1}$$

4.20. For the circuit shown in Fig. P4.20, find the scattering-matrix parameters.

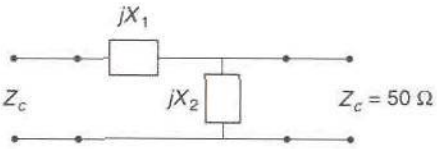


FIGURE P4.20

When $jX_1 = j25$, $jX_2 = j100$ verify that

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

$$S_{11}S_{12}^* + S_{12}S_{22}^* = 0$$

4.21. Find Z_{oc}^1, Z_{oc}^2, Z_{sc}^1, and Z_{sc}^2 for the circuit shown in Fig. P4.21.

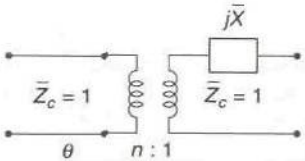


FIGURE P4.21

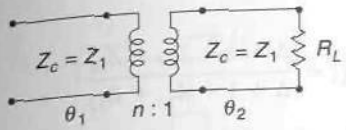


FIGURE P4.22

4.22. For the circuit shown in Fig. P4.22, *find the input reflection coefficient* Γ_{in}.

4.23. For the circuit shown in Fig. P4.23, find the expressions for Z_{in}, Γ_{in}, and the input VSWR. *Find the value of* jX *that will minimize the* VSWR.

Hint: The VSWR will be minimum when $|\Gamma_{in}|^2$ is a minimum. Why?

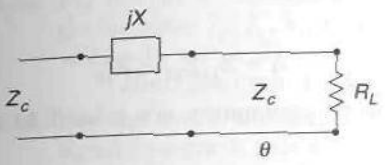


FIGURE P4.23

4.24. The field of a TE_{11} mode in a rectangular guide of width a and height b is derived from

$$h_z = C \cos \frac{\pi x}{a} \cos \frac{\pi y}{b}$$

Determine the expressions for the equivalent-transmission-line voltage V^+ and current I^+ for the two cases (1) when $Z_c = Z_w$ = wave impedance of the TE_{11} mode, (2) when $Z_c = 1$.

4.25. Apply the complex Poynting vector theorem to show that, for a one-port microwave termination, the reflection coefficient Γ *satisfies the relation*

$$(1 + \Gamma)(1 - \Gamma^*) = \frac{2j\omega(W_m - W_e) + P_l}{\frac{1}{2}|V^+|^2}$$

when the wave amplitudes are normalized, so that $VV^* = II^*$, that is, the equivalent characteristic impedance is unity.

4.26. Show that the $\mathscr{A}\mathscr{B}\mathscr{C}\mathscr{D}$ parameters for a section of transmission line of length l and characteristic impedance Z_c are given by $\mathscr{A} = \mathscr{D} = \cos \beta l$, $\mathscr{B} = jZ_c \sin \beta l$, $\mathscr{C} = jY_c \sin \beta l$.

4.27. For a section of transmission line of length l, show that the wave-amplitude transmission matrix is a diagonal matrix with elements

$$A_{11} = e^{j\beta l} \qquad A_{22} = e^{-j\beta l} \qquad A_{12} = A_{21} = 0$$

4.28. Consider the junction of two transmission lines as in Prob. 4.19. Using conventional transmission-line voltages, show that the $[A]$-matrix parameters describing the junction are $A_{11} = A_{22} = (Z_1 + Z_2)/2Z_2$, $A_{12} = A_{21} = (Z_2 - Z_1)/2Z_2$. When normalized wave amplitudes (voltages) are used, show that $A_{11} = A_{22} = (Z_1 + Z_2)/(2\sqrt{Z_1 Z_2})$, $A_{12} = A_{21} = (Z_2 - Z_1)/(2\sqrt{Z_1 Z_2})$.

4.29. Find the $[A]$-matrix parameters for a shunt susceptance jB connected across a transmission line of unit characteristic impedance. Repeat for a reactance jX connected in series with the line.

4.30. Show that when normalized voltages are used the scattering-matrix parameters of a two-port junction are given in terms of the equivalent-T-network parame-

ters by

$$S_{11} = \frac{\Delta - 1 + Z_{11} - Z_{22}}{\Delta + 1 + Z_{11} + Z_{22}}$$

$$S_{12} = S_{21} = \frac{2Z_{12}}{\Delta + 1 + Z_{11} + Z_{22}}$$

$$S_{22} = \frac{\Delta - 1 + Z_{22} - Z_{11}}{\Delta + 1 + Z_{11} + Z_{22}}$$

where

$$\Delta = Z_{11}Z_{22} - Z_{12}^2$$

4.31. Show that the T-network parameters are related to the scattering-matrix parameters as follows:

$$Z_{11} = \frac{(1 + S_{11})(1 - S_{22}) + S_{12}^2}{W}$$

$$Z_{22} = \frac{(1 - S_{11})(1 + S_{22}) + S_{12}^2}{W}$$

$$Z_{12} = \frac{2S_{12}}{W}$$

where

$$W = (1 - S_{11})(1 - S_{22}) - S_{12}^2$$

4.32. For a discontinuity in a waveguide, the following scattering-matrix parameters were measured:

$$S_{11} = \tfrac{1}{3} + j\tfrac{2}{3} \qquad S_{12} = j\tfrac{2}{3} \qquad S_{22} = \tfrac{1}{3} - j\tfrac{2}{3}$$

Find the parameters of an equivalent T network that will represent the discontinuity (Prob. 4.31).

4.33. For an E-plane step (Fig. 4.6), the following were measured:

$$S_{11} = \frac{1 - j}{3 + j} \qquad S_{22} = \frac{-(1 + j)}{3 + j}$$

An equivalent circuit of the form illustrated in Fig. P4.33 is to be used to represent the junction. Determine the susceptance jB and the ideal transformer turns ratio n:1 from the above given data.

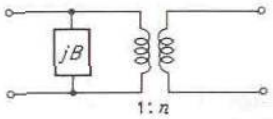


FIGURE P4.33

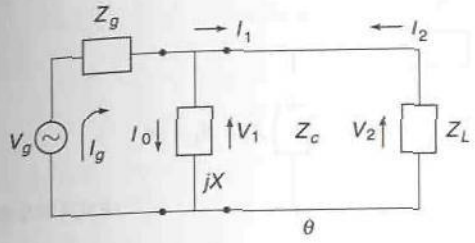


FIGURE P4.34

***4.34.** For the circuit illustrated in Fig. P4.34 construct a signal flow graph relating the variables V_g, I_g, V_1, I_1, V_2, and I_2. Use signal flow graph analysis to find the voltage V_2 across Z_L.

Hint: See Prob. 4.9 for the T-network parameters of a transmission line.

***4.35.** For the circuit illustrated in Fig. P4.35, construct a signal flow graph. Use signal flow graph reduction to derive an expression for the load voltage V_3. Use the $\mathscr{ABCD}$ chain matrix to write relationships between the variables.

Hint: See Prob. 4.26.

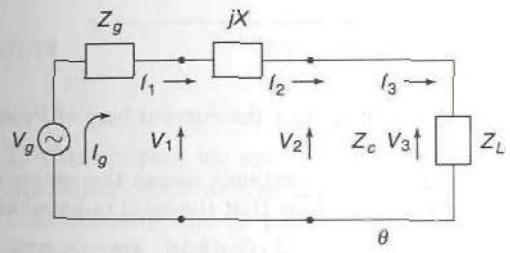


FIGURE P4.35

***4.36.** For the circuit illustrated in Fig. P4.36, construct a signal flow graph using the scattering-matrix relationships between the variables. From a reduction of the signal flow graph find the load voltage $\bar{V}_L$.

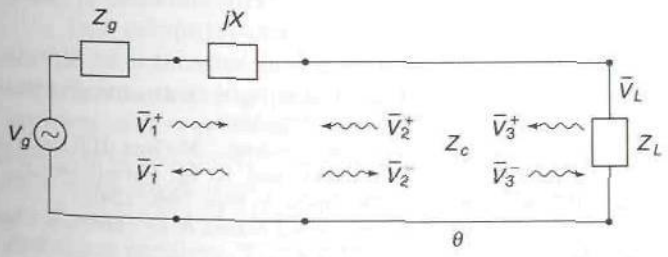


FIGURE P4.36

***4.37.** For the transmission-line circuit shown in Fig. P4.37, find the generalized scattering-matrix parameters for the series reactance jX. Use these results to derive an expression for the power delivered to Z_L. Verify your answer using a more conventional method of analysis.

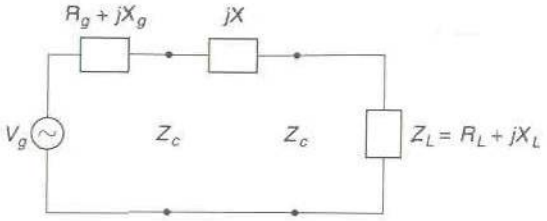


FIGURE P4.37

*4.38. Repeat Prob. 4.37 with jX replaced by a shunt element jX_s.

*4.39. Find the TE_{10} field radiated by the current loop illustrated in Fig. P4.39. Consider the loop area to be so small that (4.119) is applicable. The area of loop equals S_0.

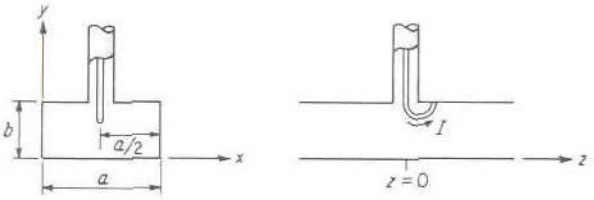


FIGURE P4.39

*4.40. Find the TE_{10} field radiated by the current loop of Prob. 4.39 if a short circuit is placed at $z = -l$.

*4.41. A linear constant current I extends across the center of a rectangular waveguide at $x = a/2$, $z = 0$. Show that the total radiated electric field is

$$E_y = \frac{-j\omega\mu_0 I}{a} \sum_{n=1}^{\infty} \frac{1}{\gamma_n} \sin \frac{n\pi}{2} \sin \frac{n\pi x}{a} e^{-\gamma_n |z|}$$

where

$$\gamma_n = j\beta_n = \left(\frac{n^2\pi^2}{a^2} - k_0^2 \right)^{1/2}$$

REFERENCES

1. Kerns, D. M.: Basis of Application of Network Equations to Waveguide Problems, *J. Res. Natl. Bur. Std.*, vol. 42, pp. 515–540, 1949.
2. Marcuvitz, N. (ed.): "Waveguide Handbook," McGraw-Hill Book Company, New York, 1951.
3. Montgomery, C. G., R. H. Dicke, and E. M. Purcell (eds.): "Principles of Microwave Circuits," McGraw-Hill Book Company, New York, 1948.
4. Pannenborg, A. E.: On the Scattering Matrix of Symmetrical Waveguide Junctions, *Philips Res. Rept.*, vol. 7, pp. 131–157, 1952.

CHAPTER 5

IMPEDANCE TRANSFORMATION AND MATCHING

In this chapter we are concerned with the important problem of impedance matching, such as the matching of an arbitrary load impedance to a given transmission line or the matching of two lines with different characteristic impedances. Methods of impedance matching to obtain maximum power transfer are presented, along with broadband design methods for quarter-wave transformers and tapered transmission-line impedance transformers. To facilitate the development of the theory, the Smith chart, a graphical aid for the solution of many transmission-line and waveguide impedance problems, is described first.

In a computer-oriented age the reader may very well question why one should be interested in a graphical aid, such as the Smith chart, to solve an impedance-matching problem. There are two basic reasons why the microwave engineer needs to be familiar with the Smith chart. One reason is that using the Smith chart to solve an impedance-matching problem shows in a very vivid way how adding reactive elements moves the impedance point around and this provides considerable insight into the impedance-matching problem. The second reason is that the Smith chart is widely used in the industry to display the performance of a microwave circuit in terms of input impedance versus frequency, VSWR or reflection coefficient versus frequency, the frequency variation of scattering-matrix parameters, etc.

In microwave amplifier design the Smith chart is indispensible as a visual aid to show how gain, noise figure, stability, and input and output matching are interrelated and how these operating characteristics depend

on the load and source impedances. Without the aid of the Smith chart, an intuitive understanding of microwave amplifier design would be much more difficult to acquire.

5.1 SMITH CHART

In Sec. 3.6 it was shown that a load impedance Z_L was transformed into an impedance

$$Z_{\text{in}} = Z_c \frac{Z_L + jZ_c \tan \beta l}{Z_c + jZ_L \tan \beta l} \tag{5.1}$$

when viewed through a length l of transmission line with characteristic impedance Z_c. This formula is valid for any waveguiding system with phase constant β, provided the impedances are properly interpreted in terms of suitably defined equivalent voltages and currents. Alternatively, the reflection coefficient $\Gamma(l)$, a distance l from the termination, is uniquely given by

$$\Gamma(l) = \frac{Z_{\text{in}}(l) - Z_c}{Z_{\text{in}}(l) + Z_c} \tag{5.2}$$

with $Z_{\text{in}}(l)$ given by (5.1). The reflection coefficient is a physical quantity that can be measured, and the normalized impedances Z_{in}/Z_c and Z_L/Z_c may therefore be appropriately defined in terms of the reflection coefficient Γ at any point l on the line and the reflection coefficient Γ_L of the load, as follows:

$$\bar{Z}_{\text{in}} = \frac{Z_{\text{in}}}{Z_c} = \frac{1 + \Gamma(l)}{1 - \Gamma(l)} = \frac{1 + \Gamma_L e^{-2j\beta l}}{1 - \Gamma_L e^{-2j\beta l}} \tag{5.3a}$$

$$\bar{Z}_L = \frac{Z_L}{Z_c} = \frac{1 + \Gamma_L}{1 - \Gamma_L} \tag{5.3b}$$

The Smith chart is a graphical representation of the impedance-transformation property of a length of transmission line as given by (5.1). Clearly, it would be impractical to plot all values of Z_L and Z_{in} on a rectangular coordinate impedance plane, with one coordinate representing the real part, or resistance, and the other coordinate representing the reactance, since this would require a semiinfinite sheet of paper. On the other hand, all values of the reflection coefficient lie within a unit circle in the reflection-coefficient plane since $|\Gamma| \leq 1$. Furthermore, each value of Γ specifies a value of normalized input impedance by means of (5.3*a*), so that there is a one-to-one correspondence between reflection coefficient and input impedance. Instead of plotting contours of constant values of the reflection coefficient, contours of constant values of input resistance and input reactance are plotted on the reflection-coefficient plane. For a given value of the reflection coefficient, the corresponding input impedance can be read di-

rectly from the plot. In addition, a movement a distance d along the line corresponds to a change in the reflection coefficient by a factor $e^{-2j\beta d}$ only. This is represented by a simple rotation through an angle $2\beta d$; so the corresponding impedance point moves on a constant radius circle through this angle to its new value. The chart thus enables the transformation of impedance along a transmission line to be evaluated graphically in an efficient and straightforward manner. A more detailed description of the chart and its use is given below. In addition, a number of matching problems are solved with the aid of the Smith chart in later sections of this chapter.

Let the reflection coefficient Γ be expressed in polar form as

$$\Gamma = \rho e^{j\theta} \tag{5.4}$$

where $\rho = |\Gamma|$ and $\theta = \angle\Gamma = \angle\Gamma_L - 2\beta l$. Let the normalized input impedance be

$$\overline{Z}_{\text{in}} = \frac{Z_{\text{in}}}{Z_c} = \overline{R} + j\overline{X} = \frac{1+\Gamma}{1-\Gamma} = \frac{1+\rho e^{j\theta}}{1-\rho e^{j\theta}} \tag{5.5}$$

From (5.5) it is readily found that in the reflection-coefficient plane (ρ, θ plane), the contours of constant $\overline{R}$ and constant $\overline{X}$ are given by (Prob. 5.1)

$$\left(u - \frac{\overline{R}}{\overline{R}+1}\right)^2 + v^2 = \frac{1}{(\overline{R}+1)^2} \tag{5.6a}$$

$$(u-1)^2 + \left(v - \frac{1}{\overline{X}}\right)^2 = \frac{1}{\overline{X}^2} \tag{5.6b}$$

where $u = \rho\cos\theta$ and $v = \rho\sin\theta$ and are rectangular coordinates in the ρ, θ plane. The above constant $\overline{R}$ and constant $\overline{X}$ contours are circles and plot as illustrated in Fig. 5.1.

For convenience in using the chart, a scale giving the angular rotation $2\beta l = 4\pi l/\lambda$ in terms of wavelength λ is attached along the circumference of the chart. Note that moving away from the load (toward the generator) corresponds to going around the chart in a clockwise direction, as illustrated in Fig. 5.2. A complete revolution around the chart is made in going a distance $l = \lambda/2$ along the transmission line. At these intervals the input impedance repeats itself. The origin for the angular scale is arbitrarily chosen at the left side of the circle.

To illustrate the use of the chart, let a line be terminated in a load impedance $\overline{R}_1 + j\overline{X}_1 = 0.5 + j0.5$. This point is located in Fig. 5.2 and labeled P_1. At a distance $l = 0.2\lambda$ away, the corresponding input impedance may be found as follows: A constant-radius circle through P_1 is constructed first. The new impedance point P_2 lies on this circle at an angle $2\beta l = 0.8\pi$

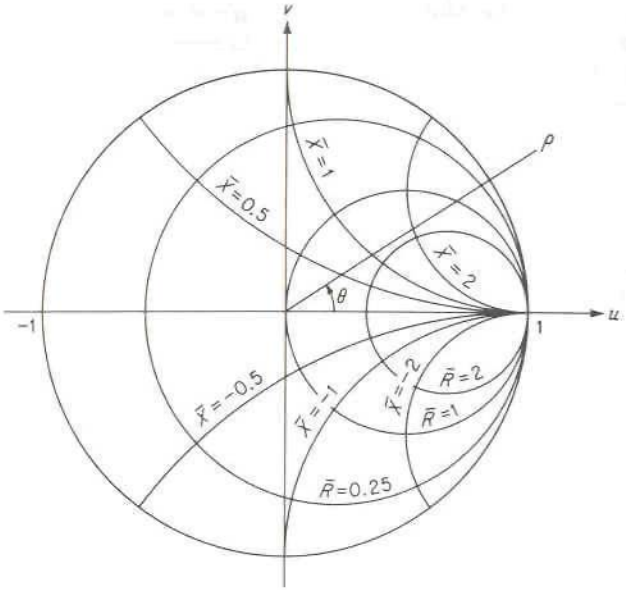


FIGURE 5.1
Constant $\bar{R}$ and $\bar{X}$ circles in the reflection-coefficient plane.

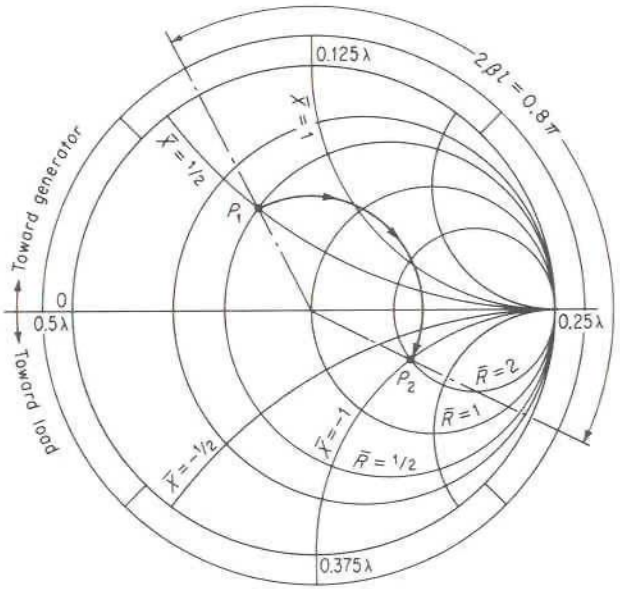


FIGURE 5.2
The Smith chart.

rad in a clockwise direction from P_1. This angular rotation is readily carried out by adding 0.2λ to the wavelength reading obtained from the intersection of the radius vector through P_1 and the angular scale at the circumference of the chart. From the chart it is found that the new value of normalized impedance is

$$\bar{R}_2 + j\bar{X}_2 = 2 - j1.04$$

If we begin at a point P_1, where the impedance is $\overline{R}_1 + j\overline{X}_1$, and move on a constant-radius circle an amount $\lambda/4$ to arrive at a point diametrically opposite, Γ_1 changes into $-\Gamma_1$ ($2\beta l$ changes by π), and we obtain an impedance

$$\overline{R}_2 + j\overline{X}_2 = \frac{1 - \Gamma_1}{1 + \Gamma_1} = \frac{1}{\overline{R}_1 + j\overline{X}_1} = \overline{G}_1 + j\overline{B}_1$$

Thus the input normalized admittance $\overline{G}_1 + j\overline{B}_1$ corresponding to a given input impedance $\overline{R}_1 + j\overline{X}_1$ may be found from the value of impedance at a point diametrically across from the first impedance point, provided $\overline{R}_2$ and $j\overline{X}_2$ are interpreted as the input conductance and susceptance. To clarify this, note that $\overline{R}_2 + j\overline{X}_2$ at P_2 is the normalized input impedance at point P_2 and equals the normalized input admittance at point P_1 at a distance $l = \lambda/4$ away.

The Smith chart may be used to find the transformation of admittances equally well. All that is required is to interpret the constant resistance and reactance contours (constant $\overline{R}$ and $j\overline{X}$ contours) as constant conductance $\overline{G}$ and susceptance $j\overline{B}$ contours. Note that a positive $\overline{X}$ corresponds to an inductive reactance but a positive $\overline{B}$ corresponds to a capacitive susceptance.

In order to facilitate the use of the Smith chart in situations where it is necessary to convert back and forth between impedances and admittances, the impedance-admittance chart is used. This chart has a second Smith chart, rotated by 180°, superimposed on the regular Smith chart. Thus one set of circles gives impedance values and the second set of rotated circles, usually shown in a different color, give the corresponding admittances values directly. For the conventional Smith chart, the admittance

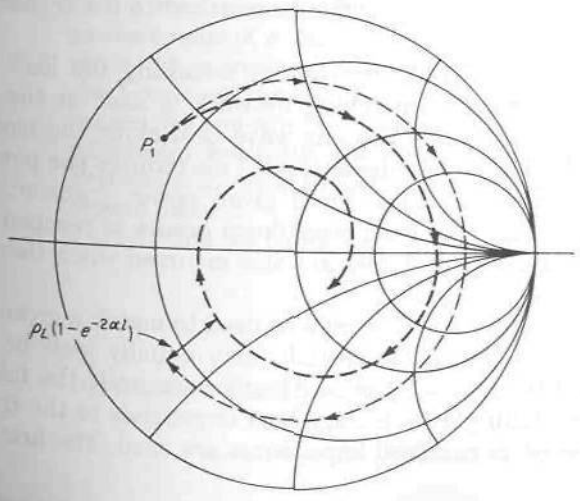


FIGURE 5.3
Inward spiraling of the impedance point on a Smith chart for a lossy transmission line.

values are obtained by rotating the impedance points by 180°. In the impedance-admittance chart, the second set of rotated circles makes it unnecessary to rotate the impedance point in order to determine the corresponding value of the admittance.

On a lossy line the reflection coefficient at any point is given by

$$\Gamma = \Gamma_L e^{-2\alpha l - 2j\beta l} = \rho_L e^{-2\alpha l - 2j\beta l + j\theta_L} \tag{5.7}$$

As we move from the load toward the generator, $\rho = \rho_L e^{-2\alpha l}$ continually decreases, and hence we move along a spiral that eventually terminates at the center, as in Fig. 5.3. In practice, we move on a constant ρ circle first through the angle $2\beta l$ and then move in radially until we are a distance $\rho_L e^{-2\alpha l}$ from the center. Many practical charts have convenient scales attached to them, so that the amount of inward spiraling is readily obtained. Note that the center of the chart represents a matched condition ($\rho = 0$).

5.2 IMPEDANCE MATCHING WITH REACTIVE ELEMENTS

When a given load is to be connected to a generator by means of a transmission line or waveguide many wavelengths long, it is preferable to match the load and generator to the transmission line or waveguide at each end of the line. There are several reasons for doing this, perhaps the most important one being the great reduction in frequency sensitivity of the match. Although the transformed load impedance as seen from the generator end of the transmission line can be matched to the generator for maximum power transfer, a small change in the operating frequency will change the electrical length βl of a long line by an appreciable fraction of π rad, and hence greatly modify the effective load impedance seen at the generator end and thus modify the matching requirements as well. To avoid this frequency sensitivity of the matching requirements, the load and generator should be individually matched to the transmission line or waveguide.

Another disadvantage of not matching the load to the transmission line is that when a matching network is used at the generator end only, there may be a large standing-wave field along the transmission line if the original load is badly mismatched. This reduces the power-handling capability of the system since, for a given power transfer, the maximum field strength before dielectric breakdown occurs is reached sooner. In addition, greater transmission losses are also incurred when there is a standing-wave current along the line.

The techniques that may be used to match a given load impedance to a transmission line or waveguide may equally well be used to match the generator to the line. Hence it suffices to limit the following discussion to that of matching an arbitrary load impedance to the transmission line. For convenience, normalized impedances are used. The first matching technique

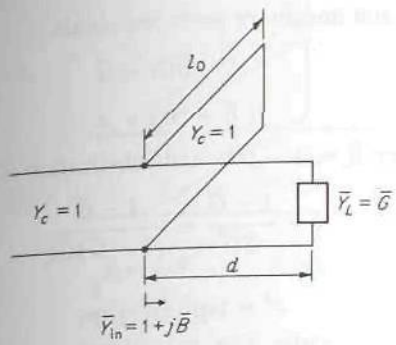


FIGURE 5.4
Single-shunt-stub matching network.

discussed employs short-circuited (or open-circuited) sections of transmission lines as reactive elements, and is referred to as *stub* matching. However, the principles involved are general in nature and may be applied to any waveguiding system by substituting suitable shunt or series reactive elements for the transmission-line stubs. A description of some typical reactive elements that may be used is given in a later section of this chapter.

Single-Stub Matching

Case 1 Shunt stub. Consider a line terminated in a pure conductive load of normalized admittance $\overline{Y}_L = \overline{G}$, as in Fig. 5.4. At some point a distance d from the load, the normalized input admittance will be $\overline{Y}_{in} = 1 + j\overline{B}$. At this point we can connect a stub with normalized input susceptance $-j\overline{B}$ across the line to yield a resultant

$$\overline{Y}_{in} = 1 + j\overline{B} - j\overline{B} = 1$$

that is, to arrive at a matched condition. The stub should be connected at the smallest value of d that will give $\overline{Y}_{in} = 1 + j\overline{B}$ in order to keep the frequency sensitivity as small as possible. The stub may be either an open-circuited or a short-circuited section of line, the latter being the most commonly used version for two-wire lines, coaxial lines, and waveguides because of ease in adjustment and better mechanical rigidity. In a microstrip circuit an open-circuited stub would be preferred since it does not require a connection to the ground plane.

To find the position d, we must solve the equation

$$\overline{Y}_{in} = 1 + j\overline{B} = \frac{\overline{Y}_L + jt}{1 + j\overline{Y}_L t} \qquad t = \tan \beta d$$

If we assume that $\overline{Y}_L = \overline{G}$ is pure real, we require

$$(1 + j\overline{B})(1 + j\overline{G}t) = \overline{G} + jt$$

or by equating real and imaginary parts, we obtain

$$1 - \overline{B}\overline{G}t = \overline{G} \tag{5.8a}$$

$$j(\overline{B} + \overline{G}t) = jt \tag{5.8b}$$

Equation (5.8b) gives $\overline{B} = (1 - \overline{G})t$, and substitution of this into (5.8a) yields

$$t = \frac{1 - \overline{G}}{\overline{B}\overline{G}} = \frac{1 - \overline{G}}{(1 - \overline{G})\overline{G}t}$$

or

$$t^2 = \tan^2 \beta d = \frac{1}{\overline{G}}$$

Replacing $\tan^2 \beta d$ by $(1 - \cos^2 \beta d)/\cos^2 \beta d$ finally gives

$$d = \frac{\lambda}{2\pi} \cos^{-1} \sqrt{\frac{\overline{G}}{1 + \overline{G}}} \tag{5.9}$$

where $\beta = 2\pi/\lambda$. Note that two principal values of d are possible, depending on which sign is chosen for the square root. An alternative relation is obtained if we replace $2\cos^2 \beta d$ by $1 + \cos 2\beta d$; thus

$$1 + \cos 2\beta d = \frac{2\overline{G}}{1 + \overline{G}}$$

and

$$\cos 2\beta d = \frac{\overline{G} - 1}{\overline{G} + 1}$$

which gives

$$d = \frac{\lambda}{4\pi} \cos^{-1} \frac{\overline{G} - 1}{\overline{G} + 1} \tag{5.10}$$

If d_1 is a solution of (5.10), then $\lambda/2 - d_1$ is another principal solution, since $\pm d_1 \pm n\lambda/2$ are all solutions of (5.10).

The value of the input susceptance $j\overline{B}$ is given by

$$\overline{B} = (1 - \overline{G})t = \frac{1 - \overline{G}}{\sqrt{\overline{G}}} \tag{5.11}$$

since $\tan^2 \beta d = 1/\overline{G}$. The required length l_0 of a short-circuited stub to give an input susceptance $-j\overline{B}$ is found from the relation

$$\overline{Y}_{in} = -j\overline{B} = -j \cot \beta l_0$$

and (5.11); thus

$$\cot \beta l_0 = \frac{1 - \overline{G}}{\sqrt{\overline{G}}}$$

or

$$l_0 = \frac{\lambda}{2\pi} \tan^{-1} \frac{\sqrt{\overline{G}}}{1 - \overline{G}} \tag{5.12}$$

where the sign of $\sqrt{\overline{G}}$ must be chosen to give the correct sign for $\overline{B}$ in (5.11). If

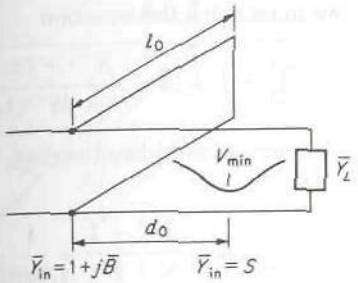


FIGURE 5.5
Location of stub relative to a voltage minimum.

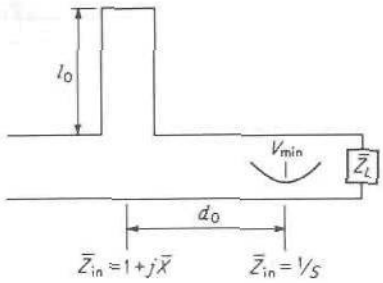


FIGURE 5.6
The series stub.

$0 < d < \lambda/4$, the positive square root should be used, whereas if the other solution, $\lambda/4 < d < \lambda/2$, is chosen, the negative square root must be used.

A similar analysis may be carried out when $\bar{Y}_L$ is complex, but it becomes more involved. The following procedure is usually followed instead. First locate a position of a voltage minimum from the load. At this point the reflection coefficient is a negative real quantity and the input admittance is pure real and given by

$$Y_{in} = \frac{1 - \Gamma}{1 + \Gamma} = \frac{1 + \rho}{1 - \rho} = S \tag{5.13}$$

where S is the standing-wave ratio on the line. Let d_0 be the distance from this voltage-minimum point to the point where $\bar{Y}_{in} = 1 + j\bar{B}$, as in Fig. 5.5. The equations to be solved for the stub position d_0 and stub length l_0 are the same as given earlier, but with S replacing $\bar{G}$. Hence

$$d_0 = \frac{\lambda}{4\pi} \cos^{-1} \frac{S - 1}{S + 1} \tag{5.14a}$$

$$l_0 = \frac{\lambda}{2\pi} \tan^{-1} \frac{\sqrt{S}}{S - 1} \tag{5.14b}$$

The position of the stub from the load is readily computed by finding the distance from the load to the V_{min} position and adding this to d_0. Note that a stub position $d_0 \pm \lambda/2$ is also a suitable one. Thus, if $d_0 - \lambda/2$ is still on the generator side of the load, the stub should be placed at this point instead of at d_0 in order to reduce the frequency sensitivity of the match.

Case 2 Series stub. At a position of a voltage minimum, $\bar{Z}_{in} = S^{-1}$. At some position d_0 from this point, $\bar{Z}_{in} = 1 + j\bar{X}$. By connecting a stub with a normalized input reactance of $-j\bar{X}$ in series with the line at this point, the resultant input impedance is reduced to unity and a matched condition is obtained. This series stub-matching network is illustrated in Fig. 5.6.

To find d_0, we must solve the equation

$$\bar{Z}_{\text{in}} = 1 + j\bar{X} = \frac{S^{-1} + j \tan \beta d_0}{1 + jS^{-1} \tan \beta d_0}$$

This is the same equation as considered earlier, with $\bar{X}, S^{-1}$ replacing $\bar{B}, S$, and thus the solutions are

$$d_0 = \frac{\lambda}{4\pi} \cos^{-1} \frac{S^{-1} - 1}{S^{-1} + 1} = \frac{\lambda}{4\pi} \cos^{-1} \frac{1 - S}{1 + S}$$

$$= \pm \frac{\lambda}{2} + \frac{\lambda}{4\pi} \cos^{-1} \frac{S - 1}{S + 1} \tag{5.15a}$$

$$\bar{X} = \left(1 - \frac{1}{S}\right) \tan \beta d_0 = \frac{S - 1}{\sqrt{S}} \tag{5.15b}$$

where the sign of $\sqrt{S}$ must be chosen to yield the correct sign for $\tan \beta d_0$; that is, for $0 < d_0 < \lambda/4$, use $+\sqrt{S}$, and for $\lambda/4 < d_0 < \lambda/2$, use $-\sqrt{S}$.

The required stub length l_0 is determined from the relation

$$j \tan \beta l_0 = -j\bar{X}$$

and hence, from (5.15*b*), we obtain

$$l_0 = \frac{\lambda}{2\pi} \tan^{-1} \frac{1 - S}{\sqrt{S}} \tag{5.16}$$

The shunt stub is most commonly used for coaxial lines because it is easy to construct a shunt stub for a coaxial line, whereas a series stub is difficult to build. A disadvantage with a single-stub-matching system is that every load requires a new stub position. The use of two stubs spaced by a fixed amount and located a fixed distance from the load may be used to overcome this disadvantage. However, a double-stub-matching system of this type will not match all possible values of load admittance. The theory of double-stub matching is presented in the next section.

5.3 DOUBLE-STUB MATCHING NETWORK

The double-stub tuner, or matching network, is illustrated schematically in Fig. 5.7. We may transform the normalized load admittance $\bar{Y}_L'$ into an equivalent load admittance $\bar{Y}_L = \bar{G}_L + j\bar{B}_L$ at the plane aa and treat the problem illustrated in Fig. 5.7*b* without loss in generality.

Let the point P_1 on the Smith chart in Fig. 5.8 represent $\bar{Y}_L$. The first stub adds a susceptance $j\bar{B}_1$ which moves P_1 along a *constant-conductance circle* to point P_2 in Fig. 5.8. At the plane bb just on the right-hand side of the second stub, the input admittance is $\bar{Y}_b = \bar{G}_b + j\bar{B}_b$, and is obtained by moving along a constant-radius circle from P_2 to P_3 through an angle $\phi = 2\beta d = 4\pi d/\lambda$ rad in a clockwise sense. The point P_3 must lie on the $\bar{G} = 1$ circle if the addition of a susceptance $j\bar{B}_2$ contributed by the second

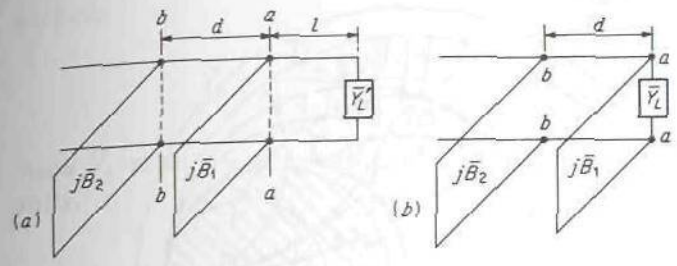


FIGURE 5.7
The double-stub tuner.

stub is to move point P_3 *into the center of the chart (matched condition)* along the $\overline{G} = 1$ circle.

From the description just given, it is clear that the first stub must add *a susceptance of just the right amount, so that after the admittance at plane aa* is transformed through a length of line d, we end up at a point on the $\overline{G} = 1$ circle. The required value of susceptance $j\overline{B}_1$ to be contributed by *the first stub may be obtained by rotating the* $\overline{G} = 1$ *circle through an angle* $-\phi$. The intersection of the rotated $\overline{G} = 1$ circle and the $\overline{G}_L$ circle determines the point P_2, and hence $j\overline{B}_1$, as illustrated in Fig. 5.9. A point P_2' would also be suitable; the location P_3' then corresponds to the admittance just to the right of the second stub.

From Fig. 5.9 it is clear that for all values of $\overline{Y}_L$ that lie within the $\overline{G} = \overline{G}_0$ circle, a match cannot be obtained since all values $\overline{G} > \overline{G}_0$ will not intersect the rotated $\overline{G} = 1$ circle. The conductance circle $\overline{G} = \overline{G}_0$ is tangent

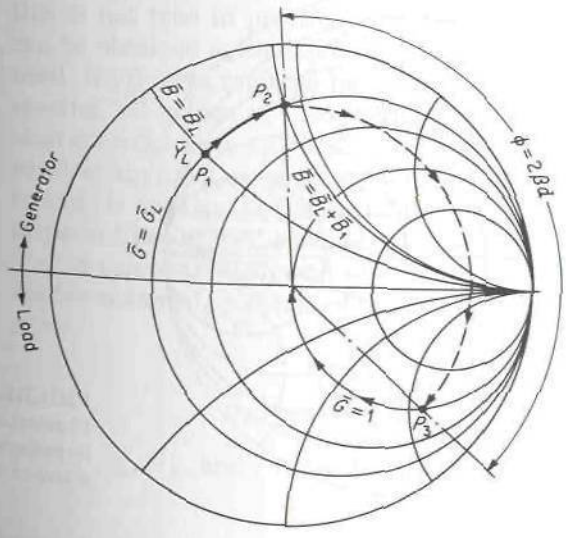


FIGURE 5.8
Graphical representation of the operation of a double-stub tuner.

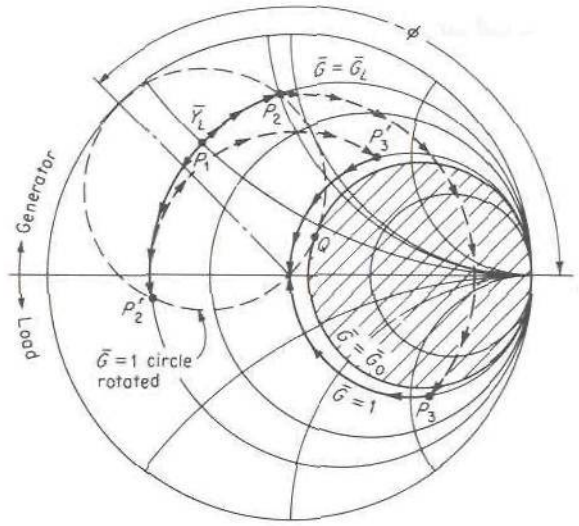


FIGURE 5.9
Graphical determination of required susceptance for the first stub in a double-stub tuner.

to the rotated $\bar{G} = 1$ circle at the point Q. It is easy to see that the smaller the distance d, the larger the range of load admittances that may be matched (see Fig. 5.10 for the case of $d = \lambda/8$, $\phi = \pi/2$). Also note that $\bar{G}_0$ will always be greater than unity; so all loads with $\bar{G}_L < 1$ can be matched.

At plane aa in Fig. 5.7 we have $\bar{Y}_L = \bar{G}_L + j\bar{B}_L$. Just to the left of the first stub we have $\bar{Y}_a = \bar{G}_L + j\bar{B}_L + j\bar{B}_1$. Just to the right of the second stub

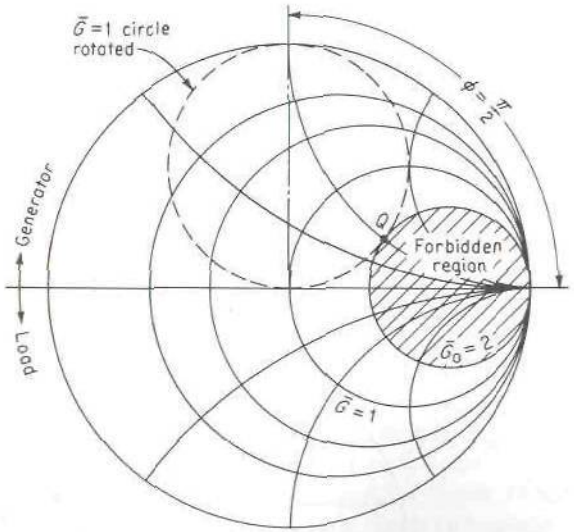


FIGURE 5.10
Illustration of range of load impedance which cannot be matched when $d = \lambda/8$.

we have

$$\bar{Y}_b = \frac{\bar{G}_L + j\bar{B}_L + j\bar{B}_1 + jt}{1 + jt\left(\bar{G}_L + j\bar{B}_L + j\bar{B}_1\right)} \qquad t = \tan \beta d \tag{5.17}$$

Since $\bar{Y}_b$ must equal $1 + j\bar{B}$, (5.17) gives, upon equating the real part to unity,

$$\bar{G}_L^2 - \bar{G}_L \frac{1 + t^2}{t^2} + \frac{\left(1 - \bar{B}_L t - \bar{B}_1 t\right)^2}{t^2} = 0 \tag{5.18a}$$

or

$$\bar{G}_L = \frac{1 + t^2}{2t^2}\left[1 \pm \sqrt{1 - \frac{4t^2\left(1 - \bar{B}_L t - \bar{B}_1 t\right)^2}{\left(1 + t^2\right)^2}}\right] \tag{5.18b}$$

In (5.18*b*) we note that the term under the radical sign equals one minus a positive quantity. Since $\bar{G}_L$ must be real, the term under the radical sign must be positive, or zero. Hence the value of the square-root term lies between zero and one. The corresponding limits on $\bar{G}_L$ are

$$0 \le \bar{G}_L \le \frac{1 + t^2}{t^2} = \frac{1}{\sin^2 \beta d}$$

For any given choice of d, the whole range of load admittances outside the circle $\bar{G}_0 = \csc^2 \beta d$ may be matched. As an example, for $d = \lambda/8$, $\beta d = \pi/4$, $\phi = 2\beta d = \pi/2$, and all values of load admittance outside the $\bar{G}_0 = 2$ circle may be matched, as shown in Fig. 5.10. For $d = \lambda/4$, all values of $\bar{Y}_L$ with $\bar{G}_L \le 1$ may be matched.

Although the theory predicts that virtually all load impedances may be matched by choosing d near zero, or $\lambda/2$, so that $\csc^2 \beta d$ becomes infinite, this is not true in practice. The maximum value of stub susceptance that can be obtained is limited by the finite attenuation of the transmission line used. If $j\beta$ were replaced by $j\beta + \alpha$, it would be found that, even with $\lambda/2$ spacing, all values of load admittance could not be matched. In addition, stub spacings near $\lambda/2$ lead to very frequency-sensitive matching networks, so that in practice spacings of $\lambda/8$ or $3\lambda/8$ are preferred. The larger spacing is used at the higher frequencies, where the wavelength is too small to permit use of $\lambda/8$ spacing.

A complete analytical solution to the double-stub matching network is readily obtained. Solution of (5.18*a*) for the susceptance $\bar{B}_1$ of the first stub gives

$$\bar{B}_1 = -\bar{B}_L + \frac{1 \pm \sqrt{(1 + t^2)\bar{G}_L - G_L^2 t^2}}{t} \tag{5.20}$$

where $\bar{B}_L$, $\bar{G}_L$, and $t = \tan \beta d$ are all known. Equating the imaginary part

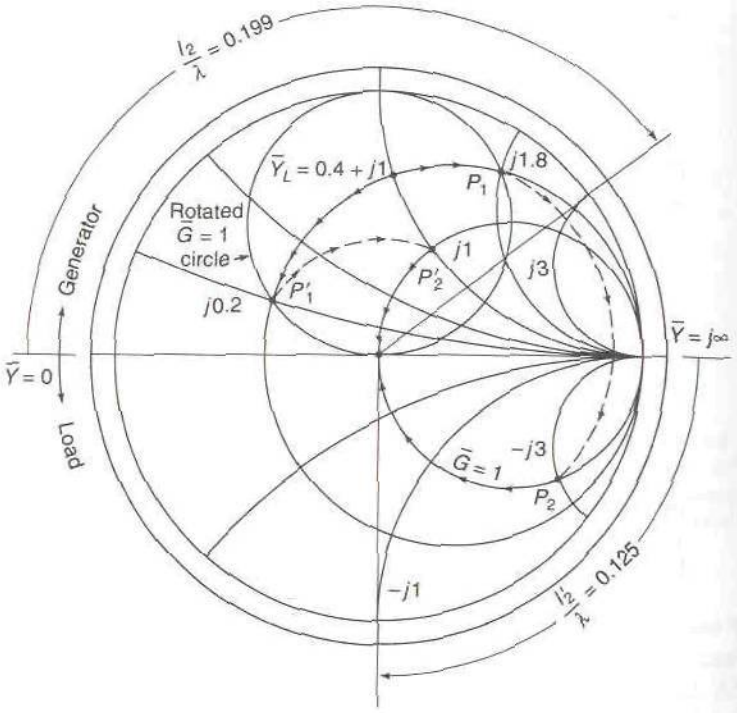


FIGURE 5.11
Illustration of the design of a double-stub matching network.

of (5.17) to $j\overline{B}$ gives

$$\overline{B} = \frac{\left(1 - \overline{B}_L t - \overline{B}_1 t\right)\left(\overline{B}_L + \overline{B}_1 + t\right) - \overline{G}_L^2 t}{\left(1 - \overline{B}_L t - \overline{B}_1 t\right)^2 + \overline{G}_L^2 t^2}$$

Substituting for $\overline{B}_1$ into this equation yields

$$\overline{B} = \frac{\mp\sqrt{\overline{G}_L(1 + t^2) - \overline{G}_L^2 t^2} - \overline{G}_L}{\overline{G}_L t} \tag{5.21}$$

The upper and lower signs in (5.20) and (5.21) go together. The susceptance of the second stub must be chosen as $-j\overline{B}$ in order to provide a matched condition.

Example 5.1. We want to design a double-stub-matching system to match a normalized load admittance $\overline{Y}_L = 0.4 + j1$. The stubs are spaced $\lambda/8$ apart.

We first construct the $\overline{G} = 1$ circle rotated by an angle $2\beta d = 4\pi\lambda/8 = \pi/2$ in the counterclockwise direction as shown in Fig. 5.11. Next we locate the point $\overline{Y}_L$ at the intersection of the $\overline{G} = 0.4$ and $\overline{B} = 1$ contours. This is the point labeled $\overline{Y}_L$ in Fig. 5.11. We can move $\overline{Y}_L$ so that it intersects the

rotated $\overline{G} = 1$ circle by moving it to P_1 or P_1' along the $\overline{G} = 0.4$ contour. At P_1, $j\overline{B} = j1.8$, so we need to add $j\overline{B}_1 = j(1.8 - 1) = j0.8$ to $\overline{Y}_L$ to get to P_1. At P_1', $j\overline{B} = j0.2$, so to move $\overline{Y}_L$ to P_1', we need a stub with susceptance $j\overline{B}_1 = j(0.2 - 1) = -j0.8$.

The next step is to rotate P_1 and P_1' by $\pi/2$ rad in the clockwise direction. The rotated points lie on the $\overline{G} = 1$ circle at P_2 and P_2'. At P_2, $j\overline{B} = -j3$, so we need a second stub with $j\overline{B}_2 = -j\overline{B} = j3$ in order to move P_2 into the origin along the $\overline{G} = 1$ contour. Similarly, in order to move P_2' into the origin, we need a second stub with $j\overline{B}_2 = -j\overline{B} = -j1$ since $j\overline{B}$ at P_2' equals $j1$.

We can also use the Smith chart to find the stub lengths. For the first solution where we needed a stub with $j\overline{B}_2 = j3$, we will assume that we use an open-circuited stub. We draw a radius vector from the origin through the point where the $j\overline{B} = j3$ contour cuts the outer boundary of the Smith chart as shown in Fig. 5.11. When we move along an open-circuited or short-circuited stub, we will be moving on the $\rho = 1$ circle or outer boundary of the Smith chart. At the open-circuit end of the stub, $\overline{Y} = 0$ and this is the point on the left-hand side of the Smith chart as shown. We begin at this point and move on the $\overline{G} = 0$ or $\rho = 1$ circle toward the generator (clockwise) until we get to the $j3$ point. From the normalized distance scale on the outer boundary of the Smith chart (not shown), we find that the required stub length is $l_2/\lambda = 0.199$.

For the second solution where we need $j\overline{B}_2 = -j1$, we will assume that we use a short-circuited stub. At the short-circuit position $\overline{Y} = j\infty$ and this is the point on the right-hand side of the chart. We begin at this point and move on the $\overline{G} = 0$ circle until we get to $-j1$ as shown. This gives $l_2'/\lambda = 0.125$ for the stub length.

The reader can verify that for stub 1 the required length is, for an open-circuited stub,

$$\frac{l_1}{\lambda} = 0.107 \qquad \frac{l_1'}{\lambda} = 0.393$$

while for a short-circuited stub,

$$\frac{l_1}{\lambda} = 0.143 \qquad \frac{l_1'}{\lambda} = 0.357$$

If series stubs are used, then we work with the normalized load impedance and the normalized reactances of the stubs. A point to keep in mind when designing the length of a stub to give a specified reactance is that at a short-circuit $\overline{Z} = 0$, while at an open-circuit $\overline{Z} = j\infty$. Thus, for an open-circuited stub, we begin at the right-hand side of the Smith chart and move clockwise on the $\rho = 1$ circle until we arrive at the desired $j\overline{X}$ point. For a shunt stub that was open-circuited, we start at the left-hand-side edge of the Smith chart.

5.4 TRIPLE-STUB TUNER

The disadvantage of not being able to match all load admittances with a double-stub tuner may be overcome by using a triple-stub tuner, as illustrated in Fig. 5.12. Stub 1 provides a susceptance $j\overline{B}_1$ such that $\overline{Y}_L + j\overline{B}_1$

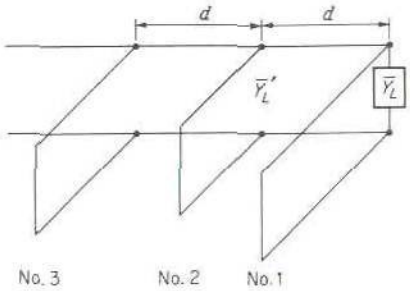


FIGURE 5.12
Triple-stub tuner.

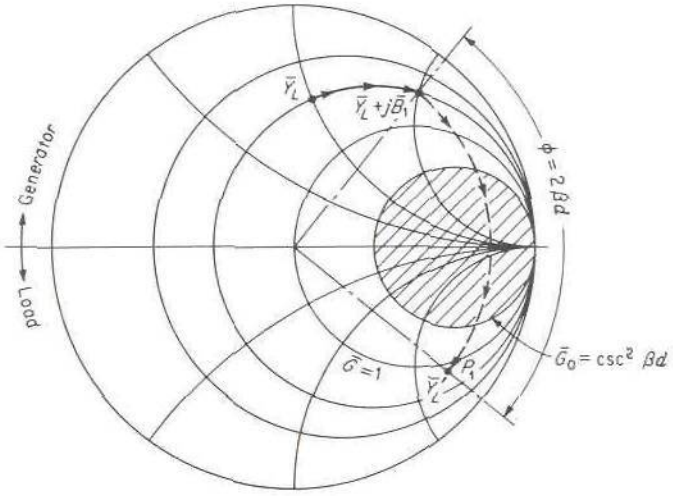


FIGURE 5.13
Transformation of Y_L into Y_L'.

transforms to some new admittance $\overline{Y}_L'$ just to the right of stub 2. Stubs 2 and 3 provide a conventional double-stub tuner for matching $\overline{Y}_L'$ to the line. These two stubs will match all values of $\overline{Y}_L'$ for which $\overline{G}_L' < \csc^2 \beta d$. Thus the function of stub 1 is to ensure that a susceptance $j\overline{B}_1$ is added to $\overline{Y}_L$ such that the transformed admittance $\overline{Y}_L'$ has a $\overline{G}_L'$ less than $\csc^2 \beta d$. To find a suitable value of $j\overline{B}_1$, we note that, after moving a distance d from $\overline{Y}_L$, the admittance $\overline{Y}_L + j\overline{B}_1$ must transform into a point P_1 outside the circle $\overline{G}_0 = \csc^2 \beta d$, as in Fig. 5.13. If we rotate the $\overline{G}_0 = \csc^2 \beta d$ circle through an angle $-\phi = -2\beta d$, we can readily see at once the range of susceptances $j\overline{B}_1$ that may be added to $\overline{Y}_L$ to keep the resulting $\overline{Y}_L'$ outside the $\overline{G}_0 = \csc^2 \beta d$ circle. The procedure is illustrated in Fig. 5.14. In this example $\overline{Y}_L$ falls within the rotated $\overline{G}_0 = \csc^2 \beta d$ circle, and hence a susceptance $j\overline{B}_1$ must be added to move the resultant load $\overline{Y}_L + j\overline{B}_1$ to some point beyond P_1 or P_2, say to P_1' or P_2', which provides for a margin of

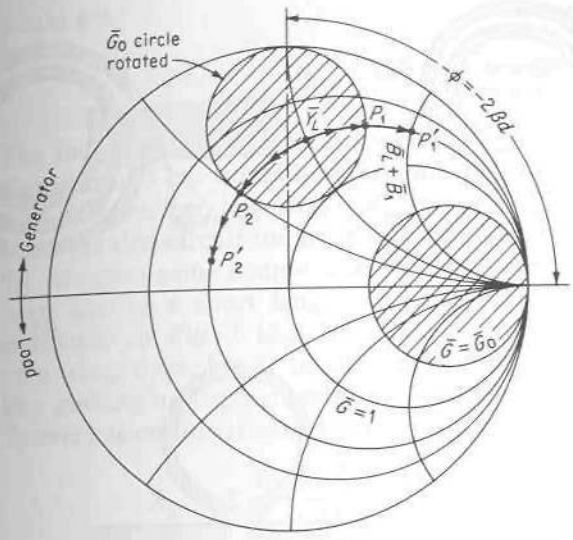


FIGURE 5.14
Graphical solution for jB_1 for a triple-stub tuner.

safety. The resultant load when transformed to the position of stub 2 will lie outside the circle $\bar{G}_0 = \csc^2 \beta d$, and hence can be matched by stubs 2 and 3.

A triple-stub tuner can match all values of load admittances. It may be considered to be two double-stub tuners in series, i.e.,

$$j\bar{B}_2 = j\bar{B}_2' + j\bar{B}_2''$$

so part of $j\bar{B}_2$ is associated with each end stub of the two double-stub tuners even though only one physical stub is present in the center. The triple-stub-matching system has more free variables so it can be optimized to increase the bandwidth.

5.5 IMPEDANCE MATCHING WITH LUMPED ELEMENTS

Lumped-parameter elements such as conventional inductors and capacitors are not compatible with coaxial transmission lines and waveguides and consequently are generally not used with these structures for impedance matching. However, the widespread use of microstrip transmission lines and the minaturization of inductors and capacitors have brought lumped-parameter elements into prominence for impedance-matching purposes. In a microstrip-line, series connections of capacitors and inductors are easily made. The shunt connection of a capacitor or inductor is somewhat more difficult since a connection through the substrate to the ground plane must be made. In the frequency range up to several gigahertz, the use of lumped capacitors and inductors for impedance-matching results in a more compact circuit.

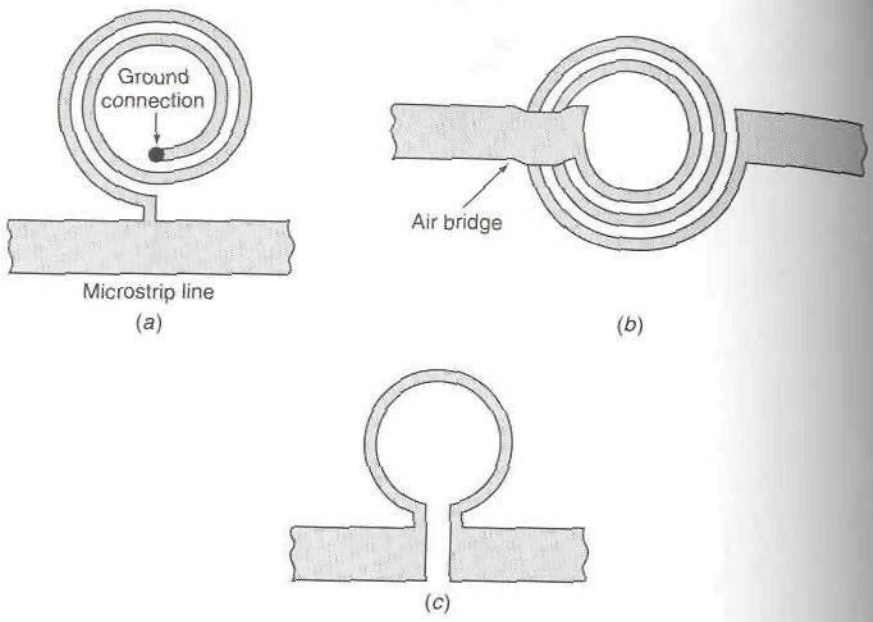


FIGURE 5.15
(*a*) A spiral inductor connected in shunt across a microstrip line; (*b*) a series-connected spiral inductor; (*c*) a series-connected single loop inductor.

For a microstrip circuit the most common form of inductor is the spiral inductor shown in Fig. 5.15*a*. For a shunt connection the center of the spiral can be connected to the ground plane. For a series connection an air bridge has to be used as shown in Fig. 5.15*b*. If only a small inductive reactance is required, a one-turn loop as shown in Fig. 5.15*c* may be used. If a spiral inductor is to behave as a lumped inductor, its total length must be a small fraction of a wavelength. An estimate of the inductive reactance can be obtained by considering the inductor to be a length l of a transmission line. By using the formulas for the characteristic impedance and propagation constant, namely,

$$Z_c = \sqrt{\frac{L}{C}} \qquad \beta = \omega\sqrt{LC}$$

we obtain

$$\omega L = \beta Z_c = k_0 Z_{c0}\ \Omega/\text{m}$$

where Z_{c0} is the characteristic impedance with air as the dielectric. As can be seen from this expression, a high-impedance line (narrow width) should be used in an inductor. As an example, if we have a 100-Ω line on a 1-mm-thick substrate with an effective dielectric constant of 4 at 2 GHz, we

would get

$$\omega L = \frac{2\pi}{\lambda_0}\sqrt{\epsilon_e} \times 100 = \frac{2\pi}{15} \times 200 \approx 83.8\ \Omega/\text{cm}$$

The inductance L is 6.67 nH/cm. A three-quarter-turn loop with a mean diameter of 1 cm would have an inductance of about 16 nH. Unfortunately, a single-turn loop this long would not function as an ideal inductor because every printed-circuit inductor has distributed capacitance associated with it. We can gain some insight into the length restriction by treating the single-turn loop as a short length of a high-impedance transmission line. With reference to Fig. 5.15*c*, let Z_0 be the characteristic impedance of the microstrip lines, let Z_c be the characteristic impedance of the transmission line making up the inductor, and let βl be the electrical length of the loop. Transmission-line theory gives

$$Z_{\text{in}} = Z_c \frac{Z_0 + jZ_c t}{Z_c + jZ_0 t} = \frac{Z_0(Z_c + jZ_0 t) + jt(Z_c^2 - Z_0^2)}{Z_c + jZ_0 t}$$

$$= Z_0 + \frac{jt(Z_c^2 - Z_0^2)(Z_c - jZ_0 t)}{Z_c^2 + Z_0^2 t^2}$$

where $t = \tan \beta l$.

Since we normally would use a large value of Z_c and keep βl small, we will assume that $Z_0^2 t^2 \ll Z_c^2$ in which case we get

$$Z_{\text{in}} \approx Z_0 + Z_0 t^2\left(1 - \frac{Z_0^2}{Z_c^2}\right) + jZ_c t\left(1 - \frac{Z_0^2}{Z_c^2}\right)$$

For a short-circuited line,

$$jX_L \approx jZ_c \tan \beta l \approx jZ_c \beta l$$

Hence

$$Z_{\text{in}} \approx Z_0 + Z_0 t^2\left(1 - \frac{Z_0^2}{Z_c^2}\right) + jX_L\left(1 - \frac{Z_0^2}{Z_c^2}\right)$$

Thus the effective inductance is reduced by the factor $1 - Z_0^2/Z_c^2$. A more serious departure from an ideal inductor behavior is the change in the real part of Z_{in}. For an ideal series-connected inductor, we would have $Z_{in} = Z_0 + jX_L$. In order that the actual Z_{in} should approximate this ideal result, we need to make the factor $t^2(1 - Z_0^2/Z_c^2)$ of the order of 0.05. In a typical situation where $Z_0 = 50\ \Omega$ and $Z_c = 100\ \Omega$, this requires that βl be no larger than 0.26. The corresponding maximum value for the length l is 0.04λ. For the earlier example where $\epsilon_e = 4$ and $f = 2$ GHz, we have $\lambda = 15/2 = 7.5$ cm. Thus l should be no greater than 0.3 cm. This leads to a maximum usable inductance of only 2 nH. We can, of course, use a larger value for l but the change in the resistive part of Z_{in} away from the ideal

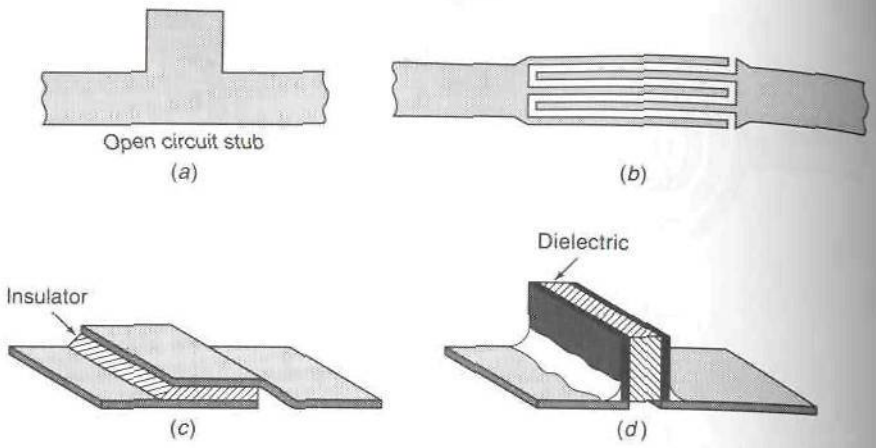


FIGURE 5.16
(*a*) A short open-circuited stub; (*b*) an interdigital capacitor; (*c*) a metal-insulator-metal (MIM) capacitor; (*d*) a chip capacitor soldered across a microstrip-line gap.

value Z_0 must then be taken into account in the design of a circuit requiring a series inductance.

The spiral inductor can provide larger values of inductance. For example, a five-turn spiral inductor approximately 1.4 mm in diameter, with a conductor width of 0.06 mm and spacing 0.038 mm, has an inductance of 25 nH at 2 GHz.†

A short open-circuited stub as shown in Fig. 5.16*a* will function as a lumped capacitor connected in shunt across a microstrip line. The distributed capacitance of a microstrip line is typically in the range $0.2\sqrt{\epsilon_e}$ to $\sqrt{\epsilon_e}$ pF/cm; so a short stub is suitable for providing a shunt capacitance up to about 1 pF. At 4 GHz a 1-pF capacitor has a reactance of about 40 Ω. The interdigital capacitor shown in Fig. 5.16*b* can provide a series capacitance up to several picofarads depending on the number of fingers used and their length. For monolithic microwave integrated circuits, the metal-insulator-metal (MIM) capacitor shown in Fig. 5.16*c* is generally used. A capacitance up to 20 pF or more can be obtained since the insulator thickness can be very small. For example, an insulator 1 mm by 1 mm and 10 μm thick and having a dielectric constant of 10 provides a capacitor with a capacitance of about 9 pF. For hybrid microstrip circuits, the chip capacitor illustrated in Fig. 5.16*d* is used. It is soldered in place and can provide a capacitance up to 100 pF or more.

†D. A. Daly, S. P. Knight, M. Caulton, and R. Ekholdt, Lumped Elements in Microwave Integrated Circuits, *IEEE Trans.*, vol. MTT-15, pp. 713–721, December, 1967.

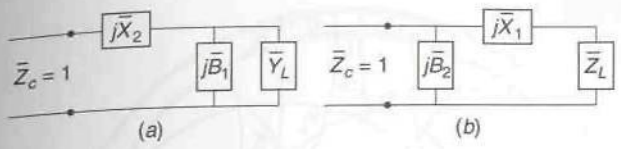


FIGURE 5.17
Two basic lumped-element matching circuits. (*a*) Circuit used when $\bar{G}_L \leq 1$; (*b*) circuit used when $\bar{R}_L \leq 1$.

The basic lumped-element impedance-matching circuit is a circuit consisting of a parallel-connected and a series-connected reactive element as shown in Fig. 5.17. The topology of the circuit is such that it is commonly referred to as an L matching network (compare with the designation of T and Π for the impedance and admittance networks for a two-port network). For the circuit shown in Fig. 5.17*a*, the shunt element jB_1 is connected in parallel with the load and the series element jX_2 is connected in series. This configuration can be used to match any load admittance Y_L having a normalized conductance $\bar{G}_L \leq 1$. The circuit shown in Fig. 5.17*b* can be used to match any load impedance having a normalized load resistance $\bar{R}_L$ less than 1. For some load impedances both $\bar{G}_L$ and $\bar{R}_L$ are less than 1 and either circuit may be used. When the normalized admittance lies inside the $\bar{G} = 1$ circle on the Smith chart, the corresponding normalized load impedance, which is the reflection of the normalized load admittance through the origin, will lie outside the $\bar{R} = 1$ circle and hence can be matched using the circuit shown in Fig. 5.17*b*. When $\bar{Z}_L$ lies inside the $\bar{R} = 1$ circle, $\bar{Y}_L$ will lie outside the $\bar{G} = 1$ circle and can be matched using the circuit shown in Fig. 5.17*a*. When both $\bar{Z}_L$ and $\bar{Y}_L$ lie outside the $\bar{R} = 1$ and $\bar{G} = 1$ circles, respectively, either matching circuit can be used.

The required values of the matching elements are easily found using the Smith chart. The procedure to be followed is described below. From this the reader will easily understand the rationale that underlies the use of the Smith chart to solve the matching problem using lumped reactance elements. Since the Smith chart procedure uses normalized immittance parameters, the first preliminary step is to determine the normalized values of Y_L and Z_L by dividing by the characteristic admittance Y_c or characteristic impedance Z_c, respectively, of the input transmission line.†

Case 1. The circuit in Fig. 5.17*a* is used. A match can be obtained only if $\bar{G}_L \leq 1$. If $\bar{G}_L > 1$ use the circuit in Fig. 5.17*b* (see Case 2).
With reference to Fig. 5.18.

1. Construct the $\bar{G} = 1$ circle rotated by 180°.
2. From the point $\bar{Y}_L$ add $j\bar{B}_1$ to move along a constant-conductance circle until the rotated $\bar{G} = 1$ circle is intersected. There are two possible solutions.

†The term immittance is used to designate either an impedance or admittance.

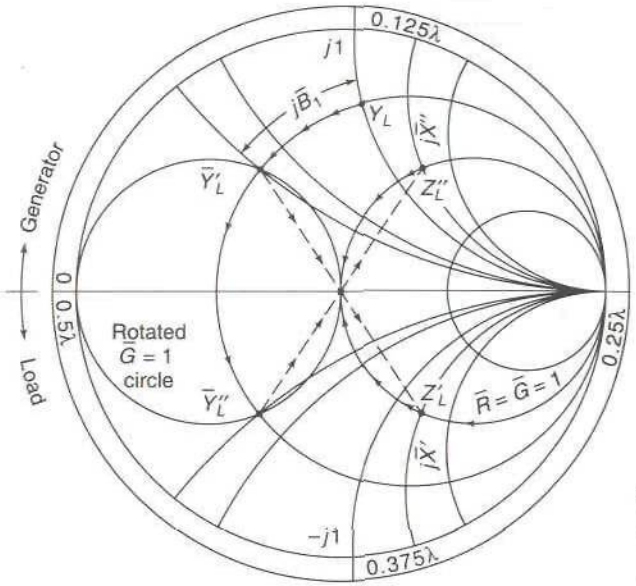


FIGURE 5.18
Illustration of steps followed in designing the matching circuit in Fig. 5.17*a* (Case 1).

The new values of $\bar{Y}_L$ are $\bar{Y}'_L$ and $\bar{Y}''_L$. Note that if $\bar{Y}_L$ is inside the $\bar{G} = 1$ circle, it cannot be moved to intersect the rotated $\bar{G} = 1$ circle because adding $j\bar{B}$ only moves $\bar{Y}_L$ on a constant-conductance circle.

3. Reflect $\bar{Y}'_L$ and $\bar{Y}''_L$ through the origin to get the corresponding impedance values $\bar{Z}'_L$ and $\bar{Z}''_L$. These lie on the $\bar{R} = 1$ circle since we made $\bar{Y}'_L$ and $\bar{Y}''_L$ lie on the rotated $\bar{G} = 1$ circle.
4. Since $\bar{Z}'_L$ and $\bar{Z}''_L$ lie on the $\bar{R} = 1$ circle, these impedance points can be moved into the origin by *subtracting* a reactance $j\bar{X}'$ or $j\bar{X}''$. The origin represents a matched condition. Hence the required value of $j\bar{X}_2$ is either $-j\bar{X}'$ or $-j\bar{X}''$. The required value of $j\bar{B}_1$ is $j(\bar{B}' - \bar{B}_L)$ or $j(\bar{B}'' - \bar{B}_L)$.

The greatest bandwidth is obtained when the reactive elements are as small as possible so that the circuit Q is as low as possible. We will return to this point after we provide the steps to be followed using the other matching circuit.

Case 2. The circuit in Fig. 5.17*b* is used. A match can be obtained only if $\bar{R}_L \leq 1$. If $\bar{R}_L > 1$ the circuit in Fig. 5.17*a* must be used (Case 1).

1. Construct the $\bar{R} = 1$ circle rotated by 180° as shown in Fig. 5.19.
2. Add $j\bar{X}_1$ to $\bar{Z}_L$ to move $\bar{Z}_L$ to intersect with the rotated $\bar{R} = 1$ circle. The motion is along a constant resistance circle. There are two solutions $\bar{Z}'_L$ and $\bar{Z}''_L$.
3. Reflect $\bar{Z}'_L$ and $\bar{Z}''_L$ through the origin to obtain $\bar{Y}'_L$ and $\bar{Y}''_L$.

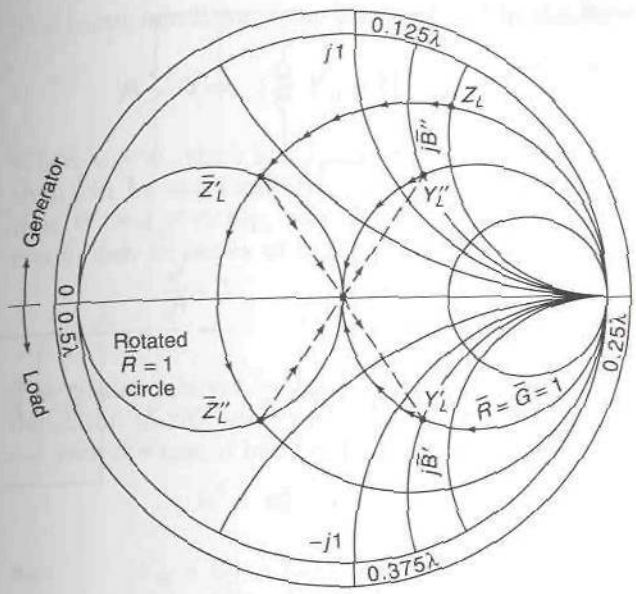


FIGURE 5.19
Illustration of steps followed in designing the matching circuit in Fig. 5.17*b* (Case 2).

4. The admittance points $\bar{Y}_L'$ and $\bar{Y}_L''$ can be moved into the origin along the $\bar{G} = 1$ circle by subtracting $j\bar{B}'$ and jB'', respectively. Hence the required value of $j\bar{B}_2$ is $-j\bar{B}'$ or $-j\bar{B}''$. The required value of $j\bar{X}_1$ is $j(\bar{X}' - \bar{X}_L)$ or $j(\bar{X}'' - \bar{X}_L)$.

Analytic solutions for the required values of the matching elements are readily derived. The analytic solutions are given in Probs. 5.17 and 5.18 but the reader has to supply the derivations.

Circuit Q and Bandwidth

When a complex load impedance has been matched to a transmission line with characteristic impedance Z_c, the input impedance looking toward the load equals Z_c. Thus the reactive elements present in Z_L and the matching network make up a resonant circuit that is loaded by R_L and Z_c. A resonant circuit has a quality factor, or Q, that can be evaluated from the general definition

$$Q = \frac{\omega(\text{average stored electric and magnetic energy})}{\text{power loss}} \tag{5.22}$$

At resonance the average stored electric energy in the capacitors equals the average stored magnetic energy in the inductors. Hence Q can be expressed as

$$Q = \frac{2\omega W_e}{P_L} = \frac{2\omega W_m}{P_L} \tag{5.23}$$

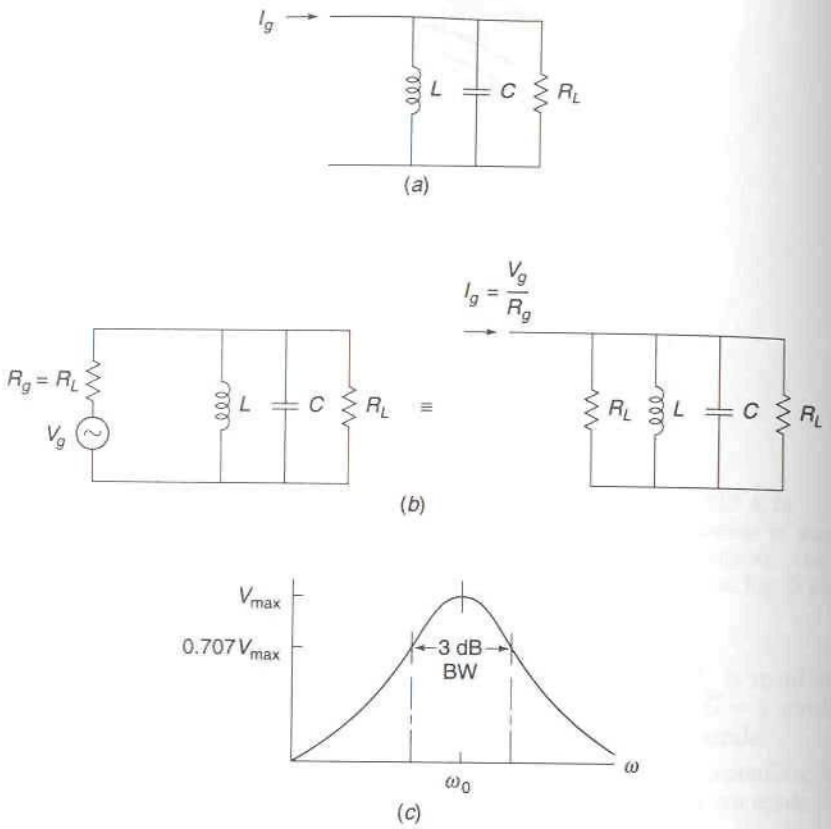


FIGURE 5.20
(*a*) Parallel *LCR* resonant circuit; (*b*) loaded resonant circuit; (*c*) frequency-response curve showing half-power bandwidth.

The bandwidth of the circuit is the frequency band over which one-half or more of the maximum power is delivered to the load. This bandwidth is called the half-power or 3-dB bandwidth and is inversely proportional to the loaded Q of the circuit.

In order to clarify the above concepts, we will analyze the parallel resonant circuit shown in Fig. 5.20. For the circuit in Fig. 5.20*a*, the voltage across R_L is given by

$$V = \frac{I_g}{Y_{\text{in}}} = \frac{I_g}{G_L + j\omega C - j/\omega L} \tag{5.24}$$

where $G_L = 1/R_L$. The resonant frequency of the circuit is given by

$$\omega_0 = \frac{1}{\sqrt{LC}} \tag{5.25}$$

The input admittance can be expressed in the form

$$Y_{\text{in}} = G_L + j\omega C\left(\frac{\omega^2 - \omega_0^2}{\omega^2}\right)$$

When $\omega = \omega_0$ then $V = I_g/G_L = I_g R_L$, which is the maximum load voltage that can be obtained. When $j\omega C(\omega^2 - \omega_0^2)/\omega^2 = jG_L$, we see from (5.24) that $|V| = I_g/\sqrt{2}\,G_L$, and hence the power in R_L is only one-half that at resonance. In terms of the circuit elements, the Q is given by

$$Q = \frac{R_L}{\omega_0 L} = R_L \omega_0 C = \frac{\omega_0 C}{G_L} \tag{5.26}$$

The reader can verify that this is the same Q as would be found using the definition (5.23) (see Prob. 5.19). When $Q \geq 10$ the frequency ω is close to ω_0 over the useful band of operation; so we have

$$\omega^2 - \omega_0^2 = (\omega - \omega_0)(\omega + \omega_0) \approx 2\omega(\omega - \omega_0)$$

and

$$Y_{\text{in}} \approx G_L + j\omega C\frac{2\omega(\omega - \omega_0)}{\omega^2} = G_L + j\omega_0 C\frac{2(\omega - \omega_0)}{\omega_0}$$

$$= G_L\left(1 + 2jQ\frac{\Delta\omega}{\omega_0}\right) \tag{5.27}$$

where $\Delta\omega = \omega - \omega_0$. From this expression it is readily seen that the 3-dB fractional bandwidth is given by

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

or $\text{BW} = 1/Q$ since the half-power points occur when $2jQ\,\Delta\omega/\omega_0 = \pm j$.

When the circuit is connected to a matched source as shown in Fig. 5.20*b*, it is clear that

$$V = \frac{I_g}{2G_L(1 + 2jQ_L\,\Delta\omega/\omega_0)} \tag{5.28}$$

where $Q_L = Q/2$ is the loaded Q of the circuit. Consequently, for the loaded circuit, the bandwidth is twice as great. A typical response curve showing the bandwidth is given in Fig. 5.20*c*.

In an impedance-matching problem, there are generally two solutions available. If we want a narrowband design, we should choose the solution that gives the largest loaded Q. On the other hand, when we want a broadband match, we should choose the circuit with the lowest loaded Q. The load impedance plus matching network will contain either one capacitor and two inductors or one inductor and two capacitors. Thus, when using (5.23) to evaluate the circuit Q, the energy stored in the single reactive element or that stored in the two opposite reactive elements must be

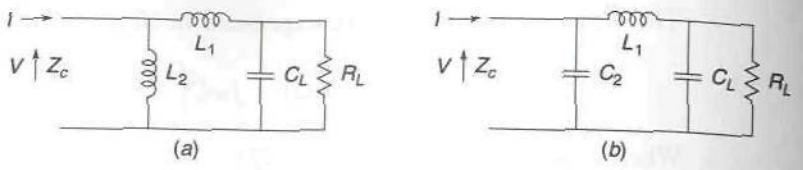


FIGURE 5.21
Two matching networks.

evaluated. We will illustrate the evaluation of circuit Q by means of two examples that show the steps involved. By making use of the facts that at resonance $Z_{in} = Z_c$ and that $P_{in} = P_L$, the evaluation of the circuit Q can be reduced to a few simple steps.

For the circuit shown in Fig. 5.21*a*, we will base the evaluation of the circuit Q on the energy stored in the capacitor C_L. The power dissipated in R_L is given by

$$P_L = \frac{1}{2}\frac{|V_L|^2}{R_L}$$

The average electric energy stored in C_L is

$$W_e = \tfrac{1}{4}|V_L|^2 C_L$$

Hence from (5.23) we get

$$Q = \frac{2\omega_0 W_e}{P_L} = \omega_0 C_L R_L$$

The loaded Q is $Q_L = Q/2$ since the system is matched at ω_0. The operating bandwidth is determined by the loaded Q.

For the circuit in Fig. 5.21*b*, we note that

$$P_{in} = \frac{1}{2}\frac{|V|^2}{Z_c} = P_L = \frac{1}{2}\frac{|V_L|^2}{R_L}$$

since $Z_{in} = Z_c$ at $\omega = \omega_0$. The energy stored in C_2 is

$$W_{e2} = \tfrac{1}{4}|V|^2 C_2$$

while that stored in C_L is

$$W_{eL} = \frac{1}{4}|V_L|^2 C_L = \frac{1}{4} C_L \frac{R_L}{Z_c}|V|^2$$

where we used the equality of P_{in} and P_L to express $|V_L|$ in terms of $|V|$.

The circuit Q is given by

$$Q = \frac{2\omega_0(W_{e2} + W_{eL})}{P_L} = \frac{\omega_0(C_2 + (R_L/Z_c)C_L)|V|^2}{|V_L|^2/R_L}$$

$$= \omega_0 C_2 Z_c + \omega_0 C_L R_L$$

This circuit Q is larger than that for the circuit in Fig. 5.21*a*. In general, the smallest circuit Q is obtained when the matching circuit contains two similar reactive elements that are opposite to the reactive element in the load; i.e., if the load contains an inductive element, the matching circuit should be made up of capacitive elements only, and vice versa.

For circuits with a loaded Q of 5 or more, the frequency response is very nearly the same as that of the parallel resonant circuit shown in Fig. 5.20, over the useful operating frequency band. Thus the 3-dB fractional bandwidth is $1/Q_L$. For circuits with a low Q, the frequency behavior is different but similar.

The discussion above has been based on the assumption that ideal lossless inductors and capacitors are used. As a general rule, lumped capacitors have very little loss and the assumption of negligible loss does not introduce significant error. However, lumped inductors do not have negligible loss, and in high-Q circuits inductor loss should be taken into account. The Q of an inductor is the ratio of the reactance ωL to the series resistance R of the inductor. Lumped inductors may have Q values in the range of 25 to several hundred. If the loaded circuit Q is less than about one-fifth of the inductor Q, then neglecting the loss in the inductor will not produce serious error. When several inductors are used in a matching circuit, the losses in the inductors should be taken into account since these losses can account for a significant fraction of the total power delivered to a circuit.

For microwave circuits the useful bandwidth is often much smaller than the 3-dB bandwidth. It is not uncommon to require an impedance match providing a VSWR of no more than 2, or even less than 1.5 in critical applications. A VSWR of 2 corresponds to a reflection coefficient of 0.333. With this mismatch a fraction of 0.889 of the incident power is delivered to the load.

The degree of mismatch is usually described in terms of the input VSWR or in terms of the return loss. The return loss is the ratio of the reflected power to the incident power, expressed in decibels; thus

$$\text{Return loss} = \text{RL} = -20 \log \rho = -20 \log \frac{\text{VSWR} - 1}{\text{VSWR} + 1}$$

A well-matched system will have a return loss of 15 dB or more. A VSWR equal to 2 gives a return loss of 9.54 dB, while a VSWR of 1.5 gives a return loss of 13.98 dB. Thus a return loss of 15 dB, corresponding to a VSWR of 1.43, does indeed represent a well-matched system.

In order to get a feeling for how narrow the bandwidth of a system is using this criterion, let us assume that the loaded Q equals 5. When the VSWR = 1.43, we have $\rho = 0.178$. We can use

$$Y_{\text{in}} \approx Y_c\left(1 + 2jQ\frac{\Delta\omega}{\omega_0}\right)$$

$$\rho = \left|\frac{Y_c - Y_{\text{in}}}{Y_c + Y_{\text{in}}}\right| = \frac{2Q\,\Delta\omega/\omega_0}{\left(4 + 4Q^2\,\Delta\omega^2/\omega_0^2\right)^{1/2}} \approx Q\frac{\Delta\omega}{\omega_0}$$

where $Q = 2Q_L$. Hence the fractional bandwidth is

$$2\frac{\Delta\omega}{\omega_0} = \frac{2\rho}{Q} = \frac{\rho}{Q_L} = 0.0356$$

This is much smaller than the 3-dB bandwidth which equals $1/Q_L$ or 0.2.

5.6 DESIGN OF COMPLEX IMPEDANCE TERMINATIONS

In microwave amplifier design, we often require a complex load impedance for the output and a complex source impedance for the input even though the final input and output connections are to 50-Ω lines. In a typical amplifier circuit as shown in Fig. 5.22*a*, the function of the input and output matching networks is to transform the 50-Ω line impedances into the required complex source impedance Z_s and load impedance Z_L. In a broadband amplifier design, the design of the matching networks can be quite complex. In narrowband amplifier design, elementary matching networks can be used.

The matching problem for an amplifier is the reverse of that for matching a complex load impedance Z_L to a transmission line. The first step is to choose a network topology to be used. In Figs. 5.22*b* to *d* we show three possible networks that could be used. Many other choices can be made. We will illustrate the use of the Smith chart to design matching networks of the type shown in Fig. 5.22. The methods used can be extended to more complex networks.

Design procedure for the circuit in Fig. 5.22*b*. The matching network in Fig. 5.22*b* consists of a transmission-line stub placed at a distance l from the input. The stub may be either an open-circuited or a short-circuited stub or a lumped susceptance since its function is to provide a normalized susceptance $j\bar{B}_1$ connected across the transmission line at a distance l from the input. The following procedure should be used to determine l and $\bar{B}_1$:

1. Locate the required value of $\bar{Z}_L$ on the Smith chart and find $\bar{Y}_L$ by reflecting $\bar{Z}_L$ through the origin. This procedure is illustrated in Fig. 5.23 for the case where $\bar{Z}_L = 0.4 - j0.2$.

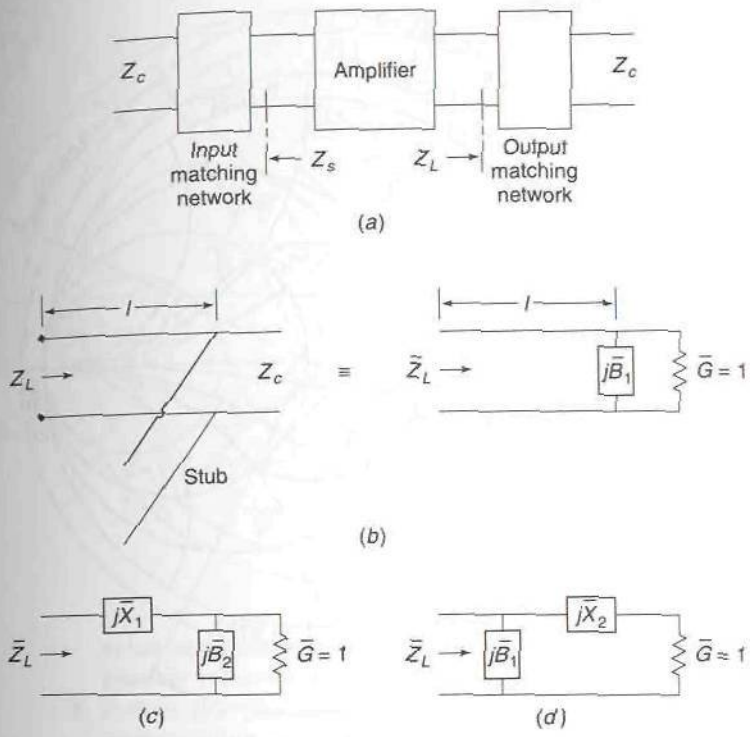


FIGURE 5.22
(*a*) Microwave amplifier circuit; (*b*) transmission-line matching network; (*c*) and (*d*) alternative matching networks.

2. Rotate the point $\bar{Y}_L$ on a constant-radius circle until it intersects the $\bar{G} = 1$ circle. There are two possible solutions. From the angle of rotation the normalized length l/λ is found. Note that the rotation is in a *counterclockwise direction since we are moving toward the load.*
3. When we move to the right of the point where $j\bar{B}_1$ is connected, $\bar{Y}_{in}$ *decreases* by $j\bar{B}_1$ to become $\bar{Y}_{in} = 1 + j\bar{B} - j\bar{B}_1$. But to the right of $j\bar{B}_1$, we require $\bar{Y}_{in} = \bar{G} = 1$ so $\bar{B}_1 = \bar{B}$. Thus the values of $j\bar{B}$ on the $\bar{G} = 1$ circle, where the rotated $\bar{Y}_L$ intersected the circle, gives the required values of $j\bar{B}_1$ for the two possible solutions.

Note that if we begin at $\bar{Y}_{in} = 1$ *and move left across* $j\bar{B}_1$, *then* $\bar{Y}_{in}$ changes to $1 + j\bar{B}_1 = 1 + j\bar{B}$. A rotation in the clockwise direction through an angle $2\beta l$ then makes $\bar{Y}_{in} = \bar{Y}_L$. Thus using the reverse of the usual matching procedure results in the desired value of $\bar{Y}_L$.

For our specific example, $l = 0.0626\lambda$ or 0.4375λ and the corresponding values required for $j\bar{B}_1$ are $j1$ or $-j1$, respectively, as shown in Fig. 5.23.

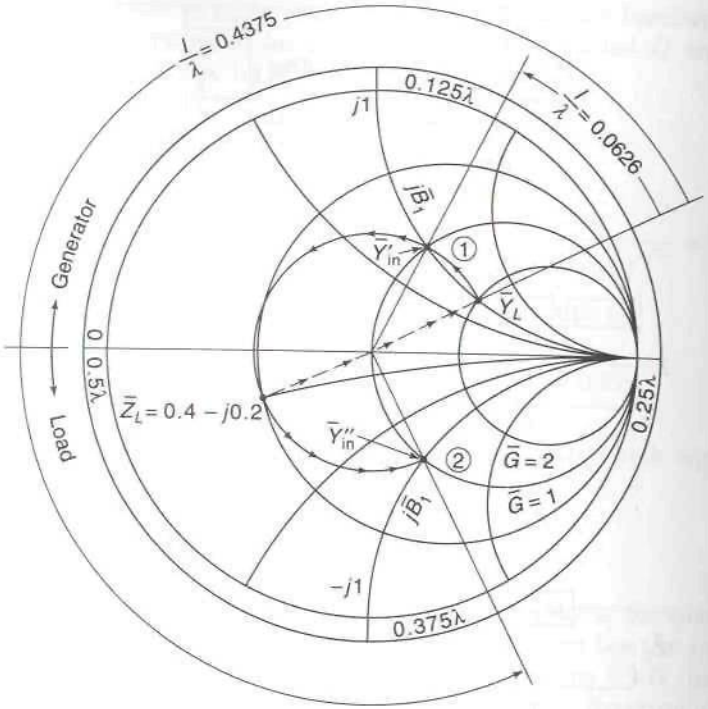


FIGURE 5.23
Illustration of design procedure for the matching network in Fig. 5.22*b*.

The analytic solution for the above matching network is

$$\bar{B}_1 = \pm\left[\frac{\bar{B}_L^2 + \bar{G}_L^2 + 1 - 2\bar{G}_L}{\bar{G}_L}\right]^{1/2} = \pm\left[\frac{\bar{R}_L^2 + \bar{X}_L^2 + 1 - 2\bar{R}_L}{\bar{R}_L}\right]^{1/2} \tag{5.29a}$$

$$\tan \beta l = \frac{\bar{G}_L - 1}{\bar{B}_L + \bar{B}_1\bar{G}_L} = \frac{\bar{R}_L^2 + \bar{X}_L^2 - \bar{R}_L}{\bar{X}_L - \bar{B}_1\bar{R}_L} \tag{5.29b}$$

Design procedure for the circuit in Fig. 5.22*c*. The circuit shown in Fig. 5.22*c* can be used only if $\bar{R}_L < 1$. The following steps should be followed to design this circuit using a Smith chart:

1. Construct the $\bar{R} = 1$ circle rotated by 180°.
2. Locate the desired impedance $\bar{Z}_L$ on the Smith chart.
3. Move $\bar{Z}_L$ in a counterclockwise direction (toward the load) on the $\bar{R}_L =$ constant circle until it intersects the rotated $\bar{R} = 1$ circle as shown in Fig. 5.24. There are two possible solutions. The required value for the reactance $j\bar{X}_1$ is given by $j\bar{X}_1 = j\bar{X}_L - j\bar{X}$, where $j\bar{X}$ equals $j\bar{X}'$ or $j\bar{X}''$ and is the

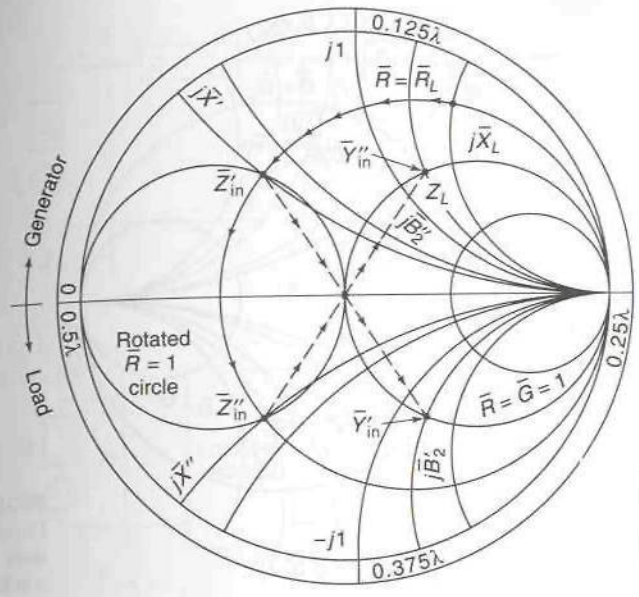


FIGURE 5.24
Illustration of design procedure for the matching network in Fig. 5.22c.

value obtained where $\bar{Z}_L$ intersects the rotated $\bar{R} = 1$ circle. The two possible values of $j\bar{X}$ are always equal in magnitude but opposite in sign.

4. Reflect the two new values of $\bar{Z}_{in}$ through the origin to obtain the corresponding values for $\bar{Y}_{in}$. This gives $\bar{Y}_{in} = 1 + j\bar{B}$. As we move to the right of the element $j\bar{B}_2$, we obtain $\bar{Y}_{in} - j\bar{B}_2$, which must be equal to 1. Hence the required value of $j\bar{B}_2$ is equal to $j\bar{B}$.

The reactance $j\bar{X}_1$ and susceptance $j\bar{B}_2$ can be realized using lumped elements or transmission-line stubs.

Design procedure for the circuit in Fig. 5.22*d*. The matching network shown in Fig. 5.22*d* can be used only when $\bar{G}_L < 1$ (can you explain why?). The network can be designed using the Smith chart by following the steps given below.

1. Construct the $\bar{G} = 1$ circle rotated by 180°.
2. Locate the desired load admittance point $\bar{Y}_L$ (reflection of $\bar{Z}_L$ through the origin).
3. Move $\bar{Y}_L$ on the $\bar{G} = \bar{G}_L$ circle in a counterclockwise direction until it intersects the rotated $\bar{G} = 1$ circle. There are two possible intersections. The required value of $j\bar{B}_1$ is given by $j\bar{B}_1 = j(\bar{B}_L - \bar{B})$, where $j\bar{B}$ equals $j\bar{B}'$ or $j\bar{B}''$, as shown in Fig. 5.25.
4. Reflect the admittances $\bar{Y}'_{in}$ and $\bar{Y}''_{in}$ through the origin to get the corresponding impedance points $\bar{Z}'_{in}$ and $\bar{Z}''_{in}$ that lie on the $\bar{R} = 1$ circle.
5. When we subtract $j\bar{X}_2$ we must get $\bar{Z}_{in} = 1$, so $j\bar{X}_2 = j\bar{X}$, where $j\bar{X}$ equals $j\bar{X}'$ or $j\bar{X}''$.

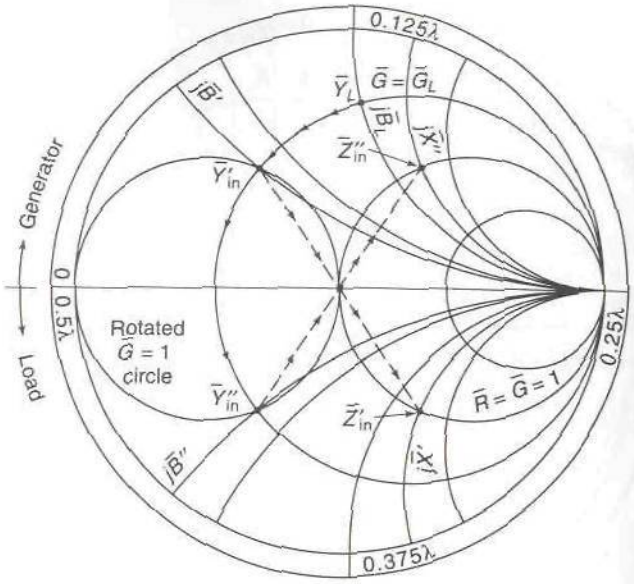


FIGURE 5.25
Illustration of design procedure for the matching network in Fig. 5.22*d*.

5.7 INVARIANT PROPERTY OF IMPEDANCE MISMATCH FACTOR

Consider the network shown in Fig. 5.26*a*. The T network between the load impedance Z_L and the source will be assumed to be lossless. The impedance looking into the network toward the load is Z_{in}. For this network the input current is

$$I = \frac{V_s}{Z_s + Z_{in}}$$

and the input power is

$$P_{in} = \frac{1}{2}\left|\frac{V_s}{Z_s + Z_{in}}\right|^2 R_{in} = \frac{1}{2}|I|^2 R_{in}$$

If $Z_{in} = Z_s^*$ then $R_{in} = R_s$, $X_{in} = -X_s$, and we get maximum power transfer from the source to the load resistance R_L. This power is

$$P = \frac{1}{2}\frac{|V_s|^2 R_{in}}{|R_{in} + R_s|^2} = \frac{1}{2}\frac{|V_s|^2}{4R_s} = P_{ava} \tag{5.30}$$

and is called the available power from the source. When $Z_{in} \neq Z_s^*$ we have an impedance mismatch and $P_{in} < P_{ava}$. We can express P_{in} in the form

$$P_{in} = \frac{1}{2}\frac{|V_s|^2 R_{in}}{|Z_s + Z_{in}|^2} = \frac{1}{2}\frac{|V_s|^2}{4R_s}\left[\frac{4R_s R_{in}}{|Z_{in} + Z_s|^2}\right] = P_{ava} M \tag{5.31}$$

where $M = 4R_{in}R_s/|Z_{in} + Z_s|^2$ is called the impedance mismatch factor.

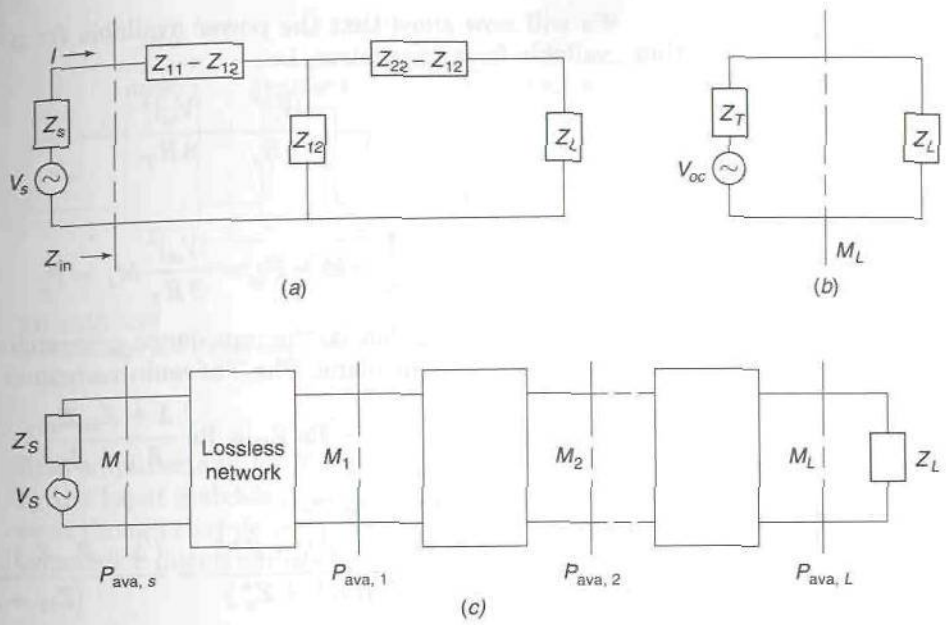


FIGURE 5.26
(*a*) A T matching network; (*b*) Thévenin equivalent network; (*c*) a cascade connection of lossless networks for which $P_{ava,\,s} = P_{ava,\,1} = P_{ava,\,2} = \cdots P_{ava,\,L}$ and $M = M_1 = M_2 = \cdots = M_L$.

We can use the Thévenin equivalent circuit shown in Fig. 5.26*b* to calculate the power delivered to R_L, which is P_L and must equal P_{in} since the network with elements Z_{11}, Z_{12}, Z_{22} is lossless.

From circuit theory we find that

$$V_{oc} = \frac{Z_{12}V_s}{Z_{11} + Z_s} \qquad Z_T = \frac{\Delta + Z_{22}Z_s}{Z_{11} + Z_s}$$

where $\Delta = Z_{11}Z_{22} - Z_{12}^2 = X_{12}^2 - X_{11}X_{22}$, upon using $Z_{11} = jX_{11}$, $Z_{22} = jX_{22}$, $Z_{12} = jX_{12}$. The available power from the network is

$$\frac{1}{2}\frac{|V_{oc}|^2}{4R_T} \qquad \text{where } Z_T = R_T + jX_T$$

The power delivered to R_L is

$$P_L = \frac{1}{2}\frac{|V_{oc}|^2}{4R_T}\frac{4R_TR_L}{|Z_T + Z_L|^2} = \frac{1}{2}\frac{|V_{oc}|^2}{4R_T}M_L = P_{in}$$

We will now show that the power available from the network equals that available from the source, i.e.,

$$\frac{|V_s|^2}{8R_s} = \frac{|V_{oc}|^2}{8R_T}$$

We must then have

$$\frac{|V_s|^2}{8R_s}M = P_{in} = \frac{|V_{oc}|^2}{8R_T}M_L = P_L$$

which gives $M = M_L$, that is, the impedance mismatch at the output plane equals that at the input plane. The Thévenin resistance is given by

$$R_T = \text{Re}\, Z_T = \text{Re}\, \frac{\Delta + Z_{22}Z_s}{Z_{11} + Z_s}$$

We can express Z_T in the form

$$Z_T = \frac{(\Delta + Z_{22}Z_s)(Z_{11}^* + Z_s^*)}{(Z_{11} + Z_s)(Z_{11}^* + Z_s^*)} = \frac{(\Delta + Z_{22}Z_s)(Z_{11}^* + Z_s^*)}{|Z_{11} + Z_s|^2}$$

The numerator equals

$$(X_{12}^2 - X_{11}X_{22} - X_{22}X_s + jX_{22}R_s)(-jX_{11} - jX_s + R_s)$$

and has a real part which is readily found to be simply $X_{12}^2 R_s$. Hence $R_T = X_{12}^2 R_s/|Z_{11} + Z_s|^2$. We now get

$$\frac{|V_{oc}|^2}{8R_T} = \frac{X_{12}^2|V_s|^2}{|Z_{11} + Z_s|^2 8R_T} = \frac{|V_s|^2}{8R_s} = P_{ava}$$

Since the available power at the output plane is the same as that from the source, the impedance mismatch $M_L = M$.

If we have several lossless networks (reciprocal) between the source and Z_L as shown in Fig. 5.26*c*, we can repeat the analysis using Thévenin equivalent circuits at each stage to find that at any plane the available power and impedance mismatch are invariant quantities.

If the networks are not reciprocal, i.e., are linear active networks, then the impedance mismatch is not invariant, e.g., an amplifier supplies more output power than is fed in at the input.

The invariance of the impedance mismatch factor in a lossless network places an important constraint on interstage matching networks in a multistage amplifier. The design of the interstage matching network must take this constraint into account.

In Fig. 5.27 we show a multistage microwave amplifier. Since transistors at microwave frequencies are often not absolutely stable, it is generally not possible to use conjugate impedance matching at the input and output of each stage. For the first stage let $Z_{in,1}$ be the impedance looking into the

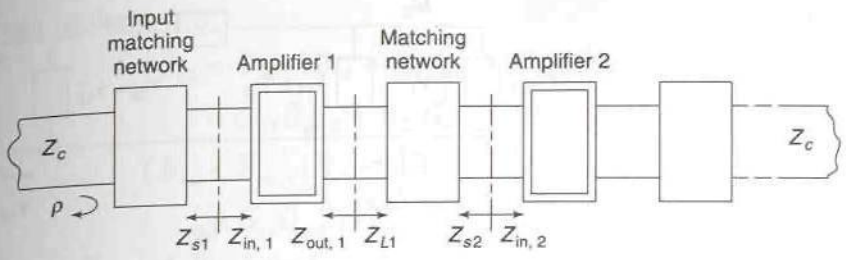


FIGURE 5.27
A multistage microwave amplifier.

first amplifier and let Z_{s1} be the source impedance presented to this stage by the input matching network (the networks are called matching networks even though they do not produce conjugate impedance matching). The input impedance mismatch is

$$M_1 = \frac{4R_{\text{in},1}R_{s1}}{|Z_{\text{in},1} + Z_{s1}|^2}$$

This same impedance mismatch must exist between the input transmission line and the input to the matching network. Hence on the input line we will have a reflection coefficient ρ that satisfies the constraint

$$1 - \rho^2 = M_1$$

Thus $\rho = (1 - M_1)^{1/2}$ and the input VSWR will be

$$\text{VSWR} = \frac{1 + \sqrt{1 - M_1}}{1 - \sqrt{1 - M_1}} \tag{5.32}$$

Consequently, an input VSWR greater than unity is unavoidable when conjugate impedance matching cannot be used. A resultant problem in microwave amplifier design is to find an optimum load impedance that will result in an input impedance $Z_{\text{in},1}$ that is close to the complex conjugate of the optimum source impedance Z_{s1} so as to minimize the input VSWR. The optimum source impedance is constrained by the requirement for a low noise figure for the input stage.

For the interstage network the matching network must transform the input impedance $Z_{\text{in},2}$ into the required load impedance Z_{L1} for stage 1 and simultaneously transform the output impedance $Z_{\text{out},1}$ of stage 1 into the required source impedance Z_{s2} for stage 2. The design of the matching network is constrained by the condition that the impedance mismatch be

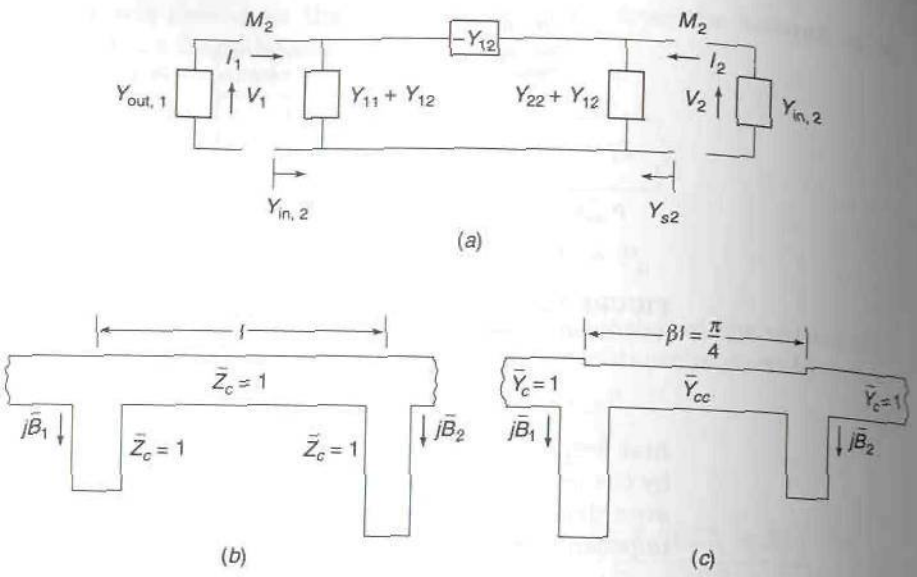


FIGURE 5.28
(*a*) A Π matching network; (*b*) and (*c*) transmission-line matching networks.

constant; thus

$$\frac{4R_{L1}R_{\text{out},1}}{|Z_{L1}+Z_{\text{out},1}|^2}=\frac{4R_{\text{in},2}R_{s2}}{|Z_{\text{in},2}+Z_{s2}|^2} \tag{5.33}$$

The design of microwave amplifiers is described in detail in Chap. 10. As part of the design procedure, the four impedances Z_{L1}, $Z_{\text{out},1}$, $Z_{\text{in},2}$, and Z_{s2} are determined so that the impedance mismatch constraint given by (5.33) is satisfied. Hence we can assume that we know the admittances Y_{L1}, $Y_{\text{out},1}$, $Y_{\text{in},2}$, and Y_{s2}. In terms of these self-consistent quantities, we can determine the parameters of the Π matching network illustrated in Fig. 5.28*a*. By conventional circuit analysis, we readily find that

$$\bar{Y}_{L1}=\bar{Y}_{11}-\frac{\bar{Y}_{12}^2}{\bar{Y}_{22}+\bar{Y}_{\text{in},2}}$$

$$\bar{Y}_{s2}=\bar{Y}_{22}-\frac{\bar{Y}_{12}^2}{\bar{Y}_{11}+\bar{Y}_{\text{out},1}}$$

For a lossless network $\bar{Y}_{11}=j\bar{B}_{11}$, $\bar{Y}_{12}=j\bar{B}_{12}$, and $\bar{Y}_{22}=j\bar{B}_{22}$. We can equate the real and imaginary parts of the above two equations. After a number of algebraic steps, the following pair of equations for $\bar{B}_{11}$ and $\bar{B}_{22}$

can be derived:

$$\left(\overline{G}_{s2} + \overline{G}_{\text{in},2}\right)\overline{B}_{11} - \left(\overline{G}_{\text{out},1} + \overline{G}_{L1}\right)\overline{B}_{22} = \overline{G}_{L1}\overline{B}_{\text{in},2} + \overline{B}_{L1}\overline{G}_{\text{in},2} - \overline{B}_{s2}\overline{G}_{\text{out},1} - \overline{G}_{s2}\overline{B}_{\text{out},1} \quad (5.34a)$$

$$\left(\overline{B}_{s2} + \overline{B}_{\text{in},2}\right)\overline{B}_{11} - \left(\overline{B}_{L1} + \overline{B}_{\text{out},1}\right)\overline{B}_{22} = \overline{G}_{s2}\overline{G}_{\text{out},1} - \overline{G}_{L1}\overline{G}_{\text{in},2} + \overline{B}_{L1}\overline{B}_{\text{in},2} - \overline{B}_{s2}\overline{B}_{\text{out},1} \quad (5.34b)$$

After $\overline{B}_{11}$ and $\overline{B}_{22}$ have been found, we use

$$\overline{B}_{12}^2 = \overline{G}_{s2}\overline{G}_{\text{out},1} + \left(\overline{B}_{11} + \overline{B}_{\text{out},1}\right)\left(\overline{B}_{22} - \overline{B}_{s2}\right) \quad (5.35)$$

to find $\overline{B}_{12}$. The expressions for $\overline{Y}_{L1}$ and $\overline{Y}_{s2}$ generate four equations which would generally not allow a Π network to be determined unless (5.33) was also satisfied.

Many matching networks can now be designed by relating their parameters to those of the Π network. Two possible transmission-line networks are shown in Figs. 5.28*b* and *c*. The parameters of the network in Fig. 5.28 are given by

$$\csc \beta l = |\overline{B}_{12}| \quad (5.36a)$$

$$j\overline{B}_1 = j\left(\overline{B}_{11} + \cot \beta l\right) \quad (5.36b)$$

$$j\overline{B}_2 = j\left(\overline{B}_{22} + \cot \beta l\right) \quad (5.36c)$$

where $j\overline{B}_1$ and $j\overline{B}_2$ are the stub susceptances. This network can only be used if $|\overline{B}_{12}| \geq 1$ since $\csc \beta l \geq 1$. When $|\overline{B}_{12}| < 1$ the network shown in Fig. 5.28*c* can be realized. Its parameters are given by

$$j\overline{B}_1 = j\left(\overline{B}_{11} + \overline{B}_{12}\right) \quad (5.37a)$$

$$j\overline{B}_2 = j\left(\overline{B}_{22} + \overline{B}_{12}\right) \quad (5.37b)$$

$$\overline{Y}_{cc} = \overline{B}_{12} \quad (5.37c)$$

where $\overline{Y}_{cc}$ is the characteristic admittance of the center one-quarter-wavelength-long section.

5.8 WAVEGUIDE REACTIVE ELEMENTS

In the place of transmission-line stubs, any other element that acts as a shunt susceptance may be used for the purpose of matching an arbitrary load impedance to a waveguide or transmission line. A number of such reactive elements for use in rectangular waveguides supporting the dominant TE_{10} propagating mode are described in this section.† The formulas given for the normalized susceptance of these elements are approximate

†Detailed information on susceptance values and equivalent circuits are given in N. Marcuvitz (ed.), "Waveguide Handbook," McGraw-Hill Book Company, New York, 1951.

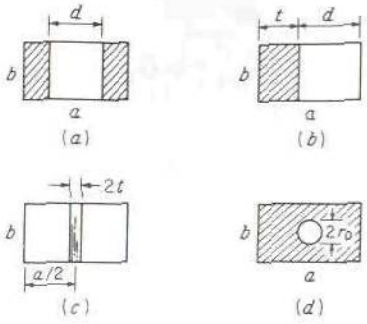


FIGURE 5.29
Shunt inductive elements. (*a*) Symmetrical diaphragm; (*b*) asymmetrical diaphragm; (*c*) thin circular post; (*d*) small circular aperture.

ones, with accuracies of the order of 10 percent or better. The derivation of these formulas requires the detailed solution of boundary-value problems and is outside the scope of this text.†

Shunt Inductive Elements

Figure 5.29 illustrates a number of rectangular waveguide elements that act as shunt inductive susceptances for the TE_{10} mode. These consist of thin metallic windows extending across the narrow dimension of the guide as in Figs. 5.29*a* and *b*, a very thin cylindrical post as in Fig. 5.29*c*, and a small circular aperture as in Fig. 5.29*d*. When a TE_{10} mode is incident on any of these discontinuities, evanescent TE_{n0} modes are excited in order to provide a total field that will satisfy the required boundary condition of a vanishing tangential electric field on the obstacle. These nonpropagating modes store predominantly magnetic energy and give the obstacle its inductive characteristics.

Approximate values for the normalized inductive susceptance of these obstacles are:

For Fig. 5.29*a*,

$$\overline{B} = \frac{2\pi}{\beta a} \cot^2 \frac{\pi d}{2a} \left(1 + \frac{a\gamma_3 - 3\pi}{4\pi} \sin^2 \frac{\pi d}{a} \right) \tag{5.38}$$

where $\beta = [k_0^2 - (\pi/a)^2]^{1/2}$ and $\gamma_3 = [(3\pi/a)^2 - k_0^2]^{1/2}$.

For Fig. 5.29*b*,

$$\overline{B} = \frac{2\pi}{\beta a} \cot^2 \frac{\pi d}{2a} \left(1 + \csc^2 \frac{\pi d}{2a} \right) \tag{5.39}$$

†N. Marcuvitz, *loc. cit.*
R. E. Collin, "Field Theory of Guided Waves," 2nd. ed., IEEE Press, Piscataway, N.J. 1991.
L. Lewin, "Theory of Waveguides," Newnes-Butterworth, London, 1975.

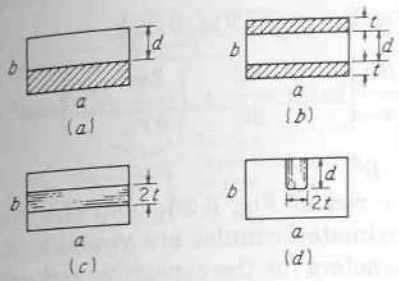


FIGURE 5.30
Shunt capacitive elements. (*a*) Asymmetrical capacitive diaphragm; (*b*) symmetrical diaphragm; (*c*) capacitive rod; (*d*) capacitive post.

For the thin inductive post of Fig. 5.29*c*,

$$\overline{B} = \frac{4\pi}{\beta a}\left[\ln\frac{a}{\pi t} - 1 + 2\left(\frac{a}{\pi t}\right)^2 \sum_{n=3,5,\ldots}^{\infty}\left(\frac{\pi}{a\gamma_n} - \frac{1}{n}\right)\sin^2\frac{n\pi t}{a}\right]^{-1} \tag{5.40}$$

where $\gamma_n = [(n\pi/a)^2 - k_0^2]^{1/2}$ and t is the post radius. For the small centered circular aperture of Fig. 5.29*d*,

$$\overline{B} = \frac{3ab}{8\beta r_0^3} \tag{5.41}$$

Shunt Capacitive Elements

Typical shunt capacitive elements that may be used for matching purposes are illustrated in Fig. 5.30. These consist of thin metal septa extending across the broad dimension of the guide to form capacitive diaphragms as in Figs. 5.30*a* and *b*, a thin circular rod extending across the guide as in Fig. 5.30*c*, and a short thin circular post extending into the guide as in Fig. 5.30*d*. The post illustrated in Fig. 5.30*d* behaves more like an *LC* series network connected across a transmission line. When the depth of penetration is between $0.7b$ and $0.9b$, it becomes resonant and acts almost like an ideal short circuit. For lengths greater than this resonant length, the post is equivalent to a shunt inductive susceptance. Actually, for a post of finite thickness, the equivalent circuit is a T network, but for small-diameter posts, the series elements in this T network are negligible (for post diameters less than about $0.05a$).

Approximate expressions for the normalized susceptance of the obstacles illustrated in Fig. 5.30 are:

For the asymmetrical diaphragm of Fig. 5.30*a*,

$$\overline{B} = \frac{4\beta b}{\pi}\left[\ln\csc\frac{\pi d}{2b} + \left(\frac{\pi}{b\gamma_1} - 1\right)\cos^4\frac{\pi d}{2b}\right] \tag{5.42}$$

where $\beta = [k_0^2 - (\pi/a)^2]^{1/2}$ and $\gamma_1 = [(\pi/b)^2 - \beta^2]^{1/2}$.

For the symmetrical diaphragm of Fig. 5.30*b*,

$$\overline{B} = \frac{2\beta b}{\pi}\left[\ln \csc \frac{\pi d}{2b} + \left(\frac{2\pi}{b\gamma_2} - 1\right)\cos^4 \frac{\pi d}{2b}\right] \tag{5.43}$$

where $\gamma_2 = [(2\pi/b)^2 - \beta^2]^{1/2}$.

For the capacitive rod of Fig. 5.30*c* and the post illustrated in Fig. 5.30*d*, no simple approximate formulas are available. Analytical expressions for the T-network parameters for the capacitive rod are given by Lewin, but are not reproduced here.†

The inductive and capacitive properties of the waveguide obstacles described above are determined by the electric and magnetic energy in the evanescent modes excited by the obstacle. For example, the inductive obstacles discussed only excite TE evanescent modes and these modes store more magnetic energy than electric energy. The evanescent modes are nonpropagating modes so they represent the local field in the vicinity of the obstacles. The evanescent modes are excited with amplitudes such that the combination of all of the evanescent modes with the incident, reflected, and transmitted dominant mode, produces a total field with a zero tangential electric field at the surface of the obstacle.

There is no way to assign a unique inductance or capacitance to these obstacles since there is no unique way to define the voltage, current, or characteristic impedance for a waveguide. Nevertheless, these obstacles do produce a reflected wave with a reflection coefficient that is correctly determined in terms of the equivalent normalized reactance and susceptance of the obstacle as connected into the equivalent-transmission-line model of the waveguide.

Waveguide Stub Tuners

An approximate equivalent of a single-stub matching network is the *sliding screw* tuner illustrated in Fig. 5.31*a*. This consists of a variable-depth screw mounted on a sliding carriage free to move longitudinally along the guide over a distance of at least a half guide wavelength. The screw penetrates into the guide through a centered narrow slot in the broad wall of the guide. This slot is cut along the current flow lines so that it has a negligible perturbing effect on the internal field. Since the position of the screw is adjustable over at least a half guide wavelength, its penetration does not need to be so great that it will behave as an inductive element; i.e., a match can be obtained with a shunt capacitive susceptance in all cases, as a review of single-stub-matching theory will verify.

†*Op. cit.*, chap. 2. This text contains many excellent and instructive derivations of equivalent circuit parameters for a variety of waveguide structures.

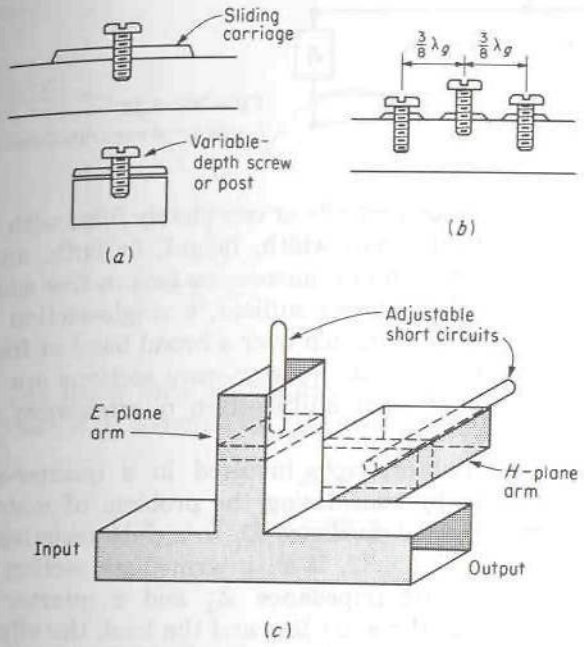


FIGURE 5.31
Waveguide tuners. (*a*) Sliding-screw tuner; (*b*) triple-screw tuner; (*c*) *E-H* tuner.

Three variable-depth screws spaced a fixed distance of about $3\lambda_g/8$ apart as in Fig. 5.31*b* are essentially equivalent to a triple-stub tuner and can match a large variety of loads even though the range of susceptance values obtainable from a single screw is limited.

Circuits that are physically more like the actual short-circuited transmission-line stub are also possible. Figure 5.31*c* illustrates combined *E*-plane and *H*-plane stubs, referred to as an *E-H* tuner. The positions of the sliding short circuits in the *E*- and *H*-plane arms are variable, so that a wide range of load impedances may be matched. The equivalent circuit of either an *E*-plane or *H*-plane junction is, however, much more elaborate than a simple shunt- or series-connected transmission line because the junctions are of the order of a wavelength in size and hence produce a very complicated field structure in their vicinity. Nevertheless, since no power flow is possible through the arms terminated in the short circuits, these do still provide adjustable reactance elements that may be used for matching purposes.

5.9 QUARTER-WAVE TRANSFORMERS

Quarter-wave transformers are primarily used as intermediate matching sections when it is desired to connect two waveguiding systems of different characteristic impedance. Examples are the connection of two transmission lines with different characteristic impedances, connection of an empty wave-

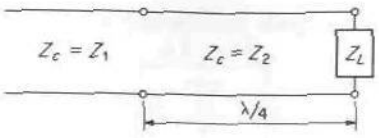


FIGURE 5.32
A quarter-wave transformer.

guide to a waveguide partially or completely filled with dielectric, connection of two guides of different width, height, or both, and the matching of a dielectric medium such as a microwave lens to free space. If a match over a narrow band of frequencies suffices, a single-section transformer may be used. To obtain a good match over a broad band of frequencies, two, three, or even more intermediate quarter-wave sections are commonly used. The optimum design of such multisection quarter-wave transformers is presented in this section.

The essential principle involved in a quarter-wave transformer is readily explained by considering the problem of matching a transmission line of characteristic impedance Z_1 to a pure resistive load impedance Z_L, as illustrated in Fig. 5.32. If an intermediate section of transmission line with a characteristic impedance Z_2 and a quarter wavelength long is connected between the main line and the load, the effective load impedance presented to the main line is

$$Z = Z_2 \frac{Z_L + jZ_2 \tan(\beta\lambda/4)}{Z_2 + jZ_L \tan(\beta\lambda/4)} = \frac{Z_2^2}{Z_L} \tag{5.44}$$

If Z_2 is chosen equal to $\sqrt{Z_1 Z_L}$, then $Z = Z_1$ and the load is matched to the main line. In other words, the intermediate section of transmission line of length $\lambda/4$ transforms the load impedance Z_L into an impedance Z_1 and hence acts as an ideal transformer of turns ratio $\sqrt{Z_1/Z_L}$. A perfect match is obtained only at that frequency for which the transformer is a quarter wavelength (or $n\lambda/2 + \lambda/4$) long.

Let θ be the electrical length of the transformer at the frequency f, that is, $\beta(f)l = \theta$, where the phase constant β has been written as a function of frequency. For a TEM wave in an air-filled line, $\beta l = 2\pi fl/c$. At any frequency the input impedance presented to the main line is

$$Z_{\text{in}} = Z_2 \frac{Z_L + jZ_2 t}{Z_2 + jZ_L t} \tag{5.45}$$

where $t = \tan\theta = \tan\beta l$. Consequently, the reflection coefficient is

$$\Gamma = \frac{Z_{\text{in}} - Z_1}{Z_{\text{in}} + Z_1} = \frac{Z_2(Z_L - Z_1) + jt(Z_2^2 - Z_1 Z_L)}{Z_2(Z_L + Z_1) + jt(Z_2^2 + Z_1 Z_L)}$$
$$= \frac{Z_L - Z_1}{Z_L + Z_1 + jt2\sqrt{Z_1 Z_L}} \tag{5.46}$$

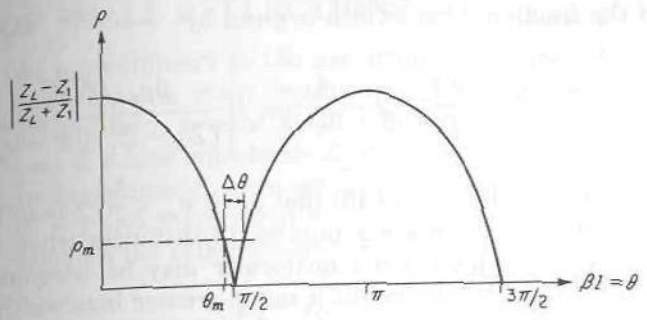


FIGURE 5.33
Bandwidth characteristic for a single-section quarter-wave transformer.

The latter form is obtained by using the relation $Z_2^2 = Z_1 Z_L$. The magnitude of Γ, denoted by ρ, is readily evaluated and is given by

$$\rho = \frac{|Z_L - Z_1|}{\left[(Z_L + Z_1)^2 + 4t^2 Z_1 Z_L\right]^{1/2}} = \frac{1}{\left[1 + \left(\dfrac{2\sqrt{Z_1 Z_L}}{Z_L - Z_1}\sec\theta\right)^2\right]^{1/2}} \tag{5.47}$$

For θ near $\pi/2$, this equation is well approximated by

$$\rho = \frac{|Z_L - Z_1|}{2\sqrt{Z_1 Z_L}}|\cos\theta| \tag{5.48}$$

A plot of ρ versus θ is given in Fig. 5.33, and this is essentially a plot of ρ versus frequency. The variation of ρ with frequency, or θ, is periodic because of the periodic variation of the input impedance with frequency; i.e., the impedance repeats its value every time the electrical length of the transformer changes by π. If ρ_m is the maximum value of reflection coefficient that can be tolerated, the useful bandwidth provided by the transformer is that corresponding to the range $\Delta\theta$ in Fig. 5.33. Because of the rapidly increasing values of ρ on either side of $\theta = \pi/2$, the useful bandwidth is small. The value of θ at the edge of the useful passband may be found from (5.47) by equating ρ to ρ_m; thus

$$\theta_m = \cos^{-1}\left|\frac{2\rho_m\sqrt{Z_1 Z_L}}{(Z_L - Z_1)\sqrt{1 - \rho_m^2}}\right| \tag{5.49}$$

In the case of a TEM wave, $\theta = \beta l = (f/f_0)(\pi/2)$, where f_0 is the frequency for which $\theta = \pi/2$. In this case the bandwidth is given by

$$\Delta f = 2(f_0 - f_m) = 2\left(f_0 - \frac{2f_0}{\pi}\theta_m\right)$$

and the fractional bandwidth is given by

$$\frac{\Delta f}{f_0} = 2 - \frac{4}{\pi} \cos^{-1} \left| \frac{2\rho_m \sqrt{Z_1 Z_L}}{(Z_L - Z_1)\sqrt{1 - \rho_m^2}} \right| \tag{5.50}$$

where that solution of (5.49) that gives $\theta_m < \pi/2$ is to be chosen.

Although there are a number of instances when the bandwidth provided by a single-section transformer may be adequate, there are also a number of situations in which much greater bandwidths must be provided for. The required increase in bandwidth can be obtained by using multisection quarter-wave transformers. The approximate theory of these multisection transformers is discussed first, in order to develop a theory that also has application in the design of other microwave devices, such as directional couplers and antenna arrays. This is followed by a discussion and presentation of results obtainable from a more exact analysis.

It should be noted that in the previous discussion it was assumed that the characteristic impedances Z_1 and Z_2 were independent of frequency. For transmission lines this is a good approximation, but for waveguides the wave impedance varies with frequency, and this complicates the analysis considerably. In addition, for both transmission lines and waveguides, there are reactive fields excited at the junctions of the different sections, brought about because of the change in geometrical cross section necessary to achieve the required characteristic impedances. These junction effects can often be represented by a pure shunt susceptance at each junction.† The susceptive elements will also vary the performance of any practical transformer from the predicted performance based on an ideal model where junction effects are neglected. In spite of all these limitations, only the theory for ideal transformers is developed here; i.e., junction effects and the frequency dependency of the equivalent characteristic impedances are neglected. Thus the theory given will only be indicative of the performance that can be obtained in the nonideal case.‡

†S. B. Cohn, Optimum Design of Stepped Transmission Line Transformers, *IRE Trans.*, vol. MTT-3, pp. 16–21, April, 1955. This paper presents an approximate theory of Chebyshev transformers, together with a method of accounting for the reactances introduced at each step.

‡For typical application to waveguide transformers, see:

R. E. Collin and J. Brown, The Design of Quarter-Wave Matching Layers for Dielectric Surfaces, *Proc. IEE*, vol. 103, pt. C, pp. 153–158, March, 1956.

L. Young, Optimum Quarter-Wave Transformers, *IRE Trans.*, vol. MTT-8, pp. 478–482, September, 1960; also Inhomogeneous Quarter-Wave Transformers of Two Sections, *ibid.*, pp. 645–649, November, 1960.

E. S. Hensperger, Broad-Band Stepped Transformers from Rectangular to Double-Ridged Waveguide, *IRE Trans.*, vol. MTT-6, pp. 311–314, July, 1958.

5.10 THEORY OF SMALL REFLECTIONS

As a preliminary to the approximate analysis of multisection quarter-wave transformers, some results pertaining to the overall reflection coefficient arising from several small reflecting obstacles are required. Consider the case of a load impedance Z_L connected to a transmission line of characteristic impedance Z_1 through an intermediate section of line of electrical length $\beta l = \theta$ and characteristic impedance Z_2, as illustrated in Fig. 5.34. For each junction the reflection and transmission coefficients are

$$\Gamma_1 = \frac{Z_2 - Z_1}{Z_2 + Z_1} \qquad \Gamma_2 = -\Gamma_1$$

$$T_{21} = 1 + \Gamma_1 = \frac{2Z_2}{Z_1 + Z_2} \qquad T_{12} = 1 + \Gamma_2 = \frac{2Z_1}{Z_1 + Z_2}$$

$$\Gamma_3 = \frac{Z_L - Z_2}{Z_L + Z_2}$$

A wave of unit amplitude is incident, and the total reflected wave has a complex amplitude Γ equal to the total reflection coefficient. When the incident wave strikes the first junction, a partial reflected wave of amplitude Γ_1 is produced. A transmitted wave of amplitude T_{21} is then incident on the second junction. A portion of this is reflected to give a wave of amplitude $\Gamma_3 T_{21} e^{-2j\theta}$ incident from the right on the first junction. A portion $T_{12}T_{21}\Gamma_3 e^{-2j\theta}$ is transmitted, and a portion $\Gamma_2\Gamma_3 T_{12}T_{21}e^{-2j\theta}$ is reflected back toward Z_L. Figure 5.35 illustrates the first few of the infinite number of multiply reflected waves that occur. The total reflected wave of amplitude Γ is the sum of all the partial waves transmitted past the first junction toward the left. This sum is given by

$$\Gamma = \Gamma_1 + T_{12}T_{21}\Gamma_3 e^{-2j\theta} + T_{12}T_{21}\Gamma_3^2\Gamma_2 e^{-4j\theta} + \cdots$$

$$= \Gamma_1 + T_{12}T_{21}\Gamma_3 e^{-2j\theta} \sum_{n=0}^{\infty} \Gamma_2^n \Gamma_3^n e^{-2jn\theta}$$

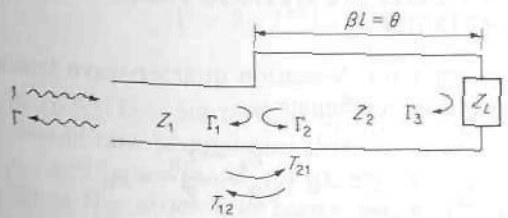


FIGURE 5.34
A microwave circuit with two reflecting junctions.

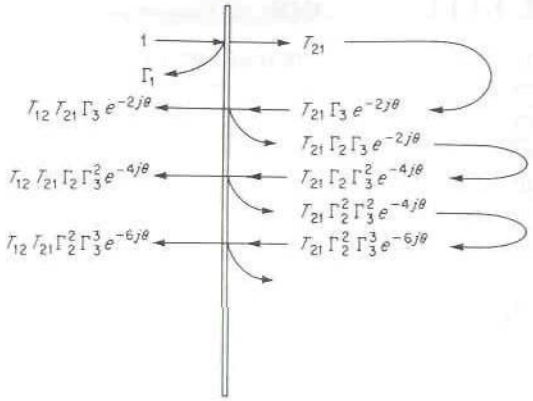


FIGURE 5.35
Multiple reflection of waves for a circuit with two reflecting junctions.

This geometric series is readily summed to give [note that $\sum_{n=0}^{\infty} r^n = (1-r)^{-1}$]

$$\Gamma = \Gamma_1 + \frac{T_{12}T_{21}\Gamma_3 e^{-2j\theta}}{1 - \Gamma_2\Gamma_3 e^{-2j\theta}}$$

Replacing T_{12} by $1 + \Gamma_2 = 1 - \Gamma_1$ and T_{21} by $1 + \Gamma_1$ gives

$$\Gamma = \frac{\Gamma_1 + \Gamma_3 e^{-2j\theta}}{1 + \Gamma_1\Gamma_3 e^{-2j\theta}} \tag{5.51}$$

If $|\Gamma_1|$ and $|\Gamma_3|$ are both small compared with unity, an excellent approximation to Γ is

$$\Gamma = \Gamma_1 + \Gamma_3 e^{-2j\theta} \tag{5.52}$$

This result states that, for small reflections, the resultant reflection coefficient is just that obtained by taking only first-order reflections into account. This is the result that will be used to obtain a first-order theory for multisection quarter-wave transformers. As an indication of the accuracy of the approximate formula, note that if $|\Gamma_1| = |\Gamma_3| = 0.2$, the error in Γ does not exceed 4 percent.

5.11 APPROXIMATE THEORY FOR MULTISECTION QUARTER-WAVE TRANSFORMERS

Figure 5.36 illustrates an N-section quarter-wave transformer. At the first junction the reflection coefficient is

$$\Gamma_0 = \frac{Z_1 - Z_0}{Z_1 + Z_0} = \rho_0 \tag{5.53a}$$

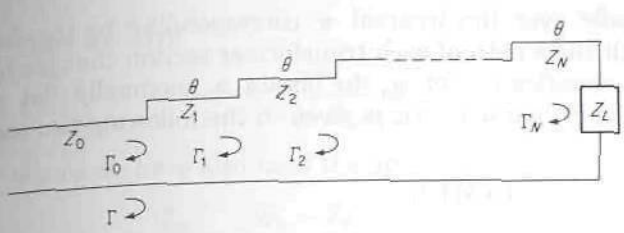


FIGURE 5.36
A multisection quarter-wave transformer.

Similarly, at the nth junction, the reflection coefficient is

$$\Gamma_n = \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} = \rho_n \tag{5.53b}$$

The last reflection coefficient is

$$\Gamma_N = \frac{Z_L - Z_N}{Z_L + Z_N} = \rho_N \tag{5.53c}$$

Note that Z_0 is a characteristic impedance, and not necessarily equal to $(\mu_0/\epsilon_0)^{1/2}$ here. Each section has the same electrical length $\beta l = \theta$, and will be a quarter wave long at the matching frequency f_0. The load Z_L is assumed to be a pure resistance, and may be greater or smaller than Z_0. In this analysis it is chosen greater, so that all $\Gamma_n = \rho_n$, where ρ_n is the magnitude of Γ_n. If Z_L is smaller than Z_0, all Γ_n are negative real numbers and the only modification required in the theory is replacing all ρ_n by $-\rho_n$.

For a first approximation the total reflection coefficient is the sum of the first-order reflected waves only. This is given by

$$\Gamma = \rho_0 + \rho_1 e^{-2j\theta} + \rho_2 e^{-4j\theta} + \cdots + \rho_N e^{-2jN\theta} \tag{5.54}$$

where $e^{-2jn\theta}$ accounts for the phase retardation introduced because of the different distances the various partial waves must travel.

At this point it is expedient to assume that the transformer is symmetrical, so that $\rho_0 = \rho_N$, $\rho_1 = \rho_{N-1}$, $\rho_2 = \rho_{N-2}$, etc. In this case (5.54) becomes

$$\Gamma = e^{-jN\theta}\left[\rho_0(e^{jN\theta} + e^{-jN\theta}) + \rho_1(e^{j(N-2)\theta} + e^{-j(N-2)\theta}) + \cdots\right] \tag{5.55}$$

where the last term is $\rho_{(N-1)/2}(e^{j\theta} + e^{-j\theta})$ for N odd and $\rho_{N/2}$ for N even. It is thus seen that *for a symmetrical transformer the reflection coefficient* Γ is given by a Fourier cosine series:

$$\Gamma = 2e^{-jN\theta}\left[\rho_0 \cos N\theta + \rho_1 \cos(N-2)\theta + \cdots + \rho_n \cos(N-2n)\theta + \cdots\right] \tag{5.56}$$

In (5.56) the last term is $\rho_{(N-1)/2}\cos\theta$ for N odd and $\frac{1}{2}\rho_{N/2}$ for N even. It should now be apparent that by a proper choice of the reflection coefficients ρ_n, and hence the Z_n, a variety of passband characteristics can be obtained. Since the series is a cosine series, the periodic function that it defines is

periodic over the interval π corresponding to the frequency range over which the length of each transformer section changes by a half wavelength. The specification of ρ_n to obtain a maximally flat and an equal-ripple passband characteristic is given in the following two sections.

5.12 BINOMIAL TRANSFORMER

A maximally flat passband characteristic is obtained if $\rho = |\Gamma|$ and the first $N-1$ derivatives with respect to frequency (or θ) vanish at the matching frequency f_0, where $\theta = \pi/2$. Such a characteristic is obtained if we choose

$$\Gamma = A(1 + e^{-2j\theta})^N \tag{5.57a}$$

for which

$$\rho = |\Gamma| = \left| A2^N (\cos\theta)^N \right| \tag{5.57b}$$

When $\theta = 0$ or π, we have $\Gamma = (Z_L - Z_0)/(Z_L + Z_0)$, and from (5.57*a*) we obtain $\Gamma = A2^N$. Thus the constant A is given by

$$A = 2^{-N} \frac{Z_L - Z_0}{Z_L + Z_0} \tag{5.58}$$

However, if we use the theory of small reflections, then the constant A should be chosen in a different way, which we will explain shortly.

Expanding (5.57*a*) by the binomial expansion gives

$$\Gamma = 2^{-N} \frac{Z_L - Z_0}{Z_L + Z_0} (1 - e^{-2j\theta})^N = 2^{-N} \frac{Z_L - Z_0}{Z_L + Z_0} \sum_{n=0}^{N} C_n^N e^{-j2n\theta} \tag{5.59}$$

where the binomial coefficients are given by

$$C_n^N = \frac{N(N-1)(N-2)\cdots(N-n+1)}{n!} = \frac{N!}{(N-n)!n!} \tag{5.60}$$

Note that $C_n^N = C_{N-n}^N$, $C_0^N = 1$, $C_1^N = N = C_{N-1}^N$, etc. Comparing (5.59) with (5.54) shows that we must choose

$$\rho_n = 2^{-N} \frac{Z_L - Z_0}{Z_L + Z_0} C_n^N = \rho_{N-n}$$

since $C_n^N = C_{N-n}^N$.

To obtain a simple solution for the characteristic impedances Z_n, it is convenient to make a further approximation. Since we have already specified that all ρ_n are to be small, we can use the result

$$\ln \frac{Z_{n+1}}{Z_n} \approx 2 \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} = 2\rho_n \tag{5.61a}$$

$$Z_{n+1} = Z_n e^{2\rho_n} \tag{5.61b}$$

Thus we have

$$\ln \frac{Z_{n+1}}{Z_n} = 2\rho_n = 2^{-N} C_n^N \ln \frac{Z_L}{Z_0} \tag{5.62}$$

where we have also used the approximation

$$\ln \frac{Z_L}{Z_0} = 2\frac{Z_L - Z_0}{Z_L + Z_0} + \frac{2}{3}\left(\frac{Z_L - Z_0}{Z_L + Z_0}\right)^3 + \cdots \approx 2\frac{Z_L - Z_0}{Z_L + Z_0}$$

When we use the theory of small reflections (5.55) gives, for $\theta = 0$,

$$\Gamma(0) = \rho_0 + \rho_1 + \rho_2 + \cdots + \rho_N$$

When we use (5.62) to evaluate the characteristic impedances, and also use $A = 2^{-(N+1)} \ln(Z_L/Z_0)$, we obtain

$$2^N A = \Gamma(0) = \frac{1}{2}\left[\ln \frac{Z_1}{Z_0} + \ln \frac{Z_2}{Z_1} + \cdots + \ln \frac{Z_L}{Z_N}\right]$$

$$= \frac{1}{2} \ln\left(\frac{Z_1}{Z_0}\frac{Z_2}{Z_1} \cdots \frac{Z_L}{Z_N}\right) = \frac{1}{2} \ln \frac{Z_L}{Z_0} \tag{5.63}$$

Thus the approximations introduced above for finding the characteristic impedances should be used along with (5.63) to evaluate $\Gamma(0)$. This will make the quarter-wave transformer designs using the theory of small reflections self-consistent.† In place of (5.58) we use the expression given above for A.

Equation (5.62) gives the solution for the logarithm of the impedances, and since these are proportional to the binomial coefficients, the transformer is called a binomial transformer. Since the theory is approximate, the range of Z_L is restricted to about

$$0.5Z_0 < Z_L < 2Z_0$$

for accurate results.

As an example, consider a two-section transformer. From (5.62) obtain

$$\ln \frac{Z_1}{Z_0} = \frac{1}{4} \ln \frac{Z_L}{Z_0} \quad \text{or} \quad Z_1 = Z_L^{1/4} Z_0^{3/4}$$

and

$$\ln \frac{Z_2}{Z_1} = \frac{1}{2} \ln \frac{Z_L}{Z_0} \quad \text{or} \quad Z_2 = Z_L^{3/4} Z_0^{1/4}$$

since $C_0^2 = 1$ and $C_1^2 = 2$. Although the approximate theory was used, it turns out that the above values for Z_1 and Z_2 for the special case of the

†The author is indebted to Dr. E. E. Altschuler for pointing this out.

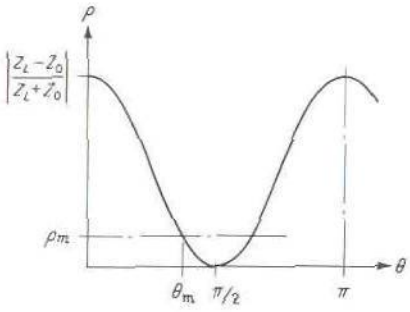


FIGURE 5.37
Passband characteristic for a maximally flat transformer.

two-section transformer are the correct nonapproximate solutions, a result that gives an indication of the accuracy of the approximate theory.

The type of passband characteristic obtained with a maximally flat transformer is illustrated in Fig. 5.37. Let ρ_m be the maximum value of ρ that can be tolerated. The angle θ_m that gives $\rho = \rho_m$ is given by

$$\theta_m = \cos^{-1}\left|\frac{2\rho_m}{\ln(Z_L/Z_0)}\right|^{1/N} \tag{5.64}$$

as obtained from (5.57*b*). In the case of transmission-line sections, $\theta = \pi f/2f_0$, and hence the fractional bandwidth is given by

$$\frac{\Delta f}{f_0} = \frac{2(f_0 - f_m)}{f_0} = 2 - \frac{4}{\pi}\cos^{-1}\left|\frac{2\rho_m}{\ln(Z_L/Z_0)}\right|^{1/N} \tag{5.65}$$

since $\theta_m = \pi f_m/2f_0$. Note that in (5.65) the solution to the inverse cosine function is chosen so that $\theta_m < \pi/2$. By comparing Figs. 5.33 and 5.37 it is clear that a multisection maximally flat transformer can provide a much greater useful bandwidth than a single-section transformer.

5.13 CHEBYSHEV TRANSFORMER

Instead of a maximally flat passband characteristic, an equally useful characteristic is one that may permit ρ to vary between 0 and ρ_m in an oscillatory manner over the passband. A transformer designed to yield an equal-ripple characteristic as illustrated in Fig. 5.38 is of this type and provides a considerable increase in bandwidth over the binomial transformer design. The equal-ripple characteristic is obtained by making ρ behave according to a Chebyshev polynomial, and hence the name Chebyshev transformer. It is possible to have ρ vanish at as many different frequencies in the passband as there are transformer sections. To see how Chebyshev polynomials may be used in the design, it is necessary to consider the basic properties of these polynomials first.

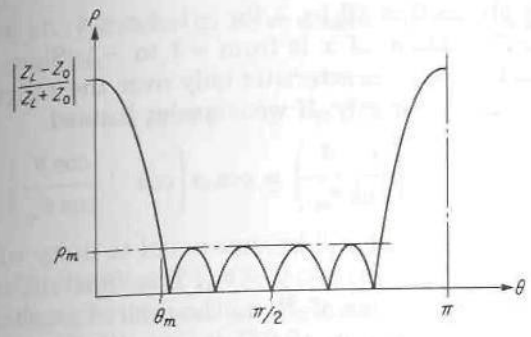


FIGURE 5.38
Equal-ripple characteristic obtained from a Chebyshev transformer.

The Chebyshev polynomial of degree n, denoted by $T_n(x)$, is an nth-degree polynomial in x. The first four polynomials and the recurrence relation are

$$T_1(x) = x$$
$$T_2(x) = 2x^2 - 1$$
$$T_3(x) = 4x^3 - 3x$$
$$T_4(x) = 8x^4 - 8x^2 + 1$$
$$T_n(x) = 2xT_{n-1} - T_{n-2}$$

The polynomials T_n oscillate between ± 1 for x in the range $|x| \leq 1$ and increase in magnitude indefinitely for x outside this range. Figure 5.39 gives a sketch of the first four polynomials. If x is replaced by $\cos\theta$, we have

$$T_n(\cos\theta) = \cos n\theta \tag{5.66}$$

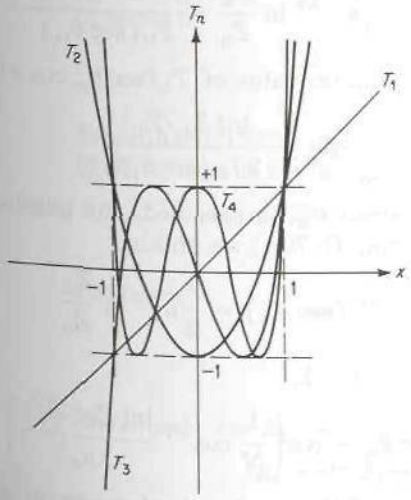


FIGURE 5.39
Chebyshev polynomials.

which clearly shows that $|T_n| \leq 1$ for $-1 \leq x \leq 1$. As θ varies from 0 to π, the corresponding range of x is from $+1$ to -1. Since we wish to make ρ have the equal-ripple characteristic only over the range θ_m to $\pi - \theta_m$, we cannot use $T_n(\cos\theta)$ directly. If we consider instead

$$T_n\left(\frac{\cos\theta}{\cos\theta_m}\right) = \cos n\left(\cos^{-1}\frac{\cos\theta}{\cos\theta_m}\right) \tag{5.67}$$

we see that the argument will become equal to unity when $\theta = \theta_m$ and will be less than unity for $\theta_m < \theta < \pi - \theta_m$. This function will therefore confine the equal-ripple oscillations of T_n to the desired passband.

The function given in (5.67) is an nth-degree polynomial in the variable $\cos\theta/\cos\theta_m$. Since $(\cos\theta)^n$ can be expanded into a series of cosine terms such as $\cos\theta, \cos 2\theta, \ldots, \cos n\theta$, it follows that (5.67) is a series of the form (5.56). Hence we may choose

$$\begin{aligned}\Gamma &= 2e^{-jN\theta}[\rho_0 \cos N\theta + \rho_1 \cos(N-2)\theta + \cdots \\ &\qquad + \rho_n \cos(N-2n)\theta + \cdots] \\ &= Ae^{-jN\theta}T_N(\sec\theta_m \cos\theta)\end{aligned} \tag{5.68}$$

where A is a constant to be determined. When $\theta = 0$, we have

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = AT_N(\sec\theta_m) \approx \frac{1}{2}\ln\frac{Z_L}{Z_0}$$

and so

$$A = \frac{\ln(Z_L/Z_0)}{2T_N(\sec\theta_m)}$$

Consequently, we have

$$\Gamma = \frac{1}{2}e^{-jN\theta}\ln\frac{Z_L}{Z_0}\frac{T_N(\sec\theta_m\cos\theta)}{T_N(\sec\theta_m)} \tag{5.69}$$

In the passband the maximum value of $T_N(\sec\theta_m\cos\theta)$ is unity, and hence

$$\rho_m = \frac{\ln(Z_L/Z_0)}{2T_N(\sec\theta_m)} \tag{5.70a}$$

If the passband, and hence θ_m, is specified, the passband tolerance ρ_m is fixed, and vice versa. From (5.70a) we obtain

$$T_N(\sec\theta_m) = \frac{1}{2}\rho_m^{-1}\ln\frac{Z_L}{Z_0}$$

or by using (5.67) for $\cos\theta = 1$,

$$\sec\theta_m = \cos\left(\frac{1}{N}\cos^{-1}\frac{\ln(Z_L/Z_0)}{2\rho_m}\right) \tag{5.70b}$$

which gives θ_m in terms of the passband tolerance on ρ, that is, ρ_m.

In order to solve (5.68) for the unknown ρ_n, we need the following results:

$$(\cos\theta)^n = 2^{-n}e^{-jn\theta}(1 + e^{2j\theta})^n = 2^{-n}e^{-jn\theta}\sum_{m=0}^{n} C_m^n e^{j2m\theta}$$

$$= 2^{-n+1}\left[C_0^n \cos n\theta + C_1^n \cos(n-2)\theta + \cdots + C_m^n \cos(n-2m)\theta + \cdots\right] \quad (5.71)$$

The last term in (5.71) is $\frac{1}{2}C_{n/2}^n$ for n even and $C_{(n-1)/2}^n \cos\theta$ for n odd. Using (5.71) and the earlier expression for $T_n(x)$, we can obtain the following:

$$T_1(\sec\theta_m \cos\theta) = \sec\theta_m \cos\theta \quad (5.72a)$$

$$T_2(\sec\theta_m \cos\theta) = 2(\sec\theta_m \cos\theta)^2 - 1 = \sec^2\theta_m(1 + \cos 2\theta) - 1 \quad (5.72b)$$

$$T_3(\sec\theta_m \cos\theta) = \sec^3\theta_m(\cos 3\theta + 3\cos\theta) - 3\sec\theta_m \cos\theta \quad (5.72c)$$

$$T_4(\sec\theta_m \cos\theta) = \sec^4\theta_m(\cos 4\theta + 4\cos 2\theta + 3) - 4\sec^2\theta_m(\cos 2\theta + 1) \quad (5.72d)$$

These results are sufficient for designing transformers up to four sections in length. A greater number of sections would rarely be required in practice.

Example 5.2 Design of a two-section Chebyshev transformer. As an example, consider the design of a two-section transformer to match a line with $Z_0 = 1$ to a line or load with $Z_L = 2$. Let the maximum tolerable value of ρ be $\rho_m = 0.05$. Using (5.70a), we obtain

$$T_2(\sec\theta_m) = 2\sec^2\theta_m - 1 = \frac{1}{3(0.05)} = 6.67$$

and hence $\sec\theta_m = 1.96$, and $\theta_m = 1.04$. Thus the fractional bandwidth that is obtained is

$$\frac{\Delta\theta}{\pi/2} = \frac{\Delta f}{f_0} = \frac{4}{\pi}\left(\frac{\pi}{2} - 1.04\right) = 0.675$$

From (5.68), (5.69), and (5.72b), we obtain [refer to the remarks following (5.56) as regards the last term in the cosine series for ρ]

$$2\rho_0 \cos 2\theta + \rho_1 = \rho_m T_2(\sec\theta_m \cos\theta)$$

$$= \rho_m \sec^2\theta_m \cos 2\theta + \rho_m(\sec^2\theta_m - 1)$$

and hence

$$\rho_0 = \tfrac{1}{2}\rho_m \sec^2\theta_m = \rho_2 = 0.099$$

$$\rho_1 = \rho_m(\sec^2\theta_m - 1) = 0.148$$

The impedances Z_1 and Z_2 are given by

$$Z_1 = e^{2\rho_0}Z_0 = 1.219 \qquad Z_2 = e^{2\rho_1}Z_1 = 1.639$$

A plot of the passband characteristic is given in Fig. 5.40. As a check on the

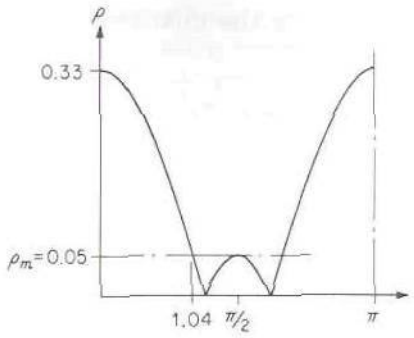


FIGURE 5.40
Passband characteristic for a two-section Chebyshev transformer with $\rho_m = 0.05$, $Z_L/Z_0 = 2$.

accuracy, we calculated Z_{in} for $\theta = 0$ using

$$Z_{in} = \frac{Z_1^2}{Z_0^2}\frac{Z_L}{Z_2^2} = 1.1063$$

From Z_{in} we obtain

$$\rho = \frac{Z_{in} - 1}{Z_{in} + 1} = 0.05047$$

which is within 1 percent of the design value.

*5.14 CHEBYSHEV TRANSFORMER (EXACT RESULTS)

An exact theory for a multisection transformer having an equal-ripple passband characteristic has also been developed (see the references at the end of this chapter). Since the analysis is rather long, only the final results for the two- and three-section transformers are given here.

In the exact theory of multisection ideal transformers, it is convenient to introduce the power loss ratio P_{LR}, which is defined as the available power (incident power) divided by the actual power delivered to the load. If the incident power is P_i, the reflected power is $\rho^2 P_i$ and the power delivered to the load is $(1 - \rho^2)P_i$. Hence

$$P_{LR} = \frac{P_i}{(1-\rho^2)P_i} = \frac{1}{1-\rho^2} \tag{5.73a}$$

and

$$\rho = \sqrt{\frac{P_{LR} - 1}{P_{LR}}} \tag{5.73b}$$

If T is the overall transmission coefficient, then $|T|^2 = 1 - \rho^2$.

For any transformer an expression for Z_{in} is readily obtained, and from this ρ, and hence P_{LR}, can be computed. When this is done it is found

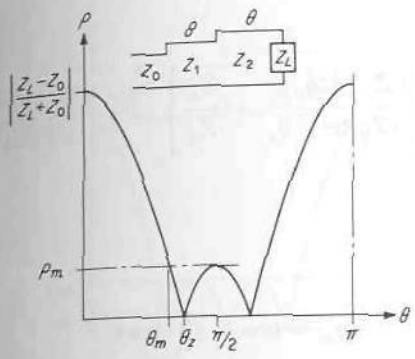


FIGURE 5.41
Passband characteristic for a two-section Chebyshev transformer.

that P_{LR} can be expressed in the form

$$P_{\text{LR}} = 1 + Q_{2N}(\cos\theta) \tag{5.74}$$

where $Q_{2N}(\cos\theta)$ is an even polynomial of degree $2N$ in $\cos\theta$, with coefficients that are functions of the various impedances Z_n. To obtain an equal-ripple characteristic, P_{LR} is now specified to be

$$P_{\text{LR}} = 1 + k^2 T_N^2(\sec\theta_m \cos\theta) \tag{5.75}$$

where k^2 is the passband tolerance on P_{LR}; that is, the maximum value of P_{LR} in the passband is $1 + k^2$, since T_N^2 has a maximum value of unity. By equating (5.74) and (5.75), algebraic equations that can be solved for the various characteristic impedances are obtained.

Figure 5.41 is a plot of ρ versus θ for a two-section transformer. For this transformer

$$P_{\text{LR}} = 1 + \frac{(Z_L - Z_0)^2}{4Z_L Z_0} \frac{(\sec^2\theta_z \cos^2\theta - 1)^2}{\tan^4\theta_z} \tag{5.76}$$

where θ_z is the value of θ at the lower zero where ρ vanishes. The maximum value of P_{LR} in the passband is

$$1 + \frac{(Z_L - Z_0)^2}{4Z_L Z_0} \cot^4\theta_z$$

and hence

$$\rho_m = \left(\frac{k^2}{1 + k^2}\right)^{1/2} \tag{5.77}$$

where $k^2 = \cot^4\theta_z(Z_L - Z_0)^2/4Z_L Z_0$. The required values of Z_1 and Z_2

are given by

$$Z_1^2 = Z_0^2\left[\frac{(Z_L - Z_0)^2}{4Z_0^2 \tan^4 \theta_z} + \frac{Z_L}{Z_0}\right]^{1/2} + \frac{(Z_L - Z_0)Z_0}{2\tan^2 \theta_z} \tag{5.78a}$$

$$Z_2 = \frac{Z_L}{Z_1} Z_0 \tag{5.78b}$$

The value of θ_m is given by

$$\theta_m = \cos^{-1}\sqrt{2}\cos\theta_z \tag{5.79a}$$

and

$$\frac{\Delta f}{f_0} = 2 - \frac{4}{\pi}\cos^{-1}\sqrt{2}\cos\theta_z \tag{5.79b}$$

provided $2\,\Delta\theta/\pi = \Delta f/f_0$ [if not, (5.79*b*) gives $2\,\Delta\theta/\pi$]. If the bandwidth is specified, then θ_z, and from (5.79*a*) θ_m, are fixed. Equation (5.77) then specifies ρ_m. On the other hand, if ρ_m is given, the bandwidth is determined.

In the limit as θ_z approaches $\pi/2$, the two zeros of ρ coalesce to give a maximally flat transformer. From (5.78) it is found that, for this case (compare with the approximate theory),

$$Z_1 = Z_L^{1/4} Z_0^{3/4} \tag{5.80a}$$

$$Z_2 = Z_L^{3/4} Z_0^{1/4} \tag{5.80b}$$

For the maximally flat transformer, the value of θ_m at the point where $\rho = \rho_m$ is given by

$$\theta_m = \cos^{-1}\cot\theta_z \tag{5.81}$$

where θ_z is the previously defined quantity for the Chebyshev transformer. Equations (5.79) and (5.81) provide a comparison of the relative bandwidths obtainable from the Chebyshev transformer and the maximally flat transformer. This comparison is illustrated in Fig. 5.42 for $N = 2$ and $N = 3$ and shows that the Chebyshev transformer can give bandwidths that are considerably greater for the same maximum tolerable value ρ_m.

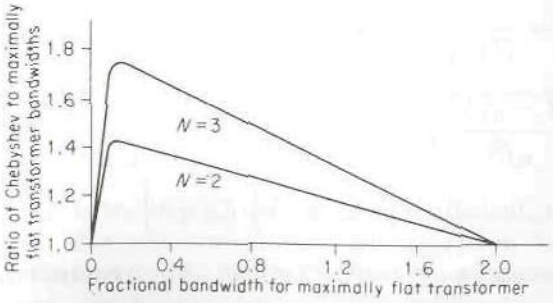


FIGURE 5.42
Comparison of bandwidths for Chebyshev and maximally flat transformers.

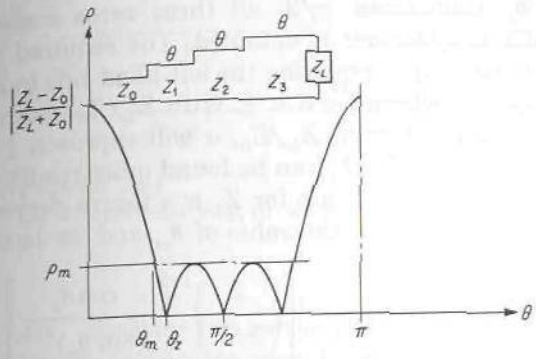


FIGURE 5.43
Passband characteristic for three-section Chebyshev transformer.

Figure 5.43 illustrates the passband characteristic for a three-section Chebyshev transformer. The power loss ratio is given by

$$P_{\mathrm{LR}} = 1 + \frac{(Z_L - Z_0)^2}{4Z_L Z_0} \frac{\left(\sec^2 \theta_z \cos^2 \theta - 1\right)^2 \cos^2 \theta}{\tan^4 \theta_z} \tag{5.82}$$

The passband tolerance k^2 is given by

$$k^2 = \frac{(Z_L - Z_0)^2}{4Z_L Z_0} \left(\frac{2 \cos \theta_z}{3\sqrt{3} \tan^2 \theta_z} \right)^2 \tag{5.83}$$

from which ρ_m may be found by using (5.77). Again the general result that specifying k^2 determines the bandwidth, and vice versa, holds. The value of θ_m is given by

$$\theta_m = \cos^{-1} \frac{2}{\sqrt{3}} \cos \theta_z \tag{5.84a}$$

and for transmission lines for which $\Delta f/f_0 = 2\,\Delta\theta/\pi$,

$$\frac{\Delta f}{f_0} = \frac{2(\pi/2 - \theta_m)}{\pi/2} = 2 - \frac{4}{\pi} \cos^{-1} \frac{2}{\sqrt{3}} \cos \theta_z \tag{5.84b}$$

The characteristic impedance Z_1 is determined by solving

$$\frac{Z_L - Z_0}{\tan^2 \theta_z} = \frac{Z_1^2}{Z_0} + 2\left(\frac{Z_L}{Z_0}\right)^{1/2} Z_1 - \frac{Z_L Z_0^2}{Z_1^2} - 2\left(\frac{Z_L}{Z_0}\right)^{1/2} Z_1^{-1} Z_0^2 \tag{5.85a}$$

and Z_2 and Z_3 are given by

$$Z_2 = (Z_L Z_0)^{1/2} \tag{5.85b}$$

$$Z_3 = \frac{Z_L Z_0}{Z_1} \tag{5.85c}$$

When θ_z approaches $\pi/2$, all three zeros coalesce at $\pi/2$, and a maximally flat transformer is obtained. The required value of Z_1 may be obtained from (5.85a) by equating the left-hand side to zero. It will be found that $Z_1 = Z_L^{\alpha} Z_0^{1-\alpha}$, where $\frac{1}{8} \leq \alpha < \frac{1}{4}$. With Z_L/Z_0 near unity, α will be close to $\frac{1}{8}$, and for large values of Z_L/Z_0, α will approach $\frac{1}{4}$. By picking various values of α, a solution for Z_1 can be found quite readily by a trial-and-error process (note that the equation for Z_1 is a fourth-degree equation). For the maximally flat transformer the value of θ_m and the bandwidth are given by

$$\theta_m = \cos^{-1}\left[\left(\frac{2}{3\sqrt{3}}\right)^{1/2} \frac{\cos\theta_z}{(\sin\theta_z)^{2/3}}\right] \tag{5.86a}$$

$$\frac{\Delta f}{f_0} = 2 - \frac{4}{\pi}\theta_m \tag{5.86b}$$

The Chebyshev transformer represents an optimum design in that no other design can give a greater bandwidth with a smaller passband tolerance. If it is assumed that some choice, other than (5.75), for the polynomial Q_{2N} in (5.74) can give a smaller passband tolerance for the same bandwidth, it will be found that a plot of the polynomial Q_{2N} will intersect the polynomial T_N^2 in at least $N + 1$ points. Since the polynomials are even in $\cos\theta$, they have at most $N + 1$ coefficients. Thus Q_{2N} must be equal to T_N^2 since they have $N + 1$ points in common. But this equality contradicts the original assumption that Q_{2N} could yield a better result, and hence proves that the Chebyshev transformer is an optimum one.

5.15 FILTER DESIGN BASED ON QUARTER-WAVE-TRANSFORMER PROTOTYPE CIRCUIT

A very interesting filter design based on the theory of multisection quarter-wave transformers was given by Young.† The quarter-wave transformer is a bandpass filter but would normally not be used as a filter since the input and output impedances are very different. In a multisection quarter-wave transformer, the impedances increase monotonically from Z_0 to Z_L. In most filter applications we desire equal input and output impedances. What Young showed was that every other impedance step in a multisection quarter-wave transformer could be replaced with an opposite impedance step. By alternating between a step up in impedance level and a step down in impedance level, we can end up with a final output impedance equal to that at the input. The filter design based on this concept is described below.

†L. Young, The Quarter-Wave Transformer Prototype Circuit, *IEEE Trans.*, vol. MTT-8, pp. 483–489, September, 1960.

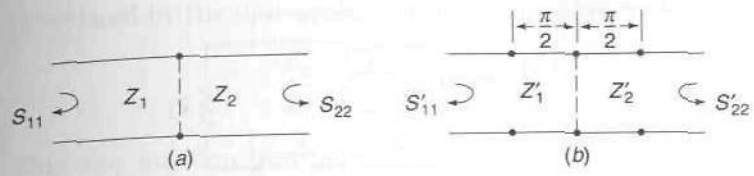


FIGURE 5.44
(*a*) An impedance step; (*b*) an equivalent junction when $Z'_2/Z'_1 = Z_1/Z_2$.

Consider the two junctions shown in Fig. 5.44. The first junction is a simple impedance step from Z_1 to Z_2. The second junction is also an impedance step from Z'_1 to Z'_2, but, in addition, it has an ideal transmission-line section of electrical length $\pi/2$ on either side. Furthermore, we assume that this electrical length does not vary with frequency. Obviously, we have introduced a nonphysical element and the reader may rightfully question whether anything useful can come from introducing such nonphysical elements. We will show that in the final filter configuration these nonphysical transmission lines can be eliminated. Thus their introduction is only to facilitate the development of the theory for the filter design. The two junctions will be fully equivalent if the scattering-matrix parameters S'_{11} and S'_{22} are the same as S_{11} and S_{22}. If the output is terminated in a matched impedance Z_2, we get

$$S_{11} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{(Z_2/Z_1) - 1}{(Z_2/Z_1) + 1}$$

for the first junction. Similarly, for a matched termination on the input side, we get

$$S_{22} = \frac{1 - (Z_2/Z_1)}{1 + (Z_2/Z_1)} = -S_{11}$$

For the second junction we use the quarter-wave-transformer formula to evaluate Z_{in} with the output matched. This gives

$$Z_{in} = \frac{(Z'_1)^2}{Z'_2}$$

from which we get

$$S'_{11} = \frac{(Z'_1)^2/Z'_2 - Z'_1}{(Z'_1)^2/Z'_2 + Z'_1}$$

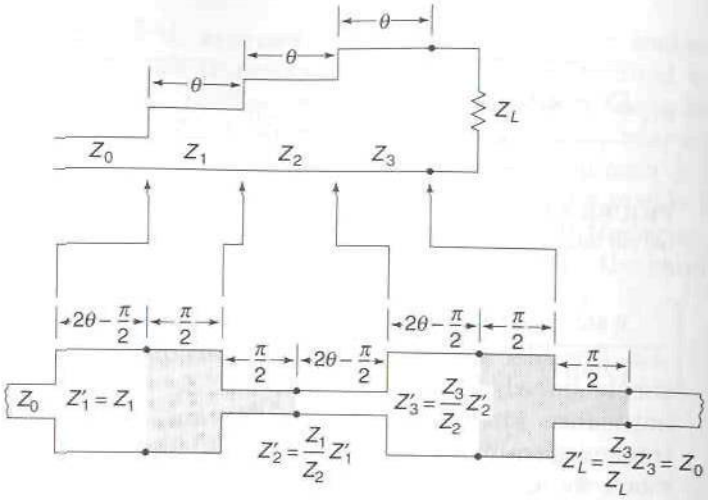


FIGURE 5.45
Microstrip filter design obtained from a multisection quarter-wave-transformer prototype circuit by replacing every other impedance step by the equivalent junction shown in Fig. 5.44*b*.

In a similar way we easily find that $S'_{22} = -S'_{11}$. By comparing the expressions for S_{11} and S'_{11}, we now conclude that the two junctions are equivalent if

$$\frac{Z_2}{Z_1} = \frac{Z_1'}{Z_2'} \tag{5.87}$$

This is precisely the property we are interested in because if Z_2/Z_1 is a step up in impedance, then Z_2'/Z_1' must be a step down in impedance level.

Our next step is now to replace every other impedance step in a multisection quarter-wave transformer by this new equivalent junction. We assume that the multisection quarter-wave transformer we use as a prototype circuit has already been designed to give a desired power loss ratio. The filter design procedure is illustrated in Fig. 5.45. In order that each section be one quarter-wave long at the center frequency, we make each transmission line have an electrical length $2\theta - \pi/2$. Thus, when $\theta = \pi/2$, $2\theta - \pi/2 = \pi/2$. In other words, we use a physical line of electrical length 2θ and a nonphysical line of electrical length $-\pi/2$ for reasons to be more fully explained later. For the first junction in the filter, we simply use

$$\frac{Z_1'}{Z_0} = \frac{Z_1}{Z_0} \tag{5.88a}$$

which is the same as in the quarter-wave transformer. The second junction

is replaced by the new equivalent junction; so we make

$$\frac{Z_2'}{Z_1'} = \frac{Z_1}{Z_2} \qquad Z_2' = \frac{Z_1 Z_1'}{Z_2} = \frac{Z_1^2}{Z_2} \tag{5.88b}$$

This new junction incorporates the two nonphysical transmission lines of electrical length $\pi/2$. The next junction is an impedance step like that in the quarter-wave transformer; so

$$\frac{Z_3'}{Z_2'} = \frac{Z_3}{Z_2} \qquad Z_3' = \frac{Z_2' Z_3}{Z_2} = \frac{Z_1^2}{Z_2^2} Z_3 \tag{5.88c}$$

The next junction is again the new equivalent junction which requires

$$\frac{Z_4'}{Z_3'} = \frac{Z_3}{Z_4} \qquad Z_4' = \frac{Z_1^2}{Z_2^2} \frac{Z_3^2}{Z_4} \tag{5.88d}$$

This procedure is continued until we have worked our way through all the impedance steps in the quarter-wave-transformer prototype circuit.

The filter illustrated in Fig. 5.45 is a three-section filter. If this filter is examined, it will be seen that each section contains a transmission line of electrical length $2\theta - \pi/2 + \pi/2 = 2\theta$, so the nonphysical lines have been eliminated. We chose each transmission line to have an electrical length of $2\theta - \pi/2$ specifically for the purpose of eliminating the nonphysical lines that are part of the new junctions that replace every other impedance step in the quarter-wave-transformer prototype circuit. However, this means that the power loss ratio of the filter is obtained by replacing θ by $2\theta - \pi/2$ in the expression for the power loss ratio for the quarter-wave transformer. If the quarter-wave transformer is a Chebyshev transformer with power loss ratio given by (5.75), then the power loss ratio for the filter is given by

$$\begin{aligned} P_{\text{LR}} &= 1 + k^2 T_N^2 \left[\sec \theta_m \cos\left(2\theta - \frac{\pi}{2}\right) \right] \\ &= 1 + k^2 T_N^2 [\sec \theta_m \sin 2\theta] \end{aligned} \tag{5.89}$$

Thus $\sin 2\theta$ replaces $\cos \theta$ as the frequency-dependent variable. At the center frequency $2\theta = \pi$, *so each filter section is one half-wavelength long.* For this reason the filter is called a half-wave filter. The band edges which occur at $\cos \theta = \pm \cos \theta_m$ *for the quarter-wave transformer now occur when* $\sec \theta_m \sin 2\theta = \pm 1$ or

$$\sin 2\theta = \pm \cos \theta_m = \pm \sin\left(\frac{\pi}{2} \pm \theta_m\right)$$

which gives

$$\beta l = 2\theta = \frac{\pi}{2} \pm \theta_m \pm n\pi$$

The upper band edge will be at $\beta l = 3\pi/2 - \theta_m$ and the lower band edge will be at $\beta l = \pi/2 + \theta_m$. Consequently, the fractional bandwidth is $\Delta(\beta l)/\pi = 1 - 2\theta_m/\pi$, which is one-half that of the quarter-wave-transformer prototype circuit. Thus the prototype circuit should be designed to have a bandwidth twice as large as that required for the filter. Apart from this change the frequency response of the filter is the same as that of the prototype circuit.

As a final point we note that, in a quarter-wave transformer with an odd number of sections, the power loss ratio is unity at the center frequency. Since

$$Z_{\text{in}} = \frac{Z_L}{Z_N^2}\frac{Z_{N-1}^2}{Z_{N-2}^2}\cdots\frac{Z_2^2}{Z_1^2} = Z_0$$

for an odd number of sections, the final impedance Z_L' in the filter equals Z_0. Hence the filter operates between impedance levels of Z_0 and Z_0. The last nonphysical transmission line that appears in the output line (see Fig. 5.45) for an odd number of sections can be deleted since this is a matched line of arbitrary length.

For an even number of sections, Z_{in} does not equal Z_0 at the center frequency because the power loss ratio does not equal unity; it equals $1 + k^2$ for this case. This means that the last impedance element in the filter, which is given by

$$Z_L' = Z_0' = \frac{Z_1^2}{Z_2^2}\frac{Z_3^2}{Z_4^2}\cdots\frac{Z_{N-1}^2}{Z_N^2}Z_L = Z_{\text{in}}$$

will be different from Z_0. By using (5.77) for ρ at the center frequency, we can solve for Z_{in} to get

$$Z_{\text{in}} = Z_0\frac{1+\rho}{1-\rho} = \left(k + \sqrt{1+k^2}\right)^2 Z_0 \tag{5.90}$$

Since k is normally very small, the output line for the filter has a characteristic impedance not quite equal to that of the input line.

In Table 5.1 we list the required values of the impedances for a Chebyshev quarter-wave transformer with three sections for several values of the passband tolerance k^2. We will use this table to design a three-section filter in the example that follows. More extensive tables are available in the literature.†

†See, for example, G. L. Matthaei, L. Young, and E. M. T. Jones, "Microwave Filters, Impedance Matching Networks, and Coupling Structures," Artech House Books, Dedham, Mass. 1980.

TABLE 5.1
Chebyshev quarter-wave-transformer data

	$\Delta f/f_0 = 0.2$		$\Delta f/f_0 = 0.4$		$\Delta f/f_0 = 0.6$	
Z_L/Z_0	Z_1/Z_0	k^2	Z_1/Z_0	k^2	Z_1/Z_0	k^2
2	1.09247	1.19×10^{-7}	1.09908	7.89×10^{-6}	1.1083	9.57×10^{-5}
4	1.19474	5.35×10^{-7}	1.20746	3.55×10^{-5}	1.23087	4.31×10^{-4}
10	1.349	1.92×10^{-7}	1.37482	1.28×10^{-4}	1.4232	1.55×10^{-3}
20	1.48359	4.29×10^{-7}	1.52371	2.85×10^{-4}	1.60023	3.45×10^{-3}
100	1.87411	2.33×10^{-6}	1.975	1.55×10^{-3}	2.17928	1.87×10^{-2}

$Z_2 = \sqrt{Z_L Z_0}$ $\quad Z_3 = Z_L Z_0 / Z_1$

Example 5.3 Filter design. *We want to design a bandpass filter with a* fractional bandwidth of 0.2 and having a VSWR of 1.02 or less in the passband. From the given VSWR we find $\rho_m = (1.02 - 1)/2.02 = 9.9 \times 10^{-3}$. By using (5.77) we get $k^2 = \rho_m^2/(1 - \rho_m^2) = 9.8 \times 10^{-5}$. The quarter-wave-transformer bandwidth must be chosen as $2 \times 0.2 = 0.4$. From Table 5.1 we find that, for $Z_L/Z_0 = 10$ and $\Delta f/f_0 = 0.4$, $k^2 = 1.28 \times 10^{-4}$. This value of k^2 would give a maximum value of VSWR equal to 1.023 in the passband. We will accept this value since an entry for $k^2 = 9.8 \times 10^{-5}$ is not given. From Table 5.1 we get $Z_1/Z_0 = 1.37482$. We will use 50-Ω input and output lines. Thus $Z_1 = 1.3748 \times 50 = 68.74$ Ω. The required values of Z_2 and Z_3 are $Z_2 = \sqrt{500 \times 50} = 158.1$ Ω, $Z_3 = 500 \times 50/68.74 = 363.69$ Ω. This completes the design of the quarter-wave-transformer prototype circuit. For the filter we need $Z_1' = Z_1 = 68.74$ Ω, and by using (5.88*b*) and (5.88*c*), $Z_2' = 29.89$ Ω, $Z_3' = 68.74$ Ω. These impedance values are readily realized for a microstrip filter.

The maximum out-of-band attenuation occurs at $\theta = 0$ *and is the same* as what is obtained when the input line is connected directly to a load $Z_L = 10Z_0$. *The reflection coefficient will be* $9/11$ *so 66.9 percent of the* incident power is reflected. The transmission through the filter is reduced by $-10 \log(1 - \rho^2) = 4.8$ dB. This is a small attenuation and shows that the particular filter configuration used here will generally not have a large out-of-band attenuation. If we use $Z_L/Z_0 = 100$, we would obtain an attenuation of 14 dB. But in this case the passband tolerance would be larger. The required passband tolerance could be achieved by using a filter with more sections. However, *there are better filter configurations to use when large out-of-band* attenuation is needed (see Chap. 8).

Junction Capacitance and Length Compensation

At an abrupt step in the width of a microstrip line, there will be an additional fringing electric field from the open-circuited portions of the wider strip as shown pictorially in Fig. 5.46*a*. The effect of this fringing field can be modeled as a shunt capacitance at the junction. The equivalent circuit of the step is shown in Fig. 5.46*b* and consists of an ideal impedance

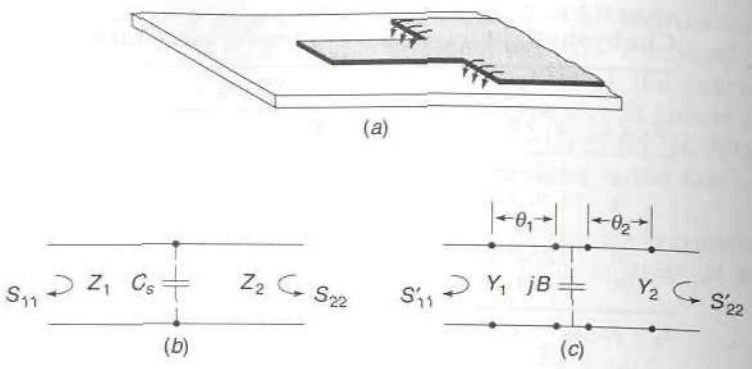


FIGURE 5.46
(a) Fringing electric field at a step change in width for a microstrip line; (b) equivalent circuit for a step-change-in-width junction; (c) an alternative equivalent circuit.

step and a shunt capacitive susceptance. The susceptance is small but does modify the performance of the filter if not properly compensated. The junction capacitance can be compensated for by changing the length of each filter section. The equivalent circuit in Fig. 5.46b will be shown to be nearly equivalent to the circuit in Fig. 5.46c. The latter consists of the ideal impedance step plus two short lengths of transmission line. The electrical lengths of these transmission lines are denoted by θ_1 and θ_2. The two electrical lengths are of opposite sign.

For the equivalent circuit in Fig. 5.46b, we have

$$S_{11} = \frac{Y_1 - Y_2 - jB}{Y_1 + Y_2 + jB} = \frac{Y_1 - Y_2}{Y_1 + Y_2} \frac{1 - \dfrac{jB}{Y_1 - Y_2}}{1 + \dfrac{jB}{Y_1 + Y_2}}$$

Since $B/(Y_1 + Y_2)$ is very small, we can use the binomial expansion to obtain

$$\begin{aligned} S_{11} &\approx \frac{Y_1 - Y_2}{Y_1 + Y_2}\left(1 - j\frac{B}{Y_1 - Y_2}\right)\left(1 - j\frac{B}{Y_1 + Y_2}\right) \\ &\approx \frac{Y_1 - Y_2}{Y_1 + Y_2}\left(1 - \frac{jB}{Y_1 - Y_2} - j\frac{B}{Y_1 + Y_2}\right) \\ &= \frac{Y_1 - Y_2}{Y_1 + Y_2}\left(1 - \frac{2jBY_1}{Y_1^2 - Y_2^2}\right) \end{aligned} \tag{5.91}$$

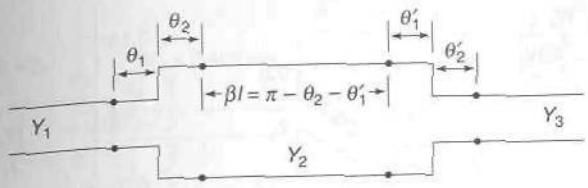


FIGURE 5.47
Length compensation of a filter section.

For the circuit in Fig. 5.46*c*, we have

$$S'_{11} = \frac{Y_1 - Y_2}{Y_1 + Y_2} e^{-2j\theta_1} \approx \frac{Y_1 - Y_2}{Y_1 + Y_2}(1 - 2j\theta_1) \tag{5.92}$$

when θ_1 is small. In a similar way we obtain

$$S_{22} \approx \frac{Y_2 - Y_1}{Y_2 + Y_1}\left(1 - \frac{2jBY_2}{Y_2^2 - Y_1^2}\right) \tag{5.93}$$

$$S'_{22} \approx \frac{Y_2 - Y_1}{Y_2 + Y_1}(1 - 2j\theta_2) \tag{5.94}$$

A comparison of S_{11} and S_{22} with S'_{11} and S'_{22} shows that the two junctions are equivalent, to the order of approximations used, if we choose

$$\theta_1 = \frac{BY_1}{Y_1^2 - Y_2^2} \tag{5.95a}$$

$$\theta_2 = -\frac{BY_2}{Y_1^2 - Y_2^2} \tag{5.95b}$$

In the half-wave filter the length of each section is now changed to compensate for the junction capacitance as shown in Fig. 5.47. The required section length is 2θ including θ_2 and θ'_1. Hence the physical length l is chosen to make

$$\beta l = 2\theta - \theta_2 - \theta'_1 = \pi - \theta_2 - \theta'_1 \tag{5.96}$$

at the center frequency. Since θ_1 and θ_2, etc., vary with frequency like $B = \omega C$, compensation is obtained at all frequencies. The parameters θ_2 and θ'_1 are small so the change in physical length of each section is small but important enough to take into account in the design of a filter.

Similar junction capacitance effects also occur in some quarter-wave-transformer realizations and can be compensated for by changing the physical length of each section by a small amount.

The application of the above method of compensation will be illustrated in Example 5.4, but first we need data for the junction capacitance. A number of investigations of junction equivalent circuits for microstrip discontinuities have been carried out, but surprisingly very little data for design purposes are available. Gupta and Gopinath have evaluated the

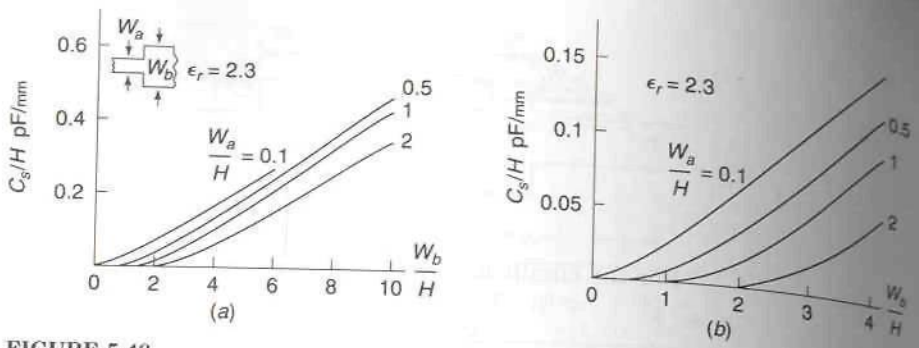


FIGURE 5.48
(*a*) Junction capacitance C_s for a step change in width for a substrate with $\epsilon_r = 2.3$; (*b*) enlarged portion of curves in (*a*).

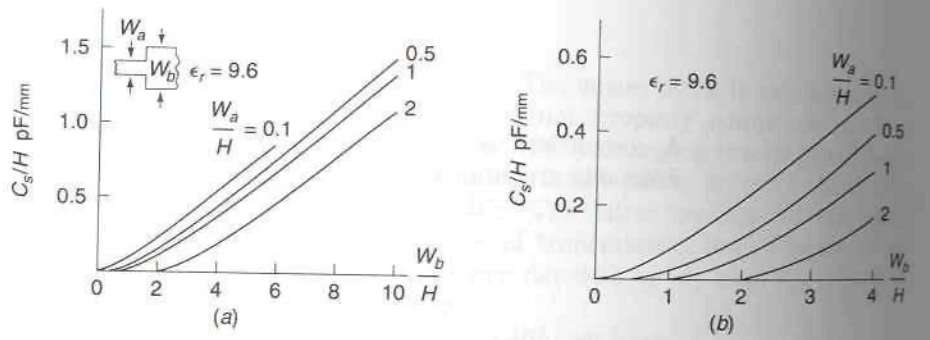


FIGURE 5.49
(*a*) Junction capacitance C_s for a step change in width for a substrate with $\epsilon_r = 9.6$; (*b*) enlarged portion of curves in (*a*).

shunt capacitance at a step in width for a microstrip line for substrate dielectric constants of 2.3, 4, 9.6, and 15.1.† Their data give the shunt capacitance normalized by the distributed capacitance per meter of the wider microstrip line. We have used these data to evaluate the shunt capacitance for the two cases where the dielectric constant of the substrate is 2.3 and 9.6. The junction capacitance C_s divided by the substrate thickness H is shown in Figs. 5.48 and 5.49.

Example 5.4 Microstrip half-wave filter. The filter described in Example 5.3 is to be built using microstrip construction. The substrate is alumina with

†C. Gupta and A. Gopinath, Equivalent Circuit Capacitance of Microstrip Change in Width, *IEEE Trans.*, vol. MTT-25, pp. 819–822, October, 1977.

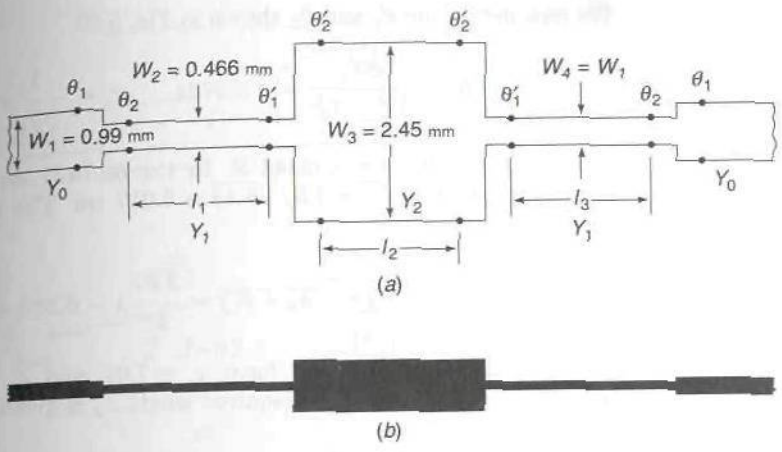


FIGURE 5.50
(*a*) Three-section half-wave filter using microstrip construction; (*b*) scaled drawing of filter.

a dielectric constant of 9.6 and is 1 mm thick. By using the computer program MSTP, we find after a few iterations the following required widths and the effective dielectric constants of each section:

$$\text{For } Z_c = 50\ \Omega \qquad W = 0.99 \text{ mm} \qquad \epsilon_e = 6.49$$
$$\text{For } Z_c = 68.74\ \Omega \qquad W = 0.466 \text{ mm} \qquad \epsilon_e = 6.18$$
$$\text{For } Z_c = 29.89\ \Omega \qquad W = 2.45 \text{ mm} \qquad \epsilon_e = 7.07$$

With reference to Fig. 5.49, we identify W_a/H and W_b/H for the junction between the 50- and 68.74-Ω lines to be

$$\frac{W_a}{H} = 0.466 \qquad \frac{W_b}{H} = 0.99$$

From Fig. 5.49*b* we estimate C_s/H to be 0.03 and since $H = 1$ mm, $C_s = 0.03$ pF. The center frequency of the filter passband is to be 4 GHz. Hence $B = \omega C_s = 7.5 \times 10^{-4}$ S. The characteristic admittances Y_0 and Y_1 are (see Fig. 5.50)

$$Y_0 = \frac{1}{50} = 0.02 \qquad Y_1 = \frac{1}{68.74} = 0.01455$$

We now use (5.95) to obtain

$$\theta_1 = \frac{BY_0}{Y_0^2 - Y_1^2} = 0.08 \qquad \theta_2 = -\frac{Y_1}{Y_0}\theta_1 = -0.058$$

For the junction between the 68.74-Ω line and the 29.89-Ω line, we have

$$\frac{W_a}{H} = 0.466 \qquad \frac{W_b}{H} = 2.45$$

From Fig. 5.49*b* we estimate C_s to be 0.18 pF which gives $B = 4.52 \times 10^{-3}$ S.

We now determine θ_1' and θ_2' shown in Fig. 5.50.

$$\theta_1' = \frac{BY_1}{Y_1^2 - Y_3^2} = -0.0724 \qquad \theta_2' = -\frac{Y_3}{Y_1}\theta_1' = 0.167$$

where $Y_3 = 1/29.89 = 0.03346$ S. In the 68.74-Ω section $\epsilon_e = 6.18$ so the wavelength is $\lambda_0/\sqrt{\epsilon_e} = 7.5/\sqrt{6.18} = 3.017$ cm. The required length l_1 for this section is

$$l_1 = \frac{\lambda}{2\pi}(\pi - \theta_2 - \theta_1') = \frac{3.27}{2\pi}\lambda = 0.52\lambda = 1.57 \text{ cm}$$

For the 29.89-Ω line, we have $\epsilon_e = 7.07$ and a wavelength equal to $7.5/\sqrt{7.07} = 2.821$ cm. The required length l_2 is given by

$$l_2 = \frac{\lambda}{2\pi}(\pi - 2\theta_2') = 0.447\lambda = 1.261 \text{ cm}$$

Note that the length corrections are about 4 and 10 percent. If the filter were designed for 2-GHz operation, the length corrections would be half as large or about 2 and 5 percent. Below 1 GHz, compensation for the junction susceptances could be ignored. A proportionally scaled drawing of the foil pattern for the filter is shown in Fig. 5.50*b*.

5.16 TAPERED TRANSMISSION LINES

In a multisection quarter-wave transformer used to match two transmission lines with different characteristic impedances, the change in impedance level is obtained in a number of discrete steps. An alternative is to use a tapered transition which has a characteristic impedance that varies continuously in a smooth fashion from the impedance of one line to that of the other line. A transition, or matching section, of this type is referred to as a tapered transmission line. An approximate theory of tapered transmission lines, analogous to the approximate theory presented earlier for multisection transformers, is readily developed. This approximate theory is presented below. A following section gives a derivation of the exact differential equation for the reflection coefficient on a tapered transmission line and also gives a brief evaluation of the validity of the approximate theory.

Figure 5.51*a* illustrates schematically a tapered transmission line used to match a line with normalized impedance unity to a load with normalized impedance $\bar{Z}_L$ (assumed to be a pure resistive load). The taper has a normalized impedance $\bar{Z}$ which is a function of the distance z along the taper. Figure 5.51*b* illustrates an approximation to the continuous taper by considering it to be made up of a number of sections of line of differential length dz and for which the impedance changes by differential amounts $d\bar{Z}$ from section to section.

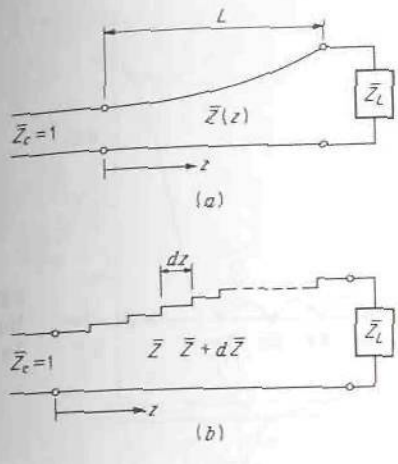


FIGURE 5.51
Tapered-transmission-line matching section.

The step change $d\bar{Z}$ in impedance at z produces a differential reflection coefficient

$$d\Gamma_0 = \frac{\bar{Z} + d\bar{Z} - \bar{Z}}{\bar{Z} + d\bar{Z} + \bar{Z}} \approx \frac{d\bar{Z}}{2\bar{Z}} = \frac{1}{2} d(\ln \bar{Z})$$

$$= \frac{1}{2}\frac{d}{dz}(\ln \bar{Z})\, dz \tag{5.97}$$

At the input to the taper, the contribution to the input reflection coefficient from this step is

$$d\Gamma_i = e^{-j2\beta z}\frac{1}{2}\frac{d}{dz}(\ln \bar{Z})\, dz$$

If it is assumed that the total reflection coefficient can be computed by summing up all the individual contributions, as was done in the approximate theory of the multisection quarter-wave transformer, the input reflection coefficient is given by

$$\Gamma_i = \frac{1}{2}\int_0^L e^{-j2\beta z}\frac{d}{dz}(\ln \bar{Z})\, dz \tag{5.98}$$

where L is the total taper length. If the variation in $\bar{Z}$ with z is known, Γ_i may be readily evaluated from the above. A problem of much greater practical importance is the synthesis problem, where $\bar{Z}(z)$ is to be determined to give Γ_i the desired characteristics as a function of frequency. Before taking up the synthesis problem, two examples of practical taper designs are presented.

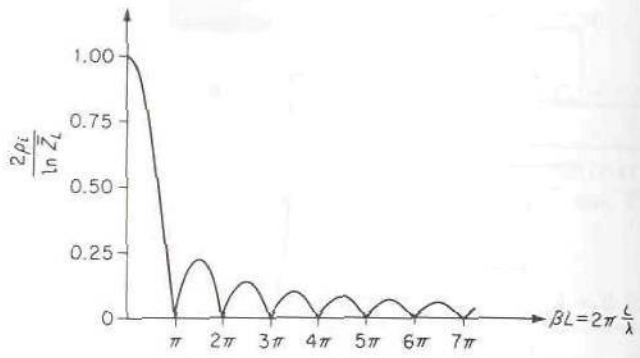


FIGURE 5.52
Input reflection coefficient for an exponential taper.

Exponential Taper

The exponential taper is one for which $\ln \bar{Z}$ varies linearly, and hence $\bar{Z}$ varies exponentially, from unity to $\ln \bar{Z}_L$; that is,

$$\ln \bar{Z} = \frac{z}{L} \ln \bar{Z}_L \tag{5.99a}$$

$$\bar{Z} = e^{(z/L)\ln \bar{Z}_L} \tag{5.99b}$$

Substituting (5.99) into (5.98) gives

$$\Gamma_i = \frac{1}{2}\int_0^L \frac{\ln \bar{Z}_L}{L} e^{-2j\beta z}\, dz = \frac{1}{2} e^{-j\beta L} \ln \bar{Z}_L \frac{\sin \beta L}{\beta L} \tag{5.100}$$

where it has been assumed that we are dealing with a transmission line for which $\beta = k = 2\pi/\lambda$ and is not a function of z. A plot of $\rho_i = |\Gamma_i|$ versus βL is given in Fig. 5.52. For a fixed length of taper, this is a plot of ρ_i as a function of frequency since $k = 2\pi f(\mu\epsilon)^{1/2}$. Note that when L is greater than $\lambda/2$, the reflection coefficient is quite small, the first minor lobe being about 22 percent of the major-lobe maximum.

Taper with Triangular Distribution

If $d(\ln \bar{Z})/dz$ is chosen as a triangular function of the form

$$\frac{d(\ln \bar{Z})}{dz} = \begin{cases} \dfrac{4z}{L^2} \ln \bar{Z}_L & 0 \le z \le \dfrac{L}{2} \\ \dfrac{4}{L^2}(L - z)\ln \bar{Z}_L & \dfrac{L}{2} \le z \le L \end{cases} \tag{5.101}$$

a matching section with more desirable properties is obtained. Integrating

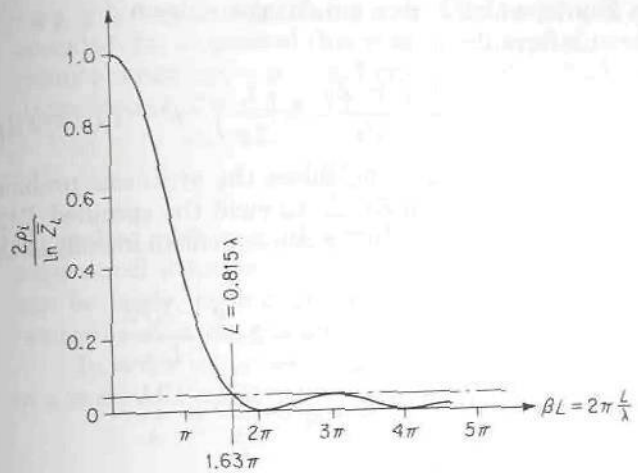


FIGURE 5.53
Input reflection coefficient for a taper with a triangular distribution of reflections.

(5.101) gives

$$\bar{Z} = \begin{cases} e^{[2(z/L)^2]\ln \bar{Z}_L} & 0 \le z \le \dfrac{L}{2} \\ e^{(4z/L - 2z^2/L^2 - 1)\ln \bar{Z}_L} & \dfrac{L}{2} \le z \le L \end{cases} \tag{5.102}$$

Substituting (5.101) into (5.98) and performing the straightforward integration give

$$\Gamma_i = \frac{1}{2} e^{-j\beta L} \ln \bar{Z}_L \left[\frac{\sin(\beta L/2)}{\beta L/2} \right]^2 \tag{5.103}$$

A plot of ρ_i versus βL is given in Fig. 5.53. Note that, by comparison with the exponential taper, this taper has a first minor-lobe maximum which is less than 5 percent of the major-lobe peak. However, this small value of reflection coefficient occurs for a taper length of about $3\lambda/2$, or for a length twice that for the exponential taper. If $\bar{Z}_L$ is considerably greater than unity, this latter taper will be preferable because of the much smaller values of ρ_i obtained for all frequencies, such that the taper length is greater than 0.815λ, which corresponds to the lower edge of the passband in Fig. 5.53.

*5.17 SYNTHESIS OF TRANSMISSION-LINE TAPERS

Equation (5.98) is repeated here for convenience:

$$\Gamma_i(2\beta) = \frac{1}{2} \int_0^L e^{-j2\beta z} \frac{d(\ln \bar{Z})}{dz} \, dz \tag{5.104}$$

This equation may be interpreted as the Fourier transform of a function

$d(\ln \bar{Z})/dz$, which is zero outside the range $0 \le z \le L$.† As such the Fourier inversion formula gives

$$\frac{1}{2}\frac{d(\ln \bar{Z})}{dz} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j2\beta z}\Gamma_i(2\beta)2\,d\beta \tag{5.105}$$

This formula, in principle, solves the synthesis problem since it gives the required value of $d(\ln \bar{Z})/dz$ to yield the specified $\Gamma_i(2\beta)$. To simplify the discussion to follow, it will be convenient to introduce the following normalized variables:

$$p = 2\pi\frac{z - L/2}{L} \tag{5.106a}$$

$$u = \frac{\beta L}{\pi} = \frac{2L}{\lambda} \tag{5.106b}$$

In this case (5.104) becomes

$$\Gamma_i = \frac{1}{2}e^{-j\beta L}\int_{-\pi}^{\pi} e^{-jpu}\frac{d(\ln \bar{Z})}{dp}\,dp \tag{5.107}$$

Now define $g(p)$ to be

$$g(p) = \frac{d(\ln \bar{Z})}{dp} \tag{5.108a}$$

and $F(u)$ by

$$F(u) = \int_{-\pi}^{\pi} e^{-jpu}g(p)\,dp \tag{5.108b}$$

Thus

$$\Gamma_i = \tfrac{1}{2}e^{-j\beta L}F(u) \tag{5.109}$$

The Fourier transform pair (5.104) and (5.105) now may be expressed as

$$F(u) = \int_{-\pi}^{\pi} e^{-jpu}g(p)\,dp \tag{5.110a}$$

$$g(p) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{jpu}F(u)\,du \tag{5.110b}$$

The synthesis problem may now be stated as follows: Specify a reflection-coefficient characteristic $F(u)$ that will give the desired taper performance and yet be such that the $g(p)$ computed by (5.110*b*) will be a function identically zero outside the range $|p| > \pi$. This latter restriction corresponds to the physical requirement that $d(\ln \bar{Z})/dz$ be different from

†The Fourier transform relation was first pointed out by F. Bolinder, Fourier Transforms in the Theory of Inhomogeneous Transmission Lines, *Proc. IRE*, vol. 38, p. 1354, November, 1950.

zero only in the range $0 \le z \le L$. Obviously, any arbitrary $F(u)$ cannot be specified, for in general this would lead to a $g(p)$ that exists over the whole infinite range $-\infty < p < \infty$. For example, if $F(u)$ were chosen to be equal to unity for $-1 \le u \le 1$ and zero otherwise, then (5.110b) would give

$$g(p) = \frac{\sin p}{\pi p} \qquad -\infty < p < \infty$$

To realize such a $g(p)$ would require an infinitely long taper, clearly an impractical solution. Before further progress with the synthesis problem can be made, restrictions to be imposed on $F(u)$, to obtain a physically realizable solution, must be deduced.

In order to derive suitable restrictions on $F(u)$, let $g(p)$ be expanded in a complex Fourier series as follows:

$$g(p) = \begin{cases} \sum_{n=-\infty}^{\infty} a_n e^{jnp} & -\pi \le p \le \pi \\ 0 & |p| > \pi \end{cases} \tag{5.111}$$

where the a_n are as yet unspecified coefficients subject to the restriction $a_n = a^*_{-n}$ so that g will be a real function. Substitution into (5.110a) now gives

$$\begin{aligned} F(u) &= 2\pi \sum_{n=-\infty}^{\infty} a_n \frac{\sin \pi(u-n)}{\pi(u-n)} \\ &= 2\pi \sum_{n=-\infty}^{\infty} a_n(-1)^n \frac{\sin \pi u}{\pi(u-n)} = 2\pi \frac{\sin \pi u}{\pi u} \sum_{n=-\infty}^{\infty} a_n(-1)^n \frac{u}{u-n} \end{aligned} \tag{5.112}$$

The coefficients a_n can be related to $F(u=n)$, for when u equals an integer n,

$$\lim_{u \to n} \frac{\sin \pi(u-n)}{\pi(u-n)} = 1$$

and

$$\lim_{u \to n} \frac{\sin \pi(u-m)}{\pi(u-m)} = 0$$

Thus $F(n) = 2\pi a_n$, or $a_n = F(n)/2\pi$; so

$$F(u) = \sum_{n=-\infty}^{\infty} F(n) \frac{\sin \pi(u-n)}{\pi(u-n)} \tag{5.113}$$

This result is a statement of the well-known sampling theorem used in communication theory and states that $F(u)$ is uniquely reconstructed from a knowledge of the sample values of $F(u)$ at $u = n$,

$$n = 0, \pm 1, \pm 2, \ldots$$

by means of the interpolation formula (5.113). One possible way to restrict $F(u)$ is now seen to be a relaxation on the specification of $F(u)$; that is, specify $F(u)$ at all integer values of u only. This, however, is not an entirely satisfactory solution, because we have no a priori knowledge that if we specify $F(u)$ at the integer values of u only, the resultant $F(u)$ given by (5.113) will be an acceptable reflection-coefficient characteristic for all values of u, even though it can be realized by a $g(p)$ given by (5.111) with $a_n = F(n)/2\pi$.

We should like to obtain greater flexibility in the choice of $F(u)$, and to see how this may be accomplished, let it be assumed that all a_n for $|n| > N$ are zero. In this case

$$g(p) = \sum_{n=-N}^{N} a_n e^{jnp} \tag{5.114a}$$

$$F(u) = 2\pi \frac{\sin \pi u}{\pi u} \sum_{n=-N}^{N} (-1)^n a_n \frac{u}{u-n} \tag{5.114b}$$

The series in (5.114b) can be recognized as the partial-fraction expansion of a function of the form

$$\frac{Q(u)}{\prod_{n=1}^{N}(u^2 - n^2)}$$

where $Q(u)$ is, apart from the restriction $Q(-u) = Q^*(u)$ so that $a_n = a^*_{-n}$, an arbitrary polynomial of degree $2N$ in u, and the denominator is the product of the N terms $(u^2 - 1), (u^2 - 4)$, etc. Using the partial-fraction-expansion formula, we have

$$\begin{aligned}\frac{Q(u)}{\prod_{n=1}^{N}(u^2 - n^2)} &= \sum_{m=-N}{}' \frac{Q(m)}{(u-m)2m\prod_{n=1,\, n\neq m}^{N}(m^2 - n^2)} + \lim_{u\to\infty} \frac{Q(u)}{u^{2N}} \\ &= \frac{uQ(u)}{u\prod_{n=1}^{N}(u^2 - n^2)} = \sum_{m=-N}{}' \frac{uQ(m)}{(u-m)2m^2\prod_{n=1,\, n\neq m}^{N}(m^2 - n^2)} \\ &\quad + \frac{Q(0)}{\prod_{n=1}^{N}(-n^2)}\end{aligned} \tag{5.115}$$

where the prime means omission of the term $m = 0$. This is clearly of the same form as the series in (5.114b), with

$$(-1)^m a_m = \frac{Q(m)}{2m^2\prod_{n=1,\, n\neq m}^{N}(m^2 - n^2)} \tag{5.116a}$$

$$a_0 = \frac{Q(0)}{\prod_{n=1}^{N}(-n^2)} \tag{5.116b}$$

The expression for $F(u)$ now can be written as

$$F(u) = 2\pi \frac{\sin \pi u}{\pi u} \frac{Q(u)}{\prod_{n=1}^{N}(u^2 - n^2)} \tag{5.117}$$

where $Q(u)$ is an arbitrary polynomial of degree $2N$ in u, subject to the restriction $Q(-u) = Q^*(u)$. This result states that the first $2N$ zeros of $\sin \pi u$, which are canceled by the denominator in (5.117), can be replaced by $2N$ new arbitrarily located zeros by proper choice of $Q(u)$. If $g(p)$ were a constant (exponential taper), we should have $F(u)$ proportional to $(\sin \pi u)/\pi u$. But with $2N + 1$ coefficients available in the expansion of $g(p)$, we are at liberty to rearrange $2N$ of the zeros of $(\sin \pi u)/\pi u$ to obtain a more desirable $F(u)$. We have now reduced the synthesis problem to one of specifying an arbitrary polynomial $Q(u)$. To illustrate the theory, two examples are discussed below.

A qualitative insight into how $Q(u)$ should be specified may be obtained by imagining that $F(u)$ is a rubber band stretched horizontally at some height above the βL or u axis. The zeros of $F(u)$ may then be thought of as points at which this rubber band is pinned down to the u axis. If the band is pinned down at a number of closely spaced points, it will not rise much above the u axis in the regions between. The corresponding reflection coefficient will then also be small in this region. At a double zero the band is pinned down in such a fashion that its slope is zero at the point as well. This results in a less rapid increase in the height with distance away from the point. The number of zeros available (the points at which the band may be pinned down) is fixed and equal to those in the $\sin \pi u$ function. The polynomial $Q(u)$ permits only a relocation of these zeros.

With reference to Fig. 5.52, which illustrates the characteristic for an exponential taper, let the zero at $\beta L = \pi$ be moved to 2π to form a double zero at this point. Likewise, let the zero at 3π be moved to form a double zero at 4π, etc. The function $Q(u)$ that will provide this shift in every other zero so as to produce double zeros at $u = \pm 2, \pm 4, \pm 6, \pm 8$, etc., can be chosen as

$$Q(u) = \prod_{n=1}^{N} (u^2 - 4n^2)^2$$

From (5.117) we obtain

$$F(u) = 2\pi \frac{\sin \pi u}{\pi u} \frac{\prod_{n=1}^{N}(u^2 - 4n^2)^2}{\prod_{n=1}^{N}(u^2 - n^2)}$$

We now wish to let N go to infinity. However, the products do not converge in this case; so we must modify the expression for $F(u)$ to the following:

$$F(u) = C \frac{\sin \pi u}{\pi u} \frac{\prod_{n=1}^{N}(1 - u^2/4n^2)^2}{\prod_{n=1}^{N}(1 - u^2/n^2)}$$

This modification is permissible since Q contains an arbitrary constant multiplier. All constants can be incorporated in the one constant C. If the following infinite-product representations for the sine functions are noted,

$$\frac{\sin \pi u}{\pi u} = \prod_{n=1}^{\infty} \left(1 - \frac{u^2}{n^2}\right)$$

$$\left[\frac{\sin(\pi u/2)}{\pi u/2}\right]^2 = \prod_{n=1}^{\infty} \left(1 - \frac{u^2}{4n^2}\right)^2$$

it is readily seen that as $N \to \infty$ we obtain

$$F(u) = C\left[\frac{\sin(\pi u/2)}{\pi u/2}\right]^2 \tag{5.118}$$

This is the reflection-coefficient characteristic for the taper with a triangular $d(\ln \bar{Z})/dz$ function discussed earlier. In the present case we have arrived at this solution for a taper with equally spaced double zeros by a direct-synthesis procedure. As Fig. 5.53 shows, the specification of double zeros holds the values of $F(u)$ (that is, ρ_t) to much smaller values in the region between zeros.

As shown earlier, the coefficients a_n in the Fourier series expansion of $g(p)$ are given by

$$a_n = \frac{1}{2\pi} F(n) = \frac{C}{2\pi}\left[\frac{\sin(n\pi/2)}{n\pi/2}\right]^2 \tag{5.119}$$

from (5.118). The reader may readily verify that the expansion (5.114a) for $g(p)$, with the above coefficients, is a triangular wave. To fix the constant C, we integrate (5.108a) to obtain

$$\int_{-\pi}^{\pi} g(p)\, dp = \int_{-\pi}^{\pi} \frac{d(\ln \bar{Z})}{dp}\, dp = \ln \bar{Z}\Big|_{-\pi}^{\pi} = \ln \bar{Z}_L$$

But from (5.111) we have

$$\int_{-\pi}^{\pi} g(p)\, dp = 2\pi a_0$$

and hence

$$a_0 = \frac{1}{2\pi} \ln \bar{Z}_L \tag{5.120}$$

From (5.119) we now find that

$$a_0 = \frac{1}{2\pi} \ln \bar{Z}_L = \frac{C}{2\pi}$$

so $C = \ln \bar{Z}_L$. With this value of C, the reflection coefficient corresponding to the $F(u)$ in (5.118) is easily verified to be the same as that given by (5.103).

As a second example, consider the synthesis of a taper with an $F(u)$ having a triple zero at $u = \pm 2$. This can be accomplished by moving the

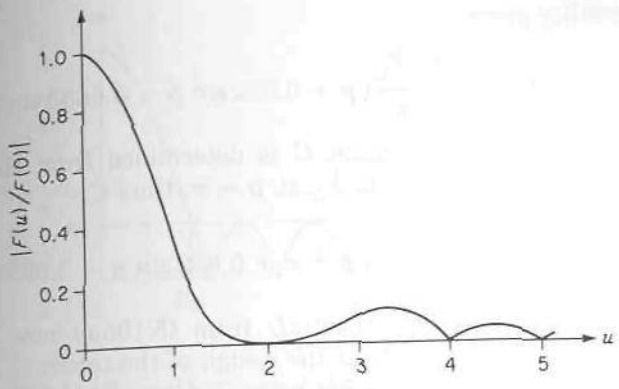


FIGURE 5.54
Reflection-coefficient characteristic for a taper with a triple zero at $|u| = 2$.

zeros at ± 1 and ± 3 *into the points* $u = \pm 2$. *The resultant reflection* coefficient should remain very small for a considerable region around $|u| = 2$. In the present case $N = 3$, and we choose

$$Q(u) = C(u^2 - 4)^3$$

Thus
$$F(u) = 2\pi C \frac{\sin \pi u}{\pi u} \frac{(u^2 - 4)^3}{(u^2 - 1)(u^2 - 4)(u^2 - 9)}$$

A plot of $|F(u)/F(0)|$ is given in Fig. 5.54. As anticipated, $F(u)$ remains small in a considerable region around the point $|u| = 2$. Since the zeros at $|u| = 3$ have been removed, $F(u)$ reaches a relatively large value at this point. However, for a range of frequencies around which $L \approx \lambda$, this taper represents a very good design.

The coefficients in the expansion for $g(p)$ are given by

$$a_0 = \frac{1}{2\pi} F(0) = \frac{1}{2\pi} \ln \bar{Z}_L$$

$$a_1 = a_{-1} = \frac{1}{2\pi} F(1) = \frac{0.316}{2\pi} \ln \bar{Z}_L$$

$$a_2 = a_{-2} = 0$$

$$a_3 = a_{-3} = \frac{1}{2\pi} F(3) = -\frac{0.098}{2\pi} \ln \bar{Z}_L$$

$$a_n = a_{-n} = 0 \qquad n > 3$$

Hence

$$g(p) = \frac{d(\ln \bar{Z})}{dp} = \frac{\ln \bar{Z}_L}{2\pi}(a_0 + 2a_1 \cos p + 2a_3 \cos 3p)$$

$$= \frac{\ln \bar{Z}_L}{2\pi}(1 + 0.632 \cos p - 0.196 \cos 3p)$$

Integrating gives

$$\ln \bar{Z} = \frac{\ln \bar{Z}_L}{2\pi}(p + 0.632 \sin p - 0.0653 \sin 3p) + C$$

The constant of integration C is determined from the requirement that $\ln \bar{Z} = 0$ at $p = -\pi$, or $\ln \bar{Z}_L$ at $p = \pi$; thus $C = \frac{1}{2} \ln \bar{Z}_L$ and

$$\ln \bar{Z} = \frac{\ln \bar{Z}_L}{2\pi}(p + \pi + 0.632 \sin p - 0.0653 \sin 3p)$$

Replacing p by $2\pi(z - L/2)/L$ from (5.106a) now specifies $\ln \bar{Z}$ as a function of z and completes the design of the taper.

The foregoing synthesis procedure must be used with some caution, stemming from the approximations involved. The theory is valid as long as $d(\ln \bar{Z})/dz$ is small; that is, $\ln \bar{Z}$ must be a slowly varying function of z in order that the reflection coefficient everywhere along the taper be small; that is, $|\Gamma(z)|^2 \ll 1$. This means that $|g(p)|$, and hence all $|a_n|$, must not be permitted to assume excessively large values. Consequently, $|F(n)|$ must not be permitted to become too large. If too many zeros are closely grouped together around a particular value of u, then outside this range $F(u)$ may become excessively large and the accuracy of the theory will suffer. Such "supermatched" designs must be avoided.

*5.18 CHEBYSHEV TAPER

If the number of sections in a Chebyshev transformer is increased indefinitely, with the overall length L kept fixed, we obtain a Chebyshev taper. This taper has equal-amplitude minor lobes and is an optimum design in the sense that it gives the smallest minor-lobe amplitudes for a fixed taper length, and conversely, for a specified minor-lobe amplitude it has the shortest length. As such it is a good taper by which to judge how far other tapers depart from an optimum design. It has been shown that, in the limit as the number of sections in a Chebyshev transformer goes to infinity, the reflection coefficient becomes†

$$\Gamma_i = \frac{1}{2} e^{-j\beta L} \ln \bar{Z}_L \frac{\cos L\sqrt{\beta^2 - \beta_0^2}}{\cosh \beta_0 L} \tag{5.121}$$

where β_0 is the value of β at the lower edge of the passband, as illustrated in Fig. 5.55. As β increases from zero to β_0, the magnitude ρ_i of Γ_i decreases to a final value of $(\ln \bar{Z}_L)/(2 \cosh \beta_0 L)$, since in this region

†R. E. Collin, The Optimum Tapered Transmission Line Matching Section, *Proc. IRE*, vol. 44, pp. 539–548, April, 1956.

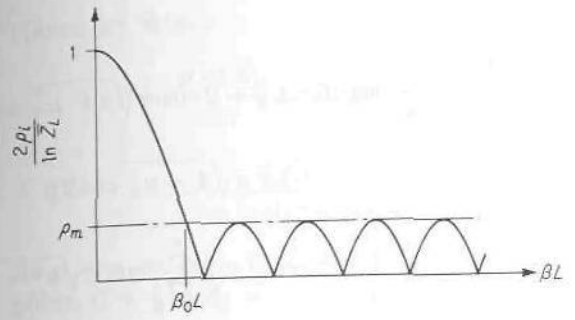


FIGURE 5.55
Reflection-coefficient characteristic for a Chebyshev taper.

$\cos L\sqrt{\beta^2 - \beta_0^2} = \cosh L\sqrt{\beta_0^2 - \beta^2}$. Beyond this point the function in the numerator is the cosine function that oscillates between ± 1 and produces the equal-amplitude minor lobes. The major-lobe to minor-lobe amplitude ratio equals $\cosh \beta_0 L$. Hence, if this is specified so as to keep ρ_i less than or equal to some maximum value ρ_m in the passband, the taper length L is fixed for a given choice of the frequency of the lower edge of the passband which determines β_0. We have

$$\cosh \beta_0 L = \frac{\ln \bar{Z}_L}{2\rho_m} \tag{5.122}$$

Conversely, if β_0 and the taper length L are given, the passband tolerance ρ_m is fixed.

The theory given earlier may be used to determine the function $g(p)$ that will produce the reflection coefficient given by (5.121). Introducing the u variable again, we find that the function $F(u)$ is

$$F(u) = (\ln \bar{Z}_L)\frac{\cos \pi\sqrt{u^2 - u_0^2}}{\cosh \pi u_0} \tag{5.123}$$

where $\pi u = \beta L$, $\pi u_0 = \beta_0 L$. The function $\cos \pi\sqrt{u^2 - u_0^2}$ can be expressed in infinite-product form as

$$\cos \pi\sqrt{u^2 - u_0^2} = \cosh \pi u_0 \prod_{n=1}^{\infty} \left[1 - \frac{u^2}{u_0^2 + \left(n - \frac{1}{2}\right)^2}\right]$$

and this is the limiting value of the polynomial $Q(u)$ in (5.117) as $N \to \infty$. The $\sin \pi u$ term has been canceled by the infinite product $\prod_{n=1}^{\infty}(1 - u^2/n^2)$ as in the first example presented on taper synthesis. However, we do not need this product expansion since $2\pi a_n = F(n)$ in any case. From (5.123) we have

$$a_n = a_{-n} = \frac{1}{2\pi}F(n) = \frac{\ln \bar{Z}_L}{2\pi}\frac{\cos \pi\sqrt{n^2 - u_0^2}}{\cosh \pi u_0} \tag{5.124}$$

Thus

$$g(p) = \frac{\ln \bar{Z}_L}{2\pi \cosh \pi u_0}\left(\cosh \pi u_0 + 2\cos \pi\sqrt{1-u_0^2}\cos p + 2\cos \pi\sqrt{4-u_0^2}\cos 2p + \cdots\right)$$

$$= \frac{\ln \bar{Z}_L}{2\pi \cosh \pi u_0}\left(\cosh \pi u_0 + 2\sum_{n=1}^{\infty} \cos \pi\sqrt{n^2-u_0^2}\cos np\right) \quad (5.125)$$

Integrating with respect to p gives

$$\ln \bar{Z} = \frac{\ln \bar{Z}_L}{2\pi \cosh \pi u_0}\left(p \cosh \pi u_0 + 2\sum_{n=1}^{\infty} \frac{\cos \pi\sqrt{n^2-u_0^2}}{n}\sin np\right) + C \quad (5.126)$$

where C is a constant of integration. To render this result more suitable for computation, it is expedient to add and subtract a similar series; i.e.,

$$\ln \bar{Z} = \frac{p \ln \bar{Z}_L}{2\pi} + \frac{\ln \bar{Z}_L}{\pi \cosh u_0\pi}\sum_{n=1}^{\infty} \frac{\cos n\pi}{n}\sin np + \frac{\ln \bar{Z}_L}{\pi \cosh u_0\pi}\sum_{n=1}^{\infty} \frac{\cos \pi\sqrt{n^2-u_0^2} - \cos n\pi}{n}\sin np + C$$

The second series converges rapidly because $\cos \pi\sqrt{n^2-u_0^2}$ approaches $\cos n\pi$ as n becomes large. The first series may be recognized as the Fourier sine series for the sawtooth function $S(p)$,

$$S(p) = \begin{cases} -\dfrac{p}{2\pi}\dfrac{\ln \bar{Z}_L}{\cosh \pi u_0} & -\pi < p < \pi \\ 0 & p = \pm\pi \end{cases} \quad (5.127)$$

and the periodic continuation of this function as illustrated in Fig. 5.56.

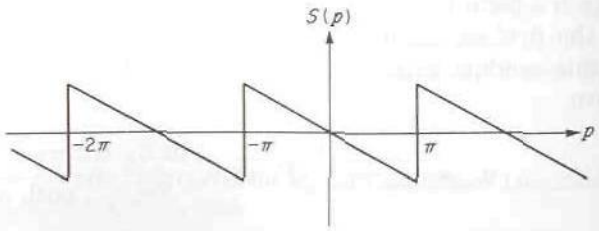


FIGURE 5.56
Sawtooth function.

Hence we obtain

$$\ln \bar{Z} = \frac{p \ln \bar{Z}_L}{2\pi} + S(p) + \frac{\ln \bar{Z}_L}{\pi \cosh \pi u_0} \sum_{n=1}^{\infty} \frac{\cos \pi\sqrt{n^2 - u_0^2} - \cos n\pi}{n} \sin np + C$$

At $p = \pi$, we have $\ln \bar{Z} = \ln \bar{Z}_L$, and since $S(p)$ and $\sin np$ are zero at this point, $C + \frac{1}{2} \ln \bar{Z}_L = \ln \bar{Z}_L$, or $C = \frac{1}{2} \ln \bar{Z}_L$. Our final result is

$$\ln \bar{Z} = \left(\frac{p}{2\pi} + \frac{1}{2} - \frac{p}{2\pi \cosh \pi u_0} \right) \ln \bar{Z}_L + \frac{\ln \bar{Z}_L}{\pi \cosh \pi u_0} \sum_{n=1}^{\infty} \frac{\cos \pi\sqrt{n^2 - u_0^2} - \cos n\pi}{n} \sin np \quad (5.128)$$

for $-\pi < p < \pi$. An interesting feature of the above result is that $\ln \bar{Z}$ changes in a stepwise fashion from 0 to $(\ln \bar{Z}_L)/(2 \cosh \pi u_0)$ as p changes from $-\pi - \epsilon$ to $-\pi + \epsilon$, where $\epsilon \ll 1$. Likewise, at the other end of the taper, $\ln \bar{Z}$ changes abruptly from a value $\ln \bar{Z}_L - (\ln \bar{Z}_L)/(2 \cosh \pi u_0)$ to $\ln Z_L$ as the point $p = \pi$ is reached. This means that the optimum taper has a step change in impedance at each end. The physical basis for this is readily understood by noting that when the frequency is very high, so that the taper is many wavelengths long, the reflection from the smooth part of the taper vanishes. Thus, in order still to maintain equal-amplitude minor lobes, the two-step changes in impedance must be provided to give a reflection coefficient

$$\rho_i = \frac{\ln \bar{Z}_L}{2} \frac{\cos \beta L}{\cosh \beta_0 L} \qquad \text{for } \beta \gg \beta_0$$

As an indication of the superiority of the Chebyshev taper, computations show that it is 27 percent shorter than the taper with $d(\ln \bar{Z})/dz$ a triangular function, for the same passband tolerance and lower cutoff frequency. If the tapers are made the same length, the Chebyshev taper provides a major- to minor-lobe ratio of 84 as compared with 21 for the taper with a triangular distribution.

*5.19 EXACT EQUATION FOR THE REFLECTION COEFFICIENT

The basic equation (5.98) for the input reflection coefficient Γ_i was derived by neglecting all multiple reflections between individual differential sections. The exact equation, derived below, enables an estimate of the range of validity of the approximate theory to be made. First of all, the differential

equation describing the total reflection coefficient Γ at any point z on the line, according to the approximate theory, is derived for later comparison.

With reference to Fig. 5.57, let $d\Gamma_0$ be the reflection coefficient arising from the change $d\bar{Z}$ in characteristic impedance in the interval dz at z. This differential reflection was shown earlier to be given by [see (5.97)]

$$d\Gamma_0 = \frac{1}{2}\frac{d(\ln \bar{Z})}{dz}\,dz$$

The total reflection coefficient at z is the sum of all differential contributions $d\Gamma_0$ from z to L and is

$$\Gamma(z) = \frac{1}{2}\int_z^L e^{-j2\beta(u-z)}\frac{d(\ln \bar{Z})}{du}\,du$$

where u is a dummy variable that measures the distance from the point $z = 0$ toward the load end. The phase angle of the reflected wave arising at u is $2\beta(z-u)$ relative to the forward propagating wave at z.

Differentiating $\Gamma(z)$ with respect to z gives

$$\frac{d\Gamma}{dz} = \frac{2j\beta}{2}\int_z^L e^{j2\beta(z-u)}\frac{d(\ln \bar{Z})}{du}\,du - \frac{1}{2}e^{-j2\beta(u-z)}\frac{d(\ln \bar{Z})}{du}\bigg|_{u=z}$$

$$= 2j\beta\Gamma - \frac{1}{2}\frac{d(\ln \bar{Z})}{dz} \tag{5.129}$$

This is the approximate differential equation for the total reflection coefficient at any point z along the taper.

To find the exact differential equation for Γ, let $\bar{Z}_{in}$ be the input impedance at z and $\bar{Z}_{in} + d\bar{Z}_{in}$ be the input impedance at $z + dz$. We then have

$$\bar{Z}_{in} = \bar{Z}\frac{\bar{Z}_{in} + d\bar{Z}_{in} + j\bar{Z}\tan(\beta\,dz)}{\bar{Z} + j(\bar{Z}_{in} + d\bar{Z}_{in})\tan(\beta\,dz)}$$

$$\approx \bar{Z}\frac{\bar{Z}_{in} + d\bar{Z}_{in} + j\bar{Z}\beta\,dz}{\bar{Z} + j(\bar{Z}_{in} + d\bar{Z}_{in})\beta\,dz}$$

$$\approx (\bar{Z}_{in} + d\bar{Z}_{in} + j\bar{Z}\beta\,dz)\left(1 - j\frac{\bar{Z}_{in}}{\bar{Z}}\beta\,dz\right)$$

$$\approx \bar{Z}_{in} + d\bar{Z}_{in} + j\beta\bar{Z}\,dz - j\frac{\bar{Z}_{in}^2}{\bar{Z}}\beta\,dz$$

upon replacing the tangent by its argument $\beta\,dz$, neglecting products of

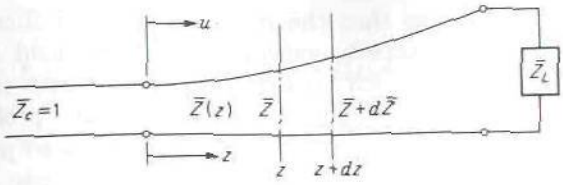


FIGURE 5.57
Tapered transmission line.

differential terms, and replacing the denominator by

$$\left[\bar{Z} + j(\bar{Z}_{\text{in}}\,\beta\,dz)\right]^{-1} \approx \bar{Z}^{-1}\left(1 - j\frac{\bar{Z}_{\text{in}}}{\bar{Z}}\beta\,dz\right)$$

The above gives

$$\frac{d\bar{Z}_{\text{in}}}{dz} = j\beta\left(\frac{\bar{Z}_{\text{in}}^2}{\bar{Z}} - \bar{Z}\right)$$

Now we also have

$$\bar{Z}_{in} = \frac{1+\Gamma}{1-\Gamma}\bar{Z}$$

and hence

$$\frac{d\bar{Z}_{\text{in}}}{dz} = \frac{1+\Gamma}{1-\Gamma}\frac{d\bar{Z}}{dz} + \frac{2\bar{Z}}{(1-\Gamma)^2}\frac{d\Gamma}{dz}$$

Combining these two equations for $d\bar{Z}_{\text{in}}/dz$ and replacing $\bar{Z}_{\text{in}}$ by $\bar{Z}(1+\Gamma)/(1-\Gamma)$ finally give

$$\frac{d\Gamma}{dz} = 2j\beta\Gamma - \frac{1}{2}(1-\Gamma^2)\frac{d(\ln\bar{Z})}{dz} \tag{5.130}$$

upon using the relation $\bar{Z}^{-1}\,d\bar{Z}/dz = d(\ln\bar{Z})/dz$. If we compare (5.130) with (5.129), we see that the approximate equation differs only by the factor $1-\Gamma^2$ *that multiplies the* $d(\ln\bar{Z})/dz$ *term in the exact equation.* If $|\Gamma^2| \ll 1$ everywhere along the line, the approximate equation would be expected to yield good results.

The exact equation (5.130) is called a Riccati equation. It is a nonlinear equation because of the term in Γ^2 and does not have a known general solution. This equation can be integrated only in certain special cases (one such case is for the exponential taper, as in Prob. 5.48). However, the practical difficulty in applying the exact theory does not stem so much from the lack of a general solution of (5.130), since numerical integration or iteration schemes can always be employed, as from being unable to specify what the characteristic impedance $\bar{Z}$ is along a general taper. If the taper is very gradual, then $\bar{Z}(z)$ can be taken as the characteristic impedance of a uniform line having the same cross-sectional dimensions as the taper has at the plane z. But for such gradual tapers the approximate theory is valid and

gives good results, so that the more complicated Riccati equation is not required. For more rapidly varying tapers, the field structure along the taper is perturbed to such an extent that no simple method of specifying $\bar{Z}(z)$ exists. In actual fact, the new boundary-value problem will have to be solved to determine $\bar{Z}(z)$, and this solution will also provide the values of the reflection coefficient along the taper. Thus one concludes that the inability to specify $\bar{Z}(z)$, except for the case of gradual tapers, makes the exact equation of minor importance in practice.

In the case of waveguide tapers, β is a function of the cross-sectional geometry and hence a function of z along the taper. In order to integrate the approximate equation (5.129), in this case, an auxiliary variable θ is introduced as follows:

$$\theta = \int_0^z 2\beta(z)\,dz \qquad d\theta = 2\beta\,dz$$

and hence

$$\frac{d}{dz} = \frac{d}{d\theta}\frac{d\theta}{dz} = 2\beta\frac{d}{d\theta}$$

We now have

$$\frac{d\Gamma}{dz} = 2\beta\frac{d\Gamma}{d\theta} = 2j\beta\Gamma - \beta\frac{d(\ln \bar{Z})}{d\theta}$$

or

$$\frac{d\Gamma}{d\theta} = j\Gamma - \frac{1}{2}\frac{d(\ln \bar{Z})}{d\theta} \tag{5.131}$$

This is readily integrated to give (multiply both sides by $e^{-j\theta}$ and note that $d\Gamma e^{-j\theta}/d\theta = -j\Gamma e^{-j\theta} + e^{-j\theta}d\Gamma/d\theta$)

$$\Gamma(\theta)e^{-j\theta}\Big|_0^{\theta_0} = -\frac{1}{2}\int_0^{\theta_0} e^{-j\theta}\frac{d(\ln \bar{Z})}{d\theta}\,d\theta$$

where

$$\theta_0 = \int_0^L 2\beta\,dz$$

If $\bar{Z} = \bar{Z}_L$ at $z = L$, then $\Gamma(\theta_0) = 0$, and since $\Gamma(0) = \Gamma_i$, we have

$$\Gamma_i = \frac{1}{2}\int_0^{\theta_0} e^{-j\theta}\frac{d(\ln \bar{Z})}{d\theta}\,d\theta \tag{5.132}$$

In terms of the new variable θ, the problem is formally the same as before. However, the specification of $\bar{Z}(\theta)$ as a function of θ may be rather difficult to translate back into a function of z. Consequently, a general synthesis procedure applied to (5.132) may be difficult to carry out, although in principle it is formally the same problem as considered earlier, except for the last step, i.e., expressing $\bar{Z}(\theta)$ as a function of z in order to determine the shape of the taper.

PROBLEMS

5.1. Show that the $\bar{R}$ = constant and $\bar{X}$ = constant contours in the reflection-coefficient plane are circles given by (5.6*a*) and (5.6*b*).

5.2. In the circuit shown in Fig. P5.2, what is the smallest value of d that will make the resistive part of Z_{in} equal to 50 Ω at the plane t_1? Find the required value of jX to make Z_{in} equal to 50 Ω at the plane t_2. Use a Smith chart to find d and jX.

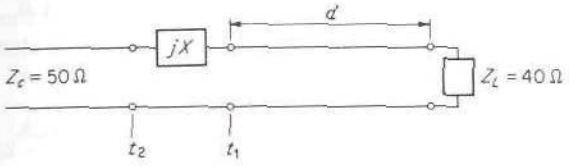


FIGURE P5.2

5.3. On a certain line terminated in a normalized load impedance $\bar{Z}_L$, it was found that the standing-wave ratio S was equal to 2 and that a voltage minimum occurred at $\lambda/4$ from the load. What is $\bar{Z}_L$? Find the position and length of a single shunt stub that will match the load to the line. Use a Smith chart.

5.4. What are the length and position required for a series stub to match the load specified in Prob. 5.3? Use a Smith chart.

5.5. *A normalized load $\bar{Z}_L = 2$ terminates a transmission line. Does a voltage* maximum or minimum occur at the load? Find the position and length of a *shunt stub that will match the load to the line at a frequency f_1, where the* wavelength is λ_1. With the stub parameters and the load fixed, let the *wavelength be increased to $1.1\lambda_1$. What is the new value of stub susceptance* and the standing-wave ratio on the line? If at the wavelength λ_1 the stub is *placed $\lambda_1/2$ farther toward the generator from its original position*, what is the standing-wave ratio on the line when the wavelength is increased to $1.1\lambda_1$? *This illustrates the greater frequency sensitivity of the match when the stub is* placed farther from the load.

5.6. A double stub with spacing 0.25λ is used to match a normalized load admittance $\bar{Y}_L = 0.5 + j1$. Find the required stub susceptances by using a Smith chart.

Answer:

Solution 1: $\bar{B}_1 = -0.5$ $L_1 = 0.176\lambda$ $\bar{B}_2 = 1$ $L_2 = 0.375\lambda$

Solution 2: $\bar{B}_1 = -1.5$ $L_1 = 0.0936\lambda$ $\bar{B}_2 = -1$ $L_2 = 0.125\lambda$

5.7. Can a normalized load admittance $\bar{Y}_L = 2.5 + j1$ be matched with a double-stub tuner with stub spacing $\lambda/10$?

5.8. What is the minimum stub spacing in a double-stub tuner that will permit a normalized load admittance $\bar{Y}_L = 1.5 + j2$ to be matched if the spacing is restricted to be greater than $\lambda/4$? If the spacing must be less than $\lambda/4$, what is the maximum stub spacing that can be used?

5.9. Show that, for a double-stub tuner with $d = \lambda/4$, the required values of stub susceptances are given by

$$B_1 = -B_L \pm [G_L(1 - G_L)]^{1/2}$$

$$B_2 = \pm\left(\frac{1 - G_L}{G_L}\right)^{1/2}$$

Hint: Take the limit of (5.20) and (5.21) as $\tan \beta d$ becomes infinite.

5.10. A normalized load admittance $\overline{Y}_L = 0.8 + j0.5$ is to be matched to a transmission line using a double-stub tuner. Use a Smith chart to find the required stub susceptances and lengths. Assume that short-circuited stubs spaced by a distance 0.4λ are used.

Answer:

Solution 1: $B_1 = -3.171$ $L_1 = 0.0486\lambda$ $B_2 = -2.994$ $L_2 = 0.0513\lambda$

Solution 2: $B_1 = -0.582$ $L_1 = 0.166\lambda$ $B_2 = 0.2416$ $L_2 = 0.288\lambda$

5.11. A line with an attenuation constant $\alpha = 0.01$ Np/m is used as a short-circuited stub. Using the formula $Y_{in} = Y_c \coth(\alpha + j\beta)l$, find the maximum value of normalized susceptance this stub can give. If $\alpha = 0.02$, what is the maximum value of stub susceptance that can be obtained? Use $l = 1$ m.

5.12. Consider the series-shunt matching circuit illustrated in Fig. P5.12. What values of series stub reactance and shunt stub susceptance will match an arbitrary load $\overline{Y}_L = \overline{G}_L + j\overline{B}_L$ a distance d away?

Hint: Note that $\overline{Y}_{in}$ at the position of the shunt stub equals $(j\overline{X} + \overline{Z}'_{in})^{-1}$, where $\overline{Z}'_{in}$ is the input impedance just to the right of the series stub. Next impose the condition $\overline{Y}_{in} = 1 + j\overline{B}$ to find the required value of $j\overline{X}$.

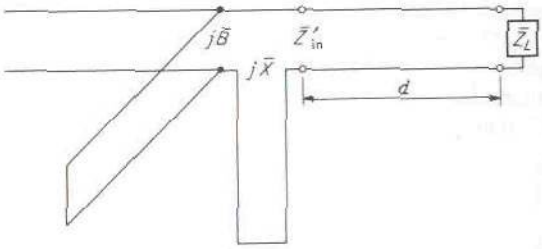


FIGURE P5.12

5.13. *Develop a graphical method, using the Smith chart, to solve Prob. 5.12.*

Hint: $\overline{Y}_{in}$ must lie on the $\overline{G} = 1$ circle. But $\overline{Y}_{in}$ is obtained by moving 180° around the chart from the $\overline{Z}_{in} = j\overline{X} + \overline{Z}'_{in}$ point. Hence, to make $\overline{Y}_{in}$ lie on the $\overline{G} = 1$ circle, choose $j\overline{X}$ to make $\overline{Z}_{in}$ lie on the $\overline{R} = 1$ circle rotated through 180°.

5.14. A horn antenna is fed from a rectangular waveguide in which the TE_{10} mode propagates. At the junction a reflection coefficient $\Gamma = 0.3e^{j\pi/4}$ is produced. What is the normalized input admittance to the horn? Find the required normalized susceptance and the spacing in guide wavelengths from the junction of an inductive diaphragm that will match the waveguide to the horn.

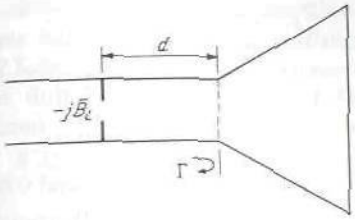


FIGURE P5.14

5.15. A load impedance $40 + j10$ is to be matched to a 50-Ω line using lumped elements. Use a Smith chart to find the parameters of a suitable matching circuit. Evaluate the loaded Q of the circuit. Two solutions are possible; choose the one with the lowest Q. Check your answers using LCMATCH.

5.16. Use a Smith chart to design a lumped-parameter matching network that will match a load admittance $Y_L = 0.04 + j0.08$ to a 50-Ω transmission line. Find both solutions and evaluate the loaded Q for each. Check your answers using LCMATCH.

5.17. For the matching circuit shown in Fig. 5.17*a*, shown that the required values of $j\bar{B}_1$ and $j\bar{X}_2$ are given by

$$j\bar{B}_1 = j\left(-\bar{B}_L \pm \sqrt{\bar{G}_L(1-\bar{G}_L)}\right)$$

$$j\bar{X}_2 = \pm j\sqrt{\frac{1-\bar{G}_L}{\bar{G}_L}}$$

Hint: Note that $\bar{Z}_{in} = j\bar{X}_2 + (j\bar{B}_1 + \bar{Y}_L)^{-1}$ must equal 1.

5.18. For the matching circuit shown in Fig. 5.17*b*, show that the required values of $j\bar{X}_1$ and $j\bar{B}_2$ are given by

$$j\bar{X}_1 = j\left(-\bar{X}_L \pm \sqrt{\bar{R}_L(1-\bar{R}_L)}\right)$$

$$j\bar{B}_2 = \pm j\sqrt{\frac{1-\bar{R}_L}{\bar{R}_L}}$$

5.19. Show that the definition (5.23) for Q gives the expressions in (5.26).

5.20. A load admittance $G_L + j\omega C_L = G_L + jB_L = 0.02 + j0.05$ is to be matched to a 50-Ω transmission line using a double-stub tuner with short-circuited shunt stubs spaced by 0.18λ. The stubs have a characteristic impedance of 40 Ω. Use the computer program STUBMACH to find the required stub lengths for the two possible solutions. Use the frequency scan option with a normalized frequency step of 0.02 and five steps below and above the matching frequency to evaluate the frequency sensitivity of the two solutions. Plot the return loss as a function of normalized frequency and find the bandwidth of each circuit for which the return loss is less than −20 dB.

5.21. Repeat Prob. 5.20 but use open-circuited stubs. Compare the bandwidths obtained with those when short-circuited stubs are used.

5.22. A load impedance $Z_L = R_L + j\omega L = R_L + jX_L = 30 + j100$ is to be matched to a 50-Ω transmission line using open-circuited series stubs spaced 0.15λ apart. The stubs have a characteristic impedance of 75 Ω. Use the computer program *STUBMACH* to find the required stub reactances and lengths. Carry out a frequency scan to determine the bandwidth over which the VSWR remains below 1.5.

Answer: The two bandwidths are 0.045 and 0.087.

5.23. (*a*) For Prob. 5.22 replace the first stub by a short-circuited shunt stub. Does this improve the bandwidth? (*b*) Replace the second stub by a short-circuited shunt stub but keep the original first stub. How does this affect the bandwidth?

Answer: (*a*) BWs are 0.01 and 0.017. (*b*) BWs are 0.03 and 0.077.

5.24. A load impedance $Z_L = 5 + j50$ is to be matched to a 50-Ω transmission line. Design a lumped-element matching network that uses a capacitor in series with the load as the first element plus a second shunt element. Use the computer program LCMATCH. Evaluate the return loss as a function of normalized frequency using normalized frequency steps of 0.01. For the two solutions plot the return loss as a function of the normalized frequency. Find the 3-dB bandwidth for the two circuits. Evaluate the loaded Q_L for the two circuits and compare the bandwidth with that given by $1/Q_L$.

Answer: $Q_L = 5$ *and* 6.5. *The bandwidths are* 0.21 *and* 0.16 and are close to the respective $1/Q_L$ values.

5.25. Repeat Prob. 5.24 but use a shunt element for the first element in the matching network (option b in the computer program LCMATCH).

5.26. A load admittance $Y_L = 0.002 + j0.1$ is to be matched to a 50-Ω transmission line. Use a matching network with the first element being a shunt element. Evaluate the return loss as a function of the normalized frequency and find the 3-dB bandwidth of the two circuits. Evaluate the loaded Q_L and compare the bandwidth with that given by $1/Q_L$. Use LCMATCH.

5.27. Repeat Prob. 5.26 but use a series element as the first element.

5.28. A load admittance $G_L = 0.002 - j0.05$ is to be matched to a 50-Ω transmission line. Find solutions for four different matching networks that can be used. Evaluate the loaded Q_L of each and identify the network giving the greatest bandwidth. Use LCMATCH.

5.29. A load impedance $Z_L = 80 - j10$ is to be matched to a 40-Ω transmission line. Find the elements for two possible matching networks and the corresponding bandwidths for which the VSWR remains below 1.2. Use LCMATCH.

5.30. A microwave amplifier requires a load termination $Z_L = 20 + j80$. Use a Smith chart to design a matching network of the form shown in Fig. 5.22*b*. The transmission line is an open-circuited one. The transmission lines have a characteristic impedance of 50 Ω.

5.31. Use a Smith chart to design a lumped-element network of the form shown in Fig. 5.22*c* that will provide a load $Z_L = 30 + j50$ for an amplifier.

5.32. What are the required length and impedance of a quarter-wave transformer that will match a 100-Ω load to a 50-Ω line at $f = 10{,}000$ MHz (air-filled line)? What is the frequency band of operation over which the reflection coefficient remains less than 0.1?

5.33. Design a two-section binomial transformer to match the load given in Prob. 5.32. What bandwidth is obtained for $\rho_m = 0.1$?

5.34. Design a three-section binomial transformer to match a 100-Ω load to a 50-Ω line. The maximum VSWR that can be tolerated is 1.1. What bandwidth can be obtained? Plot ρ versus f.

5.35. Design a two-section Chebyshev transformer to match a 100-Ω load to a 50-Ω line. The maximum voltage standing-wave ratio (VSWR) that can be tolerated is 1.2. What bandwidth is obtained? Plot ρ versus f. Use the approximate theory.

***5.36.** Use the exact theory to design a two-section Chebyshev transformer to match a normalized load $\bar{Z}_L = 5$ to a transmission line. The required fractional bandwidth is 0.6. What is the resultant value of ρ_m?

***5.37.** Design a three-section Chebyshev transformer (exact theory) to match the load specified in Prob. 5.36. The same bandwidth is required. Compute ρ_m and note the improvement obtained.

***5.38.** Let x_s be a zero of $T_N(x)$; that is, $T_N(x_s) = 0$. Let the corresponding value of $\cos\theta$ be $\cos\theta_z$. Note that $T_N(\sec\theta_m \cos\theta) = 0$ when $\sec\theta_m \cos\theta = \sec\theta_m \cos\theta_z$. Use the result

$$T_N(\cos\phi) = \cos N\phi$$

to compute the zeros of $T_2(x)$. Note that the zeros occur when $\cos 2\phi = 0$, or $\phi = \pi/4, 3\pi/4$, and that $x_s = \cos\phi_z$. Using these results together with the relation $\sec\theta_m = x_s \sec\theta_z$, show that, for a two-section transformer,

$$P_{LR} = 1 + k^2 T_2^2(\sec\theta_m \cos\theta)$$

reduces to (5.76). Note that when $\theta = 0$, $\rho^2 = (Z_L - Z_0)^2/(Z_L + Z_0)^2$, and hence

$$k^2 = \frac{(Z_L - Z_0)^2}{4Z_L Z_0 T_2^2(\sec\theta_m)}$$

5.39. For a particular application it is desired to obtain a reflection-coefficient characteristic

$$\rho_i = C\left(\theta - \frac{\pi}{2}\right)^2 \qquad 0 \le \theta \le \pi$$

The normalized load impedance equals 2. What must the value of C be? Use the approximate theory for an N-section transformer to design a four-section transformer that will approximate the above specified ρ_i. Expand ρ_i in a Fourier cosine series to determine the coefficients ρ_n. Plot the resultant ρ_i versus θ and compare with the specified characteristic. Show that the approximation to the specified ρ_i is a least-mean-square-error approximation. If the number of sections in the transformer were increased indefinitely, could the specified ρ_i be realized?

5.40. An empty rectangular waveguide is to be matched to a dielectric-filled rectangular guide by means of an intermediate quarter-wave transformer. Find the length and dielectric constant of the matching section. Use $a = 2.5$ cm, $f = 10{,}000$ MHz, $\epsilon_{r2} = 2.56$. Plot ρ_i versus frequency f. Note that the appropriate impedances to use here are the wave impedances for the TE_{10} mode. ϵ_{r2} is the dielectric constant in the output guide.

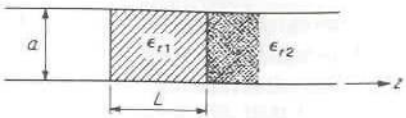


FIGURE P5.40

5.41. Design a two-section binomial transformer to match the empty guide to the dielectric-filled guide in Prob. 5.40. Use intermediate sections with dielectric constants ϵ_{r1} and ϵ_{r2}. Plot the input reflection coefficient as a function of frequency.

5.42. Find the required impedances for a three-section half-wave filter having a fractional bandwidth of 0.3 and a passband tolerance $k^2 = 1.87 \times 10^{-2}$. The input and output transmission lines have a characteristic impedance of 50 Ω.

5.43. Find the required impedances for a two-section half-wave filter having a fractional bandwidth of 0.2 and a passband tolerance $k^2 = 10^{-2}$. The input transmission line has a characteristic impedance of 50 Ω. What is the required characteristic impedance of the output transmission line? You will need to design the quarter-wave-transformer prototype circuit first.

5.44. The filter described in Prob. 5.42 is to be built using microstrip construction on an alumina substrate ($\epsilon_r = 9.6$) 1 mm thick. The center frequency is 2 GHz. Determine the line widths and lengths including compensation for junction capacitance. Use the MSTP program.

5.45. Obtain an expression for ρ_i for a taper which has $d(\ln Z)/dz = C \sin \pi(z/L)$. Determine C so that $\ln Z = \ln Z_L$ at $z = L$ and equals zero at $z = 0$. Plot ρ_i versus βL.

***5.46.** Design a taper that has double zeros at $\beta L = \pm 2\pi$ and $\pm 3\pi$. Plot $F(u)$ versus u for this taper. Determine $\ln Z$ as a function of z. Achieve the design by moving the zeros at $\pm\pi$ and $\pm 4\pi$ into the points at $\pm 2\pi$, $\pm 3\pi$.

***5.47.** Design a taper with single zeros at $\beta L = \pm\pi, \pm 1.25\pi, \pm 1.5\pi, \pm 1.75\pi$ by moving the zeros at $\pm 2\pi, \pm 3\pi, \pm 4\pi$ into the specified points. Plot the resultant $F(u)$. Determine the expansion for $\ln Z$ as a function of z for this taper. Note that the close spacing of zeros keeps ρ_i small in the range $\pi < \beta l < 2\pi$, but that removal of the zeros at $\pm 3\pi$ and $\pm 4\pi$ lets ρ_i rise to a large value in this region.

***5.48.** Show that, for the exponential taper, the exact solution of (5.130) for the input reflection coefficient is

$$\Gamma_i = \frac{A \sin(BL/2)}{B \cos(BL/2) + 2j\beta \sin(BL/2)}$$

where $A = (\ln Z_L)/L$, $B = \sqrt{4\beta^2 - A^2}$.

***5.49.** Show that the approximate differential equation for the input impedance along a slowly varying tapered line is ($Z_{in} \approx Z$ at all points)

$$\frac{dZ_{in}}{dz} = -2j\beta(Z - Z_{in})$$

Integrate this to obtain the input impedance to an exponential line terminated in a load impedance Z_L.

REFERENCES

Quarter-wave transformers

1. Collin, R. E.: Theory and Design of Wide Band Multisection Quarter-Wave Transformers, *Proc. IRE*, vol. 43, pp. 179–185, February, 1955.
2. Cohn, S. B.: Optimum Design of Stepped Transmission Line Transformers, *IRE Trans.*, vol. MTT-3, pp. 16–21, April, 1955.
3. Riblet, H. J.: General Synthesis of Quarter-Wave Impedance Transformers, *IRE Trans.*, vol. MTT-5, pp. 36–43, January, 1957.
4. Young, L.: Tables for Cascaded Homogeneous Quarter-Wave Transformers, *IRE Trans.*, vol. MTT-7, pp. 233–237, April, 1959. See also *IRE Trans.*, vol. MTT-8, pp. 243–244, for corrections.
5. Solymar, L.: Some Notes on the Optimum Design of Stepped Transmission Line Transformers, *IRE Trans.*, vol. MTT-6, pp. 374–378, October, 1958.

Tapered lines

6. Bolinder, F.: Fourier Transforms in the Theory of Inhomogeneous Transmission Lines, *Proc. IRE*, vol. 38, p. 1354, November, 1950.
7. Klopfenstein, R. W.: A Transmission Line Taper of Improved Design, *Proc. IRE*, vol. 44, pp. 31–35, January, 1956.
8. Collin, R. E.: The Optimum Tapered Transmission Line Matching Section, *Proc. IRE*, vol. 44, pp. 539–548, April, 1956.
9. Bolinder, F.: Fourier Transforms in the Theory of Inhomogeneous Transmission Lines, *Trans. Roy. Inst. Technol., Stockholm, no.* 48, 1951.
10. Matsumaru, K.: Reflection Coefficient of *E*-Plane Tapered Waveguides, *IRE Trans.*, vol. MTT-6, pp. 143–149, April, 1958.
11. Johnson, R. C.: Design of Linear Double Tapers in Rectangular Waveguides, *IRE Trans.*, vol. MTT-7, pp. 374–378, July, 1959.

CHAPTER 6

PASSIVE MICROWAVE DEVICES

A large variety of passive microwave circuit components or devices have been developed for use in both the laboratory and in microwave communication and radar systems. In this chapter we will describe the basic operating principles for a number of the most commonly used devices, such as attenuators, phase shifters, directional couplers, power dividers, hybrid junctions, and ferrite devices such as isolators or gyrators and circulators. The physical form that these devices take depends on whether waveguide, coaxial transmission line, strip line, microstrip, etc., is used as the transmission medium. The development of transistors and switching diodes for applications in the microwave frequency band has resulted in the development of a number of electronically controlled devices such as attenuators and phase shifters. Some of these electronically controlled circuit elements are also examined in this chapter.

6.1 TERMINATIONS

Two types of waveguide and transmission-line terminations are in common use. One is the matched load, and the other is a variable short circuit that produces an adjustable reactive load. These terminations are extensively used in the laboratory when measuring the impedance or scattering parameters of a microwave circuit element. The matched load provides a termination that absorbs all the incident power, and hence is equivalent to terminating the line in its characteristic impedance. The variable short circuit is a termination that reflects all the incident power. The phase of the

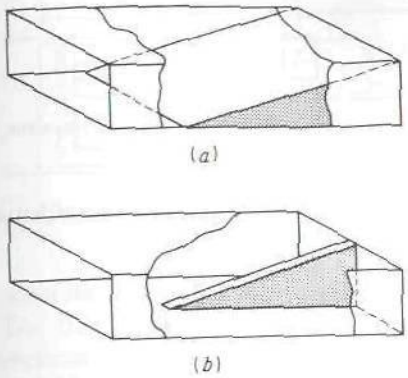


FIGURE 6.1
Matched loads for a waveguide. (*a*) Lossy wedge; (*b*) tapered resistive card.

reflected wave is varied by changing the position of the short circuit, and this is equivalent to changing the reactance of the termination.

The usual matched load for a waveguide is a tapered wedge or slab of lossy material inserted into the guide, as illustrated in Fig. 6.1. Since the material is lossy, the incident power is absorbed. Reflections are avoided by tapering the lossy material into a wedge. Thus the termination may be viewed as a lossy tapered transmission line. An overall length of one or more wavelengths is usually sufficient to provide a matched load with an input standing-wave ratio of 1.01 or smaller.

Variable Short Circuit

The simplest form of adjustable short circuit for use in a waveguide is a sliding block of copper or some other good conductor that makes a snug fit in the guide, as illustrated in Fig. 6.2. The position of the block is varied by means of a micrometer drive. This simple form of adjustable short circuit, however, is not very satisfactory in its electrical performance. The erratic contact between the sliding block and the waveguide walls causes the equivalent electrical short-circuit position to deviate in a random fashion from the physical short-circuit position which is the front face of the sliding block. In addition, some power leakage past the block may occur, and this results in a reflection coefficient less than unity. These problems may be overcome by using a choke-type plunger, as discussed below.

The choke-type plunger is an example of the use of impedance transformation properties of a quarter-wave transformer. Consider, for example, a load impedance Z_s that is approximately zero. If this impedance is

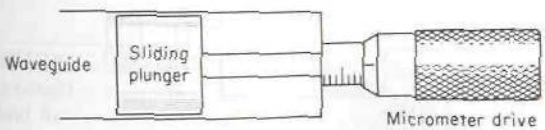


FIGURE 6.2
A simple short circuit for a waveguide.

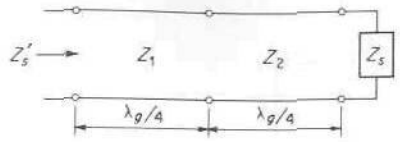


FIGURE 6.3
Two-section quarter-wave transformer.

viewed through a two-section quarter-wave transformer, as in Fig. 6.3, the impedance seen at the input is

$$Z_s' = \left(\frac{Z_1}{Z_2}\right)^2 Z_s \tag{6.1}$$

If Z_2 is chosen much greater than Z_1, the new impedance Z_s' will approximate a short circuit by a factor $(Z_1/Z_2)^2$ better than Z_s does. This is essentially the principle used in choke-type plungers. The improvement factor of course deteriorates when the frequency is changed, so that the transformer sections are no longer a quarter wave long. However, by proper design, a bandwidth of 10 percent or more can be achieved. In very critical applications more than two sections may be used. The foregoing theory is now applied to a choke-type plunger for use in a rectangular guide.

For the TE_{10} dominant mode in a rectangular guide, the surface currents on the interior wall flow up and down along the side walls and both across and in the axial direction on the broad walls. In the simple type of short circuit illustrated in Fig. 6.2, the axial current must flow across the gap between the upper and lower waveguide walls and vertically across the front face of the sliding block. The currents that flow along the side walls flow in the vertical direction and do not need to cross the gap between the waveguide walls and the front face of the plunger. Consequently, the erratic performance arises only from the axial currents flowing on the upper and lower walls. To avoid this erratic behavior, the plunger may be made in the form illustrated in Fig. 6.4. The width of the plunger is uniform and slightly less than the interior guide width. However, the height of the plunger is made nonuniform. The front section is a quarter guide wavelength long and less than the guide height b by an amount $2b_1$. The gap b_1 is made as small as possible consistent with the requirement that the front section must not touch the upper and lower waveguide walls. The second section is also a quarter guide wavelength long, but with the gap b_2 made as large as

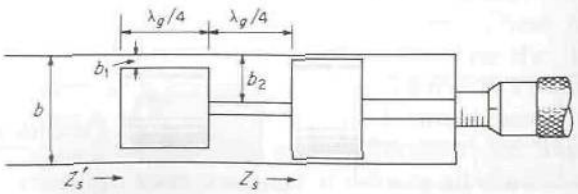


FIGURE 6.4
Choke-type adjustable short circuit (side view).

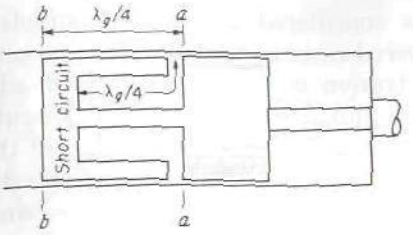


FIGURE 6.5
Alternative choke-type plunger design.

possible consistent with maintaining the mechanical strength of the plunger. The final back section makes a sliding fit in the guide. The quarter-wave sections have equivalent characteristic impedances proportional to $2b_1/b$ and $2b_2/b$ relative to that of the input guide. Thus application of (6.1) gives

$$Z_s' = \left(\frac{b_1}{b_2}\right)^2 Z_s$$

for the normalized input impedance. By good mechanical design b_2 can be made about 10 times greater than b_1, and hence an improvement in performance by a factor of 100 over the non-choke-type short circuit may be achieved.

A somewhat different design, as illustrated in Fig. 6.5, is also frequently used. In this plunger a two-section folded quarter-wave transformer is used. The inner line transforms the short-circuit impedance to an ideal open circuit at the plane *aa*. At this point, i.e., at an open-circuit or infinite impedance point, the axial current is zero. Hence there is no current present to flow across the gap between the waveguide wall and the plunger at the contact point *aa*. The next, or outer, quarter-wave transformer transforms the open-circuit impedance at *aa* into a short-circuit impedance at the front end of the plunger, i.e., at plane *bb* in Fig. 6.5. Short-circuit plungers of this type give very satisfactory performance.

The above application of quarter-wave transformers is also used in the construction of choke joints for joining two waveguide sections together, in rotary joints, for plungers used to tune cavity resonators, etc.†

6.2 ATTENUATORS

Attenuators may be of the fixed or the variable type. The first is used only if a fixed amount of attenuation is to be provided. For bridge setups used to measure transmission coefficients, the variable attenuator is used. There are many ways of constructing a variable attenuator; only one type, the

†See, for example, G. L. Ragan (ed.), "Microwave Transmission Circuits," McGraw-Hill Book Company, New York, 1948.

rotary attenuator, is considered in detail. A simple form of attenuator consists of a thin tapered resistive card, of the type used for matched loads, whose depth of penetration into the waveguide is adjustable. The card is inserted into the guide through a longitudinal slot cut in the center of the broad wall of a rectangular guide. An attenuator of this form has a rather complicated attenuation variation with depth of insertion and frequency.

A better precision type of attenuator utilizes an adjustable length of waveguide operated below its cutoff frequency. The disadvantage of this type of attenuator is that the output is attenuated by reducing the coupling between the input and output guides, and not by absorption of the incident power. As a result, a high degree of attenuation corresponds to a reflection coefficient near unity in the input guide, and this is often undesirable.

Perhaps the most satisfactory precision attenuator developed to date is the rotary attenuator, which we now examine in some detail. The basic components of this instrument consist of two rectangular-to-circular waveguide tapered transitions, together with an intermediate section of circular waveguide that is free to rotate, as in Fig. 6.6. A thin tapered resistive card is placed at the output end of each transition section and oriented parallel to the broad walls of the rectangular guide. A similar resistive card is located in the intermediate circular-guide section. The incoming TE_{10} mode in the rectangular guide is transformed into the TE_{11} mode in the circular guide with negligible reflection by means of the tapered transition. The polarization of the TE_{11} mode is such that the electric field is perpendicular to the thin resistive card in the transition section. As such, this resistive card has a negligible effect on the TE_{11} mode. Since the resistive card in the center section can be rotated, its orientation relative to the electric field of the incoming TE_{11} mode can be varied so that the amount by which this mode is attenuated is adjustable.

With reference to Fig. 6.7, let the center resistive card be oriented at an angle θ relative to the direction of the electric field polarization of the TE_{11} mode. The TE_{11} mode polarized in the x direction may be decomposed into the sum of two TE_{11} modes polarized along the u and v directions, as

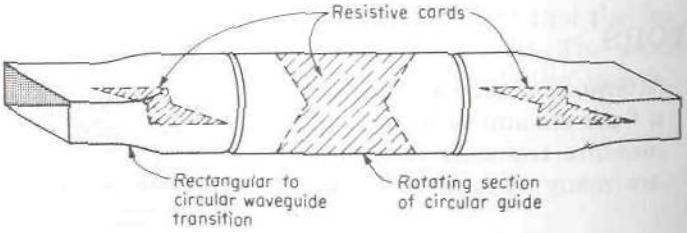


FIGURE 6.6
Basic construction of a rotary attenuator.

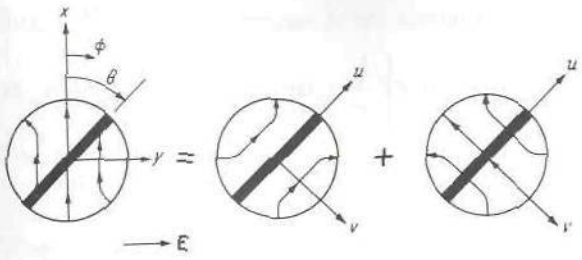


FIGURE 6.7
Decomposition of TE_{11} mode into two orthogonally polarized modes.

illustrated in Fig. 6.7. That portion which is parallel to the resistive slab will be absorbed, whereas the portion which is polarized perpendicular to the slab will be transmitted. However, upon entering the transition section, the transmitted mode is again not perpendicular to the resistive card in this section; so some additional attenuation will occur.

To derive the expression for the dependence of the attenuation on the rotation angle θ, consider the analytic expression for the TE_{11}-mode electric field. For polarization in the x direction the electric field is given by (Table 3.6)

$$\mathbf{E} = \frac{J_1(p'_{11}r/a)}{r}\mathbf{a}_r \cos\phi - \frac{p'_{11}}{a}J'_1\left(\frac{p'_{11}r}{a}\right)\mathbf{a}_\phi \sin\phi \qquad (6.2)$$

apart from the propagation factor $e^{-j\beta z}$. Since

$$\sin\phi = \sin(\phi - \theta + \theta) = \cos\theta\sin(\phi - \theta) + \sin\theta\cos(\phi - \theta)$$

and similarly

$$\cos\phi = \cos(\phi - \theta + \theta) = \cos\theta\cos(\phi - \theta) - \sin\theta\sin(\phi - \theta)$$

the above expression may be written as

$$\mathbf{E} = \cos\theta\left[\frac{J_1}{r}\mathbf{a}_r\cos(\phi - \theta) - \frac{p'_{11}}{a}J'_1\mathbf{a}_\phi\sin(\phi - \theta)\right]$$
$$- \sin\theta\left[\frac{J_1}{r}\mathbf{a}_r\sin(\phi - \theta) + \frac{p'_{11}}{a}J'_1\mathbf{a}_\phi\cos(\phi - \theta)\right] \qquad (6.3)$$

which is equivalent to referring the angle variable ϕ to a new origin at θ. The first term in brackets in (6.3) is a TE_{11} mode polarized with the electric field *along the u axis as in Fig. 6.7, and the second term in brackets is a* TE_{11} mode polarized along the v axis. Since the first part is completely absorbed, only the portion multiplied by $\sin\theta$ is transmitted into the output transition section. If we assume that the resistive card in this section is parallel to the y axis, only the component of the transmitted field which is polarized along the x axis is transmitted. We have, at the input to the

transition section,

$$\mathbf{E} = -\sin\theta\left[\frac{J_1}{r}\mathbf{a}_r \sin(\phi - \theta) + \frac{p'_{11}}{a}J'_1\mathbf{a}_\phi \cos(\phi - \theta)\right]$$

$$= \sin^2\theta\left(\frac{J_1}{r}\mathbf{a}_r \cos\phi - \frac{p'_{11}}{a}J'_1\mathbf{a}_\phi \sin\phi\right)$$

$$- \sin\theta\cos\theta\left(\frac{J_1}{r}\mathbf{a}_r \sin\phi + \frac{p'_{11}}{a}J'_1\mathbf{a}_\phi \cos\phi\right) \tag{6.4}$$

of which the first part is a TE_{11} mode polarized along the x direction and is transmitted. Comparing (6.4) with (6.2) shows that the transmitted field is reduced by a factor $\sin^2\theta$ from the amplitude of the incident field. Hence the attenuation produced is given in decibels by

$$\alpha = -20\log(\sin^2\theta) = -40\log(\sin\theta) \tag{6.5}$$

A notable feature is that the attenuation depends only on the angle of rotation θ, a result that is verified in practice to a high degree of approximation.

Electronically Controlled Attenuators

For applications in various microwave systems, it is desirable to have an attenuator whose attenuation can be controlled by the application of a suitable signal, such as a dc voltage or a bias current. Two devices that are suitable for use in an electronically controlled attenuator are the *PIN* diode and a field-effect transistor. These devices can be used as variable resistors whose resistance is controlled by the applied signal.

The basic attenuator network is a symmetrical resistive T or Π network as shown in Figs. 6.8*a* and *b*. The resistor values R_1 and R_2 are chosen so that when the attenuator is terminated in a resistance equal to the transmission-line characteristic impedance Z_c, the input is matched, that is, $Z_{in} = Z_c$, and to provide an output voltage reduction by a chosen

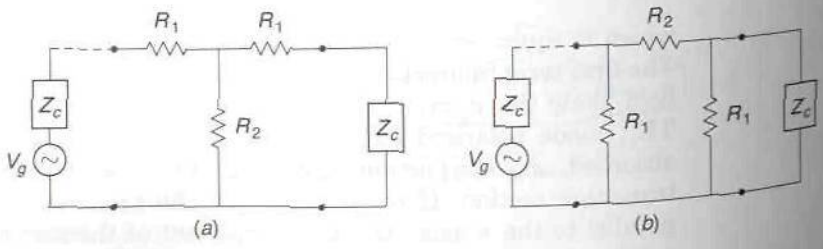


FIGURE 6.8
The two basic attenuator networks.

factor K. For the T network we have

$$R_{in} = R_1 + \frac{R_2(R_1 + Z_c)}{R_1 + R_2 + Z_c}$$

For $R_{in} = Z_c$ we get the equation

$$R_1(R_1 + 2R_2) = Z_c^2$$

When $R_{in} = Z_c$ the Thévenin impedance seen by the load equals Z_c also. The Thévenin open-circuited voltage across R_2 is

$$V_{TH} = \frac{R_2}{R_1 + R_2 + Z_c} V_g$$

The power delivered to the load is

$$P_L = \frac{1}{2}\left|\frac{V_{TH}}{2Z_c}\right|^2 Z_c = \left(\frac{R_2}{R_1 + R_2 + Z_c}\right)^2 \frac{|V_g|^2}{8Z_c}$$

The available power is $|V_g|^2/8Z_c$, so that the power attenuation K^2 is given by

$$K^2 = \left(\frac{R_2}{R_1 + R_2 + Z_c}\right)^2$$

When K has been specified, and also requiring that $R_{in} = Z_c$, we have two equations that are readily solved for the required values of R_1 and R_2. Thus we find that

$$R_1 = \frac{1 - K}{1 + K} Z_c \tag{6.6a}$$

$$R_2 = \frac{2K}{1 - K^2} Z_c \tag{6.6b}$$

For 10-dB attenuation in a 50-Ω system, we require $K = \sqrt{0.1}$ which results in $R_1 = 25.97\ \Omega$ and $R_2 = 35.14\ \Omega$. For a 3-dB attenuator, we require $R_1 = 8.58\ \Omega$ and $R_2 = 141.4\ \Omega$. Thus, in order to produce a wide range of attenuation values, R_1 and R_2 must be capable of being varied over a substantial range of values. For the Π network, R_1 and R_2 are given by

$$R_1 = \frac{1 + K}{1 - K} Z_c \tag{6.7a}$$

$$R_2 = \frac{1 - K^2}{2K} Z_c \tag{6.7b}$$

The *PIN* diode differs from a conventional *PN* junction diode by having a thin layer of *I*ntrinsic semiconductor material between the usual *P*ositive and *N*egative doped regions. The addition of the intrinsic region

reduces the junction capacitance, since the P and N regions are further apart. It also makes the forward conductivity of the diode a much more linear function of the diode bias current. In the forward bias condition, charge carriers are injected into the intrinsic layer and thus the conductivity is proportional to the injected charge, which in turn is proportional to the bias current. The equivalent circuit of the *PIN* diode is shown in Fig. 6.9*a*. The resistance R_i and capacitance C_i represent the intrinsic layer. The resistance R_j, the junction capacitance C_j, and the diffusion capacitance C_d represent the *PN* diode characteristics. The resistance R_s is that of the bulk semiconductor regions. The capacitance C_f is the stray capacitance between the terminals of the chip and can usually be neglected. For the packaged diode the equivalent circuit also includes the lead inductance L_s and the package capacitance C_p.

Under the reverse bias conditions, the diffusion capacitance is negligible and the junction resistance R_j is very large. Also the parallel impedance of R_i and C_i is negligible relative to the impedance of C_j and R_s in series; so the equivalent circuit reduces to that shown in Fig. 6.9*b*. A typical value

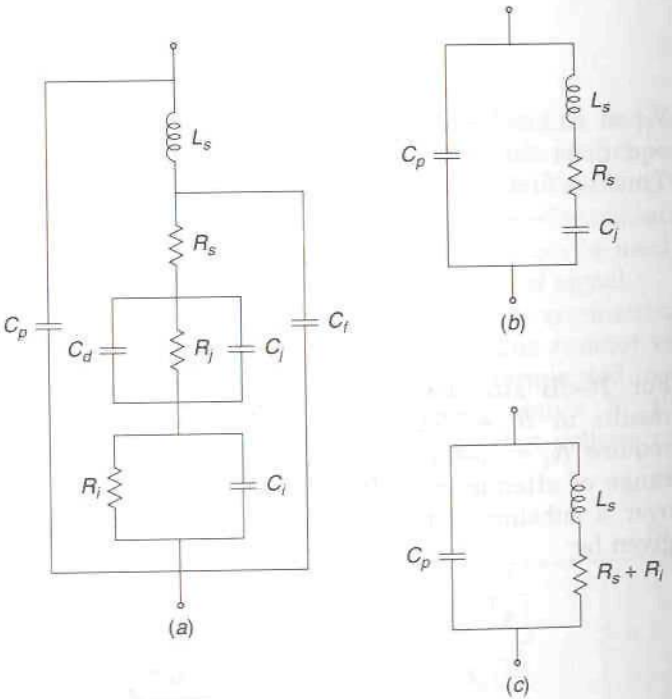


FIGURE 6.9
(*a*) Equivalent microwave circuit of a *PIN* diode; (*b*) equivalent circuit for reverse bias conditions; (*c*) equivalent circuit for forward bias conditions.

for the impedance of the diode in the off state is predominantly a capacitive reactance of $-j40$ to $-j400\ \Omega$. For forward bias conditions, the diffusion capacitance C_d is large and provides a low-reactance shunt across the junction. Also C_i is small, so that $R_i + R_s$ becomes the controlling element for the diode impedance. The equivalent circuit under forward bias conditions is shown in Fig. 6.9*c*. In the fully on state, the resistance R_i of the intrinsic layer is less than 1 Ω in a typical diode. The resistance R_i varies inversely with the bias current.

PIN diodes in the forward biased state can be used for the resistors R_1 and R_2 in the attenuator networks shown in Fig. 6.8. The presence of reactive elements in the equivalent *PIN* diode circuit will result in some inevitable input and output mismatch. The dynamic range of an attenuator using three *PIN* diodes as variable resistors is also limited to usually 10 dB or less.

A more versatile attenuator can be built by using chip resistors to construct a cascade connection of Π networks and using *PIN* diodes as switches to short-circuit the series arms for low attenuation or to be in the open state for one or more sections for various amounts of attenuation. An illustration of this type of attenuator is shown in Fig. 6.10. A basic section producing 2 dB of attenuation in a 50-Ω system would have $R_1 = 436.2\ \Omega$ and $R_2 = 11.61\ \Omega$. In the fully forward biased on state, the *PIN* diode short-circuits R_2 and, since the parallel combination of R_1 with a second R_1 is a resistance of 218.1 Ω, the loading effect on a 50-Ω line is relatively small, so that attenuation is also quite small.

A field-effect transistor with zero voltage applied between the drain and source has a channel whose resistance can be controlled by the gate voltage. Thus FETs can be used as the variable resistors in one of the basic attenuator circuits shown in Fig. 6.8. In Fig. 6.11 we show a photograph of a commercial electronically controlled attenuator built as a MMIC circuit. This attenuator will operate over the frequency range from dc to 26 GHz and can produce a maximum attenuation greater than 30 dB. The return loss is greater than 10 dB and the insertion loss is 1.5 to 2.5 dB. This particular device is suitable for use in a microstrip circuit. The attenuator uses six MESFET devices as variable resistors in a basic T attenuation circuit.

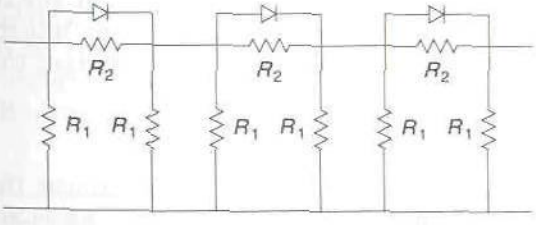


FIGURE 6.10
A switchable attenuator network. *PIN* diodes are used to switch R_2 in and out of the circuit.

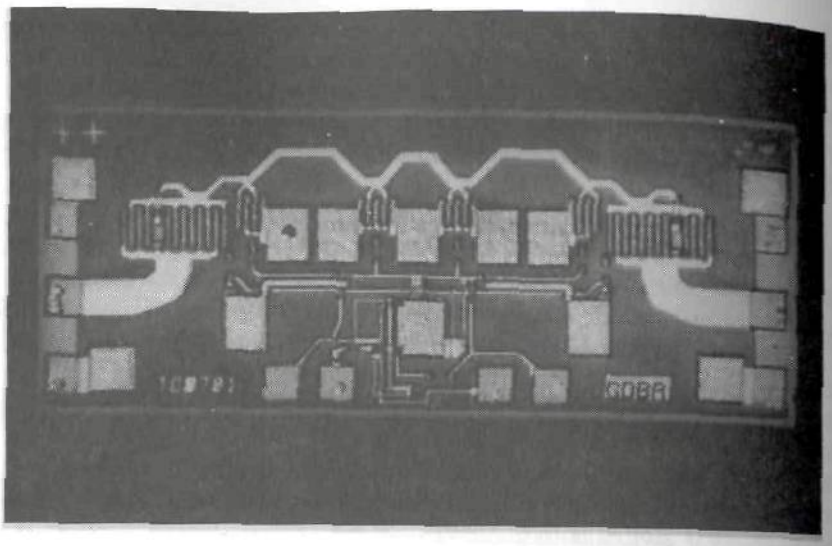


FIGURE 6.11
A photograph of a MMIC electronically controlled attenuator. (*Photograph courtesy of Ray Moskaluk, Hewlett-Packard Company.*)

6.3 PHASE SHIFTERS

A phase shifter is an instrument that produces an adjustable change in the phase angle of the wave transmitted through it. Ideally, it should be perfectly matched to the input and output lines and should produce zero attenuation. These requirements can be met to within a reasonable degree of approximation. There are a variety of designs for phase shifters of the mechanically adjustable type. The rotary phase shifter is the best in this class and is the only one we will discuss.

Electronically controlled phase shifters using *PIN* diodes as switches have become popular for use in phased-array antennas. In a phased-array antenna, there are many radiating elements, such as printed-circuit dipoles, and the radiated beam can be scanned in direction by varying the phase of excitation of each element in the array. In a large array this requires many phase shifters; so a design that is small, electronically controlled, and can be economically produced in large quantities using integrated circuit technology is desirable. The *PIN* diode phase shifters meet these requirements.

Rotary Phase Shifter

The rotary phase shifter is a precision instrument that is widely used in microwave measurements. Its basic construction is similar to that of the

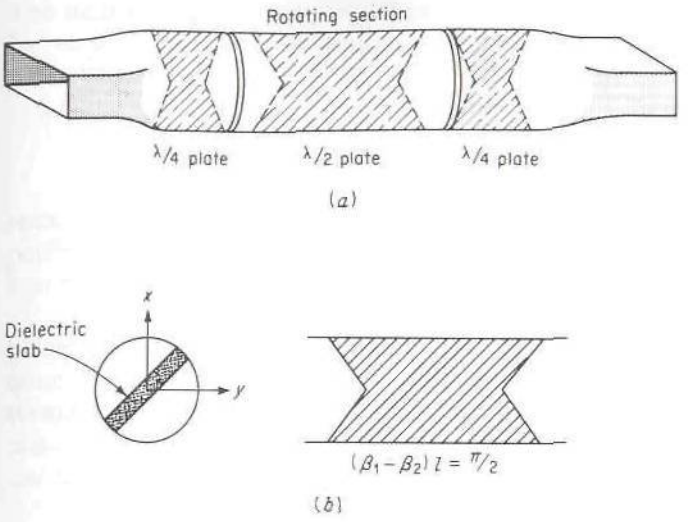


FIGURE 6.12
(*a*) Rotary phase shifter; (*b*) quarter-wave plate.

rotary attenuator, except that the center resistive card is replaced by a half-wave plate and the two outer resistive cards are replaced by quarter-wave plates. The quarter-wave plates convert a linearly polarized TE_{11} mode into a circularly polarized mode, and vice versa. The half-wave plate produces a phase shift equal to twice the angle θ through which it is rotated. The analysis of the principles of operation is given below.

A circularly polarized field is a field with x and y components of electric field that are equal in magnitude but 90° apart in time phase.† A quarter-wave plate is a device that will produce a circularly polarized wave when a linearly polarized wave is incident upon it. Figure 6.12 illustrates the basic components of the rotary phase shifter. The quarter-wave plate may be constructed from a slab of dielectric material, as illustrated in Fig. 6.12*b*. When the TE_{11} mode is polarized parallel to the slab, the propagation constant β_1 is greater than for the case when the mode is polarized perpendicular to the slab; that is, $\beta_1 > \beta_2$, where β_2 is the propagation constant for perpendicular polarization. The length l of a quarter-wave plate is chosen to obtain a differential phase change $(\beta_1 - \beta_2)l$ equal to 90°. The ends of the dielectric slab are tapered to reduce reflections to a negligible value. The half-wave plate is similar in construction, except that its length is increased to produce a differential phase change of 180°.

†Circularly polarized fields are discussed in greater detail in Sec. 6.7.

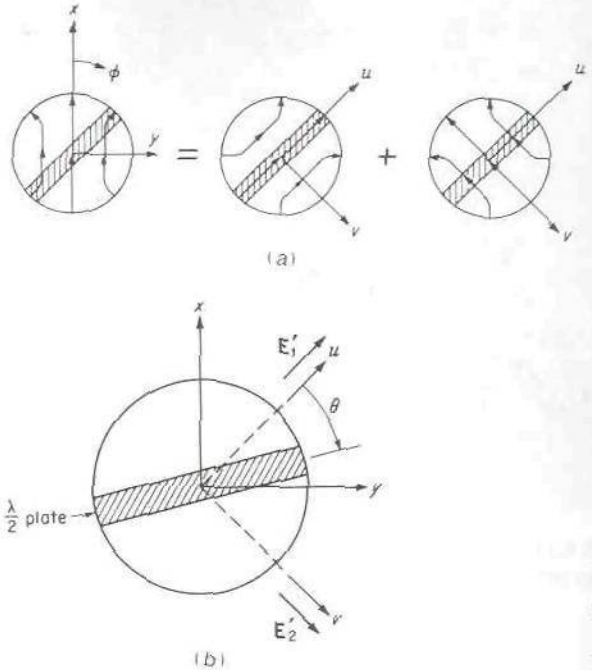


FIGURE 6.13
(*a*) Decomposition of incident TE_{11} mode; (*b*) orientation of half-wave plate.

In the rotary phase shifter, the quarter-wave plates are oriented at an angle of 45° relative to the broad wall of the rectangular guide. The incoming TE_{11} mode may be decomposed into two modes polarized parallel and perpendicular to the quarter-wave plate, as illustrated in Fig. 6.13*a*. The incident mode is assumed given by (6.2) as

$$\mathbf{E} = \frac{J_1}{r}\mathbf{a}_r \cos\phi - \frac{p'_{11}}{a}J'_1\mathbf{a}_\phi \sin\phi$$

If we replace $\cos\phi$ by

$$\cos\left(\phi - \frac{\pi}{4} + \frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}\left[\cos\left(\phi - \frac{\pi}{4}\right) - \sin\left(\phi - \frac{\pi}{4}\right)\right]$$

and $\sin\phi$ by

$$\frac{\sqrt{2}}{2}\left[\cos\left(\phi - \frac{\pi}{4}\right) + \sin\left(\phi - \frac{\pi}{4}\right)\right]$$

the above expression for the incident field may be written as

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \tag{6.8a}$$

where

$$\mathbf{E}_1 = \frac{\sqrt{2}}{2}\left[\frac{J_1}{r}\mathbf{a}_r \cos\left(\phi - \frac{\pi}{4}\right) - \frac{p'_{11}}{a}J'_1\mathbf{a}_\phi \sin\left(\phi - \frac{\pi}{4}\right)\right] \tag{6.8b}$$

$$\mathbf{E}_2 = -\frac{\sqrt{2}}{2}\left[\frac{J_1}{r}\mathbf{a}_r \sin\left(\phi - \frac{\pi}{4}\right) + \frac{p'_{11}}{a}J'_1\mathbf{a}_\phi \cos\left(\phi - \frac{\pi}{4}\right)\right] \tag{6.8c}$$

The field $\mathbf{E}_1$ is polarized parallel to the slab, and $\mathbf{E}_2$ is polarized perpendicular to the slab. After propagation through the quarter-wave plate, these fields become

$$\mathbf{E}_1' = \mathbf{E}_1 e^{-j\beta_1 l} \tag{6.9a}$$

$$\mathbf{E}_2' = \mathbf{E}_2 e^{-j\beta_2 l} = \mathbf{E}_2 e^{-j\beta_1 l} e^{-j(\beta_2 - \beta_1)l} = j\mathbf{E}_2 e^{-j\beta_1 l} \tag{6.9b}$$

since $(\beta_2 - \beta_1)l = -\pi/2$. The resultant field consists of two orthogonally polarized TE_{11} modes of equal amplitude and 90° apart in time phase, and hence constitutes a circularly polarized field.

Consider next the action of the half-wave plate on the above circularly polarized field. Let the half-wave plate be rotated by an angle θ past the quarter-wave plate, as in Fig. 6.13*b*. The field $\mathbf{E}_1' + \mathbf{E}_2'$ may be expressed in terms of TE_{11} modes polarized parallel and perpendicular to the half-wave plate by changing the origin of the angle variable ϕ to $\pi/4 + \theta$; that is, we use the relations

$$\cos\left(\phi - \frac{\pi}{4}\right) = \cos\left(\phi - \frac{\pi}{4} - \theta + \theta\right) = \cos\theta \cos\left(\phi - \theta - \frac{\pi}{4}\right) - \sin\theta \sin\left(\phi - \theta - \frac{\pi}{4}\right)$$

and

$$\sin\left(\phi - \frac{\pi}{4}\right) = \sin\theta \cos\left(\phi - \theta - \frac{\pi}{4}\right) + \cos\theta \sin\left(\phi - \theta - \frac{\pi}{4}\right)$$

Thus we obtain

$$\mathbf{E}_1' = \frac{\sqrt{2}}{2} e^{-j\beta_1 l} \left\{ \cos\theta \left[\frac{J_1}{r} \mathbf{a}_r \cos\left(\phi - \theta - \frac{\pi}{4}\right) - \frac{p_{11}'}{a} J_1' \mathbf{a}_\phi \sin\left(\phi - \theta - \frac{\pi}{4}\right) \right] - \sin\theta \left[\frac{J_1}{r} \mathbf{a}_r \sin\left(\phi - \theta - \frac{\pi}{4}\right) + \frac{p_{11}'}{a} J_1' \mathbf{a}_\phi \cos\left(\phi - \theta - \frac{\pi}{4}\right) \right] \right\} \tag{6.10a}$$

$$\mathbf{E}_2' = \frac{-j\sqrt{2}}{2} e^{-j\beta_1 l} \left\{ \cos\theta \left[\frac{J_1}{r} \mathbf{a}_r \sin\left(\phi - \theta - \frac{\pi}{4}\right) + \frac{p_{11}'}{a} J_1' \mathbf{a}_\phi \cos\left(\phi - \theta - \frac{\pi}{4}\right) \right] + \sin\theta \left[\frac{J_1}{r} \mathbf{a}_r \cos\left(\phi - \theta - \frac{\pi}{4}\right) - \frac{p_{11}'}{a} J_1' \mathbf{a}_\phi \sin\left(\phi - \theta - \frac{\pi}{4}\right) \right] \right\} \tag{6.10b}$$

The field polarized parallel to the half-wave plate has an r component varying as $\cos(\phi - \theta - \pi/4)$, whereas the perpendicularly polarized mode has an r component of electric field varying as $\sin(\phi - \theta - \pi/4)$. Hence,

from (6.10), we obtain

$$\mathbf{E}'_1 + \mathbf{E}'_2 = \mathbf{E}''_1 + \mathbf{E}''_2 \tag{6.11a}$$

where

$$\mathbf{E}''_1 = \frac{\sqrt{2}}{2} e^{-j\beta_1 l}(\cos\theta - j\sin\theta)\left[\frac{J_1}{r}\mathbf{a}_r \cos\left(\phi - \theta - \frac{\pi}{4}\right) - \frac{p'_{11}}{a} J'_1 \mathbf{a}_\phi \sin\left(\phi - \theta - \frac{\pi}{4}\right)\right]$$

$$= \frac{\sqrt{2}}{2} e^{-j\beta_1 l - j\theta}\left[\frac{J_1}{r}\mathbf{a}_r \cos\left(\phi - \theta - \frac{\pi}{4}\right) - \frac{p'_{11}}{a} J'_1 \mathbf{a}_\phi \sin\left(\phi - \theta - \frac{\pi}{4}\right)\right] \tag{6.11b}$$

$$\mathbf{E}''_2 = \frac{-j\sqrt{2}}{2} e^{-j\beta_1 l - j\theta}\left[\frac{J_1}{r}\mathbf{a}_r \sin\left(\phi - \theta - \frac{\pi}{4}\right) + \frac{p'_{11}}{a} J'_1 \mathbf{a}_\phi \cos\left(\phi - \theta - \frac{\pi}{4}\right)\right] \tag{6.11c}$$

After propagating through the half-wave plate of length $2l$, this field becomes

$$\mathbf{E}_3 = \mathbf{E}''_1 e^{-2j\beta_1 l} \tag{6.12a}$$

$$\mathbf{E}_4 = \mathbf{E}''_2 e^{-2j\beta_2 l} = \mathbf{E}''_2 e^{-2j(\beta_2 - \beta_1)l - 2j\beta_1 l} = -\mathbf{E}''_2 e^{-2j\beta_1 l} \tag{6.12b}$$

since $2(\beta_2 - \beta_1)l = -\pi$.

This new field may now be decomposed once again into two TE_{11} modes polarized parallel and perpendicular to the quarter-wave plate in the output guide. If we assume that this plate is parallel to the input quarter-wave plate, we can obtain the required decomposition by referring the angle variable ϕ to $\pi/4$ as the origin. If we follow the procedure used earlier, we obtain

$$\mathbf{E}_3 + \mathbf{E}_4 = \mathbf{E}'_3 + \mathbf{E}'_4 \tag{6.13a}$$

where

$$\mathbf{E}'_3 = \frac{\sqrt{2}}{2} e^{-3j\beta_1 l - j2\theta}\left[\frac{J_1}{r}\mathbf{a}_r \cos\left(\phi - \frac{\pi}{4}\right) - \frac{p'_{11}}{a} J'_1 \mathbf{a}_\phi \sin\left(\phi - \frac{\pi}{4}\right)\right] \tag{6.13b}$$

$$\mathbf{E}'_4 = \frac{j\sqrt{2}}{2} e^{-3j\beta_1 l - j2\theta}\left[\frac{J_1}{r}\mathbf{a}_r \sin\left(\phi - \frac{\pi}{4}\right) + \frac{p'_{11}}{a} J'_1 \mathbf{a}_\phi \cos\left(\phi - \frac{\pi}{4}\right)\right] \tag{6.13c}$$

Finally, after propagating through the second quarter-wave plate, the output field becomes

$$\mathbf{E}_0 = \mathbf{E}''_3 + \mathbf{E}''_4 \tag{6.14a}$$

where
$$\mathbf{E}_3'' = \mathbf{E}_3' e^{-j\beta_1 l} \tag{6.14b}$$
$$\mathbf{E}_4'' = \mathbf{E}_4' e^{-j\beta_2 l} = j\mathbf{E}_4' e^{-j\beta_1 l} \tag{6.14c}$$

When the fields $\mathbf{E}_3''$ and $\mathbf{E}_4''$ are combined, we obtain

$$\mathbf{E}_0 = e^{-4j\beta_1 l - 2j\theta}\left(\frac{J_1}{r}\mathbf{a}_r \cos\phi - \frac{p'_{11}}{a}J_1'\mathbf{a}_\phi \sin\phi\right) \tag{6.15}$$

which is again a linearly polarized TE_{11} mode having the same direction of polarization as the incident field given by (6.2) and (6.8). Note, however, that the phase has been changed by an amount $4\beta_1 l + 2\theta$. Thus rotation of the half-wave plate through an angle θ changes the phase of the transmitted wave by an amount 2θ. This simple dependence of the phase change on a mechanical rotation is the chief advantage of the rotary phase-shifter.

Besides dielectric slabs, the circular guide may be loaded with metallic fins or rods to produce 90° and 180° differential phase-shift sections. These methods are discussed in a paper by Fox.†

Electronically Controlled Phase Shifters

There are several basic designs that are used to build digital-type electronically controlled phase shifters. In all designs the *PIN* diodes are used to switch circuit elements in and out of the transmission path. Each switching operation adds or subtracts a finite phase-shift increment such as ±11.25°, ± 22.5°, ± 45°, ± 90°, etc. By using a cascade connection of several phase shifters, the full range 0° to 180° can be covered with a resolution equal to the smallest phase increment that is made available.

The simplest phase-shifter design uses *PIN* diode switches to switch one of two alternate transmission-line lengths into the transmission path as shown in Fig. 6.14. The bias currents can be supplied by connections at the midpoint of a half-wave open-circuited stub. The first quarter-wave section is a low-impedance stub and transforms the open-circuit impedance to a short circuit or low impedance at the point where the bias line is connected. The next quarter-wave section uses a high-impedance line and transforms the low impedance of the midpoint into a high impedance which produces negligible loading of the main transmission line. The dc return for the bias current can be obtained by connecting the input and output lines to the ground plane through short-circuited high-impedance quarter-wave-line sections as shown in Fig. 6.14.

If the two transmission-line sections have lengths l_1 and l_2, then the incremental phase change produced when line 2 is switched in to replace line 1 is $\Delta\phi = \beta(l_2 - l_1)$, where β is the propagation phase constant. This

†A. G. Fox, An Adjustable Waveguide Phase Changer, *Proc. IRE*, vol. 35, pp. 1489–1498, December, 1947.

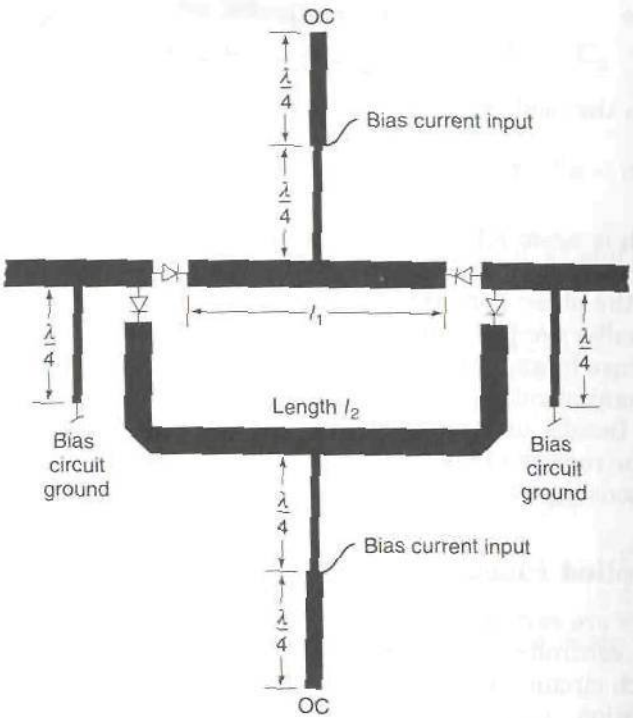


FIGURE 6.14
Incremental-line-type phase shifter.

type of phase shifter produces an incremental change in phase that depends on the frequency since β is a function of frequency. If the dispersion of the line is small, i.e., phase velocity is independent of frequency, then the increment in signal time delay is a constant and is given by $\Delta\tau \approx (l_2 - l_1)/v_p = (l_2 - l_1)\beta/\omega$. When broadband signals are to be radiated by a phased array, the use of different phases for the excitation of the various elements in the array will generally result in some signal distortion unless the phase shifters are true time-delay devices.

Since the *PIN* diodes are not ideal switches and have a finite on resistance and some series lead inductance, the use of several phase shifters in series will result in a significant overall insertion loss.

A second type of phase shifter is shown in Fig. 6.15. In this design two susceptances of equal value jB are switched in shunt with the main transmission line when the diodes are on. The stubs are used to obtain the required susceptances and are usually short-circuited at the far end to provide a dc return for the bias current. The stubs are spaced a distance l apart; so the phase shift is $\beta l = \theta$ when the diodes are off. The spacing l is chosen so that when the stubs are switched into the circuit the resultant network will still present an input and output impedance equal to Z_r to the connecting transmission lines.

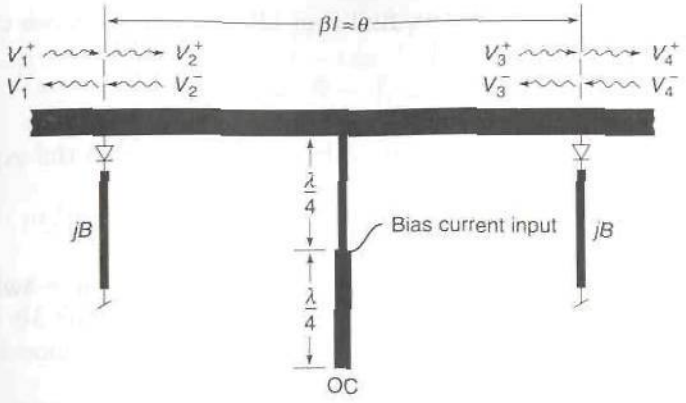


FIGURE 6.15
A phase shifter using switched reactive elements.

We can analyze this circuit in a straightforward way using wave-amplitude transmission matrices (Sec. 4.9). The wave amplitudes at various points in the circuit are designated by $V_1^+, V_1^-, V_2^+, V_2^-$, etc., as shown in Fig. 6.15. The plus superscript designates the amplitudes of waves propagating to the right, while the minus superscript is used for the amplitudes of the waves propagating to the left. The wave-amplitude transmission matrices for a normalized shunt susceptance $j\overline{B}$ and a section of transmission line of electrical length θ are

$$[A_1] = \begin{bmatrix} 1 + j\dfrac{\overline{B}}{2} & j\dfrac{\overline{B}}{2} \\ -j\dfrac{\overline{B}}{2} & 1 - j\dfrac{\overline{B}}{2} \end{bmatrix} \qquad [A_2] = \begin{bmatrix} e^{j\theta} & 0 \\ 0 & e^{-j\theta} \end{bmatrix}$$

We can now write

$$\begin{bmatrix} V_1^+ \\ V_1^- \end{bmatrix} = [A_1]\begin{bmatrix} V_2^+ \\ V_2^- \end{bmatrix} = [A_1][A_2]\begin{bmatrix} V_3^+ \\ V_3^- \end{bmatrix} = [A_1][A_2][A_3]\begin{bmatrix} V_4^+ \\ V_4^- \end{bmatrix}$$

$$= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} V_4^+ \\ V_4^- \end{bmatrix}$$

If we choose $V_4^- = 0$, then $V_4^+ = V_1^+/A_{11}$; so the transmission coefficient from port 1 to port 4 is $T_{14} = A_{11}^{-1}$. By multiplying out the matrix product shown above, it is readily found that

$$T_{14} = \left[\left(1 + j\frac{\overline{B}}{2}\right)^2 e^{j\theta} + \frac{\overline{B}^2}{4} e^{-j\theta}\right]^{-1}$$

After some algebra we find that $|T_{14}| = 1$ provided we choose

$$\tan\theta = \frac{2}{\overline{B}} \tag{6.16}$$

When we use this relationship to replace $\overline{B}/2$ in the expression for T_{14}, we obtain

$$T_{14} = -e^{j\theta} = e^{-j(\pi-\theta)} \tag{6.17}$$

Thus the change in phase when the diodes are switched on is $\Delta\phi = (\pi - \theta) - \theta = \pi - 2\theta$. For $\overline{B} = 2$, $\theta = \pi/4$ and $\Delta\phi = \pi/2$ or 90°. For $\overline{B} = 1$, $\theta = 1.107$ rad and $\Delta\phi = 53.14°$. Thus, by choosing $\overline{B}$ appropriately, a wide range of phase shifts can be obtained.

To a first approximation the diode impedance in the on state is a small series inductive reactance. This should be included as part of the susceptance that is switched into the line. In the off state the diode impedance is due to the shunt capacitance. Thus the shunt capacitance must be small if the phase shifter is to work properly. Since the off-state impedance is not always that large, an alternative design that allows some compensation for both the on-state and off-state diode impedance can be used. This alternate design is shown in Fig. 6.16. In this design two *PIN* diodes are used to open-circuit or short-circuit the tap points P and P' of two stubs that are connected in shunt to the main transmission line. The stubs are open-circuited at the remote end if shunt switches are used.

Since the condition (6.16) for a matched circuit cannot be simultaneously met for two different susceptances, this last design requires that the two susceptances $j\overline{B}_1$ and $j\overline{B}_2$ that load the main line when the diodes are on and off be small. It is preferable to choose $j\overline{B}_2 = -j\overline{B}_1$ so as to maximize the phase difference in the transmission coefficient for the two states. By spacing the stubs $\lambda/4$ apart, the reflections from the two stubs almost cancel. The stub lengths d and tap point distance d_1 can be chosen so that $\overline{B}_2 = -\overline{B}_1$ including the diode reactances in the on and off states.

When $\theta = \pi/2$ the use of wave transmission matrices shows that

$$T_{14} = \left[j\left(1 + j\frac{\overline{B}}{2}\right)^2 - j\frac{\overline{B}^2}{4} \right]^{-1} = -j\left[1 + j\overline{B} - \frac{\overline{B}^2}{2}\right]^{-1}$$

$$\approx -j(1 + j\overline{B})^{-1}$$

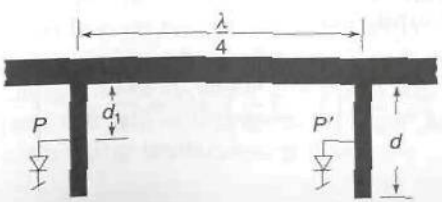


FIGURE 6.16
A phase shifter using open-circuited stubs spaced $\lambda/4$ apart. P and P' are switched into the circuit when the diodes are off and on, respectively. The bias circuit is not shown.

when $\bar{B} \ll 1$. Hence $|T_{14}|^2 = (1 + \bar{B}^2)^{-1}$ which is close to one. The phase angle of T_{14} is $-\pi/2 - \tan^{-1} \bar{B} \approx -\pi/2 - \bar{B}$. The changes in phase between the states when $\bar{B} = \bar{B}_1$ and $-\bar{B}_1$ is thus $2\bar{B}_1$. A value of $\bar{B}_1$ as large as 0.2 would not produce a significant mismatch and would result in a change in phase $\Delta\phi = 0.4 = 22.92°$ between the two states. This type of phase shifter is limited to relatively small phase shifts between states in order to keep the input VSWR small.

A variety of other phase shifter circuits are also possible. A good summary of various circuits that have been analyzed and the performance that can be obtained can be found in the literature.†

6.4 DIRECTIONAL COUPLERS

A directional coupler is a four-port microwave junction with the properties discussed below. With reference to Fig. 6.17, which is a schematic illustration of a directional coupler, the ideal directional coupler has the property that a wave incident in port 1 couples power into ports 2 and 3 but not into port 4. Similarly, power incident in port 4 couples into ports 2 and 3 but not into port 1. Thus ports 1 and 4 are uncoupled. For waves incident in port 2 or 3, the power is coupled into ports 1 and 4 only, so that ports 2 and 3 are also uncoupled. In addition, all four ports are matched. That is, if three ports are terminated in matched loads, the fourth port appears terminated in a matched load, and an incident wave in this port suffers no reflection.

Directional couplers are widely used in impedance bridges for microwave measurements and for power monitoring. For example, if a radar transmitter is connected to port 1, the antenna to port 2, a microwave crystal detector to port 3, and a matched load to port 4, the power received in port 3 is proportional to the power flowing from the transmitter to the antenna in the forward direction only. Since the reflected wave from the antenna, if it exists, is not coupled into port 3, the detector monitors the power output of the transmitter.

If the coupler is designed for 3-dB coupling, then it splits the input power in port 1 into equal powers in ports 2 and 3. Thus a 3-dB directional coupler serves as a power divider. Directional couplers with 3-dB coupling are also called hybrid junctions and are widely used in microwave mixers and as input and output couplers in balanced microwave amplifier circuits. There are many available designs and configurations for directional couplers, hybrid junctions, and power dividers. Space limitations will allow us to only examine some of these.†

---

†I. Bahl and P. Bhartia, "Microwave Solid State Circuit Design," John Wiley & Sons, Inc., New York, 1988.

K. Chang (ed.), "Handbook of Microwave and Optical Components, Microwave Solid State Components," vol. 2, John Wiley & Sons, Inc., New York, 1990.

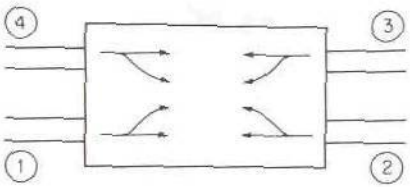


FIGURE 6.17
A directional coupler. Arrows indicate the direction of power flow.

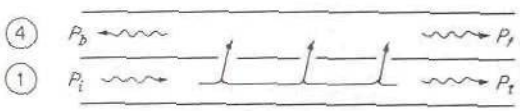


FIGURE 6.18
Aperture-coupled-waveguide directional coupler.

Directional couplers using waveguides usually consist of two rectangular waveguides with coupling apertures located in a common wall as illustrated in Fig. 6.18. Since these devices are required to operate over a band of frequencies, it is not possible to obtain ideal performance over the whole frequency band. The performance of a directional coupler is measured by two parameters, the *coupling* and the *directivity*. Let P_i be the incident power in port 1, and let P_f be the coupled power in the forward direction in arm 3. The coupling in decibels is then given by

$$C = 10 \log \frac{P_i}{P_f} \tag{6.18}$$

Ideally, the power P_b coupled in the backward direction in arm 4 should be zero. The extent to which this is achieved is measured by the directivity D, which is defined as

$$D = 10 \log \frac{P_f}{P_b} \tag{6.19}$$

The directivity is a measure of how well the power can be coupled in the desired direction in the second waveguide.

A number of properties of the ideal directional coupler may be deduced from the symmetry and unitary properties of its scattering matrix (Sec. 4.7). The least stringent definition of a directional coupler as illustrated in Fig. 6.17 is that it is a four-port junction with

$$S_{14} = S_{23} = 0$$

$S_{11} = S_{22} = 0$, that is, ports 1 and 2 matched, and the coupling elements S_{12}, S_{13}, S_{24}, and S_{34} are not equal to zero. The scattering matrix then has

†A very good overall survey of types of directional couplers, hybrid junctions, and power dividers can be found in K. Chang, "Handbook of Microwave Components, Microwave Passive Components and Antennas," vol. 1, John Wiley & Sons, Inc., New York, 1990.

the form

$$[S] = \begin{bmatrix} 0 & S_{12} & S_{13} & 0 \\ S_{12} & 0 & 0 & S_{24} \\ S_{13} & 0 & S_{33} & S_{34} \\ 0 & S_{24} & S_{34} & S_{44} \end{bmatrix}$$

If we form the product of row 1 with the complex conjugate of row 3, and also the product of row 2 with the conjugate of row 4, we obtain

$$S_{13}S_{33}^* = 0 \qquad S_{24}S_{44}^* = 0$$

because of the unitary nature of the scattering matrix. Since S_{13} and S_{24} are assumed to be nonzero, these equations show that $S_{33} = S_{44} = 0$; that is, all four ports are matched. Thus the scattering matrix becomes

$$[S] = \begin{bmatrix} 0 & S_{12} & S_{13} & 0 \\ S_{12} & 0 & 0 & S_{24} \\ S_{13} & 0 & 0 & S_{34} \\ 0 & S_{24} & S_{34} & 0 \end{bmatrix} \tag{6.20}$$

If we take the product of row 1 with the conjugate of row 4, and similarly row 2 with the conjugate of row 3, we now find that

$$S_{12}S_{24}^* + S_{13}S_{34}^* = 0 \qquad S_{12}S_{13}^* + S_{24}S_{34}^* = 0$$

If we note that $|S_{12}S_{24}^*| = |S_{12}|\,|S_{24}|$, these equations are seen to give

$$|S_{12}|\,|S_{24}| = |S_{13}|\,|S_{34}| \tag{6.21a}$$

$$|S_{12}|\,|S_{13}| = |S_{24}|\,|S_{34}| \tag{6.21b}$$

When we divide the first equation by the second equation, we obtain

$$\frac{|S_{24}|}{|S_{13}|} = \frac{|S_{13}|}{|S_{24}|}$$

or

$$|S_{13}| = |S_{24}| \tag{6.22a}$$

thus the coupling between ports 1 and 3 is the same as that between ports 2 and 4. Use of (6.22a) in (6.21a) also gives

$$|S_{12}| = |S_{34}| \tag{6.22b}$$

so that the coupling between ports 1 and 2 equals that between ports 3 and 4 also.

The product of the first row with its conjugate equals unity, so that

$$|S_{12}|^2 + |S_{13}|^2 = 1 \tag{6.23a}$$

similarly,

$$|S_{12}|^2 + |S_{24}|^2 = 1 \tag{6.23b}$$

By choosing the terminal plane in arm 1 properly, we can adjust the phase

angle of S_{12} so that S_{12} is real [see (4.53)]. Thus let S_{12} be a real positive number C_1. Similarly, by choosing the terminal plane in arm 3 properly, we can make S_{13} a positive imaginary quantity jC_2, where C_2 is real and positive. We now have

$$C_1^2 + C_2^2 = 1 \tag{6.24}$$

We can choose the reference plane in arm 4 so as to make S_{34} real and thus equal to C_1 by virtue of (6.22*b*). It is now necessary for S_{24} to be equal to jC_2 since $S_{12}S_{24}^* + S_{13}S_{34}^* = 0$, as given earlier. Thus the simplest form for the scattering matrix of an ideal directional coupler is

$$[S] = \begin{bmatrix} 0 & C_1 & jC_2 & 0 \\ C_1 & 0 & 0 & jC_2 \\ jC_2 & 0 & 0 & C_1 \\ 0 & jC_2 & C_1 & 0 \end{bmatrix} \tag{6.25}$$

where $C_2 = (1 - C_1^2)^{1/2}$ from (6.24).

It may also be shown from the unitary properties of the scattering matrix of a lossless reciprocal four-port junction that if all four ports are matched, the device must be a directional coupler.†

Directional-Coupler Designs

There are a great variety of ways of constructing directional couplers. Some of the more common aperture-coupled types are described below. Their design is based on the small-aperture-coupling theory presented in Sec. 4.13. This theory was originally developed by Bethe.‡

Bethe-Hole Coupler

The Bethe-hole directional coupler consists of two rectangular waveguides coupled by means of a small circular aperture located in the center of the common broad wall. To achieve directional coupling, the axis of the two guides must be at an angle θ, as illustrated in Fig. 6.19*a*. A variation of this design consists of a similar arrangement, with $\theta = 0$, but an offset aperture as in Fig. 6.19*b*.

The theory for the coupler in Fig. 6.19*b* was given in Sec. 4.13. An incident TE_{10} mode in guide 1, with an amplitude A, produces a normal electric dipole in the aperture plus a tangential magnetic dipole proportional and in the same direction as the magnetic field of the incident wave. In the

†C. G. Montgomery, R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits," sec. 9.10, McGraw-Hill Book Company, New York, 1948.

‡H. A. Bethe, Theory of Diffraction by Small Holes, *Phys. Rev.*, vol. 66, pp. 163–182, 1944.

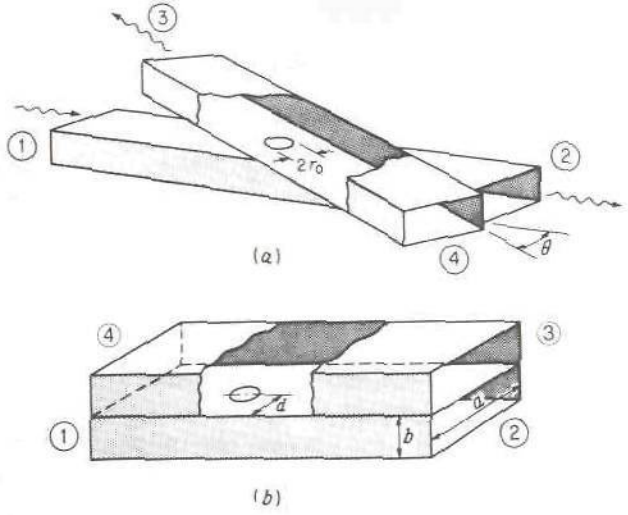


FIGURE 6.19
Bethe-hole directional coupler.

upper guide the normal electric dipole and the axial component of the magnetic dipole radiate symmetrically in both directions. The transverse component of the magnetic dipole radiates antisymmetrically. By varying the angle θ or the aperture position d, the amplitude of the fields coupled into ports 3 and 4 can be controlled. For the directional coupler shown in Fig. 6.19*a*, the optimum value for the angle θ is given by

$$\cos\theta = \frac{\lambda_g^2}{2\lambda_0^2} \tag{6.26}$$

This choice for θ will minimize the field coupled into port 4. Since the coupling is not zero, a perfect directional coupler is not obtained. A detailed analysis shows that the coupling and directivity that can be obtained are given by †

$$C \approx 20 \log \frac{1}{\bar{X}\cos\theta} \tag{6.27a}$$

$$D \approx C + 20 \log \frac{2\cos\theta}{1 + \cos\theta} \tag{6.27b}$$

where $\bar{X} = 16\pi r_0^3/3ab\lambda_g$.

The directional coupler shown in Fig. 6.19*b* was analyzed in Sec. 4.13. When the spacing d from the side wall is chosen to satisfy the following

†R. E. Collin, "Field Theory of Guided Waves," 2nd ed., chap. 7, IEEE Press, Piscataway, N.J., 1991.

relation:

$$\sin \frac{\pi d}{a} = \frac{\lambda_0}{\sqrt{6}\, a} \tag{6.28}$$

there will be zero power coupled into port 4. The coupling into port 3 is given by (6.30*a*) with $1 + \overline{X}^2$ replaced by $\sqrt{1 + \overline{X}^2}$ and $A_1 + A_3$ interchanged with $A_2 + A_4$.

Directional-coupler characteristics are also obtained by choosing d to satisfy the equation†

$$\sin \frac{\pi d}{a} = \frac{\lambda_0}{\sqrt{2(\lambda_0^2 - a^2)}} \tag{6.29}$$

When $\lambda_0 = \sqrt{2}\, a$ the aperture will be located at the center, that is, $d = a/2$. For the above condition the field coupled into port 3 is minimized and that coupled into port 4 is a maximum. The coupling and directivity that can be achieved are given by [see (4.139) and (4.140) and let $\overline{X} = -\overline{B}$]

$$C = 20 \log \frac{|C|}{|A_2 + A_4|} = 20 \log \frac{1 + \overline{X}^2}{\overline{X}} \tag{6.30a}$$

$$D = 20 \log \left| \frac{A_2 + A_4}{A_1 + A_3} \right| = 20 \log \overline{X}^{-1} \tag{6.30b}$$

where $\overline{X} = (16\pi r_0^3 / 3ab\lambda_g)\sin^2(\pi d/a)$, These formulas apply only at the design frequency.

The above results are based on the assumption that the guide wall in which the aperture is located is infinitely thin. For the normal thickness used in a waveguide wall, the coupling will be 1 to 2 dB smaller.

Example 6.1. Design of a Bethe-hole directional coupler. We will design a directional coupler based on (6.28) and (6.30*a*). The waveguide dimensions are $a = 0.9$ in, $b = 0.4$ in. The center frequency is 9 GHz for which $\beta = [k_0^2 - (\pi/a)^2]^{1/2} = 1.29$ rad/cm. From (6.28) we obtain $\sin(\pi d/a) = 3.333/2.286\sqrt{6}$ which gives $d = 0.464$ cm. Thus the center of the aperture is located at 0.464 cm from the side of the waveguide. We will design the coupler for 30-dB coupling. From (6.30*a*) this requires that $(1 + \overline{X}^2)^{1/2}/\overline{X} = 31.623$ or $\overline{X} = 0.03164$. The required aperture radius is given by

$$r_0 = \left[\frac{3ab\lambda_g \overline{X}}{16\pi \sin^2(\pi d/a)} \right]^{1/3} = 0.392 \text{ cm}$$

This is already a large aperture; so clearly it would be difficult to obtain a coupling of 20 dB. The variation of the coupling and directivity as a function of frequency can be obtained by using the calculated values for d and r_0 in the

†R. E. Collin, *loc. cit.*

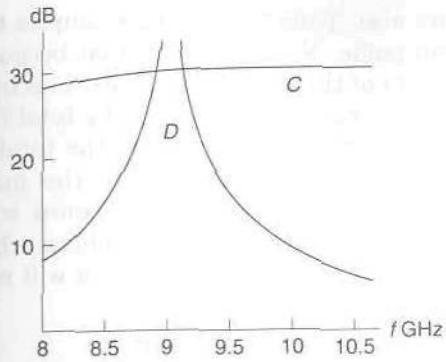


FIGURE 6.20
The coupling and directivity of the Bethe-hole directional coupler as a function of frequency for the coupler in Example 6.1.

expressions for $A_1 + A_3$ and $A_2 + A_4$ given by (4.140) and using these in (6.30). In Fig. 6.20 we show a plot of coupling C and directivity D as a function of frequency. The coupling remains almost constant. It increases from 28.46 dB at 8 GHz to 30.96 dB at 10.5 GHz. The directivity, however, drops rapidly as the frequency changes from 9 GHz. From these results we can see that the Bethe-hole coupler is a narrowband device. By using a second aperture on the opposite side of the center line the amplitudes of the coupled waves are doubled. This will increase the coupling by 6 dB without affecting the directivity. Note that increased coupling is measured by a decrease in the coupling C when expressed in decibels by (6.18).

Two-Hole Couplers

Two-hole couplers consist of two rectangular waveguides coupled by two identical apertures spaced a quarter guide wavelength $\lambda_g/4$ apart as in Fig. 6.21. The aperture may, in general, have directive properties, i.e., radiate a field with different amplitudes in the forward and reverse directions. With a wave of unit amplitude incident at port 1, let the field coupled into the second guide have an amplitude B_f in the forward direction and B_b in the backward direction. Since B_f and B_b are the amplitudes of the coupled fields for an incident wave of unit amplitude, they are called the aperture-coupling coefficients. If only a small amount of the incident power is coupled by the first aperture, the amplitude of the incident wave is essentially unity

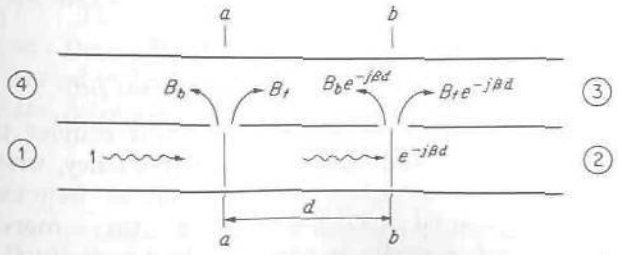


FIGURE 6.21
Two-hole directional coupler.

at the second aperture also. Thus this aperture couples the same amount of power into the second guide. Note, however, that because of the difference in path length, the phase of the field coupled by the second aperture is $-\beta d$ relative to that coupled by the first aperture. The total forward wave in the upper guide at the plane bb is $2B_f e^{-j\beta d}$, and the total backward wave at the plane aa is $B_b(1 + e^{-2j\beta d})$. Hence, since the forward-path lengths in the two guides are always the same, the forward waves always add in phase. The backward waves will add out of phase whenever $2\beta d = n\pi$, $n = 1, 3, 5, \ldots$. In particular, a value of $d = \lambda_g/4$ will result in cancelation of the backward wave. The coupling is given by

$$C = -20 \log 2|B_f| \tag{6.31a}$$

and the directivity is given by

$$D = 20 \log \frac{2|B_f|}{|B_b|\,|1 + e^{-2j\beta d}|} = 20 \log \frac{|B_f|}{|B_b|\,|\cos \beta d|}$$

$$= 20 \log \left| \frac{B_f}{B_b} \right| + 20 \log |\sec \beta d| \tag{6.31b}$$

The directivity is the sum of the inherent directivity of the single aperture plus a directivity associated with the array (in this case a two-element array only). Since B_f and B_b are the aperture-coupling parameters and are generally slowly varying functions of frequency, the coupling C is not particularly frequency-sensitive. However, the directivity is a sensitive function of frequency because of the sensitivity of the array factor sec βd.

Schwinger Reversed-Phase Coupler

The Schwinger reversed-phase coupler is designed to interchange the frequency sensitivity of the coupling C and directivity D. This is accomplished by making one aperture radiate a field which is the negative of that radiated by the other. With reference to Fig. 6.21, let the first aperture radiate fields B_f, B_b and the second aperture $-B_f$, $-B_b$. At plane bb in the upper guide, the total field is now $B_f - B_f = 0$ under all conditions. Hence port 3 is not coupled to port 1. At the plane aa the total field is, after accounting for the phase change due to propagation,

$$B_b - B_b e^{-2j\beta d} = e^{-j\beta d} B_b 2j \sin \beta d$$

Thus the coupling between ports 1 and 4 is

$$C = -20 \log 2|B_b \sin \beta d| \tag{6.32}$$

and is a maximum for $d = \lambda_g/4$. For this coupler the directivity D is theoretically infinite and independent of frequency, whereas the coupling C is quite frequency-sensitive, although not as frequency-sensitive as the directivity D given by (6.31b), since sin βd varies more slowly around $\pi/2$ than does cos βd. Actually, in practice, the directivity D is not infinite since,

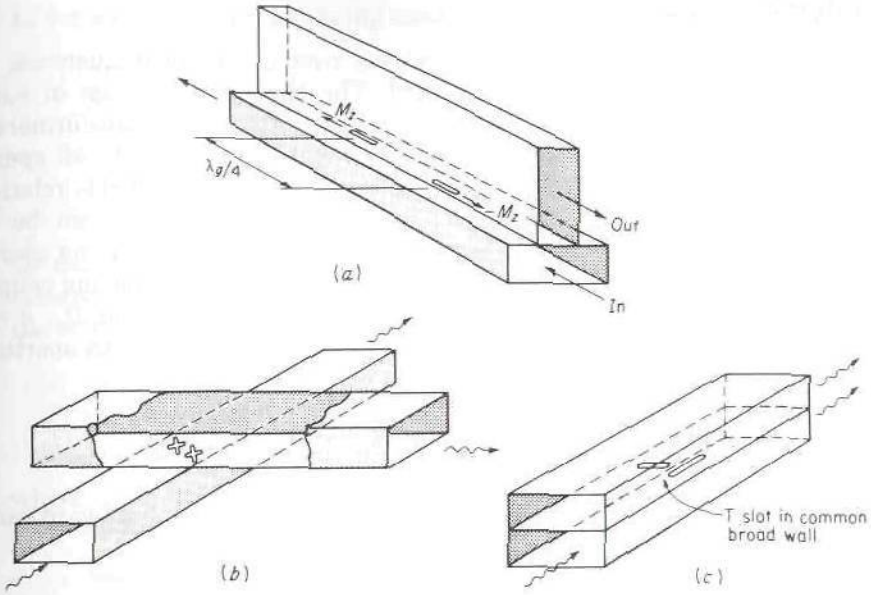


FIGURE 6.22
(*a*) Schwinger reversed-phase coupler; (*b*) Moreno crossed-guide coupler; (*c*) Riblet T-slot coupler.

in the foregoing discussion, it was assumed that the same incident field was present at each aperture, and each aperture radiated the same field into the upper guide. Because of interaction effects between the two apertures, the assumption of equal-amplitude fields coupled by both apertures is an approximation valid for a small amount of coupling only.

Figure 6.22*a* illustrates a typical reversed-phase coupler. The TE_{10} mode has a zero normal electric field and transverse magnetic field at the narrow wall, and hence the coupling to this mode in the upper guide is through the induced axial magnetic dipole moment of the aperture only. In the lower guide the axial magnetic field of the TE_{10} mode is of opposite sign on the two sides of the center, so that induced dipoles M_z and $-M_z$ are produced. These dipoles radiate symmetrically in both directions, but are phase-reversed to obtain the desired reversed-phase directional coupler.

The other double-aperture-coupled directional couplers in common use are the Moreno crossed-guide coupler and the Riblet T-slot coupler, illustrated in Fig. 6.22 also. Design nomograms for these couplers, as well as for the Schwinger reversed-phase coupler, are given in a paper by Anderson.†

†T. N. Anderson, Directional Coupler Design Nomograms, *Microwave J.*, vol. 2, pp. 34–38, May 1959.

Multielement Couplers

To achieve good directivity over a band of frequencies, couplers with many apertures may be used. The theory and design of such couplers parallel those given for multisection quarter-wave transformers in Chap. 5. Figure 6.23 illustrates an $N + 1$ element coupler with all aperture spacings equal to d. If we assume that the total power coupled is relatively small compared with the incident power, the incident wave can be considered to have essentially the same amplitude A at each coupling aperture apart from the additional phase change. Let the apertures having coupling coefficients C_n, $n = 0, 1, 2, \ldots, N$, in the forward direction, and D_n, $n = 0, 1, 2, 3, \ldots, N$, in the reverse direction. At the position of the Nth aperture the total forward wave in the upper guide is

$$B_f = Ae^{-j\beta Nd} \sum_{n=0}^{N} C_n \tag{6.33}$$

At the plane of the first aperture, the total backward wave has an amplitude

$$B_b = A \sum_{n=0}^{N} D_n e^{-j\beta 2nd} \tag{6.34}$$

The coupling and directivity are given by

$$C = -20 \log \left| \sum_{n=0}^{N} C_n \right| \tag{6.35a}$$

$$D = -20 \log \frac{|\sum_{n=0}^{N} D_n e^{-j\beta 2nd}|}{|\sum_{n=0}^{N} C_n|} = -C - 20 \log \left| \sum_{n=0}^{N} D_n e^{-j\beta 2nd} \right| \tag{6.35b}$$

In a multiaperture directional coupler the required coupling from each aperture is small so the aperture radii are then also small. The corresponding aperture susceptance $\overline{B}$ and reactance $\overline{X}$ are then also small. Thus the amplitudes of the waves coupled in the direction of ports 3 and 4 in Fig. 6.23, which are given by (4.140), can be approximated by

$$A_1 + A_3 = j\left(\frac{\overline{B}}{2} + \frac{\overline{X}}{2}\right)A$$

$$A_2 + A_4 = j\left(\frac{\overline{B}}{2} - \frac{\overline{X}}{2}\right)A$$

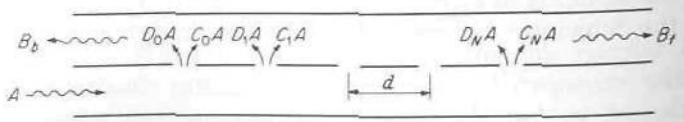


FIGURE 6.23
A multielement directional coupler.

In terms of the notation being used in this section, we can thus write

$$C_n = j\frac{\overline{B} + \overline{X}}{2} \tag{6.36a}$$

$$D_n = j\frac{\overline{B} - \overline{X}}{2} \tag{6.36b}$$

By using these expressions we obtain the advantage that for fixed aperture offsets both C_n and D_n are constants multiplied by the radius r_n cubed for the nth aperture. Hence we can express C_n and D_n in the form

$$C_n = T_f r_n^3 = T_f d_n \tag{6.37a}$$

$$D_n = T_b r_n^3 = T_b d_n \tag{6.37b}$$

where $d_n = r_n^3$. Both T_f and T_b depend on the frequency.

We can now express the directivity D in the form

$$D = -C - 20\log|T_b| - 20\log\left|\sum_{n=0}^{N} d_n e^{-j2\beta n d}\right| \tag{6.38a}$$

and express C as

$$C = -20\log|T_f| - 20\log\left|\sum_{n=0}^{N} d_n\right| \tag{6.38b}$$

In the expression for D the term $20\log|T_f/T_b|$ gives the intrinsic directivity of a single aperture. This directivity is usually small except possibly at the design frequency. Thus, if we are to achieve a broadband design, we must design the array factor

$$F = \left|\sum_{n=0}^{N} d_n e^{-j2\beta n d}\right| \tag{6.39}$$

to give a high directivity over the frequency band of interest.

In order to obtain an equal-ripple characteristic in the passband, the array factor F is made proportional to a Chebyshev polynomial. If we choose a symmetrical array, with $d_0 = d_N$, $d_1 = d_{N-1}$, etc., we obtain [Eq. (5.56)]

$$F = \left|\sum_{n=0}^{M} 2d_n \cos(N - 2n)\beta d\right| \tag{6.40}$$

where $M = (N - 1)/2$ for N odd and $N/2$ for N even. Note that for N odd there are an even number of apertures, since the first aperture has been labeled the zeroth aperture. In (6.40) the Mth term is d_M for N even and $2d_M \cos(N - 2M)\beta d$ for N odd. To obtain a Chebyshev-type response, we

now choose

$$F = \left| \sum_{n=0}^{M} 2d_n \cos(N - 2n)\theta \right| = K|T_N(\sec\theta_m \cos\theta)| \tag{6.41}$$

as in (5.68). In this equation $\theta = \beta d$ and $\sec\theta_m$ is the value of $\sec\beta d$ at the upper and lower edges of the passband. At the center of the passband, $\theta = \pi/2$, corresponding to a spacing $d = \lambda_g/4$. The constant K is chosen to give the desired value of coupling C in the center of the band, where $\theta = \pi/2$. Thus we obtain

$$\begin{aligned} C &= -20\log|T_f| \left| \sum_{n=0}^{N} d_n \right| \\ &= -20\log K|T_f|\,|T_N(\sec\theta_m)| \end{aligned} \tag{6.42}$$

since $|\sum_{n=0}^{N} d_n| = K|T_N(\sec\theta_m)|$ from (6.41).

If we use (6.38a), (6.41), and (6.42), the expression for directivity may be written as

$$\begin{aligned} D &= 20\left[\log K|T_f T_N(\sec\theta_m)| - \log|T_b| - \log K|T_N(\sec\theta_m\cos\theta)|\right] \\ &= 20\left[\log\left|\frac{T_f}{T_b}\right| + \log\left|\frac{T_N(\sec\theta_m)}{T_N(\sec\theta_m\cos\theta)}\right|\right] \end{aligned} \tag{6.43}$$

Since T_f/T_b is a function of frequency, D will not have a Chebyshev-type behavior. However, the departure from a true Chebyshev behavior will usually be small since T_f/T_b gives very little directivity, except perhaps near the center of the band. For a conservative design we choose the minimum value of D on the basis that T_f/T_b contributes negligible directivity. Certainly, for a broadband design, this will be the case at the edges of the passband. The minimum value D_m of directivity in the passband as contributed by the array factor F occurs when

$$T_N(\sec\theta_m\cos\theta) = 1$$

Hence let D_m be defined as

$$D_m = 20\log|T_N(\sec\theta_m)| \tag{6.44}$$

This equation shows that if we specify D_m, then $\sec\theta_m$ is fixed, which in turn fixes the bandwidth, and vice versa. Thus we must specify either D_m or $\sec\theta_m$, and the other is fixed. We may then solve (6.42) for the constant K in terms of the given value of the coupling C at the center of the band. From (6.41) the coefficients d_n are found.

We will outline the procedure by considering the design of a three-hole Chebyshev directional coupler utilizing offset circular apertures in the common broad wall between two rectangular waveguides as shown in Fig. 6.23. For the nth aperture the field coupled in the forward direction, with $A = 1$, is C_n where C_n is given by (6.36a). The field coupled in the reverse

direction is given by D_n in (6.36b). By using (4.139) we get

$$C_n = \left[-j\frac{2k_0^2}{3\beta ab}\sin^2\frac{\pi x_0}{a} + j\frac{4}{3ab}\left(\beta\sin^2\frac{\pi x_0}{a} + \frac{\pi^2}{\beta a^2}\cos^2\frac{\pi x_0}{a}\right)\right] r_n^3 \tag{6.45a}$$

$$D_n = \left[-j\frac{2k_0^2}{3\beta ab}\sin^2\frac{\pi x_0}{a} - j\frac{4}{3ab}\left(\beta\sin^2\frac{\pi x_0}{a} - \frac{\pi^2}{\beta a^2}\cos^2\frac{\pi x_0}{a}\right)\right] r_n^3 \tag{6.45b}$$

We have used x_0 for the aperture position so as not to confuse this with the aperture spacing d. The factors multiplying r_n^3 are T_f and T_b, respectively, and x_0 is the aperture offset measured from the waveguide side wall.

For a three-hole coupler,

$$T_N(\sec\theta_m\cos\theta) = T_2(\sec\theta_m\cos\theta) = 2\sec^2\theta_m\cos^2\theta - 1$$

From a specification of D_m, we can solve for θ_m using

$$D_m = 20\log|T_N(\sec\theta_m)|$$

which for $n = 2$ gives

$$\sec\theta_m = \frac{(10^{(D_m/20)} + 1)^{1/2}}{\sqrt{2}}$$

In general we can let $\sec\theta_m = \cosh\phi$ and use $T_N(\cosh\phi) = \cosh N\phi$; thus

$$\cosh N\phi = 10^{(D_m/20)} = y \tag{6.46a}$$

$$\phi = \frac{1}{N}\cosh^{-1}(10^{(D_m/20)}) = \frac{1}{N}\ln\left(y + \sqrt{y^2 - 1}\right) \tag{6.46b}$$

$$\theta_m = \cos^{-1}\left(\frac{1}{\cosh\phi}\right) \tag{6.46c}$$

There will be two solutions for θ_m, one less than $\pi/2$ and one greater than $\pi/2$. These two values determine the values of $\beta d = \theta_m$ at the lower and upper edges of the passband.

When θ_m has been found, then if we specify the desired coupling $C = C_0$ at the center of the band we must have, from (6.42) with $\theta = \pi/2$, $C_0 = -20\log K|T_f|\,|T_N(\sec\theta_m)|$ or

$$K = \frac{10^{(-C_0/20)}}{|T_f|\,|T_N(\sec\theta_m)|} = \frac{10^{(-C_0/20)}}{|T_f|10^{(D_m/20)}} \tag{6.47}$$

where $|T_f|$ is found from (6.45a) at the center of the band where $\beta d = \pi/2$. After we have found the constant K, we express $T_N(\sec\theta_m\cos\theta)$ as a

Fourier series. For $N = 2$ we have

$$T_2(\sec\theta_m \cos\theta) = \sec^2\theta_m \cos 2\theta + \sec^2\theta_m - 1$$

From (6.41) we can then determine the d_n. For our specific case we have

$$2d_0 \cos 2\beta d = 2d_0 \cos 2\theta = K \sec^2\theta_m \cos 2\theta$$

$$d_1 = K(\sec^2\theta_m - 1)$$

Thus $d_0 = d_2 = (K/2)\sec^2\theta_m$. Since $d_n = r_n^3$ we get

$$r_0 = r_2 = \left(\frac{K}{2}\sec^2\theta_m\right)^{1/3}$$

$$r_1 = \left(K\sec^2\theta_m - K\right)^{1/3}$$

This completes the design of the directional coupler. The coupling and directivity as a function of frequency can be evaluated using (6.42) and (6.43) and requires the evaluation of T_f, T_b, and $T_N(\sec\theta_m \cos\theta)$ at each frequency of interest. The following numerical example illustrates the performance that can be achieved.

Example 6.2 Three-Hole Chebyshev directional-coupler design. A three-hole directional coupler with a coupling of 20 dB and a minimum array factor directivity of 30 dB will be designed. The waveguide dimensions are $a = 0.9$ in, $b = 0.4$ in. The design frequency is 9 GHz. At 9 GHz the waveguide TE_{10} propagation constant is

$$\beta = \beta_0 = \left[k_0^2 - \left(\frac{\pi}{a}\right)^2\right]^{1/2} = 1.29 \text{ rad/cm}$$

Thus the aperture spacing d is $\pi/2\beta = 1.218$ cm. By using (6.46) we obtain

$$\phi = \tfrac{1}{2}\cosh^{-1}(10^{1.5}) = 2.07338$$

and $\theta_m = 1.3206$ and $\pi - 1.3206 = 1.821$.

The values of β at the lower and upper band edges are given by $(2\theta_m/\pi)\beta_0$ or 1.0845 and 1.4954. The corresponding values of k_0 are $(1.0845^2 + \pi^2/a^2)^{1/2} = 1.7506$ and $(1.4954^2 + \pi^2/a^2)^{1/2} = 2.03097$. From k_0 we obtain the frequencies at the band edges and these are 8.358 and 9.697 GHz. The fractional bandwidth is $\Delta f/f_0 = 0.149$.

The next step is the evaluation of K using (6.47). The parameter $|T_f|$ is the magnitude of the coefficient of r_n^3 in (6.45a) at the frequency 9 GHz. We will choose an aperture offset given by (6.28), that is, $x_0 = 0.46397$ cm. We then find that $K = 6.0244 \times 10^{-3}$. The aperture radii are now readily found and are $r_0 = r_2 = 0.3663$ cm and $r_1 = 0.4518$ cm.

The coupling and directivity at any frequency is given by (6.38). In Fig. 6.24 we show the overall performance of the directional coupler. The dashed curve shows the directivity contributed by the array factor F. The total directivity is greater because we chose the aperture offset such that $T_b = 0$ at 9 GHz; thus the aperture directivity contributes in a significant way to the

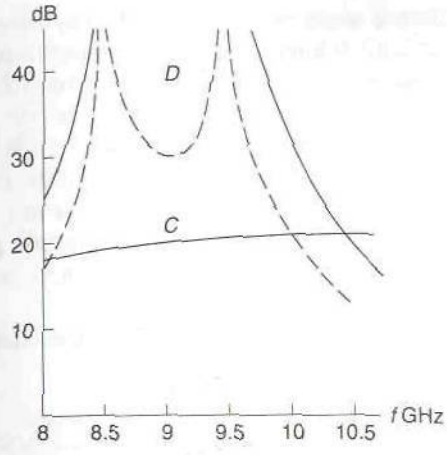


FIGURE 6.24
Performance of a three-hole Chebyshev directional coupler with $C = 20$ dB and $D_m = 30$ dB. The dashed curve shows the directivity contributed by the array factor and is never less than 30 dB in the passband between 8.358 and 9.697 GHz.

overall directivity. If we used a second set of apertures spaced a distance x_0 from the opposite side wall, the coupling would be increased by 6 dB without *changing the directivity. We could thus obtain a nominal coupling of* 14 *dB* over the frequency band.

Coupled-Line Directional Couplers

Aperture-type directional couplers are not suitable for microstrip or strip-line construction. For planar-transmission-line structures, coupled transmission lines are frequently used for building directional couplers. In Fig. 6.25 we illustrate a microstrip directional coupler that involves two coupled microstrip lines. In practice, the printed circuit board would be housed in a shielded box and coaxial-transmission-line connectors would be bonded to each microstrip line. The analysis of the coupled-line directional coupler is readily carried out by taking advantage of the fourfold symmetry of the

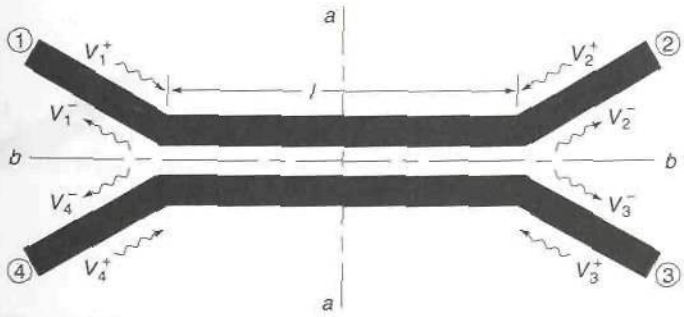


FIGURE 6.25
A microstrip coupled-line directional coupler.

structure. We can choose excitations so that the symmetry plane aa corresponds to an electric wall (short circuit) or a magnetic wall (open circuit) and also so that the symmetry plane bb corresponds to an electric wall or a magnetic wall. When bb corresponds to an electric wall, the mode of propagation on the coupled line is the odd mode which has a characteristic impedance Z_o and propagation constant β_o. When bb corresponds to a magnetic wall, the mode of propagation is the even mode which has a characteristic impedance Z_e and propagation constant β_e. The propagation constants are different because the effective dielectric constants for the two modes are different.

We will consider the following four different excitations:

(a) $$V_1^+ = V_2^+ = V_3^+ = V_4^+ = V^+$$

This case corresponds to both symmetry planes aa and bb being magnetic walls. For this case we only need to analyze the equivalent circuit of one-quarter of the structure as shown in Fig. 6.26a.

(b) $$V_1^+ = V_4^+ = V^+ \qquad V_2^+ = V_3^+ = -V^+$$

For this case the plane aa is an electric wall and the plane bb is a magnetic wall. The equivalent circuit is shown in Fig. 6.26b.

(c) $$V_1^+ = -V_4^+ = V^+ \qquad V_2^+ = -V_3^+ = V^+$$

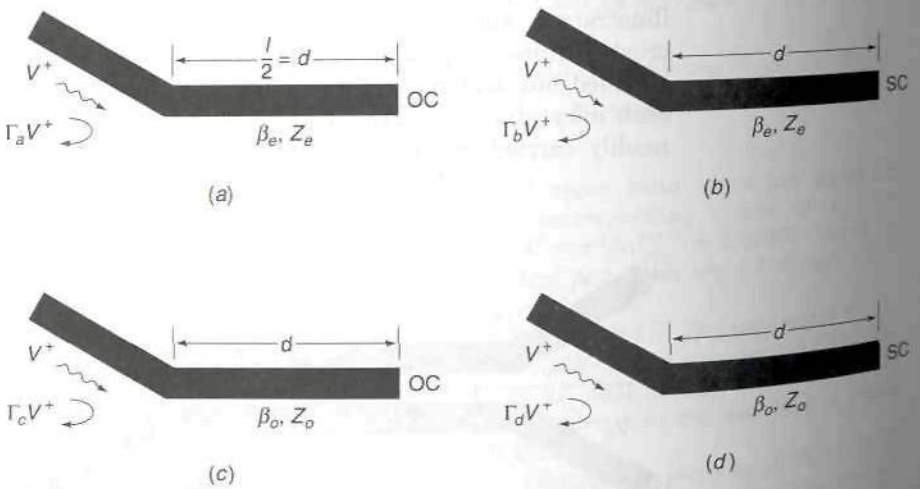


FIGURE 6.26
Equivalent circuit for one-quarter of the coupled-line directional coupler when (a) the planes aa and bb are magnetic walls, (b) aa is an electric wall and bb is a magnetic wall, (c) aa is a magnetic wall and bb is an electric wall, (d) aa and bb are both electric walls.

For this case the plane aa is a magnetic wall and the plane bb is an electric wall. The equivalent circuit for this case is shown in Fig. 6.26c.

(d) $$V_1^+ = -V_4^+ = V^+ \qquad V_2^+ = -V_3^+ = -V^+$$

For this case both symmetry planes correspond to electric walls and the equivalent circuit is that shown in Fig. 6.26d.

For case (a) it is readily seen from the equivalent circuit in Fig. 6.26c that

$$Z_{\text{in}} = -jZ_e \cot \beta_e d$$

$$\Gamma_a = \frac{Z_{\text{in}} - Z_c}{Z_{\text{in}} + Z_c} = \frac{-jZ_e \cot \beta_e d - Z_c}{-jZ_e \cot \beta_e d + Z_c}$$

where Z_c is the characteristic impedance of the input microstrip line. From symmetry considerations we have

$$V_1^- = \Gamma_a V^+ \qquad V_2^- = \Gamma_a V^+ \qquad V_3^- = \Gamma_a V^+ \qquad V_4^- = \Gamma_a V^+$$

for this case. For case (b) we have

$$Z_{\text{in}} = jZ_e \tan \beta_e d \qquad \Gamma_b = \frac{jZ_e \tan \beta_e d - Z_c}{jZ_e \tan \beta_e d + Z_c}$$

and

$$V_1^- = \Gamma_b V^+ \qquad V_2^- = -\Gamma_b V^+ \qquad V_3^- = -\Gamma_b V^+ \qquad V_4^- = \Gamma_b V^+$$

For case (c) we have

$$Z_{\text{in}} = -jZ_o \cot \beta_o d \qquad \Gamma_c = \frac{-jZ_o \cot \beta_o d - Z_c}{-jZ_o \cot \beta_o d + Z_c}$$

$$V_1^- = \Gamma_c V^+ \qquad V_2^- = \Gamma_c V^+ \qquad V_3^- = -\Gamma_c V^+ \qquad V_4^- = -\Gamma_c V^+$$

For the last case, namely case (d), we have

$$Z_{\text{in}} = jZ_o \tan \beta_o d \qquad \Gamma_d = \frac{jZ_o \tan \beta_o d - Z_c}{jZ_o \tan \beta_o d + Z_c}$$

$$V_1^- = \Gamma_d V^+ \qquad V_2^- = -\Gamma_d V^+ \qquad V_3^- = \Gamma_d V^+ \qquad V_4^- = -\Gamma_d V^+$$

We now superimpose these four solutions. The superposition of the four incident waves at each port gives $V_1^+ = 4V^+$, $V_2^+ = V_3^+ = V_4^+ = 0$. Thus only port 1 is excited. The superposition of the reflected waves gives

$$V_1^- = (\Gamma_a + \Gamma_b + \Gamma_c + \Gamma_d)V^+ = \tfrac{1}{4}(\Gamma_a + \Gamma_b + \Gamma_c + \Gamma_d)V_1^+ \qquad (6.48a$$

$$V_2^- = \tfrac{1}{4}(\Gamma_a - \Gamma_b + \Gamma_c - \Gamma_d)V_1^+ \qquad (6.48b$$

$$V_3^- = \tfrac{1}{4}(\Gamma_a - \Gamma_b - \Gamma_c + \Gamma_d)V_1^+ \qquad (6.48c$$

$$V_4^- = \tfrac{1}{4}(\Gamma_a + \Gamma_b - \Gamma_c - \Gamma_d)V_1^+ \qquad (6.48d$$

With some algebraic steps it is easy to show that

$$\Gamma_a + \Gamma_d = \frac{2(Z_c^2 t_e - Z_e Z_o t_o)}{jZ_c(Z_e - Z_o t_e t_o) - Z_c^2 t_e - Z_e Z_o t_o} \tag{6.49a}$$

where $t_o = \tan \beta_o d$ and $t_e = \tan \beta_e d$. We also readily find that

$$\Gamma_b + \Gamma_c = \frac{2(Z_c^2 t_o - Z_e Z_o t_e)}{jZ_c(Z_o - Z_e t_e t_o) - Z_c^2 t_o - Z_e Z_o t_e} \tag{6.49b}$$

For the ideal directional coupler, the two propagation constants should be equal, that is, $\beta_o = \beta_e = \beta$, in which case $t_e = t_o$. For a coupled-line coupler using strip-line construction, we have $\beta_e = \beta_o$. For a microstrip coupled-line coupler, we can approach this ideal situation by placing a dielectric sheet made from the substrate material on top of the conducting strips.† This surrounds the conductors with dielectric and will make the propagation constants equal to that for TEM waves in the solid dielectric. The use of a dielectric overlay will change the even- and odd-mode characteristic impedances but they will remain different. When $t_e = t_o$ we see from (6.49) that, provided we choose the line dimensions to make

$$Z_e Z_o = Z_c^2 \tag{6.50}$$

$\Gamma_a + \Gamma_d = 0$ and $\Gamma_b + \Gamma_c = 0$. An examination of (6.48) now shows that $V_1^- = 0$ and $V_3^- = 0$. Thus there is no reflection in port 1 and no power coupled into port 3 at any frequency. Thus the four-port junction is a directional coupler. When $t_e = t_o = t$ and we make $Z_e Z_o = Z_c^2$, then we also have $\Gamma_a + \Gamma_b - \Gamma_c - \Gamma_d = 2(\Gamma_a + \Gamma_b)$. The latter is given by

$$(\Gamma_a + \Gamma_b) = \frac{2(Z_c^2 - Z_e^2)t}{jZ_c Z_e(1 - t^2) - (Z_e^2 + Z_e^2)t}$$

This expression can be reduced to the following form:

$$\Gamma_a + \Gamma_b = \frac{-2(Z_e - Z_o)\sin 2\beta d}{j2Z_c \cos 2\beta d - (Z_e + Z_o)\sin 2\beta d}$$

The transmission coefficient into port 4 is $\frac{1}{2}(\Gamma_a + \Gamma_b)$ and is the scattering-matrix parameter $S_{41} = S_{14}$; thus

$$S_{41} = \frac{j\dfrac{Z_e - Z_o}{Z_o + Z_e}\sin 2\beta d}{\dfrac{2Z_c}{Z_e + Z_o}\cos 2\beta d + j \sin 2\beta d} \tag{6.51}$$

†M. Horno and F. Medina, Multilayer Planar Structures for High-Directivity Directional Coupler Design, *IEEE Trans.*, vol. MTT-34, pp. 1442–1449, 1986.

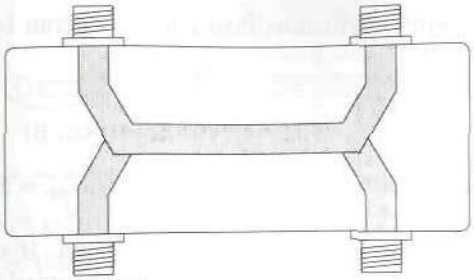


FIGURE 6.27
A strip-line coupled-line directional coupler

The parameter $(Z_e - Z_o)/(Z_e + Z_o)$ is the voltage-coupling parameter c. For this coupler the coupling is given by (note that $2d = l$)

$$C = 20 \log \frac{\left[(1 - c^2)\cos^2 \beta l + \sin^2 \beta l\right]^{1/2}}{c \sin \beta l}$$

$$= 20 \log \frac{\left[1 - c^2 \cos^2 \beta l\right]^{1/2}}{c \sin \beta l} \qquad (6.52$$

The maximum coupling occurs when $2\beta d = \beta l = \pi/2$ which corresponds to a coupled-line one-quarter wavelength long. The maximum coupling is $20 \log(1/c)$. In a microstrip coupled line it is not practical to obtain a ratio of Z_e/Z_o greater than 2; so the maximum coupling is limited to a value of 9.5 dB or perhaps up to 8 dB with careful design. By using broadside coupled strips in a strip-line configuration as shown in Fig. 6.27, a coupling of 3 dB is readily achieved. Figure 6.28 shows a plot of $|S_{41}|$ as a function of $2\beta d = \beta l$ for a coupled-line directional coupler.

The directivity of this type of coupler is given by V_4^-/V_3^- and is infinite since V_3^- is zero at all frequencies.

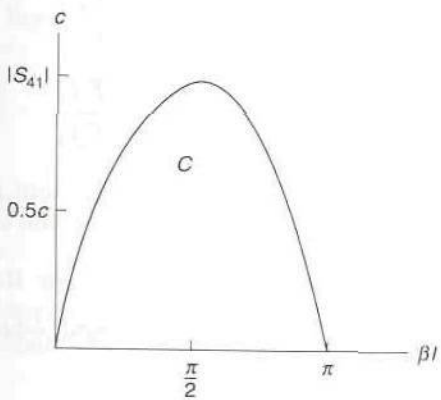


FIGURE 6.28
The variation of $|S_{41}|$ as a function of βl for coupled-line directional coupler.

The transmission coefficient into port 2 is given by

$$S_{21} = \frac{\sqrt{1 - c^2}}{\sqrt{1 - c^2}\cos \beta l + j \sin \beta l} \tag{6.53}$$

The symmetry of the structure requires $S_{11} = S_{22} = S_{33} = S_{44} = 0$, $S_{12} = S_{21} = S_{34} = S_{43}$, $S_{14} = S_{41} = S_{23} = S_{32}$, and $S_{13} = S_{24} = S_{31} = S_{42} = 0$. Thus all scattering-matrix parameters are known. If the propagation constants are not equal we can still use the coupled-line structure as a directional coupler, but it will not have infinite directivity. The coupler performance can be improved for the case $\beta_e \neq \beta_o$ by adding a small shunt capacitor between the two coupled lines at the input and output.† The bandwidth can be increased by using several sections in cascade.

Branch-Line Directional Coupler

The branch-line directional coupler shown in Fig. 6.29 is readily fabricated using microstrip construction and can be designed for 3-dB coupling without any difficulty. The analysis of this coupler is also readily carried out using the fourfold symmetry that is inherent in the structure. With proper excitation the symmetry planes *aa* and *bb* can be made to correspond to either electric or magnetic walls. If excitations are chosen the same as for cases (*a*) to (*d*) in the coupled-line coupler, then the equivalent circuits for the four excitations are those shown in Figs. 6.30*a* to *d*. From these equivalent circuits we readily obtain

$$\Gamma_a = \frac{Y_c - jY_1t_1 - jY_2t_2}{Y_c + jY_1t_1 + jY_2t_2} \tag{6.54a}$$

$$\Gamma_b = \frac{Y_ct_1 + jY_1 - jY_2t_1t_2}{Y_ct_1 - jY_1 + jY_2t_1t_2} \tag{6.54b}$$

$$\Gamma_c = \frac{Y_ct_2 - jY_1t_1t_2 + jY_2}{Y_ct_2 + jY_1t_1t_2 - jY_2} \tag{6.54c}$$

$$\Gamma_d = \frac{Y_ct_1t_2 + jY_1t_2 + jY_2t_1}{Y_ct_1t_2 - jY_1t_2 - jY_2t_1} \tag{6.54d}$$

where $t_1 = \tan \theta_1 = \tan \beta_1 d_1$ and $t_2 = \tan \theta_2 = \text{tant } \beta_2 d_2$ and Y_c, Y_1, Y_2 are the characteristic admittance of the input line, the through line, and the

†I. Bahl and P. Bhartia, "Microwave Solid State Circuit Design," John Wiley & Sons, Inc., New York, 1988.

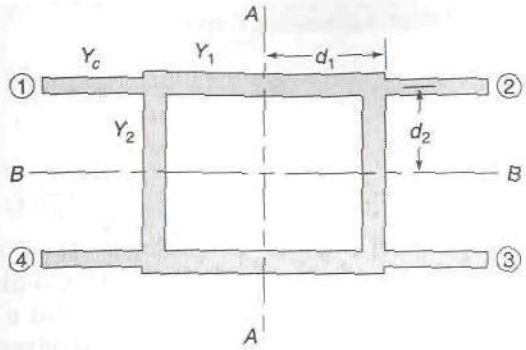


FIGURE 6.29
A branch-line directional coupler.

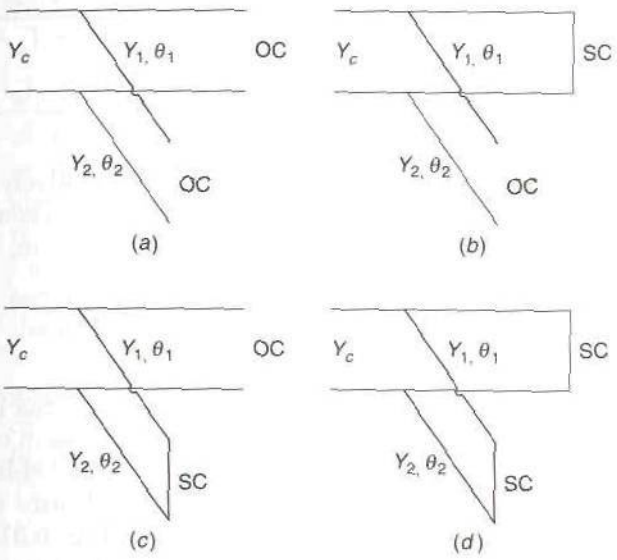


FIGURE 6.30
The equivalent circuit for one quarter section when (*a*) *aa* and *bb* are magnetic walls, (*b*) *aa* is an electric wall and *bb* is a magnetic wall, (c) *aa* is a magnetic wall and *bb* is an electric wall, (*d*) *aa* and *bb* are both electric walls.

branch line as shown in Fig.6.29. The relations (6.48) apply to the branch-line coupler also; so we have

$$S_{11} = \tfrac{1}{4}(\Gamma_a + \Gamma_b + \Gamma_c + \Gamma_d) \tag{6.55a}$$

$$S_{12} = S_{21} = \tfrac{1}{4}(\Gamma_a - \Gamma_b + \Gamma_c - \Gamma_d) \tag{6.55b}$$

$$S_{13} = S_{31} = \tfrac{1}{4}(\Gamma_a - \Gamma_b - \Gamma_c + \Gamma_d) \tag{6.55c}$$

$$S_{14} = S_{41} = \tfrac{1}{4}(\Gamma_a + \Gamma_b - \Gamma_c - \Gamma_d) \tag{6.55d}$$

If we choose $t_1 = t_2 = 1$ so that the through lines and branch lines are one-quarter wavelength lines and also choose $Y_1^2 - Y_2^2 = Y_c^2$, then we find that $S_{11} = S_{22} = S_{33} = S_{44} = 0$, $S_{14} = 0$, $S_{31} = \frac{1}{2}(\Gamma_a - \Gamma_c)$ and $S_{21} =$

$\frac{1}{2}(\Gamma_a + \Gamma_c)$. These latter expressions give

$$S_{31} = -\frac{Y_2}{Y_1} \tag{6.56a}$$

$$S_{21} = -j\frac{Y_c}{Y_1} \tag{6.56b}$$

for $t_1 = t_2 = 1$. A 3-dB coupler is obtained if we choose $Y_2 = Y_c$ and $Y_1 = \sqrt{2}\,Y_c$, a condition that is easily satisfied. A 3-dB directional coupler with the two outputs 90° out of phase is also called a 90° hybrid junction. The coupling and directivity at any frequency are given by

$$C = 20 \log \frac{1}{|S_{31}|} = 20 \log \frac{4}{|\Gamma_a + \Gamma_d - \Gamma_b - \Gamma_c|} \tag{6.57a}$$

$$D = 20 \log \left|\frac{S_{31}}{S_{41}}\right| = 20 \log \left|\frac{\Gamma_a + \Gamma_d - \Gamma_b - \Gamma_c}{\Gamma_a + \Gamma_b - \Gamma_c - \Gamma_d}\right| \tag{6.57b}$$

The branch-line directional coupler is a relatively narrowband device. However, by cascading several sections a broadband coupler can be obtained by appropriate choices for Y_1 and Y_2 for each section. Design formulas are available in the literature.†

Lange Directional Coupler‡

The last directional coupler that we will discuss is the Lange coupler shown in Fig. 6.31 This coupler uses several coupled lines in order to obtain larger coupling than what is possible in the simpler coupled-line coupler discussed earlier. The design of the coupler is such that wire connections between some of the lines are needed as shown in Fig. 6.31. This is the major shortcoming of the Lange coupler since such wire connections are not readily made in an MMIC circuit. The outstanding features of the Lange coupler are its compact size and very broadband characteristics. The Lange coupler is often used as an input coupler in balanced microwave amplifier circuits. For this application it is designed as a 3-dB coupler and the output signals are in phase quadrature, so that it is a 90° hybrid junction. Useful design formulas for a Lange coupler have been developed by Presser.§

†G. L. Matthaei, L. Young, and E. M. T. Jones, "Microwave Filters, Impedance-Matching Networks and Coupling Structures," Artech House Books, Dedham, Mass., 1980.

‡J. Lange, Interdigitated Stripline Quadrature Hybrid, *IEEE Trans.*, vol. MTT-17, pp. 1150–1151, December, 1969.

§A. Presser, Interdigitated Microstrip Coupler Design, *IEEE Trans.*, vol. MTT-26, pp. 801–805, October, 1978.

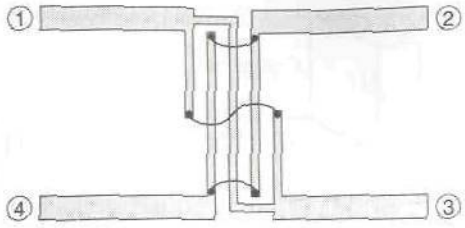


FIGURE 6.31
The Lange directional coupler.

6.5 HYBRID JUNCTIONS

Magic T

A waveguide hybrid junction, known as a magic T, is illustrated in Fig. 6.32. When a TE_{10} mode is incident in port 1, the electric field within the junction is like that sketched in Fig. 6.32*b*. This electric field has even symmetry about the midplane and hence cannot excite the TE_{10} mode in arm 4 since this mode must have an electric field with odd symmetry (shown dashed in Fig. 6.32*b*). Thus there is no coupling between ports 1 and 4. The coupling between ports 1 and 2, and 1 and 3, is clearly the same, as may be seen from the symmetry involved.

For a TE_{10} mode incident in arm 4, the electric field within the junction is sketched in Fig. 6.32*c*. Symmetry again shows that there is no coupling into port 1 (this is required by reciprocity as well). The coupling from port 4 into ports 2 and 3 is equal in magnitude but 180° out of phase. The scattering matrix of this hybrid T thus has the form

$$[S] = \begin{bmatrix} S_{11} & S_{12} & S_{12} & 0 \\ S_{12} & S_{22} & S_{23} & S_{24} \\ S_{12} & S_{23} & S_{33} & -S_{24} \\ 0 & S_{24} & -S_{24} & S_{44} \end{bmatrix}$$

since $S_{12} = S_{13}$, $S_{24} = -S_{34}$, from symmetry.

Matching elements that do not destroy the symmetry of the junction may be placed in the E-plane and H-plane arms so as to make $S_{11} = S_{44} = 0$. For a lossless structure we may then show that the unitary properties of the scattering matrix require that $S_{22} = S_{33} = 0$, so that all ports are matched. In addition, $S_{23} = 0$; so ports 2 and 3 as well as ports 1 and 4 are uncoupled. The hybrid T now becomes a directional coupler with 3-dB coupling, and is often called a *magic T*, even though there is nothing magic about its

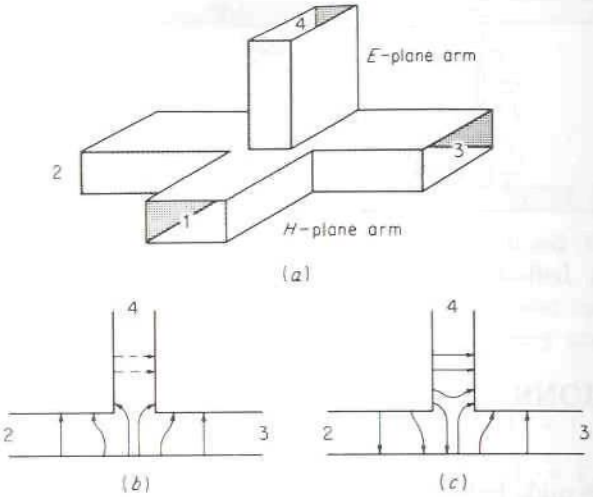


FIGURE 6.32
(*a*) Hybrid-T junction; (*b*) electric field pattern for wave incident in port 1; (*c*) electric field pattern for wave incident in port 4.

operation. The magic T is commonly used in waveguide balanced mixers and in bridge networks.

With $S_{11} = S_{44} = 0$, the scattering matrix becomes

$$[S] = \begin{bmatrix} 0 & S_{12} & S_{12} & 0 \\ S_{12} & S_{22} & S_{23} & S_{24} \\ S_{12} & S_{23} & S_{33} & -S_{24} \\ 0 & S_{24} & -S_{24} & 0 \end{bmatrix}$$

The product of the second row with its conjugate gives

$$|S_{12}|^2 + |S_{22}|^2 + |S_{23}|^2 + |S_{24}|^2 = 1 \tag{6.58a}$$

and the similar expression for row 3 is

$$|S_{12}|^2 + |S_{23}|^2 + |S_{33}|^2 + |S_{24}|^2 = 1 \tag{6.58b}$$

If we subtract these two equations, we obtain

$$|S_{22}|^2 - |S_{33}|^2 = 0 \tag{6.58c}$$

so $|S_{22}| = |S_{33}|$. From rows 1 and 4 we have

$$2|S_{12}|^2 = 1 \quad \text{or} \quad |S_{12}| = \frac{\sqrt{2}}{2}$$

$$2|S_{24}|^2 = 1 \quad \text{or} \quad |S_{24}| = \frac{\sqrt{2}}{2}$$

and thus

$$|S_{12}| = |S_{24}| = \frac{\sqrt{2}}{2} \tag{6.59}$$

Use of this relation in (6.58*a*) gives

$$1 + |S_{22}|^2 + |S_{23}|^2 = 1$$

or $|S_{22}|^2 + |S_{23}|^2 = 0$. This sum can equal zero only if both S_{22} and S_{23} vanish. From the relation (6.58) it follows that S_{33} equals zero also.

The reduced form of the scattering matrix becomes

$$[S] = \begin{bmatrix} 0 & S_{12} & S_{12} & 0 \\ S_{12} & 0 & 0 & S_{24} \\ S_{12} & 0 & 0 & -S_{24} \\ 0 & S_{24} & -S_{24} & 0 \end{bmatrix}$$

By proper choice of terminal planes in arms 1 and 4, we can make both S_{12} and S_{24} real. Thus the scattering matrix of a magic T can be exhibited in the form

$$[S] = \frac{\sqrt{2}}{2} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \tag{6.60}$$

upon using the relations (6.59).

Hybrid Ring

The branch-line coupler designed for 3-dB coupling is a 90° hybrid junction. The magic T is a 0° or 180° hybrid junction since the two outputs are in phase if port 1 is the input port and are 180° out of phase if port 3 is the input port. A 180° hybrid junction that is readily made using microstrip construction is shown in Fig. 6.33. To understand its operation, consider a wave incident in port 1. This wave splits equally into two waves traveling around the ring circuit in opposite directions. The two waves will arrive in phase at ports 2 and 4 and out of phase at port 3. Thus ports 1 and 3 are uncoupled. Similarly, ports 2 and 4 are uncoupled since the two paths coupling these ports differ in length by $\lambda/2$.

A quantitative analysis of the hybrid ring is readily carried out. Each input line has a characteristic impedance Z_c and the ring has a characteristic impedance Z_1. We will let θ_1 be the electrical length of the ring between ports 1 and 2, 2 and 3 and 3 and 4, while $\theta_2 = 3\theta_1$ represents the electrical length of the ring between ports 1 and 4. We can choose excitations such that the symmetry plane aa is either an electric wall or a magnetic wall. This will allow us to characterize one-half of the structure in terms of

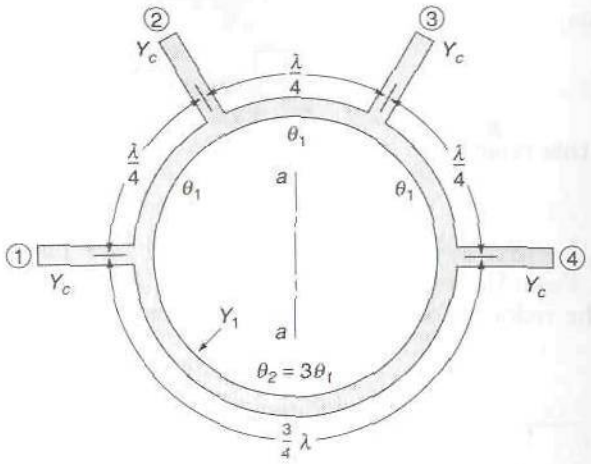


FIGURE 6.33
The hybrid ring ("rat-race").

scattering matrices for a two-port junction, ports 1 and 2. From the symmetry properties the complete four-port scattering matrix can be constructed.

Let ports 1 and 2 have incident waves V_a^+ and V_b^+. Also let ports 4 and 3 have incident waves V_a^+ and V_b^+, respectively. For this symmetrical excitation the electric field must be a maximum on the symmetry plane aa and the magnetic field must be zero. Thus aa can be replaced by a magnetic wall, i.e., the ring is open circuited on the plane aa. The equivalent circuit of half of the structure is shown in Fig. 6.34a. It consists of two input lines with characteristic impedance Z_c and an interconnecting line of electrical length θ_1 and characteristic impedance Z_1. The input port 1 is shunted by a susceptance $jB_1 = jY_1 \tan 3\theta_1/2$ due to the open-circuited section of the

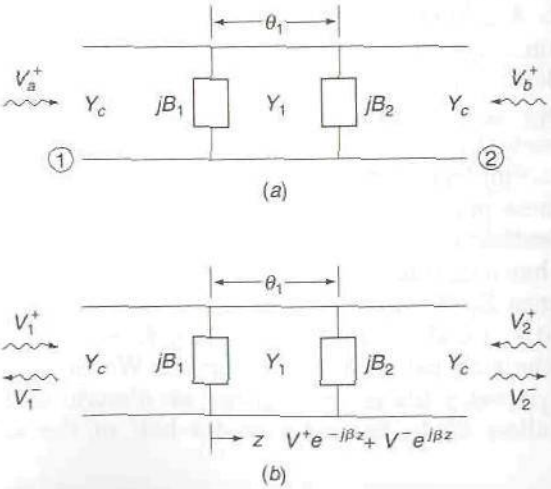


FIGURE 6.34
(a) Equivalent circuit for one-half of hybrid ring for open-circuit conditions on the symmetry plane; (b) equivalent circuit used to find the two-port scattering-matrix parameters for one-half of the hybrid-ring structure.

hybrid ring having an electrical length $3\theta_1/2$. The output port 2 is shunted by a susceptance $jB_2 = jY_1 \tan \theta_1/2$ due to the open-circuited section of the ring of electrical length $\theta_1/2$. We will let the two-port scattering-matrix parameters under open-circuit conditions on the plane aa be S_{ij}^{oc}. Thus we can write

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} S_{11}^{oc} & S_{12}^{oc} \\ S_{21}^{oc} & S_{22}^{oc} \end{bmatrix} \begin{bmatrix} V_a^+ \\ V_b^+ \end{bmatrix} \tag{6.61a}$$

and from symmetry considerations

$$\begin{bmatrix} V_4^- \\ V_3^- \end{bmatrix} = \begin{bmatrix} S_{11}^{oc} & S_{12}^{oc} \\ S_{21}^{oc} & S_{22}^{oc} \end{bmatrix} \begin{bmatrix} V_a^+ \\ V_b^+ \end{bmatrix} \tag{6.61b}$$

since port 4 is similar to port 1 and port 3 is similar to port 2.

Let us now change the incident waves at ports 4 and 3 to $-V_a^+$ and $-V_b^+$. We now have an antisymmetrical excitation and the symmetry plane is an electric wall or short circuit. For this case the equivalent circuit is the same as that in Fig. 6.34*a*, except that the shunting susceptances are due to short-circuited transmission-line sections, so jB_1 is replaced by $jB_3 = -jY_1 \cot 3\theta_1/2$ and jB_2 is replaced by $jB_4 = -jY_1 \cot \theta_1/2$. For this case we will designate the two-port scattering-matrix parameters under short-circuit conditions on the plane aa by S_{ij}^{sc}. For this odd excitation we can write

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \begin{bmatrix} S_{11}^{sc} & S_{12}^{sc} \\ S_{21}^{sc} & S_{22}^{sc} \end{bmatrix} \begin{bmatrix} V_a^+ \\ V_b^+ \end{bmatrix} \tag{6.62a}$$

and

$$\begin{bmatrix} V_4^- \\ V_3^- \end{bmatrix} = \begin{bmatrix} S_{11}^{sc} & S_{12}^{sc} \\ S_{21}^{sc} & S_{22}^{sc} \end{bmatrix} \begin{bmatrix} -V_a^+ \\ -V_b^+ \end{bmatrix} \tag{6.62b}$$

Let us now superimpose the two solutions which then gives $V_1^+ = 2V_a^+$, $V_2^+ = 2V_b^+$, $V_3^+ = V_4^+ = 0$. The superposition of the scattered waves gives

$$\begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix} = \frac{1}{2} \begin{bmatrix} S_{11}^{oc} + S_{11}^{sc} & S_{12}^{oc} + S_{12}^{sc} \\ S_{21}^{oc} + S_{21}^{sc} & S_{22}^{oc} + S_{22}^{sc} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} \tag{6.63a}$$

$$\begin{bmatrix} V_4^- \\ V_3^- \end{bmatrix} = \frac{1}{2} \begin{bmatrix} S_{11}^{oc} - S_{11}^{sc} & S_{12}^{oc} - S_{12}^{sc} \\ S_{21}^{oc} - S_{21}^{sc} & S_{22}^{oc} - S_{22}^{sc} \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \end{bmatrix} \tag{6.63b}$$

From these equations we can identify the following eight four-port scattering-matrix elements:

$$S_{11} = \tfrac{1}{2}(S_{11}^{oc} + S_{11}^{sc}) \qquad S_{12} = \tfrac{1}{2}(S_{12}^{oc} + S_{12}^{sc})$$
$$S_{21} = \tfrac{1}{2}(S_{21}^{oc} + S_{21}^{sc}) \qquad S_{22} = \tfrac{1}{2}(S_{22}^{oc} + S_{22}^{sc})$$
$$S_{31} = \tfrac{1}{2}(S_{21}^{oc} - S_{21}^{sc}) \qquad S_{32} = \tfrac{1}{2}(S_{22}^{oc} - S_{22}^{sc})$$
$$S_{41} = \tfrac{1}{2}(S_{11}^{oc} - S_{11}^{sc}) \qquad S_{42} = \tfrac{1}{2}(S_{12}^{oc} - S_{12}^{sc})$$

From symmetry considerations $S_{44} = S_{11}$, $S_{33} = S_{22}$, $S_{34} = S_{21}$. The re-

maining five elements follow from reciprocity, namely, $S_{ij} = S_{ji}$. Thus the four-port hybrid can be characterized in terms of two sets of two-port scattering-matrix parameters.

In order to obtain the two-port scattering-matrix parameters, consider the circuit shown in Fig. 6.34*b*. We assume incident and scattered waves to exist at ports 1 and 2. On the interconnecting line we assume the existence of a forward and backward propagating voltage wave given by

$$V^+ e^{-j\beta z} + V^- e^{j\beta z}$$

with associated current waves

$$V^+ Y_1 e^{-j\beta z} - V^- Y_1 e^{j\beta z}$$

The continuity of the voltage at ports 1 and 2 gives

$$V_1^+ + V_1^- = V^+ + V^- \tag{6.64a}$$

$$V_2^+ + V_2^- = V^+ e^{-j\theta_1} + V^- e^{j\theta_1} \tag{6.64b}$$

The continuity of the current at each port gives

$$(V_1^+ - V_1^-)Y_c = (V^+ - V^-)Y_1 + (V_1^+ + V_1^-)jB_1 \tag{6.64c}$$

$$(V^+ e^{-j\theta_1} - V^- e^{j\theta_1})Y_1 = -(V_2^+ - V_2^-)Y_c + (V_2^+ + V_2^-)jB_2 \tag{6.64d}$$

We can solve the first two equations for V^+ and V^-. We then substitute these solutions into the last two equations and solve for V_1^- and V_2^- in terms of V_1^+ and V_2^+. This will give us the two-port parameters S_{ij}^{oc}. After carrying out the algebra, we obtain

$$S_{11}^{oc} = \frac{1}{\Delta}\left[Y_c^2 - Y_1^2 + B_1 B_2 - Y_1(B_1 + B_2)\cot\theta_1 + jY_c(B_2 - B_1)\right] \tag{6.65a}$$

$$S_{12}^{oc} = S_{21}^{oc} = -\frac{2jY_c Y_1 \csc\theta_1}{\Delta} \tag{6.65b}$$

$$S_{22}^{oc} = S_{11}^{oc} + \frac{2jY_c(B_1 - B_2)}{\Delta} \tag{6.65c}$$

where

$$\Delta = Y_c^2 - B_1 B_2 + Y_1(B_1 + B_2)\cot\theta_1 + Y_1^2 + jY_c(B_1 + B_2 - 2Y_1\cot\theta_1)$$

$$B_1 = Y_1 \tan\frac{3\theta_1}{2} \qquad B_2 = Y_1 \tan\frac{\theta_1}{2}$$

The short-circuit two-port scattering-matrix parameters are obtained by replacing B_1 by $B_3 = -Y_1 \cot(3\theta_1/2)$ and B_2 by $B_4 = -Y_1 \cot(\theta_1/2)$.

At the center of the frequency band of interest, we choose $\theta_1 = \pi/2$ for which $B_1 = -Y_1$, $B_2 = Y_1$, $B_3 = Y_1$, and $B_4 = -Y_1$. For these conditions the expressions for the scattering-matrix parameters simplify and we

readily find that

$$S_{11}^{oc} = \frac{Y_c^2 - 2Y_1^2 + 2jY_cY_1}{Y_c^2 + 2Y_1^2} = S_{22}^{sc}$$

$$S_{22}^{oc} = \frac{Y_c^2 - 2Y_1^2 - 2jY_cY_1}{Y_c^2 + 2Y_1^2} = S_{11}^{sc}$$

$$S_{12}^{oc} = S_{12}^{sc} = \frac{-2jY_cY_1}{Y_c^2 + 2Y_1^2}$$

From these expressions we get

$$S_{11} = S_{22} = S_{33} = S_{44} = \frac{Y_c^2 - 2Y_1^2}{Y_c^2 + 2Y_1^2} \tag{6.66a}$$

$$S_{31} = S_{42} = 0 \tag{6.66b}$$

$$S_{12} = S_{34} = -S_{41} = S_{32} = \frac{-2jY_cY_1}{Y_c^2 + 2Y_1^2} \tag{6.66c}$$

We see that port 3 is uncoupled to port 1 and ports 2 and 4 are uncoupled. Also ports 2 and 4 are coupled to port 1 but the port 2 output is 180° out of phase with the port 4 output since $S_{21} = -S_{41}$. All ports will be matched if we choose $Y_c^2 = 2Y_1^2$ or

$$Z_1 = \sqrt{2}\,Z_c$$

When this latter condition is imposed, then $|S_{21}| = |S_{41}| = \sqrt{2}/2$ and we obtain a 3-dB directional coupler or 180° hybrid junction. The four-port scattering matrix reduces to

$$[S] = \frac{-j}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{bmatrix} \tag{6.67}$$

which is the same as that for the magic-T hybrid junction, apart from a different numbering of the ports and the choice of input terminal reference planes.

At frequencies away from the frequency at which $\theta_1 = \pi/2$, the performance of the hybrid ring can be found by evaluating the various scattering-matrix parameters of interest as a function of the normalized frequency variable $2\theta_1/\pi$. In Fig. 6.35 we show a plot of the couplings $|S_{12}|$, $|S_{14}|$, and $|S_{23}|$, the isolation $|S_{13}|$, and the input reflection coefficients $|S_{11}|$ and $|S_{22}|$ as a function of $2\theta_1/\pi$.

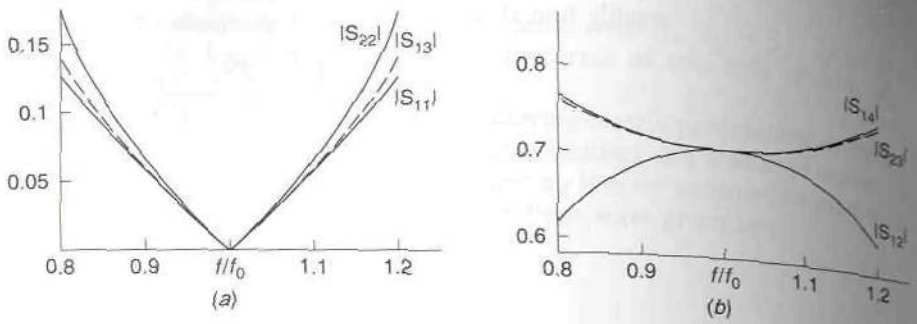


FIGURE 6.35
Hybrid-ring performance as a function of normalized frequency $f/f_0 = 2\theta_1/\pi$. (*a*) Magnitude of reflection coefficient and isolated port coupling coefficient; (*b*) magnitude of coupling coefficients.

6.6 POWER DIVIDERS

Power dividers are used to divide the input power into a number of smaller amounts of power for exciting the radiating elements in an array antenna. They are also used in balanced power amplifiers both as power dividers and power combiners.

A fundamental property of a lossless reciprocal three-port junction is that not all three ports can be simultaneously matched. If we assume that all three ports can be matched, then $S_{11} = S_{22} = S_{33}$ and the scattering matrix has the form

$$[S] = \begin{bmatrix} 0 & S_{12} & S_{13} \\ S_{12} & 0 & S_{23} \\ S_{13} & S_{23} & 0 \end{bmatrix}$$

For a lossless junction the scattering matrix is a unitary matrix. This means that the sum of the products of the elements in any row with the complex conjugate of the elements in another row is zero. For the junction under discussion this would require $S_{12}S_{23}^* = S_{13}S_{23}^* = S_{12}S_{13}^* = 0$. These equations will hold only if two of the transmission coefficients S_{12}, S_{23}, and S_{13} equals zero, in which case we do not have a functioning three-port junction. If we want to use a lossless three-port junction to split or divide the input power P_1 into fractions $\alpha P_1 = P_2$ and $(1 - \alpha)P_1 = P_3$ at ports 2 and 3, this is readily accomplished. For the three-port junction shown in Fig. 6.36, we can choose Z_2 and Z_3 so that the input port 1 is matched and the desired power split is obtained. If the input is matched, then $P_1 = \frac{1}{2}Y_1|V_1^+|^2$. Since $V_2^- = V_3^- = V_1^+$ because of the parallel connection of all three lines, we have

$$P_1 = \tfrac{1}{2}Y_2|V_1^+|^2 + \tfrac{1}{2}Y_3|V_1^+|^2$$

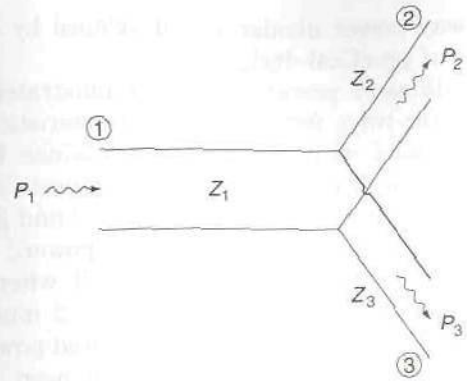


FIGURE 6.36
A lossless three-port junction used as a power divider.

For an impedance match we require $Y_1 = Y_2 + Y_3$ and, in order to obtain the desired power division, we require

$$\frac{Y_2}{Y_3} = \frac{\alpha}{1 - \alpha}$$

For example, if we want to split the input power so that $P_2 = P_1/3$ and $P_3 = 2P_1/3$, then $Y_2/Y_3 = \frac{1}{2}$ or $Y_3 = 2Y_2$. Consequently, in order to make port 1 matched, $Y_1 = Y_2 + Y_3 = 3Y_2$; so $Z_2 = 3Z_1$ and $Z_3 = 1.5Z_1$. This type of lossless power divider will not have matched output ports and since S_{23} will not be zero it also does not have isolation between the output ports. If there is a shunt susceptance at the junction, such as would occur from excitation of evanescent modes in a waveguide T or Y junction, the input port can still be matched by placing a suitable shunt-compensating susceptance at an appropriate position in the input line. It is desirable to have $S_{23} = 0$, so that reflected power at port 2 does not couple into port 3, and vice versa.

Wilkinson developed an N-way power divider that would split the input power into output power at N ports and that would also provide isolation between the output powers.† A unique feature of the Wilkinson power divider is the use of resistors connected between the various output ports. When the output ports are terminated in the correct load impedance, there is no current in the resistors; so they do not absorb any power. If one port is matched, then the reflected power from that port is partly absorbed by the resistor network and partly returned to the input, but no power is coupled into the other output ports as long as they remain properly terminated. Many different versions, including broadband designs, of the Wilkinson power divider have been developed. The multisection broadband

†E. Wilkinson, An N-Way Hybrid Power Divider, *IEEE Trans.*, vol. MTT-8, pp. 116–118 1960.

design for a two-way power divider was developed by Cohn and has been used for a number of practical designs.†

The basic Wilkinson power divider is illustrated in Fig. 6.37*a*. It consists of two quarter-wave sections with characteristic impedances Z_2 and Z_3 connected in parallel with the input line, which has a characteristic impedance Z_c. A resistor R is connected between ports 2 and 3. Let Z_{L2} and Z_{L3} be the matched terminating loads for ports 2 and 3, respectively. If we want to split the input power P_1 into output powers P_2 and P_3 so that $P_3 = K^2 P_2$ and also maintain zero current in R when ports 2 and 3 are terminated, then the output voltage V_2^- at port 2 must equal the output voltage V_3^- at port 3. In order to obtain the desired power ratio, we require $K^2|V_2^-|^2/Z_{L2} = |V_3^-|^2/Z_{L3}$ for $V_2^- = V_3^-$. Hence we need

$$K^2 Z_{L3} = Z_{L2} \tag{6.68}$$

For matched output terminations the resistor R has no effect on the operation of the circuit. In order to obtain a matched input at port 1, we require $Y_{in} = Y_c$. By using the transforming properties of the quarter-wave sections, we have

$$Y_{in} = \frac{Z_{L2}}{Z_2^2} + \frac{Z_{L3}}{Z_3^2} = Y_c \tag{6.69a}$$

From the above two equations we obtain

$$(K^2 Z_3^2 + Z_2^2) Z_{L3} = \frac{Z_2^2 Z_3^2}{Z_c} \tag{6.69b}$$

upon eliminating Z_{L2} using (6.68).

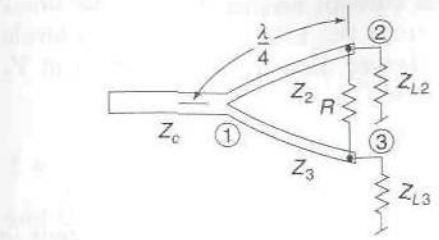
At port 1 the load impedance Z_{L2} is transformed into an admittance $Y_{in,2} = Z_{L2}/Z_2^2$ and Z_{L3} is transformed into an admittance $Y_{in,3} = Z_{L3}/Z_3^2$. The power delivered to Z_{L2} and Z_{L3} is the same as that delivered to $Y_{in,2}$ and $Y_{in,3}$, respectively. Thus, in order to get the desired power ratio, we require that

$$\tfrac{1}{2}|V_1^+|^2 Y_{in,3} = K^2 \tfrac{1}{2}|V_1^+|^2 Y_{in,2}$$

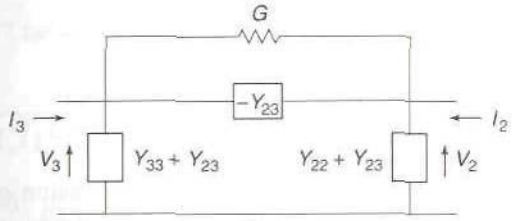
which gives, upon using (6.68),

$$Z_2 = K^2 Z_3 \tag{6.70}$$

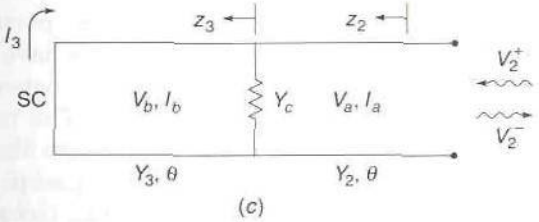
†S. B. Cohn, A Class of Broadband 3-Port TEM Hybrids, *IEEE Trans.*, vol. MTT-16, pp. 110–118, 1968.



(a)



(b)



(c)

FIGURE 6.37
(*a*) The Wilkinson power divider; (*b*) the equivalent circuit with port 1 terminated and ports 2 and 3 excited. (*c*) equivalent circuit between ports 2 and 3 with port 3 short-circuited and R removed.

At any frequency Y_{L2} transforms into

$$Y_{\text{in},2} = Y_2 \frac{Y_{L2} + jY_2 \tan\theta}{Y_2 + jY_{L2} \tan\theta}$$

and Y_{L3} transforms into

$$Y_{\text{in},3} = Y_3 \frac{Y_{L3} + jY_3 \tan\theta}{Y_3 + jY_{L3} \tan\theta} = K^2 Y_{in,2}$$

by virtue of (6.68) and (6.70). Thus the input current on line 3 will be K^2

larger than the input current on line 2. Since the line 2 and line 3 networks are identical, apart from the relative impedance levels, the load current in Y_{L3} will also be K^2 larger than the load current in Y_{L2}. Hence the ratio of the load voltages will be

$$\frac{V_{L2}}{V_{L3}} = \frac{I_{L2}Z_{L2}}{I_{L3}Z_{L3}} = \frac{I_{L2}Z_{L2}}{K^2 I_{L2}Z_{L3}} = 1$$

Since the load voltages are equal, there is no current in R at any frequency as long as ports 2 and 3 are terminated in their matched load impedances. When there is no current in R, the input admittance at port 1 is $Y_{\text{in},2} + Y_{\text{in},3} = (1 + K^2)Y_{\text{in},2}$. The input reflection coefficient is given by

$$\Gamma_{\text{in}} = \frac{Y_c - Y_{\text{in}}}{Y_c + Y_{in}}$$

$$= \frac{Y_c^2 - (1 + K^2)^2 Y_2^2}{Y_c^2 + (1 + K^2)^2 Y_2^2 + 2j(1 + K^2)Y_c Y_2 \tan\theta} \tag{6.71}$$

This result is obtained by substituting the expression given earlier for $Y_{\text{in},2}$ at any frequency. Equation (6.71) gives the interesting result that input port 1 will be matched at all frequencies if we choose

$$Y_2 = \frac{Y_c}{1 + K^2} \tag{6.72}$$

In order to analyze the coupling between ports 2 and 3, we will terminate port 1 in a load admittance Y_c. We now have a two-port network. Apart from the resistor R, we can represent the network between ports 2 and 3 by a Π network as shown in Fig. 6.37*b*. The resistor R is a shunt conductance $G = 1/R$ in parallel with $-Y_{23}$. From this network representation it is clear that ports 2 and 3 will be uncoupled if $G = Y_{23}$, so that the admittance between ports 2 and 3 vanishes. The circuit equations with G removed are

$$\begin{bmatrix} I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} Y_{22} & Y_{23} \\ Y_{23} & Y_{33} \end{bmatrix} \begin{bmatrix} V_2 \\ V_3 \end{bmatrix}$$

For $V_3 = 0$, that is, port 3 short-circuited, we have

$$Y_{23} = \left.\frac{I_3}{V_2}\right|_{V_3=0}$$

Thus we can evaluate Y_{23} by finding the short-circuit current at port 3 with port 2 excited. The transmission-line circuit to be analyzed is shown in Fig. 6.37*c*.

Let the voltage and current waves on line 2 be

$$V_a = V_a^+ e^{-j\beta_2 z_2} + V_a^- e^{j\beta_2 z_2}$$

$$I_a = \left(V_a^+ e^{-j\beta_2 z_2} - V_a^- e^{j\beta_2 z_2}\right) Y_2$$

and those on line 3 be

$$V_b = V_b^+ e^{-j\beta_3 z_3} + V_b^- e^{j\beta_3 z_3}$$

$$I_b = \left(V_b^+ e^{-j\beta_3 z_3} - V_b^- e^{j\beta_3 z_3}\right) Y_3$$

The terminal conditions at port 2 are

$$V_a^+ + V_a^- = V_2 \tag{6.73a}$$

$$(V_a^+ - V_a^-) Y_2 = I_2 \tag{6.73b}$$

At port 1 the terminal conditions are

$$V_a^+ e^{-j\theta} + V_a^- e^{j\theta} = V_b^+ + V_b^- \tag{6.73c}$$

$$\left(V_a^+ e^{-j\theta} - V_a^- e^{j\theta}\right) Y_2 = (V_b^+ + V_b^-) Y_c + (V_b^+ - V_b^-) Y_3 \tag{6.73d}$$

At port 3 we have

$$V_b^+ e^{-j\theta} + V_b^- e^{j\theta} = 0 \tag{6.73e}$$

$$\left(V_b^+ e^{-j\theta} - V_b^- e^{j\theta}\right) Y_3 = -I_3 \tag{6.73f}$$

where $\theta = \beta_2 l_2 = \beta_3 l_3$ and l_2 and l_3 are the line lengths. From the last two equations we get $I_3 = -2V_b^+ e^{-j\theta}$. By adding and subtracting (6.73c) multiplied by Y_2 to (6.73d), we can solve for V_b^+ in terms of either V_a^+ or V_a^-. We can combine the two solutions to get V_b^+ in terms of $V_a^+ + V_a^- = V_2$ and then solve for $Y_{23} = -(2V_b^+ e^{-j\theta})/V_2$. The results are

$$Y_{23} = \frac{2Y_2 Y_3}{Y_c(1 - \cos 2\theta) - j(Y_2 + Y_3)\sin 2\theta} \tag{6.74}$$

When $\theta = \pi/2$ we get $Y_{23} = Y_2 Y_3 / Y_c$. In order to uncouple ports 2 and 3, we thus require

$$G = Y_{23} = \frac{Y_2 Y_3}{Y_c} \tag{6.75a}$$

or

$$R = \frac{Z_2 Z_3}{Z_c} \tag{6.75b}$$

From (6.68), (6.69*b*), (6.70), and (6.75*b*), we can express all impedances in terms of R in the following form:

$$Z_2 = K\sqrt{RZ_c} \tag{6.76a}$$

$$Z_3 = \frac{1}{K}\sqrt{RZ_c} \tag{6.76b}$$

$$Z_{L2} = \frac{K^2R}{K^2+1} \tag{6.76c}$$

$$Z_{L3} = \frac{R}{K^2+1} \tag{6.76d}$$

The resistance R can be chosen arbitrarily. If we require port 1 to be matched at all frequencies, then R is determined by the condition given by (6.72).

We can also solve for I_2 in terms of V_2 from (6.73). The ratio I_2/V_2 for $V_3 = 0$ gives the parameter Y_{22}. We find that

$$Y_{22} = Y_2\left[\frac{Y_3 - Y_2 + (Y_3 + Y_2)\cos 2\theta + jY_c \sin 2\theta}{Y_c(\cos 2\theta - 1) + j(Y_2 + Y_3)\sin 2\theta}\right] \tag{6.77}$$

The parameter Y_{33} is obtained by replacing Y_2 by Y_3 and Y_3 by Y_2 in the above equation. The parameters Y_{22} and Y_{33} are needed in order to determine the port 2 and port 3 reflection coefficients as a function of θ or frequency.

A 3-dB power divider is obtained by choosing $K^2 = 1$. If we also specify that $Z_{L2} = Z_{L3} = Z_c$, we find from (6.76*c*) that $R = 2Z_c$. From (6.76*a*) we get $Z_2 = Z_3 = \sqrt{2}\,Z_c$ for this case. A 3-dB power divider that has port 1 matched at all frequencies requires, from (6.72), $Y_2 = Y_c/2$ or $Z_2 = 2Z_c$. From (6.76*a*) and (6.76*c*) we obtain $R = 4Z_c$ and $Z_{L2} = Z_{L3} = 2Z_c$. For this case the output lines have a characteristic impedance twice that of the input line. We could choose $R = 2Z_c$, $Z_{L2} = Z_{L3} = Z_c$ and use a quarter-wave transformer with impedance $Z_c/\sqrt{2}$ to transform the port 1 input impedance, which now equals $Z_c/2$, into an impedance Z_c. Figure 6.38*a* illustrates this type of 3-dB power divider.

For the unequal power division case if we choose Y_2 according to (6.72), the input port 1 is matched at all frequencies. From (6.76*a*) we find that $R = (1 + K^2)^2Z_c/K^2$. By using (6.76*c*) we get $Z_{L2} = (K^2 + 1)Z_c = Z_2$, and hence there is no discontinuity at the junction of line 2 and the load Z_{L2}. Also $Z_{L3} = Z_3$, so that there is no discontinuity at the junction of line 3 and its load. The only disadvantage with this design is that the two output impedances $Z_{L2} = (1 + K^2)Z_c$ and $Z_{L3} = (1 + K^2)Z_c/K^2$ are both different from Z_c. Output quarter-wave transformers can be added to transform Z_{L2} and Z_{L3} into the impedance Z_c. The required transformation ratios are $1 + K^2$ and $(1 + K^2)/K^2$ and are different. Thus the mismatch as a

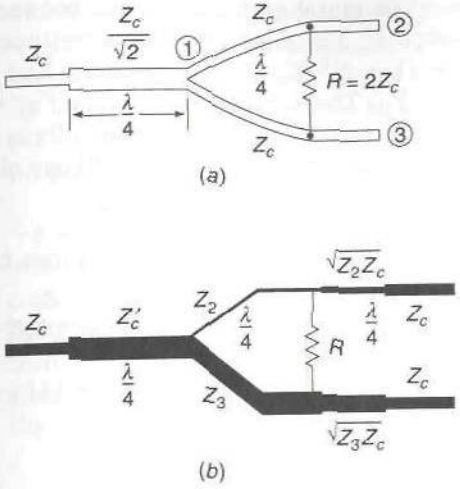


FIGURE 6.38
(*a*) A 3-dB power divider with an input quarter-wave transformer; (*b*) a broadband unequal power divider.

function of frequency will be different for the two output transformers. In order to equalize the transformation ratios and thereby obtain similar characteristics for ports 2 and 3, the output impedances should be chosen according to the relations

$$Z_{L2} = KZ_c \tag{6.78a}$$

$$Z_{L3} = \frac{Z_c}{K} \tag{6.78b}$$

We now require

$$R = \frac{1 + K^2}{K} Z_c \tag{6.78c}$$

$$Z_2 = Z_c \sqrt{K(1 + K^2)} \tag{6.78d}$$

$$Z_3 = Z_c \sqrt{\frac{1 + K^2}{K^3}} \tag{6.78e}$$

The required output quarter-wave transformers that will transform Z_{L2} and Z_{L3} into Z_c will have characteristic impedances given by $\sqrt{Z_{L2} Z_c} = \sqrt{K}\, Z_c$ and $\sqrt{Z_{L3} Z_c} = Z_c / \sqrt{K}$. A power divider designed on this basis will have a bandwidth approaching one octave.

A modified design that incorporates an input quarter-wave transformer as shown in Fig. 6.38*b* will give a significant increase in performance.† The design equations are arrived at by noting that when ports 2

†L. I. Parad and R. L. Moynihan, Split-Tee Power Divider, *IEEE Trans.*, vol. MTT-13, pp. 91–95, 1965.

and 3 are terminated in matched loads we can connect these ports at the location of the resistor R. The characteristic admittance of lines 2 and 3 in parallel is $Y_2 + Y_3 = (1 + K^2)Y_2$. The terminating load admittance is $Y_{L2} + Y_{L3} = (1 + K^2)Y_{L2} = Y_L$. The structure shown in Fig. 6.38*b* can be viewed as a two-section quarter-wave transformer. Thus, for a maximally flat input reflection coefficient characteristic (see Sec. 5.12), we should choose

$$Z_c' = Z_L^{1/4} Z_c^{3/4}$$

where Z_c' is the characteristic impedance of the input transformer, and

$$\frac{1}{(1 + K^2)Y_2} = Z_L^{3/4} Z_c^{1/4} = \frac{Z_2}{1 + K^2}$$

With $Z_{L2} = KZ_c$ and $Z_{L3} = Z_c/K$, we have $Y_L = (1 + K^2)Y_c/K$. Hence we must choose

$$Z_c' = \left(\frac{K}{1 + K^2}\right)^{1/4} Z_c \tag{6.79a}$$

$$Z_2 = K^{3/4}(1 + K^2)^{1/4} Z_c \tag{6.79b}$$

$$Z_3 = \frac{(1 + K^2)^{1/4}}{K^{5/4}} Z_c \tag{6.79c}$$

$$R = \frac{1 + K^2}{K} Z_c \tag{6.79d}$$

6.7 MICROWAVE PROPAGATION IN FERRITES

The development of ferrite materials suitable for use at microwave frequencies has resulted in a large number of microwave devices. A number of these have nonreciprocal electrical properties; i.e., the transmission coefficient through the device is not the same for different directions of propagation. An understanding of the operation of ferrite devices may be obtained once the basic nature of microwave propagation in an infinite unbounded ferrite medium is understood. In this section we consider plane-wave propagation in an infinite ferrite medium with a static biasing magnetic field B_0 present. It will be found that the natural modes of propagation in the direction of B_0 are left and right circularly polarized waves and that these modes have different propagation constants. In addition, we shall find that the permeability of the ferrite is not a single scalar quantity, but instead is a tensor, which can be represented as a matrix.

Ferrites are ceramiclike materials with specific resistivities that may be as much as 10^{14} greater than that of metals and with dielectric constants around 10 to 15 or greater. Ferrites are made by sintering a mixture of metallic oxides and have the general chemical composition $MO \cdot Fe_2O_3$, where M is a divalent metal such as manganese, magnesium, iron, zinc,

nickel, cadmium, etc., or a mixture of these. Relative permeabilities of several thousand are common. The magnetic properties of ferrites arise mainly from the magnetic dipole moment associated with the electron spin. By treating the spinning electron as a gyroscopic top, a classical picture of the magnetization process and, in particular, the anisotropic magnetic properties may be obtained.

The electron has a number of intrinsic properties such as a charge of $-e = -1.602 \times 10^{-19}$ C, a mass $w = 9.107 \times 10^{-31}$ kg, an angular momentum P equal in magnitude to $\frac{1}{2}\hbar$, or 0.527×10^{-34} J · s ($\hbar$ is Planck's constant divided by 2π), and a magnetic dipole moment m equal to one Bohr magneton, that is, $m = e\hbar/2w = 9.27 \times 10^{-24}$ A · m^2. For the electron, the angular momentum **P** and magnetic dipole moment **m** are antiparallel. The ratio of the magnetic moment to the angular momentum is called the gyromagnetic ratio γ; that is,

$$\gamma = \frac{m}{P} \tag{6.80}$$

If an electron is located in a uniform static magnetic field $\mathbf{B}_0$, a torque **T** given by

$$\mathbf{T} = \mathbf{m} \times \mathbf{B}_0 = -\gamma \mathbf{P} \times \mathbf{B}_0 \tag{6.81}$$

will be exerted on the dipole moment. This torque will cause the dipole axis to precess about an axis parallel to $\mathbf{B}_0$, as illustrated in Fig. 6.39. The equation of motion is obtained from the condition that the rate of change of angular momentum is equal to the torque and hence is

$$\frac{d\mathbf{P}}{dt} = \mathbf{T} = -\gamma \mathbf{P} \times \mathbf{B}_0 = \boldsymbol{\omega}_0 \times \mathbf{P} \tag{6.82a}$$

or

$$\gamma P B_0 \sin \phi = \omega_0 P \sin \phi = m B_0 \sin \phi \tag{6.82b}$$

where $\boldsymbol{\omega}_0$ is the vector-precession angular velocity directed along $\mathbf{B}_0$, and ϕ

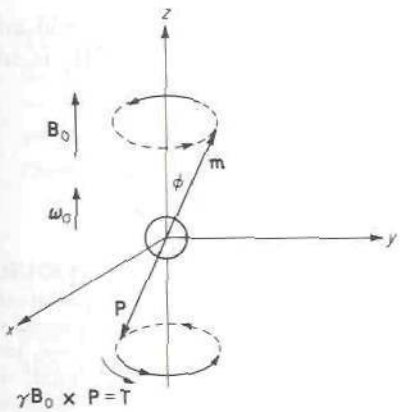


FIGURE 6.39
Free precession of spinning electron.

is the angle between $\mathbf{m}$ and $\mathbf{B}_0$. For free precession the angular velocity ω_0 is given by

$$\omega_0 = \gamma B_0$$

and is independent of the angle ϕ. The angular velocity ω_0 is often called the Larmor frequency.

If a small ac magnetic field is superimposed on the static field $\mathbf{B}_0$, the magnetic dipole moment will undergo a forced precession. Of particular interest is the case where the ac magnetic field is circularly polarized in the plane perpendicular to $\mathbf{B}_0$. A circularly polarized field results when the x and y components of the ac field are equal in magnitude and 90° apart in time phase. Thus let the ac magnetic field be given by the phasor

$$\mathbf{B}_1^- = B_1(\mathbf{a}_x + j\mathbf{a}_y) \tag{6.83a}$$

If we assume B_1 *to be real, the physical field is given by*

$$\mathbf{B}_1^- = B_1 \operatorname{Re}(\mathbf{a}_x + j\mathbf{a}_y)e^{j\omega t} = B_1(\mathbf{a}_x \cos \omega t - \mathbf{a}_y \sin \omega t) \tag{6.83b}$$

The resultant field has a constant magnitude B_1, but the orientation of the field in space changes or rotates with time. At time t the resultant field vector makes an angle

$$\tan^{-1} \frac{B_y}{B_x} = -\tan^{-1} \tan \omega t = -\omega t$$

with the axis and hence rotates at the rate $-\omega$, as in Fig. 6.40*a*. It is this rotation of the field vector in space that results in the field being called *circularly polarized. If the above ac magnetic field* is that of a wave propagating in the z direction, it is said to be left circularly polarized. If the direction of rotation is clockwise, looking in the direction of propagation, the wave is called right circularly polarized. The latter type of wave would have an ac magnetic field given by

$$\mathbf{B}_1^+ = B_1(\mathbf{a}_x - j\mathbf{a}_y) \tag{6.84}$$

With a left circularly polarized ac magnetic field superimposed on the *static field* $\mathbf{B}_0 = B_0\mathbf{a}_z$, the resultant total field $\mathbf{B}_t$ is inclined at an angle

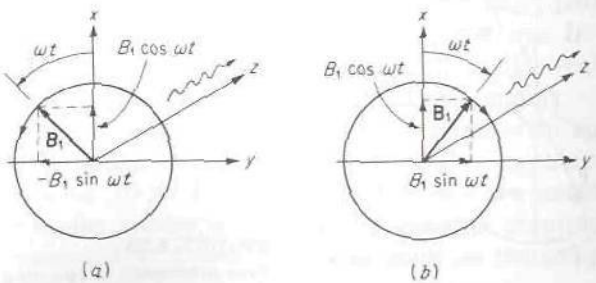


FIGURE 6.40
Magnetic field for circularly polarized waves. (*a*) Left, or negative, circular polarization; (*b*) right, or positive, circular polarization.

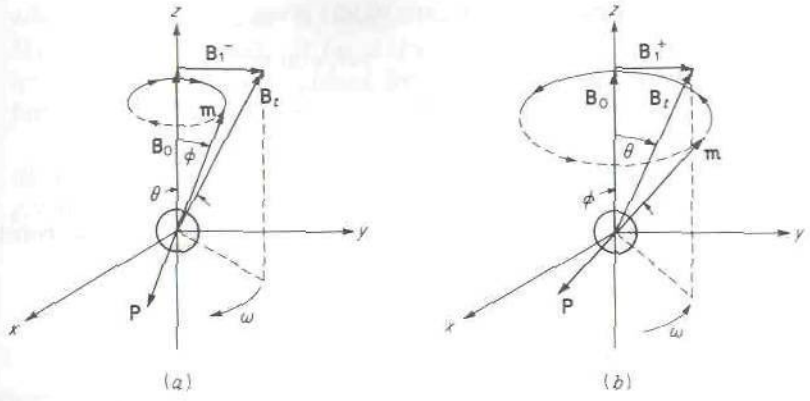


FIGURE 6.41
Forced precession of spinning electron.

$\theta = \tan^{-1}(B_1/B_0)$ with the z axis and rotates at a rate $-\omega$ about the z axis, as illustrated in Fig. 6.41*a*. Under steady-state conditions the magnetic dipole axis will be forced to precess about the z axis at the same rate. Thus the precession angle ϕ will have to be less than θ, as in Fig. 6.41*a*, in order to obtain a torque to cause precession in a counterclockwise direction. The equation of motion (6.82) gives

$$\mathbf{T} = \mathbf{m} \times \mathbf{B}_t = -\gamma \mathbf{P} \times \mathbf{B}_t = \frac{d\mathbf{P}}{dt} = -\omega \mathbf{a}_z \times \mathbf{P}$$

or

$$-\gamma P B_t \sin(\theta - \phi) = -\omega P \sin \phi$$

Expanding $\sin(\theta - \phi)$, replacing $B_t \sin \theta$ by B_1, $B_t \cos \theta$ by B_0, and solving for $\tan \phi$ gives

$$\tan \phi = \frac{\gamma B_1}{\gamma B_0 + \omega} = \frac{\gamma B_1}{\omega_0 + \omega} \tag{6.85}$$

The component of $\mathbf{m}$ which rotates in synchronism with $\mathbf{B}_1^-$ in the xy plane for the left (also called negative) circularly polarized ac field is $m^- = m \sin \phi = m_0 \tan \phi$, where $m_0 = m \cos \phi$ is the z-directed component of $\mathbf{m}$. For B_1 very small compared with B_0, the angle ϕ is small, so that m_0 is approximately equal to m. Using (6.85) gives

$$m^- = m_0 \tan \phi = \frac{\gamma m_0 B_1}{\omega_0 + \omega} \tag{6.86}$$

If we have a *right (or positive) circularly polarized ac field* superimposed on the static field $\mathbf{B}_0$, the forced precession is in a clockwise sense about the z axis. A torque giving precession in this direction is obtained only if the angle ϕ is greater than the angle θ, as in Fig. 6.41*b*. In this case

the equation of motion (6.82) gives

$$\gamma B_t \sin(\phi - \theta) = \omega \sin \phi$$

from which we obtain

$$\tan \phi = \frac{\gamma B_1}{\omega_0 - \omega} \tag{6.87}$$

The component of magnetization in the xy plane rotating in synchronism with the positive circularly polarized ac field is

$$m^+ = m_0 \tan \phi = \frac{\gamma m_0 B_1}{\omega_0 - \omega} \tag{6.88}$$

The foregoing discussion has pointed out the essential features of the motion of a single spinning electron in a magnetic field consisting of a static field along the z axis and a small circularly polarized ac field in the xy plane. A ferrite material may be regarded as a collection of N effective spinning electrons per unit volume. Since the spacing between electrons is of atomic dimensions, we may regard the density of magnetic dipoles per unit volume as a smeared-out continuous distribution from a macroscopic viewpoint. The total magnetic dipole moment per unit volume is $\mathbf{M} = N\mathbf{m}$. When the static field $\mathbf{B}_0$ is large enough to cause saturation of the magnetization in the ferrite, $\mathbf{M} = \mathbf{M}_s$. In a saturated ferrite all the spins are very tightly coupled, so that the whole sample acts essentially as a large single magnetic dipole. The magnetization $\mathbf{M}_s$ produces a contribution to the total internal $\mathbf{B}$ field according to the relation $\mathbf{B} = \mu_0(\mathbf{H}_0 + \mathbf{M}_s)$. The torque acting on $\mathbf{M}_s$ is due only to the field $\mathbf{B}_0 = \mu_0\mathbf{H}_0$, since the cross product of $\mu_0\mathbf{M}_s$ with $\mathbf{M}_s$ is zero and hence does not contribute to the torque. Thus, in the equation of motion for the magnetization, the field producing the torque is $\mu_0\mathbf{H}_t$, where $\mathbf{H}_t$ is the total static plus ac magnetic field intensity in the ferrite medium. That is,

$$\frac{d\mathbf{M}}{dt} = -\gamma(\mathbf{M} \times \mathbf{B}) = -\gamma\mu_0\mathbf{M} \times (\mathbf{H} + \mathbf{M}) = -\gamma\mu_0\mathbf{M} \times \mathbf{H}$$

If the magnetic field intensity in the ferrite is $\mathbf{H}_0 + \mathbf{H}_1^{\pm}$, where $\mathbf{H}_1^{\pm}$ is a circularly polarized ac field, the resultant ac magnetization will be given by expressions analogous to (6.86) and (6.88), but with m_0 replaced by $M_0 = Nm_0$. The total ac magnetic field in the xy plane is the field $\mu_0\mathbf{H}_1^{\pm} = \mathbf{B}_1^{\pm}$ plus the contribution from the ac magnetization. Thus the total ac fields for positive and negative circular polarization are [we are using the formula $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$]

$$\mathbf{B}^+ = \mu_0\mathbf{M}^+ + \mathbf{B}_1^+ = \mu_0(N\mathbf{m}^+ + \mathbf{H}_1^+) = \mu_0\left(1 + \frac{\mu_0\gamma M_0}{\omega_0 - \omega}\right)\mathbf{H}_1^+ \tag{6.89a}$$

$$\mathbf{B}^- = \mu_0\mathbf{M}^- + \mathbf{B}_1^- = \mu_0\left(1 + \frac{\mu_0\gamma M_0}{\omega_0 + \omega}\right)\mathbf{H}_1^- \tag{6.89b}$$

where $M_0 = M \cos \phi$, $B_1 = \mu_0 H_1$, and $\mathbf{H}_1^+ = H_1(\mathbf{a}_x - j\mathbf{a}_y)$ in (6.89a), and $\mathbf{H}_1^-$ in (6.89b) equals $H_1(\mathbf{a}_x + j\mathbf{a}_y)$, as seen from (6.84) and (6.83a). The quantity M may be replaced by the saturation magnetization M_s in the ferrite since the static field B_0 is usually large enough to cause saturation.

If we assume that $B_1 \ll B_0$, so that $M_0 \approx M_s$, the effective permeabilities for positive and negative circularly polarized ac fields are seen to be given by

$$\mu_+ = \mu_0 \left(1 + \frac{\gamma \mu_0 M_s}{\omega_0 - \omega}\right) \tag{6.90a}$$

$$\mu_- = \mu_0 \left(1 + \frac{\gamma \mu_0 M_s}{\omega_0 + \omega}\right) \tag{6.90b}$$

Plane circularly polarized TEM waves propagating in the direction of the static field $\mathbf{B}_0$ will have propagation constants

$$\beta_+ = \omega \sqrt{\epsilon \mu_+} \tag{6.91a}$$

$$\beta_- = \omega \sqrt{\epsilon \mu_-} \tag{6.91b}$$

where ϵ is the dielectric permittivity of the ferrite. The significance of the inequality of β_+ and β_- is discussed later. The results expressed by (6.91) are also derived in an alternative way later.

If small-signal conditions $B_1 \ll B_0$ are not assumed, we cannot replace M_0 by M_s. In place of (6.85) and (6.87), which give solutions for $\tan \phi$, we can solve for $\sin \phi$ to obtain, respectively,

$$\sin \phi = \frac{\tan \phi}{\sqrt{1 + \tan^2 \phi}} = \frac{\gamma B_1}{\sqrt{(\gamma B_1)^2 + (\omega_0 + \omega)^2}} \tag{6.92a}$$

$$\sin \phi = \frac{\gamma B_1}{\sqrt{(\gamma B_1)^2 + (\omega_0 - \omega)^2}} \tag{6.92b}$$

The magnetizations M^+ and M^- are given by

$$M^+ = M_s \sin \phi = \frac{\gamma \mu_0 M_s H_1}{\sqrt{(\gamma \mu_0 H_1)^2 + (\omega_0 - \omega)^2}} \tag{6.93a}$$

$$M^- = \frac{\gamma \mu_0 M_s H_1}{\sqrt{(\gamma \mu_0 H_1)^2 + (\omega_0 + \omega)^2}} \tag{6.93b}$$

It is seen that the ac magnetization depends nonlinearly on the ac field strength H_1, and hence, under large-signal conditions, μ_+ and μ_- will be functions of the applied ac field strength. The nonlinear behavior of ferrites under large-signal conditions results in the generation of harmonics of the

fundamental frequency ω. For this reason ferrites may be used as harmonic generators.†

For M^- it is clear that, if $B_1 \ll B_0$, that is, $\gamma B_1 \ll \gamma B_0 = \omega_0$, then (6.93*b*) is well approximated by

$$M^- = \frac{\gamma\mu_0 M_s H_1}{\omega_0 + \omega}$$

Similarly, if ω is not too close to the resonant frequency ω_0, we see that (6.93*a*) becomes

$$M^+ = \frac{\gamma\mu_0 M_s H_1}{\omega_0 - \omega}$$

These latter values of M^+ and M^- lead directly to the expressions (6.90) for μ_+ and μ_-. In any practical ferrite medium, damping effects are always present, so that M^+ will remain finite and small compared with M_s even when $\omega = \omega_0$. Thus, for small-signal conditions, we can assume that $M_0 \approx M_s$ in an actual ferrite medium. Damping effects are discussed in more detail later.

It will be instructive to study the propagation of a plane wave in an unbounded ferrite medium by solving Maxwell's equations directly, together with the equation of motion for the magnetization. This analysis will illustrate the general technique of linearization to be applied in the small-signal analysis of propagation through a medium such as a ferrite. However, it will not give as clear an insight into the physical reason why μ_+ and μ_- are different, as the analysis above did. That is, basically, μ_+ and μ_- differ because the precession angle ϕ must be greater than the angle θ in one case and less than θ in the other case, and hence the projection of the magnetic dipole moment onto the xy plane is different in the two cases.

Consider an infinite unbounded ferrite medium with an applied static magnetic field $\mathbf{B}_0 = \mu_0\mathbf{H}_0 = B_0\mathbf{a}_z$. Let the magnetization in the ferrite be $\mathbf{M}_s$ per unit volume when no time-varying magnetic field is applied. When a time-varying magnetic field $\mu_0\mathbf{H}$ is also applied, a time-varying component $\mathscr{M}$ of magnetization will be produced. The equation of motion for the total magnetization per unit volume is similar to that for a single electron, and hence we have

$$\begin{aligned}\frac{d(\mathbf{M}_s + \mathscr{M})}{dt} = \frac{d\mathscr{M}}{dt} &= -\gamma[(\mathbf{M}_s + \mathscr{M}) \times (\mathbf{B}_0 + \mu_0\mathscr{H})] \\ &= -\gamma\mu_0(\mathbf{M}_s \times \mathbf{H}_0 + \mathbf{M}_s \times \mathscr{H} + \mathscr{M} \times \mathbf{H}_0 + \mathscr{M} \times \mathscr{H}) \quad (6.94)\end{aligned}$$

†W. P. Ayres, P. H. Vartanian, and J. L. Melchor, Frequency Doubling in Ferrites, *J. Appl. Phys.*, vol. 27, p. 188, 1956; Microwave Frequency Doubling from 9 kmc to 18 kmc in Ferrites, *Proc. IRE*, vol. 45, pp. 643–646, May, 1957.

If small-signal conditions are assumed, i.e.,

$$|\mathscr{M}| \ll |\mathbf{M}_s| \qquad \text{and} \qquad |\mathscr{H}| \ll |\mathbf{H}_0|$$

the nonlinear term $\mathscr{M} \times \mathscr{H}$ in (6.94) may be dropped. We then obtain for the equation of motion the linearized equation

$$\frac{d\mathscr{M}}{dt} = -\gamma(\mu_0 \mathbf{M}_s \times \mathscr{H} + \mathscr{M} \times \mathbf{B}_0) \tag{6.95}$$

since $\mathbf{M}_s \times \mathbf{B}_0 = 0$, because the saturation magnetization is in the same direction as the applied static field.

Let the time dependence be $e^{j\omega t}$, and let $\mathscr{M}$ and $\mathscr{H}$ be represented by the phasors $\mathbf{M}$ and $\mathbf{H}$. From (6.95) we obtain

$$j\omega\mathbf{M} + \gamma\mathbf{M} \times \mathbf{B}_0 = j\omega\mathbf{M} + \omega_0\mathbf{M} \times \mathbf{a}_z = -\gamma\mu_0\mathbf{M}_s \times \mathbf{H}$$

where $\omega_0 = \gamma\mu_0 H_0 = \gamma B_0$. In component form we have

$$j\omega M_x + \omega_0 M_y = \gamma M_s \mu_0 H_y$$

$$j\omega M_y - \omega_0 M_x = -\gamma M_s \mu_0 H_x$$

$$j\omega M_z = 0$$

The solution of these equations for M_x, M_y, and M_z gives

$$M_x = \frac{\omega_0 \gamma\mu_0 M_s H_x + j\omega\gamma\mu_0 M_s H_y}{\omega_0^2 - \omega^2} \tag{6.96a}$$

$$M_y = \frac{\omega_0 \gamma\mu_0 M_s H_y - j\omega\gamma\mu_0 M_s H_x}{\omega_0^2 - \omega^2} \tag{6.96b}$$

$$M_z = 0 \tag{6.96c}$$

In the solution of Maxwell's equations it is convenient not to have to deal explicitly with the magnetization. The magnetization may be eliminated by introducing the magnetic susceptibility and permeability. In the scalar case this is done by means of the relations $M = \chi_m H$, $B = \mu_0(M + H) = \mu_0(1 + \chi_m)H = \mu H$. For a ferrite similar relations may be used, but χ_m and μ will not be scalar quantities. In matrix form (6.96) gives

$$\begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} = \begin{bmatrix} \chi_{xx} & \chi_{xy} & 0 \\ \chi_{yx} & \chi_{yy} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} \tag{6.97}$$

where

$$\chi_{xx} = \chi_{yy} = \frac{\omega_0 \gamma \mu_0 M_s}{\omega_0^2 - \omega^2}$$

$$\chi_{xy} = -\chi_{yx} = \frac{j\omega \gamma \mu_0 M_s}{\omega_0^2 - \omega^2}$$

The matrix with the parameters χ_{xx}, χ_{xy}, χ_{yx}, and χ_{yy} in (6.97) represents the susceptibility tensor of the ferrite. The relation between the ac **B** and **H** fields is

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$$

or

$$\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \mu_0 \begin{bmatrix} 1 + \chi_{xx} & \chi_{xy} & 0 \\ \chi_{yx} & 1 + \chi_{yy} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} \tag{6.98}$$

The matrix relating the components of **H** to **B** in (6.98) is the permeability tensor for the ferrite. It will be denoted by a boldface μ with an overbar, i.e.,

$$\bar{\boldsymbol{\mu}} = \mu_0 \begin{bmatrix} 1 + \chi_{xx} & \chi_{xy} & 0 \\ \chi_{yx} & 1 + \chi_{yy} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In shorthand notation the matrix equation (6.98) will be written as

$$\mathbf{B} = \bar{\boldsymbol{\mu}} \cdot \mathbf{H} \tag{6.99}$$

In the literature the minus sign in the equation of motion is often deleted, and this amounts to replacing γ by $-\gamma$ in the equations used in this text.

Losses present in a ferrite may be accounted for in a phenomenological way by introducing into the equation of motion a damping term that will produce a torque tending to reduce the precession angle ϕ. The following modified form of the equation has often been used in practice:

$$\frac{d\mathscr{M}}{dt} = -\gamma \mu_0 (\mathbf{M}_s + \mathscr{M}) \times (\mathbf{H}_0 + \mathscr{H}) + \frac{\alpha}{M_s} \mathbf{M}_s \times \frac{d\mathscr{M}}{dt} \tag{6.100}$$

where α is a dimensionless damping constant. With a small-signal analysis, the elements of the susceptibility matrix are now found to be†

$$\chi_{xx} = \chi_{yy} = \chi' - j\chi'' = \chi \tag{6.101a}$$

$$\chi_{xy} = -\chi_{yx} = j(K' - jK'') = jK \tag{6.101b}$$

†R. F. Soohoo, "Theory and Application of Ferrites," chap. 5, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1960.

where

$$\chi' = \frac{\omega_0\omega_m(\omega_0^2 - \omega^2) + \omega_m\omega_0\omega^2\alpha^2}{\left[\omega_0^2 - \omega^2(1+\alpha^2)\right]^2 + 4\omega_0^2\omega^2\alpha^2}$$

$$\chi'' = \frac{\omega\omega_m\alpha\left[\omega_0^2 + \omega^2(1+\alpha^2)\right]}{\left[\omega_0^2 - \omega^2(1+\alpha^2)\right]^2 + 4\omega_0^2\omega^2\alpha^2}$$

$$K' = \frac{\omega\omega_m\left[\omega_0^2 - \omega^2(1+\alpha^2)\right]}{\left[\omega_0^2 - \omega^2(1+\alpha^2)\right]^2 + 4\omega_0^2\omega^2\alpha^2}$$

$$K'' = \frac{2\omega^2\omega_0\omega_m\alpha}{\left[\omega_0^2 - \omega^2(1+\alpha^2)\right]^2 + 4\omega_0^2\omega^2\alpha^2}$$

and $$\omega_m = \mu_0\gamma M_s$$

As we derived the permeability tensor to use in the constitutive equation relating **B** and **H**, the only remaining task is to find solutions for Maxwell's equations in the form

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B} = -j\omega\bar{\boldsymbol{\mu}} \cdot \mathbf{H} \qquad (6.102a)$$

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E} \qquad (6.102b)$$

$$\nabla \cdot \mathbf{B} = \nabla \cdot \mathbf{E} = 0 \qquad (6.102c)$$

For a TEM wave propagating in the z direction, i.e., along the direction of $\mathbf{B}_0$, solutions are readily found. Let the electric field be given by

$$\mathbf{E} = \mathbf{E}_0 e^{-j\beta z}$$

where $\mathbf{E}_0$ is a constant vector in the xy plane. Equation (6.102a) gives

$$\nabla \times \mathbf{E} = -\mathbf{E}_0 \times \nabla e^{-j\beta z} = j\beta\mathbf{E}_0 \times \mathbf{a}_z e^{-j\beta z} = -j\omega\bar{\boldsymbol{\mu}} \cdot \mathbf{H}$$

Let the solution for **H** be $\mathbf{H}_0 e^{-j\beta z}$, where $\mathbf{H}_0$ is also a constant vector in the xy plane. From (6.102b) we obtain

$$j\beta\mathbf{H}_0 \times \mathbf{a}_z = j\omega\epsilon\mathbf{E}_0$$

If this equation is substituted into the equation

$$j\beta\mathbf{E}_0 \times \mathbf{a}_z = -j\omega\bar{\boldsymbol{\mu}} \cdot \mathbf{H}_0$$

so as to eliminate $\mathbf{E}_0$, we obtain

$$j\beta\frac{j\beta}{j\omega\epsilon}(\mathbf{H}_0 \times \mathbf{a}_z) \times \mathbf{a}_z = -j\omega\bar{\boldsymbol{\mu}} \cdot \mathbf{H}_0$$

Expanding the left-hand side gives (note that $\mathbf{a}_z \cdot \mathbf{H}_0 = 0$)

$$\beta^2\mathbf{H}_0 = \omega^2\epsilon\bar{\boldsymbol{\mu}} \cdot \mathbf{H}_0 \qquad (6.103a)$$

This equation may be written in the following matrix form:

$$\begin{bmatrix} \beta^2 - \omega^2\epsilon\mu_0(1+\chi) & -j\omega^2\epsilon\mu_0 K \\ j\omega^2\epsilon\mu_0 K & \beta^2 - \omega^2\epsilon\mu_0(1+\chi) \end{bmatrix} \begin{bmatrix} H_{0x} \\ H_{0y} \end{bmatrix} = 0 \quad (6.103b)$$

For a nontrivial solution for $\mathbf{H}_0$, the determinant must vanish. This condition yields the following eigenvalue equation for the propagation constant β:

$$\left[\beta^2 - \omega^2\epsilon\mu_0(1+\chi)\right]^2 - \omega^4\epsilon^2\mu_0^2 K^2 = 0$$

or

$$\beta^2 = \omega^2\epsilon\mu_0(1+\chi) \pm \omega^2\epsilon\mu_0 K \quad (6.104)$$

If we substitute for χ and K and assume a lossless ferrite, so that $\chi'' = K'' = 0$, we readily find that the two solutions for β^2 are

$$\beta^2 = \beta_+^2 = \omega^2\epsilon\mu_+ \quad (6.105a)$$

$$\beta^2 = \beta_-^2 = \omega^2\epsilon\mu_- \quad (6.105b)$$

where μ_+ and μ_- are given by (6.90).

For each eigenvalue or solution for β^2, the ratio of H_{0x} to H_{0y} is determined. For the solution β_+^2 the first equation in the pair of equations (6.103*b*) gives

$$\left[\beta_+^2 - \omega^2\epsilon\mu_0(1+\chi)\right] H_{0x} - j\omega^2\epsilon\mu_0 K H_{0y} = 0$$

or

$$\frac{H_{0x}}{H_{0y}} = j \quad (6.106a)$$

But this condition means that $\mathbf{H}_0 = H_0(\mathbf{a}_x - j\mathbf{a}_y)$, or is a positive circularly polarized wave. Similarly, the solution β_-^2 gives

$$\frac{H_{0x}}{H_{0y}} = -j \quad (6.106b)$$

which specifies a negative circularly polarized wave. Therefore it is seen that the natural modes of propagation along the direction of the static field in a ferrite are circularly polarized TEM waves. If directions of propagation other than along $\mathbf{B}_0$ were considered, it would be found that there are, again, two modes of propagation, but these are no longer circularly polarized TEM waves. For a linearly polarized wave propagating along $\mathbf{B}_0$, the plane of polarization rotates, since β_+ and β_- are not equal. This phenomenon is a nonreciprocal one, and is called Faraday rotation. It is discussed in the following section.

6.8 FARADAY ROTATION

Consider an infinite lossless ferrite medium with a static field $\mathbf{B}_0$ applied along the z direction. Let a plane TEM wave that is linearly polarized along

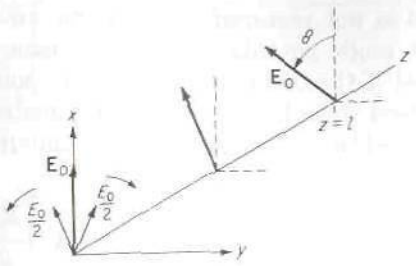


FIGURE 6.42
Faraday rotation.

the x axis at $z = 0$ be propagating in the z direction, as in Fig. 6.42. We shall show that the plane of polarization of this wave rotates as it propagates (Faraday rotation). The linearly polarized wave may be decomposed into the sum of a left and right circularly polarized wave as follows:

$$\mathbf{E} = \mathbf{a}_x E_0 = (\mathbf{a}_x + j\mathbf{a}_y)\frac{E_0}{2} + (\mathbf{a}_x - j\mathbf{a}_y)\frac{E_0}{2} \qquad z = 0 \qquad (6.107)$$

The component waves propagate with different phase constants β_+ and β_-, and hence the wave at $z = l$ becomes

$$\begin{aligned}
\mathbf{E} &= (\mathbf{a}_x + j\mathbf{a}_y)\frac{E_0}{2}e^{-j\beta_- l} + (\mathbf{a}_x - j\mathbf{a}_y)\frac{E_0}{2}e^{-j\beta_+ l} \\
&= \mathbf{a}_x \frac{E_0}{2}(e^{-j\beta_- l} + e^{-j\beta_+ l}) + j\mathbf{a}_y \frac{E_0}{2}(e^{-j\beta_+ l} - e^{-j\beta_+ l}) \\
&= \frac{E_0}{2}e^{-j(\beta_- + \beta_+)l/2}\Big[\mathbf{a}_x(e^{-j(\beta_- - \beta_+)l/2} + e^{j(\beta_- - \beta_+)l/2}) \\
&\qquad + j\mathbf{a}_y(e^{-j(\beta_- - \beta_+)l/2} - e^{j(\beta_- - \beta_+)l/2})\Big] \\
&= E_0 e^{-j(\beta_- + \beta_+)l/2}\left[\mathbf{a}_x \cos(\beta_+ - \beta_-)\frac{l}{2} - \mathbf{a}_y \sin(\beta_+ - \beta_-)\frac{l}{2}\right] \qquad (6.108)
\end{aligned}$$

This resultant wave is a linearly polarized wave that has undergone a phase delay of $(\beta_- + \beta_+)l/2$. The new plane of polarization makes an angle

$$\theta = \tan^{-1}\frac{E_y}{E_x} = \tan^{-1}\left[-\tan(\beta_+ - \beta_-)\frac{l}{2}\right] \approx -(\beta_+ - \beta_-)\frac{l}{2} \qquad (6.109)$$

with respect to the x axis. When $\omega < \omega_0$, that is, below the ferrite resonant frequency, β_+ is greater than β_- and the plane of polarization rotates counterclockwise, looking in the direction of propagation (Fig. 6.42). The rate of rotation is $(\beta_+ - \beta_-)/2$ rad/m. Rotation of 100° or more per centimeter is typical in ferrites at a frequency of 10,000 MHz/s.

If the direction of propagation is reversed, the plane of polarization continues to rotate in the same direction. Thus, if we consider the propagation of the wave described by (6.108) back to the plane $z = 0$, the original

polarization direction is not restored; instead the wave will arrive back at $z = 0$ polarized at an angle 2θ relative to the x axis. This result is easily derived by noting that if the component circularly polarized waves making up the linearly polarized wave in (6.108) are propagated from $z = l$ back to $z = 0$, they undergo additional phase delays of amount $\beta_+ l$ and $\beta_- l$ and become, at $z = 0$,

$$\mathbf{E} = (\mathbf{a}_x + j\mathbf{a}_y)\frac{E_0}{2}e^{-2j\beta_- l} + (\mathbf{a}_x - j\mathbf{a}_y)\frac{E_0}{2}e^{-2j\beta_+ l} \tag{6.110}$$

By analogy with (6.108) it is now clear that the new direction of polarization at $z = 0$ makes an angle

$$2\theta = -(\beta_+ - \beta_-)l$$

with respect to the x axis. Thus Faraday rotation is a nonreciprocal effect.

A practical ferrite medium has finite losses, and this will have a significant influence on the propagation. The propagation constants $\gamma_+ = j\beta_+ + \alpha_+$ and $\gamma_- = j\beta_- + \alpha_-$ for circularly polarized waves will have unequal attenuation constants as well as unequal phase constants. When losses are present, the propagation constants are given by (6.104) if β^2 is replaced by $-\gamma^2$; thus

$$\gamma_+ = j\omega\sqrt{\mu_0\epsilon}\,(1 + \chi' - j\chi'' + K' - jK'')^{1/2} \tag{6.111a}$$

$$\gamma_- = j\omega\sqrt{\mu_0\epsilon}\,(1 + \chi' - j\chi'' - K' + jK'')^{1/2} \tag{6.111b}$$

where χ', χ'', K', K'' are given after (6.101). The solutions of (6.111) are

$$\beta_\pm = \frac{\omega\sqrt{\mu_0\epsilon}}{\sqrt{2}}\left[1 + \chi' \pm K' + \sqrt{(1 + \chi' \pm K')^2 + (\chi'' \pm K'')^2}\right]^{1/2} \tag{6.112a}$$

$$\alpha_\pm = \omega^2\mu_0\epsilon\frac{\chi'' \pm K''}{2\beta_\pm} = \frac{\omega^2\epsilon\mu''_\pm}{2\beta_\pm} \tag{6.112b}$$

The permeabilities for circularly polarized waves are

$$\mu_\pm = \mu'_\pm - j\mu''_\pm = \mu_0(1 + \chi' - j\chi'' \pm K' \mp jK'') \tag{6.113}$$

The values of $\mu'_\pm$, $\mu''_\pm$ as given by (6.101) and (6.113) and the propagation factors $\beta_\pm$ and $\alpha_\pm$ are plotted in Figs. 6.43 and 6.44 for a typical ferrite with parameters $\omega_m = 2\pi \times 5.6 \times 10^9$, $\alpha = 0.05$, as a function of ω_0 at a frequency of 10,000 MHz. Note that ω_0 equals $2\pi \times 2.8$ MHz/Oe of applied field H_0 and that $4\pi \times 10^{-3}$ Oe is a field strength of 1 A/m. The value of ω_m chosen corresponds to a saturation magnetization of 2,000 G, or $\mu_0 M_s$ equal to 0.2 Wb/m^2. The curves in Fig. 6.44 show that α_- is always very small but that α_+ is large in the vicinity of the resonant frequency $\omega_0 \approx \omega$. For ω_0 considerably above ω, the attenuation α_+ becomes small, but in this region β_+ and β_- do not differ greatly, so that the rate of Faraday rotation would be small. At low values of ω_0, that is, small

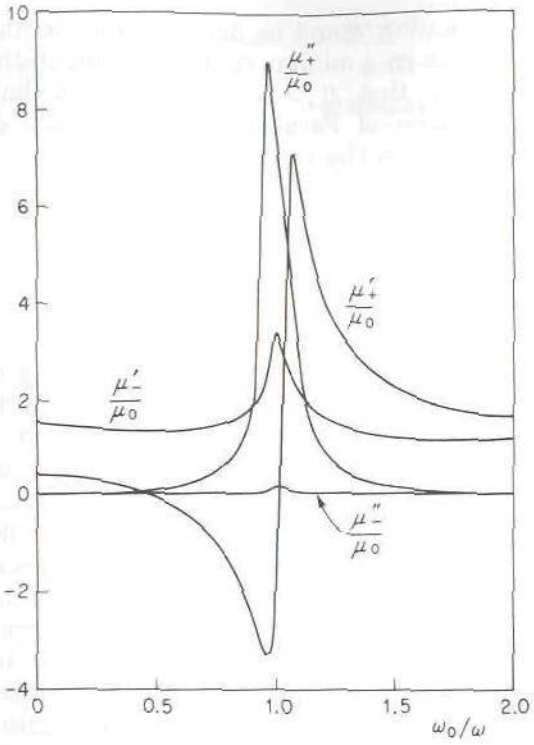


FIGURE 6.43
Real and imaginary components of permeability for circularly polarized waves in a ferrite as a function of ω/ω_0 for $\omega/2\pi = 10$ GHz, $\omega_m/2\pi = 5.6$ GHz, $\alpha = 0.05$.

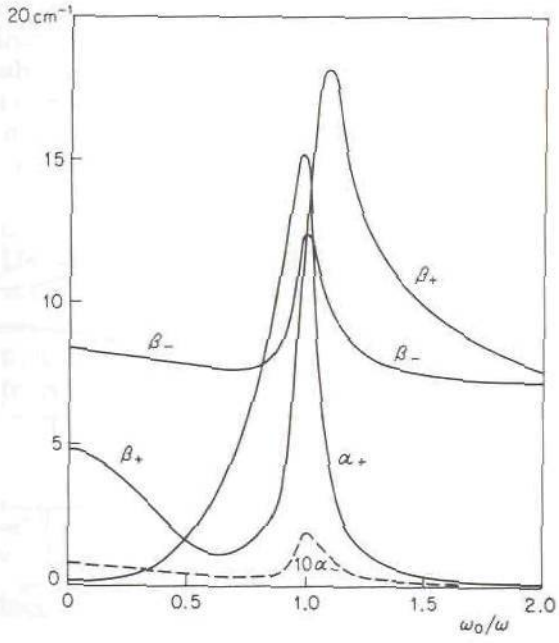


FIGURE 6.44
Propagation and attenuation constants for circularly polarized waves in a ferrite, with parameters given in Fig. 6.43 ($\epsilon = 10\epsilon_0$). Note that $10\alpha_-$ is plotted since α_- is very small.

H_0, the rate of rotation would be much greater, particularly in the region where β_+ goes through a minimum, which occurs in the range where μ_+ is negative. Note also that $\beta_+ > \beta_-$ when $\omega_0 > \omega$ but that $\beta_+ < \beta_-$ for $\omega_0 < \omega$. The direction of Faraday rotation is thus different in the two regions above and below the resonant frequency.

6.9 MICROWAVE DEVICES EMPLOYING FARADAY ROTATION

Gyrator

A gyrator is defined as a two-port device that has a relative difference in phase shift of 180° for transmission from port 1 to port 2 as compared with the phase shift for transmission from port 2 to port 1. A gyrator may be obtained by employing the nonreciprocal property of Faraday rotation. Figure 6.45 illustrates a typical microwave gyrator. It consists of a rectangular guide with a 90° twist connected to a circular guide, which in turn is connected to another rectangular guide at the other end. The two rectangular guides have the same orientation at the input ports. The circular guide contains a thin cylindrical rod of ferrite with the ends tapered to reduce reflections. A static axial magnetic field is applied so as to produce 90° Faraday rotation of the TE_{11} dominant mode in the circular guide. Consider a wave propagating from left to right. In passing through the twist the plane of polarization is rotated by 90° in a counterclockwise direction. If the ferrite produces an additional 90° of rotation, the total angle of rotation will be 180°, as indicated in Fig. 6.45. For a wave propagating from right to left, the Faraday rotation is still 90° in the same sense. However, in passing through the twist, the next 90° of rotation is in a direction to cancel the Faraday rotation. Thus, for transmission from port 2 to port 1, there is no

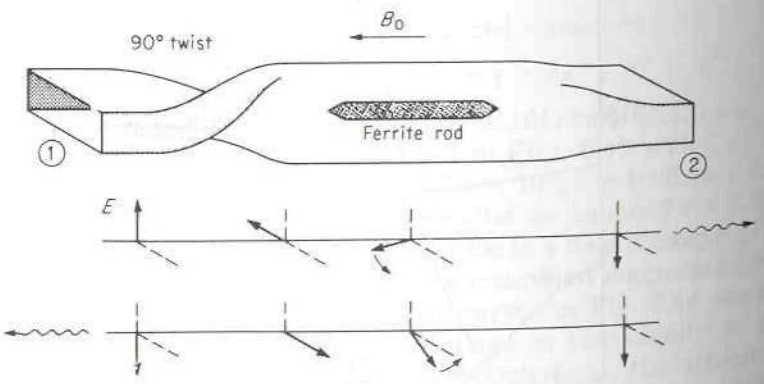


FIGURE 6.45
A microwave gyrator.

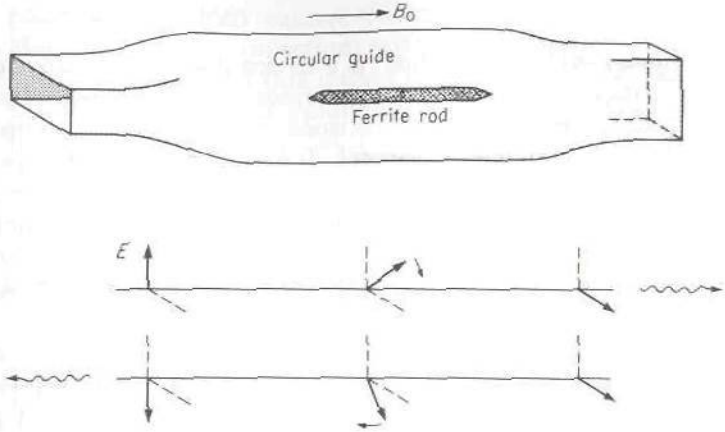


FIGURE 6.46
A gyrator without a twist section.

net rotation of the plane of polarization. The 180° rotation for transmission from port 1 to port 2 is equivalent to an additional 180° of phase shift since it reverses the polarization of the field. It is apparent, then, that the device just described satisfies the definition of a gyrator.

If the inconvenience of having the input and output rectangular guides oriented at 90° can be tolerated, a gyrator without a 90° twist section can be built. With reference to Fig. 6.46, it is seen that if the ferrite produces 90° of rotation and the output guide is rotated by 90° relative to the input guide, the emerging wave will have the right polarization to propagate in the output guide. When propagation is from port 2 to port 1, the wave arriving in guide 1 will have its polarization changed by 180°, as shown in Fig. 6.46. Hence a differential phase shift of 180° is again produced.

The solution for wave propagation in a circular guide with a longitudinal magnetized cylinder placed in the center can be carried out exactly.† However, the solution requires a great deal of algebraic manipulation, and it is very laborious to compute numerical values from the resultant transcendental equations for the propagation constants. The solution does verify that Faraday rotation takes place as would be expected, by analogy with propagation in an infinite ferrite medium.

---

†A. A. Th. M. van Trier, Guided Electromagnetic Waves in Anisotropic Media, *Appl. Sci. Res.*, vol. B3, p. 305, 1953.

M. L. Kales, Modes in Waveguides That Contain Ferrites, *J. Appl. Phys.*, vol. 24, p. 604, 1953.

Isolator

The isolator, or uniline, is a device that permits unattenuated transmission from port 1 to port 2 but provides very high attenuation for transmission in the reverse direction. The isolator is often used to couple a microwave signal generator to a load network. It has the great advantage that all the available power can be delivered to the load and yet reflections from the load do not get transmitted back to the generator output terminals. Consequently, the generator sees a matched load, and effects such as power output variation and frequency pulling (change in frequency), with variations in the load impedance, are avoided.

The isolator is similar to the gyrator in construction except that it employs a 45° twist section and 45° of Faraday rotation. In addition, thin resistive cards are inserted in the input and output guides to absorb the field that is polarized, with the electric vector parallel to the wide side of the guide, as shown in Fig. 6.47. The operation is as follows: A wave propagating from port 1 to port 2 has its polarization rotated 45° counterclockwise by the twist section and 45° clockwise by the Faraday rotator. It will emerge at port 2 with the correct polarization to propagate in the output guide. A wave propagating from port 2 to port 1 will have its plane of polarization rotated by 90° and will enter the guide at port 1 with the electric field parallel to the resistance card, and hence be absorbed. Without the resistance card, the wave would be reflected from port 1 because of the incorrect

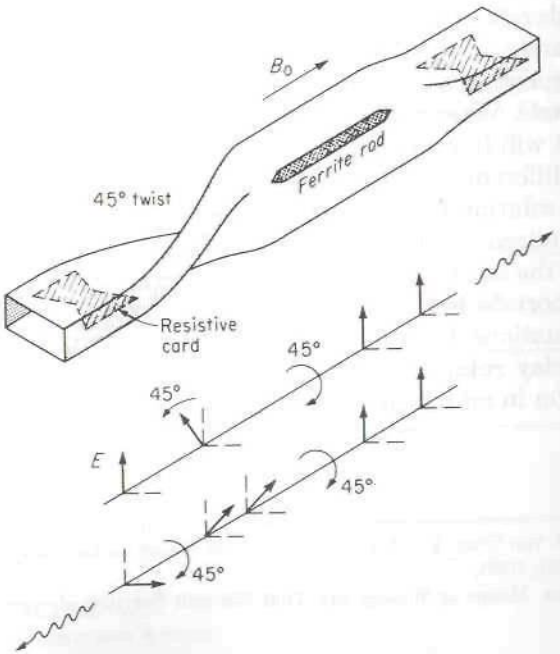


FIGURE 6.47
A Faraday-rotation isolator.

polarization, which cannot propagate in the guide constituting port 1. However, multiple reflections within the isolator will lead to transmission in both directions, and this makes it necessary to use resistance cards in both the input and output guides for satisfactory performance. Typical performance figures for an isolator are forward transmission loss of less than 1 dB, reverse attenuation of 20 to 30 dB, and bandwidth of operation approaching 10 percent.

Resonance Isolator

If the curves in Fig. 6.44 for the propagation constants of circularly polarized waves in an infinite ferrite medium are examined, it will be seen that the attenuation constant for negative circular polarization is always very small whereas that for positive circular polarization is very large in the vicinity of the resonance point $\omega_0 \approx \omega$. This property may be used as the basis for a resonance isolator by using a negative circularly polarized wave for transmission in the low-loss direction and a positive circularly polarized wave for transmission in the reverse direction. In the latter case the wave is rapidly absorbed or attenuated.

The condition for circular polarization is an inherent property of the dominant TE_{10} mode in a rectangular guide at two positions within the guide. The TE_{10}-mode fields are

$$E_y = \sin\frac{\pi x}{a} e^{+j\beta z}$$

$$j\omega\mu_0 H_x = \pm j\beta \sin\frac{\pi x}{a} e^{+j\beta z}$$

$$j\omega\mu_0 H_z = -\frac{\pi}{a}\cos\frac{\pi x}{a} e^{+j\beta z}$$

Since H_x and H_z differ in phase by 90°, circular polarization occurs when $|H_x| = |H_z|$, or when $x = x_1$, where

$$\tan\frac{\pi x_1}{a} = \pm\frac{\pi}{\beta a} = \pm\frac{\lambda_g}{2a} \tag{6.114}$$

For the solution in the range $0 < x_1 < a/2$, the ratio of H_x to H_z is

$$\frac{H_x}{H_z} = \mp j \tag{6.115a}$$

and the solution occurring for $a/2 < x_1 < a$ gives

$$\frac{H_x}{H_z} = \pm j \tag{6.115b}$$

With respect to the y axis, the solution given by (6.115a) corresponds to a negative circularly polarized field for propagation in the $+z$ direction and to

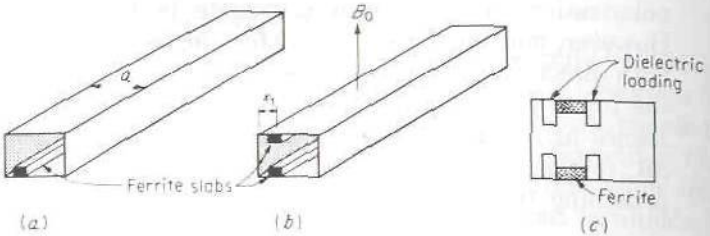


FIGURE 6.48
Rectangular-waveguide resonance isolators.

positive circular polarization for propagation in the $-z$ direction. If the solution given by (6.115*b*) is considered, the direction of polarization is reversed.

The above property of the TE_{10} mode is utilized in the resonance isolator by locating a thin ferrite slab (or two slabs, as in Fig. 6.48*b*) in a rectangular guide at a position where the RF magnetic field is circularly polarized. The ferrite is magnetized by a static field applied in the y direction, as in Fig. 6.48. Since the sense of the circular polarization depends on the direction of propagation, as (6.115) shows, it follows that, for propagation in one direction, the magnetic field is negative circularly polarized and suffers little attenuation, whereas in the reverse direction the field is positive circularly polarized and rapidly attenuated. By proper design the forward loss can be kept to under 0.5 dB/in at $\lambda_0 = 3$ cm, and the reverse loss can be as high as 6 to 10 dB/in or even more. Dielectric loading, as illustrated in Fig. 6.48*c*, gives an improved reverse-to-forward attenuation ratio.

6.10 CIRCULATORS

A circulator is a multiport device that has the property (Fig. 6.49) that a wave incident in port 1 is coupled into port 2 only, a wave incident in port 2 is coupled into port 3 only, and so on. The ideal circulator is also a matched device; i.e., with all ports except one terminated in matched loads, the input impedance of the remaining port is equal to the characteristic impedance of its input line, and hence presents a matched load.

A four-port circulator may be constructed from two magic T's or hybrid junctions and a gyrator as shown in Fig. 6.50. The gyrator produces an additional phase shift of 180° for propagation in the direction from a to b in Fig. 6.50. For propagation from b to a, and also from c to d or d to c, the electrical path lengths are equal

Consider a wave incident in port 1. This wave is split into two equal-amplitude in-phase waves propagating in the side arms of the hybrid

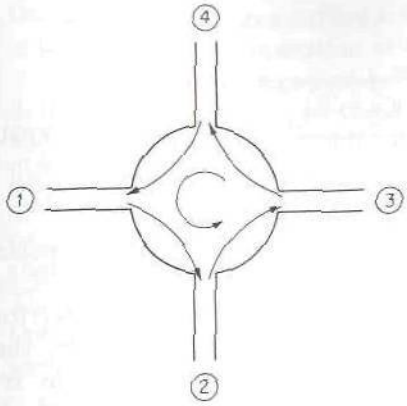


FIGURE 6.49
Schematic diagram for a four-port circulator.

junction. The waves will arrive at points a and c in phase, and hence will emerge from port 2. A wave incident in port 2 will be split into two waves, one arriving at d with a phase ϕ and the other arriving at b with a phase $\phi + \pi$ because of the presence of the gyrator. These partial waves have the right phase relationship to combine and emerge from port 3 in the hybrid junction. A wave incident in port 3 is split into two equal-amplitude waves, differing in phase by 180°, and hence will arrive at the other hybrid junction with the correct phase to combine and emerge from port 4. In a similar manner a wave incident in port 4 will split into two equal waves 180° apart in phase. But now the gyrator will restore phase equality, so that the waves will combine and emerge from port 1. Consequently, the microwave device illustrated in Fig. 6.50 has the required circulating transmission property.

A more compact form of four-port circulator may be constructed by employing 3-dB side-hole directional couplers and rectangular-waveguide nonreciprocal phase shifters. The nonreciprocal phase shifter will be described first. It consists of a thin slab of ferrite placed in a rectangular guide at a point where the ac magnetic field of the TE_{10} mode is circularly polarized, as in Fig. 6.51. A biasing field B_0 is applied in the y direction.

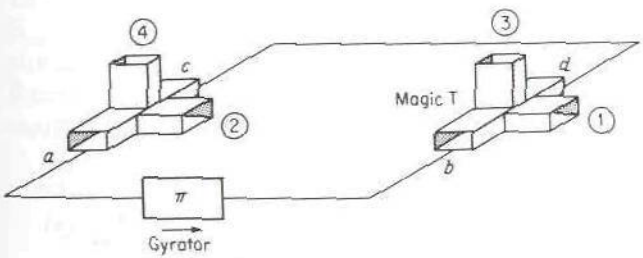


FIGURE 6.50
A four-port circulator.

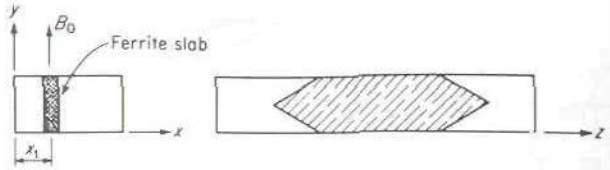


FIGURE 6.51
A nonreciprocal phase shifter.

Since the ac magnetic field is right circularly polarized at x_1 for one direction of propagation and left circularly polarized for the opposite direction of propagation, the perturbing effect of the ferrite slab is different for the two directions of propagation. Consequently, the propagation phase constant β_+ for forward propagation is different from the propagation constant β_- for reverse propagation. By choosing the length of slab so that $(\beta_+ - \beta_-)l = \pi/2$, a differential phase shift of 90° for the two directions of propagation can be achieved.

A four-port circulator utilizing two 90° nonreciprocal phase shifters is illustrated in Fig. 6.52. The phase shifters are biased, with oppositely directed static fields—an arrangement easily achieved in practice with permanent magnets, as shown in Fig. 6.52. One guide is loaded with a dielectric insert to provide an additional 90° of reciprocal phase shift. The coupling holes are arranged to provide 3 dB of coupling. The wave coupled

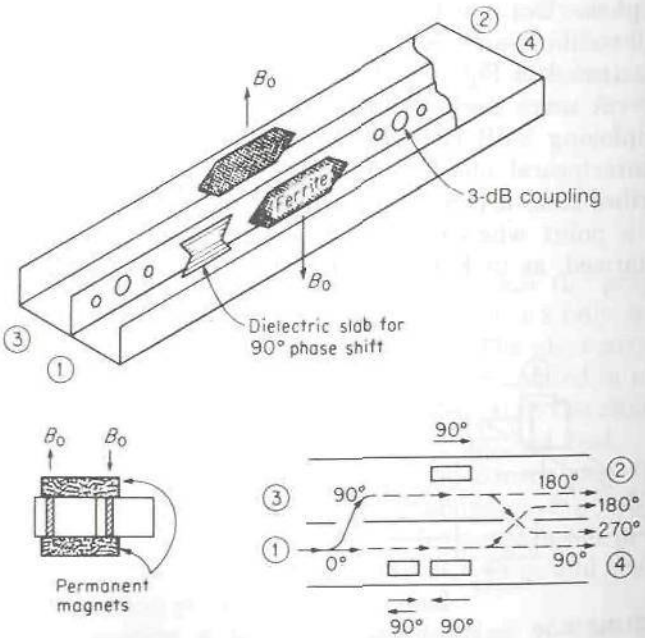


FIGURE 6.52
A compact form of four-port circulator.

through the apertures suffers a 90° change in phase, and this phase change is important in the operation of the circulator.

Consider a wave in port 1. This wave is split into two waves by the first 3-dB coupler, the wave in the upper guide undergoing a 90° phase change because of the transmission properties of an aperture. The wave in the upper guide will arrive at the second coupler with a relative phase of 180°, and the wave in the lower guide with a relative phase of 90°. The second coupler splits these waves in the manner illustrated in Fig. 6.52. It is seen that the resultant waves are out of phase in port 4 but in phase at port 2. Thus transmission is from port 1 to port 2. A similar analysis will verify that a wave incident in port 2 emerges at port 3, or, in general, that the sequence $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$ is followed.

Three-Port Circulator

Carlin has shown that any lossless, matched, nonreciprocal three-port microwave junction is a perfect three-port circulator.† This theorem is readily proved from the properties of the scattering matrix. A perfectly matched three-port junction has a scattering matrix of the form

$$[S] = \begin{bmatrix} 0 & S_{12} & S_{13} \\ S_{21} & 0 & S_{23} \\ S_{31} & S_{32} & 0 \end{bmatrix} \tag{6.116}$$

For a nonreciprocal junction the scattering matrix is no longer symmetrical; that is, $S_{ij} \neq S_{ji}$. However, if the junction is lossless, conservation of power still requires that the $[S]$ matrix be unitary. Thus (4.63*a*) will hold for any lossless microwave junction independently of whether or not the junction is reciprocal. Applying the unitary condition to (6.116) gives

$$\begin{aligned} S_{12}S_{12}^* + S_{13}S_{13}^* &= 1 \\ S_{21}S_{21}^* + S_{23}S_{23}^* &= 1 \\ S_{31}S_{31}^* + S_{32}S_{32}^* &= 1 \\ S_{13}S_{23}^* = S_{12}S_{32}^* &= S_{21}S_{31}^* = 0 \end{aligned}$$

Let us assume that $S_{21} \neq 0$. The fourth of the above equations then gives $S_{31} = 0$. The third equation now requires $|S_{32}| = 1$, and thus $S_{12} = 0$ from the fourth equation, $|S_{13}| = 1$ from the first equation, and $S_{23} = 0$ from the fourth equation again. Thus we see that $|S_{21}| = 1$ also from the second equation, so that

$$\begin{aligned} |S_{21}| = |S_{32}| = |S_{13}| &= 1 \\ S_{12} = S_{23} = S_{31} &= 0 \end{aligned}$$

†H. J. Carlin, Principles of Gyrator Networks, *Polytech. Inst. Brooklyn, Microwave Res. Inst. Symp. Ser.*, vol. 4, p. 175, 1955.

Consequently, there is perfect transmission from port 1 into port 2, from port 2 into port 3, and from port 3 into port 1. There is zero transmission in any other direction. The resultant scattering matrix of any *matched*, *lossless*, *nonreciprocal* three-port junction must then have the form

$$[S] = \begin{bmatrix} 0 & 0 & S_{13} \\ S_{21} & 0 & 0 \\ 0 & S_{32} & 0 \end{bmatrix} \tag{6.117}$$

If the locations of the terminal planes in the three input lines are properly chosen, the phase angles of S_{13}, S_{21}, and S_{32} can be made zero, and then $S_{13} = S_{21} = S_{32} = 1$.

Practical realizations for three-port circulators usually involve the symmetrical junction (Y junction) of three identical waveguides or "strip-line" type of transmission lines, together with an axially magnetized ferrite rod or disks placed at the center. Figure 6.53 illustrates both a waveguide version and a balanced strip-line version of the three-port circulator. The ferrite rod or disks are magnetized by a static B_0 field applied along the axis and give the junction the required nonreciprocal property. By placing

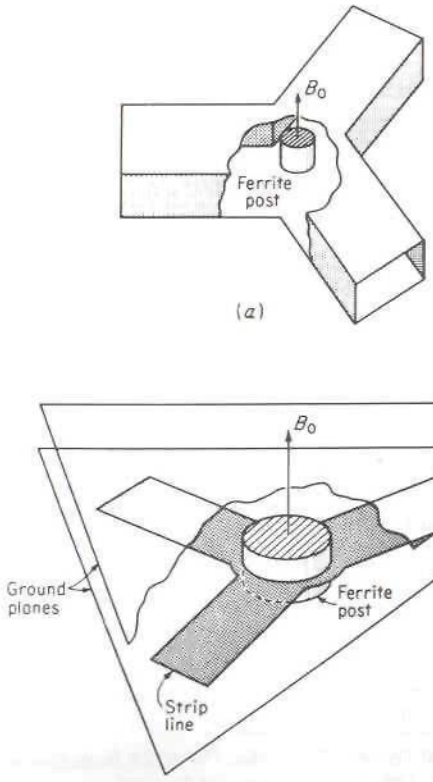


FIGURE 6.53
Three-port circulators. (*a*) Waveguide version; (*b*) strip-line circulator.

suitable tuning elements in each arm (these can be identical in each arm because of the threefold symmetry involved) the junction can be matched; that is, S_{11}, S_{22}, and S_{33} can be made zero. The analysis given above then shows that the junction must, of necessity, be a perfect circulator if all losses are negligible. Losses are, of course, always present, and this limits the performance that can be achieved. Typical characteristics that can be obtained are insertion loss of less than 1 dB, that is, $|S_{13}|, |S_{21}|, |S_{32}|$ greater than 0.89, isolation from 30 to 50 dB, and input reflection coefficients less than 0.2. The isolation that can be obtained corresponds to values of $|S_{31}|$, $|S_{12}|$, and $|S_{23}|$ in the range 0.01 to 0.03.

The junction circulator is an essential component used to separate the input and output ports in negative resistance amplifiers (see Sec. 11.5). Circulators are also used to couple a transmitter and receiver to a common antenna. Circulators ranging from miniature units that can handle a few watts of power up to units that can handle 100 or more kilowatts of average power are commercially available. Bandwidths of more than one octave have been obtained.

Field Analysis of Three-Port Circulator

The field analysis of a three-port circulator is based on the assumption that the electric field in the ferrite disk has only a single component E_z in the axial direction. The applied dc magnetic field is also in the z direction. The permeability tensor then has the same form as in rectangular coordinates; so we can write

$$\begin{bmatrix} B_r \\ B_\phi \end{bmatrix} = \mu_0 \begin{bmatrix} \mu_r & jK \\ -jK & \mu_r \end{bmatrix} \begin{bmatrix} H_r \\ H_\phi \end{bmatrix} \tag{6.118}$$

where $\mu_r = 1 + \chi_{xx}$ and $jK = \chi_{xy}$. In order to obtain a good circulator, the losses in the ferrite must be very small. If we neglect losses

$$\mu_r = \frac{\omega_0^2 - \omega^2 + \omega_0 \omega_m}{\omega_0^2 - \omega^2}$$

$$K = \frac{\omega \omega_m}{\omega_0^2 - \omega^2}$$

Equation (6.118) is easily derived from (6.98) by expressing H_x, H_y, B_x, and B_y in terms of the r and ϕ components of the fields.

From Maxwell's equations we obtain

$$\nabla \times \mathbf{E} = -\mathbf{a}_z \times \nabla E_z = \mathbf{a}_r \frac{1}{r} \frac{\partial E_z}{\partial \phi} - \mathbf{a}_\phi \frac{\partial E_z}{\partial r}$$

$$= -j\omega \bar{\boldsymbol{\mu}} \cdot \mathbf{H} = -j\omega\mu_0 \mathbf{a}_r (\mu_r H_r + jKH_\phi)$$

$$-j\omega\mu_0 \mathbf{a}_\phi (-jKH_r + \mu_r H_\phi) \tag{6.119}$$

We can solve this equation for H_r and H_ϕ and use the results in the equation

$$\nabla \times \mathbf{H} = j\omega\epsilon E_z \mathbf{a}_z$$

to obtain

$$\frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r}\frac{\partial E_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial \phi^2} + k_e^2 E_z = 0 \tag{6.120}$$

where $k_e^2 = \omega^2\mu_0(\mu_r^2 - K^2)\epsilon/\mu_r$. This equation has the same form as (3.217) for the axial electric field for TM modes in a circular waveguide, but with k_e replacing k_c. The general solution for E_z is, by analogy, of the form

$$E_z = \sum_{n=0}^{\infty} (a_n e^{-jn\phi} + b_n e^{jn\phi}) J_n(k_e r) \tag{6.121}$$

where a_n and b_n are amplitude constants and J_n is the Bessel function of order n. The corresponding solution for the ϕ component of the magnetic field is

$$H_\phi = -j\frac{\omega\epsilon}{k_e}\sum_{n=0}^{\infty}\left\{a_n\left[J_n'(k_e r) - \frac{nK}{k_e\mu_r r}J_n(k_e r)\right]e^{-jn\phi} + b_n\left[J_n'(k_e r) + \frac{nk}{k_e\mu_r r}J_n(k_e r)\right]e^{jn\phi}\right\} \tag{6.122}$$

where $J_n'(k_e r) = dJ_n(k_e r)/d(k_e r)$. We can interpret this solution as waves that circulate around the ferrite disk in the $\pm\phi$ directions. The terms multiplied by $a_n e^{-jn\phi}$ are waves circulating in the $+\phi$ direction and the terms multiplied by $b_n e^{jn\phi}$ are waves circulating in the $-\phi$ direction. Since H_ϕ is different for the two sets of waves, the ferrite disk clearly will exhibit nonreciprocal properties, an essential requirement for a circulator.

Consider now the circulator configuration shown in Fig. 6.54. Each of the three microstrip input ports are 120° apart and extend over a coupling angle ψ. In the field analysis of the circulator, it is assumed that H_ϕ is a constant over each coupling region and is zero along the remainder of the boundary of the ferrite disk. Thus at $r = a$ we let

$$H_\phi = \begin{cases} H_1 & -\dfrac{\psi}{2} \le \phi \le \dfrac{\psi}{2} \\ H_2 & \dfrac{2\pi}{3} - \dfrac{\psi}{2} \le \phi \le \dfrac{2\pi}{3} + \dfrac{\psi}{2} \\ H_3 & \dfrac{4\pi}{3} - \dfrac{\psi}{2} \le \phi \le \dfrac{4\pi}{3} + \dfrac{\psi}{2} \end{cases} \tag{6.123}$$

The assumption of a constant magnetic field H_ϕ over each coupling region is a first-order approximation to the magnetic field of the TEM mode that exists in each port. At $r = a$ the solution for H_ϕ given by (6.122) is a

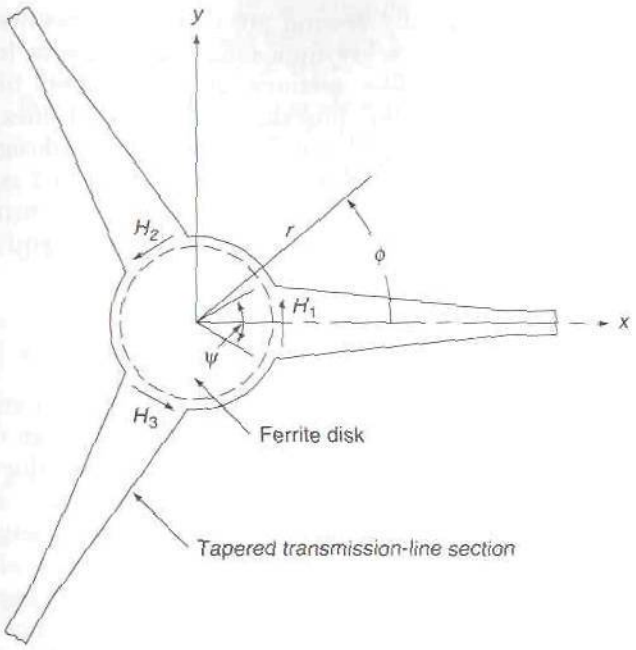


FIGURE 6.54
Three-port circulator showing coupling angle ψ and tapered transmission lines for matching.

Fourier series; so the amplitude coefficients a_n and b_n are readily found by equating (6.122) to (6.123). When the coefficients a_n and b_n have been found by Fourier analysis, the electric field given by (6.121) can be evaluated at $r = a$.

In order to obtain a circulator, the condition $E_z = 0$ at $\phi = 4\pi/3$ is imposed so that there will be no coupling into port 3. This condition establishes a constraint on the ferrite disk radius a, the magnetization M_s in the ferrite, and the coupling angle ψ. In practice, these parameters are optimized so as to obtain the best circulator performance possible, over as broad a band of frequencies as possible. It is found that the optimum coupling angle ψ is given approximately by†

$$\psi = \frac{2\pi}{1.84\sqrt{3}} \sqrt{\frac{\epsilon}{\mu_r \epsilon_0}} \frac{K}{\mu_r} \tag{6.124}$$

†H. Bosma, On Stripline Y-Circulators at UHF, *IRE Trans.*, vol. MTT-12, pp. 61–72, 1964.
Y. S. Wu and F. J. Rosenbaum, Wide-Band Operation of Microstrip Circulators, *IEEE Trans.*, vol. MTT-22, pp. 849–856, 1974.

This angle is typically around 30° to 50° and results in a strip-line width that corresponds to a low-impedance transmission line. Consequently, tapered transmission-line sections are often used to transform this low impedance into the 50-Ω impedance of the input lines. The tapered sections are illustrated in Fig. 6.54. Satisfactory circulator designs have been achieved by taking as few as 3 to 6 modes into account, that is, n up to 6.

In the above account we have outlined the principal features of the field analysis of circulators. The references at the end of the chapter provide many of the details that we have omitted.

6.11 OTHER FERRITE DEVICES

The devices utilizing ferrites for their operation described in the preceding sections represent only a small number of the large variety of devices that have been developed. In addition to the above, there are other forms of isolators, both reciprocal and nonreciprocal phase shifters, electronically controlled (by varying the current in the electromagnet that supplies the static biasing field) phase shifters and modulators, electronic switches and power limiters, etc. The nonlinear property of ferrites for high signal levels has also been used in harmonic generators, frequency mixers, and parametric amplifiers. A discussion of these devices, together with design considerations, performance data, and references to the original literature, is contained in the book by Lax and Button, listed in the references at the end of this chapter. The recent article by Rodrique gives a good survey of the present status of ferrite devices.†

PROBLEMS

6.1. Determine the values of R_1 and R_2 in the T and Π attenuator networks shown in Fig. 6.8 in order to obtain 6 dB of attenuation.

6.2. Derive the equations (6.7) for the Π attenuator network.

6.3. For the electronically controlled attenuator shown in the photograph in Fig. 6.10, the return loss is 10 dB. Calculate the input VSWR (see Sec. 5.5).

6.4. In the rotary phase shifter, show that if the output quarter-wave plate, transition, and rectangular guide are rotated by an angle θ_1, an additional phase change of θ_1 is produced in the transmitted wave.

6.5. For the phase shifter shown in Fig. 6.15, determine the length (in terms of wavelength) for the two short-circuited stubs in order to obtain an incremental phase change of 45°.

6.6. For the phase shifter shown in Fig. 6.16, assume that the *PIN* diodes act like ideal switches. When the diodes are on, the susceptance placed in shunt on the

†G. P. Rodrique, A Generation of Microwave Ferrite Devices, *Proc. IEEE*, vol. 76, pp. 121–137, 1988.

main line is $j\bar{B}_1 = -j \cot \beta d_1$. When the diodes are off, the susceptance loading the main line is $j\bar{B}_2 = j \tan \beta d$. Show that by choosing $d_1 = \lambda_0/4 + \delta$ and $d = \lambda_0/2 - \delta$, $\bar{B}_1 = -\bar{B}_2$.

6.7. For the phase shifter shown in Fig. 6.15, the transmission coefficient $T_{14} = -j[1 + j\bar{B} - \bar{B}^2/2]^{-1}$, where $j\bar{B}$ is the susceptance of the stub. By using $1 - \rho^2 = |T_{14}|^2$, find the magnitude ρ of the reflection coefficient and the input VSWR when $\bar{B} = 0.2$ and when $\bar{B} = 0.4$.

Answer: VSWR = 1.04, 1.173

6.8. For a Bethe-hole directional coupler with the two guides aligned ($\theta = 0$) and a centered aperture, why does not (6.28) give a useful solution for λ_0 as a function of a whereas (6.29) does?

6.9. Design a Bethe-hole directional coupler with a centered circular aperture. The waveguide size is $a = 0.9$ in, $b = 0.4$ in. The center frequency is 9.8 GHz. The required coupling is 30 dB. Find the aperture radius and the frequency band over which the directivity remains greater than 20 dB.

6.10. Design a Bethe-hole coupler based on (6.29) and (6.30). Assume $a = 0.9$ in, $b = 0.4$ in, $f = 9$ GHz, and $C = 30$ dB. Find the aperture position d and radius r. Evaluate C and D as a function of frequency over the band 8.5 to 10.5 GHz and compare the performance with that shown in Fig. 6.20.

6.11. Figure P6.11 illustrates two rectangular guides coupled by circular apertures in a common side wall. A TE_{10} mode of unit amplitude radiates a field of amplitude $-j\frac{4}{3}r_0^3(\pi/a)^2(ab\beta)^{-1}$ in both directions in the other guide. Design a five-hole directional coupler of the binomial type. The coupling required is 30 dB at a frequency of 10 GHz. The guide width a is 2.5 cm, and the height $b = 1.2$ cm. Find the required aperture radii and the frequency band over which the directivity D remains greater than 40 dB.

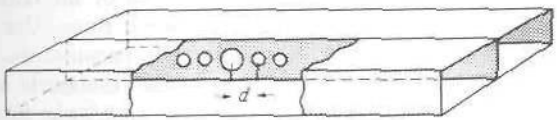


FIGURE P6.11

6.12. For the coupler described in Prob. 6.11, find the aperture radii to give a Chebyshev coupler. The minimum value of directivity required is $D_m = 50$ dB. Find the corresponding frequency band. How much greater bandwidth is obtained as compared with that of the binomial coupler of Prob. 6.11?

6.13. Design a three-hole Chebyshev directional coupler using rectangular waveguides with $a = 0.9$ in, $b = 0.4$ in. The center frequency is 9 GHz. The coupling and minimum directivity D_m contributed by the array factor is 20 dB and 30 dB at 9 GHz. The apertures are located at $x_0 = a/4$. Plot C and D as a function of frequency and compare the performance with that shown in Fig. 6.24.

6.14. Repeat the design problem given in Prob. 6.13 but instead of using $D_m = 30$ dB, the requirement is that the fractional bandwidth $\Delta\beta/\beta_0 = 0.1$, where $\beta_0 = 1.29$ at 9 GHz and $\Delta\beta = \beta_2 - \beta_1$ with $\beta_2 = \beta_0 + 0.05\beta_0$ and $\beta_1 = \beta_0 - 0.05\beta_0$. What is the resultant value of D_m for this design?

6.15. Design a three-hole Chebyshev directional coupler using centered apertures in the common broad wall between two rectangular waveguides (Fig. 6.23). The center frequency is 9.5 GHz. The waveguide width equals 0.9 in and the height equals 0.4 in. The required coupling is 30 dB and the minimum directivity is 30 dB. Find the aperture radii, spacing, and the bandwidth of the coupler. Why is this not a good design?

6.16. Design a broadside-coupled strip-line directional coupler with 3-dB coupling. See Figs. 6.27 and P6.16 for details. The ground-plane spacing is 2 cm. The input and output lines have a characteristic impedance of 50 Ω. Find the required strip width W and spacing S. Find W so as to obtain $Z_c = 50\ \Omega$ for the input and output lines. For this calculation assume that the line is centered between the ground planes. In the actual coupler the lower and upper lines are not centered, but since the spacing S is small the change in Z_c will also be small. The computer program CSTPL should be used to obtain the needed line parameters. Find the length of the coupled section when the frequency is 4 GHz. Assume that the strips have a negligible thickness.

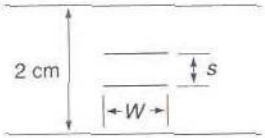


FIGURE P6.16

6.17. A branch-line coupler of the type shown in Fig. 6.29 and having 6-dB coupling at a frequency of 5 GHz is needed. The input and output transmission lines have $Z_c = 50\ \Omega$. Determine the characteristic impedances of the through lines and branch lines and their lengths in terms of wavelength. The coupler is to be built using microstrip lines on a substrate with a dielectric constant of 6 and 1 mm thick. Determine the widths of all transmission lines and the lengths of the through lines and branch lines. Use the computer program MSTP to determine the microstrip line parameters. Use the computer program MSTPD to determine the effective dielectric constants at 5 GHz and modify the transmission-line lengths so as to take into account the dispersion in the effective dielectric constant at 5 GHz.

6.18. The hybrid ring shown in Fig. 6.33 is constructed using microstrip lines on a 1-mm-thick substrate with a dielectric constant of 2.3. The input and output transmission lines have a characteristic impedance of 50 Ω. Determine the width of the transmission lines and of the ring as well as the radius of the ring. The frequency of operation is 2 GHz. Use the computer program MSTP.

6.19. Determine the impedances and the value of the resistor R for the power divider shown in Fig. 6.38. The power in port 2 is one-half of that delivered to port 3.

6.20. Use the equation of motion (6.94) to study second-harmonic generation in a ferrite. Assume that $\mathbf{M} = \mathbf{M}_1 e^{j\omega t} + \mathbf{M}_2 e^{2j\omega t}$ for the ac magnetization and that $\mathbf{H}$ has only an x component with time dependence $e^{j\omega t}$; that is, the ac magnetic field is $H_x e^{j\omega t}\mathbf{a}_x$. Neglect the third-harmonic term that occurs and show that

$$2j\omega\mathbf{M}_2 = \gamma\mu_0 H_x \mathbf{a}_x \times \mathbf{M}_1 + \gamma\mu_0 H_0 \mathbf{a}_z \times \mathbf{M}_2$$

Thus $2j\omega M_{2z} = \gamma\mu_0 H_x M_{1y}$. For M_{1y}, take the small-signal solution $\chi_{yx} H_x$ and use the value of χ_{yx} at resonance for a lossy ferrite to show that

$$M_{2z} = \frac{j\gamma\mu_0\omega_m H_x^2}{4\omega\omega_0\alpha}$$

Note that, for good efficiency, α must be small (small damping), so that the precession angle will be large at resonance.

6.21. From the unitary properties of the scattering matrix for a lossless nonreciprocal two-port microwave junction, show that it is not possible to have S_{21} zero while S_{12} is finite. Thus a lossless one-way transmission device cannot be built.

6.22. Continue the argument in the text to verify that the transmission sequence $1 \to 2 \to 3 \to 4 \to 1$ is followed in the circulator illustrated in Fig. 6.52.

6.23. Show that the scattering matrix for an ideal lossless N-port circulator can be put into the form

$$[S] = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

by choosing proper terminal-plane locations in each port.

6.24. Show that, for TEM-wave propagation in a direction perpendicular to $\mathbf{B}_0$ in an infinite ferrite medium, the two solutions are linearly polarized waves with propagation constants

$$\gamma_+ = j\omega\sqrt{\mu_0\epsilon}\left[\frac{(1+\chi)^2 - K^2}{1+\chi}\right]^{1/2}$$

$$\gamma_- = j\omega\sqrt{\mu_0\epsilon}$$

Hint: Consider propagation along x, and in one case assume $\mathbf{E}$ to have only a y component, and for the other case assume $\mathbf{E}$ to have a z component.

REFERENCES

1. Montgomery, C. G., R. H. Dicke, and E. M. Purcell: "Principles of Microwave Circuits," McGraw-Hill Book Company, New York, 1948.
2. Ragan, G. L. (ed.): "Microwave Transmission Circuits," McGraw-Hill Book Company, New York, 1948.
3. Marcuvitz, N. (ed.): "Waveguide Handbook," McGraw-Hill Book Company, New York, 1951.
4. Harvey, A. F.: "Microwave Engineering," Academic Press, Inc., New York, 1963. An outstanding handbook, offering a comprehensive survey of the whole microwave field together with an extensive bibliography covering the international literature.
5. Bahl, I., and P. Bhartia: "Microwave Circuit Design," John Wiley & Sons, Inc., New York 1988.
6. Chang, K. (ed.): "Handbook of Microwave and Optical Components, Microwave Passiv and Antenna Components," vol. 1, John Wiley & Sons, Inc., New York, 1989.
7. D. M. Pozar, "Microwave Engineering," Addison-Wesley Publishing Company, Reading Mass., 1990.

Microwave ferrites

8. Lax, B., and K. J. Button: "Microwave Ferrites and Ferrimagnetics," McGraw-Hill Book Company, New York, 1962. A very complete treatment of the physical properties of ferrites, electromagnetic propagation in ferrites, measurement techniques, and a discussion of many ferrite devices. Includes a 24-page bibliography.
9. Soohoo, R. F.: "Theory and Application of Ferrites," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1960.
10. Clarricoats, P: "Microwave Ferrites," John Wiley & Sons, Inc., New York, 1961. Gurevich, A. G.: "Ferrites at Microwave Frequencies," transl. from the Russian by A. Tybulewicz, Consultants Bureau, New York, 1963.
11. Roberts, J.: "High Frequency Application of Ferrites," D. Van Nostrand Company, Inc., Princeton, N.J., 1960.

Field theory of circulators

12. Bosma, H.: On Stripline Y-Circulators at UHF, *IRE Trans.*, vol. MTT-12, pp. 61–72, 1964.
13. Fay, C. E., and R. L. Comstock: Operation of the Ferrite Junction Circulator, *IEEE Trans.*, vol. MTT-13, pp. 15–27, 1965.
14. Wu, Y. S., and F. J. Rosenbaum: Wide-Band Operation of Microstrip Circulators, *IEEE Trans.*, vol. MTT-22, pp. 849–856, 1974.
15. Schloemann, E., and R. E. Blight: Broad-Band Stripline Circulators Based on YIG and Li-Ferrite Single Crystals, *IEEE Trans.*, vol. MTT-34, pp. 1394–1400, 1986.

CHAPTER 7

ELECTROMAGNETIC RESONATORS

7.1 RESONANT CIRCUITS

Resonant circuits are of great importance for oscillator circuits, tuned amplifiers, frequency filter networks, wavemeters for measuring frequency, etc., at all frequencies from a few hertz up to and including light frequencies. Electric resonant circuits have many features in common, and it will be worthwhile to review some of these by using a conventional lumped-parameter *RLC* parallel network as an example. Figure 7.1 illustrates a typical low-frequency resonant circuit. The resistance R is usually only an equivalent resistance that accounts for the power loss in the inductor L and capacitor C and possibly the power extracted from the resonant system by some external load coupled to the resonant circuit.

At resonance the input impedance is pure real and equal to R. This implies that the average energies stored in the electric and magnetic fields are equal, since from (2.60)

$$Z_{\text{in}} = \frac{P_l + 2j\omega(W_m - W_e)}{\frac{1}{2}II^*} \tag{7.1}$$

This equation is valid for any one-port circuit provided a suitably defined equivalent terminal current I is used. Thus resonance always occurs when $W_m = W_e$, if we define resonance to be that condition which corresponds to a pure resistive input impedance. In the present case the time-average energy stored in the electric field in the capacitor is

$$W_e = \tfrac{1}{4}VV^*C$$

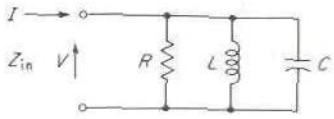


FIGURE 7.1
Lumped-parameter resonant circuit.

and that stored in the magnetic field around the inductor is

$$W_m = \frac{1}{4}LI_L I_L^* = \frac{1}{4}L\left|\frac{V}{\omega L}\right|^2 = \frac{1}{4\omega^2 L}VV^*$$

The resonant frequency ω_0 is now found by equating W_m and W_e:

$$\omega_0 = (LC)^{-1/2} \tag{7.2}$$

An important parameter specifying the frequency selectivity, and performance in general, of a resonant circuit is the quality factor, or Q. A very general definition of Q that is applicable to all resonant systems is

$$Q = \frac{\omega(\text{time-average energy stored in system})}{\text{energy loss per second in system}} \tag{7.3}$$

At resonance $W_m = W_e$ and since the peak value of electric energy stored in the capacitor is $2W_e$ and occurs when the energy stored in the inductor is zero, and vice versa, the average energy W stored in the circuit is

$$W = W_m + W_e = 2W_m = 2W_e = \tfrac{1}{2}CVV^* \tag{7.4}$$

The power loss is $\frac{1}{2}GVV^*$ and is the energy loss per second. Hence, for the circuit of Fig. 7.1,

$$Q = \frac{\omega C}{G} = \omega RC = \frac{R}{\omega L} \tag{7.5}$$

since $\omega^2 LC = 1$ at resonance and $G = R^{-1}$.

In the vicinity of resonance, say $\omega = \omega_0 + \Delta\omega$, the input impedance can be expressed in a relatively simple form. We have

$$Z_{\text{in}} = \left(\frac{1}{R} + \frac{1}{j\omega L} + j\omega C\right)^{-1} = \left(\frac{1}{R} + j\omega_0 C + j\,\Delta\omega C + \frac{1 - \Delta\omega/\omega_0}{j\omega_0 L}\right)^{-1}$$

where the approximation $1/(\omega_0 + \Delta\omega) \approx (1 - \Delta\omega/\omega_0)/\omega_0$ has been used. Since $j\omega_0 C + 1/j\omega_0 L = 0$, we obtain

$$Z_{\text{in}} = \frac{\omega_0^2 RL}{\omega_0^2 L + j2R\,\Delta\omega} = \frac{R}{1 + j2Q(\Delta\omega/\omega_0)} \tag{7.6}$$

A plot of Z_{in} as a function of $\Delta\omega/\omega_0$ is given in Fig. 7.2, and is a typical resonance curve. When $|Z_{\text{in}}|$ has fallen to 0.707 of its maximum value, its phase is 45° if $\omega < \omega_0$ and $-45°$ if $\omega > \omega_0$. From (7.6) the corresponding

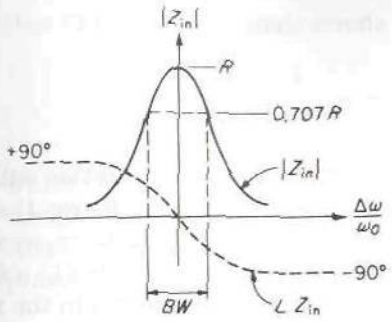


FIGURE 7.2
Z_{in} for a parallel resonant circuit.

value of $\Delta\omega$ is found to be given by

$$2Q\frac{\Delta\omega}{\omega_0} = 1$$

or

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

The fractional bandwidth BW between the $0.707R$ points is twice this; hence

$$Q = \frac{\omega_0}{2\,\Delta\omega} = \frac{1}{\text{BW}} \tag{7.7}$$

This relation provides an alternative definition of the Q; that is, the Q is equal to the fractional bandwidth between the points where $|Z_{in}|$ is equal to 0.707 of its maximum value (for a series resonant circuit this definition applies to $|Y_{in}|$).

If the resistor R in Fig. 7.1 represents the losses in the resonant *circuit only, the Q given by (7.5) is called the unloaded Q.* If *the resonant* circuit is coupled to an external load that absorbs a certain amount of power, this loading effect can be represented by an additional resistor R_L in parallel with R. The total resistance is now less, and consequently the new Q is also smaller. The Q, called the loaded Q and denoted Q_L, is

$$Q_L = \frac{RR_L/(R + R_L)}{\omega L}$$

The external Q, denoted Q_e, is defined to be the Q that would result if the resonant circuit were loss-free and only the loading by the external load were present. Thus

$$Q_e = \frac{R_L}{\omega L}$$

Use of these definitions shows that

$$\frac{1}{Q_L} = \frac{1}{Q_e} + \frac{1}{Q} \tag{7.8}$$

Another parameter of importance in connection with a resonant circuit is the damping factor δ. This parameter measures the rate at which the oscillations would decay if the driving source were removed. For a high-Q circuit, δ may be evaluated in terms of the Q, using a perturbation technique. With losses present, the energy stored in the resonant circuit will decay at a rate proportional to the average energy present at any time (since $P_l \propto VV^*$ and $W \propto VV^*$, we have $P_l \propto W$), so that

$$\frac{dW}{dt} = -2\delta W \tag{7.9}$$

or

$$W = W_0 e^{-2\delta t} \tag{7.10}$$

where W_0 is the average energy present at $t = 0$. But the rate of decrease of W must equal the power loss, so that

$$-\frac{dW}{dt} = 2\delta W = P_l$$

Consequently,

$$\delta = \frac{P_l}{2W} = \frac{\omega}{2}\frac{P_l}{\omega W} = \frac{\omega}{2Q} \tag{7.11}$$

upon using (7.3). The damping factor is seen to be inversely proportional to the Q. In place of (7.10) we now have

$$W = W_0 e^{-\omega t/Q} \tag{7.12}$$

In (7.12) Q must be replaced by Q_L if an external load is coupled to the circuit. The damping factor δ is also a measure of how fast the amplitude of oscillations in the resonant circuit can build up upon application of a driving source.

In microwave systems sections of transmission lines or metallic enclosures called cavities are used as resonators in place of the lumped-parameter circuit. The reason for this is that lumped-parameter circuits have too high losses, from both conductor loss and radiation loss, to be effective at microwave frequencies. In calculating the impedance of a microwave cavity, it is sometimes convenient to assume there are no losses present. The Q can be evaluated separately, and in terms of this parameter the impedance can be modified to account for losses by replacing the resonant frequency ω_0 by an equivalent complex resonant frequency $\omega_0(1 + j/2Q) = \omega_0 + j\delta$. Note

that (7.6) can be written as

$$Z_{\text{in}} = \frac{\omega_0 R/2Q}{j[\omega - \omega_0(1 + j/2Q)]} \tag{7.13}$$

which shows that when losses are present this is equivalent to having a complex resonant frequency $\omega_0(1 + j/2Q)$. Equation (7.13) neglects the small change in resonant frequency that occurs when small losses are present.

7.2 TRANSMISSION-LINE RESONANT CIRCUITS

Series Resonance; Short-Circuited Line

At high frequencies, usually in the range 100 to 1,000 MHz, short-circuited or open-circuited sections of transmission lines are commonly used to replace the usual lumped LC resonant circuit. It is therefore of interest to consider the order of magnitude of Q and impedance that can be obtained. It will be assumed that air-filled lines are used, so that the only losses are those due to the series resistance R of the line. This is usually the case in practice since a dielectric-filled line has some shunt conductance loss and hence would result in a lower Q.

Consider a short-circuited line of length l, parameters R, L, C per unit length, as in Fig. 7.3. Let $l = \lambda_0/2$ at $f = f_0$, that is, at $\omega = \omega_0$. For f near f_0, say $f = f_0 + \Delta f$, $\beta l = 2\pi fl/c = \pi\omega/\omega_0 = \pi + \pi\,\Delta\omega/\omega_0$, since at ω_0, $\beta l = \pi$. The input impedance is given by

$$Z_{\text{in}} = Z_c \tanh(j\beta l + \alpha l) = Z_c \frac{\tanh \alpha l + j \tan \beta l}{1 + j \tan \beta l \tanh \alpha l}$$

But $\tanh \alpha l \approx \alpha l$ since we are assuming small losses, so that $\alpha l \ll 1$. Also $\tan \beta l = \tan(\pi + \pi\,\Delta\omega/\omega_0) = \tan \pi\,\Delta\omega/\omega_0 \approx \pi\,\Delta\omega/\omega_0$ since $\Delta\omega/\omega_0$ is

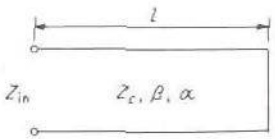


FIGURE 7.3
Short-circuited transmission-line resonator.

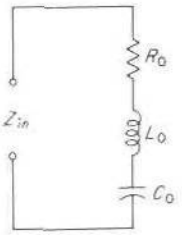


FIGURE 7.4
A series resonant circuit.

small. Hence

$$Z_{in} = Z_c \frac{\alpha l + j\pi\,\Delta\omega/\omega_0}{1 + j\alpha l\,\pi\,\Delta\omega/\omega_0} \approx Z_c\left(\alpha l + j\pi\frac{\Delta\omega}{\omega_0}\right) \tag{7.14}$$

since the second term in the denominator is very small. Now

$$Z_c = \sqrt{\frac{L}{C}}$$

$\alpha = \frac{1}{2}RY_c = (R/2)\sqrt{C/L}$, and $\beta l = \omega_0\sqrt{LC}\,l = \pi$; so $\pi/\omega_0 = l\sqrt{LC}$, and the expression for Z_{in} becomes

$$Z_{in} = \sqrt{\frac{L}{C}}\left(\frac{l}{2}R\sqrt{\frac{C}{L}} + j\,\Delta\omega\, l\sqrt{LC}\right) = \frac{1}{2}Rl + jlL\,\Delta\omega \tag{7.15}$$

It is of interest to compare (7.15) with a series $R_0L_0C_0$ circuit illustrated in Fig. 7.4. For this circuit

$$Z_{in} = R_0 + j\omega L_0\left(1 - \frac{1}{\omega^2 L_0 C_0}\right)$$

If we let $\omega_0^2 = 1/L_0C_0$, then

$$Z_{in} = R_0 + j\omega L_0\frac{\omega^2 - \omega_0^2}{\omega^2}$$

Now $\omega^2 - \omega_0^2 = (\omega - \omega_0)(\omega + \omega_0) \approx \Delta\omega\, 2\omega$ if $\omega - \omega_0 = \Delta\omega$ is small. Thus

$$Z_{in} = R_0 + j\omega L_0\frac{2\omega\,\Delta\omega}{\omega^2} = R_0 + j2L_0\,\Delta\omega \tag{7.16}$$

By comparison with (7.15), we see that in the vicinity of the frequency for which $l = \lambda_0/2$, the short-circuited line behaves as a series resonant circuit with resistance $R_0 = \frac{1}{2}Rl$ and inductance $L_0 = \frac{1}{2}Ll$. We note that Rl, Ll are the total resistance and inductance of the line; so we might wonder why the factors $\frac{1}{2}$ arise. These enter because the current on the short-circuited line is a half sinusoid, and hence the effective circuit parameters R_0, L_0 are only one-half of the total line quantities.

The Q of the short-circuited line may be defined as for the circuit of Fig. 7.4:

$$Q = \frac{\omega_0 L_0}{R_0} = \frac{\omega_0 L}{R} = \frac{\beta}{2\alpha} \tag{7.17}$$

As an alternative, the general definition (7.3) may be used. We shall evaluate the Q of the short-circuited line from this definition by means of

an approximate method valid for high-Q (i.e., low-loss) systems in order to illustrate a method of great utility in connection with many microwave devices (see Sec. 3.8, where this perturbation method is discussed). For small losses the energy stored in the system is, to first order, the same as if no losses were present. For a loss-free short-circuited line, the current on the line is a pure standing wave

$$I = I_0 \cos \beta z\, e^{j\omega t}$$

where z measures distance along the line from the short toward the input end. In a length dz, the energy stored in the magnetic field is

$$dW_m = \tfrac{1}{4} II^* L\, dz = \tfrac{1}{4} I_0^2 L \sin^2 \beta z\, dz$$

The total time-average stored energy in the magnetic field is

$$W_m = \frac{1}{4} I_o^2 L \int_0^{\lambda_0/2} \cos^2 \beta z\, dz = \frac{\lambda_0}{16} I_0^2 L$$

The energy stored in the electric field, i.e., in the line capacitance, is equal to W_m at resonance; so the time-average energy stored in the system is

$$W = W_m + W_e = \frac{\lambda_0}{8} L I_0^2$$

To a first approximation the losses do not modify the current distribution along the line. Hence the power loss is given by

$$P = \frac{1}{2} \int_0^{\lambda_0/2} RII^*\, dz = \frac{R}{2} I_0^2 \int_0^{\lambda_0/2} \cos^2 \beta z\, dz = \frac{\lambda_0}{8} R I_0^2$$

Thus, at $\omega = \omega_0$,

$$Q = \frac{\omega_0 W}{P} = \frac{\omega_0 \lambda_0 L I_0^2/8}{R I_0^2 \lambda_0/8} = \frac{\omega_0 L}{R} \tag{7.18}$$

which checks with the earlier result. Typical values of Q range from several hundred up to about 10,000. As contrasted with low-frequency lumped-parameter circuits, the practical values of Q are very much higher for microwave resonators. It should be noted that in the above analysis the losses in the short circuit have been neglected. This does not introduce appreciable error if the length l is considerably greater than the conductor spacing.

Open-Circuited Line

By means of an analysis similar to that used earlier, it is readily verified that an open-circuited transmission line is equivalent to a series resonant circuit in the vicinity of the frequency for which it is an odd multiple of a

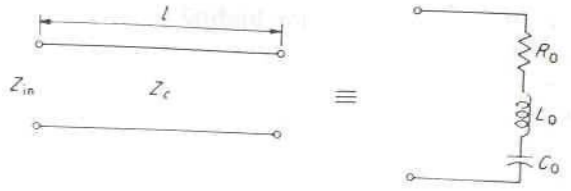


FIGURE 7.5
Open-circuited transmission-line resonator.

quarter wavelength long. The equivalent relations are (Fig. 7.5)

$$l = \frac{\lambda_0}{4} \qquad \text{at } \omega_0 \tag{7.19a}$$

$$Z_{\text{in}} \approx \left(\alpha l + j\frac{\Delta\omega}{\omega_0}\frac{\pi}{2}\right)Z_c = \frac{1}{2}Rl + j\,\Delta\omega\, Ll \tag{7.19b}$$

$$R_0 = \frac{1}{2}Rl \tag{7.19c}$$

$$L_0 = \frac{1}{2}Ll \tag{7.19d}$$

$$\omega_0^2 = (L_0C_0)^{-1} \tag{7.19e}$$

Antiresonance

Short-circuited transmission lines behave as parallel resonant circuits in the frequency range where they are close to an odd multiple of a quarter wavelength long. The same property is true of open-circuited lines that are a multiple of a half wavelength long. When they behave as parallel resonant circuits, they are said to be antiresonant.

The case of parallel resonance is best analyzed on an admittance basis. With reference to Fig. 7.6, let l equal $\lambda_0/4$ at $\omega = \omega_0$. Then

$$\beta l = \omega\sqrt{LC}\,l = \omega_0\sqrt{LC}\,l + \Delta\omega\sqrt{LC}\,l$$

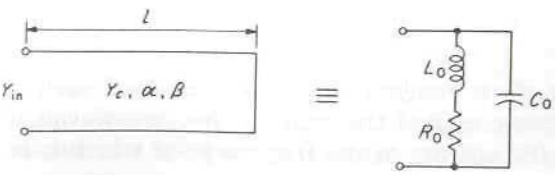


FIGURE 7.6
Short-circuited antiresonant transmission line.

at ω, and the input admittance is given by

$$\begin{aligned} Y_{\text{in}} &= Y_c \coth(\alpha + j\beta)l \\ &= Y_c \frac{1 + j \tan \beta l \tanh \alpha l}{\tanh \alpha l + j \tan \beta l} \\ &\approx Y_c \frac{1 - j\alpha l/(\Delta\omega\, l\sqrt{LC})}{\alpha l - j/(\Delta\omega\, l\sqrt{LC})} \end{aligned}$$

since

$$\begin{aligned} \tan\left(\omega_0\sqrt{LC}\,l + \Delta\omega\sqrt{LC}\,l\right) &= \tan\left(\frac{\pi}{2} + \Delta\omega\sqrt{LC}\,l\right) \\ &\approx -\left(\Delta\omega\sqrt{LC}\,l\right)^{-1} \end{aligned}$$

and $\tanh \alpha l \approx \alpha l$. A further approximation yields

$$\begin{aligned} Y_{\text{in}} &= Y_c \frac{j\,\Delta\omega\sqrt{LC}\,l + \alpha l}{1 + j\,\Delta\omega\,\alpha l^2\sqrt{LC}} \\ &\approx Y_c\left(j\,\Delta\omega\sqrt{LC}\,l + \alpha l\right) \\ &= \frac{RC}{2L} l + j\,\Delta\omega\, Cl \end{aligned} \tag{7.20}$$

after replacing Y_c by $\sqrt{C/L}$ and α by $Y_c R/2$.

For the parallel resonant circuit of Fig. 7.6, we have

$$\begin{aligned} Y_{\text{in}} &\approx \frac{1}{R_0 + j\omega L_0} + j\omega C_0 \\ &\approx \frac{j\omega C_0(R_0 + j\omega L_0) + 1}{R_0 + j\omega L_0} \\ &\approx \frac{j\omega C_0 R_0 - \omega^2 L_0 C_0 + 1}{j\omega L_0} \end{aligned}$$

since we assume $R_0 \ll \omega L_0$. If we define ω_0 by $\omega_0^2 L_0 C_0 = 1$, then

$$\begin{aligned} Y_{\text{in}} &= \frac{C_0 R_0}{L_0} + \frac{\omega_0^2 L_0 C_0 - \omega^2 L_0 C_0}{j\omega L_0} \\ &= R_0 \frac{C_0}{L_0} - jL_0 C_0 \frac{(\omega_0 - \omega)(\omega_0 + \omega)}{\omega L_0} \\ &\approx R_0 \frac{C_o}{L_0} + jC_0 2\,\Delta\omega \end{aligned} \tag{7.21}$$

Comparison with (7.20) shows that the short-circuited line in the vicinity of a quarter wavelength long is equivalent to a parallel resonance circuit with

$$\frac{C_0 R_0}{L_0} = \frac{RC}{2L} l \qquad Cl = 2C_0 \qquad \frac{R_0}{L_0} = \frac{R}{L}$$

The Q of the circuit is given by

$$Q = \frac{\omega_0 L_0}{R_0} = \frac{\omega L}{R} = \frac{\beta}{2\alpha} \tag{7.22}$$

Although sections of transmission lines behave as simple lumped-parameter resonant circuits in the vicinity of a particular resonant frequency, they are in reality a much more complicated network having an infinite number of resonance and antiresonance frequencies. The resonance frequencies occur approximately when the short-circuited line is a multiple of a half wavelength long, that is, $f_n = nc/2l$, and the antiresonance frequencies (parallel resonance) when the line is an odd multiple of a quarter wavelength long, that is, $f_n = (2n + 1)c/4l$, where n is an integer. Thus the exact equivalent circuit would consist of an infinite number of resonant circuits coupled together. However, in practice, the frequency range of interest is normally such that a simple single-resonant-frequency circuit represents the transmission-line resonator with adequate accuracy.

7.3 MICROSTRIP RESONATORS

In microwave circuits that use microstrip transmission lines, an open-circuited section of microstrip line may be used as a simple resonator. In Fig. 7.7*a* we show a microstrip resonator that is capacitively coupled to the input microstrip line. In Fig. 7.7*b* we show a similar resonator that is capacitively coupled to the side of the microstrip line. An alternative resonator configu-

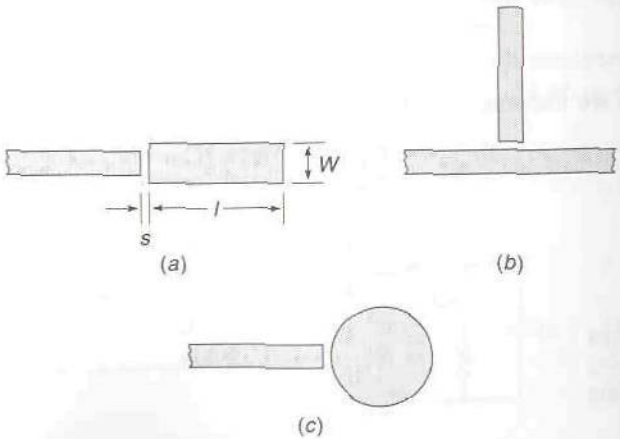


FIGURE 7.7
(*a*) A microstrip resonator consisting of an open-circuited microstrip line approximately one-half wavelength long and capacitively end-coupled to a microstrip line; (*b*) the same resonator as in (*a*) but coupled to the side of the microstrip line; (*c*) a circular disk or "microstrip patch" resonator capacitively coupled to a microstrip line.

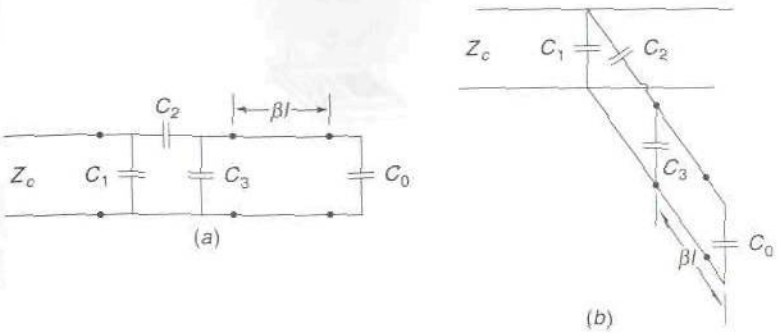


FIGURE 7.8
The equivalent circuit of the resonators shown in Figs. 7.7*a* and *b*.

ration is the circular disk that is capacitively coupled to a microstrip line as shown in Fig. 7.7*c*. The equivalent circuit for the resonators shown in Figs. 7.7*a* and *b* is a transmission line of length l with a capacitor C_0 at the open-circuit end and a capacitive Π network that represents the coupling region as shown in Fig. 7.8. The fringing electric field at the open-circuit end results in additional charge on the microstrip line near the open-circuit end and this is modeled by the capacitor C_0. The capacitor C_0 makes the electrical length of the resonator appear to be longer than the physical length l, because the line length has to be increased beyond $\lambda/2$ to provide the additional inductance that will compensate for the additional capacitance. Resonators of the type shown in Figs. 7.7*a* and *b* are commonly used in microstrip filters (see Chap. 8).

In order to design a resonator of the type shown in Fig. 7.7*a*, the open-circuit capacitance C_0 and the values of C_1, C_2, and C_3 in the coupling network must be known. These parameters have been evaluated by Silvester and Benedek using the assumption of static potential fields.† The results are accurate for frequencies up to several gigahertz for microstrip lines with substrates 1 mm thick and even at higher frequencies for thinner substrates. In Fig. 7.9 we show typical values of the open-circuit capacitance C_0/W, where W is the width of the microstrip line. In Fig. 7.10 we show typical results for C_1, C_2, and C_3 for various microstrip parameters including the gap spacing s.

†P. Silvester and P. Benedek, Equivalent Capacitances of Microstrip Open Circuits, *IEEE Trans.*, vol. MTT-20, pp. 511–516, 1972.

P. Benedek and P. Silvester, Equivalent Capacitances for Microstrip Gaps and Steps, *IEEE Trans.*, vol. MTT-20, pp. 729–733, 1972.

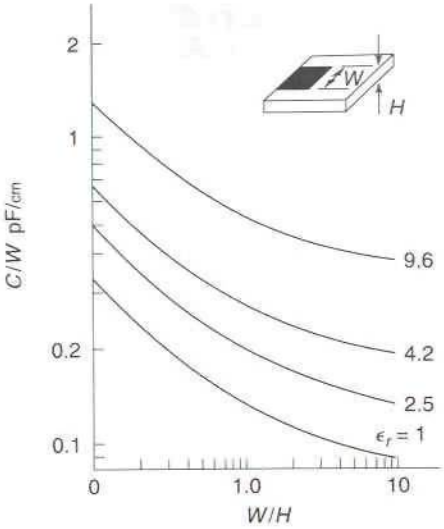


FIGURE 7.9
Microstrip-line open-circuited capacitance. (*Based on Fig. 5 of Silvester and Benedek.*)

The resonator and coupling circuit shown in Fig. 7.8*a* can be represented by the equivalent lumped-parameter circuit shown in Fig. 7.11*a*, in the vicinity of the resonant frequency. In these circuits the capacitance C_l and inductance L_l are those contributed by the transmission line and are given by

$$C_l = \tfrac{1}{2}C'l \tag{7.23a}$$

$$L_l = \tfrac{2}{\pi^2}L'l \tag{7.23b}$$

where C' and L' are the distributed capacitance and inductance per meter for the microstrip line and l is the length of the resonator. The factor of $\frac{1}{2}$ arises because the voltage and current standing waves on the line are sinusoidal, so that the stored electric and magnetic energies are reduced by one-half from what one would have for constant voltage and current wave amplitudes, and this makes the effective capacitance and inductance in the equivalent lumped-parameter model one-half of those for the line. The resistance R represents the total losses in the resonator.

For the circuit in Fig. 7.11*a*, the input admittance at the location of C_3 is given by

$$Y'_{\text{in}} = j\omega C_3 + Y_c \frac{j\omega C_0 + jY_c \tan \beta l}{Y_c - \omega C_0 \tan \beta l} + G$$

as reference to Fig. 7.8*a* clearly shows. For convenience we will denote ωC_0, ωC_1, ωC_2, and ωC_3 by B_0, B_1, B_2, and B_3. The input admittance seen

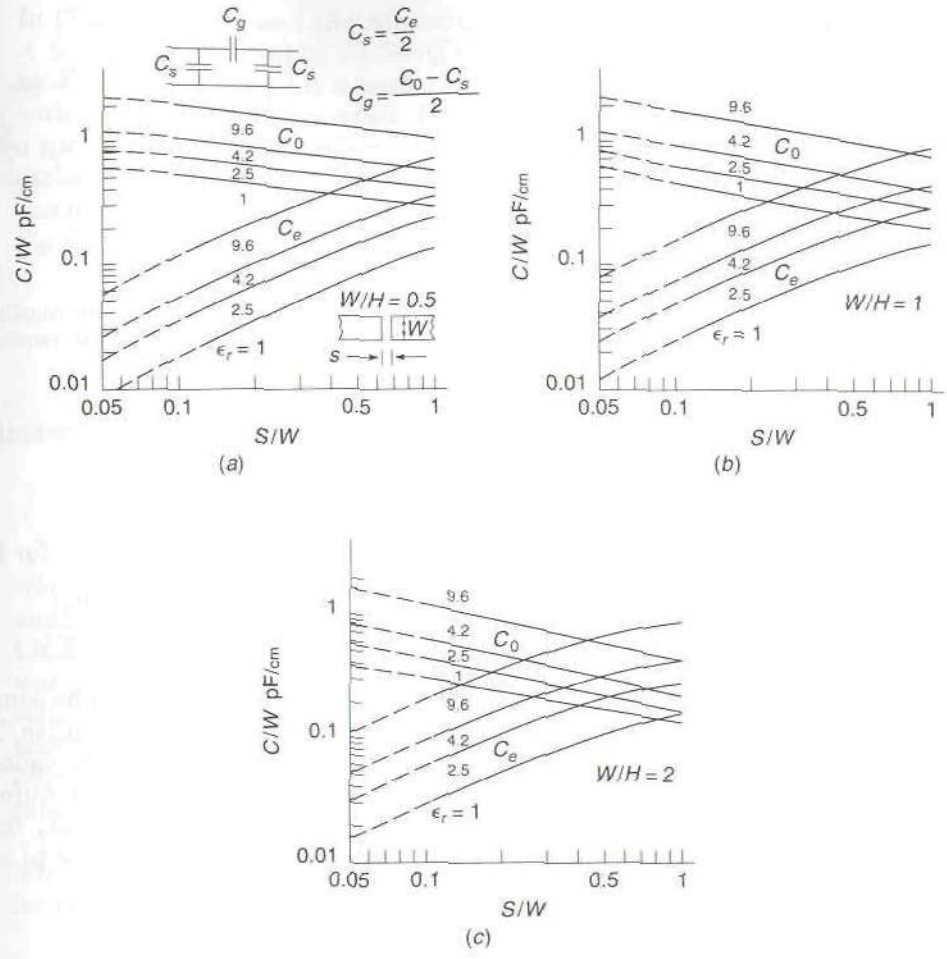


FIGURE 7.10
The Π-network capacitances for a microstrip-line gap. (*Based on Fig. 7 of Benedek and Silvester.*)

across C_1 is given by

$$Y_{in} = jB_1 + \frac{jB_2 Y'_{in}}{jB_2 + Y'_{in}}$$

$$= \frac{j(B_1 + B_2)Y'_{in} - B_1 B_2}{jB_2 + Y'_{in}} \tag{7.24}$$

In a gap coupled microstrip resonator $B_1 \ll B_2$ so it is operated as a series resonant circuit. Series resonance occurs when the imaginary part of the

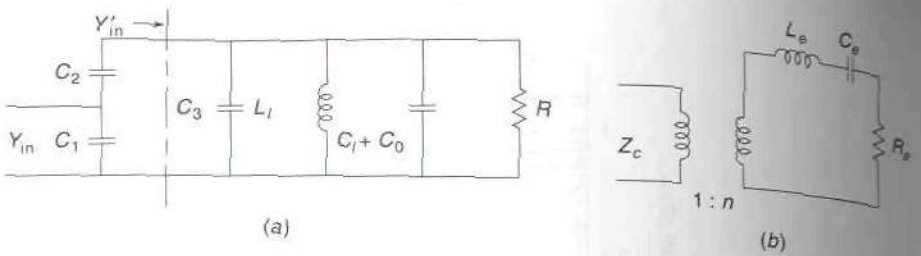


FIGURE 7.11
(*a*) Equivalent lumped-parameter circuit for the resonator and coupling network shown in Fig. 7.8*a*; (*b*) a simplified equivalent circuit incorporating an ideal transformer.

denominator vanishes. The condition for resonance is thus

$$Y'_{in} - G = jB'_{in} = -jB_2 \tag{7.25}$$

When we substitute for Y'_{in}, we obtain an equation for tan βl which is

$$\tan \beta l = \frac{Y_c(B_2 + B_0 + B_3)}{B_0(B_2 + B_3) - Y_c^2} \tag{7.26}$$

Since B_2, B_3, and B_0 are all very small, βl will be somewhat smaller than π since the expression for tan βl is small and negative. The resonant length is shorter than one-half wavelength because of the capacitive loading at the input and output ends. When we use (7.25) in the numerator in (7.24) and assume that we can approximate $G + jB'_{in}$ by $-jB_2$ for all values of ω in the vicinity of the resonant frequency, we are able to express (7.24) in the form

$$Y_{in} = \frac{B_2^2}{G + j(B'_{in} + B_2)}$$

When we compare this expression with that for the input admittance of the circuit in Fig. 7.11*b*, we see that

$$Z_{in} = \frac{1}{n^2}\left(j\omega L_e - \frac{j}{\omega C_e} + R_e\right) = \frac{1}{B_2^2}[G + j(B'_{in} + B_2)]$$

so the ideal transformer turns ratio is

$$n = B_2\sqrt{\frac{R_e}{G}} \tag{7.27}$$

where R_e can be chosen arbitrarily.

Thus we have

$$Y_{in} = B_2^2[j\omega(C_2 + C_3 + C_0) + jY_c \tan \beta l + G]^{-1} \tag{7.28}$$

In (7.28) we have used the approximation $Y'_{in} \approx j\omega C_3 + j\omega C_0 + jY_c \tan \beta l + G$ which is obtained by replacing the denominator factor $Y_c - \omega C_0 \tan \beta l$ by Y_c since $\omega C_0 \tan \beta l$ is a second-order small term near resonance because both ωC_0 and $\tan \beta l$ are small. The term $jY_c \tan \beta l$ can be represented by a parallel connection of C_l and L_l in the vicinity of resonance as discussed earlier for transmission-line resonators. Thus, in the equivalent series resonant circuit shown in Fig. 7.11*b*, the total inductance is the sum of C_l, C_0, C_3, and C_2, that is,

$$L_e = \frac{n^2}{B_2^2}[C_2 + C_3 + C_0 + C_l] \tag{7.29}$$

When the dispersion of the microstrip line can be neglected, the Q is given by (7.17) as

$$Q = \frac{\beta}{2\alpha} = \frac{R}{\omega_0 L_l} \approx \omega_0 C_l R \tag{7.30}$$

where R is the equivalent parallel resistance that will account for conductor and dielectric losses in the microstrip line. From the examples considered in Chap. 3, we found that typical values of attenuation for a microstrip line was in the range 0.5×10^{-2} to 1.5×10^{-2} Np/wavelength. Hence typical values of Q obtained from (7.30) would be 200 to 600. There is some additional loss caused by radiation of power from the open-circuited end. Measured values of Q for microstrip-line resonators are usually in the range 100 up to several hundred.

When $Y_{in} = Y_c$ at resonance, all of the incident power will be coupled into the resonator. Since $Y_{in} = n^2/R_e$ at resonance, the required value of n^2 for critical coupling is

$$n^2 = Y_c R_e$$

The parameter

$$\beta l Y_c \approx \frac{2\pi}{\lambda}\frac{\lambda}{2} Y_c = \pi Y_c = \omega_0 \sqrt{L'C'}\sqrt{\frac{C'}{L'}}\, l = 2\omega_0 C_l$$

so $R = 2Q/\pi Y_c$, as may be seen by using (7.30) for Q. Thus

$$B_2 = \sqrt{Y_c G} = Y_c \sqrt{\frac{\pi}{2Q}} \tag{7.31}$$

for critical coupling. If $Q = 157$, then $B_2 = Y_c/10$. For a high-Q resonator, only a small coupling capacitance is required to obtain critical coupling.

The loading of the resonator by the external transmission line is equivalent to a resistance $n^2 Z_c$ connected in series with R_e. Thus the

external Q_e is given by

$$Q_e = \frac{\omega_0 L_e}{n^2 Z_c} \approx \frac{\omega_0 C_l Y_c}{B_2^2} \approx \frac{Q Y_c G}{B_2^2} \tag{7.32a}$$

The coupling parameter K is given by the ratio Q/Q_e and is

$$K = \frac{B_2^2}{Y_c G} = \frac{2QB_2^2}{\pi Y_c^2} \tag{7.32b}$$

as found by using $R = 2Q/\pi Y_c$ given above (7.31). Since $B_2 = Y_c\sqrt{\pi/2Q} \gg \pi Y_c/2Q = G$ our earlier assumption that the numerator for Y_{in} in (7.24) could be approximated by B_2^2 is justified. From Fig. 7.10 it will be found that for the typical values of C_2 that are required C_1 and C_3 are negligible. A parallel resonant circuit coupled by a small series connected capacitor functions as a series resonant circuit.

For a transmission line without dispersion, that is, β is a linear function of ω, the Q of the transmission line is given by $\beta/2\alpha$. If the propagation constant β is not a linear function of ω, the transmission line is said to be dispersive. For a microstrip line the effective dielectric constant ϵ_e is a function of ω; so $\beta = \sqrt{\epsilon_e(\omega)}\, k_0$ is not a linear function of ω. In the frequency range where ϵ_e is changing quite rapidly with frequency, a different expression from that given above must be used for the Q. The power flow along the transmission line is given by Wv_g, where W Is the average stored energy per unit length and v_g is the group velocity which is given by $(d\beta/d\omega)^{-1}$. Thus $W = P/v_g = P(d\beta/d\omega)$ and the Q is given by

$$Q = \frac{\omega W}{P_l} = \frac{\omega P(d\beta/d\omega)}{2\alpha P} \approx \frac{\omega(d\beta/d\omega)}{2\alpha}$$

When there is no dispersion, $\beta = \omega/v_p$, where v_p is a frequency-independent phase velocity. For this case the general expression for Q reduces to $\beta/2\alpha$. When there is significant dispersion, the capacitance C_l in the equivalent circuit of the resonator is also a function of ω because ϵ_e is.

Circular Disk Resonator

The circular disk or planar radial resonator is shown in Fig. 7.12. An approximate analysis of this resonator is readily carried out by treating the outer boundary at $r = a$ as a perfect open-circuit boundary (magnetic wall) on which $\mathbf{n} \times \mathbf{H} = 0$. The field in the resonator will not depend on z and will have an axial electric field E_z. Thus the modes in the resonator will be like those of TM of E modes in a circular waveguide at cutoff but with the guide boundary being a magnetic wall instead of an electric wall. A suitable expression for the axial electric field is

$$E_z = CJ_n(kr)\cos n\phi \tag{7.33}$$

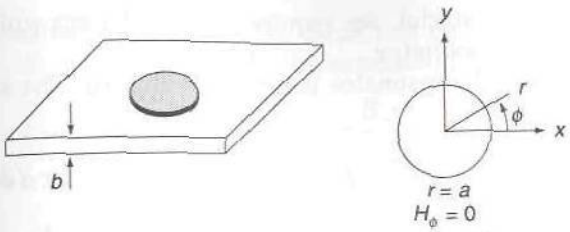


FIGURE 7.12
The circular disk resonator.

where C is an amplitude constant, J_n is the Bessel function of order n, and $k = \sqrt{\epsilon_r} k_0$ with ϵ_r being the dielectric constant of the substrate material. Since this is the only electric field component present, the magnetic field is given by

$$\nabla \times \mathbf{a}_z E_z = -\mathbf{a}_z \times \nabla E_z = -j\omega\mu_0 \mathbf{H} = -jk_0 Z_0 \mathbf{H}$$

or

$$\mathbf{H} = \frac{-j}{k_0 Z_0} \mathbf{a}_z \times \nabla E_z = \frac{-j}{k_0 Z_0} \left(\mathbf{a}_z \times \mathbf{a}_r \frac{\partial E_z}{\partial r} + \mathbf{a}_z \times \mathbf{a}_\phi \frac{1}{r} \frac{\partial E_z}{\partial \phi} \right)$$

$$= \frac{-jCk}{k_0 Z_0} \mathbf{a}_\phi J_n'(kr) \cos n\phi - \frac{jnC}{k_0 Z_0 r} \mathbf{a}_r J_n(kr) \sin n\phi \qquad (7.34)$$

where $J_n'(kr) = dJ_n(kr)/d(kr)$. In order for H_ϕ to vanish at the boundary $r = a$, we require $J_n'(ka) = 0$. The roots of this equation are given in Table 3.7. The smallest root is 1.841 corresponding to the use of the function $J_1(kr)$. The resonant frequency is obtained by equating $ka = \sqrt{\epsilon_r} k_0 a$ to 1.841; thus

$$\omega_{110} = \frac{1.841c}{\sqrt{\epsilon_r} a} \qquad (7.35a)$$

where $c = 3 \times 10^{10}$ cm/s is the speed of light.† Since $n = 1$, 1.841 is the first root, and there is no dependence on z, this dominant mode is designated as the TM_{110} mode. As an alternative we can solve for the radius a which is given by

$$a = \frac{1.841}{\sqrt{\epsilon_r} k_0} \qquad (7.35b)$$

For a resonator at a frequency of 4 GHz and using alumina with $\epsilon_r = 9.7$ as

†More accurate expressions for the resonant frequencies are given in I. Wolff and N. Knoppik, Rectangular and Circular Microstrip Disk Capacitors and Resonators, *IEEE Trans.*, vol. MTT-22, pp. 857–864, 1974.

the substrate material, we require $a = 0.7056$ cm which results in a relatively compact resonator.

The Q of the resonator is readily evaluated. The stored electric energy is given by

$$W_e = |C|^2 \frac{\epsilon_r' \epsilon_0}{4} \int_0^b \int_0^{2\pi} \int_0^a J_1^2(kr) \cos^2 \phi \, r \, dr \, d\phi \, dz$$

$$= |C|^2 \frac{\pi \epsilon_r' \epsilon_0 b}{8} a^2 \left[J_1'^2(ka) + \left(1 - \frac{1}{k^2 a^2}\right) J_1^2(ka) \right]$$

$$= |C|^2 \frac{\pi \epsilon_r' \epsilon_0 b}{8} a^2 \left(1 - \frac{1}{k^2 a^2}\right) J_1^2(ka) \tag{7.36}$$

since $J_1'(ka) = 0$. In this expression ϵ_r' is the real part of the dielectric constant and b is the substrate thickness. The stored magnetic energy W_m equals W_e at resonance.

The power loss in the dielectric is given by

$$P_{ld} = \frac{\omega \epsilon_r''}{2} \int_0^b \int_0^{2\pi} \int_0^a |E_z|^2 r \, dr \, d\phi \, dz = \frac{2\omega \epsilon_r''}{\epsilon_r'} W_e \tag{7.37}$$

The power loss in the conducting disk and ground plane is given by

$$P_{lc} = 2 \frac{R_m}{2} \int_0^{2\pi} \int_0^a \left(|J_r|^2 + |J_\phi|^2 \right) r \, dr \, d\phi$$

$$= R_m \int_0^{2\pi} \int_0^a \left(|H_\phi|^2 + |H_r|^2 \right) r \, dr \, d\phi$$

since the current density is given by $\mathbf{a}_z \times \mathbf{H}$ on the ground plane and by $-\mathbf{a}_z \times \mathbf{H}$ on the disk. After substituting from (7.34) for the magnetic field, we obtain

$$P_{lc} = |C|^2 \frac{R_m \pi}{k_0^2 Z_0^2} \int_0^a \left[k^2 J_1'^2(kr) + \frac{1}{r^2} J_1^2(kr) \right] r \, dr$$

after integrating over ϕ which gave a factor of π. We can integrate the first term by parts to obtain

$$\int_0^a \frac{dJ_1(kr)}{dr} \frac{dJ_1(kr)}{dr} r \, dr = r J_1(kr) \frac{dJ_1(kr)}{dr} \Bigg|_0^a - \int_0^a J_1(kr) \frac{d}{dr} \left[r \frac{dJ_1(kr)}{dr} \right] dr$$

Since $J_1'(ka) = 0$ the integrated term is zero. We now take note of the differential equation

$$\frac{d}{dr} r \frac{dJ_1}{dr} + \left(k^2 r - \frac{1}{r} \right) J_1 = 0$$

satisfied by $J_1(kr)$. By using this equation to express the second derivative of J_1 in terms of J_1, the expression for P_{lc} reduces to

$$P_{lc} = |C|^2 \frac{R_m \pi}{k_0^2 Z_0^2} \int_0^a k^2 r J_1^2(kr)\, dr$$

$$= |C|^2 \frac{R_m \pi k^2}{2k_0^2 Z_0^2} a^2 \left(1 - \frac{1}{k^2 a^2}\right) J_1^2(ka)$$

The Q is given by

$$Q = \frac{2\omega W_e}{P_{ld} + P_{lc}} = \frac{k_0 b}{2R_m Y_0 + \epsilon_r'' k_0 b/\epsilon_r'} \tag{7.38}$$

which is a surprisingly simple result in that there is no dependence on the Bessel function or the radius of the resonator. In (7.38) the substrate thickness is given by the parameter b.

As an example we will consider a disk resonator operating at 4 GHz. The substrate will be chosen as alumina with $\epsilon_r' = 9.7$ and $\epsilon_r'' = 0.0002$ and having a thickness $b = 1$ mm. For this resonator $k_0 b = (2\pi/75) = 0.08377$:

$$Q = \frac{8.377 \times 10^{-2}}{2 \times 16.44 \times 10^{-3}/120\pi + 2 \times 8.377 \times 10^{-6}/9.7}$$

$$= \frac{8.377}{8.72 \times 10^{-3} + 1.72 \times 10^{-4}} = 942$$

The dominant loss mechanism is conductor loss. The Q of this resonator is higher than that of a transmission-line type since the conductor losses are smaller. In practice, the Q can be expected to be less because of radiation loss and extra loss from surface roughness on the disk. There will also be a much larger azimuthal current density J_ϕ at the edge than is obtained using the approximate theory given above. For an infinitely thin disk, the current J_ϕ should exhibit the usual edge singularity and be proportional to $(a^2 - r^2)^{-1/2}$ as the edge is approached. There will also be some current flowing on the top surface of the disk and this will also increase the conductor losses.

The actual resonant frequency of the disk resonator is lower than that predicted by the simple approximate theory used above. More accurate formulas have been obtained that will give the resonant frequency to within about 1 percent. The reader is referred to the literature for a more comprehensive treatment.†

†I. Wolff and N. Knoppik, Rectangular and Circular Microstrip Disk Capacitors and Resonators, *IEEE Trans.*, vol. MTT-22, pp. 857–864, 1974.

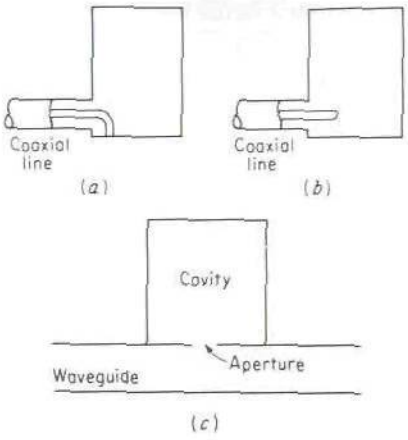


FIGURE 7.13
Cavity-coupling methods. (*a*) Loop coupling; (*b*) probe coupling; (*c*) aperture coupling.

In addition to the circular disk, other shapes such as ellipses, rings, triangles, and squares, have been considered for use as resonators, as well as coupled sections of microstrip lines.†

7.4 MICROWAVE CAVITIES

At frequencies above 1,000 MHz, transmission-line resonators have relatively low values of Q, and so it becomes preferable to use metallic enclosures, or cavities, instead. A cavity can be considered as a volume enclosed by a conducting surface and within which an electromagnetic field can be excited. The electric and magnetic energies are stored in the volume of the cavity. The finite conducting walls give rise to power loss and thus are equivalent to some effective resistance. The fields in the cavity may be excited, or coupled to an external circuit, by means of small coaxial-line probes or loops. Alternatively, the cavity may be coupled to a waveguide by means of a small aperture in a common wall. These coupling methods are illustrated in Fig. 7.13. Before considering the effects of coupling on cavity performance, the field solutions in rectangular and cylindrical cavities are presented.

Rectangular Cavity

Figure 7.14 illustrates a rectangular cavity of height b, width a, and length d. It may be considered to be a section of rectangular waveguide terminated in a short circuit at $z = d$. If d equals a multiple of a half guide wavelength

†J. Helszajn and D. S. James, Planar Triangular Resonators with Magnetic Walls, *IEEE Trans.*, vol. MTT-26, pp. 95–100, 1978.

I. Bahl and P. Bhartia, "Microwave Solid State Circuit Design," John Wiley & Sons, Inc., New York, 1988.

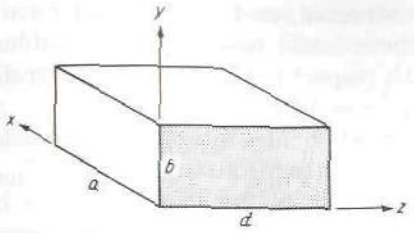


FIGURE 7.14
A rectangular cavity.

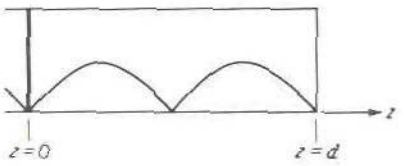


FIGURE 7.15
Standing-wave pattern in a short-circuited waveguide.

at the frequency f, the resultant standing-wave pattern is such that the x and y components of electric field are zero at $z = 0$. Consequently, a short circuit can be placed at $z = 0$ as well, as in Fig. 7.15. The resultant structure is a rectangular cavity. This description of a cavity also shows that the field solution may be obtained directly from the corresponding waveguide solutions. For the nmth TE or TM mode, the propagation constant is given by

$$\beta_{nm}^2 = k_0^2 - \left(\frac{n\pi}{a}\right)^2 - \left(\frac{m\pi}{b}\right)^2 \tag{7.39a}$$

where $k_0 = 2\pi f_0/c$. We require $\beta_{nm} d = l\pi$, where l is an integer in order for the cavity to be a multiple of a half guide wavelength long. Thus, when d is specified, β_{nm} is given by

$$\beta_{nm} = \frac{l\pi}{d} \qquad l = 1, 2, \ldots \tag{7.39b}$$

However, this relation is consistent with the earlier one only for certain discrete values of k_0. Only if $k_0 = k_{nml}$, where k_{nml} is given by

$$k_{nml} = \left[\left(\frac{l\pi}{d}\right)^2 + \left(\frac{m\pi}{b}\right)^2 + \left(\frac{n\pi}{a}\right)^2\right]^{1/2} \tag{7.40}$$

will (7.39a) and (7.39b) be satisfied. These particular values of k_0 give the resonant frequencies of the cavity; i.e.,

$$f_{nml} = \frac{ck_{nml}}{2\pi} = c\left[\left(\frac{l}{2d}\right)^2 + \left(\frac{m}{2b}\right)^2 + \left(\frac{n}{2a}\right)^2\right]^{1/2} \tag{7.41}$$

where c is the velocity of light. Note that there is a triply infinite number of resonant frequencies corresponding to different field distributions. Also note that there is more than one field solution for a given resonant frequency since (7.41) holds for both TE and TM modes. In addition, because of a lack of a preferential coordinate, in the case of a rectangular cavity, field solutions corresponding to TE and TM modes with respect to the x and y

axes could also be constructed, and these would have the same resonant frequency. However, these latter modes are just a linear combination of a TE and a TM mode with respect to the z axis and therefore do not represent a new solution.

To illustrate the method of solution for the fields in a rectangular cavity and the evaluation of the unloaded Q, the TE_{101} mode is treated in detail. If $b < a < d$, this will be the mode with the lowest resonant frequency and corresponds to the TE_{10} mode in a rectangular waveguide. The mode subscripts nml pertain to the number of half-sinusoid variations in the standing-wave pattern along the x, y, and z axes, respectively. Using (3.206), the field solution for a TE_{10} mode is

$$H_z = (A^+ e^{-j\beta_{10}z} + A^- e^{j\beta_{10}z}) \cos \frac{\pi x}{a}$$

$$H_x = \frac{j\beta_{10} a}{\pi} (A^+ e^{-j\beta_{10}z} - A^- e^{j\beta_{10}z}) \sin \frac{\pi x}{a}$$

$$E_y = -j \frac{k_0 Z_0 a}{\pi} (A^+ e^{-j\beta_{10}z} + A^- e^{j\beta_{10}z}) \sin \frac{\pi x}{a}$$

$$H_y = E_x = E_z = 0$$

where A^+ and A^- are amplitude constants for the modes propagating in the $+z$ and $-z$ directions, respectively. To make E_y vanish at $z = 0, d$, we must choose $A^- = -A^+$ so that

$$A^+ e^{-j\beta_{10}z} + A^- e^{j\beta_{10}z} = -2jA^+ \sin \beta_{10} z$$

and also choose $\beta_{10} = \pi/d$. The corresponding value of k_0 is thus

$$k_0 = k_{101} = \left[\left(\frac{\pi}{a} \right)^2 + \left(\frac{\pi}{d} \right)^2 \right]^{1/2} = \left[\left(\frac{\pi}{a} \right)^2 + \beta_{10}^2 \right]^{1/2} \tag{7.42}$$

and this determines the resonant frequency. The solution for the fields may now be expressed as follows:

$$E_y = \frac{-2A^+ k_{101} Z_0 a}{\pi} \sin \frac{\pi x}{a} \sin \frac{\pi z}{d} \tag{7.43a}$$

$$H_x = \frac{2jA^+ a}{d} \sin \frac{\pi x}{a} \cos \frac{\pi z}{d} \tag{7.43b}$$

$$H_z = -2jA^+ \cos \frac{\pi x}{a} \sin \frac{\pi z}{d} \tag{7.43c}$$

Note that the magnetic field is $\pm 90°$ out of phase relative to the electric field. This is always the case in a lossless cavity and corresponds to voltage and current being $\pm 90°$ out of phase with each other in a lossless LC circuit.

At resonance the time-average electric and magnetic energy stored in the cavity are equal. The average stored electric energy is given by

$$W_e = \frac{\epsilon_0}{4}\int_0^a\int_0^b\int_0^d E_y E_y^*\,dx\,dy\,dz$$

$$= \frac{\epsilon_0}{4\pi^2}a^3bdk_{101}^2Z_0^2|A^+|^2 \tag{7.44}$$

The reader may readily verify that

$$W_m = \frac{\mu_0}{4}\int_0^a\int_0^b\int_0^d(H_xH_x^* + H_zH_z^*)\,dx\,dy\,dz = W_e \tag{7.45}$$

In order to determine the cavity Q, the losses caused by the finite conductivity of the cavity walls must be evaluated. For small losses the surface currents are essentially those associated with the loss-free field solutions (7.43). Thus the surface current is given by

$$\mathbf{J}_s = \mathbf{n} \times \mathbf{H}$$

where $\mathbf{n}$ is a unit normal to the surface and directed into the cavity. Hence the power loss in the walls is given by

$$P_l = \frac{R_m}{2}\int_{\text{walls}} \mathbf{J}_s \cdot \mathbf{J}_s^*\,dS = \frac{R_m}{2}\int_{\text{walls}} |\mathbf{H}_{\tan}|^2\,dS \tag{7.46}$$

where $R_m = 1/\sigma\delta_s$ is the resistive part of the surface impedance exhibited by the conducting wall having a conductivity σ and for which the skin depth is $\delta_s = (2/\omega\mu\sigma)^{1/2}$. In (7.46) $\mathbf{H}_{\tan}$ is the tangential magnetic field at the surface of the cavity walls. Substituting from (7.43) into (7.46), a straightforward calculation gives

$$P_l = |A^+|^2R_m\frac{2a^3b + 2d^3b + ad^3 + da^3}{d^2} \tag{7.47}$$

With the use of (7.3), we find that the Q is given by

$$Q = \frac{\omega W}{P_l} = \frac{2\omega W_e}{P_l} = \frac{\omega k_{101}^2 Z_0^2 a^3 d^3 b\epsilon_0}{2\pi^2 R_m(2a^3b + 2d^3b + a^3d + d^3a)}$$

$$= \frac{(k_{101}ad)^3 bZ_0}{2\pi^2 R_m(2a^3b + 2d^3b + a^3d + d^3a)} \tag{7.48}$$

upon replacing $\omega Z_0\epsilon_0 = \omega\sqrt{\mu_0\epsilon_0}$ by k_{101}.

As a typical example, consider a copper cavity ($\sigma = 5.8 \times 10^7$ S/m) with $a = b = d = 3$ cm. The resonant frequency is found to be 7,070 MHz. The surface resistance $R_m = 0.022\ \Omega$, and the Q comes out equal to 12,700. The damping factor $\delta = \omega/2Q$ equals 1.74×10^6 Np/s, or about 2.5×10^{-4} Np/cycle of oscillation. Because of the high value of Q, 4,000

cycles of free oscillation can occur before the amplitude has decreased by a factor e^{-1}.

If the cavity is filled with a lossy dielectric material with permittivity $\epsilon = \epsilon' - j\epsilon''$, the time-average electric energy stored in the cavity volume is given by

$$W_e = \frac{\epsilon'}{4}\int_V |\mathbf{E}|^2\, dV \tag{7.49a}$$

The lossy dielectric has an effective conductivity $\omega\epsilon''$, and hence the power loss in the dielectric is

$$P_{ld} = \frac{1}{2}\int_V \mathbf{J}\cdot\mathbf{E}^*\, dV = \frac{\omega\epsilon''}{2}\int_V |\mathbf{E}|^2\, dV \tag{7.49b}$$

If Q_d is the Q when a lossy dielectric is present but the walls are perfectly conducting, then

$$Q_d = \frac{2\omega W_e}{P_{ld}} = \frac{\epsilon'}{\epsilon''} \tag{7.50}$$

When wall losses are also present, the net Q is Q' and given by

$$Q' = \left(\frac{1}{Q_d} + \frac{1}{Q}\right)^{-1} \tag{7.51}$$

where Q is the quality factor when lossy walls are present and $\epsilon'' = 0$. Q is given by (7.48), with ϵ_0 replaced by ϵ'. Also note that, for a cavity filled with dielectric, the resonant frequency is given by

$$f_{nml} = \sqrt{\frac{\epsilon_0}{\epsilon'}}\,\frac{c}{2\pi}k_{nml} \tag{7.52}$$

Cylindrical Cavity

The cylindrical cavity is a section of circular waveguide of length d and radius a, with short circuiting plates at each end, as in Fig. 7.16. This type of cavity is very commonly used in practice for wavemeters, to measure frequency, because of the high Q and wide frequency range of operation it

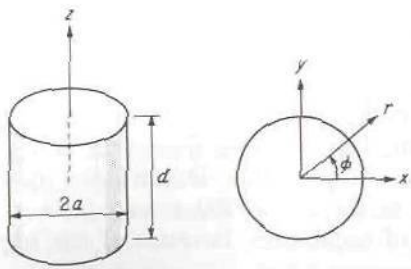


FIGURE 7.16
Cylindrical cavity.

provides. A high Q is necessary in a frequency meter in order to obtain a high degree of resolution or accuracy in the measurement of an unknown frequency. When the cavity is tuned to the frequency of the unknown source, it absorbs a maximum of power from the input line. A crystal detector coupled to the input line can be used to indicate this dip in power at resonance.

The fields in the cylindrical cavity may be determined from the corresponding waveguide solutions. The lowest resonant mode is the TE_{111} mode corresponding to the dominant TE_{11} mode in the circular guide. This mode is examined in detail below. Using the field solutions tabulated in Table 3.6 and combining a forward and backward propagating TE_{11} mode, we have

$$H_z = J_1\left(\frac{p'_{11}r}{a}\right)\cos\phi\,(A^+e^{-j\beta_{11}z} + A^-e^{j\beta_{11}z}) \tag{7.53a}$$

$$H_r = \frac{-j\beta_{11}a}{p'_{11}}J'_1\left(\frac{p'_{11}r}{a}\right)\cos\phi\,(A^+e^{-j\beta_{11}z} - A^-e^{j\beta_{11}z}) \tag{7.53b}$$

$$H_\phi = \frac{j\beta_{11}a^2}{(p'_{11})^2 r}J_1\left(\frac{p'_{11}r}{a}\right)\sin\phi\,(A^+e^{-j\beta_{11}z} - A^-e^{j\beta_{11}z}) \tag{7.53c}$$

$$E_r = \frac{jk_0Z_0a^2}{(p'_{11})^2 r}J_1\left(\frac{p'_{11}r}{a}\right)\sin\phi\,(A^+e^{-j\beta_{11}z} + A^-e^{j\beta_{11}z}) \tag{7.53d}$$

$$E_\phi = \frac{jk_0Z_0a}{p'_{11}}J'_1\left(\frac{p'_{11}r}{a}\right)\cos\phi\,(A^+e^{-j\beta_{11}z} + A^-e^{j\beta_{11}z}) \tag{7.53e}$$

$$E_z = 0 \tag{7.53f}$$

where $p'_{11} = 1.841$. To make E_r and E_ϕ vanish at $z = 0, d$, we must choose $A^- = -A^+$. The factor $A^+e^{-j\beta_{11}z} + A^-e^{j\beta_{11}z}$ becomes $-2jA^+\sin\beta_{11}z$. β_{11} must be chosen equal to π/d to make $\sin\beta_{11}d$ vanish. The resonant frequency is determined from the relation

$$k_0 = \left[\beta_{11}^2 + \left(\frac{p'_{11}}{a}\right)^2\right]^{1/2} = \left[\left(\frac{\pi}{d}\right)^2 + \left(\frac{p'_{11}}{a}\right)^2\right]^{1/2} = \frac{2\pi f_{111}}{c} \tag{7.54}$$

To find the Q of the TE_{111} mode, the time-average energy W stored in the cavity and the power loss in the walls must be calculated. The general expressions are

$$W = 2W_e = \frac{\epsilon_0}{2}\int_0^a\int_0^{2\pi}\int_0^d\left(|E_r|^2 + |E_\phi|^2\right)r\,d\phi\,dr\,dz$$

$$P_l = \frac{R_m}{2}\int_{\text{walls}}|\mathbf{H}_{\tan}|^2\,dS$$

These integrals may be evaluated by substituting the expressions given

earlier for the fields. The final result obtained for the Q is

$$Q\frac{\delta_s}{\lambda_0} = \frac{\left[1-\left(\frac{n}{p'_{nm}}\right)^2\right]\left[(p'_{nm})^2+\left(\frac{l\pi a}{d}\right)^2\right]^{3/2}}{2\pi\left[(p'_{nm})^2+\frac{2a}{d}\left(\frac{l\pi a}{d}\right)^2+\left(1-\frac{2a}{d}\right)\left(\frac{nl\pi a}{p'_{nm}d}\right)^2\right]} \tag{7.55}$$

where $l = n = m = 1$ for the TE_{111} mode. For the TE_{nml} mode, the Q is also given by (7.55), with the appropriate values of n, m, l, and p'_{nm} inserted. Note that all terms on the right are independent of frequency, and hence the Q varies as λ_0/δ_s for any given cavity and thus decreases as $f^{-1/2}$.

An analysis similar to the above may be carried out for the other TE_{nml} and TM_{nml} modes to obtain expressions for the fields. For the TE_{nml} modes it is necessary only to replace $\cos\phi$ and $\sin\phi$ by $\cos n\phi$ and $\sin n\phi$, J_1 by J_n, p'_{11} by p'_{nm}, and β_{11} by $l\pi/d$ in (7.53). Of particular interest is the TE_{011} mode for wavemeters because its Q is two to three times that of the TE_{111} mode. Another advantage of the TE_{011} mode is that $H_\phi = 0$, and hence there are no axial currents. This means that the end plate of the cavity can be free to move to adjust the cavity length d for tuning purposes without introducing any significant loss since no currents flow across the gap; i.e., the gap between the circular end plate and the cylinder wall is parallel too the current flow lines. However, the TE_{011} mode is not the dominant mode; so care must be exercised to choose a coupling scheme that does not excite the other possible modes that could resonate within the frequency tuning range of the cavity.

To determine what modes can resonate for a given value of $2a/d$ and frequency, it is convenient to construct a mode chart. For any given mode we have

$$f_{nml} = \frac{k_{nml}}{2\pi}c = \left[\left(\frac{x_{nm}}{a}\right)^2+\left(\frac{l\pi}{d}\right)^2\right]^{1/2}\frac{c}{2\pi}$$

or

$$(2af_{nml})^2 = \left(\frac{cx_{nm}}{\pi}\right)^2+\left(\frac{cl}{2}\right)^2\left(\frac{2a}{d}\right)^2 \tag{7.56}$$

where $x_{nm} = p'_{nm}$ for TE modes and p_{nm} for TM modes. Figure 7.17 gives a plot of $(2af_{nml})^2$ against $(2a/d)^2$ for several modes and constitutes a mode chart. Examination of this chart shows over what range of frequency and $2a/d$ only a single mode can resonate (in the case of degenerate modes, two modes resonate at the same frequency). For example, for $(2a/d)^2$ between 2 and 3, only the TE_{011} and TM_{111} modes can resonate in the frequency range corresponding to $(2af)^2$ between 16.3×10^8 and 20.4×10^8 (region within dashed rectangle in Fig. 7.17). If the TM_{111} mode is not excited, this

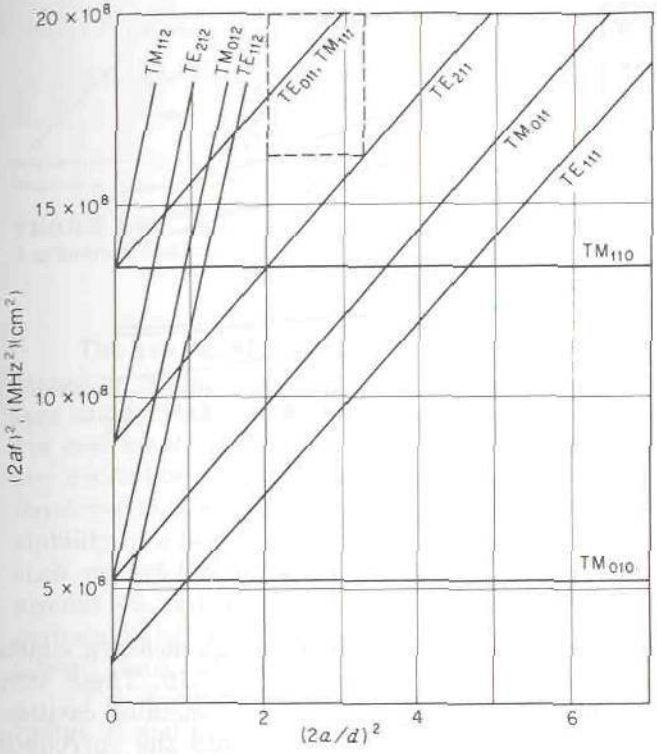


FIGURE 7.17
Mode chart for a circular cylindrical cavity.

frequency range at least can be tuned without spurious resonances from other modes occurring.

For the TM modes the Q can be evaluated to give

$$Q\frac{\delta_s}{\lambda_0} = \begin{cases} \dfrac{\left[p_{nm}^2 + (l\pi a/d)^2\right]^{1/2}}{2\pi(1 + 2a/d)} & l > 0 \\ \dfrac{p_{nm}}{2\pi(1 + a/d)} & l = 0 \end{cases} \tag{7.57}$$

Figure 7.18 gives a plot of $Q\delta_s/\lambda_0$ against $2a/d$ for several modes. Note the considerably higher value of Q obtained for the TE_{011} mode relative to that for the TE_{111} mode. Optimum Q occurs for $d \approx 2a$. At $\lambda_0 = 3$ cm, $\delta_s/\lambda_0 = 2.2 \times 10^{-5}$, and hence, from Fig. 7.18, it is apparent that typical values of Q range from 10,000 to 40,000 or more. At $\lambda_0 = 10$ cm, the corresponding values of Q would be $\sqrt{10/3}$ greater.

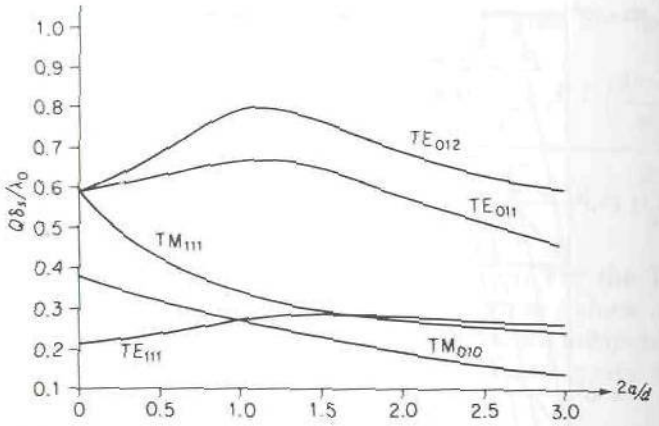


FIGURE 7.18
Q for circular-cylindrical-cavity modes.

7.5 DIELECTRIC RESONATORS

A dielectric resonator consists of a high dielectric constant cylinder, sphere, or parallelopiped as shown in Fig. 7.19. These structures will support resonant modes similar to those in metallic cavities but with the field extending beyond the boundaries into the surrounding air region. The linear dimension of the resonator is of order $\lambda_0/\sqrt{\epsilon_r}$, where λ_0 is the free-space wavelength and ϵ_r is the dielectric constant. For $\epsilon_r = 100$ the resonator is approximately one-tenth the size of a metallic cavity, and for this reason it is a very attractive alternative to other types of resonators for use in microstrip circuits such as filters and oscillator circuits. Since the required size of a dielectric resonator in order to resonate is very small relative to the free-space wavelength, the electromagnetic field outside the resonator is quasistatic and very little radiation takes place. Thus dielectric resonators will exhibit a relatively high Q if a low-loss high dielectric constant material is used. Typical values for the unloaded Q of a dielectric resonator range from around 100 to several hundred. If the resonator is enclosed in a shielding box so that there is no energy loss due to radiation, an unloaded Q approaching the intrinsic Q of the material can be obtained.

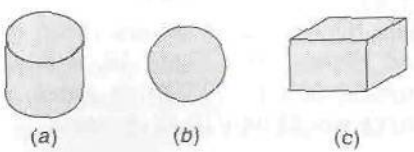


FIGURE 7.19
Basic geometrical shapes used for dielectric resonators. (*a*) Cylinder; (*b*) sphere; (*c*) parallelopiped.

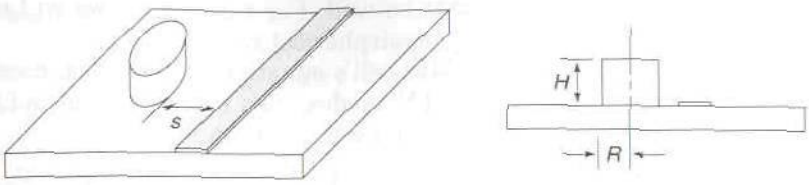


FIGURE 7.20
A cylindrical dielectric resonator coupled to a microstrip line.

The resonant frequency of a dielectric resonator depends on the dimensions of the resonator and its dielectric constant. Since these parameters change with the temperature, it is necessary to use materials that have low coefficients of expansion and relatively temperature-independent dielectric constants. In recent years a number of ceramic compounds have been developed that are suitable for use in dielectric resonators and can provide a stability of a few parts per million per degree change in temperature. One such material is barium tetratitanate which has a dielectric constant of around 40 and a loss tangent of about 0.0005. The intrinsic Q of the material is the reciprocal of the loss tangent and equals 2,000 for barium tetratitanate.

The dielectric cylinder with a height H approximately equal to the radius R and mounted on the substrate as shown in Fig. 7.20 is a popular choice for a dielectric resonator. The coupling between the resonator and a microstrip line is adjusted by adjusting the spacing S. The dominant resonant mode is a TE mode having an azimuthal electric field E_ϕ that is independent of the angle ϕ. The field has a dependence on the axial coordinate z but does not change with z rapidly enough to exhibit a standing-wave pattern along z. For this reason the mode is designated as the $TE_{01\delta}$ mode with the subscript δ signifying less than one-half cyclic variation with z.

Dielectric resonators in the shape of a cylinder or a parallelopiped do not have simple analytic solutions for the various modes of oscillation. On the other hand, there are simple solutions for the fields in a spherical resonator and also for a hemisphere mounted on a ground plane. The hemisphere on a ground plane shown in Fig. 7.21 is quite similar to the dielectric cylinder, so that an understanding of the field solution for the hemispherical resonator will provide good insight into the properties of a

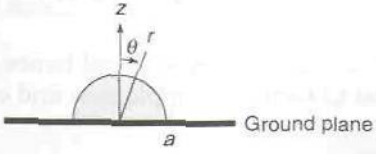


FIGURE 7.21
The hemispherical dielectric resonator on a ground plane.

cylindrical resonator as well. For this reason we will examine the solution for TE modes in a hemispherical resonator.

Solutions to Maxwell's equations in spherical coordinates can be separated into TE and TM modes with respect to the radial coordinate r. The electric field for TE modes is given by

$$\mathbf{E} = \nabla \times \mathbf{a}_r kr\psi = \mathbf{M} \tag{7.58a}$$

and the electric field for TM modes is given by

$$\mathbf{E} = \frac{1}{k}\nabla \times \nabla \times \mathbf{a}_r kr\psi = \mathbf{N} \tag{7.58b}$$

where $\psi(r, \theta, \phi)$ is a solution of the scalar Helmholtz equation in spherical coordinates, i.e.,

$$(\nabla^2 + k^2)\psi = 0 \tag{7.58c}$$

The **M** and **N** functions have the useful property that

$$\nabla \times \mathbf{M} = k\mathbf{N} \tag{7.59a}$$

$$\nabla \times \mathbf{N} = k\mathbf{M} \tag{7.59b}$$

Thus, if the electric field is described by an **M** function, the magnetic field given by $-j\omega\mu_0\mathbf{H} = \nabla \times \mathbf{E}$ will be described by an **N** function. For circularly symmetric fields, the solutions for $\psi(r, \theta)$ are

$$\psi_n(r, \theta) = z_n(kr)P_n(\cos\theta) \tag{7.60}$$

where $z_n(kr)$ is a spherical Bessel function and $P_n(\cos\theta)$ is the nth Legendre polynomial. The spherical Bessel functions are related to the cylinder Bessel functions $Z_{n+1/2}(kr)$ as follows:

$$z_n(kr) = \sqrt{\frac{\pi}{2kr}}\, Z_{n+1/2}(kr)$$

When the origin is included, the function

$$j_n(kr) = \sqrt{\frac{\pi}{2kr}}\, J_{n+1/2}(kr)$$

is used. In order to represent outward propagating waves, the spherical Hankel function

$$h_n^2(kr) = \sqrt{\frac{\pi}{2kr}}\, H_{n+1/2}^2(kr)$$

is used. The Bessel functions of order $n + \frac{1}{2}$ and hence the spherical Bessel functions can be expressed in terms of simple sine and cosine functions. For

example,

$$j_0(x) = \frac{\sin x}{x} \tag{7.61a}$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \tag{7.61b}$$

$$h_0^2(x) = \frac{je^{-jx}}{x} \tag{7.61c}$$

$$h_1^2(x) = \frac{je^{-jx}}{x^2} - \frac{e^{-jx}}{x} \tag{7.61d}$$

$$z_{n+1}(x) = \frac{2n+1}{x} z_n(x) - z_{n-1}(x) \tag{7.61e}$$

The first few Legendre polynomials are

$$P_0(\cos\theta) = 1 \tag{7.62a}$$

$$P_1(\cos\theta) = \cos\theta \tag{7.62b}$$

$$P_2(\cos\theta) = \tfrac{3}{2}\cos^2\theta - \tfrac{1}{2} = \tfrac{3}{4}\cos 2\theta + \tfrac{1}{4} \tag{7.62c}$$

For circularly symmetric modes, we can use (7.60) in (7.58) to obtain

$$\mathbf{M}_n = -k\frac{dP_n}{d\theta} z_n(kr)\mathbf{a}_\phi \tag{7.63a}$$

$$\mathbf{N}_n = \frac{n(n+1)}{r} P_n z_n(kr)\mathbf{a}_r + \frac{1}{r}\frac{dP_n}{d\theta}\frac{d(krz_n)}{d(kr)}\mathbf{a}_\theta \tag{7.63b}$$

We will use the above solutions to construct the solution for the dominant TE_{012} mode for the hemispherical resonator. The electric field has an E_ϕ component only. The ground plane is located at $\theta = \pi/2$, so that E_ϕ must vanish at $\theta = \pi/2$. The $\mathbf{M}_1$ mode does not vanish at $\theta = \pi/2$ but the $\mathbf{M}_2$ mode does; so the first mode that can exist for the hemispherical resonator on a ground plane is the TE_{012} mode. A suitable solution for the electric field is

$$\mathbf{E} = -A_1 k\frac{dP_2}{d\theta} j_2(kr)\mathbf{a}_\phi \qquad r \le a \tag{7.64a}$$

$$\mathbf{E} = -A_2 k_0\frac{dP_2}{d\theta} h_2^2(k_0 r)\mathbf{a}_\phi \qquad r > a \tag{7.64b}$$

where $k = \sqrt{\epsilon_r}k_0$ inside the sphere and A_1 and A_2 are amplitude constants. The solution for the magnetic field is given by

$$\mathbf{H} = \frac{j}{k_0 Z_0}\nabla \times \mathbf{E} = \frac{j}{k_0 Z_0}\nabla \times \mathbf{M} = \frac{j}{k_0 Z_0} k\mathbf{N}$$

and hence is given by

$$\mathbf{H} = A_1 \frac{j6k}{k_0 Z_0 r} P_2 j_2(kr)\mathbf{a}_r + A_1 \frac{jk}{k_0 Z_0 r} \frac{dP_2}{d\theta} \frac{d(krj_2(kr))}{d(kr)} \mathbf{a}_\theta \qquad r \le a \tag{7.65a}$$

$$\mathbf{H} = A_2 \frac{j6}{Z_0 r} P_2 h_2^2(k_0 r)\mathbf{a}_r + A_2 \frac{j}{Z_0 r} \frac{dP_2}{d\theta} \frac{d(k_0 r h_2^2(k_0 r))}{d(k_0 r)} \mathbf{a}_\theta \qquad r > a \tag{7.65b}$$

At $r = a$ the tangential fields E_ϕ and H_θ must be continuous across the boundary. Thus we must have

$$A_1 k j_2(ka) = A_2 k_0 h_2^2(k_0 a) \tag{7.66a}$$

$$A_1 k \frac{d(krj_2(kr))}{d(kr)}\bigg|_{r=a} = A_2 k_0 \frac{d(k_0 r h_2^2(k_0 r))}{d(k_0 r)}\bigg|_{r=a} \tag{7.66b}$$

All the other factors in the expressions for the fields are the same and cancel. By using (7.61*e*) we find that

$$j_2(kr) = \left(\frac{3}{k^3 r^3} - \frac{1}{kr}\right)\sin kr - \frac{3}{k^2 r^2}\cos kr \tag{7.67a}$$

$$h_2^2(k_0 r) = \left(\frac{3j}{k_0^3 r^3} - \frac{3}{k_0^2 r^2} - \frac{j}{k_0 r}\right) e^{-jk_0 r} \tag{7.67b}$$

With the aid of these relations, the boundary conditions can be expressed as

$$A_1 k\left[\left(\frac{3}{k^3 a^3} - \frac{1}{ka}\right)\sin ka - \frac{3}{k^2 a^2}\cos ka\right] - A_2 k_0\left[\left(\frac{3j}{k_0^3 a^3} - \frac{3}{k_0^2 a^2} - \frac{j}{k_0 a}\right) e^{-jk_0 a}\right] = 0 \tag{7.68a}$$

$$A_1 k\left[\left(\frac{3}{ka} - \frac{6}{k^2 a^3}\right)\sin ka + \left(\frac{6}{k^2 a^2} - 1\right)\cos ka\right] - A_2 k_0\left[\left(-\frac{6j}{k_0^3 a^3} + \frac{6}{k_0^2 a^2} + \frac{3j}{k_0 a} - 1\right) e^{-jk_0 a}\right] = 0 \tag{7.68b}$$

In order for this homogeneous set of equations to have a solution for A_1 and A_2, the determinant must vanish. By equating the determinant to zero, we obtain the eigenvalue equation for the resonant wave number k_0 or

$k = k_0\sqrt{\epsilon_r}$. This equation is

$$(\eta^2 + j\eta x - 1)\sin x - \left[(\eta^2 - 1)x + j\eta x^2 - \tfrac{1}{3}x^3\right]\cos x = 0 \quad (7.69)$$

where we have let $\epsilon_r = \eta^2$ and $k_0 a = ka/\eta = x/\eta$. Since the roots of (7.69) are complex, the solution for x is not straightforward to carry out. An iterative method can be used to advantage. Since the dielectric constant is normally quite large, the approximate solution of (7.69) is obtained by keeping only the terms multiplied by η^2. Thus we have

$$\sin x - x\cos x = 0$$

which has the solution $x = 4.4934 = x_0$. If we denote the eigenvalue equation (7.69) as $f(x)$, then a Taylor series expansion about the point x_0 gives

$$f(x) = f(x_0) + \left.\frac{df}{dx}\right|_{x_0}(x - x_0)$$

We can equate $f(x)$ to zero and then get the following improved solution

$$x = x_0 - \frac{f(x_0)}{df/dx_0}$$

$$= x_0 + \frac{(j\eta x_0^2 - \frac{1}{3}x_0^3 - x_0)\cos x_0 - (j\eta x_0 - 1)\sin x_0}{(\eta^2 x_0 + j\eta x_0^2 + j\eta - \frac{1}{3}x_0^3 - x_0)\sin x_0 - (j\eta x_0 - x_0^2)\cos x_0} \quad (7.70)$$

This second approximation can be called x_0 and used in (7.70) to obtain a third-order approximation. However, the second-order solution is usually quite accurate as the following example illustrates.

If the dielectric constant $\epsilon_r = 49$, then the second-order solution obtained using (7.70) is $ka = x = x_0 - 0.02619 + j0.019497 = 4.4672 + j0.019497$. The complex resonant wave number is given by $k_0 = x/\eta a$. Since we have chosen η as real, the resonant frequency is complex only because the oscillating resonator loses energy by radiation. The radiation Q_r is given by

$$Q_r = \frac{\text{Re } ka}{2\,\text{Im } ka} \quad (7.71)$$

and for our example equals 114.56. If the dielectric loss tangent were no greater than 0.001, the dielectric losses would reduce the Q by about 10 percent only; so the radiation loss is the major loss. By enclosing the resonator in a shielding box, a much higher unloaded Q can be obtained. However, when a shielding box is used, care must be exercised to avoid exciting cavity modes in the box.

For a dielectric constant of 49 and a resonant frequency of 10 GHz, the radius of the hemispherical resonator is $a = 4.4672/\eta k_0 = (4.4672/2\pi\eta)\lambda_0 = 0.1016\lambda_0$ or 0.3047 cm. Thus the resonator is very compact. The parameter $k_0 a = 0.638$ is quite small, so that the electromag-

netic field outside the resonator is approximately quasistatic. The constant A_2 can be expressed in terms of A_1 using (7.66a) and approximating $ka = x$ by 4.4672, that is, neglecting the imaginary part. Outside the resonator the electric and magnetic fields are given by

$$E_\phi = \frac{3}{2}A_2 k_0 \sin 2\theta \left(\frac{3j}{k_0^3 r^3} - \frac{3}{k_0^2 r^2} - \frac{j}{k_0 r} \right) e^{-jk_0 r} \tag{7.72a}$$

$$H_r = j\frac{3}{2}A_2 k_0 Y_0 (3\cos 2\theta + 1) \left(\frac{3j}{k_0^4 r^4} - \frac{3}{k_0^3 r^3} - \frac{j}{k_0^2 r^2} \right) e^{-jk_0 r} \tag{7.72b}$$

$$H_\theta = j\frac{3}{2}A_2 k_0 Y_0 \sin 2\theta \left(\frac{6j}{k_0^4 r^4} - \frac{6}{k_0^3 r^3} - \frac{3j}{k_0^2 r^2} + \frac{1}{k_0 r} \right) e^{-jk_0 r} \tag{7.72c}$$

The terms in E_ϕ and H_θ that vary like $1/k_0 r$ represent the radiation field and account for the power loss due to radiation.

If a cover plate at a height c above the ground plane is installed, the radiation field will be suppressed provided $c < \lambda_0/2$. Since E_ϕ must now vanish at $z = 0$ and $z = c$, the field outside the resonator can be expressed in terms of cylindrical waves. The lowest-order wave will have the form

$$\sin \frac{\pi z}{c} H_0^2 \left(\rho \sqrt{k_0^2 - \left(\frac{\pi}{c} \right)^2} \right)$$

where ρ is the radial coordinate $\sqrt{x^2 + y^2}$ and H_0^2 is the cylinder Hankel function. When ρ is large

$$H_0^2 \left(\rho \sqrt{k_0^2 - \left(\frac{\pi}{c} \right)^2} \right) \sim \left[\frac{2}{\pi \rho \sqrt{k_0^2 - (\pi/c)^2}} \right]^{1/2} e^{-j\rho\sqrt{k_0^2 - (\pi/c)^2} + j\pi/4}$$

This is a nonpropagating wave when $k_0 < \pi/c$ or $c < \lambda_0/2$. The cover plate will not have much influence on the field close to the resonator since the terms that vary like $1/k_0^2 r^2$ and higher inverse powers of $k_0 r$ will be small at the location of the cover plate when c is several times larger than a. Thus when a cover plate is used, we can approximate the fields by (7.72) but dropping the radiation terms. When the radiation from the resonator is suppressed, the unloaded Q will be much larger since the losses are now those arising from the loss in the dielectric and from the currents on the ground plane that are associated with the resonating mode of the resonator.

The magnetic field from the resonator will link the nearby microstrip line as illustrated in Fig. 7.22. As a result, an oscillating resonator will excite propagating waves on the microstrip line. These waves represent a flow of power and produce the external loading on the resonator. The resonator can be represented by a parallel resonant circuit that is coupled to the microstrip line by means of an ideal transformer with a turns ratio n:1 as shown in Fig. 7.23. The resistor R must have a value that will make

FIGURE 7.22
Magnetic coupling between a dielectric resonator and a microstrip line.

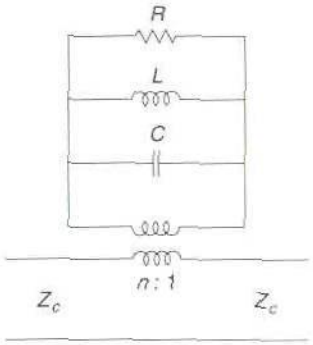


FIGURE 7.23
Equivalent circuit for a dielectric resonator coupled to a microstrip line.

ωCR equal to the unloaded Q, and L must have a value such that $\omega^2 LC = 1$, at the resonant frequency. These two conditions do not establish the impedance level or absolute value of R. We can choose R arbitrarily as long as the turns ratio n:1 is adjusted to give the correct coupling between the resonator and the microstrip line. At resonance a series resistance n^2R is coupled into the transmission line, and this makes the input reflection coefficient equal to

$$\Gamma = \frac{(n^2R + Z_c) - Z_c}{(n^2R + Z_c) + Z_c} = \frac{n^2R}{n^2R + 2Z_c}$$

when the output line is terminated in a matched load. By measuring Γ the parameter

$$\frac{n^2R}{Z_c} = \frac{2\Gamma}{1 - \Gamma}$$

is readily determined and this provides the additional information that is needed to specify the equivalent circuit. In some cases the turns ratio can also be evaluated analytically but in general it is preferable to measure it.†

A number of papers have provided analysis of the cylindrical dielectric resonator. In Fig. 7.24 we show the dominant TE_{012}-mode eigenvalue kR as a function of the ratio of the resonator height H to the radius R for a cylindrical resonator mounted on a ground plane. When the dielectric constant is greater than 20, the eigenvalue kR is essentially independent of the dielectric constant. This independence was also found for the hemi-

†Y. Komatsu and Y. Murakami, Coupling Coefficient Between Microstrip Line and dielectric Resonator, *IEEE Trans.*, vol. MTT-31, pp. 34–40, 1983.

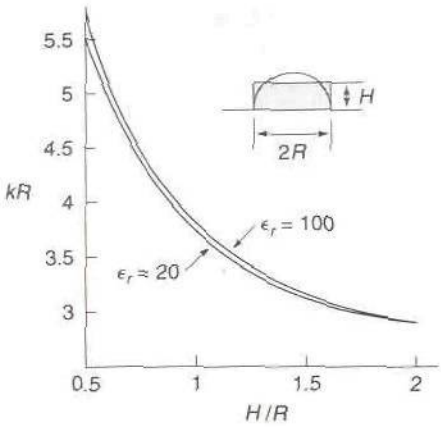


FIGURE 7.24
The eigenvalue kR for the TE_{012} mode for a cylindrical resonator on a ground plane. (*Based on Fig. 3 in R. E. Collin and D. Ksienski, Boundary Element Method for Dielectric Resonators and Waveguides, Rad. Sci., vol. 22, pp. 1155–1167, 1987.*)

spherical resonator since, to a first approximation, (7.69) gave $ka = 4.4934$ independent of ϵ_r. For a given cylindrical resonator, there exists an equivalent hemispherical resonator that will have the same resonant frequency. If we let $kR = x_c$ for the dielectric resonator and let $ka = x_s$ for the hemispherical resonator, then since k is the same we must have

$$\frac{a}{R} = \frac{x_s}{x_c}$$

From Fig. 7.24 we find that a dielectric resonator with radius R equal to a must have a height H equal to $0.7R$ in order to have the same resonant frequency as a hemispherical resonator with the same radius. The inset in Fig. 7.24 shows the two equivalent resonator cross sections and clearly they are very similar.

When a dielectric resonator is mounted on the substrate material instead of on a ground plane, the resonant frequency will change by a small amount. The electromagnetic field around the resonator will also change but will remain similar to that for a resonator on a ground plane. The quasistatic field outside the resonator is similar to that of a magnetic quadrapole. For a cylindrical or spherical resonator in free space, the

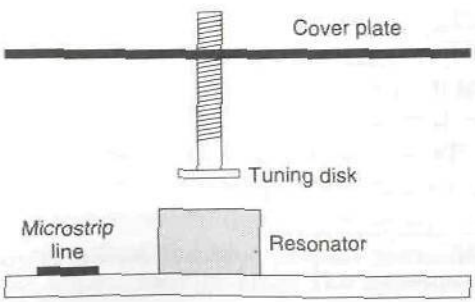


FIGURE 7.25
A method for tuning a dielectric resonator. The position of the metal disk relative to the end face of the resonator is adjustable.

dominant mode has a local field similar to that of an oscillating magnetic dipole. The image of a magnetic dipole in a ground plane, along with the original magnetic dipole, creates a magnetic quadrapole. Thus, when a resonator is mounted close to a ground plane, there is a significant change in the field around the resonator.

The resonant frequency of a dielectric resonator in a shielded enclosure can be changed by a small amount by varying the position of a small metal disk relative to the end face of the resonator as shown in Fig. 7.25. When dielectric resonators are used as part of a filter, some form of tuning for adjusting the resonant frequency is normally required.

7.6 EQUIVALENT CIRCUITS FOR CAVITIES

In this section the equivalent circuits of cavities coupled to transmission lines and waveguides are examined. A complete treatment is not given because of the complexity of the problem in general (see Sec. 7.9 for further details). Instead, a number of specific examples are examined in order to indicate the type of results that are obtained.

Aperture-Coupled Cavity

As an example of an aperture-coupled cavity, consider the rectangular cavity coupled to a rectangular guide by means of a small centered circular hole in the transverse wall at $z = 0$, as illustrated in Fig. 7.26. As indicated earlier, a small circular aperture in a transverse wall behaves as a shunt inductive susceptance with a normalized value given by (5.41) as

$$\bar{B}_L = \frac{3ab}{8\beta r_0^3} \tag{7.73}$$

where a is the guide width, b is the guide height, r_0 is the aperture radius, and $\beta = [k_0^2 - (\pi/a)^2]^{1/2}$ is the propagation constant for the TE_{10} mode. The equivalent circuit of the aperture-coupled cavity is thus a short-circuited transmission line of length d shunted by a normalized susceptance $\bar{B}_L$.

To analyze this coupled cavity, we shall assume initially that there are no losses. The modifications required when small losses are present are given later. The cavity will exhibit an infinite number of resonances, and the input impedance $\bar{Z}_{in}$ will have an infinite number of zeros interlaced by an infinite number of poles, this being the general behavior of a distributed-parameter one-port microwave network. If we are interested in a resonance corresponding to a high value of $\bar{Z}_{in}$, infinite in the case of no loss, we should examine the nature of $\bar{Z}_{in}$ in the vicinity of one of its poles. This case corresponds to a parallel resonant circuit.

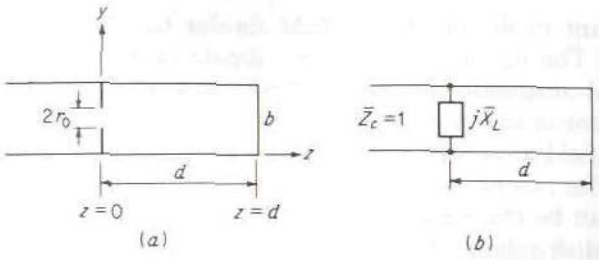


FIGURE 7.26
(*a*) Aperture-coupled rectangular cavity; (*b*) equivalent transmission-line circuit.

The input impedance is given by the parallel impedance of $j\bar{X}_L$ and $j \tan \beta d$ and is

$$\bar{Z}_{\text{in}} = \frac{-\bar{X}_L \tan \beta d}{j\bar{X}_L + j \tan \beta d} \tag{7.74}$$

where $j\bar{X}_L = (-j\bar{B}_L)^{-1}$. The antiresonances occur when the denominator vanishes, i.e., at the poles of $\bar{Z}_{\text{in}}$, or when

$$\bar{X}_L = -\tan \beta d = \frac{8r_0^3 \beta d}{3abd} \tag{7.75}$$

To solve this equation for the values of β that yield resonances, graphical methods are convenient. By plotting the two sides of (7.75) as functions of βd, the points of intersection yield the solutions for βd. When β is known, the resonant frequency may be found from the relation

$$\frac{\omega}{2\pi} = f = \frac{c}{2\pi}\left[\beta^2 + \left(\frac{\pi}{a}\right)^2\right]^{1/2} \tag{7.76}$$

This graphical solution is illustrated in Fig. 7.27. Note that there are an infinite number of solutions. Normally, $\bar{X}_L$ is very small, so that the value of βd for the fundamental mode is approximately equal to π. The higher-order modes occur for $\beta d = \beta_n d \approx (n - \frac{1}{2})\pi$ when n, an integer, is large.

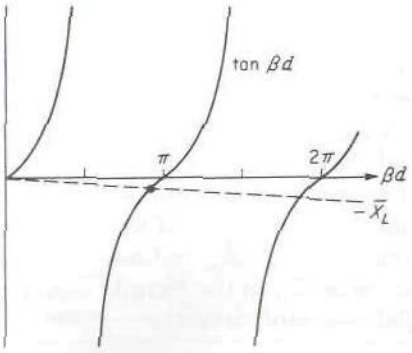


FIGURE 7.27
Graphical solution for resonant frequency of aperture-coupled cavity.

The value of β for the first mode will be denoted by β_1, and the corresponding value of ω by ω_1, as determined by putting $\beta = \beta_1$ in (7.76).

An infinite number of equivalent lumped-parameter networks can be used to represent $\bar{Z}_{in}$ in the vicinity of ω_1. Usually, the simplest possible network is used. This equivalent network must be chosen so that its input impedance $\bar{Z}$ equals $\bar{Z}_{\text{in}}$ at ω_1. Likewise, for small variations $\Delta\omega$ about ω_1, the two impedances must be equal. A general procedure for specifying this equivalence is obtained by expanding the impedance functions in a power series in $\omega - \omega_1 = \Delta\omega$ about ω_1 and equating these series term by term. Since $\bar{Z}_{\text{in}}$ has a pole at $\omega = \omega_1$, the Taylor expansion cannot be applied directly to $\bar{Z}_{\text{in}}$. However, it may be applied to $(\omega - \omega_1)\bar{Z}_{\text{in}}(\omega)$ to give

$$(\omega - \omega_1)\bar{Z}_{\text{in}}(\omega) = \lim_{\omega \to \omega_1} (\omega - \omega_1)\bar{Z}_{\text{in}}(\omega) + \frac{d}{d\omega}(\omega - \omega_1)\bar{Z}_{\text{in}}\Big|_{\omega_1} (\omega - \omega_1) + \frac{1}{2}\frac{d^2}{d\omega^2}(\omega - \omega_1)\bar{Z}_{\text{in}}\Big|_{\omega_1} (\omega - \omega_1)^2 + \cdots$$

We now obtain

$$\bar{Z}_{\text{in}}(\omega) = \frac{\lim_{\omega \to \omega_1}(\omega - \omega_1)\bar{Z}_{\text{in}}(\omega)}{\omega - \omega_1} + \frac{d}{d\omega}(\omega - \omega_1)\bar{Z}_{\text{in}}\Big|_{\omega_1} + \cdots \quad (7.77)$$

A similar expansion of $\bar{Z}$ gives (note that $\bar{Z}$ must have a pole at ω_1 also)

$$\bar{Z}(\omega) = \frac{\lim_{\omega \to \omega_1}(\omega - \omega_1)\bar{Z}(\omega)}{\omega - \omega_1} + \frac{d}{d\omega}(\omega - \omega_1)\bar{Z}\Big|_{\omega_1} + \cdots \quad (7.78)$$

Expansions of this type are called Laurent series expansions, and the coefficient of the $(\omega - \omega_1)^{-1}$ term is called the residue at the pole ω_1. These two series must be made equal term for term up to the highest power in $\omega - \omega_1 = \Delta\omega$ required to represent $\bar{Z}_{\text{in}}$ with sufficient accuracy in the frequency range of interest. For a microwave cavity, the Q is usually so high, and the frequency range $\Delta\omega/\omega_1$ of interest is approximately the range between the two points where $|\bar{Z}_{\text{in}}|$ equals 0.707 of its maximum. This latter fractional frequency band is equal to $1/Q$, and hence $\Delta\omega/\omega$ is so small that normally only the first term in the expansion (7.77) would be required to represent $\bar{Z}_{\text{in}}$ with sufficient accuracy in the vicinity of ω_1. In the present case a simple parallel LC circuit would be sufficient to represent $\bar{Z}_{\text{in}}(\omega)$ around ω_1.

In order to specify the values of L and C, we must evaluate the first terms in (7.77) and (7.78). For the LC circuit we have

$$\bar{Z} = \frac{j\omega L}{1 - \omega^2 LC}$$

We now choose $\omega_1^2 LC = 1$ to produce a pole at ω_1 for $\bar{Z}$. Hence

$$\bar{Z} = \frac{j\omega\omega_1^2 L}{\omega_1^2 - \omega^2}$$

and

$$\lim_{\omega \to \omega_1} (\omega - \omega_1)\bar{Z} = \left.\frac{-j\omega\omega_1^2 L}{\omega + \omega_1}\right|_{\omega_1} = \frac{-j\omega_1^2 L}{2}$$

Thus we have

$$\bar{Z}(\omega) = \frac{-j\omega_1^2 L}{2(\omega - \omega_1)} \tag{7.79}$$

for ω near ω_1.

To evaluate the behavior of $\bar{Z}_{\text{in}}$ near ω_1, we can place ω equal to ω_1 in the numerator in (7.74). The denominator is first expanded in a Taylor series in β about β_1 to give

$$\bar{X}_L(\beta) + \tan \beta d \approx \bar{X}_L(\beta_1) + \tan \beta_1 d + \left.\left(\frac{d\bar{X}_L}{d\beta} + \frac{d \tan \beta d}{d\beta}\right)\right|_{\beta_1} (\beta - \beta_1)$$

$$= \left(\frac{\bar{X}_{L1}}{\beta_1} + d \sec^2 \beta_1 d\right)(\beta - \beta_1)$$

since $\bar{X}_L(\beta_1) = \bar{X}_{L1} = -\tan \beta_1 d$ and $d\bar{X}_L/d\beta = (1/\beta)\bar{X}_L$. Next we expand β in terms of ω about ω_1 to give

$$\beta \approx \beta_1 + \left.\frac{d\beta}{d\omega}\right|_{\omega_1} (\omega - \omega_1)$$

If we denote $d\beta/d\omega$ at ω_1 by β_1', we see that $\bar{Z}_{\text{in}}$ can be expressed as

$$\bar{Z}_{\text{in}} = \frac{j\bar{X}_{L1} \tan \beta_1 d}{\left[\bar{X}_{L1} + \beta_1 d(1 + \tan^2 \beta_1 d)\right](\beta_1'/\beta_1)(\omega - \omega_1)}$$

upon replacing $\sec^2 \beta_1 d$ by $1 + \tan^2 \beta_1 d$. Replacing $\tan \beta_1 d$ by $-\bar{X}_{L1}$ now gives

$$\bar{Z}_{\text{in}} = -j\frac{\bar{X}_{L1}^2}{\left[\bar{X}_{L1} + \beta_1 d(1 + \bar{X}_{L1}^2)\right](\beta_1'/\beta_1)(\omega - \omega_1)}$$

Normally, $\bar{X}_{L1} \ll 1$, and since $\beta_1 d \approx \pi$, we can make further approximations to obtain (we shall verify that $\bar{X}_{L1} \ll 1$ later)

$$\bar{Z}_{\text{in}} = -j\frac{\bar{X}_{L1}^2}{\beta_1' d(\omega - \omega_1)} \tag{7.80}$$

Comparison with (7.79) shows that we must choose

$$\frac{\omega_1^2 L}{2} = \frac{\bar{X}_{L1}^2}{\beta_1' d}$$

or

$$L = \frac{2\bar{X}_{L1}^2}{\omega_1^2 \beta_1' d} \tag{7.81}$$

The capacitance C is determined by the condition $\omega_1^2 LC = 1$ given earlier.

Up to this point we have neglected the losses in the cavity. For a high-Q cavity these may be accounted for simply by replacing the resonant frequency ω_1 by a complex resonant frequency $\omega_1(1 + j/2Q)$, as indicated in Sec. 7.1. That is, the natural response of a lossy cavity is proportional to $e^{-\delta t + j\omega_1 t}$, and not to $e^{j\omega_1 t}$, where $\delta = \omega_1/2Q$. This is equivalent to having a complex resonant frequency $\omega_1(1 + j/2Q)$. The field in the cavity is, apart from some local fringing because of the presence of the aperture, a TE_{101} mode. Its Q was evaluated in Sec. 7.4, and is given by (7.48). For the lossy case we then have

$$\bar{Z}_{\text{in}} \approx -j \frac{\bar{X}_{L1}^2}{\beta_1' d(\omega - \omega_1 - j\omega_1/2Q)} \tag{7.82}$$

At resonance ($\omega = \omega_1$), we now obtain a pure resistive impedance $\bar{R}_{\text{in}}$ given by

$$\bar{R}_{\text{in}} = \bar{Z}_{\text{in}} = \frac{2\bar{X}_{L1}^2 Q}{\omega_1 \beta_1' d} \tag{7.83}$$

If we want the cavity to be matched to the waveguide at resonance, we must choose the aperture reactance $\bar{X}_{L1}$ so that $\bar{R}_{\text{in}} = 1$; that is,

$$\bar{X}_{L1} = \left(\frac{\omega_1 \beta_1' d}{2Q}\right)^{1/2} \tag{7.84}$$

This matched condition is referred to as critical coupling. If $\bar{R}_{\text{in}}$ is greater than the characteristic impedance of the input line (unity in the case of normalized impedances), the cavity is said to be overcoupled, whereas if $\bar{R}_{\text{in}}$ is smaller, the cavity is undercoupled. If $\bar{R}_{\text{in}}$ is the normalized input resistance at resonance for a parallel resonant cavity, then $\bar{R}_{\text{in}}$ is defined as the coupling parameter K. In the case of series resonance, the coupling parameter equals the input normalized conductance at resonance.

For the rectangular cavity discussed in Sec. 7.4, with $a = b = d = 3$ cm, we found $f_1 = 7{,}070$ MHz and $Q = 12{,}700$. For this cavity

$$\beta_1' = \frac{\omega_1}{\beta_1 c^2} \approx 4.7 \times 10^{-11} \text{ s/cm}$$

and (7.84) gives $\bar{X}_{L1} \approx 0.0157$ for critical coupling. The corresponding

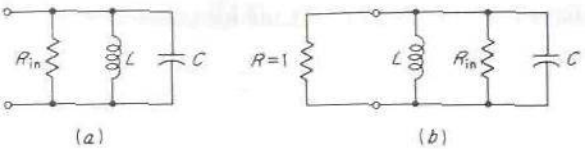


FIGURE 7.28
Equivalent circuit for aperture-coupled cavity.

aperture radius r_0 from (7.73) is found to be 0.37 cm. Note that $\bar{X}_{L1} \ll 1$, so that our earlier approximation in neglecting $\bar{X}_{L1}$ compared with unity is justified. Also note that a solution of (7.84) for the required value of $\bar{X}_{L1}$ to give critical coupling must, in general, be carried out simultaneously with the solution of (7.75) for the resonant frequency ω_1. However, for a high-Q cavity, ω_1 may be approximated with negligible error by the frequency corresponding to $\beta d = \pi$ in (7.84). This was done in the above calculation.

For the lossy cavity the equivalent circuit must include a resistance $\bar{R}_{in}$ in parallel with L and C as illustrated in Fig. 7.28*a*. The reader may readily verify that the input impedance $\bar{Z}$ now becomes, for ω near ω_1,

$$\bar{Z} = -j\frac{\omega_1^2 L}{2(\omega - \omega_1 - j\omega_1/2Q)} \tag{7.85}$$

where $Q = \bar{R}_{in}/\omega_1 L$.

Since the cavity is coupled to an input waveguide, the cavity terminals are loaded by an impedance equal to the impedance seen looking toward the generator from the aperture plane. If the generator is matched to the waveguide, a normalized resistance of unity is connected across the cavity terminals, as in Fig. 7.28*b*. The external Q_e is given by (Sec. 7.1)

$$Q_e = \frac{1}{\omega_1 L} \tag{7.86a}$$

and the loaded Q_L by

$$Q_L = \left(\frac{1}{Q} + \frac{1}{Q_e}\right)^{-1} = \frac{\bar{R}_{in}}{(1 + \bar{R}_{in})\omega_1 L} \tag{7.86b}$$

The loaded and unloaded quality factors are related as follows:

$$\frac{1}{Q_L} = \frac{1}{Q} + \frac{1}{Q_e} = \frac{(1 + \bar{R}_{in})\omega_1 L}{\bar{R}_{in}} = \omega_1 L + \frac{\omega_1 L}{\bar{R}_{in}} = \frac{1}{Q}(1 + K) \tag{7.87}$$

or

$$Q = (1 + K)Q_L$$

In general, the coupling parameter K may be defined as

$$K = \frac{Q}{Q_e} \tag{7.88}$$

For a parallel resonant circuit this is seen to give

$$K = \frac{\overline{R}_{in}}{\omega_1 L}\frac{\omega_1 L}{1}$$

which agrees with the earlier definition. Likewise, for a series resonant circuit with normalized input conductance $\overline{G}_{in}$, the unloaded and loaded Q's are given by $Q = \omega_1 L \overline{G}_{in}$, $Q_e = \omega_1 L$, and hence $K = \overline{G}_{in} = Q/Q_e$ again. The coupling parameter is a measure of the degree of coupling between the cavity and the input waveguide or transmission line.

The external Q, Q_e, is sometimes called the radiation Q. The reason for this is that the cavity may be considered to radiate power through the aperture into the input waveguide. This power loss by radiation through the aperture is equal to the power lost in the normalized unit resistance connected across the resonator terminals in the equivalent circuit illustrated in Fig. 7.28*b*.

For the rectangular cavity under discussion, the next-higher frequency at which a resonance occurs corresponds to a series-type resonance at which $|\overline{Z}_{in}|$ is a minimum. In the loss-free case, $\overline{Z}_{in} = 0$, and from (7.74) this is seen to correspond to $\tan \beta d = 0$, or $\beta d = \pi$. At a series resonance, $\overline{Y}_{in}$ has a pole, and consequently an analysis similar to that presented for $\overline{Z}_{in}$ is applied to $\overline{Y}_{in}$. It is readily found that in the vicinity of $\omega = \omega_2$, where ω_2 is the value of ω that makes $\beta d = \pi$,

$$\overline{Y}_{in} = -j\frac{1}{\beta_2'(\omega - \omega_2 - j\omega_2/2Q)}$$

In this case the aperture has no effect, as would be expected, because the standing-wave pattern in the cavity is now such that the transverse electric field is zero at the aperture plane. The input admittance near ω_2 is just that of a short-circuited guide near a half guide wavelength long. This resonance is not of any practical interest since it corresponds to a very loosely coupled cavity.

Loop-Coupled Cavity

Figure 7.29 illustrates a cavity that is coupled to a coaxial line by means of a small loop. Since the loop is very small, the current in the loop can be considered to be constant. Any mode in the cavity that has a magnetic field

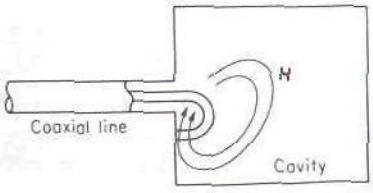


FIGURE 7.29
Loop-coupled cavity.

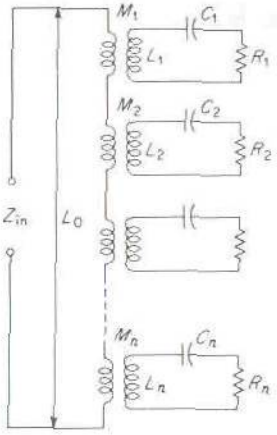


FIGURE 7.30
Equivalent circuit of loop-coupled cavity.

with flux lines that thread through the loop will be coupled by the loop. However, at any particular frequency ω, only that mode which is resonant at this frequency will be excited with an appreciable amplitude. The fields excited in the cavity by the current I flowing in the loop can be found by solving for a vector potential arising from the current I. From the vector potential the magnetic field, and hence the flux passing through the loop, may be found. For a unit current let the magnetic flux of the nth mode that threads through the loop be ψ_n. This is then equal to the mutual inductance M_n between the coupling loop and the nth mode. Each mode presents an impedance equivalent to that of a series LCR circuit to the coupling loop. Thus a suitable equivalent circuit is an infinite number of series LCR circuits coupled by mutual inductance to the input coaxial line, as illustrated in Fig. 7.30. The input impedance is thus of the form

$$Z_{\text{in}} = j\omega L_0 + j\sum_{n=1}^{\infty} \frac{\omega^3 M_n^2 C_n}{1 - \omega^2 L_n C_n + j\omega C_n R_n} \tag{7.89}$$

where L_0 is the self-inductance of the coupling loop. If we define the resonance frequencies ω_n by $\omega_n^2 L_n C_n = 1$ and the unloaded Q of the nth mode by $Q_n = \omega_n L_n / R_n$, we can rewrite (7.89) as

$$Z_{\text{in}} = j\omega L_0 + j\sum_{n=1}^{\infty} \frac{\omega^3 M_n^2}{L_n(\omega_n^2 - \omega^2 + j\omega\omega_n/Q_n)} \tag{7.90}$$

If $\omega \approx \omega_n$, then all terms in the series in (7.90) except the nth term are small. Thus, in the vicinity of the nth resonance,

$$\begin{aligned} Z_{\text{in}} &= j\omega L_0 + \frac{j\omega^3 M_n^2}{L_n(\omega_n^2 - \omega^2 + j\omega\omega_n/Q_n)} \\ &\approx j\omega L_0 - j\frac{\omega_n^2 M_n^2}{2L_n(\omega - \omega_n - j\omega_n/2Q_n)} \end{aligned} \tag{7.91}$$

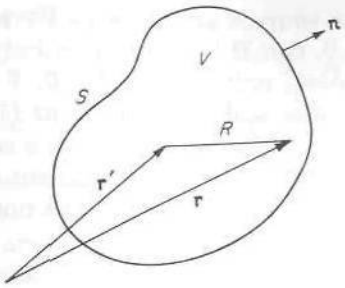


FIGURE 7.31
Illustration for Helmholtz's theorem.

The equivalent circuit now reduces to a single *LCR* series circuit mutually coupled to the input line. For efficient excitation of a given mode, the loop should be located at a point where this mode provides a maximum flux linkage.

The preceding results represent a formal solution to the loop-coupled cavity. In order actually to specify the circuit parameters, the boundary-value problem for the fields excited in a cavity by a given loop current must be solved. Also the Q's of the various cavity modes must be determined. For simple cavity shapes these calculations can be carried out with reasonable accuracy. However, they are too lengthy to include here.

⋆7.7 FIELD EXPANSION IN A GENERAL CAVITY

Considerable insight into the general properties of an electromagnetic cavity can be obtained by examining the problem of expanding an arbitrary electromagnetic field into a complete set of modes in a cavity of unspecified shape. Such general modal expansions are required in order to determine the fields excited in a cavity by an arbitrary source. These expansions are also required in the evaluation of the input impedance or admittance of cavities coupled to external transmission lines or waveguides.

A fundamental theorem which is basic to general cavity theory is Helmholtz's theorem.† This theorem states that in a volume V bounded by a closed surface S as in Fig. 7.31 a general vector field $\mathbf{P}(x, y, z)$ is given by

$$\mathbf{P}(x,y,z) = \nabla\left[\int_V \frac{-\nabla' \cdot \mathbf{P}(\mathbf{r}')}{4\pi R}\,dV' + \oint_S \frac{\mathbf{P}(\mathbf{r}') \cdot \mathbf{n}}{4\pi R}\,dS'\right] + \nabla \times \left[\int_V \frac{\nabla' \times \mathbf{P}(\mathbf{r}')}{4\pi R}\,dV' + \oint_S \frac{\mathbf{P}(\mathbf{r}') \times \mathbf{n}}{4\pi R}\,dS'\right] \quad (7.92)$$

where $R = |\mathbf{r} - \mathbf{r}'|$. Thus the volume sources for $\mathbf{P}$ are given by $-\nabla' \cdot \mathbf{P}$

†R. E. Collin, "Field Theory of Guided Waves," 2nd ed., Mathematical Appendix, IEEE Press, Piscataway, N. J., 1991

and $\nabla' \times \mathbf{P}$, and the surface sources are given by $\mathbf{P} \cdot \mathbf{n}$ and $\mathbf{P} \times \mathbf{n}$. If, and only if, $\nabla \cdot \mathbf{P} = 0$, $\mathbf{n} \cdot \mathbf{P} = 0$, can $\mathbf{P}$ be derived entirely from the curl of a suitable vector potential. Also, only if $\mathbf{n} \times \mathbf{P} = 0$, $\nabla \times \mathbf{P} = 0$, can $\mathbf{P}$ be derived from the gradient of a scalar potential, as (7.92) shows. If both $\nabla \cdot \mathbf{P} = 0$ and $\nabla \times \mathbf{P} = 0$ in V, then $\mathbf{P}$ is said to be a source-free field in V. In this case $\mathbf{P}$ can always be derived from the gradient of a scalar potential, but this potential must be multivalued if $\mathbf{n} \times \mathbf{P}$ does not equal zero, in some situations. This statement will be clarified later.

In setting up a suitable set of modes in which to expand a vector field $\mathbf{P}$ inside a given volume V, it is necessary to know the boundary conditions that must be imposed on these modes in order that a unique set of modes may be obtained. This uniqueness is needed so that when a solution for the field has been obtained we are assured of the uniqueness of that solution. Since the electromagnetic field satisfies Helmholtz's equation, we are concerned with obtaining a unique solution to this equation.

Consider $\nabla^2 \mathbf{A}_n + k_n^2 \mathbf{A}_n = 0$. We wish to determine the type of boundary conditions that must be imposed on $\mathbf{A}_n$ such that the solution to this equation is unique for a specific eigenvalue k_n. Assume that a second solution $\mathbf{B}_n$ exists such that $\nabla^2 \mathbf{B}_n + k_n^2 \mathbf{B}_n = 0$ also. The difference solution $\mathbf{C}_n = \mathbf{A}_n - \mathbf{B}_n$ satisfies $\nabla^2 \mathbf{C}_n + k_n^2 \mathbf{C}_n = 0$ also. We shall require of $\mathbf{B}_n$ the conditions $\nabla \cdot \mathbf{B}_n = \nabla \cdot \mathbf{A}_n$, $\nabla \times \mathbf{B}_n = \nabla \times \mathbf{A}_n$ in V, so that both $\mathbf{A}_n$ and $\mathbf{B}_n$ have the same volume sources. Then

$$\nabla \cdot \mathbf{C}_n = \nabla \times \mathbf{C}_n = 0$$

in V. Consider

$$k_n^2 \int_V |\mathbf{C}_n|^2 \, dV = -\int_V \mathbf{C}_n \cdot \nabla^2 \mathbf{C}_n \, dV$$

$$= \int_V (\mathbf{C}_n \cdot \nabla \times \nabla \times \mathbf{C}_n - \mathbf{C}_n \cdot \nabla \nabla \cdot \mathbf{C}_n) \, dV$$

Using the relations

$$\nabla \cdot (\mathbf{C}_n \times \nabla \times \mathbf{C}_n) = |\nabla \times \mathbf{C}_n|^2 - \mathbf{C}_n \cdot \nabla \times \nabla \times \mathbf{C}_n$$

and

$$\nabla \cdot (\mathbf{C}_n \nabla \cdot \mathbf{C}_n) = |\nabla \cdot \mathbf{C}_n|^2 + \mathbf{C}_n \cdot \nabla \nabla \cdot \mathbf{C}_n$$

we obtain

$$k_n^2 \int_V |\mathbf{C}_n|^2 \, dV = \int_V |\nabla \times \mathbf{C}_n|^2 \, dV + \int_V |\nabla \cdot \mathbf{C}_n|^2 \, dV$$

$$- \oint_S \mathbf{n} \cdot (\mathbf{C}_n \nabla \cdot \mathbf{C}_n + \mathbf{C}_n \times \nabla \times \mathbf{C}_n) \, dS \qquad (7.93)$$

The volume integrals on the right vanish, and to make $\int_V |\mathbf{C}_n|^2 \, dV = 0$, which implies $\mathbf{C}_n = 0$, we must make the two surfaces integrals vanish. This may be accomplished by making $\mathbf{A}_n$ and $\mathbf{B}_n$ satisfy the boundary

conditions

$$\nabla \cdot \mathbf{A}_n = \mathbf{n} \times \mathbf{A}_n = 0 \qquad \text{on } S \tag{7.94a}$$

or

$$\mathbf{n} \cdot \mathbf{A}_n = \mathbf{n} \times \nabla \times \mathbf{A}_n = 0 \qquad \text{on } S \tag{7.94b}$$

and similarly for $\mathbf{B}_n$. These are the boundary conditions that will be used for the electric- and magnetic-type modes, respectively. Other possibilities are $\mathbf{n} \cdot \mathbf{A}_n = \mathbf{n} \times \mathbf{A}_n = 0$ on S or $\nabla \cdot \mathbf{A}_n = \mathbf{n} \times \nabla \times \mathbf{A}_n = 0$ on S. The imposition of these boundary conditions gives a unique solution. Note that two conditions must be specified for a vector function.

Cavity Field Expansions in Terms of Short-Circuit Modes

The short-circuit modes are those corresponding to the field solutions inside ideal (perfectly conducting walls), totally closed cavities. For these modes the electric field modes $\mathbf{E}_n$ satisfy the boundary condition $\mathbf{n} \times \mathbf{E}_n = 0$ on S, and the magnetic field modes satisfy the boundary condition $\mathbf{n} \cdot \mathbf{H}_n = 0$ on S. In addition, as (7.93) and (7.94) show, we must impose the additional boundary conditions $\nabla \cdot \mathbf{E}_n = 0$ and $\mathbf{n} \times \nabla \times \mathbf{H}_n = 0$ on S in order to obtain unique solutions.

Sometimes it is convenient to consider a cavity as having a surface part of which is perfectly conducting and part of which acts as a perfect open circuit. In this case the boundary conditions for $\mathbf{E}_n$ and $\mathbf{H}_n$ are interchanged on the perfect open-circuit portion of the boundary. In practice, it is difficult to solve the cavity problem when the same boundary conditions do not apply on the whole surface. For this reason we restrict our analysis to the use of short-circuit modes, since these are readily found for the common types of cavities encountered.

There are three basic types of cavities to be considered, as illustrated in Fig. 7.32. Type 1 is a simply connected volume with a single surface, whereas type 2 is a simply connected volume with a multiple (double) surface. Finally, type 3 is a multiply connected volume with a single surface. Examples of the latter type are toroids and short-circuited coaxial lines.

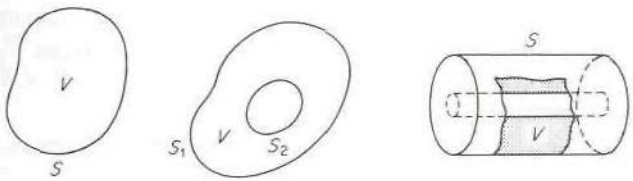


FIGURE 7.32
Basic types of cavities.

Consideration of these basic cavity types is of importance in connection with the existence of zero frequency modes, as will be seen.

Electric Field Expansion

In general, both $\nabla \times \mathbf{E}$ and $\nabla \cdot \mathbf{E}$ are nonzero; so we require two sets of modes: a solenoidal set with zero divergence and nonzero curl and an irrotational set with zero curl but nonzero divergence. The solenoidal modes are defined by

$$\begin{aligned} \nabla^2 \mathbf{E}_n + k_n^2 \mathbf{E}_n &= 0 \qquad \text{in } V \\ \mathbf{n} \times \mathbf{E}_n &= 0 \qquad \text{on } S \\ \nabla \cdot \mathbf{E}_n &= 0 \qquad \text{in } V \text{ and on } S \end{aligned} \tag{7.95}$$

and the irrotational modes by $l_n \mathbf{F}_n = \nabla \Phi_n$, where

$$\begin{aligned} \nabla^2 \Phi_n + l_n^2 \Phi_n &= 0 \qquad \text{in } V \\ \Phi_n &= \text{const on } S \qquad \text{that is, } \mathbf{n} \times \nabla \Phi_n = 0 \text{ on } S \\ \nabla \times \mathbf{F}_n &= 0 \qquad \text{in } V \end{aligned} \tag{7.96}$$

where k_n, l_n are eigenvalues for the problem. The constant is chosen as zero, except for the $n = 0$ mode, which has $l_0 = 0$. The above definition for $\mathbf{F}_n$ is chosen so that

$$\int_V \Phi_n^2 \, dV = -\frac{1}{l_n^2} \int_V \Phi_n \nabla^2 \Phi_n \, dV = -\frac{1}{l_n^2} \int_V (\nabla \cdot \Phi_n \nabla \Phi_n - \nabla \Phi_n \cdot \nabla \Phi_n) \, dV$$

$$= \int_V \mathbf{F}_n \cdot \mathbf{F}_n \, dV - \frac{1}{l_n^2} \oint_S \Phi_n \frac{\partial \Phi_n}{\partial n} \, dS = \int_V \mathbf{F}_n \cdot \mathbf{F}_n \, dV$$

Note that only one boundary condition is imposed on the $\mathbf{F}_n$ since scalar functions Φ_n are uniquely determined if either Φ_n or $\partial \Phi_n / \partial n$ is specified on S. The modes are assumed normalized, so that

$$\int_V \Phi_n^2 \, dV = \int_V \mathbf{F}_n \cdot \mathbf{F}_n \, dV = \int_V \mathbf{E}_n \cdot \mathbf{E}_n \, dV = 1 \tag{7.97}$$

For $n = 0$ the eigenvalues l_0, k_0 are assumed to be the zero eigenvalues. The corresponding eigenfunctions $\mathbf{E}_0, \mathbf{F}_0$ are the zero-frequency modes. We shall show that $\mathbf{E}_0$ cannot be distinguished from $\mathbf{F}_0$; so only the latter will be retained. We have $\nabla^2 \mathbf{E}_0 = 0 = \nabla \nabla \cdot \mathbf{E}_0 - \nabla \times \nabla \times \mathbf{E}_0$, or $\nabla \times \nabla \times \mathbf{E}_0 = 0$, since $\nabla \cdot \mathbf{E}_0 = 0$ by hypothesis. Thus

$$-\int_V \mathbf{E}_0 \cdot \nabla \times \nabla \times \mathbf{E}_0 \, dV = 0$$

$$= \int_V (\nabla \cdot \mathbf{E}_0 \times \nabla \times \mathbf{E}_0 - \nabla \times \mathbf{E}_0 \cdot \nabla \times \mathbf{E}_0) \, dV$$

This gives $\int_V |\nabla \times \mathbf{E}_0|^2 \, dV = \oint_S \mathbf{n} \cdot \mathbf{E}_0 \times \nabla \times \mathbf{E}_0 \, dS = 0$ since $\mathbf{n} \times \mathbf{E}_0 = 0$. Thus we must have $\nabla \times \mathbf{E}_0 = 0$, which implies $\mathbf{E}_0 = \nabla f$, where f is a scalar function which is constant on the surface S. This latter relation holds for $\mathbf{F}_0$ as well, so that $\mathbf{E}_0$ can be discarded as long as $\mathbf{F}_0$ is retained. The zero-frequency mode $\mathbf{F}_0$ will be defined by

$$\begin{aligned} \mathbf{F}_0 &= \nabla \Phi_0 \\ \nabla \times \mathbf{F}_0 &= 0 \\ \mathbf{n} \times \mathbf{F}_0 &= 0 \qquad \text{on } S \qquad \text{that is, } \Phi_0 = \text{const on } S \\ \nabla^2 \Phi_0 &= 0 \qquad \text{since } l_0 = 0 \end{aligned} \tag{7.98}$$

For cavity types 1 and 3, a solution for Φ_0 other than a constant does not exist. Hence the $\mathbf{F}_0$ mode is present only for the type 2 cavity, with Φ_0 having different constant values on S_1 and S_2. This mode is just the static electric field that may exist between two conducting bodies at different potentials.

Orthogonality Properties

Nondegenerate $\mathbf{E}_n$ and $\mathbf{F}_n$ modes are orthogonal among themselves and with each other. Consider

$$\begin{aligned} &\int_V (\mathbf{E}_m \nabla^2 \mathbf{E}_n - \mathbf{E}_n \nabla^2 \mathbf{E}_m) \, dV \\ &= (k_m^2 - k_n^2) \int_V \mathbf{E}_n \cdot \mathbf{E}_m \, dV \\ &= \int_V (\mathbf{E}_n \cdot \nabla \times \nabla \times \mathbf{E}_m - \mathbf{E}_m \cdot \nabla \times \nabla \times \mathbf{E}_n) \, dV \\ &= \int_V [\nabla \times \mathbf{E}_n \cdot \nabla \times \mathbf{E}_m - \nabla \times \mathbf{E}_m \cdot \nabla \times \mathbf{E}_n \\ &\qquad - \nabla \cdot (\mathbf{E}_n \times \nabla \times \mathbf{E}_m - \mathbf{E}_m \times \nabla \times \mathbf{E}_n)] \, dV \\ &= \oint_S (\mathbf{n} \times \mathbf{E}_m \cdot \nabla \times \mathbf{E}_n - \mathbf{n} \times \mathbf{E}_n \cdot \nabla \times \mathbf{E}_m) \, dS = 0 \end{aligned}$$

Hence, if $k_m^2 \neq k_n^2$,

$$\int_V \mathbf{E}_m \cdot \mathbf{E}_n \, dV = \delta_{nm} \tag{7.99}$$

where the Kronecker delta $\delta_{nm} = 0$ if $n \neq m$ and equals unity for $n = m$.

For the modes $\mathbf{F}_m$ we show first of all that the Φ_n are orthogonal. We have

$$\int_V (\Phi_n \nabla^2 \Phi_m - \Phi_m \nabla^2 \Phi_n)\, dV = (l_n^2 - l_m^2) \int_V \Phi_n \Phi_m\, dV$$

$$= \oint_S \left(\Phi_n \frac{\partial \Phi_m}{\partial n} - \Phi_m \frac{\partial \Phi_n}{\partial n} \right) dS = 0$$

so that

$$\int_V \Phi_n \Phi_m\, dV = 0 \qquad \text{when } l_n^2 \neq l_m^2$$

Next consider

$$\int_V \nabla \cdot \Phi_n \nabla \Phi_m\, dV = \oint_S \Phi_n \frac{\partial \Phi_m}{\partial n}\, dS = 0 = \int_V (\nabla \Phi_n \cdot \nabla \Phi_m + \Phi_n \nabla^2 \Phi_m)\, dV$$

$$= l_n l_m \int_V \mathbf{F}_n \cdot \mathbf{F}_m\, dV - l_m^2 \int_V \Phi_n \Phi_m\, dV$$

Since the latter integral is zero, we have

$$\int_V \mathbf{F}_n \cdot \mathbf{F}_m\, dV = \int_V \Phi_n \Phi_m\, dV = \delta_{nm} \tag{7.100}$$

for n and m not both equal to zero. If $m = 0$, $n \neq 0$, the proof of orthogonality still holds. However, the normalization for the $\mathbf{F}_0$ mode will be chosen as

$$\int_V \mathbf{F}_0 \cdot \mathbf{F}_0\, dV = \int_V |\nabla \Phi_0|^2\, dV = 1 \tag{7.101}$$

To show that the $\mathbf{E}_n$ and $\mathbf{F}_m$ modes are orthogonal, consider

$$\nabla \cdot \mathbf{F}_m \times \nabla \times \mathbf{E}_n = (\nabla \times \mathbf{F}_m) \cdot (\nabla \times \mathbf{E}_n) - \mathbf{F}_m \cdot \nabla \times \nabla \times \mathbf{E}_n$$

$$= -k_n^2 \mathbf{F}_m \cdot \mathbf{E}_n$$

since $\nabla \times \mathbf{F}_m = 0$, $\nabla \times \nabla \times \mathbf{E}_n = \nabla \nabla \cdot \mathbf{E}_n - \nabla^2 \mathbf{E}_n = k_n^2 \mathbf{E}_n$. We now obtain

$$k_n^2 \int_V \mathbf{F}_m \cdot \mathbf{E}_n\, dV = -\oint_S \mathbf{n} \cdot \mathbf{F}_m \times \nabla \times \mathbf{E}_n\, dS$$

$$= -\oint_S \mathbf{n} \times \mathbf{F}_m \cdot \nabla \times \mathbf{E}_n\, dS = 0 \tag{7.102}$$

For $n = m = 0$, the modes $\mathbf{F}_0$ and $\mathbf{E}_0$ are not orthogonal, being in fact identical. Since the $\mathbf{F}_m$ are orthogonal to the $\mathbf{E}_n$, it follows that both sets are needed, including $\mathbf{F}_0$ in general, to expand an arbitrary electric field.

Magnetic Field Expansion

To expand an arbitrary magnetic field (we consider the possibility of $\mathbf{n} \cdot \mathbf{H} \neq 0$ over a portion of the surface to exist), we shall set up a dual set of modes analogous to those used to expand the electric field. The solenoidal modes are defined by

$$\begin{aligned} \nabla^2 \mathbf{H}_n + k_n^2 \mathbf{H}_n &= 0 \\ \nabla \cdot \mathbf{H}_n &= 0 \\ \mathbf{n} \cdot \mathbf{H}_n &= 0 \qquad \text{on } S \\ \mathbf{n} \times \nabla \times \mathbf{H}_n &= 0 \qquad \text{on } S \end{aligned} \tag{7.103}$$

and the irrotational modes are defined by

$$\begin{aligned} p_n \mathbf{G}_n &= \nabla \psi_n \\ \nabla^2 \psi_n + p_n^2 \psi_n &= 0 \\ \frac{\partial \psi_n}{\partial n} &= 0 \qquad \text{on } S \\ \nabla \times \mathbf{G}_n &= 0 \end{aligned} \tag{7.104}$$

When $p_n = 0$, we have (assuming this occurs for $n = 0$) $\nabla \times \mathbf{G} = 0$, $\nabla^2 \psi_0 = \nabla \cdot \mathbf{G}_0 = 0$, $\mathbf{n} \cdot \mathbf{G}_0 = 0$. Helmholtz's theorem now states that $\mathbf{G}_0$ can be derived from the curl of a vector potential, say $\mathbf{G}_0 = \nabla \times \mathbf{A}_0$, where $\nabla \times \nabla \times \mathbf{A}_0 = 0$. Since $\mathbf{G}_0$ has also been assumed to be given by $\mathbf{G}_0 = \nabla \psi_0$, it follows that ψ_0 must be multivalued. The function $\mathbf{G}_0$ corresponding to a static magnetic field can exist in the type 3 cavity only. For example, in a short-circuited coaxial line,

$$\mathbf{H} = \frac{I\mathbf{a}_\theta}{2\pi r} = \nabla \frac{I\theta}{2\pi} = \frac{\mathbf{a}_\theta}{r} \frac{\partial}{\partial \theta} \frac{I\theta}{2\pi} = \frac{I\mathbf{a}_\theta}{2\pi r} = -\nabla \times \frac{I\mathbf{a}_z}{2\pi} \ln r$$

Note that the scalar potential $I\theta/2\pi$ is multivalued. If $k_n = 0$ for $n = 0$, then (7.103) gives $\nabla \times \nabla \times \mathbf{H}_0 = 0$. A volume integral of $\mathbf{H}_0 \cdot \nabla \times \nabla \times \mathbf{H}_0$ similar to that employed earlier in connection with $\mathbf{F}_0$ and $\mathbf{E}_0$ shows that $\nabla \times \mathbf{H}_0 = 0$. Hence $\mathbf{H}_0$ has the same properties as $\mathbf{G}_0$; that is, $\nabla \times \mathbf{H}_0 = \nabla \cdot \mathbf{H}_0 = 0$, $\mathbf{n} \cdot \mathbf{H}_0 = 0$ on S. Therefore we shall not retain the $\mathbf{H}_0$ mode, but keep the $\mathbf{G}_0$ mode for the zero-frequency mode.

Orthogonality Properties

We assume normalization according to the following:

$$\int_V \mathbf{H}_n \cdot \mathbf{H}_n \, dV = \int_V \mathbf{G}_n \cdot \mathbf{G}_n \, dV = \int_V \psi_n^2 \, dV = 1 \tag{7.105a}$$

$$\int_V \mathbf{G}_0 \cdot \mathbf{G}_0 \, dV = \int_V |\nabla \psi_0|^2 \, dV = 1 \tag{7.105b}$$

By methods paralleling that used for the electric-type modes, the following orthogonality properties may be derived:

$$\int_V \mathbf{H}_n \cdot \mathbf{H}_m \, dV = \int_V \mathbf{H}_n \cdot \mathbf{G}_m \, dV = \int_V \mathbf{G}_n \cdot \mathbf{G}_m \, dV = \int_V \psi_n \psi_m \, dV = 0 \tag{7.106}$$

for $n \neq m$. For $n = m$ we also have

$$\int_V \mathbf{H}_n \cdot \mathbf{G}_n \, dV = 0 \tag{7.107}$$

Relationship Between $\mathbf{E}_n$ and $\mathbf{H}_n$ Modes

The eigenvalues for both the $\mathbf{E}_n, \mathbf{H}_n$ modes were designated as k_n because they are, in fact, equal. Furthermore, we can show that

$$\nabla \times \mathbf{E}_n = k_n \mathbf{H}_n \qquad \nabla \times \mathbf{H}_n = k_n \mathbf{E}_n \tag{7.108}$$

The curl of the first relation gives

$$\nabla \times \nabla \times \mathbf{E}_n = \nabla \nabla \cdot \mathbf{E}_n - \nabla^2 \mathbf{E}_n = k_n \nabla \times \mathbf{H}_n = k_n^2 \mathbf{E}_n = -\nabla^2 \mathbf{E}_n$$

by using the second relation. Similarly,

$$\nabla \times \nabla \times \mathbf{H}_n = -\nabla^2 \mathbf{H}_n = k_n \nabla \times \mathbf{E}_n = k_n^2 \mathbf{H}_n$$

Hence (7.108) is consistent with the Helmholtz equation of which the $\mathbf{E}_n, \mathbf{H}_n$ are solutions. Furthermore, the boundary condition on $\mathbf{E}_n$ is $\mathbf{n} \times \mathbf{E}_n = 0$ on S. This implies $\mathbf{n} \times \nabla \times \mathbf{H}_n = 0$ on S, which is the boundary condition that has been imposed on the $\mathbf{H}_n$ functions. Also, we have $\mathbf{n} \cdot \mathbf{H}_n = 0$ on S, which implies $\mathbf{n} \cdot \nabla \times \mathbf{E}_n = 0$ on S. Now

$$\mathbf{n} \cdot \nabla \times \mathbf{E}_n = \mathbf{n} \cdot \left[\left(\nabla - \mathbf{n} \frac{\partial}{\partial n} \right) \times (\mathbf{E}_n - \mathbf{n}\mathbf{n} \cdot \mathbf{E}_n) + \left(\nabla - \mathbf{n} \frac{\partial}{\partial n} \right) \times \mathbf{n}\mathbf{n} \cdot \mathbf{E}_n + \mathbf{n} \frac{\partial}{\partial n} \times \mathbf{E}_n \right] = \mathbf{n} \cdot \nabla_t \times \mathbf{E}_{tn} = 0$$

since only the term $\nabla_t \times \mathbf{E}_{tn}$ is in the direction of $\mathbf{n}$. Here t denotes components tangent to the surface S. Since $\mathbf{E}_{tn} = 0$ on S, $\mathbf{n} \cdot \nabla \times \mathbf{E}_n = 0$ on S and is consistent with the boundary condition $\mathbf{n} \cdot \mathbf{H}_n = 0$. Thus (7.108) is the only possible relation between the $\mathbf{E}_n, \mathbf{H}_n$ modes. A volume integral of

$$\nabla \cdot \mathbf{E}_n \times \nabla \times \mathbf{E}_n = -\mathbf{E}_n \cdot \nabla \times \nabla \times \mathbf{E}_n + |\nabla \times \mathbf{E}_n|^2 = -k_n^2 (|\mathbf{E}_n|^2 - |\mathbf{H}_n|^2)$$

shows that the normalization of $\mathbf{H}_n$ is consistent with the normalization of the $\mathbf{E}_n$ also. It should now be apparent that the $\mathbf{E}_n, \mathbf{F}_n$ and $\mathbf{H}_n, \mathbf{G}_n$ modes have the properties that enable them to represent the electric and magnetic

fields, respectively. An arbitrary field would require an infinite sum of these modes for its expansion.

*7.8 OSCILLATIONS IN A SOURCE-FREE CAVITY

Consider a type 1 cavity with perfectly conducting walls and free of all currents and charges. We wish to determine the possible modes of oscillation. Let the fields in the cavity be expressed in terms of infinite series of the form

$$\mathbf{E} = \sum_n e_n(t)\mathbf{E}_n + \sum_n f_n(t)\mathbf{F}_n \tag{7.109a}$$

$$\mathbf{H} = \sum_n h_n(t)\mathbf{H}_n + \sum_n g_n(t)\mathbf{G}_n \tag{7.109b}$$

where e_n, f_n, h_n, and g_n are amplitude factors that are functions of time. Since $\mathbf{n} \times \mathbf{E} = \mathbf{n} \cdot \mathbf{H} = 0$ on S and the mode functions satisfy similar boundary conditions, the series are uniformly convergent and may be differentiated term by term. Maxwell's curl equations thus give

$$\nabla \times \mathbf{E} = \sum_n e_n \nabla \times \mathbf{E}_n = \sum_n e_n k_n \mathbf{H}_n = -\mu \sum_n \frac{\partial h_n}{\partial t}\mathbf{H}_n - \mu \sum_n \frac{\partial g_n}{\partial t}\mathbf{G}_n$$

$$\nabla \times \mathbf{H} = \sum_n h_n \nabla \times \mathbf{H}_n = \sum_n h_n k_n \mathbf{E}_n = \epsilon \sum_n \frac{\partial e_n}{\partial t}\mathbf{E}_n + \epsilon \sum_n \frac{\partial f_n}{\partial t}\mathbf{F}_n$$

If the first equation is scalar multiplied by $\mathbf{H}_n$ and $\mathbf{G}_n$ in turn and integrated over the cavity volume, we obtain

$$e_n k_n = -\mu \frac{\partial h_n}{\partial t} \tag{7.110a}$$

$$\frac{\partial g_n}{\partial t} = 0 \tag{7.110b}$$

by virtue of the orthogonality properties (7.106). Similarly, when the second equation is scalar multiplied by $\mathbf{E}_n$ and $\mathbf{F}_n$ in turn and integrated over the cavity volume, we obtain

$$h_n k_n = \epsilon \frac{\partial e_n}{\partial t} \tag{7.111a}$$

$$\frac{\partial f_n}{\partial t} = 0 \tag{7.111b}$$

From (7.110a) and (7.111a) we obtain

$$\frac{\partial^2 e_n}{\partial t^2} + \frac{k_n^2}{\mu\epsilon} e_n = 0 \qquad \frac{\partial^2 h_n}{\partial t^2} + \frac{k_n^2}{\mu\epsilon} h_n = 0$$

The solution for e_n is

$$e_n = e^{j\omega_n t} \tag{7.112a}$$

and from (7.111a) the solution for h_n is then

$$h_n = \frac{jk_n}{\omega_n \mu} e_n = j\sqrt{\frac{\epsilon}{\mu}}\, e_n = j\sqrt{\frac{\epsilon}{\mu}}\, e^{j\omega_n t} \tag{7.112b}$$

where $\omega_n = k_n(\mu\epsilon)^{-1/2}$ is the resonant frequency for the nth mode.

In the absence of volume sources, the $\mathbf{F}_n$ and $\mathbf{G}_n$ for $n \neq 0$ do not exist. However, the zero-frequency modes $\mathbf{F}_0$ and $\mathbf{G}_0$ may exist in cavity types 2 and 3, respectively. These modes are independent of the $\mathbf{E}_n, \mathbf{H}_n$ modes. The nth free oscillation in the cavity is given by

$$\mathbf{E} = e_n \mathbf{E}_n = \mathbf{E}_n e^{j\omega_n t} \tag{7.113a}$$

$$\mathbf{H} = j\sqrt{\frac{\epsilon}{\mu}}\, \mathbf{H}_n e^{j\omega_n t} \tag{7.113b}$$

These results are valid if the material in the cavity is lossy as well, *provided* ϵ and μ are taken as complex quantities. In this case ω_n is complex, with the imaginary part representing a damping of the mode.

Cavity with Lossy Walls

Consider a cavity with finite conducting walls on which

$$\mathbf{n} \times \mathbf{E} = Z_m \mathbf{n} \times \mathbf{J}_s = Z_m \mathbf{H}_t \tag{7.114}$$

where $\mathbf{H}_t$ is the tangential magnetic field and $\mathbf{n}$ is a unit outward directed normal. The surface impedance $Z_m = (1 + j)/\sigma\delta_s$. Let the fields be expanded as follows:

$$\mathbf{E}e^{j\omega t} = \sum_n e_n \mathbf{E}_n e^{j\omega t} + \sum_n f_n \mathbf{F}_n e^{j\omega t} \tag{7.115a}$$

$$\mathbf{H}e^{j\omega t} = \sum_n h_n \mathbf{H}_n e^{j\omega t} + \sum_n g_n \mathbf{G}_n e^{j\omega t} \tag{7.115b}$$

where e_n, f_n, h_n, and g_n are amplitude constants independent of time. We have assumed a time variation $e^{j\omega t}$ so that the concept of a surface impedance Z_m can be applied to account for the finite conductivity of the walls.

In the present case $\mathbf{n} \times \mathbf{E}$ and $\mathbf{n} \cdot \mathbf{H}$ do not vanish on S, and since the modes in which $\mathbf{E}$ and $\mathbf{H}$ are expanded satisfy the boundary conditions $\mathbf{n} \times \mathbf{E}_n = \mathbf{n} \times \mathbf{F}_n = 0$ on S and $\mathbf{n} \cdot \mathbf{H}_n = \mathbf{n} \cdot \mathbf{G}_n = 0$ on S, the series expansions for $\mathbf{E}$ and $\mathbf{H}$ will not be uniformly convergent at the boundary S. Consequently, the curl of (7.115a) cannot be evaluated term by term. To overcome this difficulty, we use the divergence theorem (essentially an

integration by parts) to obtain

$$\int_V \nabla \cdot \mathbf{E} \times \mathbf{H}_n \, dV = \int_V (\nabla \times \mathbf{E}) \cdot \mathbf{H}_n \, dV - \int_V \nabla \times \mathbf{H}_n \cdot \mathbf{E} \, dV$$

$$= \oint_S \mathbf{n} \times \mathbf{E} \cdot \mathbf{H}_n \, dS$$

Replacing $\nabla \times \mathbf{H}_n$ by $k_n \mathbf{E}_n$ and $\nabla \times \mathbf{E}$ by $-j\omega\mu\mathbf{H}$ and using the expansions (7.115) and the orthogonal properties of the eigenfunctions now give

$$-j\omega\mu h_n - k_n e_n = \oint_S \mathbf{n} \times \mathbf{E} \cdot \mathbf{H}_n \, dS = Z_m \oint_S \mathbf{H} \cdot \mathbf{H}_n \, dS \quad (7.116a)$$

Note that $\mathbf{H}_n \cdot \mathbf{H} = \mathbf{H}_n \cdot \mathbf{H}_t$ since $\mathbf{n} \cdot \mathbf{H}_n = 0$. Similarly, we find

$$\int_V \nabla \cdot \mathbf{E}_n \times \mathbf{H} \, dV = \int_V (\nabla \times \mathbf{E}_n \cdot \mathbf{H} - \nabla \times \mathbf{H} \cdot \mathbf{E}_n) \, dV$$

$$= \oint_S \mathbf{n} \times \mathbf{E}_n \cdot \mathbf{H} \, dS = 0$$

Replacing $\nabla \times \mathbf{E}_n$ by $k_n \mathbf{H}_n$ and $\nabla \times \mathbf{H}$ by $j\omega\epsilon\mathbf{E}$ and using the expansion (7.115) yield

$$j\omega\epsilon e_n = k_n h_n \quad (7.116b)$$

This result is the same as that obtained by taking the curl of (7.115b) term by term. However, (7.116a) cannot be obtained by taking the curl of (7.115a) term by term because of its nonuniform convergence.

From the two relations (7.116) separate expressions for e_n and h_n may be obtained. For h_n we find

$$h_n = \frac{j\omega\epsilon}{k^2 - k_n^2} Z_m \oint_S \mathbf{H} \cdot \mathbf{H}_n \, dS \quad (7.117)$$

where $k^2 = \omega^2\mu\epsilon$.

Let us now assume that the field is essentially that of the nth mode and that this mode is not degenerate, i.e., no other mode has the same eigenvalue k_n. Then we have $\mathbf{H} \approx h_n \mathbf{H}_n$, and the surface integral becomes

$$h_n Z_m \oint_S \mathbf{H}_n \cdot \mathbf{H}_n \, dS = h_n Z_m \oint_S |\mathbf{H}_n|^2 \, dS$$

since $\mathbf{H}_n$ is real. The power loss in the walls for the nth mode is

$$P_l = \frac{R_m}{2} \oint_S |\mathbf{H}_n|^2 \, dS$$

and the average stored magnetic energy is

$$W_m = \frac{\mu}{4} \int_V |\mathbf{H}_n|^2 \, dV = \frac{\mu}{4}$$

The Q for the nth mode is

$$Q_n = \frac{2\omega W_m}{P_l}$$

and thus

$$\oint_S |\mathbf{H}_n|^2 \, dS = \frac{2P_l}{R_m} = \frac{4\omega W_m}{Q_n R_m} = \frac{\mu\omega}{Q_n R_m} \tag{7.118}$$

For (7.117) we now get

$$h_n = \frac{j\omega\epsilon(1+j)R_m}{k^2 - k_n^2} \frac{\mu\omega h_n}{Q_n R_m}$$

a relation that can hold only if

$$-\frac{k^2(1-j)}{k^2 - k_n^2} \frac{1}{Q_n} = 1$$

or

$$k = \frac{k_n}{(1 + 1/Q_n - j/Q_n)^{1/2}} \approx k_n\left(1 - \frac{1}{2Q_n} + \frac{j}{2Q_n}\right) \tag{7.119}$$

We thus find that, for the cavity with lossy walls, the resonant frequency differs from the no-loss resonant frequency ω_n by a factor $1 - 1/2Q_n$. In addition, a damping constant $\delta = \omega_n/2Q_n$ is introduced. In terms of ω and δ, (7.119) gives

$$j\omega - \delta = j\omega_n\left(1 - \frac{1}{2Q_n}\right) - \frac{1}{2Q_n} \tag{7.120}$$

Degenerate Modes

The volume orthogonality of the $\mathbf{G}_n$ and $\mathbf{H}_m$ modes hold even if $p_n = k_m$, as an examination of the method used in the proof will show. However, if $k_n = k_m$, then the proof of the volume orthogonality of the $\mathbf{H}_n$ and $\mathbf{H}_m$ modes breaks down. In this case

$$\int_V \mathbf{H}_n \cdot \mathbf{H}_m \, dV$$

may or may not vanish. If the integral does not vanish, the two modes are coupled together, since the average magnetic energy stored in the two modes will contain a nonzero interaction term arising from the above integral.

In addition to volume coupling between degenerate modes, it is also possible to have coupling arising from finite wall losses. The nature of this surface coupling is examined below.

If two modes $\mathbf{H}_n$ and $\mathbf{H}_m$ are degenerate, so that $k_n = k_m$, and if in addition

$$\oint_S \mathbf{H}_n \cdot \mathbf{H}_m \, dS \neq 0 \tag{7.121}$$

then the power loss associated with these two modes will contain a cross-interaction term arising from the above integral. In this case the two modes are said to be coupled together by the finite surface impedance of the cavity walls. It is not possible to have just one of these modes present since the presence of one mode automatically couples the other mode. However, for most practical cavities such coupling does not exist. Nevertheless, the possibility of mode coupling must be kept in mind since, if it exists, both modes must be included in any calculation of energy stored, power loss, and Q.

In the case where k_n and k_m are not equal, the surface integral may be shown to vanish for rectangular and cylindrical cavities. Although a general proof is not available, we should anticipate that this is a general property of nondegenerate modes.

The problem of coupled degenerate modes may be circumvented by introducing new modes that are linear combinations of the old degenerate modes in such a fashion that they are uncoupled. If $\mathbf{H}_n$ and $\mathbf{H}_m$ are degenerate coupled modes, choose new modes

$$\mathbf{H}'_n = c_1\mathbf{H}_n + c_2\mathbf{H}_m \tag{7.122a}$$

$$\mathbf{H}'_m = d_1\mathbf{H}_n + d_2\mathbf{H}_m \tag{7.122b}$$

with c_i and d_i chosen so that

$$\oint_S \mathbf{H}'_n \cdot \mathbf{H}'_m \, dS = 0 \tag{7.123a}$$

$$\int_V \mathbf{H}'_n \cdot \mathbf{H}'_m \, dV = 0 \tag{7.123b}$$

$$\int_V |\mathbf{H}'_n|^2 \, dV = \int_V |\mathbf{H}'_m|^2 \, dV = 1 \tag{7.123c}$$

These new modes are uncoupled and can exist independently of each other in the lossy cavity. For these new uncoupled modes the Q may be evaluated for each mode individually, since the cross-coupling term in the expression for power loss has been made equal to zero and, similarly, the cross-coupling term in the expression for stored magnetic energy has been made equal to zero. If more than two modes are degenerate, a similar procedure may be applied to find a new set of uncoupled normalized modes. In a general

discussion we may therefore assume that all the degenerate modes have been chosen so that they are uncoupled.

*7.9 EXCITATION OF CAVITIES

In this section we consider the application of the modal expansion of the field in a cavity to the problem of finding the field excited by magnetic and electric dipoles, which may represent a current loop or probe, respectively. In addition, a small aperture may be described in terms of equivalent electric and magnetic dipoles as well. Thus the theory to be developed will be sufficiently general to treat the three common methods of coupling a cavity to an external waveguide or coaxial transmission line.

Let a cavity contain infinitesimal electric and magnetic dipoles

$$\mathbf{P} = \mathbf{P}_0 e^{j\omega t}\delta(\mathbf{r} - \mathbf{r}_0) \tag{7.124a}$$

$$\mathbf{M} = \mathbf{M}_0 e^{j\omega t}\delta(\mathbf{r} - \mathbf{r}_0) \tag{7.124b}$$

at a point whose position is defined by the vector $\mathbf{r}_0$.

The three-dimensional delta function $\delta(\mathbf{r} - \mathbf{r}_0)$ symbolizes that the dipoles are localized at the point $\mathbf{r} = \mathbf{r}_0$. This delta function is defined in such a manner that, for an arbitrary vector $\mathbf{A}$ which is continuous at $\mathbf{r}_0$, we have

$$\int_V \mathbf{A}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}_0)\, dV = \mathbf{A}(r_0)$$

when the point $\mathbf{r}_0$ is included in the volume V (Sec. 2.11).

By analogy with the following equations governing polarization in material bodies,

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) \qquad \mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}$$

it is seen that Maxwell's equations become

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B} = -j\omega\mu_0\mathbf{H} - j\omega\mu_0\mathbf{M}_0\delta(\mathbf{r} - \mathbf{r}_0) \tag{7.125a}$$

$$\nabla \times \mathbf{H} = j\omega\mathbf{D} = j\omega\epsilon_0\mathbf{E} + j\omega\mathbf{P}_0\delta(\mathbf{r} - \mathbf{r}_0) \tag{7.125b}$$

where a time factor $e^{j\omega t}$ has been suppressed.

We now use the general expansion (7.115) for the fields to give

$$\mathbf{E} = \sum_n e_n\mathbf{E}_n + \sum_n f_n\mathbf{F}_n \tag{7.126a}$$

$$\mathbf{H} = \sum_n h_n\mathbf{H}_n + \sum_n g_n\mathbf{G}_n \tag{7.126b}$$

To find the expansion coefficients e_n, h_n, we follow the derivation leading to (7.116) but note that $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{H}$ are replaced by the right-hand sides of (7.125). Thus

$$\int_V (\nabla \times \mathbf{E}) \cdot \mathbf{H}_n \, dV - \int_V \nabla \times \mathbf{H}_n \cdot \mathbf{E} \, dV$$

$$= Z_m \oint_S \mathbf{H} \cdot \mathbf{H}_n \, dS$$

$$= \int_V [-j\omega\mu_0 \mathbf{H} - j\omega\mu_0 \mathbf{M}_0 \delta(\mathbf{r} - \mathbf{r}_0)] \cdot \mathbf{H}_m \, dV - k_n \int_V \mathbf{E}_n \cdot \mathbf{E} \, dV$$

$$= -j\omega\mu_0 h_n - k_n e_n - j\omega\mu_0 \mathbf{M}_0 \cdot \mathbf{H}_n(\mathbf{r}_0) \qquad (7.127a)$$

and similarly,

$$k_n h_n = j\omega\epsilon_0 e_n + j\omega \mathbf{P}_0 \cdot \mathbf{E}_n(\mathbf{r}_0) \qquad (7.127b)$$

To obtain an equation for g_n, consider

$$\int_V \nabla \cdot \mathbf{E} \times \mathbf{G}_n \, dV = \int_V (\mathbf{G}_n \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{G}_n) \, dV$$

$$= \oint_S \mathbf{n} \times \mathbf{E} \cdot \mathbf{G}_n \, dS = Z_m \oint_S \mathbf{H} \cdot \mathbf{G}_n \, dS$$

Using (7.125*a*) for $\nabla \times \mathbf{E}$, the expansion (7.126*b*), and the orthogonal properties of the modes $\mathbf{H}_n$ and $\mathbf{G}_n$ now gives

$$j\omega\mu_0 g_n + j\omega\mu_0 \mathbf{M}_0 \cdot \mathbf{G}_n(\mathbf{r}_0) = -Z_m \oint_S \mathbf{H} \cdot \mathbf{G}_n \, dS \qquad (7.128)$$

since $\nabla \times \mathbf{G}_n$ is zero.

In a similar fashion, use of the relation

$$\int_V \nabla \cdot \mathbf{F}_n \times \mathbf{H} \, dV = \int_V (\mathbf{H} \cdot \nabla \times \mathbf{F}_n - \mathbf{F}_n \cdot \nabla \times \mathbf{H}) \, dV$$

$$= -\int_V \mathbf{F}_n \cdot \nabla \times \mathbf{H} \, dV = \oint_S \mathbf{n} \times \mathbf{F}_n \cdot \mathbf{H} \, dV = 0$$

together with (7.125*b*) and (7.126*a*) yields

$$j\omega\epsilon_0 f_n = -j\omega \mathbf{P}_0 \cdot \mathbf{F}_n(\mathbf{r}_0) \qquad (7.129)$$

We now have the following equations for the expansion coefficients e_n, h_n, f_n, and g_n:

$$j\omega\epsilon_0 e_n = k_n h_n - j\omega \mathbf{P}_0 \cdot \mathbf{E}_n(\mathbf{r}_0) \tag{7.130a}$$

$$h_n = \frac{-1}{k_0^2 - k_n^2}\left(j\omega k_n \mathbf{P}_0 \cdot \mathbf{E}_n + k_0^2 \mathbf{M}_0 \cdot \mathbf{H}_n - j\omega\epsilon_0 Z_m \oint_S \mathbf{H} \cdot \mathbf{H}_n \, dS\right) \tag{7.130b}$$

$$j\omega\epsilon_0 f_n = -j\omega \mathbf{P}_0 \cdot \mathbf{F}_n \tag{7.130c}$$

$$j\omega\mu_0 g_n = -j\omega\mu_0 \mathbf{M}_0 \cdot \mathbf{G}_n - Z_m \oint_S \mathbf{H} \cdot \mathbf{G}_n \, dS \tag{7.130d}$$

where $k_0^2 = \omega^2\mu_0\epsilon_0$ and (7.130a) has been used in order to eliminate e_n and obtain (7.130b).

In many practical problems dealing with cavities, the above equations may be simplified. Usually, ω is very nearly equal to a particular resonant frequency ω_n. As (7.130b) shows, all coefficients h_m, $m \neq n$, will then be small compared with h_n. Thus all the coefficients e_m, $m \neq n$, will also be small compared with e_n, and the field is predominantly that described by the $\mathbf{H}_n, \mathbf{E}_n$ mode. In the surface integrals, which represent small perturbations from the loss-free solution, we may approximate $\mathbf{H}$ by $h_n\mathbf{H}_n$ without appreciable error. In addition, in the equation for g_n, we may neglect the surface-integral term to a first approximation. We may then use the relation

$$\oint_S \mathbf{H}_n \cdot \mathbf{H}_n \, dS = \frac{2P_l}{R_m} = \frac{\omega\mu_0}{Q_n R_m}$$

derived earlier if the $\mathbf{H}_n$ modes have been chosen so that they are uncoupled. In place of (7.130) we now obtain the following simplified equations

$$\left[k_0^2\left(1 + \frac{1-j}{Q_n}\right) - k_n^2\right] h_n = -j\omega k_n \mathbf{P}_0 \cdot \mathbf{E}_n - k_0^2 \mathbf{M}_0 \cdot \mathbf{H}_n$$

or

$$h_n = \frac{j\omega k_n \mathbf{P}_0 \cdot \mathbf{E}_n + k_0^2 \mathbf{M}_0 \cdot \mathbf{H}_n}{k_n^2 - k_0^2\left(1 + \frac{1-j}{Q_n}\right)} \tag{7.131a}$$

$$e_n = -\frac{j\omega\mu_0(k_n \mathbf{M}_0 \cdot \mathbf{H}_n + j\omega \mathbf{P}_0 \cdot \mathbf{E}_n)}{k_n^2 - k_0^2\left(1 + \frac{1-j}{Q_n}\right)} \tag{7.131b}$$

$$\epsilon_0 f_n = -\mathbf{P}_0 \cdot \mathbf{F}_n \tag{7.131c}$$

$$g_n = -\mathbf{M}_0 \cdot \mathbf{G}_n \tag{7.131d}$$

Equations (7.131a) and (7.131b) may be used for all h_m and e_m. Since we have assumed ω to be equal or nearly equal to ω_n, the denominator may be

replaced by $k_0^2 - k_m^2$ for $m \neq n$. For n we may factor the denominator to give

$$\left[k_n - k_0\left(1 + \frac{1-j}{Q_n}\right)^{1/2}\right]\left[k_n + k_0\left(1 + \frac{1-j}{Q_n}\right)^{1/2}\right]$$

$$\approx -2k_n\left[k_0 - k_n\left(1 - \frac{1-j}{2Q_n}\right)\right]$$

since Q_n is very large.

Usually, we are primarily interested in the strength of excitation of the resonant mode $\mathbf{E}_n, \mathbf{H}_n$. Its excitation coefficients are given by (7.131*a*) and (7.131*b*). The coefficients f_n and g_n describe the local field that exists around the dipole sources. This field is a quasistatic field in its configuration. The excitation of fields in cavities by volume distributions of currents may be solved by the same method outlined above. As an example, consider a cavity with a volume distribution of current $\mathbf{J}(\mathbf{r}_0)e^{j\omega_n t}$. A differential element of current may be considered equivalent to an electric dipole $\mathbf{P}$ with a moment given by $\mathbf{J}/j\omega$. Thus, from a current element $\mathbf{J}(\mathbf{r}_0)\delta(\mathbf{r} - \mathbf{r}_0)$ located at the point $\mathbf{r}_0$, the amplitudes of the nth mode are given by (7.131) divided by $j\omega$. When the current varies with the frequency ω_n of the nth resonant mode, only this mode is excited with a large amplitude. The amplitude of the electric field of the nth mode due to the volume distribution of current is found by superposition, i.e., adding up the contributions from each current element. Thus we have

$$e_n = -j\omega\mu_0 \frac{\int_V \mathbf{J}(\mathbf{r}_0) \cdot \mathbf{E}_n(\mathbf{r}_0)\, dV_0}{k_n^2 - k_0^2\left(1 + \dfrac{1-j}{Q_n}\right)} \tag{7.132}$$

where the integration is taken over the volume of the current distribution. A similar expression holds for the amplitude constant h_n.

The fields excited in cavities by volume distributions of currents may also be solved in terms of a vector potential function. A specific application of the use of the vector potential is given in Chap. 9 in connection with the klystron tube; so we do not consider this method here.

*7.10 CAVITY PERTURBATION THEORY

The resonant frequency of a cavity can be varied over a small range by inserting a small adjustable screw into the cavity as shown, for example, in Fig. 7.33*a*. A small obstacle, such as a dielectric sphere, placed in a cavity will also change the resonant frequency and the Q of the cavity if the dielectric is lossy. Small obstacles, when acted upon by electric and magnetic fields that are uniform over the volume occupied by the obstacle, can be

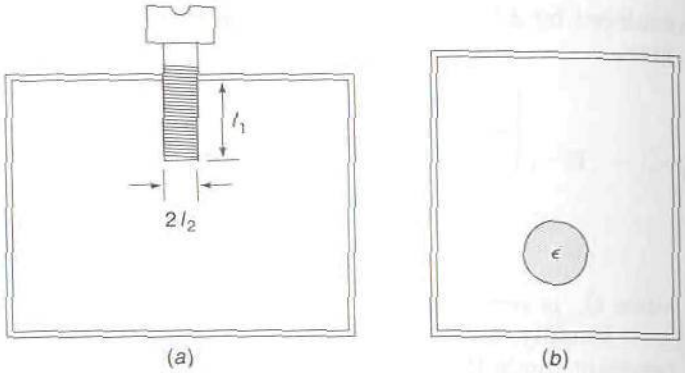


FIGURE 7.33
(a) An adjustable screw used to tune a cavity; (b) a small dielectric sphere placed in a cavity.

characterized in terms of induced electric and magnetic dipole moments. For example, a dielectric sphere of radius l and with a dielectric constant ϵ_r will have a total electric dipole moment $\mathbf{P}$ given by

$$\mathbf{P} = 4\pi l^3 \frac{\epsilon_r - 1}{\epsilon_r + 2} \epsilon_0 \mathbf{E}_0 = \alpha_e \epsilon_0 \mathbf{E}_0 \tag{7.133}$$

where $\mathbf{E}_0$ is the electric field at the location of the dielectric sphere when it is absent. The parameter α_e is called the electric polarizability of the dielectric sphere. For other obstacles we can write, in a similar way,

$$\mathbf{P} = \alpha_e \epsilon_0 \mathbf{E}_0 \tag{7.134a}$$

$$\mathbf{M} = \alpha_m \mathbf{H}_0 \tag{7.134b}$$

where $\mathbf{M}$ is the induced magnetic dipole moment and α_m is the magnetic polarizability. In Table 7.1 we list the polarizabilities for a number of obstacles that have the shape of prolate or oblate spheroids or their degenerate forms. The polarizabilities depend on whether the field acts along the major axis or minor axis.

When the dominant mode in the cavity is the nth resonant mode, the fields $\mathbf{E}_0$ and $\mathbf{H}_0$ in (7.134) represent the fields of the nth cavity mode. The additional field radiated by the induced dipole moments have amplitudes given by (7.131). If we assume that the field in the cavity is essentially only the perturbed nth mode, then the dipole moments are given by

$$\mathbf{P} = e_n[\alpha_{eu}\epsilon_0 E_{nu}\mathbf{a}_u + \alpha_{ev}\epsilon_0 E_{nv}\mathbf{a}_v] \tag{7.135a}$$

$$\mathbf{M} = h_n[\alpha_{mu} H_{nu}\mathbf{a}_u + \alpha_{mv} H_{nv}\mathbf{a}_v] \tag{7.135b}$$

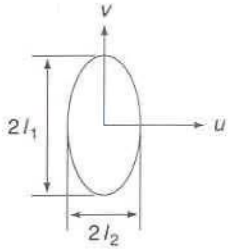
Let $\mathbf{E}_n \cdot \mathbf{P} = e_n \epsilon_0 \Delta_e$ and $\mathbf{H}_n \cdot \mathbf{M} = h_n \Delta_m$, that is,

$$\Delta_e = \alpha_{eu} E_{nu}^2 + \alpha_{ev} E_{nv}^2 \tag{7.136a}$$

$$\Delta_m = \alpha_{mu} H_{nu}^2 + \alpha_{mv} H_{nv}^2 \tag{7.136b}$$

TABLE 7.1
Polarizabilities of common obstacles

Prolate spheroid



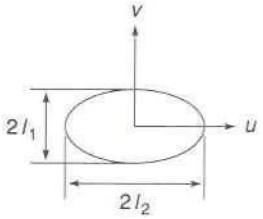
$$\alpha_{ev} = \frac{(\epsilon_r - 1)V}{(\epsilon_r - 1)L + 1} \qquad \alpha_{mv} = \frac{(\mu_r - 1)V}{(\mu_r - 1)L + 1}$$

$$\alpha_{eu} = \frac{2(\epsilon_r - 1)V}{1 + \epsilon_r - (\epsilon_r - 1)L} \qquad \alpha_{mu} = \frac{2(\mu_r - 1)V}{1 + \mu_r - (\mu_r - 1)L}$$

$$L = \frac{1 - e^2}{2e^3}\left[\ln\frac{1 + e}{1 - e} - 2e\right]$$

$$e = \left(1 - \frac{l_2^2}{l_1^2}\right)^{1/2} \qquad V = \frac{4}{3}\pi l_1 l_2^2$$

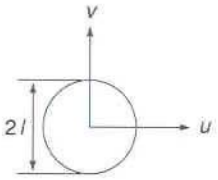
Oblate spheroid



$$L = \frac{1}{e^2} - \frac{\sqrt{1 - e^2}}{e^3}\tan^{-1}\frac{e}{\sqrt{1 - e^2}}$$

$$e = \left(1 - \frac{l_1^2}{l_2^2}\right)^{1/2}$$

Sphere

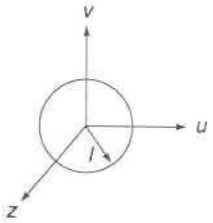


$$\alpha_{ev} = \alpha_{eu} = 4\pi l^3\frac{\epsilon_r - 1}{\epsilon_r + 2}$$

$$\alpha_{mv} = \alpha_{mu} = 4\pi l^3\frac{\mu_r - 1}{\mu_r + 2} \qquad L = \frac{1}{3}$$

For conducting obstacles let $\epsilon_r \to \infty$ and set $\mu_r = 0$

Circular disk (metallic)



$$\alpha_{eu} = \alpha_{ev} = \frac{16}{3}l^3 \qquad \alpha_{mu} = \alpha_{mv} = 0$$

$$\alpha_{ez} = 0 \qquad \alpha_{mz} = -\frac{8}{3}l^3$$

Equations (131a) and (131b) give

$$h_n - \frac{jk_0 Y_0 k_n \Delta_e e_n + k_0^2 \Delta_m h_n}{k_n^2 - k_0^2\left(1 + \dfrac{1-j}{Q_n}\right)} = 0$$

$$e_n - \frac{-jk_0 Z_0 k_n \Delta_m h_n + k_0^2 \Delta_e e_n}{k_n^2 - k_0^2\left(1 + \dfrac{1-j}{Q_n}\right)} = 0$$

This is a homogeneous set of equations and has a solution only if the determinant vanishes. By equating the determinant to zero, we obtain an equation for k_0^2 which determines the perturbed resonant frequency of the nth mode when the obstacle is placed in the cavity. When we neglect terms in Δ_e^2, Δ_m^2, and $\Delta_e \Delta_m$, it is readily found that†

$$k_0 = k_n\left[1 - \frac{\delta + \Delta_e + \Delta_m}{2}\right] \tag{7.137}$$

where $\delta = (1-j)/Q_n$. Normally Δ_e and Δ_m are both very small; so the approximation is a very good one in many practical cases.

As an example of the application of this formula, consider the cylindrical cavity with a tuning screw of length l_1 and radius l_2 as shown in Fig. 7.33a. Let the mode in the cavity be the TM_{010} mode with electric field given by

$$E_z = e_{010} E_{010} = e_{010} \frac{J_0(p_{01} r/a)}{\sqrt{\pi d}\, a J_1(p_{01})}$$

where d is the length of the cavity and a is the cavity radius. We can approximate the tuning screw as one-half of a prolate spheroid. When $l_1 \gg l_2$ the induced electric dipole moment (see Table 7.1) is

$$P_z = \tfrac{1}{2}\alpha_{eu} \epsilon_0 E_z$$

when the tuning screw is located along the axis of the cavity. The factor $\frac{1}{2}$ occurs because only one-half of the spheroid is present. The magnetic field along the axis is zero; hence (7.137) gives

$$k_0 = k_{010}\left[1 - \frac{\alpha_{eu}}{2} \frac{1}{\pi d a^2 J_1^2(p_{01})}\right]$$

when we assume that the cavity Q is infinite so that $\delta = 0$. Let us assume that $a = 2$ cm, $d = 3$ cm, $l_1 = 1$ cm, and $l_2 = 0.1$ cm. The unperturbed

†R. E. Collin, "Field Theory of Guided Waves," 2nd ed., chap. 5, IEEE Press, Piscataway, N.J., 1991.

cavity resonant wave number is $k_{010} = p_{01}/a = 2.405/2 = 1.2025$ rad/cm. The polarizability α_{eu} is given by

$$\alpha_{eu} = \frac{4}{3}\pi l_1 l_2^2 \frac{2l_1^2 e^3}{l_2^2\left[\ln\dfrac{1+e}{1-e} - 2e\right]}$$

where $e = (1 - l_2^2/l_1^2)^{1/2}$. For our example α_{eu} equals 2.065×10^{-2}. The perturbed resonant wave number is thus given by

$$k_0 = k_{010}\left(1 - \frac{2.065 \times 10^{-2}}{2\pi \times 3 \times 4 \times (0.5191)^2}\right) = k_{010}(1 - 0.001)$$

The perturbed resonant frequency is thus 0.1 percent lower with the tuning screw present.

By measuring the change in the resonant frequency and Q of a cavity when a lossy dielectric sphere is placed in the cavity, the complex permittivity of the dielectric material can be determined (see Prob. 7.20).

PROBLEMS

7.1. Show that, on a short-circuited coaxial line one-half wavelength long, the time-average stored electric and magnetic energies are equal. Use the expressions for the fields given by (3.81).

7.2. For the folded coaxial line, show that $b = \sqrt{ad}$ in order for the characteristic impedance of the inner and outer lines to be the same. This is a common form of line for use in high-frequency oscillators. The effective length l is about twice the physical length. At a frequency of 300 MHz, what must $l/2$ be in order that $l = \lambda_0/4$? If $2d = 5$ cm and $2a = 2$ cm, what must $2b$ equal for equal characteristic impedances? For a copper line ($\sigma = 5.8 \times 10^7$ S/m), find the Q and input impedance at resonance. For d fixed, what are the optimum values of a and b that will make Q a maximum?

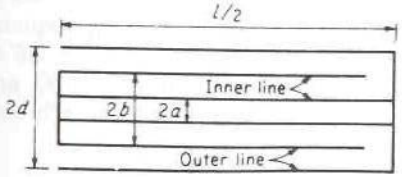


FIGURE P7.2

7.3. Verify that an open-circuited transmission line behaves as a series resonant circuit in the vicinity of the frequency for which it is a quarter wavelength long. Obtain an expression for the input impedance at resonance.

7.4. A short-circuited two-wire line is made of copper. The conductor diameter is 1 cm, the spacing is 3 cm, and the length is 40 cm. Find the antiresonant frequency, the Q, and the input resistance at resonance.

7.5. Design a microstrip-line resonator like that shown in Fig. 7.8*a*. The microstrip input line and resonator are both of width $W = 1$ mm, the substrate thickness is 1 mm and made from alumina with $\epsilon_r = 9.7$ and loss tangent equal to 2×10^{-4}. The microstrip is made from copper 0.01 mm thick. Find the required capacitance C_2 for critical coupling and the gap spacing s. What is the length l and Q for the resonator? The frequency of operation is 4 GHz. How much shorter than one-half wavelength do you have to make l because of capacitive loading at the two ends? The gap capacitance for $\epsilon_r = 9.6$ can be used.

7.6. Design a circular disk resonator operating in the TM_{110} mode at 6 GHz. The substrate to be used has a dielectric constant $\epsilon_r' - j\epsilon_r'' = 6 - j0.005$ and is 1 mm thick. The disk is made from copper. Determine the required radius a of the disk and the Q of the resonator.

7.7. Find the resonant frequency and Q of a copper rectangular cavity of dimensions $a = b = d = 10$ cm for the TE_{101} mode.

7.8. A cylindrical cavity of radius $a = 2$ cm and a length of 6 cm is filled with a dielectric with permittivity $\epsilon = (2.5 - j0.0001)\epsilon_0$. The cavity is made of copper. Find the resonant frequency and Q for the TE_{111} mode. Note that, in the expression for resonant frequency, the velocity of light in free space, c, must be replaced by

$$c\sqrt{\frac{\epsilon_0}{\epsilon'}}$$

***7.9.** Use the results given by (4.25) and (4.26) to show that the Q of a cavity is given by

$$Q = \frac{\omega}{2R}\frac{\partial X}{\partial \omega} = \frac{\omega}{2G}\frac{\partial B}{\partial \omega}$$

where $Z_{in} = R + jX$ and $Y_{in} = G + jB$ are the input impedance and admittance at the terminals. Verify that these formulas give the usual results for a series RCL network and for a parallel RCL network.

7.10. A rectangular cavity of dimensions a, b, d is coupled to a rectangular guide through a capacitive slit. The guide width is a, and the height is b. Obtain an equation for determining the first antiresonant frequency. Find the required slit susceptance for critical coupling. For $a = 2b = 2.5$ cm, $d = 2.5$ cm, and a copper cavity, compute the resonant frequency Q and slit opening t for critical coupling. Use the formula $\bar{B}_c = (2\beta b/\pi)\ln \csc(\pi t/2b)$ for the slit susceptance. What is the loaded Q?

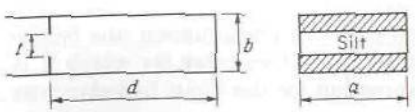


FIGURE P7.10

7.11. For a capacitive diaphragm in a rectangular guide of the dimensions given in Prob. 7.10, obtain an equivalent circuit to represent the susceptance function $\overline{B}_c = (2\beta b/\pi)\ln \csc(\pi t/2b)$ correct to terms up to $\Delta\omega = \omega - \omega_1$ in the vicinity of the frequency ω_1.

Hint: Expand β in a Taylor series about ω_1 and choose a series LC circuit.

7.12. Design a rectangular cavity of length d, height $b = 1.2$ cm, width $a = 2.5$ cm that will resonate at 10,000 MHz. The cavity is critically coupled to a rectangular guide of dimensions a by b. Specify the cavity length d and the radius of the centered circular aperture. Determine the unloaded and loaded Q's if the cavity is made of copper.

7.13. For the aperture-coupled rectangular cavity discussed in the text, let the incident power at resonance be 100 mW. The cavity is critically coupled. Evaluate the peak value of the electric field in the incident wave and in the cavity field. How does the peak amplitude of the cavity field depend on the cavity Q? (See Prob. 7.12 for cavity dimensions.)

7.14. A hemispherical resonator is to be used at 10 GHz. It is made from a dielectric with a dielectric constant of 100 and is mounted on a ground plane. Find the required radius of the hemisphere and the radiation Q_r.

***7.15.** A cavity is excited by an impressed electric field $\mathbf{E}_0$ tangent to an aperture surface S_a cut in the cavity wall S. Use the relation $\int_V \nabla \cdot \mathbf{E} \times \mathbf{G}_n \, dV$ to show that the amplitude g_n is given by

$$-j\omega\mu_0 g_n = \oint_S \mathbf{n} \times \mathbf{E} \cdot \mathbf{G}_n \, dS = \int_{S_a} \mathbf{n} \times \mathbf{E}_0 \cdot \mathbf{G}_n \, dS$$

for a cavity with perfectly conducting walls. Using the general expansion (7.115*b*), show that an alternative expression is

$$-j\omega\mu_0 g_n = -j\omega\mu_0 \int_V \mathbf{H} \cdot \mathbf{G}_n \, dV = \frac{-j\omega\mu_0}{p_n} \oint_S \mathbf{n} \cdot \mathbf{H}\psi_n \, dS$$

upon putting $\mathbf{H} \cdot \nabla\psi_n = \nabla \cdot \mathbf{H}\psi_n$ since $\nabla \cdot \mathbf{H} = 0$ and using the divergence theorem. This last relation shows that the $\mathbf{G}_n$ modes are excited whenever $\mathbf{n} \cdot \mathbf{H}$ does not vanish over S.

***7.16.** Show that the two expressions for g_n in Prob. 7.15 are identical.

Hint: Consider

$$\mathbf{n} \cdot \nabla \times \mathbf{E}\psi_n = \psi_n \mathbf{n} \cdot \nabla \times \mathbf{E} - \mathbf{n} \cdot \mathbf{E} \times \nabla\psi_n = -j\omega\mu_0 \psi_n \mathbf{n} \cdot \mathbf{H} - \mathbf{n} \times \mathbf{E} \cdot \nabla\psi_n$$

and use Stokes' law to show that

$$\oint_S \mathbf{n} \cdot \nabla \times \mathbf{E}\psi_n \, dS = 0$$

***7.17.** Find the eigenfunctions $\mathbf{E}_n$, $\mathbf{H}_n$, $\mathbf{F}_n$, and $\mathbf{G}_n$ for a rectangular cavity of dimensions a, b, c.

***7.18.** Obtain a modal expansion similar to (7.126*a*) and (7.130) for the electromagnetic field in a cavity excited by a volume distribution of electric current $\mathbf{J}(\mathbf{r})$.

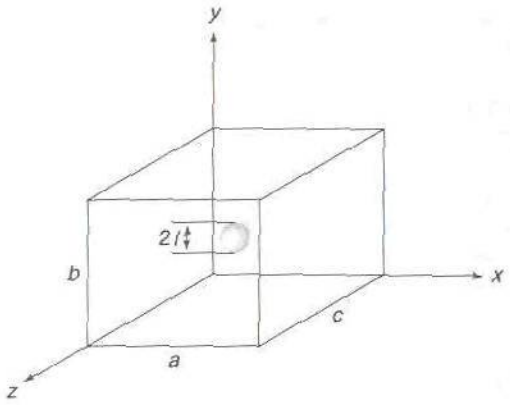


FIGURE P7.19

***7.19.** (*a*) For the box cavity illustrated in Fig. P7.19, show that the normalized electric field mode function for the TE_{101} mode is

$$\mathbf{E}_{101} = \sqrt{\frac{2}{abc}} \sin\frac{\pi x}{a} \sin\frac{\pi z}{c}\mathbf{a}_y$$

(*b*) Let a metal sphere of radius l be placed in the center of the cavity. Show that when $a = b = c$ and $l = a/20$ that the TE_{101}-mode resonant frequency is lowered by $\pi/20$ percent.

***7.20.** A dielectric sphere of radius l and complex relative permittivity $\epsilon_r = \epsilon_r' - j\epsilon_r''$ is placed in a cavity. Show that the perturbed resonant frequency for the nth cavity mode is given by

$$k_0 = k_n\left[1 - \frac{1}{2Q_n} + 2\pi l^3 \mathbf{E}_n \cdot \mathbf{E}_n \frac{(\epsilon_r' - 1)(\epsilon_r' + 2) + (\epsilon_r'')^2}{(\epsilon_r' + 2)^2 + (\epsilon_r'')^2} + \frac{j}{2}\left(\frac{1}{Q_n} + \frac{1}{Q_d}\right)\right]$$

where the cavity Q due to the lossy dielectric is given by

$$Q_d = \frac{(\epsilon_r' + 2)^2 + (\epsilon_r'')^2}{2\pi l^2 \mathbf{E}_n \cdot \mathbf{E}_n 3\epsilon_r''}$$

REFERENCES

1. Montgomery, C. G., R. H. Dicke, and E. M. Purcell: "Principles of Microwave Circuits," McGraw-Hill Book Company, New York, 1948.
2. Ragan, G. L. (ed.): "Microwave Transmission Circuits," McGraw-Hill Book Company, New York, 1948.
3. Slater, J. C.: "Microwave Electronics," D. Van Nostrand Company, Inc., Princeton, N.J., 1950.

4. Goubau, G.: "Electromagnetic Waveguides and Cavities," chap. 2, Pregamon Press, New York, 1961.
5. Kurokawa, K.: The Expansions of Electromagnetic Fields in Cavities, *IRE Trans.*, vol. MTT-6, pp. 178–187, April, 1958.
6. Kajfez, D., and P. Guillon: "Dielectric Resonators," Artech House Books, Dedham, Mass., 1986.
7. Van Bladel, J.: The Excitation of Dielectric Resonators of Very High Permittivity, *IEEE Trans.*, vol. MTT-23, pp. 208–217, 1975.
8. Gastine, M., Courtois, L. and J. L. Dormann: Electromagnetic Resonances of Free Dielectric Spheres, *IEEE Trans.*, vol. MTT-15, pp. 694–700, 1967.
9. Chow, K. K.: On the Solution and Field Pattern of Cylindrical Dielectric Resonators, *IEEE Trans.*, vol. MTT-14, p. 439, 1966.

CHAPTER 8

PERIODIC STRUCTURES AND FILTERS

Waveguides and transmission lines loaded at periodic intervals with identical obstacles, e.g., a reactive element such as a diaphragm, are referred to as periodic structures. The interest in waveguiding structures of this type arises from two basic properties common to all periodic structures, namely, (1) passband-stopband characteristics and (2) support of waves with phase velocities much less than the velocity of light. The passband-stopband characteristic is the existence of frequency bands throughout which a wave propagates unattenuated (except for incidental conductor losses) along the structure separated by frequency bands throughout which the wave is cut off and does not propagate. The former is called a passband, and the latter is referred to as a stopband. The passband-stopband property is of some interest for its frequency filtering aspects.

The ability of many periodic structures to support a wave having a phase velocity much less than that of light is of basic importance for traveling-wave-tube circuits. In a traveling-wave tube, efficient interaction between the electron beam and the electromagnetic field is obtained only if the phase velocity is equal to the beam velocity. Since the latter is often no greater than 10 to 20 percent of the velocity of light, considerable slowing down of the electromagnetic wave is required. Periodic structures suitable for use in traveling-wave tubes are discussed in this chapter. The actual principles of operation of the tube are covered in Chap. 9.

The last part of the chapter is devoted to an introduction to microwave filter theory. A complete treatment of all aspects of filter theory and design would be much too lengthy to include in this text. However, sufficient

material is covered to provide a background so that the technical literature can be read without difficulty.

8.1 CAPACITIVELY LOADED TRANSMISSION-LINE-CIRCUIT ANALYSIS

To introduce a number of basic concepts, methods of analysis, and typical properties of periodic structures, we shall consider a simple example of a capacitivity loaded transmission line. For a physically smooth transmission line, such as a coaxial line, the phase velocity is given by

$$v_p = (LC)^{-1/2} = (\mu_0 \epsilon_r \epsilon_0)^{-1/2} \tag{8.1}$$

where ϵ_r is the dielectric constant of the medium surrounding the conductor. A significant reduction in phase velocity can be achieved in a smooth line only by increasing ϵ_r. This method has the great disadvantage that the cross-sectional dimensions of the line must also be reduced to avoid the propagation of higher-order modes. The phase velocity cannot be decreased by increasing the shunt capacity C per unit length because any change in the line configuration to increase C automatically decreases the series inductance L per unit length, since $LC = \mu_0 \epsilon$. However, by removing the restriction that the line should by physically smooth, an effective increase in the shunt capacitance per unit length can be achieved without a corresponding decrease in the series inductance L. That is, lumped shunt capacitance may be added at periodic intervals without affecting the value of L. If the spacing between the added lumped capacitors is small compared with the wavelength, it may be anticipated that the line will appear to be electrically smooth, with a phase velocity

$$v_p = \left[\left(C + \frac{C_0}{d}\right)L\right]^{-1/2} \tag{8.2}$$

where C_0/d is the amount of lumped capacitance added per unit length (a capacitor C_0 added at intervals d). The following analysis will verify this conclusion.

One method of obtaining shunt capacitive loading of a coaxial transmission line is to introduce thin circular diaphragms at regular intervals, as in Fig. 8.1. The diaphragms may be machined as an integral part of the center conductor. The fringing electric field in the vicinity of the diaphragm increases the local storage of electric energy and hence may be accounted for, from a circuit viewpoint, by a shunt capacitance. The local field can be described in terms of the incident, reflected, and transmitted dominant TEM mode and a superposition of an infinite number of higher-order E modes. If the cylinder spacing $b - a$ is small compared with the wavelength, the higher-order modes are evanescent and decay to a negligible value in a distance of the order of $b - a$ away from the diaphragm in either

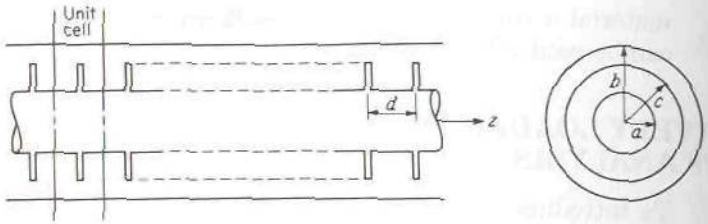


FIGURE 8.1
Capacitive loading of a coaxial line by means of thin circular diaphragms.

direction. An approximate expression for the shunt susceptance of the diaphragm is†

$$\overline{B} = \frac{B}{Y_c} = \frac{8(b-c)^2}{\lambda_0 c} \frac{\ln(b/a)}{[\ln(b/c)]^2} \ln \csc\left(\frac{\pi}{2}\frac{b-c}{b-a}\right) \tag{8.3}$$

where $Y_c = [60 \ln(b/a)]^{-1}$ is the characteristic admittance of an air-filled coaxial line. The expression for $\overline{B}$ is accurate for $b - a \leq 0.1\lambda_0$. In this low-frequency region, $\overline{B}$ has a frequency dependence directly proportional to ω. At higher frequencies $\overline{B}$ will have a more complicated frequency dependence, although the thin diaphragm can still be represented by a shunt susceptance.

The circuit, or network, analysis of a periodic structure involves constructing an equivalent network for a single basic section or unit cell of the structure first. This is followed by an analysis to determine the voltage and current waves that may propagate along the network consisting of the cascade connection of an infinite number of the basic networks. For the structure of Fig. 8.1, an equivalent network of a basic section is a shunt normalized susceptance $\overline{B}$ with a length $d/2$ of transmission line on either side, as in Fig. 8.2*a*. Figure 8.2*b* illustrates the voltage-current relationships at the input and output of the nth section in the infinitely long cascade connection.

The relationships between the input variables V_n, I_n and the output variables V_{n+1}, I_{n+1} are readily found by using the $\mathscr{ABCD}$ transmission matrix discussed in Sec. 4.9. The V_n and I_n are the total voltage and current amplitudes, i.e., the sum of the contributions from the incident and reflected TEM waves at the terminal plane. The circuit for a unit cell may be broken down into three circuits in cascade, namely, a section of transmission line of length $d/2$ (electrical length $\theta/2 = k_0 d/2$), followed by a shunt

†N. Marcuvitz (ed.), "Waveguide Handbook," p. 229, McGraw-Hill Book Company, New York, 1951.

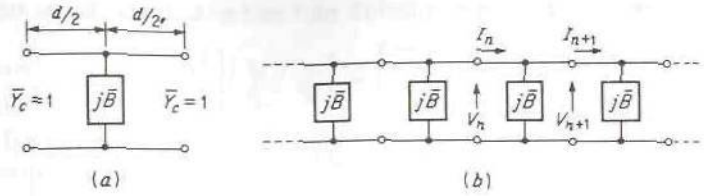


FIGURE 8.2
(a) Equivalent circuit for unit cell of loaded coaxial line; (b) cascade connection of basic unit-cell networks.

susceptance $\bar{B}$, which in turn is followed by another length of transmission line. The $\mathscr{A}\mathscr{B}\mathscr{C}\mathscr{D}$ matrix for each of these individual networks is, respectively (Prob. 4.26),

$$\begin{bmatrix} \cos\frac{\theta}{2} & j\sin\frac{\theta}{2} \\ j\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ j\bar{B} & 1 \end{bmatrix} \begin{bmatrix} \cos\frac{\theta}{2} & j\sin\frac{\theta}{2} \\ j\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}$$

The transmission matrix for the unit cell is obtained by the chain rule [see (4.75)], i.e., the product of the above three matrices, and hence we have

$$\begin{bmatrix} V_n \\ I_n \end{bmatrix} = \begin{bmatrix} \cos\frac{\theta}{2} & j\sin\frac{\theta}{2} \\ j\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ j\bar{B} & 1 \end{bmatrix} \begin{bmatrix} \cos\frac{\theta}{2} & j\sin\frac{\theta}{2} \\ j\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} V_{n+1} \\ I_{n+1} \end{bmatrix}$$

$$= \begin{bmatrix} \cos\theta - \frac{\bar{B}}{2}\sin\theta & j\left(\frac{\bar{B}}{2}\cos\theta + \sin\theta - \frac{\bar{B}}{2}\right) \\ j\left(\frac{\bar{B}}{2}\cos\theta + \sin\theta + \frac{\bar{B}}{2}\right) & \cos\theta - \frac{\bar{B}}{2}\sin\theta \end{bmatrix} \begin{bmatrix} V_{n+1} \\ I_{n+1} \end{bmatrix} \tag{8.4}$$

Note that $\mathscr{A} = \mathscr{D}$, which is always true for a symmetrical network, i.e., a symmetrical unit cell.

If the periodic structure is capable of supporting a propagating wave, it is necessary for the voltage and current at the $(n + 1)$st terminal to be equal to the voltage and current at the nth terminal, apart from a phase delay due to a finite propagation time. Thus we assume that

$$V_{n+1} = e^{-\gamma d} V_n \tag{8.5a}$$

$$I_{n+1} = e^{-\gamma d} I_n \tag{8.5b}$$

where $\gamma = j\beta + \alpha$ is the propagation constant for the periodic structure. In

terms of the transmission matrix for a unit cell, we now have

$$\begin{bmatrix} V_n \\ I_n \end{bmatrix} = \begin{bmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{C} & \mathscr{D} \end{bmatrix} \begin{bmatrix} V_{n+1} \\ I_{n+1} \end{bmatrix} = e^{\gamma d} \begin{bmatrix} V_{n+1} \\ I_{n+1} \end{bmatrix}$$

or

$$\left(\begin{bmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{C} & \mathscr{D} \end{bmatrix} - \begin{bmatrix} e^{\gamma d} & 0 \\ 0 & e^{\gamma d} \end{bmatrix} \right) \begin{bmatrix} V_{n+1} \\ I_{n+1} \end{bmatrix} = 0 \tag{8.6}$$

This equation is a matrix eigenvalue equation for γ. A nontrivial solution for V_{n+1}, I_{n+1} exists only if the determinant vanishes. Hence

$$\begin{vmatrix} \mathscr{A} - e^{\gamma d} & \mathscr{B} \\ \mathscr{C} & \mathscr{D} - e^{\gamma d} \end{vmatrix} = \mathscr{A}\mathscr{D} - \mathscr{B}\mathscr{C} + e^{2\gamma d} - e^{\gamma d}(\mathscr{A} + \mathscr{D}) = 0 \tag{8.7}$$

For a reciprocal network the determinant $\mathscr{A}\mathscr{D} - \mathscr{B}\mathscr{C}$ of the transmission matrix equals unity (Sec. 4.9); so we obtain

$$\cosh \gamma d = \frac{\mathscr{A} + \mathscr{D}}{2} \tag{8.8}$$

For the capacitively loaded coaxial line, (8.8), together with (8.4), yields

$$\cosh \gamma d = \cos \theta - \frac{\overline{B}}{2} \sin \theta \tag{8.9}$$

When $|\cos \theta - (\overline{B}/2)\sin \theta| < 1$, we must have $\gamma = j\beta$ and $\alpha = 0$; that is,

$$\cos \beta d = \cos \theta - \frac{\overline{B}}{2} \sin \theta \tag{8.10a}$$

When the right-hand side of (8.9) is greater than unity, $\gamma = \alpha$ and $\beta = 0$; so

$$\cosh \alpha d = \cos \theta - \frac{\overline{B}}{2} \sin \theta > 1 \tag{8.10b}$$

Finally, when the right-hand side of (8.9) is less than -1, we must have $\gamma d = j\pi + \alpha$, so that

$$\begin{aligned} \cosh \gamma d &= \cosh(j\pi + \alpha d) = -\cosh \alpha d \\ &= \cos \theta - \frac{\overline{B}}{2} \sin \theta < -1 \end{aligned} \tag{8.10c}$$

It is apparent, then, that there will be frequency bands for which unattenuated propagation can take place separated by frequency bands in which the wave is attenuated. Note that propagation in both directions is possible since $-\gamma$ is also a solution.

A detailed study of the passband-stopband characteristic is made in Sec. 8.6. For the present we shall confine our attention to the low-frequency limiting value of β. When $d \ll \lambda_0$, $\theta = k_0 d$ is small, and βd will then also

be small. Replacing $\cos\theta$ by $1 - \theta^2/2$ and $\sin\theta$ by θ in (8.10a) gives

$$\cos\beta d \approx 1 - \frac{\beta^2 d^2}{2} = 1 - \frac{k_0^2 d^2}{2} - \frac{\bar{B} k_0 d}{2}$$

Using the relations $k_0^2 = \omega^2 \mu_0 \epsilon_0 = \omega^2 LC$ and $\bar{B} = B/Y_c = \omega C_0 (L/C)^{1/2}$, where $\omega C_0 = B$, we obtain

$$\beta^2 = \omega^2 LC + \frac{\omega^2 L C_0}{d}$$

and hence

$$\beta = \omega \sqrt{L\left(C + \frac{C_0}{d}\right)} \tag{8.11}$$

Therefore we find that, at low frequencies where $d \ll \lambda_0$, the loaded line behaves as an electrically smooth line with a shunt capacitance $C + C_0/d$ per unit length. The increase in β results in a reduction of the phase velocity by a factor k_0/β.

Another parameter of importance in connection with periodic structures is the normalized characteristic impedance $\bar{Z}_B$ presented to the voltage and current waves at the reference terminal plane, i.e., input terminals of a unit cell. An expression for $\bar{Z}_B$ may be obtained from (8.6), which may be written as

$$(\mathscr{A} - e^{\gamma d}) V_{n+1} = -\mathscr{B} I_{n+1}$$
$$-\mathscr{C} V_{n+1} = (\mathscr{D} - e^{\gamma d}) I_{n+1}$$

Hence

$$\frac{Z_B}{Z_c} = \bar{Z}_B = \frac{V_{n+1}}{I_{n+1}} = \frac{-\mathscr{B}}{\mathscr{A} - e^{\gamma d}} = -\frac{\mathscr{D} - e^{\gamma d}}{\mathscr{C}} \tag{8.12}$$

Replacing $2e^{\gamma d}$ by $\mathscr{A} + \mathscr{D} \pm [(\mathscr{A} + \mathscr{D})^2 - 4]^{1/2}$ from (8.7), we obtain

$$\bar{Z}_B^{\pm} = \frac{2\mathscr{B}}{\mathscr{D} - \mathscr{A} \pm \sqrt{(\mathscr{A} + \mathscr{D})^2 - 4}} \tag{8.13a}$$

where the upper and lower signs refer to propagation in the $+z$ and $-z$ directions, respectively. We are using the convention that the positive directions of V_n and I_n are those indicated in Fig. 8.2, independent of the direction of propagation. For a symmetrical network, $\mathscr{A} = \mathscr{D}$, and since $\mathscr{A}\mathscr{D} - \mathscr{B}\mathscr{C} = 1$, we have $\mathscr{A}^2 - 1 = \mathscr{B}\mathscr{C}$. In this case (8.13$a$) reduces to

$$\bar{Z}_B^{\pm} = \frac{2\mathscr{B}}{\pm\sqrt{4\mathscr{A}^2 - 4}} = \pm\sqrt{\frac{\mathscr{B}}{\mathscr{C}}} \tag{8.13b}$$

In general, for a lossless structure, $\bar{Z}_B^- = -(\bar{Z}_B^+)^*$ in the passband, since $|\mathscr{A} + \mathscr{D}| < 2$, as (8.8) shows.

If the unit cell is represented by a T network with parameters $\bar{Z}_{11}$, $\bar{Z}_{12}$, and $\bar{Z}_{22}$, then, by using the relations between the $\mathscr{A}\mathscr{B}\mathscr{C}\mathscr{D}$ parameters and the impedance parameters given in Sec. 4.9, we can also show that

$$\cosh \gamma d = \frac{\bar{Z}_{11} + \bar{Z}_{22}}{2\bar{Z}_{12}} \tag{8.14}$$

$$\bar{Z}_B = \frac{\bar{Z}_{11} - \bar{Z}_{22}}{2} \pm \bar{Z}_{12} \sinh \gamma d \tag{8.15}$$

The waves that may propagate along a periodic structure are often called Bloch waves by analogy with the quantum-mechanical electron waves that may propagate through a periodic crystal lattice in a solid. It is for this reason that we have denoted the characteristic impedance as $\bar{Z}_B$ for the Bloch wave. The voltage and current at the nth terminal plane will be denoted by $V_{Bn}^{\pm}$, $I_{Bn}^{\pm}$ for the Bloch waves from now on instead of by the quantities V_n, I_n. The + and − signs refer to Bloch waves propagating in the $+z$ and $-z$ directions, respectively. We shall also adopt the convention that the positive direction of current flow for Bloch waves is always in the $+z$ direction; thus $I_B^+ = \bar{Y}_B^+ V_B^+$ and $I_B^- = \bar{Y}_B^- V_B^-$. However, for a symmetrical structure such that $\mathscr{A} = \mathscr{D}$, we shall have $\bar{Y}_B^- = -\bar{Y}_B^+ = -(\bar{Z}_B^+)^{-1}$.

If (8.13) is used, we find that, for the loaded coaxial line,

$$\bar{Z}_B = \sqrt{\frac{\mathscr{B}}{\mathscr{C}}} = \sqrt{\frac{2 \sin \theta + \bar{B} \cos \theta - \bar{B}}{2 \sin \theta + \bar{B} \cos \theta + \bar{B}}} \tag{8.16}$$

In the low-frequency limit, where we can replace $\sin \theta$ by

$$\theta = k_0 d = \omega d \sqrt{LC}$$

and $\cos \theta$ by 1, we obtain

$$\bar{Z}_B = \sqrt{\frac{2\theta}{2\theta + 2\bar{B}}} = \sqrt{\frac{C}{C + C_0/d}}$$

and thus

$$Z_B = \bar{Z}_B Z_c = \sqrt{\frac{L}{C + C_0/d}} \tag{8.17}$$

Again we see that, in the low-frequency limit, the loaded line is electrically smooth and the characteristic impedance is modified in the anticipated manner by the effective increase in the shunt capacitance per unit length.

The characteristic impedance of a periodic structure is not a unique quantity since it depends on the choice of terminal planes for a unit cell. If the terminal planes are shifted a distance l in the $-z$ direction, the new

characteristic impedance becomes

$$\bar{Z}'_B = \frac{\bar{Z}_B + j \tan k_0 l}{1 + j\bar{Z}_B \tan k_0 l} \tag{8.18}$$

8.2 WAVE ANALYSIS OF PERIODIC STRUCTURES

Periodic structures may be analyzed in terms of the forward- and backward-propagating waves that can exist in each unit cell with about the same facility as the network approach gives. In the wave approach the wave-amplitude transmission matrix $[A]$ discussed in Sec. 4.9 is used.

With reference to Fig. 8.3, let the amplitudes of the forward- and backward-propagating waves at the nth and $(n + 1)$st terminal plane be c_n^+, c_n^-, c_{n+1}^+, and c_{n+1}^-. The c_{n+1}^+, c_{n+1}^- are related to the c_n^+, c_n^- by the wave-amplitude transmission matrix as follows:

$$\begin{bmatrix} c_n^+ \\ c_n^- \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} c_{n+1}^+ \\ c_{n+1}^- \end{bmatrix} \tag{8.19}$$

The solution for a Bloch wave requires $c_{n+1}^+ = e^{-\gamma d} c_n^+$ and $c_{n+1}^- = e^{-\gamma d} c_n^-$. Hence (8.19) becomes

$$\begin{bmatrix} A_{11} - e^{\gamma d} & A_{12} \\ A_{21} & A_{22} - e^{\gamma d} \end{bmatrix} \begin{bmatrix} c_{n+1}^+ \\ c_{n+1}^- \end{bmatrix} = 0 \tag{8.20}$$

A nontrivial solution for c_{n+1}^+, c_{n+1}^- is obtained only if the determinant vanishes. Consequently, the eigenvalue equation for γ is

$$A_{11}A_{22} - A_{12}A_{21} + e^{2\gamma d} - e^{\gamma d}(A_{11} + A_{22}) = 0$$

or

$$\cosh \gamma d = \frac{A_{11} + A_{22}}{2} \tag{8.21}$$

since the determinant of the transmission matrix, that is, $A_{11}A_{22} - A_{12}A_{21}$, equals 1 when normalized wave amplitudes are used.

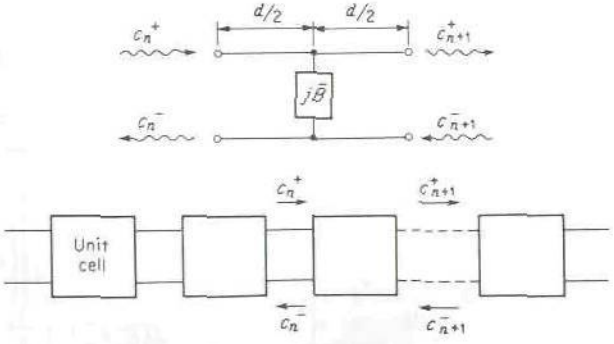


FIGURE 8.3
Wave amplitudes in a periodic structure.

The Bloch wave which can propagate in the periodic structure is made up from forward- and backward-propagating normal transmission-line or waveguide waves that exist between discontinuities. When γ has been determined from (8.21), the ratio of c_n^- to c_n^+ is fixed. This ratio is called the characteristic reflection coefficient Γ_B. Thus the transverse electric field of the Bloch wave will have an amplitude

$$V_{B0} = c_0^+ + c_0^- = c_0^+(1 + \Gamma_B)$$

at the zeroth terminal plane and an amplitude

$$V_{Bn} = c_n^+ + c_n^- = c_n^+(1 + \Gamma_B) = c_0^+(1 + \Gamma_B)e^{-\gamma nd} \tag{8.22a}$$

at the nth terminal plane. The transverse magnetic field of the Bloch wave will have an amplitude

$$I_{Bn} = c_0^+(1 - \Gamma_B)e^{-\gamma nd} \tag{8.22b}$$

at the nth terminal plane.

The characteristic reflection coefficient may be found from the pair of equations (8.20) by eliminating $e^{-\gamma d}$ by the use of (8.21). It is usually more convenient to express Γ_B in terms of $\bar{Z}_B$ by using the relation $\bar{Z}_B = (1 + \Gamma_B)/(1 - \Gamma_B)$. Thus we have

$$\Gamma_B^{\pm} = \frac{\bar{Z}_B^{\pm} - 1}{\bar{Z}_B^{\pm} + 1} \tag{8.23}$$

where the + and − signs refer to Bloch waves propagating in the $+z$ and $-z$ directions, respectively.

The above wave formulation is now applied to the capacitively loaded transmission line discussed earlier. The unit cell is chosen as in Fig. 8.3. The wave-amplitude transmission matrices for the three sections of the unit cell are (Sec. 4.9 and Prob. 8.7)

$$\begin{bmatrix} e^{jk_0 d/2} & 0 \\ 0 & e^{-jk_0 d/2} \end{bmatrix} \begin{bmatrix} \dfrac{2 + j\bar{B}}{2} & j\dfrac{\bar{B}}{2} \\ -j\dfrac{\bar{B}}{2} & \dfrac{4 + \bar{B}^2}{2(2 + j\bar{B})} \end{bmatrix}$$

and another matrix like the first one. The $[A]$ matrix for the unit cell is obtained by multiplying the three component matrices together; thus

$$[A] = \begin{bmatrix} e^{j\theta/2} & 0 \\ 0 & e^{-j\theta/2} \end{bmatrix} \begin{bmatrix} \dfrac{2 + j\bar{B}}{2} & j\dfrac{\bar{B}}{2} \\ -j\dfrac{\bar{B}}{2} & \dfrac{4 + \bar{B}^2}{2(2 + j\bar{B})} \end{bmatrix} \begin{bmatrix} e^{j\theta/2} & 0 \\ 0 & e^{-j\theta/2} \end{bmatrix}$$

where $\theta = k_0 d$. After multiplication we obtain

$$[A] = \begin{bmatrix} \dfrac{2 + j\overline{B}}{2} e^{j\theta} & j\dfrac{\overline{B}}{2} \\ -j\dfrac{\overline{B}}{2} & \dfrac{4 + \overline{B}^2}{2(2 + j\overline{B})} e^{-j\theta} \end{bmatrix} \tag{8.24}$$

Making use of (8.21), we find that

$$\cosh \gamma d = \frac{(4 + \overline{B}^2)e^{-j\theta} + (2 + j\overline{B})^2 e^{j\theta}}{4(2 + j\overline{B})} = \cos\theta - \frac{\overline{B}}{2}\sin\theta$$

which is the same as (8.9) obtained earlier.

8.3 PERIODIC STRUCTURES COMPOSED OF UNSYMMETRICAL TWO-PORT NETWORKS

The capacitively loaded coaxial transmission line can be considered as made up of symmetrical sections by choosing terminal planes midway between each diaphragm. For other choices of terminal-plane positions the unit cell would be unsymmetrical, and its equivalent T network would then also be unsymmetrical. Other types of periodic structures are composed of intrinsically unsymmetrical unit cells such that there is no terminal-plane location that will reduce them to a symmetrical structure. Several unsymmetrical structures are illustrated in Fig. 8.4.

For nonsymmetrical structures the Bloch-wave characteristic impedance is given by (8.15), which we rewrite as

$$\overline{Z}_B^+ = \zeta + \overline{Z} \tag{8.25a}$$

$$\overline{Z}_B^- = \zeta - \overline{Z} \tag{8.25b}$$

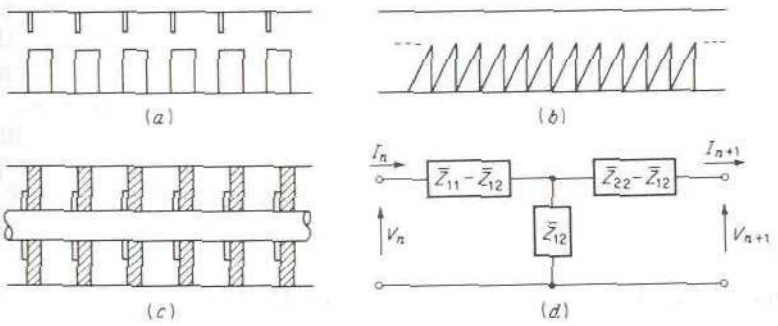


FIGURE 8.4
Periodic structures with unsymmetrical unit cells. (a, b) Rectangular waveguide loaded with thick unsymmetrical diaphragms; (c) coaxial line loaded with diaphragms and dielectric rings; (d) equivalent T network of a unit cell.

where

$$\zeta = \frac{\bar{Z}_{11} - \bar{Z}_{22}}{2} \tag{8.26a}$$

$$\bar{Z} = \pm\bar{Z}_{12}\sinh\gamma d = \pm j\bar{Z}_{12}\sin\beta d \tag{8.26b}$$

and the sign is to be chosen so that $\bar{Z}$ has a positive real part. The phase constant β is given by

$$\cos\beta d = \frac{\bar{Z}_{11} + \bar{Z}_{22}}{2\bar{Z}_{12}} \tag{8.26c}$$

in the propagation band. The physical length of a unit cell is d. The quantities $\bar{Z}_B^{\pm}$ and βd are often called the iterative parameters of the T network. A consequence of the nonsymmetry of the unit cell is that $\bar{Z}_B^{+}$ is different from $\bar{Z}_B^{-}$.

Let the voltage of the Bloch wave at the zeroth terminal plane be $V_{B0}^{\pm}$, where the signs + or − refer to Bloch waves propagating in the $+z$ and $-z$ direction, respectively. The corresponding Bloch-wave current is $I_{B0}^{\pm} = V_{B0}^{\pm}/\bar{Z}_B^{\pm}$. At the nth terminal plane the Bloch-wave voltages and currents will be

$$V_{Bn}^{\pm} = V_{B0}^{\pm}e^{\mp\gamma nd} \tag{8.27a}$$

$$I_{Bn}^{\pm} = \frac{V_{B0}^{\pm}}{\bar{Z}_B^{\pm}}e^{\mp\gamma nd} \tag{8.27b}$$

Recall that we are taking the positive direction of current flow to be in the $+z$ direction, independent of the direction of propagation for the Bloch waves.

If the restriction is made that the only points at which the voltages and currents will be specified are the terminal planes, the periodic structure has properties similar to any uniform transmission line or waveguide. As such, transmission-line theory can be applied to study the effects of terminating a periodic structure in an arbitrary load impedance, to design matching sections for periodic structures, etc. These applications are discussed in the following two sections.

8.4 TERMINATED PERIODIC STRUCTURES

Figure 8.5 illustrates a periodic structure terminated in a load impedance $\bar{Z}_L$ at the Nth terminal plane. The total voltage and current at the nth terminal plane will be a superposition of an incident and reflected Bloch

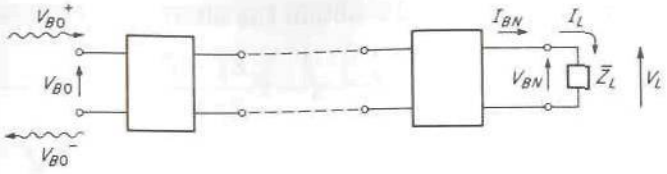


FIGURE 8.5
Periodic structure terminated in a load $\bar{Z}_L$.

wave; thus

$$V_{Bn} = V_{B0}^+ e^{-j\beta nd} + V_{B0}^- e^{j\beta nd} \tag{8.28a}$$

$$\begin{aligned} I_{Bn} &= I_{B0}^+ e^{-j\beta nd} + I_{B0}^- e^{j\beta nd} \\ &= V_{B0}^+ \bar{Y}_B^+ e^{-j\beta nd} + V_{B0}^- \bar{Y}_B^- e^{j\beta nd} \end{aligned} \tag{8.28b}$$

where $\bar{Y}_B = \bar{Z}_B^{-1}$. At the Nth terminal plane we must have

$$V_L = V_{BN} = \bar{Z}_L I_{BN} = \bar{Z}_L I_L$$

and hence

$$V_{BN}^+ + V_{BN}^- = \bar{Z}_L\left(\bar{Y}_B^+ V_{BN}^+ + \bar{Y}_B^- V_{BN}^-\right) \tag{8.29}$$

The reflection coefficient Γ_L of the load for Bloch waves is, from (8.29),

$$\begin{aligned} \Gamma_L = \frac{V_{BN}^-}{V_{BN}^+} &= -\frac{\bar{Z}_L \bar{Y}_B^+ - 1}{\bar{Z}_L \bar{Y}_B^- - 1} = -\frac{\bar{Z}_B^-}{\bar{Z}_B^+}\frac{\bar{Z}_L - \bar{Z}_B^+}{\bar{Z}_L - \bar{Z}_B^-} \\ &= \frac{\bar{Z} - \zeta}{\bar{Z} + \zeta}\frac{\bar{Z}_L - \bar{Z} - \zeta}{\bar{Z}_L + \bar{Z} - \zeta} \end{aligned} \tag{8.30}$$

For a symmetrical structure $\zeta = 0$ and the expression for Γ_L reduces to the usual form.

The Bloch-wave reflection coefficient at the nth terminal plane is

$$\Gamma_n = \frac{V_{Bn}^-}{V_{Bn}^+} = \frac{V_{BN}^- e^{-j(N-n)\beta d}}{V_{BN}^+ e^{j(N-n)\beta d}} = \Gamma_L e^{-j2(N-n)\beta d} \tag{8.31}$$

The input impedance at the nth terminal plane is

$$\begin{aligned} \bar{Z}_n = \frac{V_{Bn}^+ + V_{Bn}^-}{I_{Bn}^+ + I_{Bn}^-} &= \frac{V_{Bn}^+(1 + \Gamma_n)}{V_{Bn}^+ \bar{Y}_B^+ + V_{Bn}^- \bar{Y}_B^-} \\ &= \frac{1 + \Gamma_n}{\bar{Y}_B^+ + \Gamma_n \bar{Y}_B^-} = \frac{\bar{Z}_B + \bar{Z}_B^-(1 + \Gamma_n)}{\bar{Z}_B^- + \bar{Z}_B^+ \Gamma_n} \end{aligned} \tag{8.32}$$

From (8.32) we can also obtain the alternative expressions

$$\bar{Z}_n - \zeta = \bar{Z}\frac{\bar{Z}_B^- - \bar{Z}_B^+\Gamma_n}{\bar{Z}_B^- + \bar{Z}_B^+\Gamma_n} \tag{8.33a}$$

$$\Gamma_n = -\frac{\bar{Z}_B^-}{\bar{Z}_B^+}\frac{(\bar{Z}_n - \zeta) - \bar{Z}}{(\bar{Z}_n - \zeta) + \bar{Z}} \tag{8.33b}$$

If (8.31) is used to express Γ_n in terms of Γ_L and (8.30) is used to express Γ_L in terms of $\bar{Z}_L$, we find that (8.33*a*) gives

$$\bar{Z}_n - \zeta = \bar{Z}\frac{\bar{Z}_L - \zeta + j\bar{Z}\tan(N-n)\beta d}{\bar{Z} + j(\bar{Z}_L - \zeta)\tan(N-n)\beta d} \tag{8.34}$$

This equation gives the transformation of impedance along a periodic structure. It differs somewhat from the usual transmission-line formula when the unit cell is unsymmetrical, so that $\zeta \neq 0$.

For a Bloch wave propagating in the $+z$ direction, the periodic structure must be terminated in a load $\bar{Z}_L = \bar{Z}_B^+ = \zeta + \bar{Z}$ to avoid a reflected wave. Similarly, the matched-load termination for a Bloch wave propagating in the $-z$ direction is $-\bar{Z}_B^- = \bar{Z} - \zeta$. The two characteristic Bloch-wave impedances are the iterative impedances for the T network of the unit cell. With voltages and currents chosen as in Fig. 8.6, it is readily shown that an impedance $\bar{Z}_B^+$ connected at terminals 2 is transformed into itself at terminals 1. Similarly, an impedance $-\bar{Z}_B^-$ connected at terminals 1 is transformed into itself at terminals 2. It is for this reason that $\bar{Z}_B^\pm$ is called an iterative (repeating) impedance. For a lossless T network, ζ is pure imaginary and $\bar{Z}$ is pure real in the propagation band. Ambiguity in the sign of $\bar{Z}$ as given by (8.15) or (8.26*b*) may be avoided by noting that, in a passband, $\bar{Z}$ is real and positive, in order to be consistent with our choice of positive direction for current. We must have positive real power transmission, and hence

$$P = \text{Re}\tfrac{1}{2}V_B^+(I_B^+)^* = \text{Re}\tfrac{1}{2}|I_B^+|^2 Z_B^+ = \tfrac{1}{2}|I_B^+|^2\bar{Z} > 0 \tag{8.35}$$

Another criterion that may be used is the one requiring the reactive part of

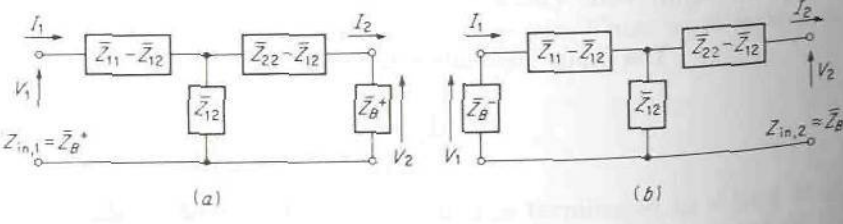


FIGURE 8.6
Iterative impedance properties of a T network.

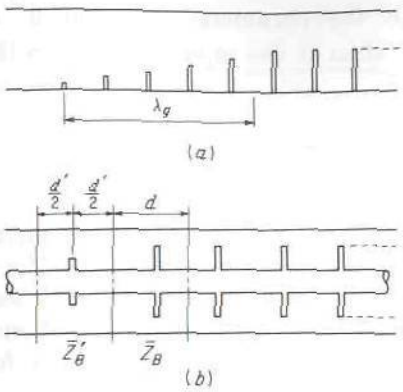


FIGURE 8.7
(*a*) Tapered transition matching section for a diaphragm-loaded rectangular guide; (*b*) quarter-wave transformer matching of a capacitively loaded coaxial line.

$\bar{Z}_B^+$ to have a positive derivative with respect to ω (Sec. 4.3). There is also reactive power in a Bloch wave in a passband, and this is given by

$$P_{\text{reactive}} = \tfrac{1}{2}|I_B^+|^2\zeta \tag{8.36}$$

Complex power for a Bloch wave propagating in the $-z$ direction is given by $-\frac{1}{2}V_B^-(I_B^-)^*$ because of our choice of direction for positive current.

8.5 MATCHING OF PERIODIC STRUCTURES

If a periodic structure is connected to a smooth transmission line or waveguide, some means of matching the periodic structure to the input waveguide must be provided to avoid reflection of the incident power. A situation encountered quite frequently is the one where the periodic structure is identical with the input waveguide apart from the periodic loading. One way of providing a matched transition from the unloaded to the loaded waveguide is to use a tapered intermediate section. The matching taper section is similar to the loaded waveguide except that the periodic loading is gradually reduced to zero over a distance of about a wavelength. Figure 8.7*a* illustrates a tapered transition in a rectangular waveguide connected to a similar guide periodically loaded with diaphragms.

Any of the matching networks discussed in Chap. 5 may also be used to match a periodically loaded guide to an unloaded guide. For example, at some distance $d'/2$ in front of the first terminal plane for the periodic structure, the characteristic admittance $\bar{Y}_B$ of the periodic structure is transformed into an admittance $1 - j\bar{B}'$, so that placing a shunt susceptance $j\bar{B}'$ at this point provides a matched transition. Matching by means of a shunt susceptance may be viewed as an application of the quarter-wave transformer matching technique. The unit cell consisting of the shunt susceptance $j\bar{B}'$ plus a length $d'/2$ of transmission line (or waveguide) on either side, as in Fig. 8.7*b*, may be considered as part of an infinite periodic structure with a propagation phase constant β' and a normalized character-

istic impedance $\bar{Z}'_B$. If the parameters $\bar{B}'$ and d' are chosen so that $\beta' d' = \pi/2$, $\bar{Z}'_B = \bar{Z}_B^{1/2}$, then at the input terminal to the matching section, (8.34) gives

$$\bar{Z}_{\text{in}} = \frac{(\bar{Z}'_B)^2}{\bar{Z}_B} = 1 \tag{8.37}$$

Note that the matching section is a symmetrical structure, so that $\bar{Z}'_B = \bar{Z}'$ and $\zeta' = 0$. We also require that $\bar{Z}_B$ be real in order for (8.37) to have a real solution for $\bar{Z}'_B$. When these conditions are met, we see that the matching section behaves essentially as a quarter-wave transformer. For symmetrical structures the required values of $\bar{B}'$ and d' may be found from (8.8) and (8.13*b*); thus, as $\cos \beta' d' = \cos(\pi/2) = 0$, we have

$$\mathscr{A}' = \mathscr{D}' = 0 \tag{8.38a}$$

$$(\bar{Z}'_B)^2 = \frac{\mathscr{B}'}{\mathscr{C}'} = \bar{Z}_B \tag{8.38b}$$

For the capacitively loaded transmission line, we obtain, by using (8.4) and (8.16) applied to the matching section,

$$2 \cot k_0 d' = \bar{B}' \tag{8.39a}$$

$$(\bar{Z}'_B)^2 = \frac{2 \sin k_0 d' + \bar{B}' \cos k_0 d' - \bar{B}'}{2 \sin k_0 d' + \bar{B}' \cos k_0 d' + \bar{B}'} = \tan^2 \frac{k_0 d'}{2} = \bar{Z}_B \tag{8.39b}$$

when (8.39*a*) is used to eliminate $\bar{B}'$. The above results are the equivalent of those derived in Chap. 5, i.e., given by (5.8) and (5.9).

To obtain a match over a wide frequency band, more elaborate matching networks must be used since a single shunt susceptance usually does not provide a match over a wide frequency band. Broadband matching is complicated by the fact that the characteristic impedance $\bar{Z}_B$ of a periodic structure is a function of frequency. No general technique exists for designing broadband matching networks because of the general nature of $\bar{Z}_B$. Each periodic structure must be considered by itself so that the frequency variation in $\bar{Z}_B$ can be incorporated into the design. For this reason, the matching problem is not discussed any further.

8.6 k_0-β DIAGRAM

We now turn to a detailed study of the passband-stopband characteristics of the capacitively loaded coaxial transmission line discussed in the earlier sections. The information contained in the eigenvalue equation for the propagation constant β in a periodic structure is usually plotted on a k_0-β (or ω-β) plane. The curves of β versus k_0 show immediately the frequency bands for propagation and also the stopbands in which no propagation takes

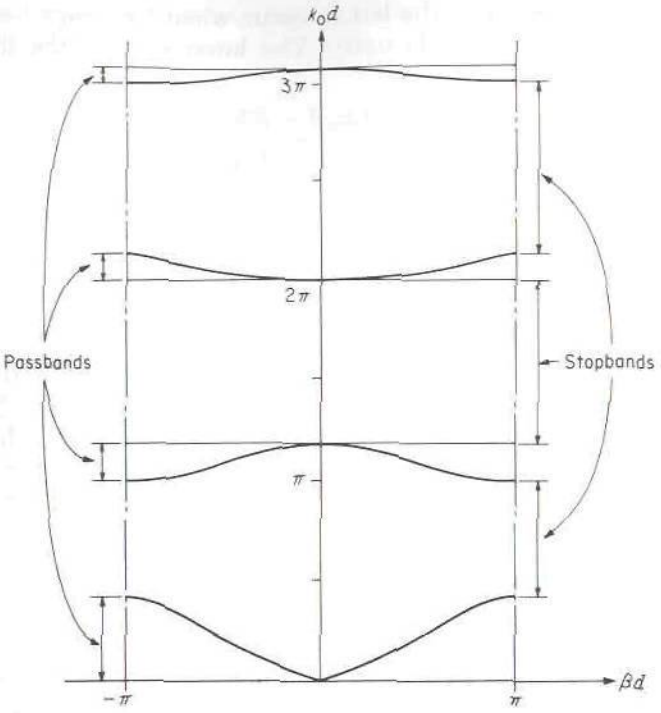


FIGURE 8.8
k_0d-βd diagram for a capacitively loaded coaxial line, $\overline{B} = 2k_0d$.

place. The resultant plot is called the k_0-β diagram, or the Brillouin diagram.†

For the capacitively loaded coaxial line, (8.9) gave

$$\cos \beta d = \cos k_0 d - \frac{\overline{B}}{2} \sin k_0 d = \cos k_0 d - K k_0 d \sin k_0 d \quad (8.40)$$

where $\overline{B}/2 = \omega C_0/2Y_c$ has been expressed as Kk_0d. Curves of k_0d versus βd are sketched in Fig. 8.8 for $K = 1$, that is, for $\overline{B} = 2k_0d$. A low-frequency passband exists for $0 < k_0d < 0.416\pi$. This passband is followed by a stopband and further alternating passbands and stopbands. As k_0d becomes large, the loading is increased, since $\overline{B}$ increases with k_0. This has the effect of decreasing the width of the passbands in terms of frequency.

†Named after Brillouin, who used diagrams of this sort to illustrate the energy-band structure in periodic crystalline media.

The edges of the bands occur when the magnitude of the right-hand side of (8.40) exceeds unity. The lower edge of the first passband occurs when $0 < k_0 d < \pi$ and

$$\cos k_0 d - K k_0 d \sin k_0 d = -1$$

This equation may be solved for $k_0 d$ to give

$$\cot \frac{k_0 d}{2} = K k_0 d \tag{8.41a}$$

$$\cos \frac{k_0 d}{2} = 0 \tag{8.41b}$$

The corresponding principal value of βd is π, and the values of $k_0 d$ obtained from (8.41) mark the edges of all the bands for this value of βd. The edges of the bands where $\beta d = 0$ are obtained by equating (8.40) to unity, in which case we obtain

$$\tan \frac{k_0 d}{2} = -K k_0 d \tag{8.42a}$$

$$\sin \frac{k_0 d}{2} = 0 \tag{8.42b}$$

One edge of the passband always occurs when the spacing between discontinuities equals one-half wavelength in the unloaded waveguide, in the present case, when $k_0 d$ is a multiple of π. When the spacing between discontinuities equals one-half wavelength, they may all be lumped together, with the result that the line becomes effectively loaded at a single point by an infinite susceptance (or reactance). Clearly, power transmission along the periodic structure must reduce to zero at this frequency.

Only the principal value of βd is plotted in Fig. 8.8. In addition, $\beta d + 2n\pi$, where n is an arbitrary integer, are solutions. These other solutions are the propagation constants of the spatial harmonics into which the Bloch wave may be expanded. The spatial harmonics are discussed in Sec. 8.8.

The k_0-β diagram for other types of periodic structures exhibit features similar to those in Fig. 8.8. For example, if the capacitive loading is replaced by inductive shunt loading, the relative locations of the passbands and stopbands are interchanged. The zero-frequency region will be a stopband since the shunt inductors will short-circuit the line at zero frequency.

*8.7 GROUP VELOCITY AND ENERGY FLOW

The phase velocity for a Bloch wave in a periodic structure is given by

$$v_p = \frac{\omega}{\beta} = \frac{k_0}{\beta} c = \frac{k_0 d}{\beta d} c \tag{8.43}$$

With reference to Fig. 8.9, it is seen that $k_0 d / \beta d$ is the slope of the line

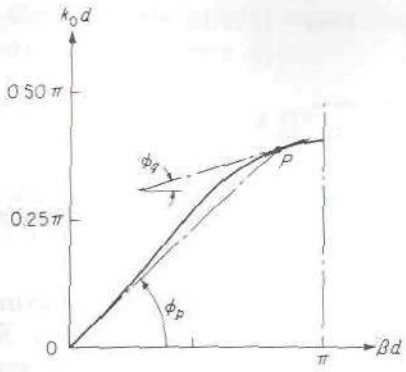


FIGURE 8.9
Enlarged drawing of first passband for a capacitively loaded transmission line, $\overline{B} = 2k_0 d$.

from the origin to a point P on the $k_0 d$-βd diagram. Since β is a function of ω, the periodic structure has frequency dispersion. The group velocity v_g as given by (Sec. 3.19)

$$v_g = \frac{d\omega}{d\beta} = c\frac{d(k_0 d)}{d(\beta d)} \tag{8.44}$$

is therefore different from the phase velocity. Again referring to Fig. 8.9, it is seen that the group velocity is equal to the slope of the tangent to the curve of k_0 versus β multiplied by the velocity of light c. Thus we have

$$v_p = c \tan \phi_p \qquad v_g = c \tan \phi_g$$

where ϕ_p and ϕ_g are the angles given in Fig. 8.9.

For the capacitively loaded coaxial line, use of the eigenvalue equation for β, that is, (8.10a), enables us to obtain

$$v_g = c\frac{dk_0}{d\beta} = c\frac{d\theta}{d(\beta d)} = \frac{c \sin \beta d}{\left(\dfrac{\overline{B}}{2k_0 d} + 1\right)\sin k_0 d + \dfrac{\overline{B}}{2}\cos k_0 d} \tag{8.45}$$

This expression shows that the group velocity becomes zero when $\beta d = 0$ or π, except when k_0 also equals zero. Thus, as the edges of the passbands are approached, the group velocity goes to zero.

The group velocity is also the signal velocity for any signal consisting of a sufficiently narrow band of frequencies such that β can be approximated by a linear function of ω throughout the band. The signal delay τ for propagation through a unit cell is given by

$$\tau = \frac{d}{v_g} \tag{8.46}$$

In Sec. 3.19 it was shown that for a waveguide, which is a dispersive medium, the velocity of energy flow in a propagating wave was equal to the group velocity. The same result will be shown to hold for a lossless periodic structure also.

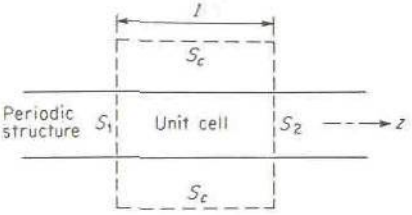


FIGURE 8.10
A unit cell of a periodic structure.

Consider a unit cell of a lossless periodic structure as in Fig. 8.10. A surface S is chosen to consist of surfaces S_1 and S_2 at the input and output terminal planes plus a cylindrical surface S_c surrounding the structure. If the periodic structure is enclosed by a perfectly conducting waveguide, the surface S_c coincides with the waveguide wall. If the periodic structure is an open-boundary structure, the surface S_c is that of a cylinder with infinite radius. In both cases $\mathbf{n} \times \mathbf{E}$ vanishes on S_c, so that the Poynting vector is zero over this surface. For generality we shall let the unit cell contain regions with frequency-dispersive material, i.e., material with parameters μ and ϵ that are functions of ω.

In the derivation of Foster's reactance theorem in Sec. 4.3, it was shown that [see (4.24*a*)]

$$\oint_S \left(\mathbf{E} \times \frac{\partial \mathbf{H}^*}{\partial \omega} + \frac{\partial \mathbf{E}^*}{\partial \omega} \times \mathbf{H} \right) \cdot d\mathbf{S} = -j \int_V \left(\mathbf{H} \cdot \mathbf{H}^* \frac{\partial \omega \mu}{\partial \omega} + \mathbf{E} \cdot \mathbf{E}^* \frac{\partial \omega \epsilon}{\partial \omega} \right)$$

$$= -4j(W_m + W_e) \tag{8.47a}$$

since the latter integral is equal to four times the time-average energy stored in the volume bounded by S. Since the Poynting vector is zero on S_c and $d\mathbf{S}$ is directed inward, we have

$$\int_{S_1} \left(\mathbf{E}_1 \times \frac{\partial \mathbf{H}_1^*}{\partial \omega} + \frac{\partial \mathbf{E}_1^*}{\partial \omega} \times \mathbf{H}_1 \right) \cdot \mathbf{a}_z \, dS$$

$$- \int_{S_2} \left(\mathbf{E}_2 \times \frac{\partial \mathbf{H}_2^*}{\partial \omega} + \frac{\partial \mathbf{E}_2^*}{\partial \omega} \times \mathbf{H}_2 \right) \cdot \mathbf{a}_z \, dS = -4j(W_m + W_e) \tag{8.47b}$$

where $\mathbf{E}_1, \mathbf{H}_1$ are the fields at terminal plane 1 and $\mathbf{E}_2, \mathbf{H}_2$ are the fields at terminal plane 2. For a Bloch wave, $\mathbf{E}_2 = \mathbf{E}_1 e^{-j\beta l}$, where βl is the phase shift through a unit cell of length l. We thus find that

$$\mathbf{E}_2 \times \frac{\partial \mathbf{H}_2^*}{\partial \omega} + \frac{\partial \mathbf{E}_2^*}{\partial \omega} \times \mathbf{H}_2 = \mathbf{E}_1 \times \frac{\partial \mathbf{H}_1^*}{\partial \omega} + jl \frac{d\beta}{d\omega} \mathbf{E}_1 \times \mathbf{H}_1^*$$

$$+ \frac{\partial \mathbf{E}_1^*}{\partial \omega} \times \mathbf{H}_1 + jl \frac{d\beta}{d\omega} \mathbf{E}_1^* \times \mathbf{H}_1$$

Consequently, (8.47*b*) gives (note that the integral over S_2 can be evaluated as an integral over S_1)

$$-2jl\frac{d\beta}{d\omega}\mathrm{Re}\int_{S_1}\mathbf{E}_1\times\mathbf{H}_1^*\cdot\mathbf{a}_z\,dS=-4jl\frac{d\beta}{d\omega}P=-4j(W_m+W_e)\quad(8.47b)$$

where $P=\frac{1}{2}\mathrm{Re}\int_{S_1}\mathbf{E}_1\times\mathbf{H}_1^*\cdot\mathbf{a}_z\,dS$ *is the power transmitted across a terminal plane.* We now see that

$$v_g=\frac{d\omega}{d\beta}=\frac{P}{(W_m+W_e)/l}\qquad(8.48)$$

But the energy density $(W_m+W_e)/l$ in a unit cell multiplied by the velocity of energy flow is equal to the power P, and therefore the group velocity is the velocity of energy flow.

8.8 FLOQUET'S THEOREM AND SPATIAL HARMONICS

It has been noted that in an infinite periodic structure the field of a Bloch wave repeats at every terminal plane except for a propagation factor $e^{-\gamma d}$, *where d is the length of a unit cell.* Since the choice of location of a terminal plane within a unit cell is arbitrary, we see that the field at any point in a unit cell will take on exactly the same value at a similar point in any other unit cell except for a propagation factor $e^{-\gamma d}$ from one cell to the next. Thus, if the field in the unit cell between $0\le z\le d$ is $\mathbf{E}(x,y,z),\mathbf{H}(x,y,z)$, the field in the unit cell located in the region $d\le z\le 2d$ must be

$$e^{-\gamma d}\mathbf{E}(x,y,z-d),e^{-\gamma d}\mathbf{H}(x,y,z-d)$$

Consequently, the field in a periodic structure is described by a solution of the form

$$\mathbf{E}(x,y,z)=e^{-\gamma z}\mathbf{E}_p(x,y,z)\qquad(8.49a)$$

$$\mathbf{H}(x,y,z)=e^{-\gamma z}\mathbf{H}_p(x,y,z)\qquad(8.49b)$$

where $\mathbf{E}_p$ and $\mathbf{H}_p$ *are periodic functions of z with period d; for example,*

$$\mathbf{E}_p(x,y,z+nd)=\mathbf{E}_p(x,y,z)\qquad(8.49c)$$

The possibility of expressing the field in a periodic structure in the form given by (8.49) is often referred to as Floquet's theorem.† From (8.49*a*) we

†Actually, Floquet's work dealt with differential equations with periodic coefficients. The case of periodic boundary conditions is an extension of that work.

see that the electric field at $z_1 + d$ is related to the field at z_1 as follows:

$$\mathbf{E}(x, y, z_1 + d) = e^{-\gamma(z_1+d)}\mathbf{E}_p(x, y, z_1 + d)$$
$$= e^{-\gamma(z_1+d)}\mathbf{E}_p(x, y, z_1) = e^{-\gamma d}\mathbf{E}(x, y, z_1)$$

which has, indeed, the correct repetitive properties of a Bloch wave.

Any periodic function such as $\mathbf{E}_p(x, y, z)$ may be expanded into an infinite Fourier series; thus

$$\mathbf{E}_p(x, y, z) = \sum_{n=-\infty}^{\infty} \mathbf{E}_{pn}(x, y)e^{-j2n\pi z/d} \tag{8.50}$$

where $\mathbf{E}_{pn}$ are vector functions of x and y. Multiplying both sides by $e^{j2m\pi z/d}$ and integrating over a unit cell, i.e., from $z = 0$ to d, give

$$\mathbf{E}_{pm}(x, y) = \frac{1}{d}\int_0^d \mathbf{E}_p(x, y, z)e^{j2m\pi z/d}\, dz \tag{8.51}$$

since the exponential functions form a complete orthogonal set; i.e.,

$$\int_0^d e^{-j2n\pi z/d}e^{j2m\pi z/d}\, dz = \begin{cases} 0 & m \neq n \\ d & m = n \end{cases}$$

The field in a periodic structure can now be represented as

$$\mathbf{E}(x, y, z) = \sum_{n=-\infty}^{\infty} \mathbf{E}_{pn}(x, y)e^{-j\beta z - j2n\pi z/d}$$
$$= \sum_{n=-\infty}^{\infty} \mathbf{E}_{pn}(x, y)e^{-j\beta_n z} \tag{8.52}$$

where $\gamma = j\beta$ and $\beta_n = \beta + 2n\pi/d$. Each term in this expansion is called a spatial harmonic (or a Hartree harmonic) and has a propagation phase constant β_n. Some of the β_n will be negative whenever the integer n is sufficiently negative. The corresponding phase velocity of the nth spatial harmonic is

$$v_{pn} = \frac{\omega}{\beta_n} = \frac{\omega}{\beta + 2n\pi/d} \tag{8.53}$$

and will be negative whenever β_n is negative. The group velocity of the nth harmonic is

$$v_{gn} = \frac{d\omega}{d\beta_n} = \left(\frac{d\beta_n}{d\omega}\right)^{-1} = \left(\frac{d\beta}{d\omega}\right)^{-1} = v_g \tag{8.54}$$

and is the same for all harmonics. From the above relations it is seen that some of the spatial harmonics (approximately one-half) have phase and group velocities that are directed in opposite directions. This property is made use of in the backward-wave traveling-wave-tube oscillator. The term backward wave, or reverse wave, is often used to refer to a wave with

oppositely directed phase and group velocities. The voltage and current waves can, of course, also be expanded into an infinite set of spatial harmonics (Prob. 8.10).

Although a Bloch wave can be expanded into an infinite set of spatial harmonics, all the spatial harmonics must be simultaneously present in *order that the total field may satisfy all the boundary conditions.* The eigenvalue equation for β for a periodic structure always yields solutions $\beta_n = \beta + 2n\pi/d$, in addition to the fundamental solution. *These other* possible solutions are clearly the propagation constants of the spatial harmonics. A complete $k_0 d$-βd diagram thus exhibits $k_0 d$ as a periodic function of βd; that is, the βd curve is continued periodically outside the range $-\pi \leq \beta d \leq \pi$. The slope of the line from the origin to any point on the curve still gives the phase velocity, and the slope of the tangent to the curve gives the group velocity, when multiplied by c.

8.9 PERIODIC STRUCTURES FOR TRAVELING-WAVE TUBES

Traveling-wave tubes require a structure capable of supporting an electromagnetic wave with a phase velocity equal to the velocity of the electron beam. Since the latter is usually much smaller than the velocity of light, the required structure is commonly referred to as a slow-wave structure. A common type of slow-wave circuit used in traveling-wave tubes is the helix. The helix is treated in the following two sections, and hence this section is restricted to a discussion of some of the other types of slow-wave periodic structures suitable for use in traveling-wave tubes.

A periodic slow-wave structure often used for the linear magnetron tube is the vane-type structure illustrated in Fig. 8.11. It consists essentially of a corrugated plane with thick teeth. It will be instructive to apply Floquet's theorem and carry out an analysis of this structure in order to illustrate the general techniques employed. Edge effects at $x = \pm a/2$ will be neglected for simplicity; i.e., we shall treat the structure as being infinitely wide. If a is large compared with the spacing b, and this in turn is small compared with λ_0, the edge effects will not produce a significant

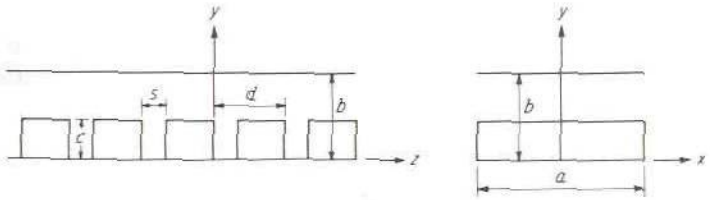


FIGURE 8.11
Vane-type, or corrugated-plane, periodic structure.

change in the characteristics of the ideal structure. For use in a magnetron, a strong axial electric field is required, and hence we shall examine the possibility of having TM- or E-type modes.

For TM modes having no variation with x, the field components may be expressed in terms of the single magnetic field component H_x that is present. We have

$$\nabla \times \mathbf{H} = -\mathbf{a}_x \times \nabla H_x = j\omega\epsilon_0 \mathbf{E}$$

and so

$$E_y = -j\frac{Z_0}{k_0}\frac{\partial H_x}{\partial z} \tag{8.55a}$$

$$E_z = j\frac{Z_0}{k_0}\frac{\partial H_x}{\partial y} \tag{8.55b}$$

The field H_x is a solution of

$$\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + k_0^2\right) H_x = 0 \tag{8.56}$$

According to Floquet's theorem, the field H_x can be expressed in the form $e^{-j\beta z}\psi(y, z)$, where $\psi(y, z)$ is periodic in z with a period d. Hence we shall assume that

$$H_x = \sum_{n=-\infty}^{\infty} f_n(y) e^{-j\beta_n z}$$

where $\beta_n = \beta + 2n\pi/d$ and the $f_n(y)$ are functions of y to be determined. The substitution of this series into (8.56) shows that the $f_n(y)$ are solutions of

$$\frac{d^2 f_n(y)}{dy^2} - (\beta_n^2 - k_0^2) f_n(y) = 0 \tag{8.57}$$

Above the corrugations, i.e., in the region $c \le y \le b$, we must choose the f_n so that E_z will vanish on the perfectly conducting wall at $y = b$. Thus we require $df_n/dy = 0$ at $y = b$. Since solutions to (8.57) are $\sinh h_n y$ and $\cosh h_n y$, where $h_n = (\beta_n^2 - k_0^2)^{1/2}$, we choose

$$f_n(y) = a_n \cosh h_n(b - y)$$

where a_n is a constant. At $y = b$, this function has a zero derivative. For the fields H_x and E_z in the region above the corrugations, we now have

$$H_x = \sum_{n=-\infty}^{\infty} a_n \cosh h_n(b - y) e^{-j\beta_n z} \tag{8.58a}$$

$$E_z = -j\frac{Z_0}{k_0} \sum_{n=-\infty}^{\infty} a_n h_n \sinh h_n(b - y) e^{-j\beta_n z} \tag{8.58b}$$

upon using (8.55b).

As a next step we must obtain a suitable expansion for H_x in each corrugation, or slot. If $H_1(y, z)$ is the field in the slot extending from $z = 0$ to $z = s$ and for $0 \le y \le c$, then the field in the nth slot beginning at $z = nd$ will be $e^{-j\beta_n d}H_1(y, z - nd)$ according to Floquet's theorem. Therefore we need to concentrate on one slot only. We must determine H_x so that E_y will vanish at $z = 0$ and s and also so that E_z will vanish at $y = 0$. A suitable expansion to use is

$$H_x = \sum_{m=0}^{\infty} g_m(y)\cos\frac{m\pi z}{s}$$

since $d[\cos(m\pi z/s)]/dz$ vanishes at $z = 0$ and s. If this expansion is substituted into (8.56), we find that

$$\frac{d^2 g_m(y)}{dy^2} - \left[\left(\frac{m\pi}{s}\right)^2 - k_0^2\right] g_m(y) = 0$$

Normally, $s \ll \lambda_0$, so that $m\pi/s > k_0$ for $m \neq 0$. Thus appropriate solutions that have a zero derivative at $y = 0$ are

$$g_m(y) = b_m \cosh l_m y$$

where $l_m = [(m\pi/s)^2 - k_0^2]^{1/2}$ and b_m is a constant. For $m = 0$, the solution is $g_0(y) = b_0 \cos k_0 y$, and this part of the solution corresponds to a TEM standing wave in the slot. This mode has $E_y = 0$. In the first slot we can thus write

$$H_x = \sum_{m=0}^{\infty} b_m \cosh l_m y \cos\frac{m\pi z}{s} \tag{8.59a}$$

$$E_z = j\frac{Z_0}{k_0}\sum_{m=0}^{\infty} b_m l_m \sinh l_m y \cos\frac{m\pi z}{s} \tag{8.59b}$$

The final step in the analysis is to determine the expansion coefficients a_n and b_m by imposing boundary conditions at the plane $y = c$ separating the two regions. We require the tangential electric and magnetic fields to be continuous across the gap $y = c$, $0 \le z \le s$. In addition, we require the tangential electric field to vanish on the upper faces of the teeth, i.e., at $y = c$ for $s \le z \le d$, or $nd + s \le z \le (n + 1)d$ in general. Using (8.58) and (8.59), we see that the boundary conditions require

$$\sum_{n=-\infty}^{\infty} a_n e^{-j\beta_n z}\cosh h_n(b - c) = \sum_{m=0}^{\infty} b_m \cosh l_m c \cos\frac{m\pi z}{s} \qquad 0 \le z \le s \tag{8.60a}$$

$$\sum_{n=-\infty}^{\infty} a_n h_n e^{-j\beta_n z}\sinh h_n(b - c) = \begin{cases} -\displaystyle\sum_{m=0}^{\infty} b_m l_m \sinh l_m c \cos\frac{m\pi z}{s} & 0 \le z \le s \\ 0 & s \le z \le d \end{cases} \tag{8.60b}$$

If we multiply (8.60*b*) by $e^{j\beta z}$, we obtain

$$\sum_{n=-\infty}^{\infty} a_n h_n \sinh h_n(b-c) e^{-j2n\pi z/d} = \begin{cases} -\sum_{m=0}^{\infty} b_m e^{j\beta z} l_m \sinh l_m c \cos\dfrac{m\pi z}{s} & 0 \le z \le s \\ 0 & s \le z \le d \end{cases} \tag{8.60c}$$

Now the coefficients in a Fourier series are uniquely determined only if the function which the series is to represent is specified for the complete interval over which the series is orthogonal. The functions $e^{-j2n\pi z/d}$ are orthogonal over the range 0 to d, and thus, since the left-hand side of (8.60*c*) is specified for all z over one period, we can obtain unique expressions for the a_n in terms of the b_m from (8.60*c*). Note that this is not true for (8.60*a*), which holds only in the region $0 \le z \le s$. If we multiply (8.60*c*) on both sides by $e^{j2r\pi z/d}$ and integrate from 0 to d, we obtain (r is an integer)

$$\begin{aligned} da_r h_r \sinh h_r(b-c) &= -\sum_{m=0}^{\infty} b_m l_m \sinh l_m c \int_0^s e^{j(\beta+2r\pi/d)z} \cos\frac{m\pi z}{s}\, dz \\ &= \sum_{m=0}^{\infty} b_m l_m \sinh l_m c \frac{j(\beta+2\pi r/d)\left[(-1)^m e^{j\beta_r s}-1\right]}{(\beta+2\pi r/d)^2-(m\pi/s)^2} \end{aligned} \tag{8.61}$$

since $\int_0^d e^{(-j2\pi z/d)(n-r)}\, dz = 0$ for $n \ne r$ and equals d for $n = r$. The above represents an infinite set of equations, i.e., one for each value of r.

Although (8.60*a*) is not a unique equation for the a_n, it does specify the b_m uniquely in terms of the a_n. The a_n have already been expressed in terms of the b_m; so we may regard them as known. Multiplying (8.60*a*) by $\cos(r\pi z/s)$ and integrating from 0 to s give

$$\begin{aligned} \frac{s}{\epsilon_{0r}} b_r \cosh l_r c &= \sum_{n=-\infty}^{\infty} a_n \cosh h_n(b-c) \int_0^s e^{-j\beta_n z} \cos\frac{r\pi z}{s}\, dz \\ &= -j\sum_{n=-\infty}^{\infty} a_n \cosh h_n(b-c) \frac{\left(\beta+\dfrac{2n\pi}{d}\right)\left[1-(-1)^r e^{-j(\beta+2n\pi/d)s}\right]}{(\beta+2n\pi/d)^2-(r\pi/s)^2} \end{aligned} \tag{8.62}$$

where the Neumann factor $\epsilon_{0r} = 1$ for $r = 0$ and equals 2 for $r > 0$. This is also an infinite set of equations since $r = 0, 1, 2, \ldots, \infty$. Equations (8.61) and (8.62) constitute two linear systems of equations involving the a_n and b_m.

If the solutions for the a_n as given by (8.61) are substituted into (8.62), the result is a homogeneous set of equations for the b_m. For a nontrivial solution, the determinant of this homogeneous set of equations must vanish. Setting the determinant equal to zero yields the eigenvalue equation for β. However, the sets of equations are of infinite order, so that, in practice, an exact solution is not possible. Therefore we shall find only a first approximation to the exact eigenvalue equation.

If the slot spacing s is small compared with λ_0, it seems reasonable to expect that the field in the slot can be approximated by the TEM standing-wave field alone. Thus we shall take all b_m except b_0 equal to zero. If we lump all the constants in (8.61) together and replace r by n, the equation is of the form

$$a_n = \sum_{m=0}^{\infty} b_m R_{mn} \qquad n = 0, \pm 1, \ldots \tag{8.63a}$$

Likewise, (8.62) is an equation of the form

$$b_m = \sum_{n=-\infty}^{\infty} a_n T_{nm} \qquad m = 0, 1, 2, \ldots \tag{8.63b}$$

Replacing a_n by (8.63a) gives

$$b_m = \sum_{n=-\infty}^{\infty} \sum_{m=0}^{\infty} b_m R_{mn} T_{nm} \qquad m = 0, 1, 2, \ldots \tag{8.64}$$

The determinant of this infinite set of homogeneous equations, when equated to zero, gives the exact eigenvalue equation for β. When we take all b_m except b_0 equal to zero, we obtain instead

$$b_0 = \sum_{n=-\infty}^{\infty} b_0 R_{0n} T_{n0}$$

or

$$1 - \sum_{n=-\infty}^{\infty} R_{0n} T_{n0} = 0 \tag{8.65}$$

for a first approximation to the eigenvalue equation. Now R_{0n} are all the constants in (8.61), multiplying b_0 when the equation is solved for a_n and with r replaced by n. Likewise, T_{n0} is the constant relating b_0 to the a_n in (8.62). When these values for R_{0n} and T_{0n} are substituted into (8.65), we obtain

$$\frac{1}{k_0 d \tan k_0 c} = \frac{s}{d} \sum_{n=-\infty}^{\infty} \left[\frac{\sin(\beta_n s/2)}{\beta_n s/2} \right]^2 \frac{1}{h_n d \tanh h_n (b - c)} \tag{8.66}$$

For slow waves β is much larger than k_0, and hence h_n can be replaced by β_n in this equation with negligible error. In this case the right-hand side is not dependent on k_0. By evaluating the right-hand side for a range of assumed values for β, the corresponding value of k_0 may be found by

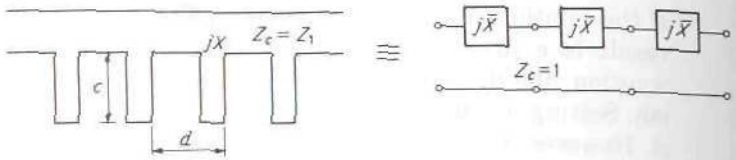


FIGURE 8.12
Simplified equivalent circuit for the vane structure of Fig. 8.11.

solving (8.66). The numerical work is straightforward. Typical results are given by Hutter.†

A reasonably accurate description of the dispersion curve relating β to k_0 may also be obtained from a simple transmission-line analysis. The region above the corrugations is essentially a parallel-plate transmission line (strip line) with a characteristic impedance $Z_1 = Z_0(b - c)$ per unit width. The slots are short-circuited transmission-line stubs connected in series with the main line at periodic intervals d. The stubs present a reactance

$$jX = jZ_0 s \tan k_0 c$$

to the main line. The equivalent circuit of the structure is therefore of the form shown in Fig. 8.12. This periodic circuit may be analyzed in the same way that the capacitively loaded transmission line was. It is readily found that the eigenvalue equation for β is

$$\begin{aligned} \cos \beta d &= \cos k_0 d - \frac{X}{2Z_1} \sin k_0 d \\ &= \cos k_0 d - \frac{s}{2(b-c)} \tan k_0 c \sin k_0 d \end{aligned} \tag{8.67}$$

This equation is quite accurate as long as $s \ll d$ and also much smaller than λ_0.

For frequencies such that $0 < k_0 c < \pi/2$, the loading is inductive, and for $\pi/2 < k_0 c < \pi$, it is capacitive, etc. A typical dispersion curve for the case $s = 2(b - c)$ and $d = 0.83c$ is given in Fig. 8.13. For these dimensions the phase velocity is reduced by a factor of about 3 only. Much greater reduction factors are obtained by making $s/(b - c)$ larger so as to increase the normalized characteristic impedance of the stubs. For comparison, the results from (8.66) for $d = 2s$ are plotted also (broken curve), verifying the accuracy of (8.67). The first cutoff occurs approximately when the stubs become resonant, i.e., for $k_0 c = \pi/2$. Increasing c will therefore also reduce

†R. G. E. Hutter, "Beam and Wave Electronics in Microwave Tubes," sec. 7.4, D. Van Nostrand Company, Inc., Princeton, N.J., 1960.

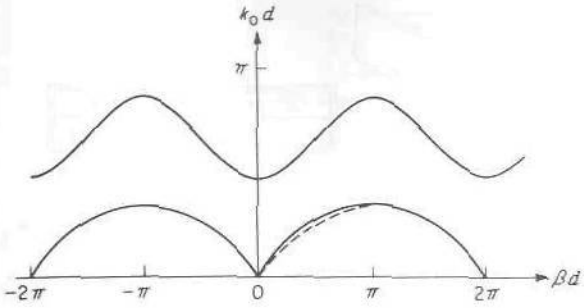


FIGURE 8.13
$k_0 d$-βd diagram for a stub-loaded transmission line.

the phase velocity in the first passband, since βd will equal π for a smaller value of k_0.

The foregoing analysis is typical for periodic structures that cannot be represented by simple transmission-line circuits. The essential steps to be followed are summarized:

1. Obtain suitable field expansions in each region of the periodic structure. This involves solution of the Helmholtz equation and the use of Floquet's theorem.
2. Impose appropriate boundary conditions on the fields at all common boundaries separating the different regions. In general, it will be found that both E and H modes may be required in order to satisfy the boundary conditions.
3. By Fourier analysis convert the boundary conditions into algebraic equations for the amplitude constants.
4. The system of algebraic equations can be written in the form of a homogeneous set of equations. Equating the determinant to zero gives the eigenvalue equation for β. Since the equations are usually infinite in order, some assumption must be made as regards the number of nonzero amplitude constants that will be chosen. Equating the higher-order amplitude constants to zero results in an approximate eigenvalue equation.

Periodic Structures for Millimeter-Wave Traveling-Wave Tubes

At millimeter wavelengths a helix has too small a diameter to be a useful slow-wave structure. Various forms of tape ladder lines, interdigital tape lines, and meander tape lines are preferred. Illustrations of these structures are given in Fig. 8.14. A discussion of these structures together with typical

†A. F. Harvey, Periodic and Guiding Structures at Microwave Frequencies, *IRE Trans.*, vol. MTT-8, pp. 30–61, January, 1960.

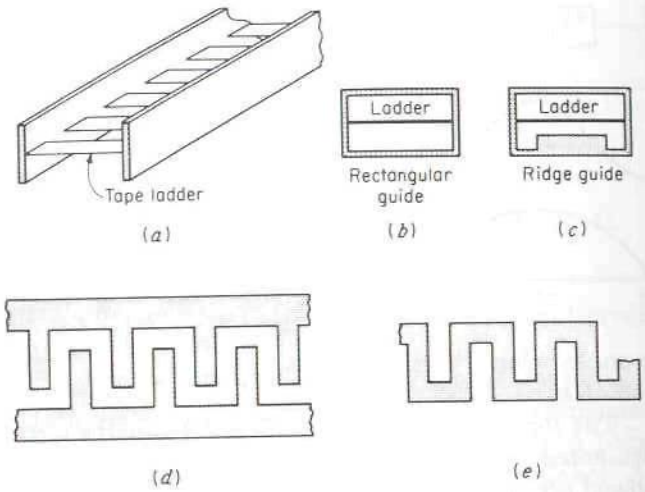


FIGURE 8.14
(*a–c*) Tape ladder lines; (*d*) interdigital tape line; (*e*) meander tape line.

k_0-β curves, is given in a paper by Harvey,† which also provides references to the original analysis of these structures.

The two structures shown in Figs. 8.14*d* and *e* are complementary, or dual, structures. That is, the meander line is obtained by interchanging the open region and the conducting region in the interdigital line. For a complementary structure of this type, we can show that the field is also a dual solution and hence both structures have exactly the same k_0-β dispersion curve. A detailed discussion of the dual properties is given below.

Let the interdigital line be located in the xy plane as in Fig. 8.15. Let us consider a mode of propagation having an electric field for which the transverse (x and y components) field $\mathbf{E}_t$ is an even function of z, that is, the same on the upper and lower sides of the structure. Since $\nabla \cdot \mathbf{E} = 0$, we now have $\partial E_z/\partial z = -\nabla_t \cdot \mathbf{E}_t$, and hence $\partial E_z/\partial z$ is an even function of z, and E_z must then be an odd function of z. From the curl equation $\nabla \times \mathbf{E} = -j\omega\mu_0\mathbf{H}$, we can readily conclude that $\mathbf{H}_t$ must be an odd function of z and H_z an even function of z. The field structure is illustrated in Fig. 8.15*a*. The field $\mathbf{E}_t$ will vanish on the conducting surface, and since $\mathbf{H}_t$ is an odd function of z, it must vanish on the open part of the xy plane. At the conductor surface, $|\mathbf{H}_t|$ will equal one-half the total current density in the line since the total change in $|\mathbf{H}_t|$ across the conductor must equal the total current density.

A dual field $\mathbf{E}', \mathbf{H}'$ given by

$$\mathbf{E}' = \pm Z_0\mathbf{H} \qquad \mathbf{H}' = \mp Y_0\mathbf{E}$$

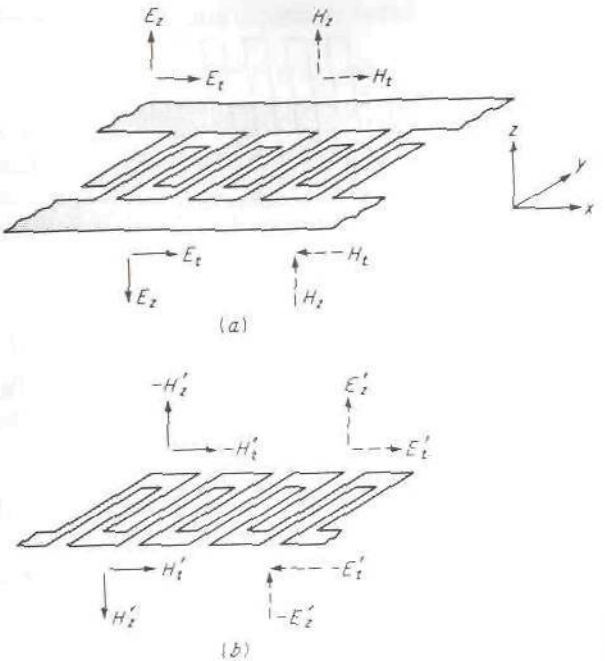


FIGURE 8.15
Illustration of dual properties of interdigital and meander lines.

is easily shown to satisfy Maxwell's equations

$$\nabla \times \mathbf{E}' = -j\omega\mu_0\mathbf{H}' \qquad \nabla \times \mathbf{H}' = j\omega\epsilon_0\mathbf{E}'$$

if the field $\mathbf{E}, \mathbf{H}$ does. The dual field is a solution to the meander-line problem (Fig. 8.15*b*), provided we choose the dual solution

$$\mathbf{E}' = Z_0\mathbf{H} \qquad \mathbf{H}' = -Y_0\mathbf{E} \tag{8.68a}$$

above the meander-line plane and the solution

$$\mathbf{E}' = -Z_0\mathbf{H} \qquad \mathbf{H}' = Y_0\mathbf{E} \tag{8.68b}$$

below the structure. In both regions the primed fields satisfy Maxwell's equations. The field $\mathbf{E}'_t$ will vanish on the conducting portions of the meander line since the field $\mathbf{H}_t$ was zero in the open regions of the interdigital line. Similarly, $\mathbf{H}'_t$ vanishes over the open regions of the meander-line plane since the field $\mathbf{E}_t$ was zero on the conducting surfaces of the interdigital line. All boundary conditions being satisfied, the solution is complete. It may now be concluded that both structures must have the same k_0-β dispersion curve. It should be noted, however, that duality applies only if *the two structures are exact complements*; i.e., *superimposing the two* structures must result in the whole xy plane being a single conducting sheet. The sides of the interdigital line must therefore extend to $y = \pm\infty$, and *the line must be infinitely long. However, in practice, the field is*

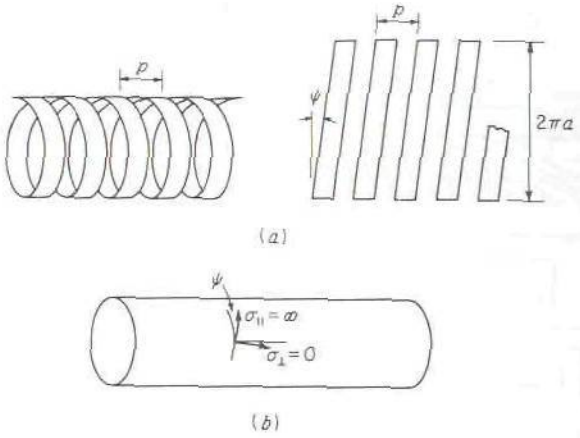


FIGURE 8.16
(*a*) A tape helix; (*b*) sheath helix.

confined to the vicinity near the cuts, so that the sides do not have to extend much beyond the toothed region before they can be terminated with negligible disturbance of the field. The duality principle used above is often referred to as Babinet's principle.

8.10 SHEATH HELIX

The sheath helix is an approximate model of a tape helix. The tape helix, illustrated in Fig. 8.16*a*, consists of a thin ribbon, or tape, wound into a helical structure. The pitch is denoted by p, and the pitch angle by ψ. If the spacing between turns and the ribbon width are made to approach zero, the resultant structure becomes electrically smooth. At the boundary surface $r = a$, the boundary conditions for the electric field may be approximated by the conditions that the conductivity in the direction along the tape (direction of current) is infinite, whereas that in the direction perpendicular to the tape is zero. The use of these boundary conditions permits a solution for the electromagnetic field guided by the helix to be obtained with relative ease. This anisotropic conducting cylinder model of a tape helix is called the sheath helix, illustrated in Fig. 8.16*b*. The field solution, derived below, will show that the sheath helix supports a slow wave with a phase velocity $v_p = c \sin \psi$. The wave may be considered to propagate along the helical conductor with a velocity c, and hence progresses along the axial direction z with a phase velocity $c \sin \psi$. The sheath-helix model is valid at low frequencies, where p is much smaller than λ_0. At higher frequencies a more realistic model must be used, and the existence of spatial harmonics then becomes apparent, as shown in Sec. 8.11.

The field solution for the helix consists of both E and H modes since these are coupled together by the boundary conditions at $r = a$. Along the direction of the tape, the tangential electric field must vanish, since the

conductivity in this direction is taken as infinite; thus

$$E_{\phi 1} \cos \psi + E_{z1} \sin \psi = E_{\phi 2} \cos \psi + E_{z2} \sin \psi = 0 \qquad (8.69a)$$

where the subscripts 1 and 2 refer to the field components in the two regions $r \leq a$ and $r \geq a$. The component of electric field on the cylindrical surface $r = a$ and perpendicular to the tape must be continuous since the conductivity is taken as zero in this direction. Hence

$$E_{z1} \cos \psi - E_{\phi 1} \sin \psi = E_{z2} \cos \psi - E_{\phi 2} \sin \psi \qquad (8.69b)$$

The component of **H** tangent to the tape must also be continuous since no current flows perpendicular to the tape; so a third boundary condition is

$$H_{z1} \sin \psi + H_{\phi 1} \cos \psi = H_{z2} \sin \psi + H_{\phi 2} \cos \psi \qquad (8.69c)$$

Expansions for the E and H modes in the two regions $r \lessgtr a$ may be obtained in terms of the axial field components E_z and H_z, as shown in Secs. 3.7 and 3.18. The axial fields $E_z = e_z(r, \phi)e^{-j\beta z}$, $H_z = h_z(r, \phi)e^{-j\beta z}$ are solutions of

$$\nabla_t^2 e_z + (k_0^2 - \beta^2)e_z = 0$$

Since we anticipate slow-wave solutions for which $\beta^2 > k_0^2$, the solutions involve Bessel functions with imaginary arguments, that is, $J_n(r\sqrt{k_0^2 - \beta^2})$ and $Y_n(r\sqrt{k_0^2 - \beta^2})$. In place of these functions, the modified Bessel functions $I_n(r\sqrt{\beta^2 - k_0^2})$, $K_n(r\sqrt{\beta^2 - k_0^2})$ are more convenient to use. These functions are related to the J_n and Y_n functions as follows:

$$I_n(x) = j^{-n} J_n(jx) \qquad (8.70a)$$

$$K_n(x) = \frac{\pi}{2} j^{n+1}[J_n(jx) + jY_n(jx)] \qquad (8.70b)$$

For small values of x, the K_n functions approach infinity in a logarithmic fashion, and hence only the I_n functions are used in the region $r < a$. For r large, the asymptotic forms

$$I_n(x) \sim \frac{1}{\sqrt{2\pi x}} e^x \qquad (8.71a)$$

$$K_n(x) \sim \sqrt{\frac{\pi}{2x}}\, e^{-x} \qquad (8.71b)$$

are valid. Since we require a field that decays for large r, only the functions K_n are employed in the region $r > a$. Suitable expansions for e_z and h_z in

the two regions are now seen to be

$$e_z = \begin{cases} \sum_{n=-\infty}^{\infty} a_n e^{-jn\phi} I_n(hr) & r \le a \\ \sum_{n=-\infty}^{\infty} b_n e^{-jn\phi} K_n(hr) & r \ge a \end{cases}$$

$$h_z = \begin{cases} \sum_{n=-\infty}^{\infty} c_n e^{-jn\phi} I_n(hr) & r \le a \\ \sum_{n=-\infty}^{\infty} d_n e^{-jn\phi} K_n(hr) & r \ge a \end{cases}$$

where $h = (\beta^2 - k_0^2)^{1/2}$ and a_n, b_n, c_n, and d_n are unknown amplitude constants.

For the sheath-helix model it is possible to find a solution for a field that satisfies the boundary conditions (8.69) for each integer n. We are primarily interested in the solution $n = 0$, which has circular symmetry. If we make use of (3.67), (3.68), and (3.72), together with the relations

$$\frac{dI_0(hr)}{d(hr)} = I_1(hr) \qquad \frac{dK_0(hr)}{d(hr)} = -K_1(hr)$$

we find that the field in the two regions can be expressed as follows:

For $r < a$,

$$E_z = a_0 I_0(hr) e^{-j\beta z} \quad E_r = \frac{j\beta}{h} a_0 I_1(hr) e^{-j\beta z} \quad E_\phi = -\frac{j\omega\mu_0}{h} c_0 I_1(hr) e^{-j\beta z}$$

$$H_z = c_0 I_0(hr) e^{-j\beta z} \quad H_r = \frac{j\beta}{h} c_0 I_1(hr) e^{-j\beta z} \quad H_\phi = \frac{j\omega\epsilon_0}{h} a_0 I_1(hr) e^{-j\beta z} \tag{8.72a}$$

For $r > a$,

$$E_z = b_0 K_0(hr) e^{-j\beta z} \quad E_r = -\frac{j\beta}{h} b_0 K_1(hr) e^{-j\beta z} \quad E_\phi = \frac{j\omega\mu_0}{h} d_0 K_1(hr) e^{-j\beta z}$$

$$H_z = d_0 K_0(hr) e^{-j\beta z} \quad H_r = -\frac{j\beta}{h} d_0 K_1(hr) e^{-j\beta z} \quad H_\phi = -\frac{j\omega\epsilon_0}{h} b_0 K_1(hr) e^{-j\beta z} \tag{8.72b}$$

for the $n = 0$ mode.

If the above expressions for the fields are substituted into the boundary conditions (8.69), the result is four homogeneous equations for the constants a_0, b_0, c_0, and d_0. A nontrivial solution exists only if the determinant vanishes. Equating the determinant to zero results in the

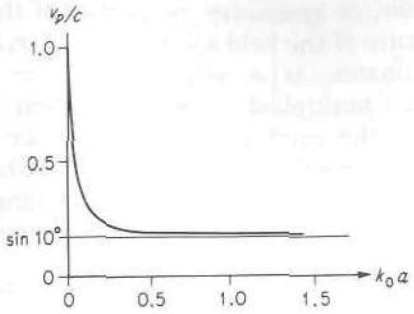


FIGURE 8.17
Phase-velocity reduction factor for a sheath helix with pitch angle $\psi = 10°$.

eigenvalue equation for β, which is

$$\frac{K_1(ha)I_1(ha)}{K_0(ha)I_0(ha)} = \frac{(ha)^2\tan^2\psi}{(k_0a)^2} \tag{8.73}$$

For ha greater than 10, the ratio K_1I_1/K_0I_0 rapidly approaches unity. In this region (8.73) gives $h = k_0\cot\psi$, from which we obtain

$$\beta = \left(k_0^2 + h^2\right)^{1/2} = k_0\csc\psi \tag{8.74}$$

The resultant phase velocity v_p is

$$v_p = \frac{\omega}{\beta} = \frac{k_0}{\beta}c = c\sin\psi \tag{8.75}$$

and is reduced by the factor $\sin\psi$. A plot of v_p/c as a function of k_0a is given in Fig. 8.17 as determined by the solution of (8.73) with ψ equal to 10°. For k_0a greater than 0.25, the phase velocity is well approximated by (8.75). In the frequency range where $v_p = c\sin\psi$, the group velocity v_g is also equal to $c\sin\psi$, and there is no frequency dispersion.

*8.11 SOME GENERAL PROPERTIES OF A HELIX

The tape helix consists of a thin ribbon of metal wound into a helical structure, as shown in Fig. 8.16*a*. A helix may also be constructed by the use of a round wire. The parameters describing the helix are the pitch p, or turn-to-turn spacing, the diameter $2a$, and the pitch angle ψ. These parameters are given in Fig. 8.16*a*, which shows a developed view of a tape helix.

The helix is a periodic structure with respect to translation by a distance p along the axis and also with respect to rotation through an arbitrary angle θ, followed by a translation $p\theta/2\pi$ along the axial direction. In other words, an infinitely long helix translated along the z axis by a distance p or rotated by an angle θ and then shifted by a distance $p\theta/2\pi$ along z will coincide with itself.

The above periodic, or symmetry, properties of the helix place certain restrictions on the nature of the field solutions. If $\mathbf{E}_1(r, \phi, z)$, where r and ϕ are cylindrical coordinates, is a solution for the electric field, then $\mathbf{E}_1(r, \phi + \theta, z + p\theta/2\pi)$ multiplied by a propagation factor $e^{-j\beta(p\theta/2\pi)}$ is another solution, since the point $r, \phi + \theta, z + p\theta/2\pi$ is indistinguishable from the point r, ϕ, z. The solution $\mathbf{E}_1(r, \phi, z)$ must be periodic in ϕ and, apart from a propagation factor $e^{-j\beta z}$, must also be periodic in z with period p. Hence $\mathbf{E}_1$ may be expanded in the double Fourier series

$$\mathbf{E}_1(r, \phi, z) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \mathbf{E}_{1,mn}(r) e^{-jm\phi - j2n\pi z/p} e^{-j\beta z} \tag{8.76}$$

where $\mathbf{E}_{1,mn}(r)$ are vector functions of r corresponding to the usual amplitude constants in a Fourier series. The relationship between translation and rotation noted above requires that $e^{j\beta z}\mathbf{E}_1(r, \phi, z)$ does not change when ϕ, z are replaced by $\phi + \theta, z + p\theta/2\pi$. Thus, in (8.76), we require

$$\begin{aligned} e^{-jm(\phi+\theta) - j2n\pi(z+p\theta/2\pi)/p} &= e^{-jm(\phi+\theta) - jn\theta - j2n\pi z/p} \\ &= e^{-jm\phi - j2n\pi z/p} \end{aligned}$$

This condition will hold only if $m = -n$. Consequently, for a helix, the double Fourier series expansion for the electric field reduces to a single series of the form

$$\mathbf{E}_1(r, \phi, z) = \sum_{n=-\infty}^{\infty} \mathbf{E}_{1,n}(r) e^{-jn(2\pi z/p - \phi)} e^{-j\beta z} \tag{8.77}$$

The solution for a helix proceeds by expanding the field in the two regions $r < a$ and $r > a$ into an infinite series of E and H modes expressed in cylindrical coordinates. The boundary conditions at $r = a$ will couple the E and H modes together, so that pure E or H modes cannot exist independently. For the nth term in (8.77), the radial dependence in the region $r > a$ will accord with the modified Bessel function of the second kind, that is, $K_n(h_n r)$, of order n and with an argument

$$h_n r = \left[\left(\beta + \frac{2n\pi}{p} \right)^2 - k_0^2 \right]^{1/2} r$$

The K_n functions are asymptotic to $(\pi/2h_n r)^{1/2} e^{-h_n r}$ for r large; so the field will decay exponentially as long as all h_n are real, i.e., for all $(\beta + 2n\pi/p)^2$ greater than k_0^2. When the field decays exponentially it corresponds to a surface-wave mode guided by the helix.

At a given frequency only certain discrete values of β, say β_m, are possible solutions. For each value of β_m, corresponding to a particular mode of propagation, the field is given by a Fourier series of the form

$$\mathbf{E}_m(r, \phi, z) = \sum_{n=-\infty}^{\infty} \mathbf{E}_{m,n}(r) e^{-jn(2\pi z/p - \phi) - j\beta_m z} \tag{8.78}$$

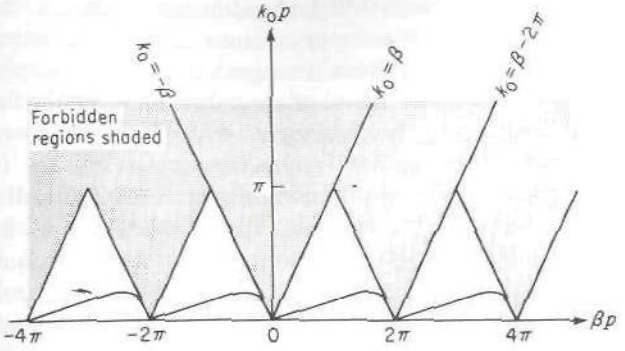


FIGURE 8.18
Illustration of allowed and forbidden regions in the k_0-β diagram for a helix.

Each term in this expansion is called a spatial harmonic, and has a propagation phase constant $\beta_m + 2n\pi/p$. On a k_0p-βp diagram the region above the lines $k_0 = \pm\beta$ is a forbidden region, as shown in Fig. 8.18, since it corresponds to a situation where $h_0 = (\beta^2 - k_0^2)^{1/2}$ is imaginary and the $n = 0$ spatial harmonic does not decay in the radial direction. Since we also require $|\beta + 2n\pi/p|$ to be greater than k_0, all possible allowed values of β, corresponding to bound surface-wave modes, are further restricted to lie in the unshaded triangular regions in Fig. 8.18. The boundaries of these regions are marked by the lines

$$k_0 = \pm\left(\beta \pm \frac{2n\pi}{p}\right)$$

where n is an integer. In the forbidden regions the propagation constant turns out to be complex rather than pure real, a feature which is different from that of a normal cutoff mode.

A first approximation to the solution for a tape helix is obtained by assuming that the current is directed along the direction of the tape only, is uniform across the width of the tape, and has a propagation factor $e^{-j\beta z}$. A typical k_0-β curve obtained on this basis and with $\psi = 10°$ is shown in Fig. 8.18. For further results the paper by Sensiper or the book by Watkins may be consulted.†

8.12 INTRODUCTION TO MICROWAVE FILTERS

The ideal filter network is a network that provides perfect transmission for all frequencies in certain passband regions and infinite attenuation in the

†S. Sensiper, Electromagnetic Wave Propagation on Helical Structures, *Proc. IRE*, vol. 43, pp. 149–161, February, 1955.

D. A. Watkins, "Topics in Electromagnetic Theory," John Wiley & Sons, Inc., New York, 1958.

stopband regions. Such ideal characteristics cannot be obtained, and the goal of filter design is to approximate the ideal requirements to within an acceptable tolerance. Filters are used in all frequency ranges to provide as nearly perfect transmission as possible for signals falling within desired passband frequency ranges, together with rejection of those signals and noise outside the desired frequency bands. Filters fall into three main categories, namely, (1) low-pass filters that transmit all signals between zero frequency and some upper limit ω_c and attenuate all frequencies above the cutoff value ω_c, (2) high-pass filters that pass all frequencies above a lower cutoff value ω_c and reject all frequencies below ω_c, and (3) bandpass filters that pass all frequencies in a range ω_1 to ω_2 and reject frequencies outside this range. The complement to the bandpass filter, i.e., the band-rejection filter, which attenuates frequencies in the range ω_1 to ω_2, is also of interest in certain applications.

At low frequencies the "building blocks" for filters are ideal inductors and capacitors. These elements have very simple frequency characteristics, and a very general and complete synthesis procedure has been developed for the design of filters utilizing them. It is possible to synthesize directly filters with a wide variety of prescribed frequency characteristics. The filter design problem at microwave frequencies where distributed parameter elements must be used is much more complicated, and no complete theory or synthesis procedure exists. The complex frequency behavior of microwave circuit elements makes it virtually impossible to develop a general and complete synthesis procedure. However, in spite of these added complications at microwave frequencies, a number of useful techniques have been developed for the design of microwave filters. The case of narrowband filters is particularly straightforward since many microwave elements will have frequency characteristics essentially like those of an ideal inductive or capacitive reactance over a limited frequency range. In this case a low-frequency prototype filter may be used as a model. The microwave filter is realized by replacing all inductors and capacitors by suitable microwave circuit elements that have similar frequency characteristics over the frequency range of interest. For this reason a good deal of the effort in microwave filter design has been based directly on the application of low-frequency filter-synthesis techniques.

There are essentially two low-frequency filter-synthesis techniques in common use. These are referred to as the image-parameter method (and variations thereof, such as the constant-k and m-derived filters) and the insertion-loss method. The image-parameter method provides a filter design having the required passband and stopband characteristics, but does not specify the exact frequency characteristics over each region. The insertion-loss method begins with a complete specification of a physically realizable frequency characteristic, and from this a suitable filter network is synthesized. The image-parameter method suffers from the shortcomings that a good deal of cut-and-try procedures must often be resorted to in order to

obtain an acceptable overall frequency characteristic. For this reason the insertion-loss method is preferable and is the only method considered in detail in the following sections. The image-parameter method is only briefly outlined, in order to show its relationship to the properties of periodic networks as already discussed.

The labor involved in filter synthesis is largely obviated by the use of certain frequency transformations and element normalizations. These enable high-pass and band-pass filters operating over arbitrary frequency bands and between arbitrary resistive load terminations to be obtained from a basic low-pass filter design. The characteristics of any filter will, of course, be modified by the losses that are present in all physical network elements. To incorporate the effect of lossy elements into the synthesis procedure makes the synthesis theory a great deal more involved; so this is usually not done. At microwave frequencies losses can be kept reasonably small, to the extent that most filter designs based on the use of lossless elements do perform satisfactorily.

The aim of the following sections is to present the essential features of low-frequency filter synthesis, frequency transformations, normalized filter design, and the applications of these techniques to microwave filter design. A number of typical microwave filters are also discussed. An extensive account of all aspects of microwave filter design is beyond the scope of this text. However, a number of selected references are given where further details may be found.

8.13 IMAGE-PARAMETER METHOD OF FILTER DESIGN

Filters designed by the image-parameter method have many features in common with those of periodic structures. As noted in the previous sections, a cascade connection of lossless two-port networks behaves similar to a transmission line. For unsymmetrical networks two characteristic impedances $\overline{Z}_B^{\pm} = \pm\overline{Z} + \zeta$ occur, and each section has a propagation factor $e^{\pm\gamma d}$. A periodic structure of this form has passband and stopband characteristics and is therefore a bandpass filter. However, the proper load termination to prevent reflections is $\overline{Z}_B$ and is complex when $\zeta \neq 0$. Usually, a filter must operate between resistive load terminations, and it would not be possible to have matched input and output terminations in this case unless ζ were zero, i.e., unless symmetrical networks were used or unless matching sections were used at the input and output. For this reason the image-parameter method of filter design is based on considerations somewhat different from those which have been discussed for periodic structures.

Consider a single two-port network with parameters $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, and $\mathscr{D}$. Let the output be terminated in a load Z_{i2}, and let the input be terminated in a load Z_{i1}, as in Fig. 8.19. For particular values of Z_{i1} and Z_{i2}, known as

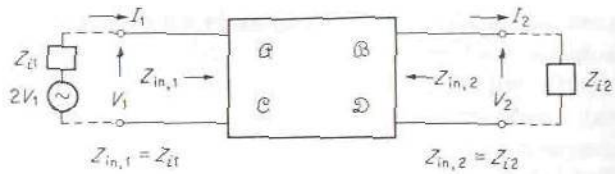


FIGURE 8.19
Image parameters for a two-port network.

the image impedances, the input impedance at port 1 equals Z_{i1} and that at port 2 equals Z_{i2}. These impedances then provide matched terminations for the two-port network, and if they are real, they also provide a maximum power transfer when the generator has an internal impedance equal to the image impedance. The governing equations for the two-port network are

$$V_1 = \mathscr{A}V_2 + \mathscr{B}I_2 \qquad I_1 = \mathscr{C}V_2 + \mathscr{D}I_2$$

and hence

$$\frac{V_1}{I_1} = Z_{\text{in},1} = \frac{\mathscr{A}V_2 + \mathscr{B}I_2}{\mathscr{C}V_2 + \mathscr{D}I_2} = \frac{\mathscr{A}Z_{i2} + \mathscr{B}}{\mathscr{C}Z_{i2} + \mathscr{D}} \tag{8.79}$$

If we solve for V_2 and I_2 in terms of V_1 and I_1, we obtain

$$V_2 = \mathscr{D}V_1 - \mathscr{B}I_1 \qquad I_2 = -\mathscr{C}V_1 + \mathscr{A}I_1$$

We thus have

$$-\frac{V_2}{I_2} = Z_{\text{in},2} = -\frac{\mathscr{D}V_1 - \mathscr{B}I_1}{-\mathscr{C}V_1 + \mathscr{A}I_1} = \frac{\mathscr{D}Z_{i1} + \mathscr{B}}{\mathscr{C}Z_{i1} + \mathscr{A}} \tag{8.80}$$

The requirement that $Z_{i1} = Z_{\text{in},1}$ and $Z_{i2} = Z_{\text{in},2}$ gives

$$Z_{i1}(\mathscr{C}Z_{i2} + \mathscr{D}) = \mathscr{A}Z_{i2} + \mathscr{B} \qquad Z_{i2}(\mathscr{C}Z_{i1} + \mathscr{A}) = \mathscr{D}Z_{i1} + \mathscr{B}$$

A simultaneous solution of these equations gives

$$Z_{i1} = \sqrt{\frac{\mathscr{A}\mathscr{B}}{\mathscr{C}\mathscr{D}}} \tag{8.81a}$$

$$Z_{i2} = \sqrt{\frac{\mathscr{D}\mathscr{B}}{\mathscr{A}\mathscr{C}}} \tag{8.81b}$$

Also we find that $Z_{i2} = (\mathscr{D}/\mathscr{A})Z_{i1}$.

If a generator with internal impedance Z_{i1} is connected at port 1 and the output port 2 is terminated in a load Z_{i2}, the voltage and current transfer ratios are readily found from the relations

$$V_2 = \mathscr{D}V_1 - \mathscr{B}I_1 = \left(\mathscr{D} - \frac{\mathscr{B}}{Z_{i1}}\right)V_1$$

$$I_2 = -\mathscr{C}V_1 + \mathscr{A}I_1 = (-\mathscr{C}Z_{i1} + \mathscr{A})I_1$$

where V_1 is the voltage across the network terminals at port 1 (the generator voltage is $2V_1$). Thus we find that

$$\frac{V_2}{V_1} = \sqrt{\frac{\mathscr{D}}{\mathscr{A}}}\,(\sqrt{\mathscr{A}\mathscr{D}} - \sqrt{\mathscr{B}\mathscr{C}}) \tag{8.82a}$$

$$\frac{I_2}{I_1} = \sqrt{\frac{\mathscr{A}}{\mathscr{D}}}\,(\sqrt{\mathscr{A}\mathscr{D}} - \sqrt{\mathscr{B}\mathscr{C}}) \tag{8.82b}$$

In a similar manner the transfer constants from port 2 to port 1 are found to be [or from (8.82)]

$$\frac{V_1}{V_2} = \sqrt{\frac{\mathscr{A}}{\mathscr{D}}}\,(\sqrt{\mathscr{A}\mathscr{D}} + \sqrt{\mathscr{B}\mathscr{C}}) \tag{8.83a}$$

$$\frac{I_1}{I_2} = \sqrt{\frac{\mathscr{D}}{\mathscr{A}}}\,(\sqrt{\mathscr{A}\mathscr{D}} + \sqrt{\mathscr{B}\mathscr{C}}) \tag{8.83b}$$

The image propagation factor $e^{-\gamma}$ is defined as

$$e^{-\gamma} = \sqrt{\mathscr{A}\mathscr{D}} - \sqrt{\mathscr{B}\mathscr{C}} \tag{8.84a}$$

whence it is found that

$$e^{\gamma} = \sqrt{\mathscr{A}\mathscr{D}} + \sqrt{\mathscr{B}\mathscr{C}} \tag{8.84b}$$

and

$$\cosh\gamma = \sqrt{\mathscr{A}\mathscr{D}} \tag{8.84c}$$

$$\sinh\gamma = \sqrt{\mathscr{B}\mathscr{C}} \tag{8.84d}$$

The factor $(\sqrt{\mathscr{A}/\mathscr{D}})^2$ is interpreted as an impedance transformation ratio and may be viewed as an ideal transformer of turns ratio $\sqrt{\mathscr{A}/\mathscr{D}}$.

For a lossless network, $\mathscr{A}$ and $\mathscr{D}$ are real and $\mathscr{B}$ and $\mathscr{C}$ are imaginary. In the passband of a filter, γ is pure imaginary and equal to $j\beta$, and this occurs for $|\mathscr{A}\mathscr{D}| < 1$, as (8.84c) shows. Also, in the passband, the image impedances are pure real, whereas in a stopband they are pure imaginary, as the following considerations show. In a passband, $\mathscr{B}$ and $\mathscr{C}$ must be of the same sign, so that $\mathscr{B}\mathscr{C} = j|\mathscr{B}|j|\mathscr{C}| = -|\mathscr{B}\mathscr{C}|$ will make sinh γ in (8.84d) pure imaginary; that is, $\gamma = j\beta$. Thus, in (8.81), the quantity under the square root will be real and positive since $\mathscr{A}\mathscr{D}$ must be positive to give a real solution for cosh γ. Hence the image impedances are real in a passband.

If N two-port networks are connected in cascade and these have propagation constants γ_n, $n = 1, 2, \ldots, N$, and voltage transformation ratios

$$T_1 = \sqrt{\frac{\mathscr{D}_1}{\mathscr{A}_1}} \qquad T_2, \ldots, T_n, \ldots, T_N$$

and the output section is terminated in an impedance equal to its output image impedance, the overall voltage transfer ratio is

$$\frac{V_N}{V_1} = \frac{V_N}{V_{N-1}} \cdots \frac{V_2}{V_1} = T_1 T_2 \cdots T_N e^{-\gamma_1 - \gamma_2 \cdots - \gamma_N} = \prod_{n=1}^{N} T_n e^{-\gamma_n} \quad (8.85)$$

provided also that the output image impedance of any one section is equal to the input image impedance of the adjacent section. With this filter network terminated in a load impedance Z_{iN} equal to the image impedance of the output section, and with the generator at the input having an internal impedance Z_{i1}, the overall network is matched for maximum power transfer. The filter operates between impedance levels of Z_{iN} and Z_{i1}, which provide an overall impedance-ratio change of amount

$$\frac{Z_{iN}}{Z_{i1}} = \frac{Z_{iN}}{Z_{iN-1}} \cdots \frac{Z_{i2}}{Z_{i1}} = \prod_{n=1}^{N} T_n^2 \quad (8.86)$$

If symmetrical two ports are used, $\mathscr{A} = \mathscr{D}$ and $Z_{i1} = Z_{i2}$, and both are equal to the Bloch-wave characteristic impedance Z_B^+. For a symmetrical network no transformation or change in impedance level is obtained. The filter consisting of N symmetrical sections terminated in load impedances equal to the image impedance Z_i behaves exactly like an infinite periodic structure, with its characteristic passband and stopband features.

In the image-parameter method of filter design, the two-port parameters $\mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D}$ are chosen to provide for the required passbands and stopbands. In addition, the image parameters are also chosen equal to the terminating impedances at the center of the passband. The shortcomings of the filter are now apparent, namely, the image impedances are functions of frequency and do not remain equal to the terminating impedances over the whole desired passband. This results in some loss in transmission (loss due to mismatch) within the passband, an amount that cannot be prescribed or determined before the filter has been designed. In addition, there is no means available for controlling the rate at which the attenuation increases with frequency beyond the edges of the passband, apart from increasing the number of filter sections. Nevertheless, many useful microwave filters have been designed on this basis.†

†S. B. Cohn, chaps. 26 and 27 in Radio Research Laboratory Staff, "Very High Frequency Techniques," vol. 2 McGraw-Hill Book Company, New York, 1947.

For a discussion of image-parameter methods at low frequencies, see E. A. Guillemin, "Communication Networks," vol. 2, John Wiley & Sons, Inc., New York, 1935.

8.14 FILTER DESIGN BY INSERTION-LOSS METHOD

The power loss ratio of a network was defined in Sec. 5.14 as the available, or incident, power divided by the actual power delivered to the load; thus

$$P_{LR} = \frac{1}{1 - \Gamma\Gamma^*} = \frac{1}{1 - \rho^2} \tag{8.87}$$

where Γ is the input reflection coefficient for a lossless network terminated in a *resistive load impedance* $Z_L = R_L$. *The insertion loss, measured in* decibels, is

$$L = 10 \log P_{LR} \tag{8.88}$$

when the terminating resistive load impedance equals the internal impedance of the generator at the input end. In general, the insertion loss is defined as the ratio of the power delivered to the load when connected directly to the generator to the power delivered when the filter is inserted.

The insertion-loss method of filter design begins by specifying the power loss ratio P_{LR} or the magnitude of the reflection coefficient $|\Gamma| = \rho$ as a function of ω. A network that will give the desired power loss ratio is then synthesized. This procedure is seen to be essentially the same as was followed in the synthesis of quarter-wave transformers in Secs. 5.12 and 5.13. Indeed, the multisection quarter-wave transformer may be considered a particular type of bandpass filter. It must be kept in mind, however, that a completely arbitrary $\Gamma(\omega)$ as a function of ω cannot be chosen since it may not correspond to a physical network. The restrictions to be imposed on Γ are known as the conditions for physical realizability, and some of these are discussed below.

For a passive network it is clear that the reflected power cannot exceed the incident power, and hence one restriction on $\Gamma(\omega)$ is

$$|\Gamma(\omega)| \leq 1 \tag{8.89}$$

If the normalized input impedance of the network is

$$\bar{Z}(\omega) = \bar{R}(\omega) + j\bar{X}(\omega)$$

we have

$$\Gamma(\omega) = \frac{\bar{Z}_{in} - 1}{\bar{Z}_{in} + 1} = \frac{\bar{R}(\omega) - 1 + j\bar{X}(\omega)}{\bar{R}(\omega) + 1 + j\bar{X}(\omega)}$$

As shown in Sec. 4.4, $\bar{R}$ is an even function of ω and $\bar{X}$ is an odd function of ω. Hence

$$\Gamma(-\omega) = \frac{\bar{R}(\omega) - 1 - j\bar{X}(\omega)}{\bar{R}(\omega) + 1 - j\bar{X}(\omega)} = \Gamma^*(\omega)$$

and thus

$$|\Gamma(\omega)|^2 = \rho^2(\omega) = \Gamma\Gamma^* = \Gamma(\omega)\Gamma(-\omega) \tag{8.90}$$

It is apparent from this relation that $\rho^2(\omega) = \rho^2(-\omega)$ is an even function of ω and must therefore contain only even powers of ω. Now any low-frequency impedance function (impedance of a network made up of resistors, capacitors, and inductors) can be expressed as the ratio of two polynomials in ω. Consequently, Γ can also be expressed as the ratio of two polynomials. It follows that $\rho^2(\omega)$ can then be expressed in the form

$$\rho^2(\omega) = \frac{M(\omega^2)}{M(\omega^2) + N(\omega^2)} = \frac{(\overline{R} - 1)^2 + \overline{X}^2}{(\overline{R} + 1)^2 + \overline{X}^2} \tag{8.91}$$

where M and N are real and nonnegative polynomials in ω^2. The power loss ratio can now be expressed as

$$P_{\text{LR}} = 1 + \frac{M(\omega^2)}{N(\omega^2)} = 1 + \frac{\left[\overline{R}(\omega) - 1\right]^2 + \left[\overline{X}(\omega)\right]^2}{4\overline{R}(\omega)} \tag{8.92}$$

The last result in (8.92) shows that $N(\omega^2)$ must be an even polynomial in ω since it equals $4\overline{R}(\omega)$. Hence we write $N(\omega^2) = Q^2(\omega)$, which is clearly an even polynomial in ω. If we denote $M(\omega^2)$ by the even polynomial $P(\omega^2)$ instead, we have

$$P_{\text{LR}} = 1 + \frac{P(\omega^2)}{Q^2(\omega)} \tag{8.93}$$

The conditions specified on P_{LR} up to this point are necessary conditions in order that the network may be physically realizable. It may be shown that the condition that the power loss ratio P_{LR} be expressible in the form (8.93) is also a sufficient condition for the network to be realizable.† In succeeding sections we consider suitable forms for the polynomials P and Q and the types of networks required to yield the corresponding power loss ratio.

8.15 SPECIFICATION OF POWER LOSS RATIO

There are virtually an unlimited number of different forms that could be specified for the power loss ratio and be realized as a physical network. However, many of these networks could be anticipated to be very complex and hence of little practical utility. The power loss ratios that have been found most useful for microwave filter design are those that give a maxi-

†G. L. Ragan (ed.), "Microwave Transmission Circuits," sec. 9.13, McGraw-Hill Book Company, New York, 1948.

mally flat passband response and those that give an equal-ripple, or Chebyshev, response in the passband. Such passband-response characteristics correspond to those of the binomial and Chebyshev multisection quarter-wave transformers discussed in Chap. 5. The maximally flat filter (commonly called a Butterworth filter) and the Chebyshev filter are described below for the low-pass case only. In a following section it is shown that high-pass and bandpass filter characteristics may be obtained from the low-pass filter response by suitable frequency transformations, or mappings.

Maximally Flat Filter Characteristic

The power loss ratio for a maximally flat low-pass filter is obtained by choosing the polynomial Q equal to unity and choosing $P(\omega^2)$ equal to $k^2(\omega/\omega_c)^{2N}$. Hence we have

$$P_{LR} = 1 + k^2\left(\frac{\omega}{\omega_c}\right)^{2N} \tag{8.94}$$

The passband is the region from $\omega = 0$ to the cutoff value ω_c. The maximum value of P_{LR} in the passband is $1 + k^2$, and for this reason k^2 is called the passband tolerance. For $\omega > \omega_c$ the power loss ratio increases indefinitely at a rate dependent on the exponent $2N$, which in turn is related to the number of filter sections employed. A typical filter characteristic is illustrated in Fig. 8.20 for $N = 2$.

Chebyshev Filter

The power loss ratio for the equal-ripple, or Chebyshev, filter is chosen as

$$P_{LR} = 1 + k^2 T_N^2\left(\frac{\omega}{\omega_c}\right) \tag{8.95}$$

where $T_N(\omega/\omega_c)$ is the Chebyshev polynomial of degree N discussed in Sec.

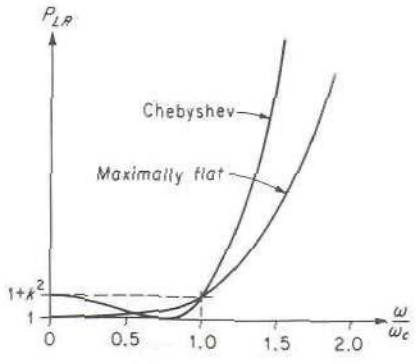


FIGURE 8.20
Low-pass-filter response for maximally flat and Chebyshev filters for $N = 2$.

5.13. Recall that

$$T_N\left(\frac{\omega}{\omega_c}\right) = \cos\left(N \cos^{-1} \frac{\omega}{\omega_c}\right)$$

and thus $T_N(\omega/\omega_c)$ oscillates between ± 1 for $|\omega/\omega_c| \leq 1$ and increases monotonically for ω/ω_c greater than unity. The power loss ratio will oscillate between 1 and $1 + k^2$ in the passband, equals $1 + k^2$ at the cutoff frequency, and will increase monotonically for $\omega > \omega_c$. A typical response curve is shown in Fig. 8.20 for $N = 2$. One particularly striking feature of the Chebyshev response curve compared with the maximally flat curve is its much greater rate of rise beyond the cutoff point. This means that the corresponding filter has a much sharper cutoff region separating the passband and stopband, which is usually a desired characteristic.

For ω/ω_c large, the power loss ratio for the Chebyshev filter approaches

$$P_{\text{LR}} \approx \frac{k^2}{4}\left(\frac{2\omega}{\omega_c}\right)^{2N} \tag{8.96}$$

Compared with the maximally flat response characteristic this is larger by a factor 2^{2N-2}. In fact, no other polynomial $P(\omega^2)$ yielding a passband tolerance of k^2 can yield a rate of increase of P_{LR} greater than that given by (8.96). Conversely, for a specified rate of increase in the power loss ratio beyond the cutoff frequency, the Chebyshev polynomial gives the smallest passband tolerance. In this sense the Chebyshev filter represents an optimum design. The proof is similar to that used in Sec. 5.14 to prove the optimum properties of the Chebyshev quarter-wave transformer.

When the power loss ratio is equal to $1 + k^2$, the magnitude of the reflection coefficient at the input is

$$\rho = \frac{k}{(1 + k^2)^{1/2}}$$

The input voltage standing-wave ratio is given by

$$S = \frac{1 + \rho}{1 - \rho} = \frac{(1 + k^2)^{1/2} + k}{(1 + k^2)^{1/2} - k}$$

If P is chosen as unity and Q is set equal to $k^2 T_N^2(\omega/\omega_c)$, a high-pass filter having Chebyshev behavior in the attenuation or stopband is obtained. It is also possible to choose P and Q so as to give Chebyshev behavior in both the passband and the stopband. The required network is, however, usually too complex to be realized in a satisfactory manner with microwave circuit elements.

8.16 SOME LOW-PASS-FILTER DESIGNS

The maximally flat and Chebyshev low-pass-filter power loss ratios discussed in the preceding section can be realized by means of a ladder network of capacitors and inductors in the form illustrated in Fig. 8.21. The load impedance is chosen equal to 1 Ω, and the generator impedance as R. The circuit in Fig. 8.21*b* is the dual of the circuit in Fig. 8.21*a*. Both circuits can be designed to give the same power loss ratio. For maximally flat or Chebyshev response in the passband, the ladder network is symmetric for an odd number of elements. This is also true for N even, in the case of the maximally flat filter. The element values are denoted by g_k, and are the same in both circuits. However, the required generator impedance R' in the network of Fig. 8.21*b* is equal to $1/R$. For N odd both R and R' equal unity.

If we let Z_{in} be the input impedance at the plane aa in Fig. 8.21*a*, the reflection coefficient will be

$$\Gamma = \frac{Z_{in} - R}{Z_{in} + R}$$

In terms of Z_{in} and R, the power loss ratio is readily computed to be

$$P_{LR} = 1 + \frac{|Z_{in} - R|^2}{2R(Z_{in} + Z_{in}^*)} \tag{8.97}$$

At $\omega = 0$, all capacitors appear as infinite impedances and all inductors as zero impedances, and hence $Z_{in} = 1$. For a maximally flat filter or a

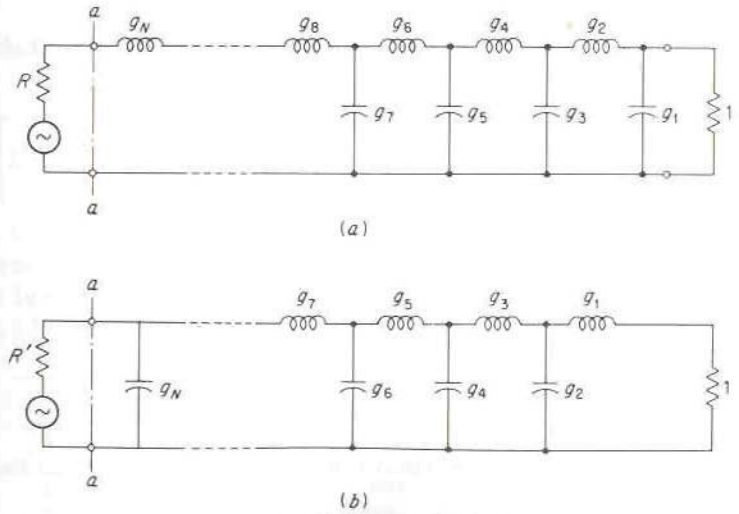


FIGURE 8.21
Low-pass ladder-prototype-filter networks.

Chebyshev filter with N odd, we must have unity power loss ratio at $\omega = 0$. This requires that we choose the generator impedance equal to unity. For a Chebyshev filter with N even, we have $P_{LR} = 1 + k^2$ at $\omega = 0$, and hence

$$1 + k^2 = 1 + \frac{(1 - R)^2}{4R}$$

or

$$R = 2k^2 + 1 - \sqrt{4k^2(1 + k^2)} = \frac{1}{R'} \tag{8.98}$$

The required values of the elements g_k in the ladder network can be obtained very readily by solving for Z_{in} and equating the power loss ratio as given by (8.97) to the desired power loss ratio for N up to 3 or 4. As an example, consider the case of $N = 2$ for the circuit in Fig. 8.21*a*. We readily find that

$$Z_{in} = j\omega L + \frac{1}{1 + j\omega C}$$

where $C = g_1$ and $L = g_2$. Using (8.97) gives

$$P_{LR} = 1 + \frac{(1 - R)^2 + \omega^2(L^2 + C^2R^2 - 2LC) + \omega^4L^2C^2}{4R}$$

To make P_{LR} equal unity at $\omega = 0$ and to obtain maximally flat response, we must choose $R = 1$ and $L^2 + C^2 - 2LC = 0$. If we specify cutoff to occur at $\omega = 1$ with a passband tolerance of k^2, we also have

$$1 + k^2 = 1 + \frac{L^2C^2}{4}$$

or $LC = 2k$. hence we find $L = C = (2k)^{1/2}$. For Chebyshev response we equate P_{LR} to

$$1 + k^2T_2^2\left(\frac{\omega}{\omega_c}\right) = 1 + k^2\left[2\left(\frac{\omega}{\omega_c}\right)^2 - 1\right]^2$$

in order to determine L and C.

For large values of N the above procedure is very laborious to carry out. In place of this direct substitution scheme general solutions have been worked out.† In addition, tables of element values, i.e., values for the g_k, have been prepared by a number of people.‡

†V. Belevitch, Chebyshev Filters and Amplifier Networks, *Wireless Eng.*, vol. 29, pp. 106–107, April, 1952.

‡L. Weinberg, Network Design by Use of Modern Synthesis Techniques and Tables, *Proc. Natl. Electron. Conf.*, vol. 12, 1956.

S. Cohn, Direct Coupled Resonator Filters, *Proc. IRE*, vol. 45, pp. 187–196, February, 1957.

See also "The Microwave Engineers' Handbook," Horizon House, Inc., 1963–1964.

For the maximally flat network with a power loss ratio

$$P_{\text{LR}} = 1 + \omega^{2N} \tag{8.99}$$

the element values are given by

$$R = 1 \tag{8.100a}$$

$$g_k = 2 \sin \frac{2k - 1}{2N} \pi \qquad k = 1, 2, \ldots, N \tag{8.100b}$$

where g_k is the value either of inductance in henries or of capacitance in farads. Each end of the filter is terminated in a 1-Ω resistance.

For a Chebyshev low-pass filter with $\omega_c = 1$, the element values are given by

$$g_{N+1} = \begin{cases} 1 & N \text{ odd} \qquad (8.101a) \\ 2k^2 + 1 - 2k\sqrt{1 + k^2} & N \text{ even} \qquad (8.101b) \end{cases}$$

When element g_N is a capacitor, $g_{N+1} = R$, but when g_N is an inductor, $g_{N+1} = 1/R$.

$$g_k = \frac{4a_{k-1}a_k}{b_{k-1}g_{k-1}} \tag{8.101c}$$

where

$$a_k = \sin \frac{2k - 1}{2N} \pi$$

$$b_k = \sinh^2 \frac{\beta}{2N} + \sin^2 \frac{k\pi}{N}$$

$$\beta = \ln \frac{\sqrt{1 + k^2} + 1}{\sqrt{1 + k^2} - 1}$$

$$g_1 = \frac{2a_1}{\sinh \beta/2N}$$

Numerical values for the g_k are given in Tables 8.1 and 8.2 for N up to 5, $\omega_c = 1$, and a passband tolerance k^2 equal to 0.0233 (a 0.1-dB ripple in

TABLE 8.1
Values of g_k for maximally flat filter

	N			
k	2	3	4	5
1	1.414	1.00	0.7654	0.6180
2	1.414	2.00	1.848	1.618
3		1.00	1.848	2.000
4			0.7654	1.618
5				0.6180

TABLE 8.2
Values of g_k for Chebyshev filter with $k^2 = 0.0233$

	N			
k	2	3	4	5
1	0.8430	1.0315	1.1088	1.1468
2	0.6220	1.1474	1.3061	1.3712
3		1.0315	1.7703	1.9750
4			0.8180	1.3712
5				1.1468

the passband) for the Chebyshev filter. More extensive tables are given in the references cited.

8.17 FREQUENCY TRANSFORMATIONS

The low-pass filter with cutoff at $\omega_c = 1$ and terminated in a 1-Ω load impedance may be used as a basis for the design of high-pass and bandpass filters with arbitrary resistive load termination. For this reason it is referred to as a prototype filter. For the purpose of this section, it is convenient to denote the frequency variable for the low-pass prototype filter by ω'. The power loss ratio may be expressed in the form

$$P_{LR} = 1 + P(\omega'^2) \tag{8.102}$$

for maximally flat and Chebyshev responses.

If it is desirable to have a load termination R_L different from 1 Ω, the required filter is obtained by multiplying all other reactances and the generator resistance by a factor R_L. That is, the prototype-filter reactances can be viewed as normalized with respect to R_L. The new values for the inductances and capacitances are

$$L'_k = R_L L_k \tag{8.103a}$$

$$C'_k = \frac{C_k}{R_L} \tag{8.103b}$$

$$R' = R_L R \tag{8.103c}$$

where R' denotes the new value of R here, and not the generator resistance in the dual circuit of Fig. 8.21*b*. In the discussion of frequency transformations below, we assume a 1-Ω termination since the change in impedance level of the filter can be made as a last step in the design process.

If we replace the frequency variable ω' by a new variable ω according to

$$\omega' = f(\omega)$$

the power loss ratio becomes

$$P_{LR} = 1 + P(\omega'^2) = 1 + P[f^2(\omega)] \tag{8.104}$$

As a function of ω' or f, this power loss ratio is that of the low-pass prototype filter, but as a function of ω, it has a different characteristic, depending on how $f(\omega)$ is chosen. A number of different frequency transformations, or mappings, are considered below.

Frequency Expansion

If it is desirable to change the cutoff frequency from unity to some other value ω_c, we choose

$$f(\omega) = \frac{\omega}{\omega_c}$$

and thus

$$P_{LR} = 1 + P\left(\frac{\omega^2}{\omega_c^2}\right) \tag{8.105}$$

Cutoff occurs when the argument ω/ω_c equals unity or when $\omega = \omega_c$. The series reactances and shunt susceptances in the prototype filter must be replaced by new reactances and susceptances

$$jX_k' = j\left(\frac{\omega}{\omega_c}\right)L_k \qquad jB_k' = j\left(\frac{\omega}{\omega_c}\right)C_k$$

when ω' is changed to ω/ω_c in order to yield the power loss ratio given by (8.105). Examination of the latter equations shows that the new values of the L_k and C_k must be

$$L_k' = \frac{L_k}{\omega_c} \tag{8.106a}$$

$$C_k' = \frac{C_k}{\omega_c} \tag{8.106b}$$

The power loss ratios in terms of ω' and ω are compared in Figs. 8.22*a* and *b*.

Low-Pass to High-Pass Transformation

A high-pass filter is obtained by choosing $f(\omega)$ equal to $-\omega_c/\omega$ to yield a power loss ratio

$$P_{LR} = 1 + P\left(\frac{\omega_c^2}{\omega^2}\right) \tag{8.107}$$

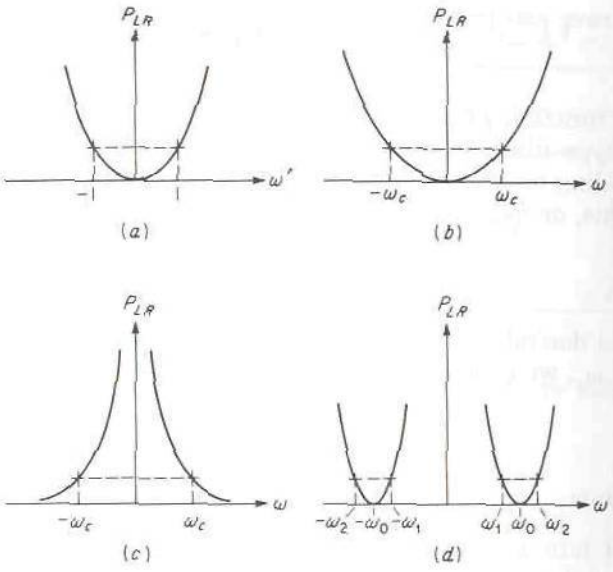


FIGURE 8.22
(*a*) Low-pass prototype-filter *response*; (*b*) frequency expansion; (*c*) low-pass-to-high-pass transformation; (*d*) low-pass-to-bandpass transformation.

This frequency transformation maps the point $\omega' = 0$ into the points $\omega = \pm\infty$, the points $\omega' = \pm 1$ into $\omega = \mp\omega_c$, and the points $\omega' = \pm\infty$ into $\omega = 0$. The effect is to interchange the passband and stopband regions, as illustrated by Figs. 8.22*a* and *c*. To obtain a power loss ratio of the form (8.107), the series reactances and shunt susceptances in the new filter must be chosen as follows:

$$jX'_k = -j\frac{\omega_c}{\omega}L_k \qquad jB'_k = -j\frac{\omega_c}{\omega}C_k$$

From these equations it is seen that all inductances L_k must be replaced by capacitances C'_k and the C_k must be replaced by inductances L'_k in the following manner:

$$C'_k = \frac{1}{\omega_c L_k} \tag{8.108a}$$

$$L'_k = \frac{1}{\omega_c C_k} \tag{8.108b}$$

Low-Pass to Bandpass Transformation

To obtain a bandpass filter, consider a change of variable according to

$$\omega' = f(\omega) = \frac{\omega_o}{\omega_2 - \omega_1}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \tag{8.109}$$

FIGURE 8.23
Series and parallel tuned circuits.

This equation may be solved for ω to give

$$\omega = \omega' \frac{\omega_2 - \omega_1}{2} \pm \sqrt{\omega'^2 \left(\frac{\omega_2 - \omega_1}{2} \right)^2 + \omega_0^2}$$

If we also choose $\omega_0^2 = \omega_1 \omega_2$, we obtain

$$\omega = \omega' \frac{\omega_2 - \omega_1}{2} \pm \frac{1}{2} \sqrt{\omega'^2 (\omega_2 - \omega_1)^2 + 4\omega_1 \omega_2} \tag{8.110}$$

The point $\omega' = 0$ is seen to map into the points $\omega = \pm\omega_0$, and $\omega' = \pm 1$ maps into the four points $\pm(\omega_2 - \omega_1)/2 \pm (\omega_2 + \omega_1)/2 = \pm\omega_2$ and $\pm\omega_1$. Thus the prototype-filter passband between ± 1 maps into passbands extending from ω_1 to ω_2 and $-\omega_1$ to $-\omega_2$, which represent bandpass filters *with band centers at $\pm\omega_0$ equal to the geometric mean of ω_1 and ω_2, as in* Fig. 8.22*d*.

The required filter elements may be deduced by considering the frequency behavior of series and parallel connections of L and C, as in Fig. 8.23. For the series circuit we have

$$jX' = j\omega L + \frac{1}{j\omega C} = j\sqrt{\frac{L}{C}} \left(\omega\sqrt{LC} - \frac{1}{\omega\sqrt{LC}} \right)$$

and for the parallel circuit we have

$$jB' = j\omega C + \frac{1}{j\omega L} = j\sqrt{\frac{C}{L}} \left(\omega\sqrt{LC} - \frac{1}{\omega\sqrt{LC}} \right)$$

If we make $\sqrt{LC}$ equal to ω_0^{-1}, we obtain

$$jX' = j\sqrt{\frac{L}{C}} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) \qquad jB' = j\sqrt{\frac{C}{L}} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)$$

The required frequency transformation may now be seen to be obtained if we replace the reactance $jX_k = j\omega L_k$ and the susceptance $jB_k = j\omega C_k$ in the prototype filter by series and parallel tuned circuits such that

$$jX_k' = j\sqrt{\frac{L_k'}{C_k'}} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) = jL_k \frac{\omega_0}{\omega_2 - \omega_1} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)$$

$$jB_k' = j\sqrt{\frac{C_k'}{L_k'}} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) = jC_k \frac{\omega_0}{\omega_2 - \omega_1} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)$$

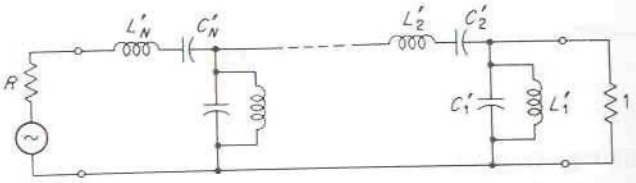


FIGURE 8.24
A bandpass-filter network.

This requires that we choose

$$L'_k C'_k = \frac{1}{\omega_0^2} = \frac{1}{\omega_1 \omega_2}$$

$$\sqrt{\frac{L'_k}{C'_k}} = \omega_0 L'_k = \frac{\omega_0 L_k}{\omega_2 - \omega_1} \tag{8.111a}$$

or

$$L'_k = \frac{L_k}{\omega_2 - \omega_1} \tag{8.111b}$$

for the series circuit. For the parallel circuit we must choose

$$L'_k C'_k = \frac{1}{\omega_0^2} = \frac{1}{\omega_1 \omega_2}$$

$$\sqrt{\frac{C'_k}{L'_k}} = \omega_0 C'_k = \frac{\omega_0 C_k}{\omega_2 - \omega_1} \tag{8.112a}$$

or

$$C'_k = \frac{C_k}{\omega_2 - \omega_1} \tag{8.112b}$$

The resultant filter network is illustrated in Fig. 8.24.

Period Bandpass Mapping

The general concept of frequency mapping may be applied in a variety of other ways as well. All that is necessary is to replace the reactances $j\omega' L_k$ and susceptances $j\omega' C_k$ in the prototype filter by other reactance and susceptance functions having ω' replaced by a new function $f(\omega)$ of ω. As a further example we shall consider the effect of replacing all L_k by short-circuited transmission-line stubs of length l and characteristic impedance Z_k and the capacitors C_k by open-circuited stubs of length l and characteristic

admittance Y_k. The new reactance and susceptance functions become†

$$jX'_k = jZ_k \tan\left(\frac{\omega}{c}l\right) = jZ_k \tan\theta = jg_k \tan\theta$$

$$jB'_k = jY_k \tan\theta = jg_k \tan\theta$$

This filter has a power loss ratio given by

$$P_{LR} = 1 + P(\tan^2\theta) \tag{8.113}$$

The frequency transformation $\omega' = \tan(\omega l/c) = \tan\theta$ *maps the whole* ω' axis periodically into intervals of length π in the θ domain or of length $\pi c/l$ in the ω domain. This filter would be satisfactory at low frequencies where junction effects at the points where the stubs are connected to the main line are negligible and where the separation between stubs could be kept very small in terms of wavelength.‡

At microwave frequencies the alternative occurrence of series and parallel tuned circuits in the filter derived from the low-pass prototype filter by a frequency transformation is an undesirable feature. It is difficult to construct a filter of this type using microwave elements. However, it is possible to convert the filter with series tuned elements into an equivalent filter containing only parallel tuned circuits in a cascade connection, or to convert the circuit into one containing only series tuned circuits. The desired transformation may be obtained by the use of impedance and admittance inverters or quarter-wave transformers, as discussed in the next section.

8.18 IMPEDANCE AND ADMITTANCE INVERTERS

An impedance inverter is an ideal quarter-wave transformer. A load impedance connected at one end is seen as an impedance that has been inverted with respect to the characteristic impedance squared at the input. Impedance inverters may be used to convert a bandpass-filter network of the type shown in Fig. 8.25 into a network containing only series tuned circuits. By using admittance inverters the bandpass filter can be converted into a network using only parallel tuned circuits. Furthermore, by choosing the inverters correctly all of the inductors and capacitors can be chosen to have the same values. Thus impedance and admittance inverters enable us

†This transformation is known as Richard's theorem.

‡If a filter is made up of transmission lines of commensurate length and resistors, the only frequency variable occurring is $\tan(\omega l/c)$. In this case the frequency transformation $\omega' = \tan(\omega l/c)$ permits the use of conventional low-frequency-network synthesis techniques to be applied.

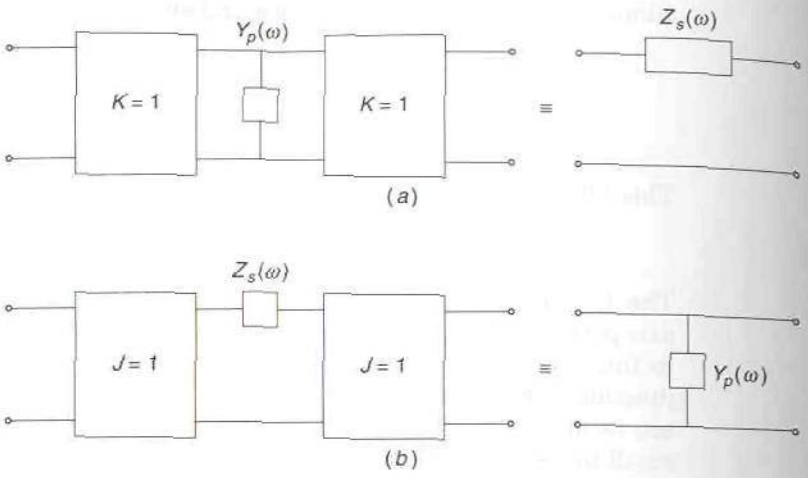


FIGURE 8.25
(*a*) Impedance inverter used to convert a parallel admittance into an equivalent series impedance; (*b*) admittance inverter used to convert a series impedance into an equivalent parallel admittance.

to use identical resonators, either series or parallel tuned, throughout the network.

In our initial discussion of inverters, we will assume that they are ideal frequency-independent elements. Later on we will consider the practical realization of inverters and how to compensate for their frequency-dependent characteristics.

Consider the parallel admittance element $Y_p(\omega)$ with an ideal impedance inverter with characteristic impedance K connected on both sides as shown in Fig. 8.25*a*. A short circuit at the output will be transformed to an open circuit in parallel with Y_p. The input impedance is given by

$$Z_{\text{in}} = \frac{K^2}{Z_p} = K^2 Y_p = Y_p = Z_s$$

Thus the shunt element with two impedance inverters converts the shunt admittance into an equivalent series impedance $Z_s(\omega) = Y_p(\omega)$. If Y_p is a parallel tuned resonator with $Y_p = j\omega C - j/\omega L = j\omega C(1 - \omega_0^2/\omega^2)$, it is converted into a series tuned circuit with $Z_s = j\omega L(1 - \omega_0^2/\omega^2)$ with the inductance L in henries having the same numerical value as the capacitance C in farads. If we want to convert an admittance $Y_1 = j\omega C_1(1 - \omega_0^2/\omega^2)$ into a particular series tuned circuit with arbitrary induc-

tance L, then we must choose K so that

$$K^2 j\omega C_1\left(1 - \frac{\omega_0^2}{\omega^2}\right) = j\omega L\left(1 - \frac{\omega_0^2}{\omega^2}\right)$$

or

$$K = \sqrt{\frac{L}{C_1}}$$

Thus we can transform a shunt element into a series element having the same dependence on frequency and an arbitrary impedance level.

Consider next a series element $Z_s(\omega)$ with an admittance inverter with characteristic admittance J connected on both sides as shown in Fig. 8.25*b*. An open circuit at the output is transformed into a short circuit in series with $Z_s(\omega)$ so the input admittance will be

$$Y_{\text{in}} = \frac{J^2}{Y_s} = J^2 Z_s = Y_p$$

Thus the series element with the two admittance inverters will appear from the terminals to be an equivalent shunt admittance. If Z_s is the impedance $j\omega L(1 - \omega_0^2/\omega^2)$ of a series tuned circuit, it is converted into a parallel tuned circuit with admittance $Y_p = j\omega L(1 - \omega_0^2/\omega^2) = j\omega C(1 - \omega_0^2/\omega^2)$, *where C is numerically equal to L. For a series tuned circuit* $j\omega L_1(1 - \omega_0^2/\omega^2)$ to be transformed into an arbitrary parallel tuned circuit with $Y_p = j\omega C(1 - \omega_0^2/\omega^2)$, we must choose $J^2 L_1 = C$ or

$$J = \sqrt{\frac{C}{L_1}}$$

From another point of view we see that a series element can be replaced by a shunt element with an impedance inverter at the input and output. Similarly, a shunt element can be replaced by a series element with an admittance inverter connected at each port. The admittance and impedance functions for parallel tuned and series tuned resonators can be expressed in the form

$$Y = j\omega C\left(1 - \frac{\omega_0^2}{\omega^2}\right) = j\sqrt{\frac{C}{L}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)$$

$$Z = j\omega L\left(1 - \frac{\omega_0^2}{\omega^2}\right) = j\sqrt{\frac{L}{C}}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)$$

The factor $\sqrt{C/L}$ is the admittance level of the resonator and $\sqrt{L/C}$ is the impedance level. The remaining factor is the frequency variable.

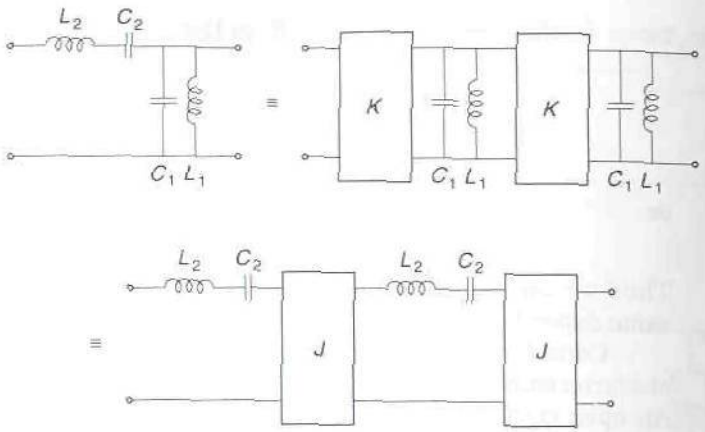


FIGURE 8.26
Use of inverters to convert a section of a filter into a network using only identical parallel or series tuned resonators.

We will illustrate the use of inverters to convert the circuit shown in Fig. 8.26 into one with two identical parallel tuned resonators or one with two identical series tuned resonators. For the first case we choose K so that $K^2 j\omega C_1(1 - \omega_0^2/\omega^2) = j\omega L_2(1 - \omega_0^2/\omega^2)$ which requires that $K = \sqrt{L_2/C_1}$. For the second case we choose J so that $J^2 j\omega L_2(1 - \omega_0^2/\omega^2) = j\omega C_1(1 - \omega_0^2/\omega^2)$ which requires $J = \sqrt{C_1/L_2}$. Impedance and admittance level changing can be accomplished by using different input and output inverters as shown in Fig. 8.27. For example, in Fig. 8.27*a* the parallel admittance appears as a series element $K_1^2 Y_p$ at the left-side port and as a series element $K_2^2 Y_p$ at the right-side port. Similarly, in Fig. 8.27*b* the series element Z_s is made to appear as a parallel admittance $J_1^2 Z_s$ at the left-side port and as an admittance $J_2^2 Z_s$ at the right-side port. We can modify the impedance and admittance inverters to accommodate an arbitrary change in the impedance and admittance levels of a resonator. In Fig. 8.27*c* the impedance level $\sqrt{L_1/C_1}$ of the resonator is changed to $\sqrt{L_0/C_0}$ by changing the impedance K' of the inverters to K, where K is chosen so that

$$\frac{K'^2}{\sqrt{L_1/C_1}} = \frac{K^2}{\sqrt{L_0/C_0}} = \frac{K'^2}{\omega_0 L_1} = \frac{K^2}{\omega_0 L_0}$$

or

$$K = K'\sqrt{\frac{L_0}{L_1}}$$

From the terminals the new circuit is equivalent to the old one. In Fig. 8.27*d* we show a similar transformation of the admittance $\sqrt{C_1/L_1}$ of the parallel tuned resonator to a value $\sqrt{C_0/L_0}$ obtained by replacing the

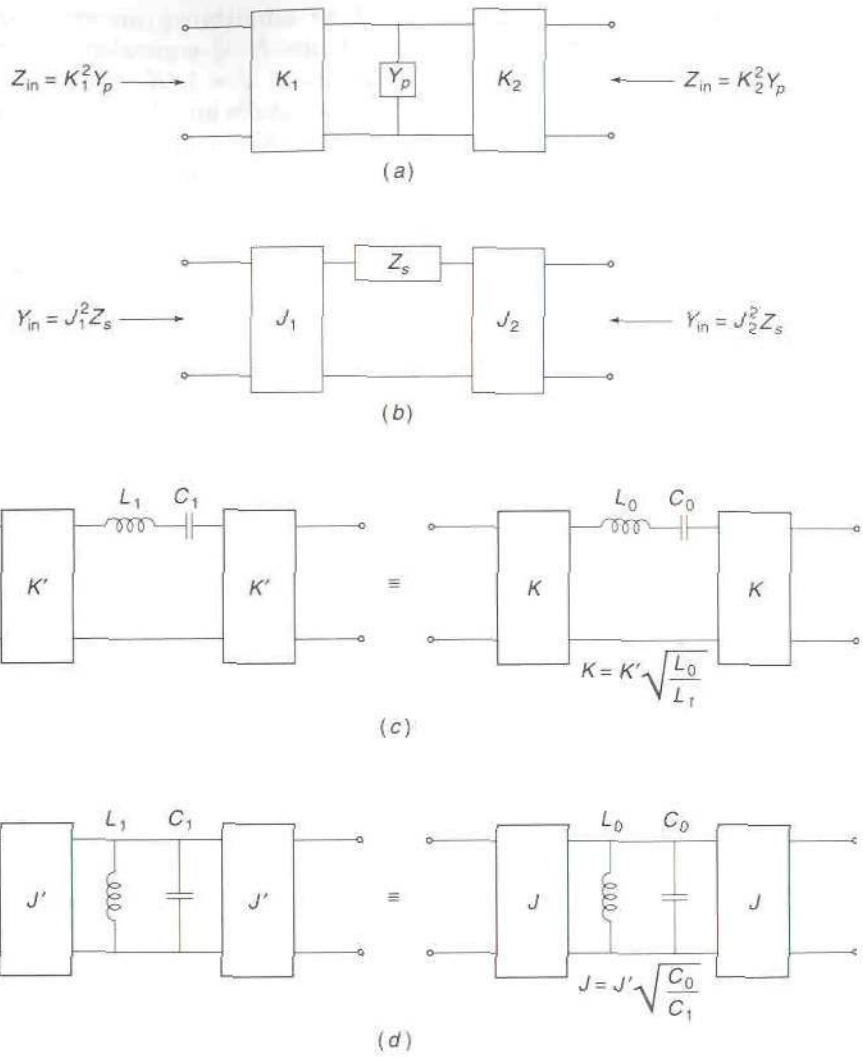


FIGURE 8.27

(*a*) Impedance level changing property of two impedance inverters; (*b*) admittance level changing properties of two admittance inverters; (*c*) equivalent basic sections with a series tuned resonator; (*d*) equivalent basic section with a parallel tuned resonator.

inverters with characteristic admittance J' by new inverters with characteristic admittance $J = J'\sqrt{C_0/C_1}$.

Impedance and admittance inverters are ideal quarter-wave transformers. There is no basic difference in their inverting properties. The only distinction that we make is to use the symbol K to denote the characteristic impedance of the impedance inverter and we use J to denote the character-

istic admittance of an admittance inverter. An impedance inverter with characteristic impedance K is equivalent to an admittance inverter with characteristic admittance $J = 1/K$. When $K = J = 1$ we have a unity inverter and it does not make any difference whether we identify this as an *impedance inverter with* $K = 1$ *or as an admittance* inverter with $J = 1$.

In Fig. 8.28*a* we show a bandpass filter terminated in a load resistance R_L and a source resistance R_g. We will use impedance inverters to convert this circuit into one using identical series tuned resonators with elements L_0 and C_0 and terminated at both ends in a resistance R. The transformation process is easiest to understand by carrying it out as a succession of several intermediate transformations. We first introduce unity inverters to convert all of the parallel tuned resonators into series tuned resonators. Thus, with all $J_i = K_i$ equal to unity, we obtain the network shown in Fig. 8.28*b*, where $L_1' = C_1$ and $L_3' = C_3$. We now change the impedance level of resonators 1 and 3 to $\sqrt{L_0/C_0}$ by changing the first and second inverters to inverters with $K_1' = K_2' = \sqrt{L_0/L_1'} = \sqrt{L_0/C_1}$ and the third and fourth inverters to inverters with $K_3' = K_4' = \sqrt{L_0/L_3'} = \sqrt{L_0/C_3}$. The new network is shown in Fig. 8.28c. This impedance shifting is the same as that illustrated in Fig. 8.27*c*.

As a next step we change the impedance level of resonator 2 from $\sqrt{L_2/C_2} = \omega_0 L_2$ to $\omega_0 L_0$ by changing inverters 2 and 3 so that

$$\frac{K_2^2}{\omega_0 L_0} = \frac{K_2'^2}{\omega_0 L_2} \qquad \frac{K_3^2}{\omega_0 L_0} = \frac{K_3'^2}{\omega_0 L_2}$$

and hence

$$K_2 = K_2'\sqrt{\frac{L_0}{L_2}} = \sqrt{\frac{L_0^2}{L_2 C_1}} \qquad K_3 = K_3'\sqrt{\frac{L_0}{L_2}} = \sqrt{\frac{L_0^2}{L_2 C_3}}$$

The resultant network is shown in Fig. 8.28*d*.

In order to change the load termination from R_L to R, we modify the first inverter so that

$$\frac{K_1^2}{R} = \frac{K_1'^2}{R_L}$$

and thus choose

$$K_1 = K_1'\sqrt{\frac{R}{R_L}} = \sqrt{\frac{L_0 R}{R_L C_1}}$$

The last modification needed to complete the transformation is to change the fourth inverter to one with an impedance

$$K_4 = K_4'\sqrt{\frac{L_0}{L_4}} = \sqrt{\frac{L_0^2}{L_4 C_3}}$$

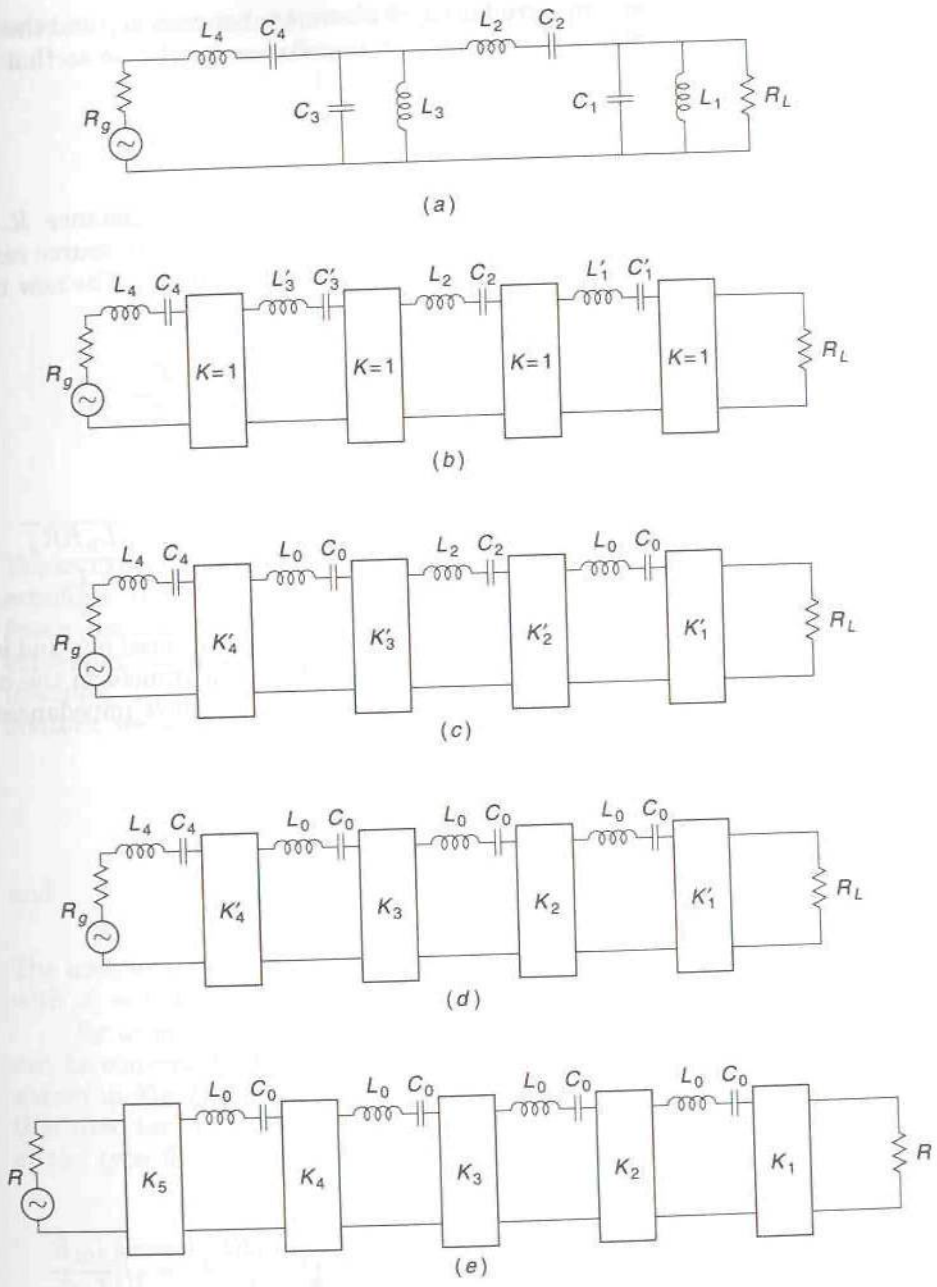


FIGURE 8.28
Illustration of successive transformations used to convert the bandpass filter in (*a*) into a network with identical series tuned resonators and equal input and output terminating resistance R as shown in (*e*).

and to introduce a fifth inverter between R_g and the fourth resonator. If we place an inverter with impedance K_5' chosen so that

$$\frac{K_5'^2}{R} = R_g$$

we can replace R_g by the new source resistance R. After introducing the fifth inverter to accommodate the change in source resistance, we must then change it again when L_4 is replaced by L_0. The new impedance K_5 must be chosen so that

$$\frac{K_5^2}{L_0} = \frac{K_5'^2}{L_4}$$

and hence

$$K_5 = K_5'\sqrt{\frac{L_0}{L_4}} = \sqrt{\frac{L_0 R R_g}{L_4}}$$

The final transformed network is the desired one and is shown in Fig. 8.28*e*.

The required inverters that will transform the original filter network to the final form shown in Fig. 8.28*e* have impedances given by

$$K_1 = \sqrt{\frac{L_0 R}{R_L C_1}} \tag{8.114a}$$

$$K_2 = \sqrt{\frac{L_0^2}{L_2 C_1}} \tag{8.114b}$$

$$K_3 = \sqrt{\frac{L_0^2}{L_2 C_3}} \tag{8.114c}$$

$$K_4 = \sqrt{\frac{L_0^2}{L_4 C_3}} \tag{8.114d}$$

$$K_5 = \sqrt{\frac{L_0 R R_g}{L_4}} = \sqrt{\frac{L_0 R}{L_4 G_g}} \tag{8.114e}$$

If instead of specifying identical resonators we choose L_{01}, L_{02}, L_{03}, and L_{04} for the new inductances, R_{00} for the new load resistance, and R_{05} for the new source resistance (this is viewed as a parallel element), a similar

series of transformations would give

$$K_1 = K_{10} = \sqrt{\frac{L_{01}R_{00}}{C_1 R_L}} \tag{8.115a}$$

$$K_2 = K_{21} = \sqrt{\frac{L_{02}L_{01}}{L_2 C_1}} \tag{8.115b}$$

$$K_3 = K_{32} = \sqrt{\frac{L_{03}L_{02}}{C_3 L_2}} \tag{8.115c}$$

$$K_4 = K_{43} = \sqrt{\frac{L_{04}L_{03}}{L_4 C_3}} \tag{8.115d}$$

$$K_5 = K_{54} = \sqrt{\frac{R_{05}L_{04}}{G_g L_4}} \tag{8.115e}$$

These expressions exhibit a regular pattern and can, by induction, be extended to a network with a larger number of resonators or to one with fewer resonators. For example, if we have three resonators then in the expression for K_{43} we interpret L_{04} as the new source resistance R_{04} and interpret L_4 as the old source resistance. For a network with five resonators, we would obtain

$$K_{54} = \sqrt{\frac{L_{05}L_{04}}{C_5 L_4}}$$

and

$$K_{65} = \sqrt{\frac{R_{06}L_{05}}{R_g C_5}}$$

The inverters shown in Fig. 8.28 can also be viewed as admittance inverters with $J_i = 1/K_i$.

By using admittance inverters the bandpass filter shown in Fig. 8.28*a* can be converted to a network that uses only parallel tuned resonators as shown in Fig. 8.29. The analysis can be carried out in a manner similar to that used for impedance inverters as a series of elementary transformations of the type illustrated in Figs. 8.26*b* and 8.27*d*. The required expressions

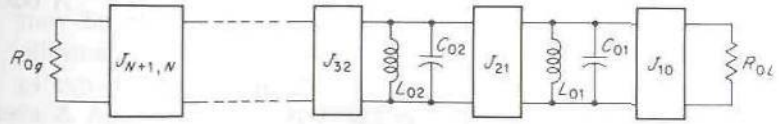


FIGURE 8.29
Equivalent filter network obtained by use of admittance inverters.

for the characteristic admittances of the inverters can be obtained, as expected, by interchanging the role of inductance and resistance with those of capacitance and conductance. It is readily found that

$$J_1 = J_{10} = \sqrt{\frac{C_{01}G_{00}}{C_1R_L}} \tag{8.116a}$$

$$J_2 = J_{21} = \sqrt{\frac{C_{02}C_{01}}{L_2C_1}} \tag{8.116b}$$

$$J_3 = J_{32} = \sqrt{\frac{C_{03}C_{02}}{C_3L_2}} \tag{8.116c}$$

$$J_4 = J_{43} = \sqrt{\frac{C_{04}C_{03}}{L_4C_3}} \tag{8.116d}$$

$$J_5 = J_{54} = \sqrt{\frac{G_{05}C_{04}}{C_5R_g}} \tag{8.116e}$$

If we denote the load and source terminations by R_L and R_g in the unmodified bandpass filter and by R_{0L} and R_{0g} for the transformed network, then the general expressions for $K_{k,k-1}$ and $J_{k,k-1}$ for a filter with N resonators are

$$K_{10} = \sqrt{\frac{L_{01}R_{0L}}{C_1R_L}} \tag{8.117a}$$

$$K_{k,k-1} = \sqrt{\frac{L_{0k}L_{0k-1}}{L_kC_{k-1}}} \qquad k \text{ odd} \tag{8.117b}$$

$$K_{k,k-1} = \sqrt{\frac{L_{0k}L_{0k-1}}{C_kL_{k-1}}} \qquad k \text{ even} \tag{8.117c}$$

$$K_{N+1,N} = \sqrt{\frac{R_{0g}L_{0N}}{R_LC_N}} \qquad N \text{ odd} \tag{8.117d}$$

$$K_{N+1,N} = \sqrt{\frac{R_{0g}R_gL_{0N}}{L_N}} \qquad N \text{ even} \tag{8.117e}$$

and

$$J_{10} = \sqrt{\frac{C_{01}}{C_1 R_{0L} R_L}} \tag{8.118a}$$

$$J_{k,k-1} = \sqrt{\frac{C_{0k} C_{0k-1}}{C_k L_{k-1}}} \qquad k \text{ odd} \tag{8.118b}$$

$$J_{k,k-1} = \sqrt{\frac{C_{0k} C_{0k-1}}{L_k C_{k-1}}} \qquad k \text{ even} \tag{8.118c}$$

$$J_{N+1,N} = \sqrt{\frac{C_{0N}}{R_{0g} R_g C_N}} \qquad N \text{ odd} \tag{8.118d}$$

$$J_{N+1,N} = \sqrt{\frac{C_{0N} R_g}{R_{0g} L_N}} \qquad N \text{ even} \tag{8.118e}$$

The impedance and admittance inverters used in the preceding analysis were assumed to be ideal frequency-independent quarter-wave transformers of electrical length $\pi/2$ independent of frequency. Such ideal inverters do not, of course, exist. Nevertheless, the theory can be applied in practice. For very narrowband bandpass filters, say bandwidths of 1 or 2 percent, a quarter-wave length of transmission line or waveguide does not depart appreciably from the ideal inverter having a $\pi/2$ electrical length. A filter designed on the basis of ideal inverters would have a response very close to the theoretical response in this case. For greater bandwidths the departure of a quarter-wave transformer from the ideal can be incorporated into the design by splitting the transformer into an ideal one with two additional short lengths of transmission line on either side to account for the excess or deficit in phase shift from the ideal phase shift of 90°. For example, a transformer of length $0.3\lambda_0$ at a wavelength of λ_0 can be treated as a transformer of length $0.25\lambda_0$ plus sections of line of length $0.025\lambda_0$ on both sides. The excess length of line may be incorporated into and made part of the resonant circuit on either side of the inverter.

Lumped-element circuits that act as impedance inverters and admittance inverters are illustrated in Fig. 8.30 along with their equivalent characteristic impedance or admittance. These circuits involve negative elements and are frequency-dependent. However, the negative element may be absorbed into the elements of the adjacent resonant circuits to eliminate them from the overall network. The resultant filter then consists of tuned circuits coupled by single reactive elements, as illustrated in Fig. 8.31. Application of these techniques to microwave filter design is discussed in the following sections.

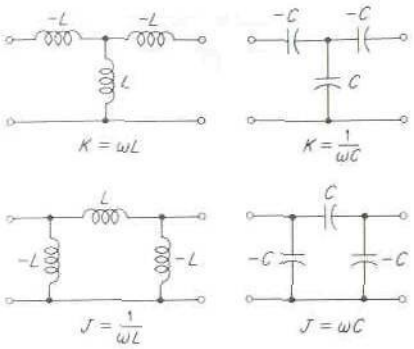


FIGURE 8.30
Lumped-element inverters.

Example 8.1. In this example we will design a two-resonator filter with a final form as shown in Fig. 8.32*a*. The filter is to be used in an intermediate-frequency amplifier with center frequency of 10 MHz, a bandwidth of 0.5 MHz and having a load and source termination of 1,000 Ω. The passband ripple is 2 dB.

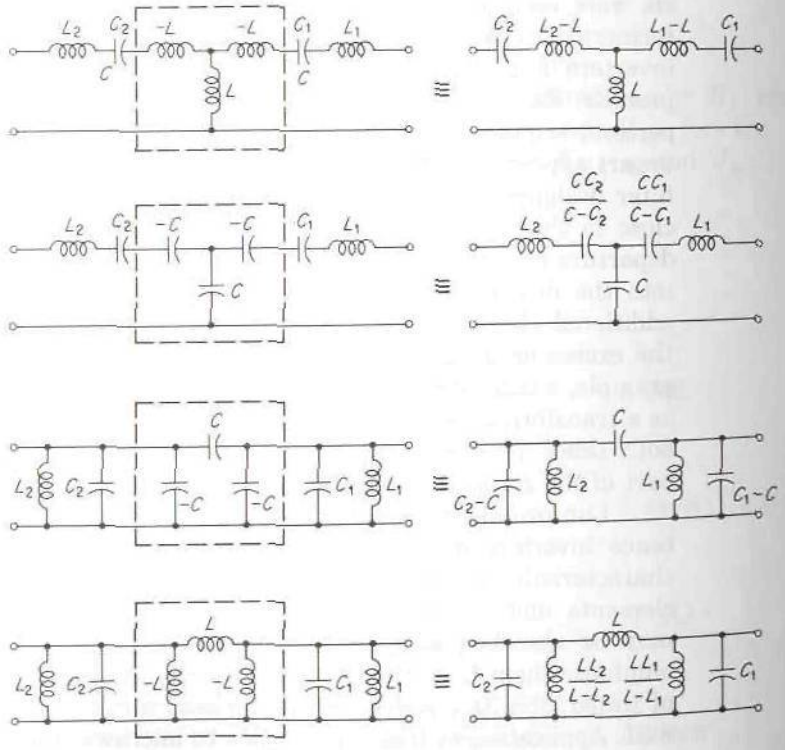


FIGURE 8.31
Reactance-coupled resonator-type circuits obtained by use of lumped-element inverters.

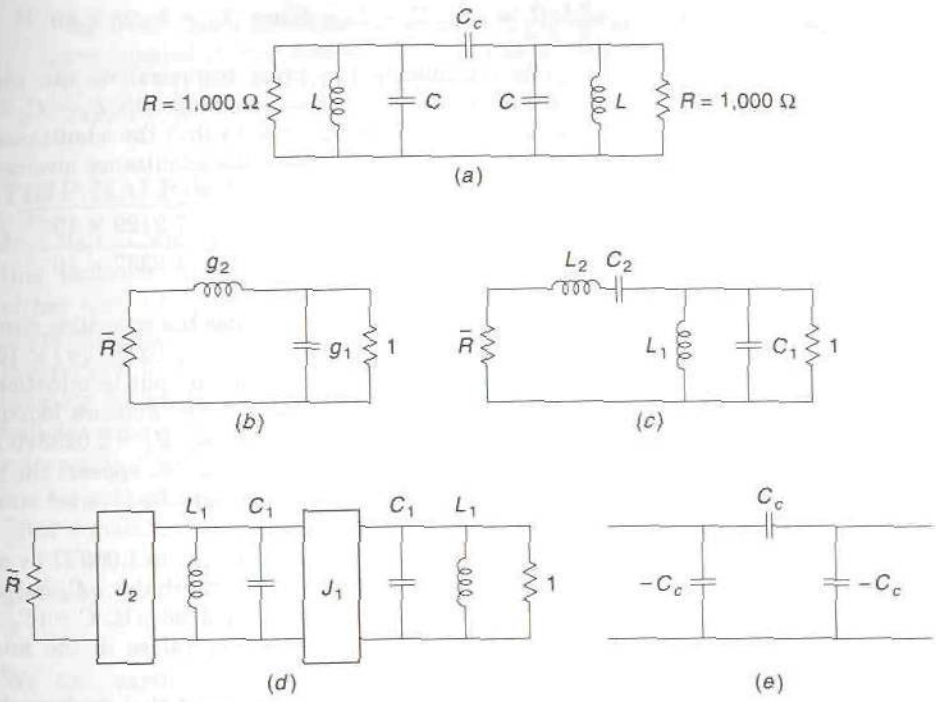


FIGURE 8.32
(*a*) Bandpass filter; (*b*) low-pass prototype circuit; (*c*) bandpass prototype circuit; (*d*) transformed bandpass prototype circuit; (*e*) admittance inverter used to obtain the circuit shown in (*a*).

We begin the design by first designing a low-pass prototype circuit with a cutoff radian frequency $\omega_c = 1$ as shown in Fig. 8.32*b*. For a 2-dB ripple the maximum power loss ratio in the passband is $10^{0.2} = 1.5849 = 1 + k^2$. The parameter β is

$$\beta = \ln \frac{\sqrt{1 + k^2} + 1}{\sqrt{1 + k^2} - 1} = 2.1661$$

From (8.101) we obtain $\bar{R} = 2k^2 + 1 - 2k\sqrt{1 + k^2} = 0.244175$ for the low-pass filter input termination. We also obtain from (8.101) $a_1 = \sin \pi/4 = 0.7071$, $a_2 = \sin 3\pi/4 = 0.7071$, $b_1 = 1.32306$, $b_2 = 0.32305$, and $g_1 = 2.4881$ and $g_2 = 0.6075$. Hence, for the low-pass prototype circuit in Fig. 8.32*b*, we have $C_1 = g_1 = 2.4881$ F and $L_2 = g_2 = 0.6075$ H. The values of ω at the band edges are ω_1 and ω_2. Since $\omega_1\omega_2 = \omega_0^2$ and $\omega_2 - \omega_1 = 2\pi \times 0.5 \times 10^6$, we get $\omega_2 = \omega_1 + \pi \times 10^6 = \omega_0^2/\omega_1$. By solving this equation we obtain $\omega_1 = 6.1281 \times 10^7$, $\omega_2 = 6.4423 \times 10^7$. The corresponding frequencies are 9.7532 and 10.2532 MHz. For the bandpass-filter prototype circuit, we use (8.111*b*) and (8.112*b*) to obtain $C_1 = 2.4881/\pi \times 10^6 = 0.79199$ μF and $L_2 = 0.6075/\pi \times 10^6 = 0.19337$ μH. The elements L_1 and C_2 are found using

$\omega_0^2 L_1 C_1 = \omega_0^2 L_2 C_2 = 1$ and are $L_1 = 3.198 \times 10^{-10}$ H and $C_2 = 1.3099 \times 10^{-9}$ F.

By introducing two unity inverters, we can convert the series tuned resonator to a parallel tuned resonator with $L_2' = C_2 = 1.3099 \times 10^{-9}$ H and $C_2' = L_2 = 0.19337\ \mu$F. In order to shift the admittance level of this resonator from $\omega_0 C_2'$ to $\omega_0 C_1$, we change the admittance inverters to ones with

$$J_1 = J_2 = \sqrt{\frac{C_1}{C_2'}} = \sqrt{\frac{7.9199 \times 10^{-7}}{1.9337 \times 10^{-7}}} = 2.0238$$

For the admittance inverters we use the capacitive circuit shown in Fig. 8.32*e* for which $J_1 = J_2 = \omega_0 C_c$; thus $C_c = (2.0238/2\pi) \times 10^{-7} = 0.0322\ \mu$F. Since we want to use the same input and output terminations, we change J_2 to J_2' so that $\overline{R} = 1$ appears the same to the network looking through the original inverter; thus $J_2'^2 \times 1 = J_2^2 \times \overline{R}$ so $J_2' = 2.0238\sqrt{0.244175} = 1$. This last result shows that a 1-Ω source resistance appears the same to the network as the 0.244175-Ω resistor seen through the inverter with $J = J_2$. Thus we can discard the input inverter.

We now scale the impedance level to 1,000 Ω by multiplying L_1 by 1,000 and dividing C_1 by 1,000. We also absorb the $-C_c$ in the remaining inverter as part of C_1. Thus we obtain $L = 0.3198\ \mu$H, $C = (C_1 - C_c)/1{,}000 = 760$ pF, and $C_c = 32.2$ pF, for the element values in the final filter shown in Fig. 8.32*a*.

The double-tuned coupled circuit that we have designed is a classic one that has been widely used in low-frequency radio IF amplifiers. The power transmission coefficient is equal to $1/P_{LR}$ and is shown in Fig. 8.33. If the circuit is designed for a maximally flat response, it is said to be critically coupled. When the response has a dip in the center, the circuit is overcoupled. The coupling coefficient for the circuit equals $C_c/(C + C_c)$. For critical coupling the coupling coefficient equals $1/Q = \omega_0 L/R = 0.02$. For the circuit that we designed, the coupling coefficient equals 0.0406 so clearly the circuit is overcoupled. Two other well-known properties of this classic circuit are that

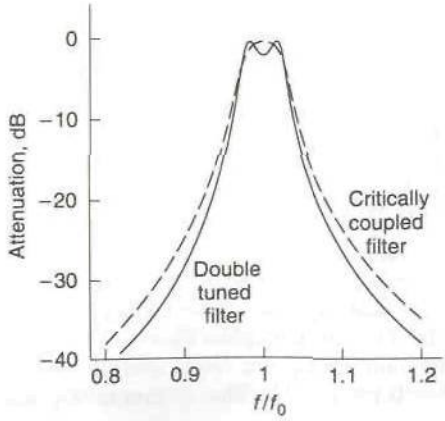


FIGURE 8.33
Power transmission coefficient for the double-tuned filter shown in Fig. 8.32*a*. The dashed curve shows the response for a critically coupled circuit with a 3-dB bandwidth equal to $0.05 f_0$.

the 3-dB fractional bandwidth equals $\sqrt{2}/Q$ for critical coupling, and for an overcoupled circuit having a 3-dB dip at ω_0, the coupling coefficient equals $(1 + \sqrt{2})$ times critical coupling and the 3-dB fractional bandwidth equals $2\sqrt{2.414}/Q$.

8.19 A MICROSTRIP HALF-WAVE FILTER

In Chap. 7 we showed that open-circuited or short-circuited transmission-line sections one-quarter or one-half wavelength long were equivalent to either series or parallel tuned lumped-parameter LC circuits. Thus it is not surprising to find that sections of transmission lines are widely used as resonators in filters. Such structures are particularly appropriate for microstrip filters because of easy fabrication and low cost. In this section we will describe a microstrip filter that uses one-half wavelength open-circuited microstrip lines for the resonators. The resonators are coupled by means of the capacitance of the gap between the resonator sections. A typical half-wave filter consisting of three resonator sections is shown in Fig. 8.34.

For an open-circuited transmission line the input admittance is given by

$$Y_{\text{in}} = jY_c \tan \beta l$$

We can expand Y_{in} in a Taylor series about the frequency ω_0 where $\beta(\omega_0)l = \pi$; thus

$$Y_{\text{in}} \approx jY_c l \frac{d\beta}{d\omega}\bigg|_{\omega_0} [\sec^2 \beta(\omega_0)l](\omega - \omega_0)$$

$$= jY_c l \beta_0'(\omega - \omega_0) = jY_c \frac{\pi \omega_0 \beta_0'}{\beta_0} \frac{\omega - \omega_0}{\omega_0}$$

where β_0' equals $d\beta/d\omega$ evaluated at $\omega = \omega_0$. For a parallel tuned LC circuit,

$$Y_{\text{in}} = j\omega C - \frac{j}{\omega L} = j\omega C\left(1 - \frac{\omega_0^2}{\omega^2}\right)$$

$$= j\omega_0 C \frac{1}{\omega\omega_0}(\omega - \omega_0)(\omega + \omega_0) \approx 2j\omega_0 C \frac{\omega - \omega_0}{\omega_0}$$

$$= j2\sqrt{\frac{C}{L}} \frac{\omega - \omega_0}{\omega_0}$$

Thus, for frequencies within around ± 10 percent of ω_0, the open-circuited transmission line is equivalent to a parallel tuned LC circuit if we choose

$$\sqrt{\frac{C}{L}} = \frac{\omega_0 \pi \beta_0'}{2\beta_0} Y_c \tag{8.119}$$

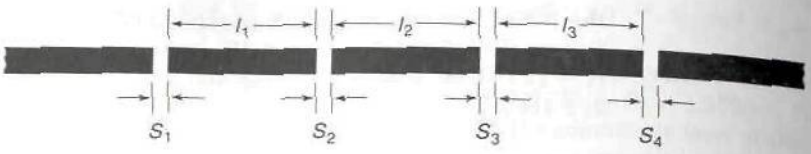


FIGURE 8.34
A microstrip half-wave filter with three resonators.

In the half-wave filter the open-circuited transmission lines can be used as equivalent parallel tuned resonators. We can choose all of the resonators to be identical provided we insert appropriate admittance inverters between each resonator.

The network shown in Fig. 8.35 functions as an admittance inverter. It consists of two transmission lines with *negative* electrical length θ and a series capacitor C_g in the center. For this network a load admittance Y_L connected at one end is transformed into

$$Y'_{\text{in}} = Y_c \frac{Y_L - jY_c t}{Y_c - jY_L t}$$

at the location of $B = \omega C_g$ and where $t = \tan\theta$. Just to the left of C_g, we have

$$Y''_{\text{in}} = \frac{jBY'_{\text{in}}}{jB + Y'_{\text{in}}}$$

and at the input

$$Y_{\text{in}} = Y_c \frac{Y''_{\text{in}} - jY_c t}{Y_c - jY''_{\text{in}} t} = \frac{jY_L Y_c^2\left[B(1-t^2) - Y_c t\right] + Y_c^3 t(2B - Y_c t)}{Y_L(Y_c^2 + 2BY_c t) + jY_c^2\left[B(1-t^2) - Y_c t\right]}$$

We now equate the coefficient of Y_L in the numerator and the constant term in the denominator to zero. Both conditions give

$$\frac{\omega C_g}{Y_c} = \frac{B}{Y_c} = \frac{t}{1-t^2} = \frac{1}{2}\tan 2\theta \qquad (8.120a)$$

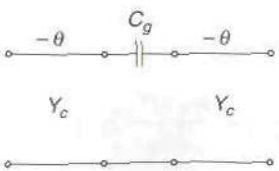


FIGURE 8.35
A simple admittance inverter used in the microstrip half-wave filter.

With this condition imposed

$$Y_{\text{in}} = \frac{Y_c^2 t(2B - Y_c t)}{(Y_c + 2Bt)Y_L} = \frac{Y_c^2 \tan^2 \theta}{Y_L} \tag{8.120b}$$

upon substituting for B. When the relations specified by (8.120) hold, we obtain an admittance inverter with

$$J = Y_c \tan \theta \tag{8.121}$$

The fact that the network in (8.35) involves transmission-line sections with negative electrical lengths does not cause any difficulty since we can add a length θ at each end of the transmission line and then add lengths of electric length $-\theta$ to serve as part of the inverter network. Since both $B = \omega C_g$ and θ are proportional to ω the inverter is not an ideal one. However, for narrowband filters J does not change very much over the passband that the filter must operate.

In the next example we carry out the design of a three-resonator filter of the type we just described.

Example 8.2 Design of a three-section half-wave bandpass filter. We will design a filter with a center frequency of 10 GHz, a bandwidth of 1 GHz, and having a passband tolerance $k^2 = 0.1$ which gives a maximum VSWR equal to 1.863.

We first design the low-pass prototype circuit shown in Fig. 8.36*a*. From (8.101) we readily find that $g_1 = C_1 = g_3 = C_3 = 1.5062$ and $g_2 = L_2 = 1.1151$. For the bandpass-filter prototype network shown in Fig. 8.36*b*, we require $C_1' = C_1/(\omega_2 - \omega_1) = C_3' = (1.5062/2\pi) \times 10^{-9} \approx 239.72$ pF, $L_2' = L_2/(\omega_2 - \omega_1) = (1.1151/2\pi) \times 10^{-9} = 177.47$ pH. By introducing unity inverters we can convert the series tuned resonator to a parallel tuned resonator with $C_2'' = L_2' = 177.47$ pF. We now change these inverters into inverters with J chosen so that $J^2/C_1' = 1/C_2''$ or $J = \sqrt{C_1'/C_2''} = \sqrt{C_1'/L_2'} = 1.16221$. The network in Fig. 8.36*d* is the final one after we scale all impedances by a factor of 50 so that the input and output transmission lines will have a characteristic impedance of 50 Ω instead of 1 Ω. Thus we require $C = (239.72/50)$ pF $= 4.7944$ pF, and $J = 1.16221/50 = 0.023244$. The ratio of the resonator admittance $\sqrt{C/L} = \omega_0 C$ to Y_c of the input and output lines is $2\pi \times 10^{10} \times 4.7944 \times 10^{-12}/0.02 = 15.062$.

If we use 50-Ω transmission-line sections for our resonators, then the ratio of the resonator admittance given by (8.119) to the characteristic admittance of the input and output lines is $\omega_0 \pi \beta_0'/2\beta_0$. If we neglect dispersion, then $\beta_0 = \sqrt{\epsilon_e}\,\omega_0/c$ and $\beta_0'/\beta_0 = 1/\omega_0$, where $c = 3 \times 10^8$ m/s is the free-space velocity of light. In this case the ratio equals $\pi/2$ which is different from the required ratio of 15.062. This problem is overcome by inserting an admittance inverter between the input line and the filter and also one between the output line and the filter. These inverters are chosen so that the line admittance Y_c will appear to be 15.062 times smaller than the

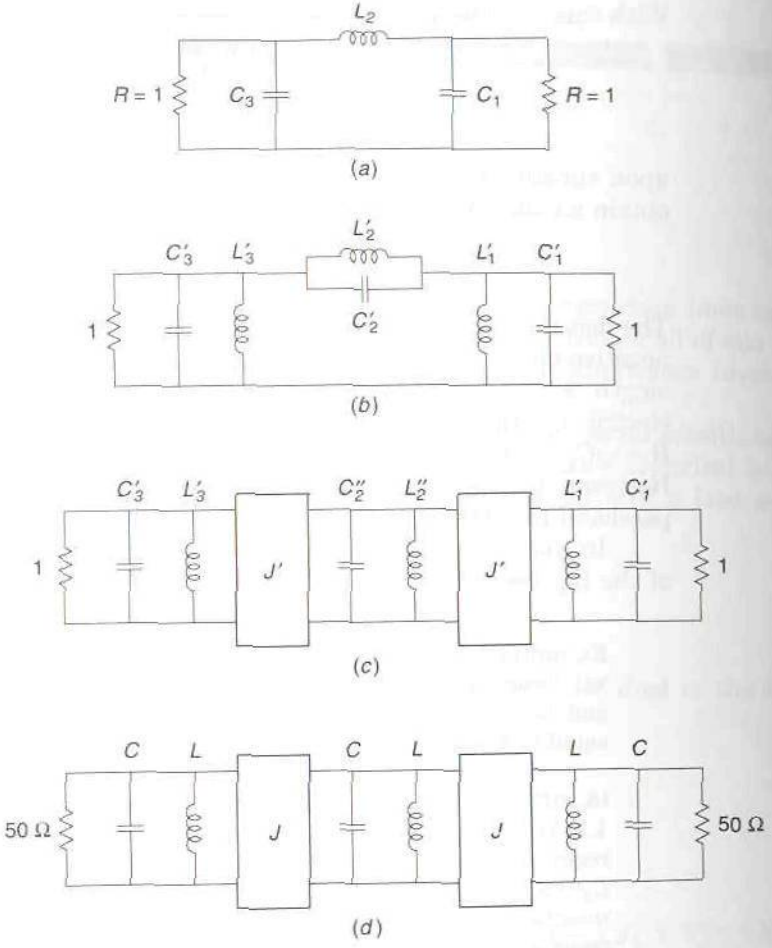


FIGURE 8.36
(*a*) Low-pass prototype circuit; (*b*) bandpass filter; (*c*) transformed bandpass filter; (*d*) final filter network.

resonator admittance $\pi Y_c/2$. Thus we require

$$\frac{J_0^2}{Y_c} = \frac{\pi Y_c}{2 \times 15.062} \qquad \text{or} \qquad J_0 = \sqrt{\frac{\pi}{2 \times 15.062}}\, Y_c = 0.3229 Y_c$$

If dispersion is important then we note that

$$\frac{d\beta}{d\omega} = \frac{d\sqrt{\epsilon_e}\,\omega/c}{d\omega} = \frac{\sqrt{\epsilon_e}}{c} + \frac{\omega}{2\sqrt{\epsilon_e}\,c}\frac{d\epsilon_e}{d\omega}$$

which now gives

$$Y = \frac{\pi}{2}\left(1 + \frac{\omega_0}{\epsilon_e}\frac{d\epsilon_e}{d\omega}\right) Y_c \tag{8.122}$$

in place of (8.119) for the resonator admittance. This expression can be evaluated once the substrate material and microstrip width W have been chosen to give a 50-Ω transmission line. The frequency dependence of ϵ_e can be evaluated using Kobayshi's formula (3.176). For simplicity, we will neglect dispersion effects.

Our last step is to design the required inverters. For the two middle inverters we require the ratio of J to the resonator admittance to be $0.023244/\omega_0 C = 0.023244/(2\pi \times 0.047944) = 0.07716$. Hence we must scale J from the value 0.023244 to $0.023244 \times \pi Y_c/2\omega_0 C$ since we are using resonators with admittances equal to $\pi Y_c/2$. Our new and final value for J is $0.07716\pi Y_c/2 = 0.1212 Y_c$. From (8.121) we find that we require $\theta = \tan^{-1} 0.1212 = 0.1206$. From (8.120$a$) we obtain $B = 0.01 \tan 2\theta = 0.00245$ and hence a gap capacitance C_g of $0.00245/\omega_0 = 0.039$ pF.

For the input and output inverters, $J_0 = 0.323 Y_c = Y_c \tan \theta_0$ so $\theta_0 = 0.3124$ and $B_0 = 0.01 \tan 2\theta_0 = 0.00721$ which gives a gap capacitance C_{g0} of 0.1147 pF.

In this example we determined the inverter parameters from basic principles in order to illustrate again the theory involved. We could have made the evaluations more directly by using the general formulas (8.118). Thus we would have

$$J_{10} = \sqrt{\frac{C_{01}}{C_1 R_{0L} R_L}} = J_{32} = J_0$$

$$J_{21} = \sqrt{\frac{C_{02} C_{01}}{L_2 C_1}} = J$$

where $C_{01} = C_{02} = \pi Y_c/2\omega_0$ is the chosen value for the resonator capacitances and $C_1 = C_1'$ and $L_2 = L_2'$ are the values used in the prototype bandpass-filter circuit. The reader can readily verify that the above expressions give the same results that we found after a series of steps.

The equivalent circuit of a symmetrical microstrip gap (equal width input and output lines) is a symmetrical capacitive Π network shown in Fig. 8.37a. Therefore, when we introduce a gap capacitance C_g for our inverters, we will also introduce a shunt capacitance C_s at the ends of the two coupled lines. These capacitances are equivalent to a small effective lengthening of each line. For the two circuits in Fig. 8.37b, we have

$$Y_{in} = j\omega C_s = jY_c \tan \beta \Delta l \approx jY_c \beta \Delta l$$

and hence the line will appear to be

$$\Delta l = \frac{\omega C_s}{\beta Y_c} \tag{8.123}$$

longer due to the capacitive loading. Thus we need to reduce the physical length of each resonator by an amount proportional to the capacitive loading at each end. For the two middle gaps we will let Δl be the length reduction required and for the two outer lengths we will let the length reduction be denoted by Δl_0.

The designed filter is symmetrical about the midsection. In Fig. 8.38a we show one-half of the filter with the various parameters that we have

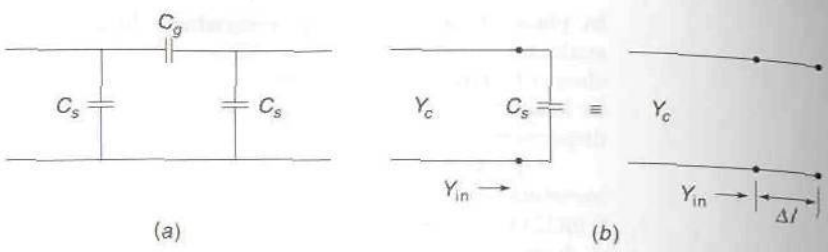


FIGURE 8.37
(*a*) Equivalent circuit of a gap in a microstrip line; (*b*) equivalent transmission-line section used to represent shunt capacitive loading.

determined. At each end of the transmission-line sections, we add and subtract lines of length θ and $-\theta$ or θ_0 and $-\theta_0$. The lines with negative electrical lengths are then part of the required inverter circuits. The additional lengths θ and θ_0 are absorbed as part of the resonators and require that the physical length of each resonator be shorter by the corresponding amounts. We must also account for the line lengthening due to capacitive end loading. The length of the first resonator is chosen so that $\beta_0 l_1 + \beta_0(\Delta l_0 + \Delta l) + \theta_0 + \theta = \pi$; thus

$$l_1 = \frac{\pi - \theta_0 - \theta}{\beta_0} - \Delta l_0 - \Delta l$$

Similarly, we find that the required physical length of the second resonator is

$$l_2 = \frac{\pi - 2\theta}{\beta_0} - 2\,\Delta l$$

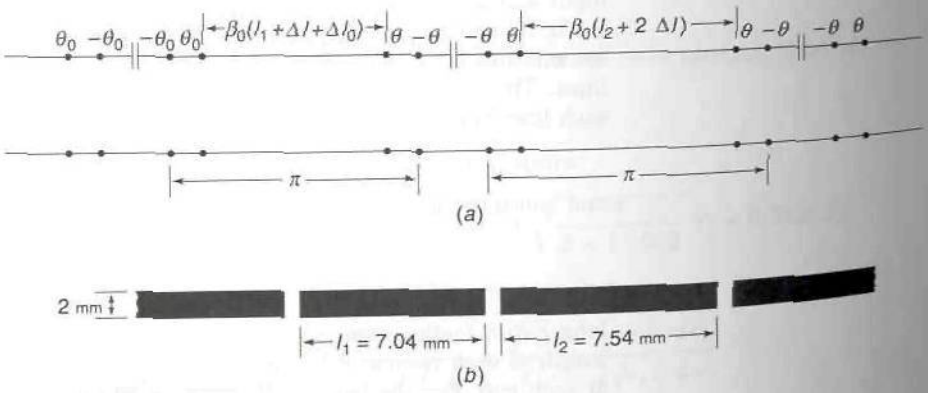


FIGURE 8.38
(*a*) Half-wave filter equivalent circuit; (*b*) metalization pattern for half-wave filter using microstrip construction on a substrate with thickness 0.5 mm and dielectric constant 4.2.

The third resonator has the same length as the first one. We can ignore the effects of Δl_0 and θ_0 at the input and output lines since the length of these lines are arbitrary.

We will complete the design by assuming that the filter will be built using microstrip construction and a substrate 1 mm thick and with a dielectric constant $\epsilon_r = 4.2$. This thickness of substrate at 10 GHz results in negligible dispersion of the characteristic impedance. The required strip width is 2 mm for a 50-Ω line. The low-frequency effective dielectric constant is 3.2. By using Kobayshi's formula we find that $\epsilon_e = 3.37$ at 10 GHz. The dielectric constant is about 5 percent larger at 10 GHz and since $\beta = \sqrt{\epsilon_e} k_0$ this will result in about a 2.5 percent shorter resonator length than what would have been specified if we used the low-frequency value of ϵ_e. If we did not change the resonator lengths, the filter center frequency would have been about 2.5 percent below the design value. Since the filter bandwidth is 10 percent, this amount of detuning is significant so a satisfactory design would not have resulted if we had ignored the dispersion in ϵ_e.

For the two outer gaps we need a capacitance of $0.1147/2 = 0.0557$ pF/mm. For the two inner gaps we need a capacitance of $0.039/2 = 0.0195$ pF/mm. From the data given in Fig. 7.10*c*, we see that the required gap spacings will be very small, to the extent that the shunt capacitance C_s is negligible; so we do not need to adjust the resonator lengths to compensate for C_s. When C_e in Fig. 7.10*c* can be neglected, the gap capacitance $C_g = C_0/2$. For the inner gaps we thus require $C_0 = 0.39$ pF/cm, which is obtained using a gap spacing of about $0.22W = 0.44$ mm. The data in Fig. 7.10*c* do not give values for the capacitance for S/W less than 0.05 for which $C_0 = 0.6$ pF/cm. For the outer gaps we require $C_0 = 1.147$ pF/cm. The data in Fig. 7.10 apply to infinitely thin strips. For very small gaps the strip thickness will be comparable to the spacing, so that the parallel plate capacitance of the end faces must be accounted for. If we use 1-oz copper-clad board, the strip thickness will be approximately 0.036 mm. If the gap spacing were 0.05 mm, the parallel plate capacitance between the two end faces would be

$$C_{pp} = \epsilon_0 \frac{0.036 \times 2 \times 10^{-3}}{0.05} = 0.0127 \text{ pF}$$

for a strip 2 mm wide. This amount of capacitance is about 10 percent of the required amount. Since the data for determining the outer gap spacing are not available, they need to be found either experimentally or numerically before the filter design can be completed.

The resonator lengths can be specified since we can neglect Δl and Δl_0. Thus for the center resonator we require

$$l_2 = \frac{\pi - 2\theta}{\beta} = \frac{\pi - 2 \times 0.1206}{\sqrt{3.37}\, 2\pi/3} = 0.754 \text{ cm}$$

The required length for the two outer resonators is

$$l_1 = \frac{\pi - \theta - \theta_0}{\beta} = \frac{\pi - 0.1206 - 0.3124}{\beta} = 0.704 \text{ cm}$$

The metalization pattern for the filter is shown in Fig. 8.38*b*.

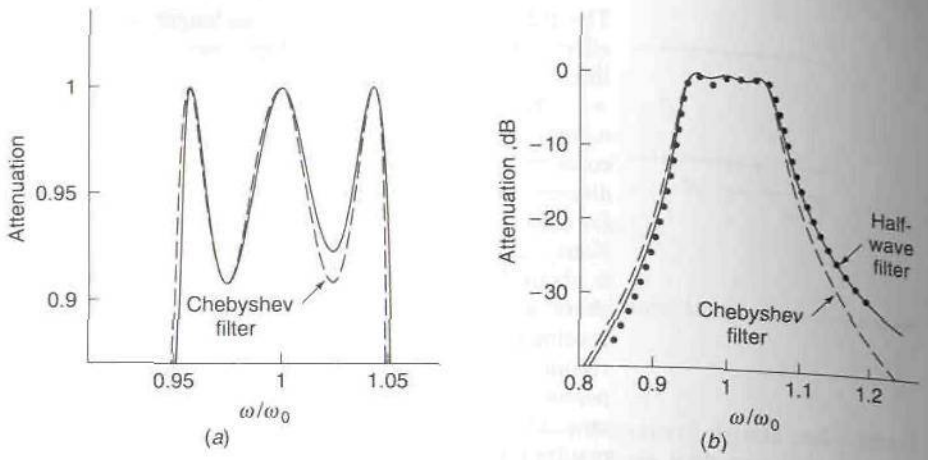


FIGURE 8.39
(*a*) Comparison of attenuation of half-wave filter and ideal Chebyshev filter in the passband; (*b*) comparison of half-wave and ideal Chebyshev filter attenuations as a function of normalized frequency over an extended frequency band. The dotted curve shows the deterioration in the filter response when the two outer gap capacitances are reduced in value by 10 percent.

From the attenuation of the microstrip line, which is approximately 0.0029 Np/cm, we obtain the intrinsic resonator $Q = \beta/2\alpha = 665$. The external Q of the first resonator is $\omega_0 C/G = \omega_0 C Y_c/J_0^2 = (\pi Y_c^2/2)/(0.323 Y_c)^2 = 15.06$. Since this external Q is much smaller than the unloaded resonator Q, the losses in the resonators will have a negligible effect on the performance of the filter.

The calculated insertion loss or attenuation for the filter is shown in Fig. 8.39 as a function of the normalized frequency ω/ω_0. Also shown is the response or attenuation of the ideal Chebyshev filter which has the power loss ratio

$$P_{\text{LR}} = 1 + k^2 T_3^2(\omega') = 1 + k^2(4\omega'^3 - 3\omega')^2$$

where $\omega' = (\omega/\omega_0)(\omega/\omega_0 - \omega_0/\omega) = (\omega^2 - \omega_0^2)/\omega_0^2$. Since $P_{\text{LR}} = 1/(1-\rho^2)$, where ρ is the magnitude of the input reflection coefficient, the power transmission coefficient $1 - \rho^2$ equals $1/P_{\text{LR}}$. The filter insertion loss or attenuation in decibel units is given by $10 \log P_{\text{LR}}$. In Fig. 8.39*a* we show the filter attenuation and that of the ideal Chebyshev filter in the passband. It can be seen that the filter response is very close to that of the ideal Chebyshev filter with the main difference being a somewhat smaller attenuation at $\omega/\omega_0 = 1.025$. In Fig. 8.39*b* the two responses are shown in decibel units over the normalized frequency range 0.8 to 1.25. The response of the half-wave filter is remarkably close to that of the Chebyshev filter with the exception of somewhat greater attenuation below the center frequency and less attenuation above the center frequency (33.8 dB versus 40.9 dB at $\omega/\omega_0 = 1.25$). The half-wave filter was designed on the basis that the admittance function $jY_c \tan \beta l$ for an open-circuited transmission line could be approximated by

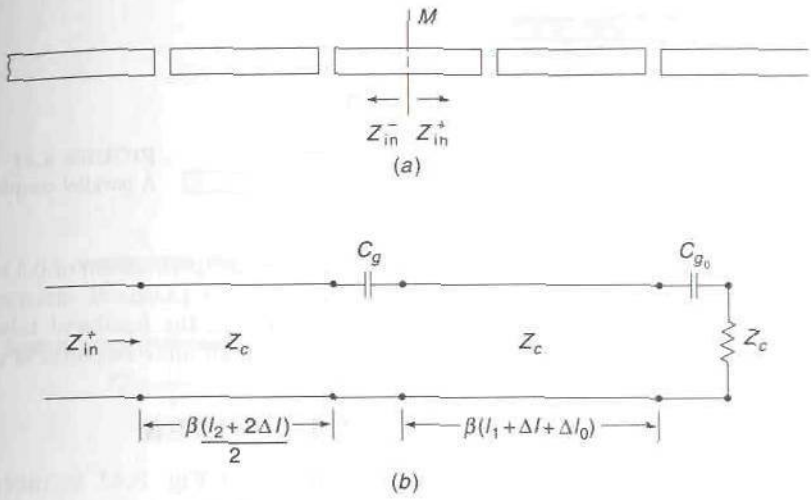


FIGURE 8.40
(*a*) Illustration of point at which the impedance mismatch is evaluated; (*b*) equivalent circuit used to evaluate Z_{in}^+.

$jY_c\pi(\omega - \omega_0)/\omega_0$ in the vicinity of the frequency at which the resonator was one-half wavelength long. Thus it is not surprising that there should be some difference in the attenuation at frequencies that are far from the center frequency. The results shown in Fig. 8.39 do, however, show that the approximations made are acceptable and lead to useful filter designs. If we use $\omega' = (\omega - \omega_0)/\omega_0$ in place of the expression (8.109) in the Chebyshev polynomial, we will obtain a symmetrical response about the point ω_0 and a closer agreement with the response of the half-wave filter. The latter has a symmetrical response curve because $\tan\theta$ is symmetrical about $\theta = 0$.

The impedance mismatch M at the input to the filter is equal to $1 - \rho^2$. In Sec. 5.7 we showed that in a lossless network the impedance mismatch is invariant throughout the network. Hence we can evaluate it at the center of the filter. If we let Z_{in}^+ and Z_{in}^- be the impedance at the center of the filter when looking toward the right and left as shown in Fig. 8.40*a*, then the impedance mismatch is given by

$$M = \frac{4R_{in}^+ R_{in}^-}{|Z_{in}^+ + Z_{in}^-|^2} = \frac{(R_{in}^+)^2}{|Z_{in}^+|^2}$$

since the filter is symmetrical so $Z_{in}^- = Z_{in}^+$. The evaluation of Z_{in}^+ is easily carried out using the circuit model shown in Fig. 8.40*b*. This method was used to evaluate the filter attenuation shown in Fig. 8.39 and is much simpler than evaluating the input impedance for the whole filter network.

The construction of a filter requires good control on the electrical parameters if the theoretical performance is to be met. In Fig. 8.39*b* we show the response of the filter when the two outer gap capacitances are reduced in

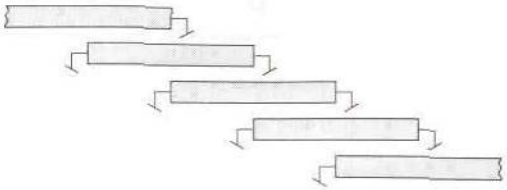


FIGURE 8.41
A parallel coupled microstrip filter.

value by 10 percent from the design specification of 0.1147 to 0.1032 pF. From the figure it can be seen that the passband attenuation at $\omega/\omega_0 = 0.98$ increased to 1.43 dB which exceeds the passband tolerance of 0.414 dB by about 1 dB. Nevertheless, the overall filter response is still quite good.

8.20 MICROSTRIP PARALLEL COUPLED FILTER

The parallel coupled filter shown in Fig. 8.41 is more compact than the half-wave filter described in the previous section. Since the coupling between resonators occurs over a quarter-wave-long side of each resonator, the slot width is larger and the tolerance on the slot width is not as critical. The ends of each resonator section may be open-circuited or short-circuited. The design of parallel coupled filters is readily carried out by using an equivalent circuit of the filter which is easily designed. This equivalent circuit for each pair of coupled resonators is derived below.

Consider the pair of coupled microstrip lines shown in Fig. 8.42*a*. The strips will be considered to have unequal widths. When the voltage applied to each strip is the same (even mode) as shown in Fig. 8.42*b*, the currents on the two strips will not be the same because of the different widths. We will let the currents be $I_1 = Y_e^a V$ and $I_2 = Y_e^b V$, where Y_e^a and Y_e^b represent the characteristic admittance of strip a and strip b relative to the ground plane. For the odd mode with a voltage V applied to strip a and $-V$ applied to strip b, the currents are given by $I_1 = Y_o^a V$ and $I_2 = -Y_o^b V$, where Y_o^a and Y_o^b are the characteristic admittances of the two strips for the odd mode. We will show that the coupled-line circuit shown in Fig. 8.42*a* is equivalent to the circuit shown in Fig. 8.42*c*, where

$$Y_1 = \tfrac{1}{2}\sqrt{(Y_o^a - Y_e^a)(Y_o^b - Y_e^b)} \tag{8.124a}$$

$$Y_2 = \tfrac{1}{2}(Y_e^a + Y_o^a) - Y_1 \tag{8.124b}$$

$$Y_3 = \tfrac{1}{2}(Y_e^b + Y_o^b) - Y_1 \tag{8.124c}$$

The equivalence will be established by showing that the input admittance at port 1 with port 2 terminated in an open circuit or a short circuit is the same for both circuits.

The voltage and current waves on the circuit shown in Fig. 8.42*a* can be expressed in terms of a superposition of odd and even modes with voltages V_o and V_e. Thus on strip a we express the voltage and current

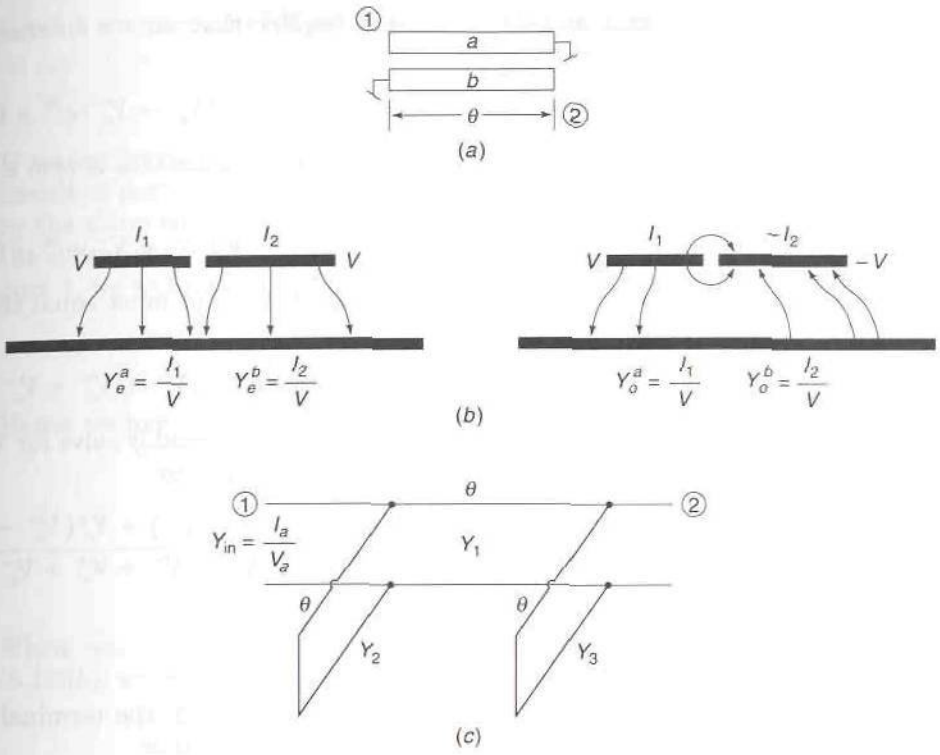


FIGURE 8.42
(*a*) A pair of coupled microstrip lines; (*b*) illustration of even and odd modes; (*c*) equivalent circuit of coupled strips shown in (*a*).

waves in the form

$$V_a(z) = V_e^+ e^{-j\beta z} + V_e^- e^{j\beta z} + V_o^+ e^{-j\beta z} + V_o^- e^{j\beta z}$$

$$I_a(z) = Y_e^a V_e^+ e^{-j\beta z} - Y_e^a V_e^- e^{j\beta z} + Y_o^a V_o^+ e^{-j\beta z} - Y_o^a V_o^- e^{j\beta z}$$

It is assumed that both modes have the same propagation constants. On strip b the voltage and current waves are

$$V_b(z) = V_e^+ e^{-j\beta z} + V_e^- e^{j\beta z} - V_o^+ e^{-j\beta z} - V_o^- e^{j\beta z}$$

$$I_b(z) = Y_e^b V_e^+ e^{-j\beta z} - Y_e^b V_e^- e^{j\beta z} - Y_o^b V_o^+ e^{-j\beta z} + Y_o^b V_o^- e^{j\beta z}$$

When $z = 0$ we must have $V_b(0) = 0$; thus

$$(V_e^+ + V_e^-) - (V_o^+ + V_o^-) = 0 \qquad (8.125a)$$

since the end at $z = 0$ on strip b is short-circuited. If we place a short

circuit at port 2, we will require that at $z = l$, where $\beta l = \theta$, $V_b(l) = 0$ which gives

$$(V_e^+ - V_o^+)e^{-j\theta} + (V_e^- - V_o^-)e^{j\theta} = 0 \tag{8.125b}$$

On strip a we must have $V_a(l) = 0$ since this is also a short-circuited end; thus

$$(V_e^+ + V_o^+)e^{-j\theta} + (V_e^- + V_o^-)e^{j\theta} = 0 \tag{8.125c}$$

The last terminal condition is that $V_a(0)$ must equal the applied voltage at port 1 which is V_1. Hence

$$V_1 = V_a(0) = V_e^+ + V_e^- + V_o^+ + V_o^- \tag{8.125d}$$

From the above four equations we can readily solve for V_e^+, V_e^-, V_o^+, and V_o^-. The input admittance at port 1 is given by

$$Y_{\text{in, sc}} = \frac{I_1}{V_1} = \frac{Y_e^a(V_e^+ - V_e^-) + Y_o^a(V_o^+ - V_o^-)}{V_e^+ + V_e^- + V_o^+ + V_o^-}$$

$$= -j\frac{Y_o^a + Y_e^a}{2}\cot\theta \tag{8.126}$$

When we place an open circuit at port 2, the terminal condition $V_b(l) = 0$ given by (8.125b) is replaced by $I_b(l) = 0$ or

$$Y_e^b(V_e^+e^{-j\theta} - V_e^-e^{j\theta}) - Y_o^b(V_o^+e^{-j\theta} - V_o^-e^{j\theta}) = 0$$

The other terminal conditions remain the same. After solving for the voltage amplitudes for this case and using (8.126), we find that

$$Y_{\text{in, oc}} = j\frac{(Y_e^a - Y_o^a)(Y_e^b - Y_o^b)}{2(Y_e^b + Y_o^b)}\tan\theta$$

$$- j\frac{(Y_e^aY_o^b + Y_o^aY_e^b)}{Y_e^b + Y_o^b}\cot\theta \tag{8.127}$$

For the network shown in Fig. 8.42c, a straightforward evaluation shows that when port 2 is short-circuited

$$Y_{\text{in, sc}} = -j(Y_1 + Y_2)\cot\theta \tag{8.128a}$$

and when port 2 is open-circuited

$$Y_{\text{in, oc}} = j\frac{Y_1^2}{Y_1 + Y_3}\tan\theta - j\frac{Y_1Y_2 + Y_2Y_3 + Y_1Y_3}{Y_1 + Y_3}\cot\theta \tag{8.128b}$$

From a comparison of (8.126) and (8.128a), we see that

$$Y_1 + Y_2 = \frac{Y_o^a + Y_e^a}{2}$$

If we excite the networks at port 2 and place a short circuit or an open circuit at port 1, the expressions for $Y_{\text{in, sc}}$ and $Y_{\text{in, oc}}$ are the same as given by the above equations but with the superscripts a and b interchanged and the subscripts 1 and 3 interchanged. Thus, for the short-circuit condition at port 1, we will find that

$$Y_1 + Y_3 = \frac{Y_o^b + Y_e^b}{2}$$

Hence we have

$$Y_2 = \frac{Y_o^a + Y_e^a}{2} - Y_1$$

$$Y_3 = \frac{Y_o^b + Y_e^b}{2} - Y_1$$

When we use the expression for $Y_1 + Y_3$ in the coefficient for $\tan\theta$ in (8.128b), we easily find that

$$Y_1^2 = \tfrac{1}{2}\sqrt{(Y_o^a - Y_e^a)(Y_o^b - Y_e^b)}$$

Thus we have shown that the circuit in Fig. 8.42a is equivalent to that in Fig. 8.42c. For the special case when the coupled strips have the same width, $Y_e^b = Y_e^a = Y_e$ and $Y_o^b = Y_o^a = Y_o$ and we obtain

$$Y_1 = \frac{Y_o - Y_e}{2} \qquad (8.129a)$$

$$Y_2 = Y_3 = Y_e \qquad (8.129b)$$

Consider now the filter structure shown in Fig. 8.43a. By replacing each coupled strip pair by its equivalent circuit in the form shown in Fig. 8.42c, we arrive at the circuit shown in Fig. 8.43b. This circuit is readily reduced to the one shown in Fig. 8.43c by combining the admittances of the stubs that are connected in parallel. A Chebyshev filter based on the circuit shown in Fig. 8.43c is readily designed.

For parallel coupled filters using microstrip construction, it is desirable to use open-circuited coupled microstrip sections in place of short-circuited coupled sections. A parallel coupled filter using open-circuited coupled transmission-line sections is shown in Fig. 8.44a. The equivalent circuit for this filter is shown in Fig. 8.44b. A basic coupled section is shown in Fig. 8.44c and its equivalent circuit is shown in Fig. 8.44d.

The equivalent circuit for the basic section is readily obtained from the set of equations (8.124) through (8.129). If we regard all voltage variables as representing currents, current variables as representing voltages, admittances as representing impedances, and open circuits/short circuits as representing short circuits/open circuits, then all the equations and terminal conditions remain the same. Thus we obtain immediately

$$Z_1 = \tfrac{1}{2}\sqrt{(Z_e^a - Z_o^a)(Z_e^b - Z_o^b)} \tag{8.130a}$$

$$Z_2 = \tfrac{1}{2}(Z_e^a + Z_o^a) - Z_1 \tag{8.130b}$$

$$Z_3 = \tfrac{1}{2}(Z_e^b + Z_o^a) - Z_1 \tag{8.130c}$$

and when $Z_e^a = Z_e^b$, $Z_o^a = Z_o^b$,

$$Z_1 = \frac{Z_e - Z_o}{2} \tag{8.131a}$$

$$Z_2 = Z_3 = Z_o \tag{8.131b}$$

The equivalent circuit shown in Fig. 8.44*b* for the filter is obtained by replacing each coupled section by its equivalent circuit shown in Fig. 8.44*d* and combining the adjacent series-connected transmission-line stubs. This filter and its equivalent network is the dual of the one shown in Figs. 8.42 and 8.43. In extracting the square root of (8.130*a*) for the case of symmetrical strips, we choose the positive root which is the one given by (8.131*a*) since $Z_e > Z_o$.

For both of the parallel coupled filters described, each section is $\lambda/4$ long at the center frequency and hence each transmission line is $\lambda/2$ long. In the equivalent circuits of these filters, we have transmission-line stubs separated by $\lambda/4$ transformers. These quarter-wave transformers function as nonideal admittance and impedance transformers. In the circuit shown in Fig. 8.43*c*, each short-circuited stub is the approximate equivalent of a parallel tuned *LC* resonator, while in Fig. 8.44*b* each open-circuited stub is the approximate equivalent of a series tuned *LC* resonator.

Equations for designing Chebyshev bandpass filters, with bandwidths up to one octave, using parallel coupled strips have been derived by Matthaei.† The filter design equations were obtained by setting the image impedance and phase of each section of the network shown in Fig. 8.43*c* equal to that for a bandpass-filter prototype circuit of the form shown in Fig. 8.29 at the center frequency ω_0 and at the passband edge where $\omega = \omega_1$. The design equations for the filter circuit shown in Fig. 8.44*b* are the dual of those for the filter circuit in Fig. 8.43*c*. We will give the design equations below without derivation.

†G. L. Matthaei, Design of Wide-Band (and Narrow-Band) Band-Pass Microwave Filters on the Insertion Loss Basis, *IRE Trans.*, vol. MTT-8, pp. 580–593, 1960.

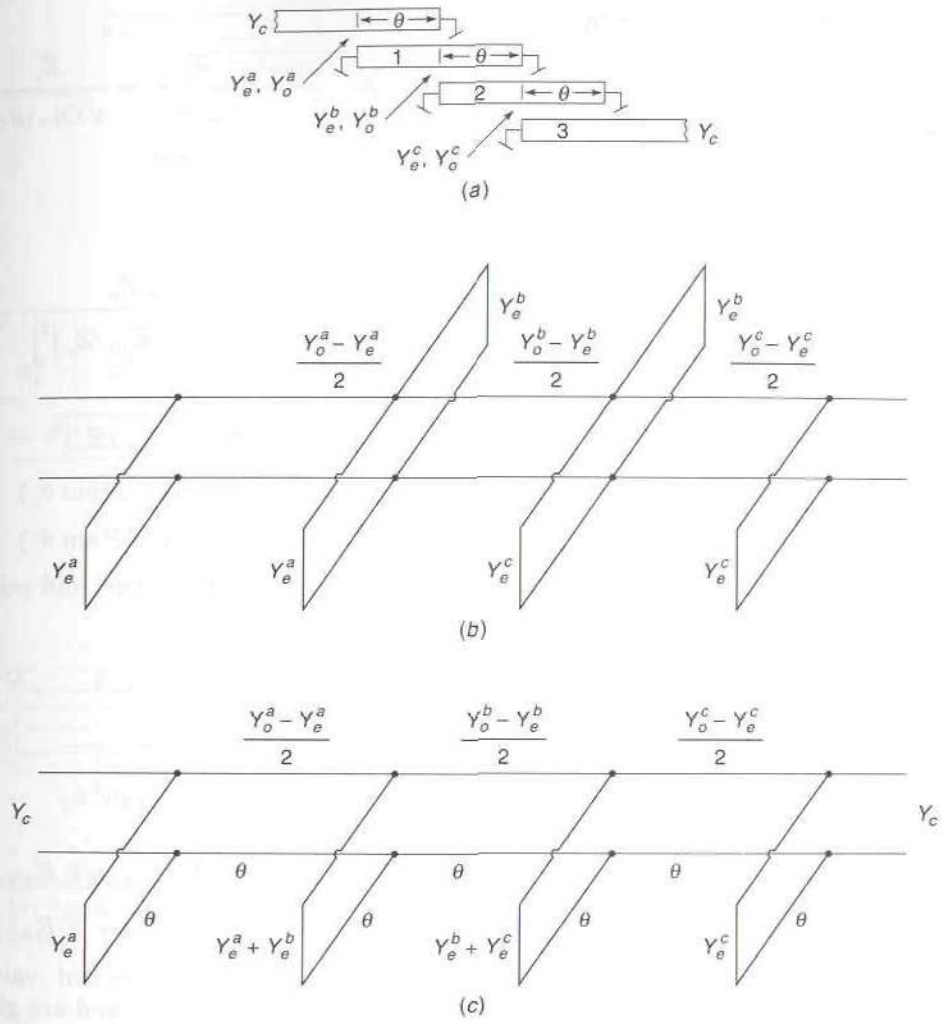


FIGURE 8.43
(*a*) A parallel coupled filter; (*b*) equivalent circuit; (*c*) reduced equivalent circuit.

For a Chebyshev filter with N sections, there are $N + 1$ impedance inverters and $N + 1$ even- and odd-mode line impedances to specify. The filter is assumed to be terminated in input and output lines with characteristic impedance Z_c. Each resonator has an electrical length of π at the center frequency ω_0. The frequency at the lower edge of the passband is ω_1 and $\beta l_e = \theta_1$ at this frequency, where l_e is the effective length of each resonator after correcting for the capacitive end loading at each open-circuited end. The impedances of the input and output impedance inverters are

given by

$$K_{10} = K_{N+1,N} = \frac{Z_c}{\sqrt{g_0 g_1}} = \frac{Z_c}{\sqrt{g_{N+1} g_N}} \tag{8.132a}$$

The following parameters are also required:

$$\theta_1 = \frac{\pi \omega_1}{2\omega_0} \tag{8.132b}$$

$$P \sin \theta_1 = \frac{K_{10}/Z_c}{\left[\frac{1}{2}\tan\theta_1 + (K_{10}/Z_c)^2\right]^{1/2}} \tag{8.132c}$$

$$s = \frac{Z_c}{\frac{1}{2}\tan\theta_1 + (K_{10}/Z_c)^2} \tag{8.132d}$$

$$Z_e^1 = Z_e^{N+1} = Z_c(1 + P \sin\theta_1) \tag{8.132e}$$

$$Z_o^1 = Z_o^{N+1} = Z_c(1 - P \sin\theta_1) \tag{8.132f}$$

The remaining impedance inverters and even- and odd-mode impedances are given by

$$K_{k+1,k} = \frac{Z_c}{\sqrt{g_k g_{k+1}}} \qquad k = 1, 2, \ldots, N-1 \tag{8.133a}$$

$$\hat{N}_{k+1,k} = \sqrt{\left(\frac{K_{k+1,k}}{Z_c}\right)^2 + \tfrac{1}{4}\tan^2\theta_1} \tag{8.133b}$$

$$Z_e^{k+1} = Z_e^{N-k+1} = s\left(\hat{N}_{k+1,k} + Y_c K_{k+1,k}\right) \tag{8.133c}$$

$$Z_o^{k+1} = Z_o^{N-k+1} = s\left(\hat{N}_{k+1,k} - Y_c K_{k+1,k}\right) \tag{8.133d}$$

In the above equations the g_k are the element values for a loss-pass prototype filter with cutoff frequency $\omega_c = 1$, and are given by (8.101).

For the dual network all $K_{k+1,k}$ are replaced by $J_{k+1,k}$, all Z_e^k are replaced by Y_o^k, and all Z_o^k are replaced by Y_e^k.

In the next example we will illustrate the application of the above equations to the design of a two-section Chebyshev filter.

Example 8.3 Parallel coupled Chebyshev bandpass filter. The filter will be designed for a 30 percent bandwidth, a passband tolerance $k^2 = 0.5$, and a center frequency of 6 GHz. The input and output lines have a characteristic impedance of 50Ω. The filter is illustrated in Fig. 8.45*a*.

The element values for the low-pass prototype circuit are calculated using (8.101) and are $g_0 = 1$, $g_1 = 2.33754$, $g_2 = 0.62634$, and $g_3 = 1/R = 3.7321$. The frequency at the lower band edge is $f_0 - 0.15 f_0 = 0.85 f_0$. The value of θ_1 is $0.85\pi/2$ and $\sin\theta_1 = 0.97237$, $\tan\theta_1 = 4.165$. By using (8.132) and (8.133) we obtain $K_{10} = 0.65406 Z_c$, $K_{21} = 0.826447 Z_c$, $P \sin\theta_1 = 0.4128$,

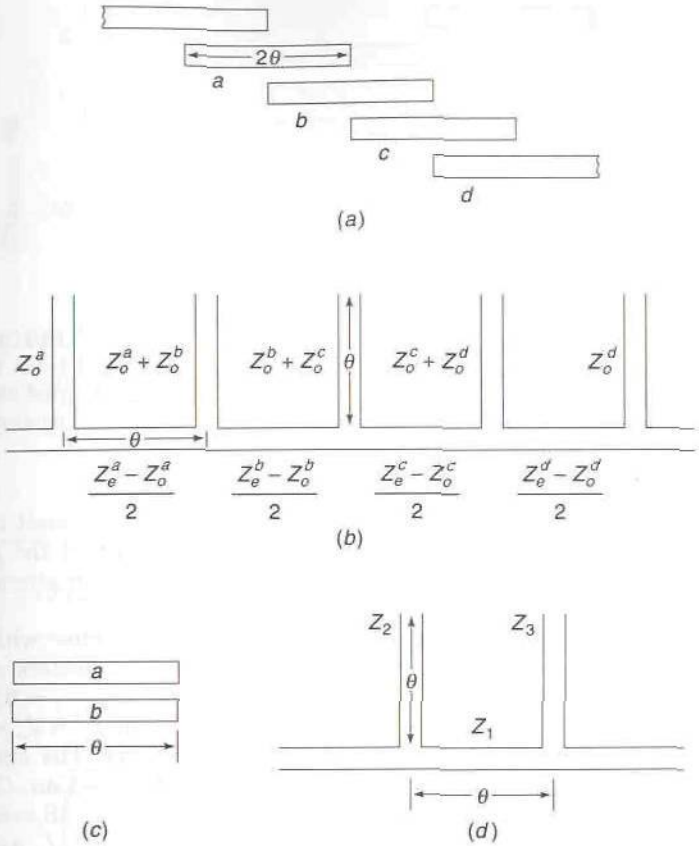


FIGURE 8.44
(*a*) Parallel coupled filter using open-circuited coupled-line sections; (*b*) equivalent circuit of filter; (*c*) basic coupled-line section; (*d*) equivalent circuit of basic section.

$s = 19.9165$, and $\hat{N}_{12} = 2.8865$. The following even- and odd-mode impedances are found

$$Z_e^1 = 70.64 \qquad Z_o^1 = 29.36 \qquad Z_e^2 = 73.95 \qquad Z_o^2 = 41.03$$

In the equivalent transmission-line circuit shown in Fig. 8.45*b*, the line impedances are found to be $Z_1 = Z_o^1$, $Z_2 = (Z_e^1 - Z_o^1)/2$, $Z_3 = Z_o^1 + Z_o^2$, $Z_4 = (Z_e^2 - Z_o^2)/2$, $Z_5 = Z_3$, $Z_6 = Z_2$, and $Z_7 = Z_1$. For this filter it is a straightforward transmission-line circuit analysis problem to evaluate the impedances $Z_{\text{in}}^- = Z_{\text{in}}^+$ at the center of the filter and evaluate the impedance mismatch $(R_{\text{in}}^-)^2/|Z_{\text{in}}^-|^2$. In Fig. 8.46 we show the computed filter response with that of the Chebyshev filter with power loss ratio $1 + k^2 T_2^2(\omega')$ and using $\omega' = (\omega - \omega_0)/\omega_0$ as the frequency variable.

The design equations are approximate ones only. Since we are specifying a 30 percent bandwidth for a filter using only two sections, it is not surprising

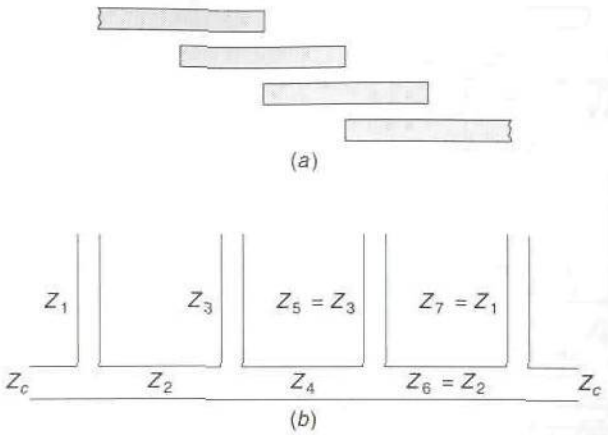


FIGURE 8.45
(*a*) A two-section parallel coupled filter; (*b*) equivalent transmission-line circuit.

that the actual filter performance does not quite meet the specifications. The response curve shown in Fig. 8.46 indicates that the actual filter has a 25 percent bandwidth but a more rapid increase in attenuation than the ideal Chebyshev filter has.

We have also designed a three-section filter with the same passband tolerance and bandwidth. The calculated parameters are $g_1 = g_3 = 2.5547$, $K_{10} = K_{43} = 0.62565$, $K_{21} = K_{32} = 0.672$, $P \sin \theta_1 = 0.39776$, $s = 20.21$, $Z_e^1 = 69.89$, $Z_o^1 = 30.11$, $Z_e^2 = Z_e^3 = 57.807$, and $Z_o^2 = Z_o^3 = 30.645$. In Fig. 8.47 we show the computed response of this filter. The agreement is now much closer to that of the corresponding three-section Chebyshev filter. The difference in the attenuation does not exceed 0.3 dB over the frequency range $0.7f_0$ to $1.3f_0$. At the band edges where $f = 0.85f_0$ and $1.15f_0$, the parallel coupled filter has an attenuation of 2 dB instead of the theoretical design value of 1.761 dB, which is a difference of only 0.24 dB.

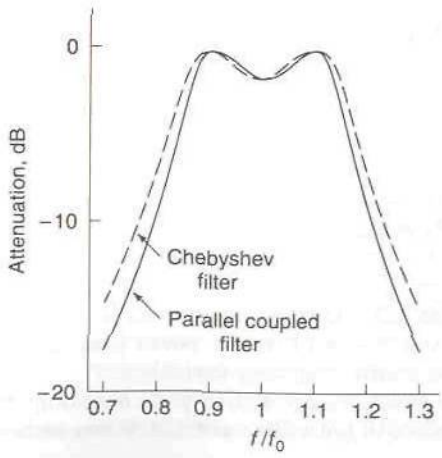


FIGURE 8.46
A comparison of the response of the two-section parallel coupled filter with that of the ideal Chebyshev filter. The passband tolerance is 1.76 dB and the fractional bandwidth equals 0.3.

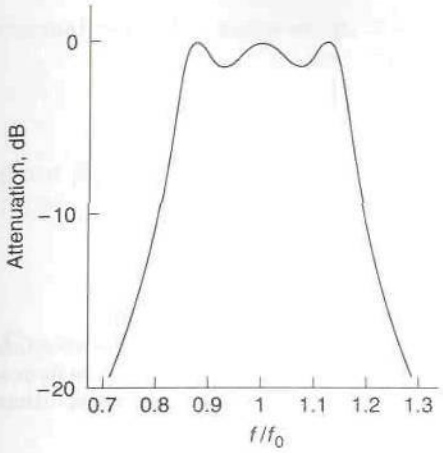


FIGURE 8.47
Response of a three-section parallel coupled filter. The passband tolerance is 1.76 dB and the fractional bandwidth is 0.3.

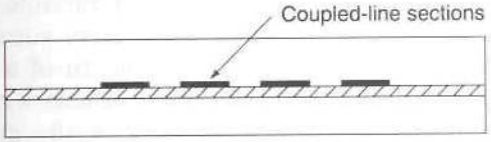


FIGURE 8.48
Shielded suspended microstrip line used for constructing parallel coupled filters.

The design formulas for parallel coupled filters are based on the assumption that the even- and odd-mode phase velocities are equal. Thus, if conventional microstrip-line construction is used, it would be necessary to use a dielectric overlay so as to equalize the two phase velocities. The parallel coupled filter can also be built using strip-line or shielded suspended microstrip line as shown in Fig. 8.48. The suspended microstrip line employs a thin low dielectric constant sheet on which the transmission-line sections are printed. Since the dielectric sheet is thin, it has about the same effect on the phase velocity of both the even and odd modes.

For parallel coupled strips a ratio Z_e/Z_o less than 3 is needed in order to avoid very close spacing between the strips. For the filters discussed in Example 8.3, the largest ratio was 2.4, so that these filters can be built using edge-coupled strips. When tighter coupling is required, the adjacent strips can be overlapped by printing every other transmission-line section on the opposite sides of the supporting dielectric sheet.

8.21 QUARTER-WAVE-COUPLED CAVITY FILTERS

Quarter-wave-coupled cavity filters are similar to the filter discussed in the preceding section except that the transmission-line stubs are replaced by cavities. The filter is realized in practice by placing diaphragms in a waveguide. To understand the basis for design in the narrowband case, we must first consider the equivalent circuit of a section of waveguide loaded with two identical diaphragms a distance l_k apart, as shown in Fig. 8.49.

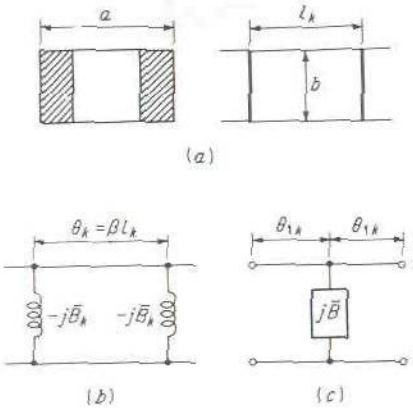


FIGURE 8.49
(*a*) Rectangular waveguide loaded with two inductive diaphragms to form a cavity; (*b*) exact equivalent circuit; (*c*) approximate equivalent circuit.

For a waveguide, the important frequency variable is not ω but rather $(\beta/k_0)\omega = \beta c$, since waveguide diaphragms have susceptances that vary very nearly as β or β^{-1} and the electrical length of a section of guide is proportional to β. The normalized frequency variable $\omega/\omega_0 = \lambda_0/\lambda$ is therefore replaced by $\lambda_{g0}/\lambda_g = \beta/\beta_0$, where λ_{g0} is the guide wavelength at $\omega = \omega_0$ and λ_g is the corresponding value at any ω. Consequently, in all design formulas, we replace ω by βc, where c is the velocity of light.

The exact equivalent circuit for a waveguide loaded with two identical inductive diaphragms with normalized susceptance $-j\overline{B}_k$ is shown in Fig. 8.49*b*. For filter design according to the methods developed in preceding sections, we must replace the exact equivalent circuit by an approximate shunt circuit. Mumford has shown that an equivalent circuit of the form illustrated in Fig. 8.49*c* has the same frequency characteristics as the exact equivalent circuit has over a narrow band of frequencies around ω_0.† The results obtained by Mumford are presented here without derivation. The derivation is straightforward, and may be found in Mumford's paper. The shunt susceptance $\overline{B}$ is expressed in the form

$$\overline{B} = \sqrt{\frac{C}{L}}\left(\frac{\beta}{\beta_0} - \frac{\beta_0}{\beta}\right) \approx 2\sqrt{\frac{C}{L}}\,\frac{\Delta\beta}{\beta_0}$$

where $\Delta\beta = \beta - \beta_0$ is small. When a resonant circuit of this type is connected across a transmission line, it is loaded by a shunt conductance of

†W. W. Mumford, Maximally Flat Filters in Waveguides, *Bell System Tech. J.*, vol. 27, pp. 684–714, October 1948.

normalized value unity on each side. The loaded Q of the circuit is thus

$$Q_k = \frac{1}{2}(\beta_0 c)C = \frac{1}{2}\sqrt{\frac{C}{L}} \tag{8.134}$$

since $\beta_0 c = (LC)^{-1/2}$. Hence we may express $\bar{B}$ in terms of the loaded Q; thus

$$\bar{B} = 4Q_k \frac{\Delta\beta}{\beta_0} \tag{8.135}$$

The value obtained for Q_k by Mumford is

$$Q_k = \frac{\pi - \tan^{-1}(2/\bar{B}_k)}{2 \sin^{-1} \dfrac{2}{\left(\bar{B}_k^4 + 4\bar{B}_k^2\right)^{1/2}}}$$

$$\approx \frac{\left(\bar{B}_k^4 + 4\bar{B}_k^2\right)^{1/2}}{4} \tan^{-1} \frac{2}{\bar{B}_k} \tag{8.136}$$

since $\bar{B}_k$ is large compared with unity for a narrowband (high-Q) filter. The required diaphragm spacing l_k to give perfect transmission through the cavity at $\omega = \omega_0$ is given by

$$\tan \beta_0 l_k = -\frac{2}{\bar{B}_k} \tag{8.137}$$

The two sections of line with electrical length θ_{1k} in the circuit of Fig. 8.49*c* are to be chosen so that

$$\beta_0 l_k + 2\theta_{1k} = \theta_k + 2\theta_{1k} = \pi \tag{8.138}$$

at the frequency ω_0. These additional lengths of line in the equivalent circuit of a single cavity are absorbed into and made part of the quarter-wave coupling lines in the filter.

The design of maximally flat and Chebyshev filters with N odd is straightforward. If the prototype circuit of Fig. 8.29 is used, it is only necessary to make

$$Q_k = \frac{1}{2}\sqrt{\frac{C_{0k}}{L_{0k}}} \tag{8.139}$$

and to choose C_{0k}, L_{0k} so that $C_{0k}L_{0k} = (\beta_0 c)^{-2}$ and all $J_{k,k-1}$ equal unity. The section of waveguide between cavity k and $k + 1$ has an electrical length equal to $\pi/2$. Since this includes θ_{1k+1} and θ_{1k} from the adjacent cavities, the physical length of the quarter-wave coupling line between

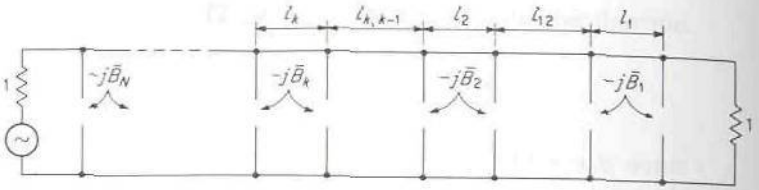


FIGURE 8.50
Quarter-wave-coupled waveguide-cavity filter.

cavities k and $k+1$ will be

$$l_{k,k+1} = \frac{1}{\beta_0}\left(\frac{\pi}{2} - \theta_{1k} - \theta_{1k+1}\right)$$

$$= \frac{\lambda_{g0}}{2\pi}\left(\frac{\theta_k + \theta_{k+1}}{2} - \frac{\pi}{2}\right) = \frac{l_k + l_{k+1}}{2} - \frac{\lambda_{g0}}{4} \tag{8.140}$$

upon using (8.138). A schematic illustration of the filter is given in Fig. 8.50. Formulas for the required diaphragm dimensions to yield the specified value of $\bar{B}_k$ are given in Sec. 5.8. The power loss ratio for the filter is obtained by replacing ω/ω_0 by β/β_0. For a Chebyshev filter it is given by [see (8.104) and (8.109)]

$$P_{\text{LR}} = 1 + k^2 T_N^2\left[\frac{\beta_0}{\beta_2 - \beta_1}\left(\frac{\beta}{\beta_0} - \frac{\beta_0}{\beta}\right)\right] \tag{8.141}$$

where β_2 and β_1 are the values of β at the edges of the passband. If β_1 and β_2 are specified, then

$$\beta_0 = \sqrt{\beta_1 \beta_2} \tag{8.142}$$

It should also be noted that for the waveguide filter, ω_0, ω_1, and ω_2 in the design formulas (8.111) and (8.112) must be replaced by $\beta_0 c$, $\beta_1 c$, and $\beta_2 c$, where c is the velocity of light; i.e., replace $k_0 = \omega/c$ by β. To illustrate the procedure, we shall evaluate the required susceptance $-j\bar{B}_1$ for the first cavity in the filter shown in Fig. 8.50.

Assume that a five-element filter will be needed. The response is to be of the Chebyshev type with a passband tolerance $k^2 = 0.0233$. The waveguide to be used has a width of $a = 0.9$ in. The passband is to extend from $f_1 = 10{,}000$ MHz to $f_2 = 10{,}400$ MHz. The corresponding values of $k_0 = \omega/c$ are 2.1 and 2.18 rad/cm. The values of β_1 and β_2 are thus

$$\beta_1 = \left[(2.1)^2 - \left(\frac{\pi}{a}\right)^2\right]^{1/2} = (4.4 - 1.89)^{1/2} = 1.59$$

$$\beta_2 = \left[(2.18)^2 - 1.89\right]^{1/2} = 1.7$$

The center of the band occurs at $\beta_0 = (\beta_1\beta_2)^{1/2} = 1.64$, which gives $f_0 = 10{,}200$ MHz. From Table 8.2 we find $g_1 = 1.1468$. Using (8.112b) gives

$$C_1' = \frac{\beta_0}{\beta_2 - \beta_1}\frac{C_1}{\beta_0 c} = \frac{\beta_0}{\beta_2 - \beta_1}\frac{g_1}{\beta_0 c}$$

Since J_{10} is to equal unity and $R_L = R_{0L} = 1$, we have $C_{01} = C_1'$ from (8.118a). Using (8.139), we obtain

$$Q_1 = \frac{1}{2}\sqrt{\frac{C_{01}}{L_{01}}} = \frac{1}{2}(\beta_0 c)C_{01} = \frac{\beta_0}{\beta_2 - \beta_1}\frac{g_1}{2}$$

For the kth resonator we should obtain

$$Q_k = \frac{\beta_0}{\beta_2 - \beta_1}\frac{g_k}{2} \tag{8.143}$$

For Q_1 we obtain 8.56, and from this result we can determine $\overline{B}_1$ by using (8.136). For $\overline{B}_k$ large, we can replace $\tan^{-1}(2/\overline{B}_k)$ by $2/\overline{B}_k$, and we then find that

$$\overline{B}_k = 2\left(Q_k^2 - 1\right)^{1/2} \tag{8.144}$$

Thus we find that $\overline{B}_1 = 17$. The required diaphragm dimensions can now be determined, and also the cavity length from (8.137). The above procedure has to be repeated for each cavity in the filter chain.

8.22 DIRECT-COUPLED CAVITY FILTERS

Direct-coupled cavity filters have the advantage that the physical structure is more compact than the corresponding quarter-wave-coupled cavity filter. A design procedure for direct-coupled cavity filters that is accurate for bandwidths up to 20 percent has been developed by Cohn.† Cohn's design method is based on the use of the network in Fig. 8.28 as a prototype. The design formulas will be presented here without the detailed derivation.

The waveguide cavity and its equivalent circuit shown in Figs. 8.49a and b may also be represented by a Π network shunted with inductive susceptances at each end, as in Fig. 8.51. The two shunt susceptances $\overline{B} = -\cot(\theta_k/2)$ may be neglected compared with $\overline{B}_k$ since $\overline{B}_k$ will be large, and θ_k is nearly equal to π; so $\overline{B}$ is small compared with unity. The series arm $\overline{X}$ is thus used as the series resonant circuit in the prototype filter.

For impedance inverters Cohn uses the shunt inductive reactance plus two short sections of waveguide (equivalent transmission lines), as shown in

†S. B. Cohn, Direct Coupled Resonator Filters, *Proc. IRE*, vol. 45, pp. 187–196, February, 1957.

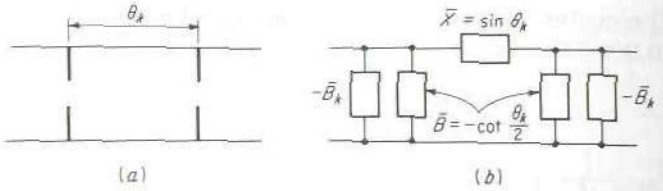


FIGURE 8.51
A waveguide cavity and its equivalent circuit.

Fig. 8.52. For this circuit the impedance inverting properties are obtained if

$$\theta_{1k} = -\frac{1}{2}\tan^{-1}\frac{2}{\bar{B}_k} \tag{8.145a}$$

$$\bar{B}_k = \frac{1 - K^2}{K} \tag{8.145b}$$

where K is the characteristic impedance of the quarter-wave impedance inverter. With θ_{1k} and $\bar{B}_k$ determined at a frequency ω_0, it is found that the inverter does not depart appreciably from its ideal characteristics over a 20 percent band.

In the vicinity of $\omega = \omega_0$, where $\theta_k = \pi$, the series reactance $\bar{X}$ behaves as

$$\bar{X} = \sin\theta_k = \sin(\theta_k - \pi + \pi)$$

$$\approx -(\theta_k - \pi) = -(\beta - \beta_0)l = \frac{\beta_0 - \beta}{\beta_0}\pi \approx -\frac{\pi}{2}\left(\frac{\beta}{\beta_0} - \frac{\beta_0}{\beta}\right)$$

where $\beta_0 l = \pi$. This frequency behavior is similar (apart from the sign, which is immaterial) to that for a series resonant circuit for which $X = \sqrt{L/C}(\omega/\omega_0 - \omega_0/\omega) \approx 2\sqrt{L/C}(\omega - \omega_0)/\omega_0$ if ω/ω_0 is replaced by the new frequency variable β/β_0.

When the negative line lengths of the impedance inverters are absorbed as part of the cavity length, the physical length of the kth cavity becomes

$$l_k = \frac{\lambda_{g0}}{2} + \frac{\lambda_{g0}}{2\pi}(\theta_{1k} + \theta_{1k+1}) \tag{8.146}$$

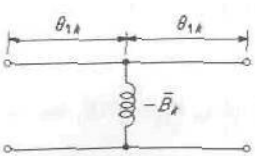


FIGURE 8.52
An impedance inverter.

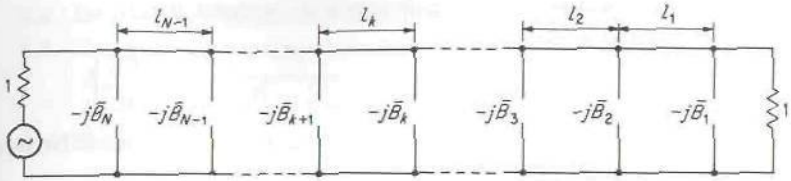


FIGURE 8.53
A direct-coupled waveguide-cavity filter.

In the prototype circuit of Fig. 8.28, we must choose all $\sqrt{L_{0k}/C_{0k}}$ equal to $\pi/2$ to obtain a correspondence with the type of series resonant circuit employed here. In addition, we choose

$$C_{0k}L_{0k} = (\beta_0 c)^{-2} \tag{8.147}$$

The impedance-inverter parameters as given by (8.115) thus become known in terms of the C_k and L_k, which are related to the g_k in the low-pass prototype. From the known values of the $Z_{k+1,k}$,the shunt susceptances $\bar{B}_k$ may be found. The filter is illustrated schematically in Fig. 8.53. The design formulas obtained as outlined above are

$$\bar{B}_1 = \frac{1 - w/g_1}{\sqrt{w/g_1}} \tag{8.148a}$$

$$\bar{B}_2 = \frac{1}{w}\left(1 - \frac{w^2}{g_1 g_2}\right)\sqrt{g_1 g_2} \tag{8.148b}$$

$$\bar{B}_k = \frac{1}{w}\left(1 - \frac{w^2}{g_k g_{k-1}}\right)\sqrt{g_k g_{k-1}} \tag{8.147c}$$

$$\bar{B}_N = \frac{1 - wR/g_{N-1}}{\sqrt{wR/g_{N-1}}} \quad R = 1 \text{ for } N \text{ even} \tag{8.148d}$$

where

$$w = \frac{\pi}{2}\frac{\beta_2 - \beta_1}{\beta_0}$$

and the g_k are the element values from the low-pass prototype filter. Note that $R = 1$ for N even, and also for N odd in the case of maximally flat filters; otherwise R is given by (8.98). The length of the kth cavity at $\beta = \beta_0$ is

$$l_k = \frac{\lambda_{g0}}{2} - \frac{\lambda_{g0}}{4\pi}\left(\tan^{-1}\frac{2}{\bar{B}_{k+1}} + \tan^{-1}\frac{2}{\bar{B}_k}\right) \tag{8.149}$$

The power loss ratio is given by substituting

$$\frac{\beta_0}{\beta_2 - \beta_1}\left(\frac{\beta}{\beta_0} - \frac{\beta_0}{\beta}\right)$$

for ω' in the low-pass prototype filter response.† Note also that $\beta_0^2 = \beta_1\beta_2$.

8.23 OTHER TYPES OF FILTERS

In the preceding sections we have discussed only a few specific filters in order to illustrate the insertion loss method of filter design. In addition to the maximally flat and Chebyshev-type filters, there exist two other types that have the feature that, in addition to producing equal-ripple response in the passband, they produce a number of attenuation poles outside the passband. These filters are the elliptic filters and the generalized Chebyshev response filters. By producing attenuation poles, i.e., frequencies at which the attenuation is infinite, the skirts of the filter response curve are much steeper so the attenuation will increase much faster outside the passband. It is much more difficult to find microwave circuit components that can be used to implement elliptic-type filters so these filters are not very commonly used.

Another important filter topic that we have not addressed in the phase response of a filter. In order that the filter should not produce a distortion of the signal transmitted through it, the amplitude response should be independent of frequency and the phase response should be a linear function of ω so that the group delay will be a constant. If we express the transmission coefficient of the filter in the form $A(\omega)e^{-j\phi(\omega)}$, then the time delay experienced by a signal propagating through the filter is given by $d\phi/d\omega$ as shown in Sec. 3.19. Thus ϕ should be a linear function of ω so as to give the same time delay for each frequency component of the signal. Real filters only satisfy this criterion in an approximate way. For critical applications, especially when broadband signals are involved, it may be necessary to insert a phase equalization circuit in series with the filter.

The reader is referred to the references at the end of this chapter for a discussion of a number of filter topics that we could not include because of space limitations.

PROBLEMS

8.1. Find the $\mathscr{ABCD}$ matrix for the following networks: (*a*) a shunt susceptance $j\bar{B}$, (*b*) a series reactance $j\bar{X}$, (*c*) a shunt reactance $j\bar{X}_1$ followed by a series reactance $j\bar{X}_2$.

†Cohn uses a somewhat different frequency variable, which, however, for the bandwidths considered, is essentially the same as we have used.

8.2. Derive the relations (8.14) and (8.15).

8.3. Consider a T network terminated in a load Z. Evaluate the input impedance Z_{in} and show that the condition that Z transforms into Z_{in}, that is, $Z_{in} = Z$, leads to the characteristic values $Z_B^{\pm}$ for the periodic structure.

8.4. Show that the eigenvalue equation for the propagation constant of a Bloch wave on a transmission line loaded at intervals d with a series reactance $j\bar{X}$ is $\cosh \gamma d = \cos k_0 d - (\bar{X}/2)\sin k_0 d$.

8.5. Show that (8.21) may be expressed in the form

$$\cosh \gamma d = \frac{\cos \phi}{|S_{12}|}$$

where ϕ is the phase angle of S_{12}, and S_{12} is the scattering-matrix off-diagonal element for the unit cell (Sec. 4.8).

8.6. Express Γ_B in terms of the A_{ij} by noting the similarity between (8.20) and (8.6) and that between Γ_B and $\bar{Z}_B^{-1}$.

8.7. Show that the wave-amplitude transmission matrix for a shunt susceptance $j\bar{B}$ is

$$[A] = \begin{bmatrix} \dfrac{2 + j\bar{B}}{2} & j\dfrac{\bar{B}}{2} \\ -j\dfrac{\bar{B}}{2} & \dfrac{4 + \bar{B}^2}{2(2 + j\bar{B})} \end{bmatrix}$$

8.8. A load $\bar{Z}_L$ on an ordinary transmission line gives a reflection coefficient $\Gamma_L' = (\bar{Z}_L - 1)/(\bar{Z}_L + 1)$. Show that (8.30), giving Γ_L for a Bloch wave, may be expressed as

$$\Gamma_L = \frac{1 + \Gamma_B^-}{1 + \Gamma_B^+} \frac{\Gamma_B^+ - \Gamma_L'}{\Gamma_L' - \Gamma_B^-}$$

where $\Gamma_B^{\pm}$ are the characteristic reflection coefficients of the component waves making up the Bloch wave.

8.9. For Bloch waves in the capacitively loaded coaxial line, show that the TEM voltage waves between any two consecutive diaphragms are given by

$$V = V_n^+ e^{-jk_0(z-nd)} + V_n^- e^{jk_0(z-nd)} = V_n^+ e^{-jk_0(z-nd)} + \Gamma_B^+ V_n^+ e^{jk_0(z-nd)}$$

for the Block wave propagating in the $+z$ direction, and

$$V = V_n^+ e^{-jk_0(z-nd)} + \Gamma_B^- V_n^+ e^{jk_0(z-nd)}$$

for the Bloch wave propagating in the $-z$ direction. The zeroth terminal plane has been chosen as $z = 0$ and $V_B^+ = V_n^+(1 + \Gamma_B^+)$, $V_B^- = V_n^+(1 + \Gamma_B^-)$.

8.10. Consider an infinite transmission line loaded with shunt capacitive susceptances $j\bar{B}$ at $z = nd$, $n = -\infty$ to ∞. Show that the current and voltage waves that make up a Bloch wave are given by

$$V_B = V^+ e^{-jk_0 z} + V^- e^{jk_0 z}$$

$$I_B = I^+ e^{-jk_0 z} + I^- e^{jk_0 z} = V^+ e^{-jk_0 z} - V^- e^{jk_0 z}$$

where $V^- = -V^+(1 - e^{-j\theta+j\beta d})/(1 - e^{j\theta+j\beta d})$ and $\theta = k_0 d$. Let $V_B(z)$ equal $V_p(z)e^{-j\beta z}$ and expand $V_p(z)$ into an infinite series of spatial harmonics. Note that the relation between V^+ and V^- may be obtained by using the condition that $V_B(z = d) = e^{-j\beta d}V_B(0)$ and that β is given by (8.10*a*).

8.11. For the sheath helix show that the eigenvalue equation for the nth mode is

$$\frac{(h^2a^2 + n\beta a \cot\psi)^2}{(k_0 ha^2 \cot\psi)^2} = \frac{K_n'(ha)I_n'(ha)}{K_n(ha)I_n(ha)}$$

8.12. Consider an N-section filter made up of a capacitively loaded coaxial line of N unit cells. The filter is terminated in a resistive load $\bar{R}$ equal to the image impedance at zero frequency, i.e., equal to 1. The generator at the input has an internal resistance equal to $\bar{R}$. Show that the power delivered to the load is given by

$$P = \frac{V_g^2}{|\bar{R} + \bar{Z}_{in}|^2} \operatorname{Re} \bar{Z}_{in}$$

where V_g is the generator voltage and $\bar{Z}_{in}$ is given by

$$\bar{Z}_{in} = \bar{Z}_i \frac{\bar{R} + \bar{Z}_i \tanh \gamma Nd}{\bar{Z}_i + \bar{R} \tanh \gamma Nd}$$

and $\bar{Z}_i$, γd are the image parameters at any frequency. In the passband where $\bar{Z}_i$ is real and $\tanh \gamma d = j \tan \beta d$, verify that

$$P = \frac{V_g^2 \bar{Z}_i^2 \bar{R}(1 + t^2)}{4\bar{R}^2\bar{Z}_i^2 + (\bar{R}^2 + \bar{Z}_i^2)t^2} = \frac{V_g^2(1 + t^2)}{4\bar{R}(1 + t^2) + \bar{R}\left(\dfrac{\bar{R}^2 - \bar{Z}_i^2}{\bar{R}\bar{Z}_i}\right)^2 t^2} \qquad t = \tan N\beta d$$

Thus show that the power loss ratio becomes

$$P_{LR} = \frac{V_g^2/4\bar{R}}{P} = 1 + \left(\frac{\bar{R}^2 - \bar{Z}_i^2}{2\bar{R}\bar{Z}_i}\right)^2 \sin N\beta d$$

Plot P_{LR} as a function of ω in the passband for the case where $\bar{B} = 3k_0 d$, $\bar{R} = 1$, and $N = 4$. See (8.10*a*) and (8.16) for expressions giving $\bar{Z}_i = \bar{Z}_B$ and βd. Verify that in the stopband the power loss ratio is given by

$$P_{LR} = 1 + \frac{1}{4}\left|\frac{\bar{R}}{\bar{Z}_i} - \frac{\bar{Z}_i}{\bar{R}}\right|^2 \sinh^2 N\alpha d$$

8.13. (*a*) For the two circuits shown in Fig. P8.13*a* show that the impedance mismatch is the same.

(*b*) Similarly, show that the two circuits shown in Fig. P8.13*b* have the same mismatch.

Thus in a filter the insertion loss does not change when an inverter with $K = Z_c$ or $J = Y_c$ is inserted at either end of the filter.

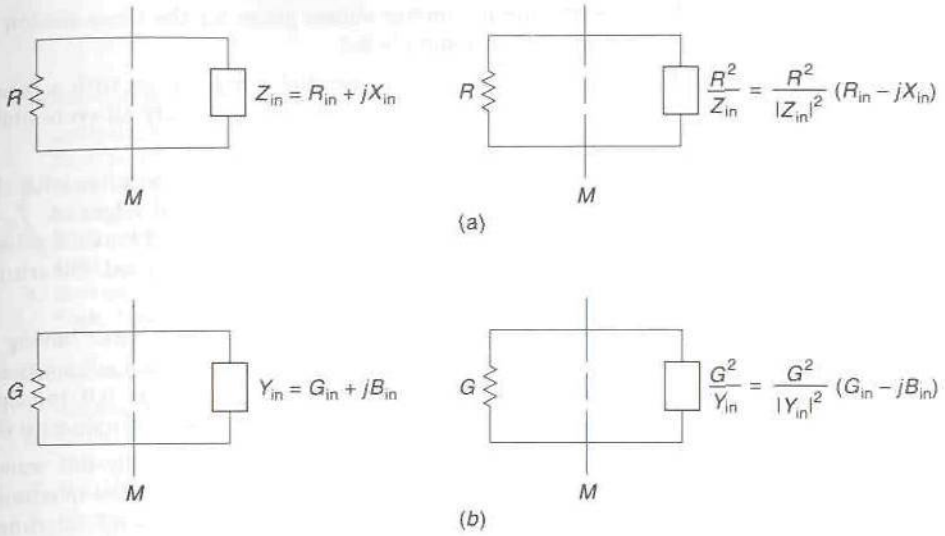


FIGURE P8.13

8.14. Design a two-section lumped-element bandpass filter similar to that in Example 8.1. Assume that $k^2 = 1$ and that the bandwidth equals $0.1f_0$. The terminating resistances equal 1,000 Ω and $f_0 = 1$ MHz. Verify that the coupling coefficient defined in Example 8.1 equals $(1 + \sqrt{2})/Q$, where Q is the loaded resonator Q.

8.15. Design a three-section lumped-element filter of the form shown in Fig. P8.15. Assume a passband tolerance $k^2 = 0.5$, a bandwidth equal to $0.1f_0$, terminating resistances of 1,000 Ω, and a center frequency of 455 kHz. Plot the filter response using $P_{LR} = 1 + k^2 T_3^2(\omega')$, where ω' is given by (8.109).

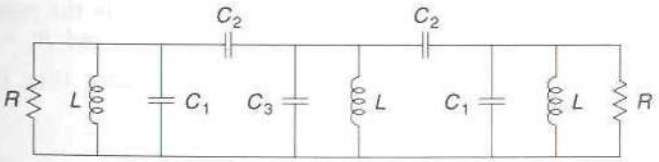


FIGURE P8.15

8.16. Design a two-section half-wave filter with a 5 percent bandwidth and a passband tolerance $k^2 = 0.2$. The frequency of operation is 5 GHz. The filter uses 50-Ω transmission lines. Specify the three required gap capacitances.

8.17. Design a three-section half-wave filter with the same specifications as in Prob. 8.16. Specify the required gap capacitances.

8.18. Design a two-section parallel coupled filter with a bandwidth of $0.1f_0$ and a passband tolerance $k^2 = 0.2$. Specify the required even- and odd-mode characteristic impedances for each section. Assume that $Z_c = 50$ Ω.

8.19. Verify the parameter values given for the three-section parallel coupled filter discussed in Example 8.3.

8.20. Design a three-section parallel coupled filter with a bandwidth equal to $0.2 f_0$ and a passband tolerance $k^2 = 0.1$. Specify all even- and odd-mode characteristic impedances. Assume $Z_c = 50\ \Omega$.

8.21. Design a three-cavity quarter-wave-coupled filter with the following specifications: waveguide width $a = 0.9$ in, band edges at $f_1 = 10{,}000$ MHz, $f_2 = 10{,}400$ MHz, passband tolerance $k^2 = 0.0233$, Chebyshev response. Inductive diaphragms with circular holes are to be used. Determine the hole radii and diaphragm spacings.

8.22. Design a four-cavity direct-coupled cavity filter having Chebyshev response. The passband tolerance is $k^2 = 0.0233$, band edges occur at $f_1 = 9{,}500$ MHz, $f_2 = 10{,}500$ MHz, and the guide width is 0.9 in. Specify the diaphragm dimensions and spacing. Use any convenient inductive diaphragm.

8.23. Design a four-cavity direct-coupled maximally flat waveguide filter with the specifications given in Prob. 8.22. Note that the maximally flat filter designed from the low-pass prototype has a passband tolerance of 1. To obtain a passband tolerance of k^2 between β_1 and β_2, the design must be carried out for a wider bandwidth, say β_1' to β_2'. Thus we should have

$$\left[\frac{\beta_0}{\beta_2' - \beta_1'}\left(\frac{\beta}{\beta_0} - \frac{\beta_0}{\beta}\right)\right]^8 = 1 \qquad \text{for } \beta = \beta_1', \beta_2'$$

Also $\beta_1'\beta_2' = \beta_0^2$. Determine β_2' and β_1' so that

$$\left[\frac{\beta_0}{\beta_2' - \beta_1'}\left(\frac{\beta}{\beta_0} - \frac{\beta_0}{\beta}\right)\right]^8 = k^2 \qquad \text{when } \beta = \beta_1 \text{ and } \beta_2$$

If the design is carried out using these values of β_1' and β_2', the required passband tolerance of k^2 will be maintained in the passband between β_1 and β_2. Show that, in general, $\beta_1'\beta_2' = \beta_1\beta_2 = \beta_0^2$ and $\beta_2' - \beta_1' = (\beta_2 - \beta_1)k^{-1/N}$.

8.24. For the circuits in Figs. 8.49*b* and *c* show that the normalized image impedances are given by

$$\left(1 - \overline{B}_k^2 - 2\overline{B}_k \cot\theta_k\right)^{-1/2} \qquad \text{and} \qquad \left(\frac{2\sin 2\theta_{1k} + \cos 2\theta_{1k} - \overline{B}}{2\sin 2\theta_{1k} + \cos 2\theta_{1k} + \overline{B}}\right)^{1/2}$$

When $\omega = \omega_0$, $\overline{B} = 0$, show that $\overline{B}_k$ must be related to θ_k by (8.137) to make $\overline{Z}_i$ equal to unity. Show that the image phase constants ϕ for the two circuits are given by

$$\cos\phi = \cos\theta_k + \overline{B}_k \sin\theta_k \qquad \text{and} \qquad \cos\phi = \cos 2\theta_{1k} - \frac{\overline{B}}{2}\sin 2\theta_{1k}$$

and will be equal at ω_0 if $\theta_k + 2\theta_{1k} = \pi$.

REFERENCES

Periodic structures

1. Brillouin, L.: "Wave Propagation in Periodic Structures," 2d ed., Dover Publications, Inc., New York, 1953.
2. Slater, J. C.: "Microwave Electronics," D. Van Nostrand Company, Inc., Princeton, N.J., 1950.
3. Watkins, D. A.: "Topics in Electromagnetic Theory," John Wiley & Sons, Inc., New York, 1958.
4. Bevensee, R. M.: "Electromagnetic Slow Wave Systems," John Wiley & Sons, Inc., New York, 1964.

Microwave filters

5. Rhodes, J. D.: "Theory of Electrical Filters," John Wiley & Sons, Inc., New York, 1976.
6. Malherbe, J. A. G.: "Microwave Transmission Line Filters," Atech House Books, Dedham, Mass., 1979.
7. Matthews, H. (ed.): "Surface Wave Filters," John Wiley & Sons, Inc., New York, 1977.
8. Alseyab, S. A.: A Novel Class of Generalized Chebyshev Low-Pass Prototype for Suspended Stripline Filters," *IEEE Trans.*, vol. MTT-30, pp. 1341–1347, 1982.
9. Mobbs, C. I., and J. D. Rhodes,: A Generalized Chebyshev Suspended Substrate Stripline Bandpass Filter, *IEEE Trans.*, vol. MTT-31, pp. 397–402, 1983.
10. Williams, A. E.: A Four-Cavity Elliptic Waveguide Filter," *IEEE Trans.*, vol. MTT-18, pp. 1109–1114, 1970.
11. Makimoti, M., and S. Yamashita: Bandpass Filters Using Parallel Coupled Stripline Stepped Impedance Resonators, *IEEE Trans.*, vol. MTT-28, pp. 1413–1417, 1980.
12. Wenzel, R. J.: Exact Theory of Inter-Digital Bandpass Filters and Related Coupled Structures, *IEEE Trans.*, vol. MTT-13, pp. 558–575, 1965.
13. Cohn, S. B.: Microwave Bandpass Filter Containing High-Q Dielectric Resonators, *IEEE Trans.*, vol. MTT-16, pp. 218–227, 1968.

CHAPTER 9

MICROWAVE TUBES

9.1 INTRODUCTION

Microwave tubes are the prime signal sources in high-power radar systems. The magnetron is the tube most frequently used and can provide many kilowatts of continuous-wave (CW) output power and a megawatt or more of peak power with pulsed operation. Magnetrons are also used for industrial heating applications and in microwave ovens for consumer use. The traveling-wave-tube amplifier with power outputs up to 10 W or more is the workhorse in satellite communications. The klystron tube can function as an oscillator or as an amplifier. It can be designed for either low or high output power applications. In low-power applications the klystron was once widely used as the local oscillator in microwave receivers but has now been replaced by solid-state oscillators. Solid-state oscillators are replacing microwave tubes in many low-power transmitter applications also. Even though many of the applications for microwave tubes have been taken over by solid-state devices, the requirements for high power can only be met by microwave tubes, so they are an essential device for many systems.

Conventional low-frequency tubes, such as triodes, fail to operate at microwave frequencies because the electron transit time from the cathode to the grid becomes an appreciable fraction of the period of the sinusoidal signal to be amplified. In other words, propagation times becomes significant, and the same limitations that are inherent in low-frequency circuits are present in low-frequency tubes also. Microwave tubes must be designed to utilize the wave-propagation phenomena to best advantage.

Broadly speaking, there are two basic types of microwave tubes, those that employ electromagnetic cavities (klystrons and some magnetrons) and

those that employ slow-wave circuits (traveling-wave tubes). Both types of tubes utilize an electron beam on which space-charge waves and cyclotron waves can be excited. The space-charge waves are primarily longitudinal oscillations of the electrons and interact with the electromagnetic fields in cavities and slow-wave circuits to give amplification. The properties of cavities and slow-wave circuits have already been discussed. What remains to be done is to examine the propagation of space-charge waves on electron beams and then to consider the interactions that take place between electron beams and the fields in cavities and slow-wave circuits.

The purpose of this chapter is to examine the nature of electron beams and the space-charge waves that they can support. In addition, the interaction of the beam with a microwave cavity or slow-wave circuit is to be studied in order to explain the operating principles of a number of different microwave tubes. Space does not permit a detailed treatment of the many different varieties of microwave tubes in existence. We shall concentrate on fundamentals that form, more or less, the basic operating principles of all microwave tubes.

Two approaches may be used in analyzing the dynamic behavior of the electron beam. The earliest approach used was the ballistic, or lagrangian, approach. In this method the motion of an individual electron is studied in detail, and it is assumed that all other electrons behave in a similar way. The ballistic approach has the advantage of permitting certain nonlinear, or large-signal, effects to be treated fairly easily.

The other approach is the field approach, sometimes called the eulerian, or hydrodynamical, approach. In this method the electron beam is essentially treated as a charged fluid. Field variables that describe the velocity, charge density, ac current, etc., at an arbitrary point as a function of time are introduced. However, no attempt is made to follow the motion of a single electron. The field approach, which leads to the space-charge waves, is more unifying and lends itself to the treatment of all different types of microwave tubes within the same general mathematical framework. Therefore only the field approach is used in this text.

An exact analysis of a microwave tube would be very difficult and laborious to carry out. As in any other physical problem, it is necessary to introduce a number of simplifying assumptions in order to arrive at a mathematical model that can be analyzed without too many complications. The success of a simplified theory must then be judged by the extent to which it predicts and agrees with experimental results.

The first few sections of this chapter discuss a number of models used for the electron beam and the propagation of space-charge waves on these beams. The governing equations are Maxwell's equations and Newton's laws, together with the Lorentz force equation. The equation of motion for a charge element is a nonlinear equation, but may be linearized by assuming small-signal conditions; i.e., all ac quantities are small compared with dc quantities. We shall consider only the small-signal situation since this will

suffice to develop the operating principles of microwave tubes. Large-signal analysis is a great deal more difficult, and the theory, in general, is not fully developed.

After treating the dynamics of the electron beam, the klystron and traveling-wave tube are examined in detail. A number of other tubes are also discussed, but in a more qualitative way.

9.2 ELECTRON BEAMS WITH dc CONDITIONS

By means of a suitable electron gun consisting of a cathode, accelerating electrodes, and focusing electrodes, a beam of electrons with essentially a uniform velocity v_0 can be produced.† Figure 9.1 is a schematic illustration of a cylindrical electron beam with a radius a. If the potential difference through which the electron is accelerated is V, the velocity v_0 is given by

$$v_0 = \left(\frac{2Ve}{m}\right)^{1/2} = 5.93 \times 10^5 V^{1/2} \text{ m/s} \tag{9.1}$$

where $-e$ is the electron charge and m is the mass of the electron. For $V = 1{,}000$ V, $v_0 = 1.87 \times 10^7$ m/s $= 0.0625c$. The beam perveance is defined by the quantity $IV^{-1/2}$, where I is the total beam current.

The coulomb repulsive force, or dc space-charge force, will tend to cause the electron beam to disperse, i.e., cause outward radial motion of the electrons. The space-charge force will be proportional to the density of the beam, i.e., to the number of electrons per unit volume. For the usual density of beams employed in microwave tubes (10^{12} to 10^{15} electrons per cubic meter), the dispersion of the beam due to space-charge forces is negligible if the drift space is short (d is small in Fig. 9.1). This condition exists in many klystrons, but in traveling-wave tubes the beam must travel over distances which are so long that considerable dispersion may take place unless some means of keeping the beam together or focused is employed. The means by which the dc space-charge forces are counteracted leads to three commonly used beam models. These models are discussed below.

Ion-Neutralized Beam

Even with the high vacuum employed in a microwave tube, a great many neutral gas particles are still present. Many of these gas molecules become ionized by means of collisions with the relatively high energy electrons. The

†The design of electron guns is not treated in this text. For a discussion of these see:
J. R. Pierce, "Theory and Design of Electron Beams," D. Van Nostrand Company, Inc., Princeton, N.J., 1950.
K. R. Spangenberg, "Vacuum Tubes," McGraw-Hill Book Company, New York, 1948.

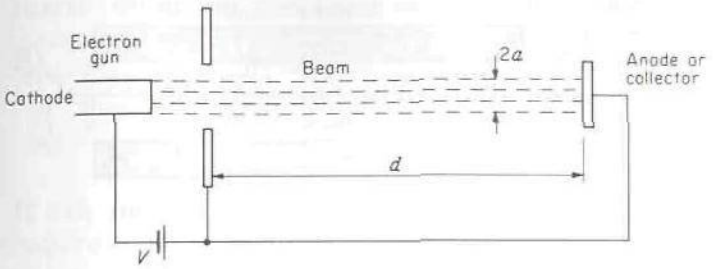


FIGURE 9.1
A cylindrical electron beam.

presence of positive ions will tend to neutralize the negative space charge of the electron beam. The positive ions, however, need not be considered in the interaction of a high-frequency electromagnetic field with the beam because their mass is at least 1,800 times greater than the electron's mass, and hence the ac motion of the ions is negligible by comparison with that of the electrons.

Although all electron beams are ion-neutralized to some extent, complete electron space-charge neutralization is rarely achieved. However, for the purpose of mathematical analysis, a completely ion-neutralized electron beam is sometimes postulated as a model. Beam spreading due to space-charge forces is discussed in Spangenberg's book.

Beam with Axially Confined Flow

If a very large static magnetic field $\mathbf{B}_0$ in the direction of the beam velocity is applied, the effect is to constrain the electrons from moving in the radial direction. The space-charge forces tend to impart a radial velocity to the electron. The magnetic field $\mathbf{B}_0$ produces a force $-e\mathbf{v}_r \times \mathbf{B}_0$, which causes the electrons to execute circular motion about the magnetic field lines and thus prevents the beam from dispersing in the radial direction.

In the magnetically focused beam the field $\mathbf{B}_0$ has its flux lines threading through the cathode surface, as in Fig. 9.2*a*. Some electron diffusion across the magnetic field lines will occur, but if $\mathbf{B}_0$ is made large enough, the amount of beam dispersion can be kept small.

For the purpose of mathematical analysis, it is convenient to assume that $\mathbf{B}_0$ is made infinite since in this case no electron motion in a transverse direction can take place. The analysis of the behavior of the beam under ac conditions is thereby greatly simplified since electron motion can now occur only in the axial direction (one-dimensional motion). The axially confined flow model is commonly used in the treatment of traveling-wave tubes.

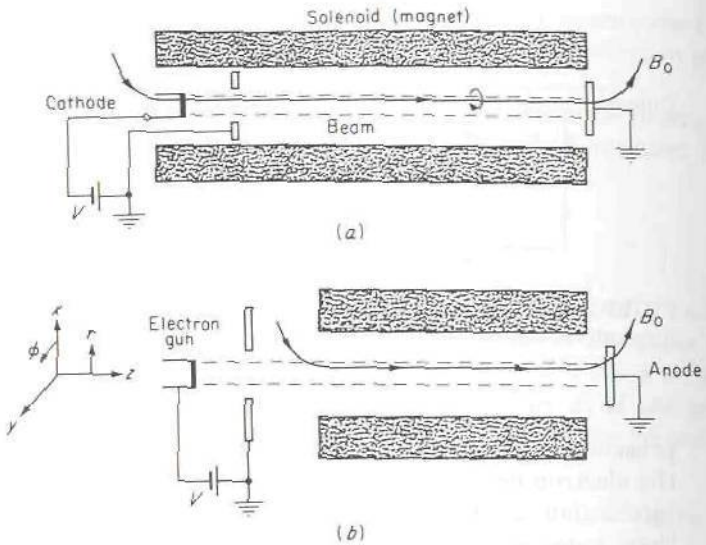


FIGURE 9.2
(*a*) Magnetic focusing for axially confined flow; (*b*) magnetic focusing for Brillouin flow.

Brillouin Flow

In Brillouin flow (to be described), the axial magnetic field $\mathbf{B}_0$ is not permitted to thread through the cathode surface. Since the field lines are continuous, they must move away from the beam region in the radial direction near the cathode, as shown in Fig. 9.2*b*. When the beam enters the magnetic field region, it is given a uniform rotation at the Larmor frequency $\omega_l = eB_0/2m$ by the magnetic field. In cyclindrical coordinates, r, ϕ, z, the equation of motion for an electron,

$$m\frac{d\mathbf{v}}{dt} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

may be written in component form as

$$\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 = -\frac{e}{m}\left(E_r + rB_0\frac{d\phi}{dt}\right) \tag{9.2a}$$

$$r\frac{d^2\phi}{dt^2} + 2\frac{dr}{dt}\frac{d\phi}{dt} = \frac{e}{m}B_0\frac{dr}{dt} \tag{9.2b}$$

$$\frac{d^2z}{dt^2} = 0 \qquad \frac{dz}{dt} = v_0 \tag{9.2c}$$

in the region where $\mathbf{B} = B_0\mathbf{a}_z$ and is uniform. It is assumed that $E_\phi = E_z = 0$. The radial electric field may be found by using Gauss' law. If the dc

charge density of the beam is $-\rho_0$, then $2\pi rD_r = -\pi r^2\rho_0$, or $E_r = -r\rho_0/2\epsilon_0$. The radial space-charge force on an electron is thus $-eE_r = re\rho_0/2\epsilon_0$. If $d^2\phi/dt^2 = 0$, we find from (9.2*b*) that

$$\frac{d\phi}{dt} = \omega_l = \frac{eB_0}{2m} \tag{9.3}$$

If this solution is to satisfy (9.2*a*) and also make d^2r/dt^2 vanish, we require

$$r\omega_l^2 = \frac{e}{m}\left(-\frac{r\rho_0}{2\epsilon_0} + rB_0\omega_l\right)$$

or

$$\omega_l^2 = \frac{e\rho_0}{2m\epsilon_0} = \frac{\omega_p^2}{2} \tag{9.4}$$

where $\omega_p = (e\rho_0/m\epsilon_0)^{1/2}$ is called the plasma frequency. Typical values of ω_p for beams used in microwave tubes range from 10^7 to 10^9. If the focusing field B_0 is chosen to satisfy (9.4), there will be no radial acceleration of the electrons. The equilibrium condition in the radial direction is actually a balance of outward radial forces $-eE_r$ due to space charge and $m\omega_l^2 r$ due to centrifugal acceleration against the inward magnetic radial force $e\omega_l rB_0$. Electron-beam flow under these conditions is referred to as Brillouin flow.

Although we have given the conditions for steady-state Brillouin flow within the uniform $\mathbf{B}_0$ field region, we did not show that a beam leaving a cathode with a velocity $v_0\mathbf{a}_z$ will assume Brillouin-flow characteristics as it enters into the uniform $\mathbf{B}_0$ field region through the nonuniform field region in front of the cathode. To show this requires demonstrating that the change in angular momentum of the beam from an initial value of zero to its final value for Brillouin flow is equal to the time integral of the torque ev_0B_{0r} produced by the radial magnetic field component in the nonuniform region. The reader is referred to Brillouin's original paper for the derivation.†

The conditions required for Brillouin flow can be achieved in practice. Even if the beam is partially ion-neutralized, as long as ρ_0 is not zero, a value for B_0 such that (9.4) holds can be found. However, the behavior of a beam with Brillouin flow under ac conditions is more difficult to treat since transverse motion of the electrons is permitted. For this reason the ideal axially confined flow model is more commonly used.

---

†L. Brillouin, A Theorem of Larmor and Its Importance for Electrons in Magnetic Fields, *Phys. Rev.*, vol. 67, p. 260, 1945.

W. G. Dow, Nonuniform D.C. Electron Flow in Magnetically Focused Cylindrical Beams, *Advan. Electron. Electron Phys.*, vol. 10, 1958.

Pierce, *op. cit.*

In the magnetron-type (M-type) traveling-wave tube, a planar sheet beam is used. For this type of beam an analogous flow, referred to as planar Brillouin flow, can take place. The properties of sheet beams are discussed in Sec. 9.11, dealing with M-type traveling-wave tubes, and hence are not covered in this section.

9.3 SPACE-CHARGE WAVES ON BEAMS WITH CONFINED FLOW

This section is devoted to an analysis of space-charge waves on an axially confined electron beam inside a cylindrical waveguide of radius b. The radius of the beam is a, as in Fig. 9.3. Small-signal conditions are assumed.

The beam is considered to be uniform in density in a cross-sectional plane. The dc charge density is $-\rho_0$, and the axial velocity is v_0. The dc current density in the z direction is $J_0 = -\rho_0 v_0$. The dc parameters ρ_0, v_0, and J_0 *are independent of space and time coordinates.* Under ac conditions with time dependence $e^{j\omega t}$, there will be ac components of charge density, velocity, and current that vary with time and the spatial coordinates. These ac components are denoted by ρ, $\mathbf{v}$, and $\mathbf{J}$. The ac fluctuation in electron density from the dc value N will be denoted by n.

The electromagnetic field satisfies the equations

$$\nabla \times \mathbf{E} = -j\omega\mu_0 \mathbf{H} \tag{9.5a}$$

$$\nabla \times \mathbf{H} = j\omega\epsilon_0 \mathbf{E} + \mathbf{J} \tag{9.5b}$$

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{9.5c}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{9.5d}$$

$$\nabla \cdot \mathbf{J} = -j\omega\rho \tag{9.5e}$$

A unit volume of the beam with charge density $-\rho_0 + \rho$ and charge-mass ratio η equal to that for electrons, that is, $\eta = e/m$, has a motion governed by the equation

$$(N + n)m\frac{d\mathbf{v}_t}{dt} = (\rho - \rho_0)(\mathbf{E} + \mathbf{v}_t \times \mathbf{B} + \mathbf{v}_t \times \mathbf{B}_0) \tag{9.5f}$$

where $N + n$ is the number of electrons per unit volume and $\mathbf{v}_t = \mathbf{v}_0 + \mathbf{v}$ is the total velocity. For small-signal conditions, $|\mathbf{B}| \ll |\mathbf{B}_0|$; so the force term

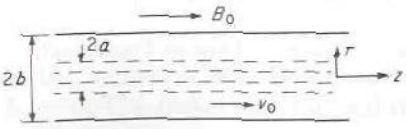


FIGURE 9.3
Electron beam inside a cylindrical waveguide.

$\mathbf{v}_t \times \mathbf{B}$ can be neglected compared with $\mathbf{v}_t \times \mathbf{B}_0$. In addition, since $B = \mu_0 H \approx \mu_0 Y_0 E = E/c$, we see that $|\mathbf{v}_t \times \mathbf{B}|$ is smaller than $|\mathbf{E}|$ by a factor v_t/c. Hence ac magnetic forces are negligible. The total velocity $\mathbf{v}_t$ is a function of x, y, z, and t. In addition, the position x, y, z of a charge element is a function of time. Consequently,

$$\frac{d\mathbf{v}_t}{dt} = \frac{\partial \mathbf{v}_t}{\partial t} + \frac{\partial \mathbf{v}_t}{\partial x}\frac{dx}{dt} + \frac{\partial \mathbf{v}_t}{\partial y}\frac{dy}{dt} + \frac{\partial \mathbf{v}_t}{\partial z}\frac{dz}{dt} = \frac{\partial \mathbf{v}_t}{\partial t} + (\mathbf{v}_t \cdot \nabla)\mathbf{v}_t \quad (9.6)$$

since $\mathbf{v}_t = \mathbf{a}_x\, dx/dt + \mathbf{a}_y\, dy/dt + \mathbf{a}_z\, dz/dt$. Thus expansion of $d\mathbf{v}_t/dt$ leads to a nonlinear term $(\mathbf{v}_t \cdot \nabla)\mathbf{v}_t$, depending on v_t^2. However, for an ac velocity $\mathbf{v}$ that is small compared with the dc velocity $\mathbf{v}_0$, we have

$$[(\mathbf{v} + \mathbf{v}_0) \cdot \nabla](\mathbf{v} + \mathbf{v}_0) = (\mathbf{v} + \mathbf{v}_0) \cdot \nabla\mathbf{v} \approx (\mathbf{v}_0 \cdot \nabla)\mathbf{v} \quad (9.7)$$

since $\mathbf{v}_0$ is constant and the second-order term $(\mathbf{v} \cdot \nabla)\mathbf{v}$ is negligible and may be dropped. Thus we obtain

$$(N + n)m\left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}_0 \cdot \nabla)\mathbf{v}\right] = (\rho - \rho_0)(\mathbf{E} + \mathbf{v} \times \mathbf{B}_0)$$

But the terms involving n and ρ are products of two ac quantities and may be dropped for small-signal conditions. Hence the first-order linearized equation of motion becomes

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}_0 \cdot \nabla)\mathbf{v} = -\eta(\mathbf{E} + \mathbf{v} \times \mathbf{B}_0) \quad (9.8)$$

since $Ne = \rho_0$ and $e/m = \eta$.

For the cylindrical beam under consideration, we also have $\mathbf{v}_0 = \mathbf{a}_z v_0$ and a time dependence $e^{j\omega t}$. If we let $\mathbf{B}_0$ approach infinity, the transverse components of $\mathbf{v}$ must vanish, so that the term $\mathbf{v} \times \mathbf{B}_0$ in (9.8) will vanish. Thus $\mathbf{v}$ has a component in the z direction only, and (9.8) gives

$$j\omega v + v_0 \frac{\partial v}{\partial z} = -\eta E_z \quad (9.9)$$

When $\mathbf{v}$ has only a z component, the ac current density $\mathbf{J}$ has only a z component since the total current is

$$\begin{aligned}\mathbf{J}_0 + \mathbf{J} &= (-\rho_0 + \rho)(\mathbf{v}_0 + \mathbf{v}) = -\rho_0\mathbf{v}_0 + (\rho\mathbf{v}_0 - \rho_0\mathbf{v}) + \rho\mathbf{v} \\ &\approx -\rho_0\mathbf{v}_0 + (\rho\mathbf{v}_0 - \rho_0\mathbf{v}) \end{aligned} \quad (9.10)$$

after dropping the second-order term $\rho\mathbf{v}$, which is the product of two small ac quantities. The dc and ac currents are thus

$$\mathbf{J}_0 = -\rho_0\mathbf{v}_0 \quad (9.11a)$$

$$\mathbf{J} = \rho\mathbf{v}_0 - \rho_0\mathbf{v} \quad (9.11b)$$

From the continuity equation (9.5e) we obtain

$$\frac{\partial J}{\partial z} = -j\omega\rho \tag{9.12}$$

Equations (9.9), (9.11b), and (9.12) permit us to express J as a function of E_z. Maxwell's equations (9.5a) and (9.5b) may then be solved in the usual manner to obtain wave solutions.

Since we are looking for wave solutions, we may assume that all ac quantities have a z dependence $e^{-j\beta z}$. In this case (9.9) and (9.12) give

$$(j\omega - j\beta v_0)v = -\eta E_z \tag{9.13a}$$

$$j\beta J = j\omega\rho \tag{9.13b}$$

For convenience, ω/v_0 will be denoted by β_0, which may be interpreted as the dc propagation constant for the beam. Using (9.11b) and (9.13), we find that

$$J = -j\frac{\omega_p^2}{\omega}\frac{\beta_0^2\epsilon_0 E_z}{(\beta_0 - \beta)^2} \tag{9.14}$$

where $\omega_p^2 = \rho_0\eta/\epsilon_0$ is the plasma frequency squared, and $-\rho_0$ is the dc electron charge density of the beam.

To solve Maxwell's equations for the beam inside a cylindrical guide, it will be convenient to introduce the vector potential. For a mode having azimuthal symmetry (no ϕ dependence), all boundary conditions can be satisfied by a vector potential having only a z component $A_z(r, z) = \psi(r)e^{-j\beta z}$. The equation satisfied by A_z is

$$\nabla^2 A_z + k_0^2 A_z = -\mu_0 J$$

From A_z we obtain

$$E_z = -j\omega A_z + \left.\frac{\nabla\nabla\cdot\mathbf{A}_z}{j\omega\mu_0\epsilon_0}\right|_z = \frac{k_0^2 + \partial^2 A_z/\partial z^2}{j\omega\mu_0\epsilon_0} = \frac{k_0^2 - \beta^2}{j\omega\mu_0\epsilon_0}A_z$$

Using (9.14) to express E_z in terms of J gives

$$\mu_0 J = -\left(\frac{\omega_p}{\omega}\right)^2 \beta_0^2 \frac{k_0^2 - \beta^2}{(\beta_0 - \beta)^2}A_z \tag{9.15}$$

The Helmholtz equation for A_z now becomes

$$\nabla_t^2 A_z + p^2 A_z = 0 \qquad 0 \le r \le a \tag{9.16a}$$

$$\nabla_t^2 A_z - h^2 A_z = 0 \qquad a \le r \le b \tag{9.16b}$$

where we have replaced ∇^2 by $\nabla_t^2 - \beta^2$ and put

$$p^2 = -\beta^2 + k_0^2 + \left(\frac{\omega_p}{\omega}\right)^2 \beta_0^2 \frac{\beta^2 - k_0^2}{(\beta_0 - \beta)^2}$$

$$= -(\beta^2 - k_0^2)\left[1 - \left(\frac{\omega_p}{\omega}\right)^2 \left(\frac{\beta_0}{\beta_0 - \beta}\right)^2\right] \tag{9.17a}$$

$$h^2 = \beta^2 - k_0^2 \tag{9.17b}$$

after using (9.15) to express J in terms of A_z.

The analysis, when completed, will show that the space-charge waves are slow waves, with $\beta \approx \beta_0 \gg k_0$, and hence p and h will be real. With no ϕ variation, (9.16) reduces to

$$\frac{d^2\psi}{dr^2} + \frac{1}{r}\frac{d\psi}{dr} + \begin{Bmatrix} p^2 \\ -h^2 \end{Bmatrix} \psi = 0 \tag{9.18}$$

where $A_z = \psi(r)e^{-j\beta z}$. The equation for ψ is Bessel's equation of order zero, and the solutions are $J_0(pr)$, $Y_0(pr)$, $J_0(jhr)$, and $Y_0(jhr)$. Instead of using the Bessel functions with imaginary argument, we use the modified Bessel functions $I_0(hr)$, $K_0(hr)$. In the region $0 \le r \le a$ we cannot use Y_0 since it becomes infinite. Therefore we let

$$\psi(r) = C_1 J_0(pr) \qquad 0 \le r \le a$$

$$\psi(r) = C_2 I_0(hr) + C_3 K_0(hr) \qquad a \le r \le b$$

where C_1, C_2, C_3 are arbitrary constants. The axial electric field must vanish at $r = b$ and must be continuous at $r = a$. These conditions hold for A_z, and hence for ψ also. Thus

$$C_1 J_0(pa) = C_2 I_0(ha) + C_3 K_0(ha) \tag{9.19a}$$

$$0 = C_2 I_0(hb) + C_3 K_0(hb) \tag{9.19b}$$

Besides E_z, the only other field components present are E_r and H_ϕ. These are given in terms of A_z by

$$E_r = -\frac{\beta}{\omega\mu_0\epsilon_0}\frac{\partial A_z}{\partial r} = -\frac{\beta c}{k_0}\frac{\partial A_z}{\partial r}$$

$$H_\phi = -\frac{1}{\mu_0}\frac{\partial A_z}{\partial r} = \frac{k_0}{\beta} Y_0 E_r$$

Continuity of H_ϕ at $r = a$ requires

$$C_1 p J_0'(pa) = C_2 h I_0'(ha) + C_3 h K_0'(ha) \tag{9.20}$$

where the prime denotes differentiation with respect to the argument pa or ha. In order for (9.19) and (9.20) to have a nontrivial solution for C_1, C_2,

and C_3, the determinant of the coefficients must vanish. Thus we find that

$$p\frac{J_0'(pa)}{J_0(pa)} = h\frac{K_0(hb)I_0'(ha) - K_0'(ha)I_0(hb)}{K_0(hb)I_0(ha) - K_0(ha)I_0(hb)} \tag{9.21}$$

This transcendental equation, together with the relations (9.17), which give

$$p^2 = -h^2 + \left(\frac{\omega_p}{\omega}\right)^2 \frac{\beta_0^2 h^2}{\left(\beta_0 - \sqrt{h^2 + k_0^2}\right)^2} \tag{9.22}$$

determines the propagation constant β.

Two special cases are now examined. First consider the case where b and a are made very large. Then, since

$$I_0(x) \sim \frac{e^x}{\sqrt{2\pi x}} \qquad K_0(x) \sim \sqrt{\frac{\pi}{2x}}\, e^{-x}$$

for large x, we find that (9.21) gives $p \tan(pa - \pi/4) = h$. But since we are letting a go to infinity, the only possible solution independent of a is $p = h = 0$. From (9.17) we then obtain the following solutions for β:

$$\beta = \pm k_0 \tag{9.23a}$$

$$\beta = \beta_0\left(1 \pm \frac{\omega_p}{\omega}\right) \tag{9.23b}$$

But with $p = h = 0$, $\beta = \pm k_0$, all field components vanish as reference to the equations given earlier for E_z, H_ϕ, and E_r shows. Thus this is a trivial solution. The other solutions $\beta = \beta_0(1 \pm \omega_p/\omega)$ correspond to the space-charge waves. The wave velocities are

$$\frac{\omega}{\beta} = \frac{v_0}{1 \pm \omega_p/\omega} \approx v_0\left(1 \mp \frac{\omega_p}{\omega}\right)$$

since $\omega_p \ll \omega$ for conditions that are typical in microwave tubes. The wave velocities are slightly greater and slightly smaller than the dc beam velocity v_0. The two waves are called the fast and slow space-charge waves. For $p = 0$, both E_r and H_ϕ vanish but E_z remains finite. The space-charge waves may thus be viewed as a longitudinal oscillation of the electrons in the beam. When $\omega = \omega_p$, one solution corresponds to $\beta = 0$, that is, no propagation. Thus it is seen that the plasma frequency is a natural frequency of oscillation for electrons in an infinite beam. Since the transverse fields are zero, the space-charge waves are not changed even if B_0 is finite as long as the beam has infinite radius.

As a second special case consider the situation $b = a$ so that the beam fills the waveguide. From (9.21) we now see that the right-hand side becomes infinite since the denominator vanishes. Thus we require $J_0(pa) = 0$. Hence pa takes on values typical of those for TM_{0m} modes in a circular guide. The lowest-order solution is (Sec. 3.18) $pa = 2.405$, or in general

$pa = p_{0m}$, where the p_{0m} are given in Table 3.5. Using (9.22), we now find that

$$\left(\frac{p_{0m}}{a}\right)^2 = -h^2\left[1 - \left(\frac{\omega_p}{\omega}\right)^2 \frac{\beta_0^2}{\left(\beta_0 - \sqrt{h^2 + k_0^2}\right)^2}\right]$$

$$= \left(k_0^2 - \beta^2\right)\left[1 - \left(\frac{\omega_p}{\omega}\right)^2 \frac{\beta_0^2}{(\beta_0 - \beta)^2}\right] \tag{9.24}$$

For the field waves we expect β to be approximately equal to k_0. Then, since $\omega_p \ll \omega$ and $\beta_0 \gg k_0$, an approximate solution of (9.24) is

$$\beta^2 \approx k_0^2 - \left(\frac{p_{0m}}{a}\right)^2 \tag{9.25}$$

This is the unperturbed propagation constant for TM_{0m} modes in a cylindrical guide. A correction to β may be obtained by using the solution given by (9.25) in the term multiplied by ω_p^2 in (9.24).

Of greater interest are the space-charge waves for which $\beta \approx \beta_0$. For these $k_0^2 \ll \beta^2$; so (9.24) may be approximated by

$$\left(\frac{p_{0m}}{a}\right)^2 = -\beta^2\left[1 - \left(\frac{\omega_p}{\omega}\right)^2 \frac{\beta_0^2}{(\beta_0 - \beta)^2}\right]$$

which is a quadratic equation in β^2. To obtain an approximate solution, let $\beta = \beta_0(1 + \delta)$, where δ will be small. Then we obtain

$$\left(\frac{p_{0m}}{a}\right)^2 = -\beta_0^2\left[1 - \left(\frac{\omega_p}{\omega}\right)^2 \frac{1}{\delta^2}\right]$$

which gives

$$\delta = \frac{\pm(\omega_p/\omega)\beta_0}{\left[\beta_0^2 + (p_{0m}/a)^2\right]^{1/2}}$$

and hence

$$\beta = \beta_0\left[1 \pm \frac{\omega_p}{\omega}\left(1 + \frac{p_{0m}^2}{\beta_0^2 a^2}\right)^{-1/2}\right] \tag{9.26}$$

Note that δ is small, which justifies the approximations made. We may express (9.26) in the same form as (9.23*b*) by introducing an effective plasma frequency ω_q given by

$$\omega_q = \omega_p\left(1 + \frac{p_{0m}^2}{\beta_0^2 a^2}\right)^{-1/2} = F\omega_p \tag{9.27}$$

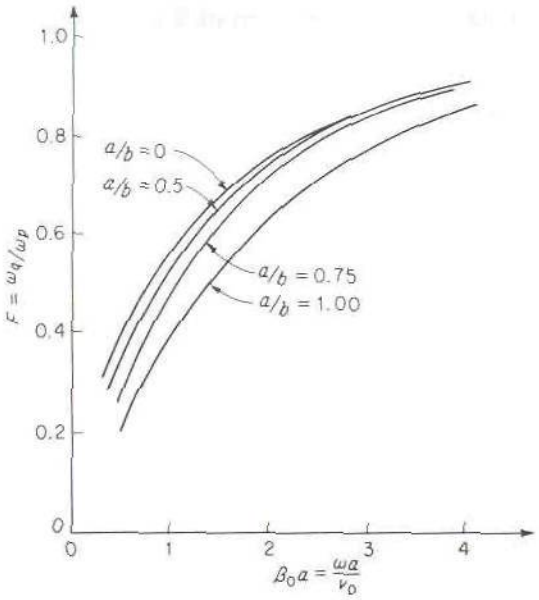


FIGURE 9.4
Space-charge reduction factor for a cylindrical electron beam of radius a, velocity $\mathbf{v}_0$, inside a circular guide of radius b. The data apply to the dominant TE_{01} space-charge mode.

where F is called the space-charge reduction factor; for example, $\omega_q^2 = F^2\rho_0\eta/\epsilon_0$; so the effective space charge is $F^2\rho_0$. Hence we can write

$$\beta = \beta_0\left(1 \pm \frac{\omega_q}{\omega}\right) \tag{9.28}$$

For the beam completely filling the guide, there are again a slow and a fast space-charge wave, with velocities slightly greater and slightly smaller than the beam velocity v_0. However, the effective plasma frequency is reduced because of transverse variations in the field. Nevertheless, the space-charge waves have very small transverse field components E_r, H_ϕ. Only the axial electric field E_z is large.

In the general case, when $a \neq b$, the solution for β is tedious. The final results may be expressed in the form (9.28) for the space-charge waves by introducing the effective plasma frequency or the space-charge reduction factor. Some typical results computed from curves given by Branch and Mihran are shown in Fig. 9.4.†

The space-charge-wave theory was first developed by Hahn and Ramo in 1939. Since that time space-charge waves under a variety of conditions have been studied. The references cited at the end of this chapter will

†G. M. Branch and T. G. Mihran, Plasma Frequency Reduction Factors in Electron Beams, *IRE Trans.*, vol. ED-2, pp. 3–11, 1955.

provide an introduction to the literature on this topic. For the analysis of the ordinary, or *O*-type, traveling-wave tube, the model discussed above is accurate enough to describe the main operating characteristics.

9.4 SPACE-CHARGE WAVES ON UNFOCUSED BEAMS

Many low-power klystrons employ electron beams without magnetic field focusing when the distance (drift-space length) the beam must travel is short. The propagation of space-charge waves on this type of beam is therefore of interest. We shall consider a beam of radius a and with dc parameters $-\rho_0, v_0\mathbf{a}_z$. For space-charge waves with axial symmetry, the only field components present are E_r, E_z, and H_ϕ. The governing equations are Maxwell's equations (9.5) and the force equation (9.8), with $\mathbf{B}_0$ equated to zero.†

For space-charge waves we may assume a z dependence $e^{-j\beta z}$. In component form (9.5a) to (9.5c) and (9.8) are

$$j\beta E_r + \frac{\partial E_z}{\partial r} = j\omega\mu_0 H_\phi \tag{9.29a}$$

$$j\beta H_\phi = j\omega\epsilon_0 E_r + J_r \tag{9.29b}$$

$$\frac{1}{r}\frac{\partial}{\partial r} rH_\phi = j\omega\epsilon_0 E_z + J_z \tag{9.29c}$$

$$\frac{1}{r}\frac{\partial}{\partial r} rE_r - j\beta E_z = \frac{\rho}{\epsilon_0} \tag{9.29d}$$

$$v_r = \frac{j\eta E_r}{\omega - \beta v_0} \tag{9.29e}$$

$$v_z = \frac{j\eta E_z}{\omega - \beta v_0} \tag{9.29f}$$

In addition, we have the relation

$$\mathbf{J} = \rho\mathbf{v}_0 - \rho_0\mathbf{v}$$

which gives

$$J_r = -\rho_0 v_r = \sigma E_r \tag{9.30a}$$

$$J_z = -\rho_0 v_z + \rho v_0 = \sigma E_z + \rho v_0 \tag{9.30b}$$

where (9.29e) and (9.29f) have been used, and the effective conductivity σ

†Static space-charge forces are assumed to be negligible, which is a valid assumption for a low-density beam and a short drift space. Alternatively, the beam may be assumed to be ion-neutralized.

has been introduced as follows:

$$\sigma = \frac{-j\eta\rho_0}{\omega - \beta v_0} = \frac{-j\epsilon_0\omega_p^2}{\omega - \beta v_0} \tag{9.31}$$

When we make use of (9.30), the continuity equation $\nabla \cdot \mathbf{J} = -j\omega\rho$ is found to give

$$\sigma\left(\frac{1}{r}\frac{\partial}{\partial r}rE_r - j\beta E_z\right) - j\beta v_0\rho = -j\omega\rho$$

If (9.29d) is used to replace the term in parentheses, we obtain

$$\left(\frac{\sigma}{\epsilon_0} - j\beta v_0 + j\omega\right)\rho = 0 \tag{9.32}$$

If the ac charge density ρ does not vanish, we must have

$$\frac{\sigma}{\epsilon_0} - j\beta v_0 + j\omega = 0$$

This requires that β be given by

$$\beta = \frac{\omega \pm \omega_p}{v_0} = \beta_0\left(1 \pm \frac{\omega_p}{\omega}\right)$$

The corresponding wave solutions are the space-charge waves in an infinite beam for which $E_r = H_\phi = 0$. For a beam with finite radius the boundary conditions at $r = a$ cannot be satisfied with these values of β. Consequently, the space-charge waves that we are looking for must have a different value of β and, in addition, must have a zero ac space-charge density ρ, so that (9.32) will hold. For these waves, (9.30) now gives $\mathbf{J} = \sigma\mathbf{E}$, or

$$\mathbf{J} = j\omega\epsilon_0\frac{\sigma}{j\omega\epsilon_0}\mathbf{E}$$

Maxwell's curl equation for $\mathbf{H}$ becomes

$$\nabla \times \mathbf{H} = j\omega\epsilon_0\mathbf{E} + \mathbf{J} = j\omega\epsilon_0\left(1 + \frac{\sigma}{j\omega\epsilon_0}\right)\mathbf{E} = j\omega\epsilon\mathbf{E}$$

where the effective permittivity ϵ of the beam is given by

$$\epsilon = \epsilon_0\left(1 + \frac{\sigma}{j\omega\epsilon_0}\right) = \epsilon_0\left[1 - \frac{\omega_p^2}{\omega(\omega - \beta v_0)}\right] \tag{9.33}$$

By using the effective permittivity ϵ, the beam may be treated as a dielectric cylinder. The equation satisfied by E_z is thus

$$\nabla^2 E_z + k^2 E_z = \nabla_t^2 E_z + (k^2 - \beta^2) E_z = 0 \tag{9.34}$$

where

$$k^2 = \omega^2 \mu_0 \epsilon = k_0^2 \left[1 - \frac{\omega_p^2}{\omega(\omega - \beta v_0)} \right]$$

For space-charge waves we anticipate slow waves, for which $\beta \gg k_0$. Hence a suitable solution for E_z is

$$E_z = C_1 I_0(pr) \qquad 0 \le r \le a \tag{9.35a}$$

$$E_z = C_2 K_0(hr) \qquad r \ge a \tag{9.35b}$$

where $p = (\beta^2 - k^2)^{1/2}$, $h = (\beta^2 - k_0^2)^{1/2}$, and C_1, C_2 are amplitude constants to be determined.

The boundary conditions at $r = a$ are different from those for a beam with confined flow, for the following reasons.† The electrons have a radial ac velocity, and hence the boundary of the beam does not remain at its dc position $r = a$ since electrons will oscillate back and forth about the dc boundary. The resultant boundary thus becomes rippled, as shown in Fig. 9.5*a*. The positive charge shown in this figure represents a deficit of negative charge. The effect of the rippled boundary on the radial electric field under small-signal conditions may be accounted for by replacing the rippled boundary by a layer of surface charge. This is, of course, exactly what is done in the case of a dielectric boundary in which polarization charge oscillates back and forth about a mean boundary surface. The surface charge is given in terms of the dielectric polarization $\mathbf{P}$ by $\mathbf{P} \cdot \mathbf{n} = (\mathbf{D} - \epsilon_0 \mathbf{E}) \cdot \mathbf{n}$. For the electron beam the corresponding surface-charge density ρ_s arises from two causes, namely, charge flowing toward the boundary because of the radial current J_r and charge carried to a given point z owing to surface charge moving with the beam. That is, the rate of increase of surface charge is (Fig. 9.5*b*)

$$\frac{\partial \rho_s}{\partial t} = j\omega \rho_s = J_r - v_0 \frac{\partial \rho_s}{\partial z} = J_r + j\beta v_0 \rho_s$$

The term $-v_0 \partial \rho_s / \partial z$ arises as follows: Let the surface charge density at a point z be ρ_s. Then at the point $z - dz$ the charge density is approximately $\rho_s - (\partial \rho_s / \partial z)\, dz$. The rate at which charge is carried away from the point z on the boundary is $v_0 \rho_s$, and the rate at which charge from the adjacent

†See also W. C. Hahn, Small Signal Theory of Velocity Modulated Electron Beams, *Gen. Elec. Rev.*, vol. 42, pp. 258–270, 1939.

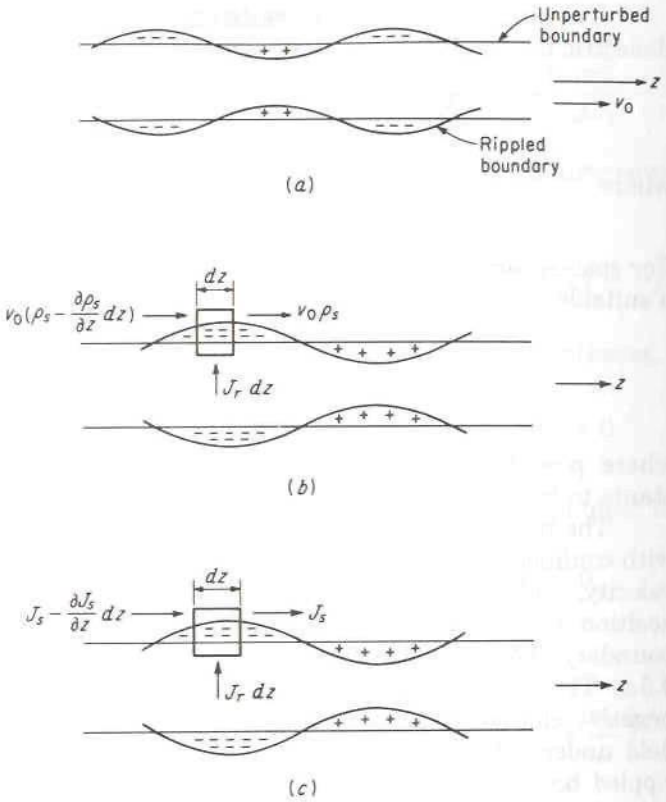


FIGURE 9.5
Boundary conditions at the surface of a rippled beam.

point $z - dz$ flows toward z is $v_0[\rho_s - (\partial \rho_s / \partial z)\, dz]$. The net rate of accumulation of charge in an interval dz, due to the finite beam velocity v_0, is thus $-v_0(\partial \rho_s / \partial z)\, dz = j\beta v_0 \rho_s\, dz$, since the z dependence is $e^{-j\beta z}$. The charge density is obtained by dividing by dz. Our final expression for the equivalent surface charge density is

$$\rho_s = \frac{J_r}{j(\omega - \beta v_0)} = \frac{\sigma E_r}{j(\omega - \beta v_0)} = \frac{-\eta \rho_0 E_r}{(\omega - \beta v_0)^2} \tag{9.36}$$

The amount of charge which has crossed the unperturbed boundary is $-\rho_0 r$, where r is the ac displacement of a unit volume of charge. Since

$$\frac{dr}{dt} = v_r = \frac{\partial r}{\partial t} + (\mathbf{v}_0 \cdot \nabla) r = j(\omega - \beta v_0) r$$

we see that

$$\rho_s = -\rho_0 r = \frac{-\rho_0 v_r}{j(\omega - \beta v_0)} = \frac{J_r}{j(\omega - \beta v_0)}$$

which is an alternative derivation of (9.36).

The boundary condition to be applied to the radial electric field is

$$E_{2r} - E_{1r} = \frac{\rho_s}{\epsilon_0} = \frac{-\eta\rho_0 E_{1r}}{\epsilon_0(\omega - \beta v_0)^2} = \frac{-\omega_p^2 E_{1r}}{(\omega - \beta v_0)^2} \tag{9.37}$$

where the subscripts 1 and 2 refer to the fields in the regions $r < a$ and $r > a$, respectively. The above boundary condition may also be expressed in the form

$$\epsilon_0 E_{2r} = \left[\epsilon_0 + \frac{\sigma}{j(\omega - \beta v_0)}\right] E_{1r}$$

This latter result is similar to that which holds at the boundary of a dielectric cylinder except that ω is replaced by $\omega - \beta v_0$ because of the uniform motion of the cylinder, in the z direction, with velocity v_0.

Associated with the equivalent surface charge is an equivalent surface current of density $\mathbf{J}_s = J_s \mathbf{a}_z$. To obtain an expression for $\mathbf{J}_s$, consider Fig. 9.5c. The total rate at which charge flows into a small region of length dz on the boundary is

$$J_r\, dz - \frac{\partial J_s}{\partial z}\, dz$$

and must equal the rate of increase $j\omega\rho_s\, dz$ of surface charge density in an interval dz. Consequently,

$$-\frac{\partial J_s}{\partial z} = j\beta J_s = j\omega\rho_s - J_r = j\beta v_0 \rho_s$$

or

$$J_s = \rho_s v_0 \tag{9.38}$$

after replacing J_r by $j(\omega - \beta v_0)\rho_s$ from (9.36). For $\omega \gg \omega_p$, the total surface current is usually much larger than the total volume current, and it is therefore very important to include it.

The boundary condition to be applied to H_ϕ is

$$H_{2\phi} - H_{1\phi} = J_s = \rho_s v_0 \tag{9.39}$$

When H_ϕ satisfies this discontinuity relation, the boundary condition on the radial electric field is also satisfied. With the above boundary conditions we are now able to complete the solution to our beam problem.

When we combine (9.29a) and (9.29b), we obtain

$$H_\phi = \frac{j\omega\epsilon_0 + \sigma}{\beta^2 - k_0^2 + j\omega\mu_0\sigma}\frac{\partial E_z}{\partial r} = \frac{j\omega\epsilon}{\beta^2 - k^2}\frac{\partial E_z}{\partial r}$$

Referring to (9.35), we thus find that

$$H_\phi = H_{1\phi} = \frac{j\omega\epsilon}{p}C_1 I_0'(pr) = \frac{j\omega\epsilon}{p}C_1 I_1(pr) \qquad r < a \quad (9.40a)$$

$$H_\phi = H_{2\phi} = \frac{j\omega\epsilon_0}{h}C_2 K_0'(hr) = -\frac{j\omega\epsilon_0}{h}C_2 K_1(hr) \qquad r > a \quad (9.40b)$$

We require E_z to be continuous at $r = a$ and H_ϕ to satisfy the condition (9.39). Therefore

$$C_1 I_0(pa) = C_2 K_0(ha) \quad (9.41a)$$

$$-\frac{j\omega\epsilon_0}{h}C_2 K_1(ha) - \frac{j\omega\epsilon}{p}C_1 I_1(pa) = \rho_s v_0 = \frac{\sigma v_0}{j(\omega - \beta v_0)}E_{1r}$$

From (9.29b) we find $E_{1r}(j\omega\epsilon_0 + \sigma) = j\beta H_{1\phi}$, and hence the second boundary condition becomes

$$\frac{C_2 K_1(ha)}{h} = -\frac{C_1 I_1(pa)}{p}\left[1 - \frac{\omega_p^2}{(\omega - \beta v_0)^2}\right] \quad (9.41b)$$

Dividing (9.41b) by (9.41a) gives

$$\frac{I_1(pa)}{pI_0(pa)}\left[1 - \frac{\omega_p^2}{(\omega - \beta v_0)^2}\right] = -\frac{K_1(ha)}{hK_0(ha)} \quad (9.42)$$

The propagation constant β is determined by a solution of this equation, together with the relation

$$\beta^2 = p^2 + k^2 = p^2 + k_0^2\left[1 - \frac{\omega_p^2}{\omega(\omega - \beta v_0)}\right] = h^2 + k_0^2 \quad (9.43)$$

For space-charge waves, $\beta \gg k_0$, so that h and p can be replaced by β, an approximation that may be used to simplify (9.42) and (9.43).

The results obtained by solving (9.42) and (9.43) may be expressed in the form (9.28):

$$\beta = \beta_0\left(1 \pm \frac{\omega_q}{\omega}\right)$$

At microwave frequencies, ω_q/ω is usually in the range 0.01 to 0.1; so β

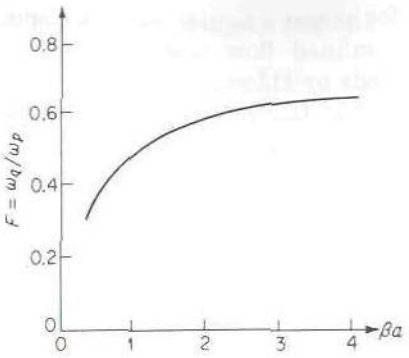


FIGURE 9.6
Plasma-frequency reduction factor F for an unfocused cylindrical electron beam as a function of βa.

differs from β_0 by only a few percent or less. In Fig. 9.6 the plasma-frequency reduction factor $F = \omega_q/\omega_p$ is plotted as a function of βa. If p and h are replaced by β in (9.42), that equation may be put into the form

$$\beta a = \beta_0 a\left[1 \pm \frac{\omega_p}{\omega}\left(\frac{K_1 I_0}{K_0 I_1} + 1\right)^{-1/2}\right]$$

from which it is seen that

$$F = \frac{\omega_q}{\omega_p} = \left[1 + \frac{K_1(\beta a) I_0(\beta a)}{K_0(\beta a) I_1(\beta a)}\right]^{-1/2} \tag{9.44}$$

This expression was evaluated to obtain F, as given in Fig. 9.6. The data in Fig. 9.6 may be used to solve for the corresponding values of $\beta_0 a$; that is,

$$\beta_0 a = \frac{\beta a}{1 \pm F(\omega_p/\omega)} \tag{9.45}$$

In a beam with confined flow only axial ac convection currents are permitted. For this reason an ac charge density ρ exists since the current J_z is a function of z and must have associated with it space-charge-density fluctuations. For the unfocused beam, ac radial convection currents are also present, and this makes it possible for the total current $\mathbf{J} + j\omega\epsilon_0\mathbf{E}$ to be solenoidal, i.e., to form continuous closed flow lines without terminating in ac space charge.

9.5 ac POWER RELATIONS

The ac power associated with the space-charge waves on an electron beam is of importance for understanding the gain mechanism of traveling-wave

tubes and, of course, for power calculations. The small-signal power theorem for beams with confined flow was first derived by Chu.† Further extensions have been made by Haus and Bobroff‡ and Klüver.§ The derivation is straightforward, but the interpretation of the various terms that enter in is not always clear, depending on the type of beam model being considered.

Maxwell's equations for the fields associated with the beam are

$$\nabla \times \mathbf{E} = -j\omega\mu_0\mathbf{H} \qquad \nabla \times \mathbf{H} = j\omega\epsilon_0\mathbf{E} + \mathbf{J}$$

where the small-signal ac current density is given by (the surface current is included for generality)

$$\mathbf{J} = -\rho_0\mathbf{v} + \rho\mathbf{v}_0 + \rho_s\mathbf{v}_0$$

In addition, we have the equation of motion

$$j\omega\mathbf{v} + (\mathbf{v}_0 \cdot \nabla)\mathbf{v} = -\eta(\mathbf{E} + \mathbf{v} \times \mathbf{B}_0)$$

Expanding the following expression and using the above equations, we obtain

$$\begin{aligned}\nabla \cdot (\mathbf{E} \times \mathbf{H}^*) &= \mathbf{H}^* \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H}^* \\ &= -j\omega\mu_0\mathbf{H} \cdot \mathbf{H}^* + j\omega\epsilon_0\mathbf{E} \cdot \mathbf{E}^* - \mathbf{E} \cdot \mathbf{J}^* \qquad (9.46)\end{aligned}$$

The continuity equation in the interior of the beam may be written as

$$\begin{aligned}\mathbf{v}_0\nabla \cdot \mathbf{J} &= -j\omega\rho\mathbf{v}_0 = -j\omega(\rho\mathbf{v}_0 - \rho_0\mathbf{v}) - j\omega\rho_0\mathbf{v} \\ &= -j\omega\mathbf{J} - j\omega\rho_0\mathbf{v}\end{aligned}$$

If we multiply the complex conjugate of this equation by $\mathbf{v}/\eta$, we obtain

$$\frac{\mathbf{v} \cdot \mathbf{v}_0}{\eta}\nabla \cdot \mathbf{J}^* = \frac{j\omega}{\eta}\mathbf{v} \cdot \mathbf{J}^* + \frac{j\omega\rho_0}{\eta}\mathbf{v} \cdot \mathbf{v}^*$$

Multiplying the equation of motion by $\mathbf{J}^*/\eta$ gives

$$\mathbf{J}^* \cdot \mathbf{E} + \mathbf{J}^* \cdot \mathbf{v} \times \mathbf{B}_0 = -\frac{j\omega}{\eta}\mathbf{v} \cdot \mathbf{J}^* - \frac{\mathbf{J}^*}{\eta} \cdot (\mathbf{v}_0 \cdot \nabla)\mathbf{v}$$

Since the transverse components of $\mathbf{J}$ and $\mathbf{v}$ are in the same direction, the term

$$\mathbf{J}^* \cdot \mathbf{v} \times \mathbf{B}_0 = \mathbf{J}^* \times \mathbf{v} \cdot \mathbf{B}_0 = 0$$

---

†L. J. Chu, A Kinetic Power Theorem, paper presented at the IRE-PGED Electron Tube Research Conference, Durham, N.H., June, 1951.

‡H. A. Haus and D. Bobroff, Small Signal Power Theorem for Electron Beams, *J. Appl. Phys.*, vol. 28, pp. 694–703, June, 1957.

§J. W. Klüver, Small Signal Power Conservation Theorem for Irrotational Electron Beams, *J. Appl. Phys.*, vol. 29, pp. 618–622, April, 1958.

The addition of the above two equations thus gives

$$\mathbf{J}^* \cdot \mathbf{E} - \frac{j\omega\rho_0}{\eta}\mathbf{v} \cdot \mathbf{v}^* = -\frac{\mathbf{v} \cdot \mathbf{v}_0}{\eta}\nabla \cdot \mathbf{J}^* - \frac{\mathbf{J}^*}{\eta} \cdot (\mathbf{v}_0 \cdot \nabla)\mathbf{v}$$

If we introduce the term $\nabla \cdot (\mathbf{v} \cdot \mathbf{v}_0\mathbf{J}^*) = \mathbf{v} \cdot \mathbf{v}_0\nabla \cdot \mathbf{J}^* + \mathbf{J}^* \cdot \nabla(\mathbf{v} \cdot \mathbf{v}_0)$, we obtain for the right-hand side

$$-\nabla \cdot \left(\frac{\mathbf{v} \cdot \mathbf{v}_0}{\eta}\mathbf{J}^*\right) - \frac{1}{\eta}\left[\mathbf{J}^* \cdot (\mathbf{v}_0 \cdot \nabla)\mathbf{v} - \mathbf{J}^* \cdot \nabla\mathbf{v} \cdot \mathbf{v}_0\right]$$

By expanding the bracketed term in rectangular coordinates, it is found to reduce to

$$\frac{v_0}{\eta}\mathbf{a}_z \times \mathbf{J}^* \cdot \nabla \times \mathbf{v}$$

when $\mathbf{v}_0 = v_0\mathbf{a}_z$. Hence we have

$$\mathbf{J}^* \cdot \mathbf{E} - \frac{j\omega\rho_0}{\eta}\mathbf{v} \cdot \mathbf{v}^* = -\nabla \cdot \left(\frac{\mathbf{v} \cdot \mathbf{v}_0}{\eta}\mathbf{J}^*\right) - \frac{v_0}{\eta}\mathbf{a}_z \times \mathbf{J}^* \cdot \nabla \times \mathbf{v} \quad (9.47)$$

The ac kinetic-power theorem is obtained by adding (9.47) to (9.46) to obtain

$$\nabla \cdot \left(\mathbf{E} \times \mathbf{H}^* - \frac{\mathbf{v} \cdot \mathbf{v}_0}{\eta}\mathbf{J}^*\right) = -j\omega\mu_0\mathbf{H} \cdot \mathbf{H}^* - j\omega\frac{\rho_0}{\eta}\mathbf{v} \cdot \mathbf{v}^* + j\omega\epsilon_0\mathbf{E} \cdot \mathbf{E}^* + \frac{v_0}{\eta}\mathbf{a}_z \times \mathbf{J}^* \cdot \nabla \times \mathbf{v} \quad (9.48)$$

The term $(\rho_0/2\eta)\mathbf{v} \cdot \mathbf{v}^*$ is the ac kinetic-energy density in the beam since $\rho_0/\eta = (\rho_0/e)m$ is the mass density per unit volume.

Let us now specialize to the case of a beam with confined flow for which $\mathbf{J} = \mathbf{a}_z J_z$ and the last term on the right-hand side of (9.48) vanishes. The real part of the above equation then gives

$$\operatorname{Re} \nabla \cdot \left(\mathbf{E} \times \mathbf{H}^* - \frac{\mathbf{v} \cdot \mathbf{v}_0}{\eta}\mathbf{J}^*\right) = 0 \quad (9.49)$$

The volume integral over a volume enclosing the beam between cross sections S_1 and S_2 as in Fig. 9.7 may be converted into a surface integral

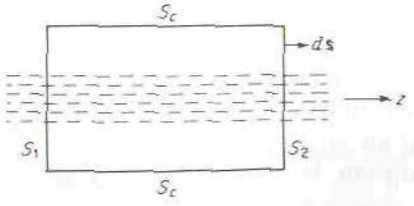


FIGURE 9.7
A section of an electron beam.

over the surface $S_1 + S_2 + S_c$. Thus the following power-conservation theorem is obtained:

$$\operatorname{Re}\oint_S \frac{1}{2}\mathbf{E}\times\mathbf{H}^*\cdot d\mathbf{S} = -\operatorname{Re}\int_{S_2}\frac{-\mathbf{v}\cdot\mathbf{v}_0}{2\eta}\mathbf{J}^*\cdot\mathbf{a}_z\,dS + \operatorname{Re}\int_{S_1}\frac{-\mathbf{v}\cdot\mathbf{v}_0}{2\eta}\mathbf{J}^*\cdot\mathbf{a}_z\,dS \tag{9.50}$$

The term $(-\mathbf{v}\cdot\mathbf{v}_0)/\eta = -(m/e)\mathbf{v}\cdot\mathbf{v}_0 = V_k$ has the dimensions of a voltage, and is called the kinetic voltage. It is the term that gives the conversion of kinetic energy of the beam into electromagnetic energy. In order that a net amount of electromagnetic energy may flow out of the surface S, the quantity

$$\operatorname{Re}\oint_S \frac{1}{2}V_k\mathbf{J}^*\cdot d\mathbf{S}$$

must be negative.

A fuller appreciation of the above relations may be obtained by considering the application to space-charge waves in an axially confined infinite cross-section electron beam. For the slow and fast space-charge waves, the ac kinetic power may be obtained by using (9.13a), (9.14), and (9.49). It is found that

$$\operatorname{Re}\left(\frac{1}{2}V_{kf}J_f^*\right) = \operatorname{Re}\frac{-v_0v_fJ_f^*}{2\eta} = -\frac{\omega\omega_p^2\epsilon_0}{2v_0^2(\beta_f-\beta_0)^3}|E_{zf}|^2$$

$$\operatorname{Re}\left(\frac{1}{2}V_{ks}J_s^*\right) = -\frac{\omega\omega_p^2\epsilon_0}{2v_0^2(\beta_s-\beta_0)^3}|E_{zs}|^2$$

Since $\beta_s > \beta_0 > \beta_f$, the slow space-charge wave has a negative ac kinetic power, whereas the fast space-charge wave has a positive ac kinetic power. If the beam is excited, with only the slow space-charge wave, the significance of the negative ac kinetic power is that some of the dc flow energy of the beam has been extracted and converted into negative ac energy by the excitation process. The negative ac kinetic energy must in turn be converted into electromagnetic energy flow in order to maintain power conservation. In the discussion of the traveling-wave tube, it will be seen that the slow space-charge wave is the one that produces amplification.

9.6 VELOCITY MODULATION

The preceding sections have established the existence of space-charge waves on electron beams. We must now examine the problem of exciting these waves, i.e., producing an ac velocity modulation on the beam. In klystron tubes velocity modulation is commonly produced by passing the beam

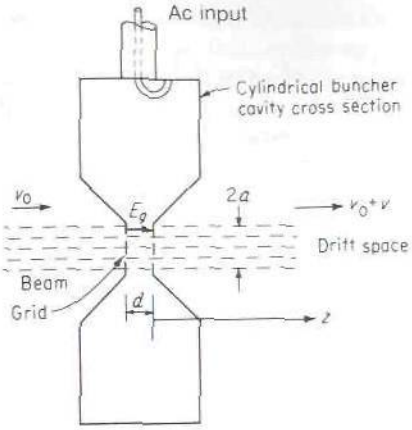


FIGURE 9.8
Velocity modulation of an electron beam.

through two closely spaced grids located at the center of a cylindrical reentrant cavity, as in Fig. 9.8. The particular form of cavity used is chosen in order to satisfy the requirement of high ac electric field strength across the grids (which requires small grid spacing) and yet maintain a high cavity Q. The latter requires a large volume–surface area ratio. If we let $\mathrm{Re}(E_g e^{j\omega t}) = E_g \cos \omega t$ be the cavity electric field across the gap (often referred to as the buncher gap), those electrons entering the gap when $E_g \cos \omega t$ is directed in the negative z direction will be accelerated and will leave with a velocity greater than v_0. Electrons entering the gap region when $E_g \cos \omega t$ is directed in the positive z direction are slowed down and will leave with a velocity less than v_0. It is apparent, then, that an applied ac electric field between two parallel grids will velocity-modulate an electron beam. The analytical details of the modulation process are discussed below.†

We shall consider an unfocused electron beam of the type discussed in Sec. 9.4. Let the cavity field in the gap region be $E_g \cos \omega t$. Electrons will traverse the gap with essentially the entrance velocity v_0. If the time at which a particular electron passes the midplane $z = -d/2$ is t_1, then the field in the cavity at time t when this electron is at a position $z = -d/2 + v_0(t - t_1)$ is

$$E_g \cos \omega t = E_g \cos \frac{\omega}{v_0}\left(z + \frac{d}{2} + v_0 t_1\right)$$

The work done by the cavity field on the electron during its transit through

†This analysis is based on a ballistic formulation, and not on a field approach, since the former is more straightforward.

the gap is

$$W = \int_{-d}^{0} -eE_g \cos \beta_0 \left(z + \frac{d}{2} + v_0 t_1 \right) dz$$

$$= -eE_g d \frac{\sin(\beta_0 d/2)}{\beta_0 d/2} \cos \omega t_1 \tag{9.51}$$

where $\beta_0 = \omega/v_0$ is the dc propagation constant for the beam. The beam-coupling parameter M is defined to be

$$M = \frac{\sin(\beta_0 d/2)}{\beta_0 d/2} \tag{9.52}$$

For an electron passing the midplane at time t, the work done on the electron is clearly given by

$$-eE_g \, dM \cos \omega t$$

The work done on an electron results in an increase in its kinetic energy. If the exit velocity from the buncher cavity is $v_0 + v_z$, we have

$$\tfrac{1}{2}m(v_0 + v_z)^2 - \tfrac{1}{2}mv_0^2 = \tfrac{1}{2}m(2v_0 v_z + v_z^2) \approx mv_0 v_z = -eE_g \, dM \cos \omega t \tag{9.53}$$

since for small-signal conditions, $v_z \ll v_0$. In complex form (9.53) may be written as

$$v_z e^{j\omega t} = \frac{-\eta}{v_0} E_g \, dM \, e^{j\omega t}$$

Thus the axial ac beam velocity at the exit grid has a value

$$v_z = \frac{-\eta}{v_0} M(E_g d) \tag{9.54}$$

The foregoing first-order analysis predicts that there will be zero average work done in bunching the electron beam since the average of (9.51) over one period from t_1 to $t_1 + 1/f$ is zero. This result is not correct, and in actual fact, a net amount of average work is required to velocity-modulate the beam. To determine the average work done, a second-order analysis must be performed.† The principal effect of requiring a finite amount of work to velocity-modulate the beam can be represented by an equivalent shunt conductance loading the buncher cavity (beam loading of the buncher cavity). The magnitude of this shunt conductance is typically such as to

†K. R. Spangenberg, "Vacuum Tubes," chap. 17, McGraw-Hill Book Company, New York, 1948.

M. Chodorow and C. Suskind, "Fundamentals of Microwave Electronics," chap. 3, McGraw-Hill Book Company, New York, 1964.

reduce the unloaded Q of the buncher cavity by a factor of 2 or so. However, even though the first-order analysis given above is not sufficiently accurate to give the beam-loading equivalent conductance, it does give a satisfactory answer for the velocity modulation of the beam, which was the information we were interested in obtaining from the analysis.

Since we now know the velocity modulation of the beam at the exit grid of the buncher cavity, we are in a position to evaluate the amplitudes of the space-charge waves that will be excited on the beam in the drift space beyond the buncher cavity. We shall treat the case of an unfocused beam in detail. The case of a beam with confined flow is somewhat easier to analyze and is the model usually assumed in the analysis of the klystron, even though it is not the type of beam used in most klystrons. However, it turns out that the results for the unfocused beam and the beam with confined flow are essentially equivalent, the main difference being that, for the unfocused beam, the major contribution to the ac current comes from the equivalent surface current on the beam, whereas for the beam with confined flow, the ac current is a volume current distributed over the cross section of the beam. During the course of the analysis, the results for the case of the beam with confined flow will be given for comparative purposes.

In the drift space $z > 0$, space-charge waves will be launched because of the ac velocity modulation of the entering beam. At the plane of the exit grid, the radial electric field of the space-charge waves is short-circuited and must be zero. This condition can be met by a suitable combination of the fast and slow space-charge waves. If we let E_{rs} and E_{rf} be the radial electric field of the slow and fast space-charge waves, we require $E_{rs} = -E_{rf}$ at $z = 0$. But E_{rs} and E_{rf} depend on r according to the first-order modified Bessel function $I_1(pr)$, where p is different for the fast and slow waves. However, for typical beams, $p \approx \beta \approx \beta_0$; so the radial dependence can be taken as $I_1(\beta_0 r)$ with negligible error, in which case the boundary conditions at $z = 0$ can be satisfied without introducing higher-order space-charge modes.

The required boundary conditions on E_r can be met if we choose the two space-charge waves to have the same amplitude and to combine in phase for the ac velocity v_z at $z = 0$. Thus let v_{zs} and v_{zf} be the amplitudes of the slow and fast space-charge-wave axial velocities, so that we may write

$$v_z = v_{zs}e^{-j\beta_s z} + v_{zf}e^{-j\beta_f z} = v_{zf}(e^{-j\beta_f z} + e^{-j\beta_s z})$$

where $\beta_s = \beta_0(1 + \omega_q/\omega)$ and $\beta_f = \beta_0(1 - \omega_q/\omega)$, and ω_q is the effective plasma frequency equal to $F\omega_p$. Introducing these expressions gives

$$v_z = 2v_{zf} \cos \beta_q z\, e^{-j\beta_0 z} \tag{9.55}$$

where $\beta_q = \beta_0\omega_q/\omega = \omega_q/v_0$. Note that v_{zf} is a function of r according to $I_0(\beta_0 r)$. However, $\beta_0 a$ is small, so that v_{zf} is almost constant. Thus we may

equate $2v_{zf}$ at $r = 0$ to v_z as given by (9.54) to obtain

$$v_z = \frac{-\eta}{v_0} MV_g \cos \beta_q z I_0(\beta_0 r) e^{-j\beta_0 z} \tag{9.56}$$

for the ac axial velocity at any point in the drift space $z > 0$. In (9.56) we have put V_g for the exciting gap voltage $E_g d$.

For the unfocused beam the ac space charge density ρ is zero, and from (9.30*b*) we find $J_z = -\rho_0 v_z$. Consequently, the ac axial current density in the drift space is

$$J_z = \frac{\eta \rho_0}{v_0} MV_g \cos \beta_q z I_0(\beta_0 r) e^{-j\beta_0 z} \tag{9.57}$$

The surface current J_s will be evaluated later, and will turn out to be more important than the volume current.

To find E_z, we use (9.29*f*) to obtain

$$E_{zs} = \frac{\omega - \beta_s v_0}{j\eta} v_{zs} = -\frac{\omega_q}{j\eta} v_{zs} = -\frac{\omega_q}{j\eta} v_{zf}$$

$$E_{zf} = \frac{\omega - \beta_f v_0}{j\eta} v_{zf} = \frac{\omega_q}{j\eta} v_{zf}$$

Hence the axial electric field is given by

$$E_z = \frac{\omega_q}{j\eta} v_{zf} (e^{-j\beta_f z} - e^{-j\beta_s z}) = \frac{2\omega_q}{\eta} v_{zf} e^{-j\beta_0 z} \sin \beta_q z$$

Introducing the earlier expression for v_{zf} gives

$$E_z = -\beta_q MV_g I_0(\beta_0 r) \sin \beta_q z e^{-j\beta_0 z} \tag{9.58}$$

Note that E_z vanishes at the exit grid, where $z = 0$.

From (9.29*d*) we have

$$\frac{1}{r}\frac{\partial}{\partial r} rE_{rs} = j\beta_s E_{zs} = -(\beta_0 + \beta_q)\frac{\omega_q}{\eta} v_{zf} \approx -\frac{\omega_q}{\eta}\beta_0 v_{zf}$$

$$\frac{1}{r}\frac{\partial}{\partial r} rE_{rf} = j\beta_f E_{zf} = (\beta_0 - \beta_q)\frac{\omega_q}{\eta} v_{zf} \approx \frac{\omega_q}{\eta}\beta_0 v_{zf}$$

$$\frac{1}{r}\frac{\partial}{\partial r} r(E_{rs} + E_{rf}) = -2\beta_q \frac{\omega_q}{\eta} v_{zf} \approx 0$$

since $\beta_0 \gg \beta_q$. The boundary conditions on E_r are consequently satisfied to a very good degree of approximation. If desired, a small adjustment of the amplitudes of the two space-charge waves could make $E_{rs} + E_{rf}$ vanish

exactly at a particular value of r. However, since we have already approximated $I_1(pr)$ by $I_1(\beta_0 r)$ for both waves, the present approximation of dropping β_q relative to β_0 is consistent with our earlier assumptions.

If we examine the expressions for J_z and v_z, we see that, because of the beating or interference between the two space-charge waves, the ac current and velocity vary according to $\cos \beta_q z$ in the drift space. Maximum volume current density occurs when

$$z = n\frac{\lambda_q}{2} \tag{9.59}$$

where $\lambda_q = 2\pi/\beta_q$ is the space-charge wavelength and n is an integer.

If we had considered a beam with confined flow as discussed in Sect. 9.3, we should have

$$J_z = -\frac{\omega\rho_0 v_z}{\omega - \beta v_0}$$

by using (9.11*b*) and (9.12). In this case

$$J_{zs} = \frac{\omega\rho_0 v_{zs}}{\omega_q} \qquad J_{zf} = -\frac{\omega\rho_0 v_{zf}}{\omega_q}$$

If v_z varies according to $\cos \beta_q z$, then J_z will vary as $\sin \beta_q z$ for a beam with confined flow. The different behavior of the two beam models arises because of the zero ac volume space charge density in the unfocused beam, a condition that can exist because radial ac motion of the electrons is permitted. The axial current for the two beam models is given by

$$J_z = -\rho_0 v_z + \rho v_0 \qquad J_z = -\rho_0 v_z$$

The space charge ρ changes the relationship between J_z and v_z from $J_z = -\rho_0 v_z$ to $J_z = -\rho_0 v_z \omega/(\omega - \beta v_0)$, which in turn produces the difference in z variation of the current amplitude. In some klystrons where a high power and long drift spaces are used, magnetic focusing is employed. In this case the confined-flow-beam model would be the appropriate one to use. For the beam with confined flow, the axial current density at $r = 0$ would be found to be

$$J_z = \frac{j\eta\rho_0}{v_0}\frac{\omega}{\omega_q} MV_g \sin \beta_q z\, e^{-j\beta_0 z} \tag{9.60}$$

The peak current density is a factor ω/ω_q greater than the maximum current density for the unfocused beam as given by (9.57). This suggests that the confined-flow beam is superior. This is usually not the case because the surface current J_s for the unfocused beam may contribute in a very substantial way to the total ac axial current, as the following analysis will show.

The surface current density is given by (9.36) and (9.38). For the two space-charge waves, we obtain

$$J_{ss} = -\frac{\eta\rho_0 v_0}{(\omega - \beta_s v_0)^2} E_{rs} = -\frac{\eta\rho_0 v_0}{\omega_q^2} E_{rs}$$

$$J_{sf} = -\frac{\eta\rho_0 v_0}{\omega_q^2} E_{rf}$$

When we combine (9.29a) and (9.29b), we find

$$E_r = \frac{j\beta}{p^2}\frac{\partial E_z}{\partial r} = \frac{j\beta}{p} C_1 I_1(pr) \approx jC_1 I_1(\beta_0 r)$$

where $C_1 I_0(pr) = E_z$. Consequently,

$$J_s = -\frac{j\eta\rho_0 v_0}{\omega_q^2} I_1(\beta_0 a)\left(C_{1f} e^{-j\beta_f z} + C_{1s} e^{-j\beta_s z}\right)$$

Now $E_{zs} = -(\omega_q/j\eta)v_{zf}$ and $E_{zf} = (\omega_q/j\eta)v_{zf}$; so we have $C_{1s} = -(\omega_q/j\eta)v_{zf}$, $C_{1f} = (\omega_q/j\eta)v_{zf}$, where $2v_{zf} = -(\eta/v_0)MV_g$. The final expression for J_s becomes

$$J_s = \frac{j\eta\rho_0}{\omega_q} MV_g I_1(\beta_0 a)\sin\beta_q z\, e^{-j\beta_0 z} \tag{9.61}$$

We shall now compare the relative contributions to the total axial current. The total surface current is given by

$$I_s = 2\pi a \frac{j\eta\rho_0}{\omega_q} MV_g I_1(\beta_0 a)\sin\beta_q z\, e^{-j\beta_0 z} \tag{9.62a}$$

and the total volume current flowing in the axial direction is

$$I_z = \pi a \frac{2\eta\rho_0}{\beta_0 v_0} MV_g I_1(\beta_0 a)\cos\beta_q z\, e^{-j\beta_0 z} \tag{9.62b}$$

where we have used the result

$$\int_0^a I_0(\beta_0 r)2\pi r\, dr = \frac{2\pi a}{\beta_0} I_1(\beta_0 a)$$

The ratio of the peak amplitudes is

$$\frac{(I_s)_{\max}}{(I_z)_{\max}} = \frac{\omega}{\omega_q} \tag{9.63}$$

This gives the very interesting result that the total surface current in the case of an unfocused beam is a factor ω/ω_q larger than the total volume current. In fact, the total surface current given by (9.62a) is equal to the total volume current in a small-radius confined-flow beam, that is, $\pi a^2 J_z$,

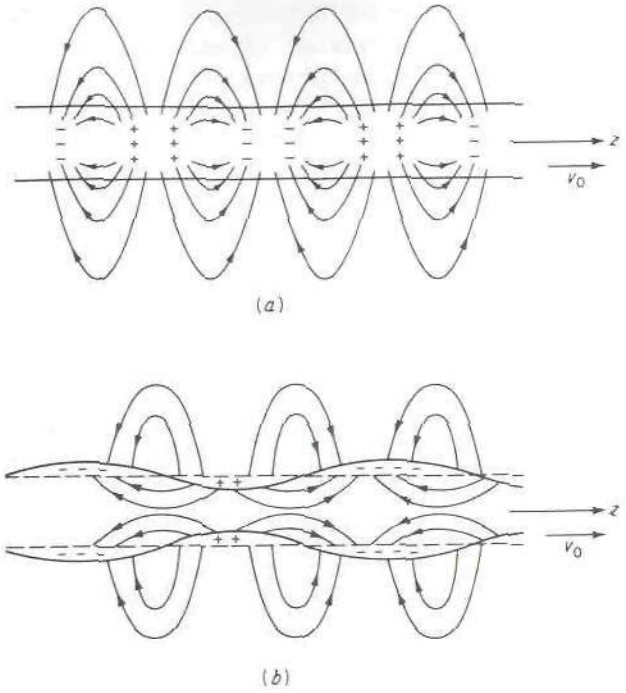


FIGURE 9.9
Electric field lines associated with ac space-charge bunching in (*a*) a beam with confined flow, (*b*) an unfocused beam.

where J_z as given by (9.60) for confined flow has a peak amplitude equal to that of I_s if we replace $I_1(\beta_0 a)$ by $\beta_0 a/2$. We therefore conclude that both types of beams are about equally efficient, at least for short drift spaces, where beam spreading would not be important. It is apparent that ac space-charge bunching is an important mechanism in the production of high-density ac currents in the velocity-modulated beam. For confined flow the ac space-charge bunches are formed within the beam, and for the unfocused beam the ac space-charge bunches appear on the beam surface in the form of a rippled boundary. A sketch of the electric field associated with the two beam models is given in Fig. 9.9. The positive charge shown is only an equivalent charge that accounts for a net migration of electrons out of the region, leaving a net negative charge density less than ρ_0, which can be viewed as a superposition of a small positive charge density on the constant dc background density $-\rho_0$.

In judging the relative amplitudes of the volume current and surface current in an unfocused beam, the ratio ω_q/ω must be known. Usually ω_q does not differ by more than a factor of $\frac{1}{2}$ or so from the plasma radian frequency ω_p. In Fig. 9.10, $f_p = \omega_p/2\pi$ is plotted as a function of the beam

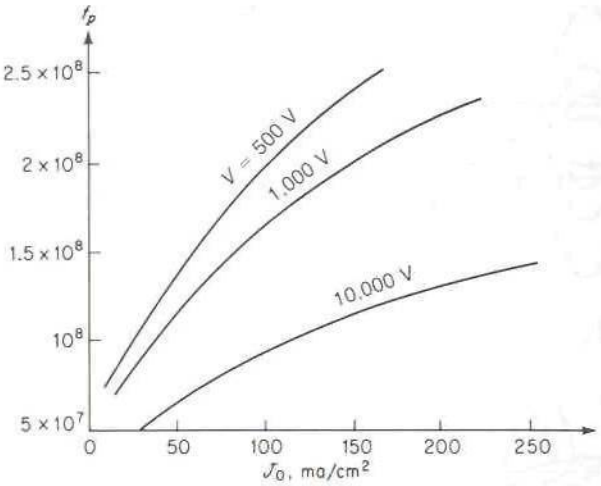


FIGURE 9.10
Plasma frequency as a function of beam current density.

current density J_0 in milliamperes per square centimeter for several values of beam-accelerating voltage V. Note that f_p is proportional to $J_0^{1/2}/V^{1/4}$.

9.7 TWO-CAVITY KLYSTRON

A schematic illustration of a typical two-cavity klystron amplifier is shown in Fig. 9.11. The first cavity is excited by the input signal, which can be coupled to the cavity by a coaxial-line loop or a waveguide aperture. The first cavity acts as the buncher and velocity-modulates the beam. The

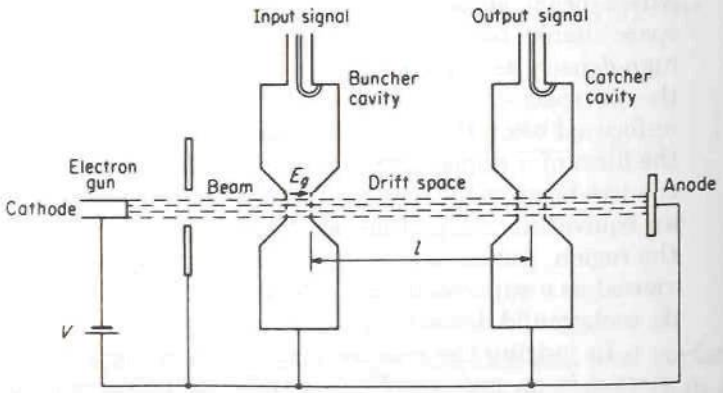


FIGURE 9.11
The two-cavity klystron amplifier.

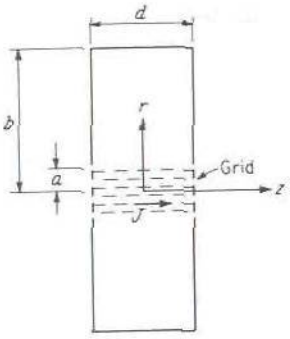
second cavity is separated from the buncher by a drift space of length l, which should ideally be chosen so that the ac current at the second (sometimes called the catcher) cavity is a maximum. The second cavity is thus excited by the ac signal impressed on the beam in the form of a velocity modulation with a resultant production of an ac current. The ac current on the beam is such that the level of excitation of the second cavity is much greater than that in the buncher cavity, and hence amplification takes place. The output signal is taken from the second cavity. If desired, a portion of the amplified output can be fed back to the buncher cavity in a regenerative manner to obtain self-sustained oscillations.

One form of klystron analysis begins with an assumed lumped-parameter equivalent circuit for the output cavity and evaluates the current flowing in this equivalent circuit from an electron beam passing through the cavity by calculating the rate of change with time of the charge induced on the grids at the center of the cavity. This analysis gives a correct picture of klystron behavior, but it fails to illustrate the mechanism of the electromagnetic field interaction with the beam as it actually takes place within the cavity. A more satisfactory approach is to begin with a field analysis that will eventually lead to an equivalent circuit and the basis for a circuit-type analysis of the problem. This is the approach presented below. We first evaluate the field set up in a cavity by the passage of an ac current through the cavity. This leads to an equivalent circuit for the cavity in the vicinity of one of the resonant frequencies. The next step is to evaluate the response of the cavity (or its equivalent circuit) to the passage of a velocity-modulated electron beam on which the ac current is in the form of a propagating current wave. This leads to a definition of the beam coupling coefficient, which is a measure of how effective the modulated electron beam is in exciting a response in the cavity. The third step, which we have presented in Sec. 9.6, is the evaluation of the ac current produced on an initially unmodulated electron beam when it passes through a cavity in which an oscillating electric field exists. These three phases of the analysis substantially provide the complete picture of the operation of a klystron.

Excitation of a Cylindrical Cavity

In a klystron cavity it is desirable to have a small grid spacing in order to make the beam coupling parameter $M = [\sin(\beta_0 d/2)]/(\beta_0 d/2)$ close to unity. The transit angle $\beta_0 d = \omega d/v_0$ should be kept small. In addition, a high cavity Q is desired, and this leads to the use of a reentrant cavity. The analysis of the modes in a cavity of this configuration is difficult, and therefore we shall consider instead an ordinary cylindrical cavity. The principal features involved in the excitation of the latter type of cavity by a modulated electron beam are the same as for the reentrant-type cavity.

The cylindrical cavity to be studied is illustrated in Fig. 9.12. The cavity radius is b, and the cavity length is d. Two small cylindrical holes are



8**FIGURE 9.12**
Cylindrical cavity excited by an axial current.

cut in the center and replaced by grids to allow an electron beam to pass through. The beam radius is a, and is considered very small compared with the cavity radius b. We shall first study the excitation of this cavity by an axial ac current of the form

$$J_z = Je^{j\omega t} \qquad 0 \le r \le a \tag{9.64}$$

Later on we shall consider a traveling-wave current $Je^{j\omega t - j\beta_0 z}$ of the type existing on a velocity-modulated beam and shall find that, for the latter, the cavity response is modified by the beam coupling parameter M.

In view of the uniformity of the current in the z direction and the axial symmetry, only TM_{0m0} modes are excited. These have E_r, E_ϕ, H_r, and H_z equal to zero. It is convenient to introduce the vector potential A_z, which is a solution of

$$\nabla^2 A_z + k_0^2 A_z = \begin{cases} -\mu_0 J & 0 \le r \le a \\ 0 & r > a \end{cases} \tag{9.65}$$

At $r = b$ we have $A_z = 0$, so that E_z will vanish on the boundary. From A_z we obtain

$$E_z = -j\omega A_z \tag{9.66}$$

$$H_\phi = -\frac{1}{\mu_0}\frac{\partial A_z}{\partial r} \tag{9.66b}$$

since there is no z or ϕ variation.

The natural modes of the cavity are solutions of the equation

$$\nabla^2 A_{z,0m0} + k_{0m0}^2 A_{z,0m0} = 0 \tag{9.67}$$

where $k_{0m0} = \omega_{0m0}(\mu_0\epsilon_0)^{1/2}$ and ω_{0m0} is the resonant frequency of the TM_{0m0} mode. With no z or ϕ variation, ∇^2 becomes

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial}{\partial r}$$

and the solutions to (9.67) are Bessel functions. That is,

$$A_{z,0m0} = C_m J_0\left(\frac{p_{0m} r}{b}\right) \tag{9.68}$$

where p_{0m} is chosen so that $J_0(p_{0m}) = 0$, C_m is an arbitrary constant, and for no axial variation $k_{0m0} = p_{0m}/b$ (Sec. 7.4).

For a solution to (9.65) we may choose]

$$A_z = \sum_{m=1}^{\infty} C_m J_0\left(\frac{p_{0m} r}{b}\right) \tag{9.69}$$

since the Bessel functions are analogous to the sine and cosine functions and may be used as such in a Fourier series expansion of the vector potential. When we substitute (9.69) into (9.65), we obtain

$$\sum_{m=1}^{\infty} C_m (k_{0m0}^2 - k_0^2) J_0\left(\frac{p_{0m} r}{b}\right) = \mu_0 J \qquad 0 \le r \le a \tag{9.70}$$

since (9.69) is a solution of (9.67).

The following orthogonality property holds:

$$\int_0^b J_0\left(\frac{p_{0m} r}{b}\right) J_0\left(\frac{p_{0n} r}{b}\right) r\, dr = \begin{cases} 0 & n \neq m \\ \dfrac{b^2}{2} J_1^2(p_{0m}) & n = m \end{cases} \tag{9.71}$$

Thus, if we multiply both sides of (9.70) by $rJ_0(p_{0m} r/b)$ and integrate, we obtain

$$\begin{aligned} C_m &= \frac{2\mu_0}{(k_{0m0}^2 - k_0^2) b^2 J_1^2(p_{0m})} \int_0^a J J_0\left(\frac{p_{0m} r}{b}\right) r\, dr \\ &= \frac{2\mu_0 J a J_1(p_{0m} a/b) c^2}{p_{0m} b J_1^2(p_{0m})(\omega_{0m0} - \omega)(\omega_{0m0} + \omega)} \end{aligned} \tag{9.72}$$

after replacing k_0 by ω/c, and similarly for k_{0m0}. This equation holds for all values of m. We see immediately from this expression that only for those modes for which $\omega \approx \omega_{0m0}$ will the excitation amplitude be large. In addition, we see that, if $\omega = \omega_{0m0}$ for a particular value of m, C_m becomes infinite. An infinite response occurs because the cavity is ideal and is driven at one of its natural resonant frequencies. A practical cavity has a finite Q and will not respond with an infinite amplitude. As shown in Chap. 7, the effect of a finite Q is to replace the resonant frequency by

$$\omega_{0m0}\left(1 + \frac{j}{2Q_{0m0}}\right)$$

where Q_{0m0} is the Q of the TM_{0m0} mode. The unloaded cavity Q is given

by (7.57) as

$$Q_{0m0} = \frac{p_{0m}c}{(1 + b/d)\omega_{0m0}\delta_s} \tag{9.73}$$

Thus the excitation amplitudes for a cavity with finite Q are given by

$$C_m = -\frac{2\mu_0 aJc^2 J_1(p_{0m}a/b)}{p_{0m}bJ_1^2(p_{0m})(\omega_{0m0} + \omega)(\omega - \omega_{0m0} - j\omega_{0m0}/2Q_{0m0})} \tag{9.74}$$

If we choose $\omega = \omega_{010}$, then C_1 will be large and all the other C_m will be small. In this case

$$A_z = C_1 J_0\left(\frac{p_{01}r}{b}\right) = -\frac{2j\mu_0 aJc^3 J_1(p_{01}a/b)Q}{b^2 J_1^2(p_{01})\omega^3} J_0\left(\frac{p_{01}r}{b}\right) \tag{9.75}$$

where ω and Q now refer to the TM_{010} mode. Note that A_z is proportional to the Q, and hence a high Q is desirable. If the cavity is coupled to an external load, we must replace Q by the loaded Q, Q_L.

The total ac current is $I = \pi a^2 J$. Also, since $a \ll b$, we can replace $J_1(p_{01}a/b)$ by $p_{01}a/2b$. In the vicinity of the resonant frequency ω_{010}, that is, for $\omega = \omega_{010} + \Delta\omega$, we find that the electric field E_z is given by

$$E_z = -j\omega A_z = \frac{jIJ_0(p_{01}r/b)}{2\pi b^2 \epsilon_0 J_1^2(p_{01})(\Delta\omega - j\omega_{010}/2Q)} \tag{9.76}$$

If we introduce an equivalent voltage V_0 as the line integral of E_z across the cavity gap at $r = 0$, that is, $V_0 = -E_z d$, we may define an admittance Y_c for the cavity as follows:

$$Y_c = \frac{I}{V_0} = \frac{2\pi b^2}{d}\epsilon_0 J_1^2(p_{01})\left(j\Delta\omega + \frac{\omega_{010}}{2Q}\right) \tag{9.77}$$

For a lumped-parameter LCG_0 circuit with resonant frequency $\omega_{010} = (LC)^{-1/2}$ as in Fig. 9.13, we have

$$Y_{\text{in}} = G_0 + 2jC\Delta\omega = G_0\left(1 + j\frac{2\Delta\omega}{\omega_{010}}Q\right)$$

where $Q = \omega_{010}C/G_0$. Comparison with (9.77) shows that the equivalent cavity conductance G_0 is given by (note that $\epsilon_0 = Y_0/c$)

$$G_0 = \pi b^2 Y_0 J_1^2(p_{01})\frac{\omega_{010}}{dQc} \tag{9.78}$$

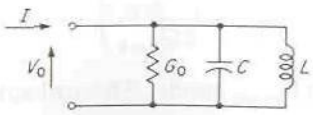


FIGURE 9.13
Equivalent circuit of excited cavity with no loading.

The above equivalent circuit thus seems a possible circuit to represent the cavity in the vicinity of the resonant frequency ω_{010} of the TM_{010} mode. However, we must show that it correctly accounts for the properties of the cavity even in the absence of the beam current I.

For the TM_{010} mode, the energy stored in the electric field is given by [we assume $E_z = J_0(p_{01}r/b)$]

$$W_e = \frac{\epsilon_0}{4}\int_0^d\int_0^{2\pi}\int_0^b J_0^2\left(\frac{p_{01}r}{b}\right) r\,dr\,d\phi\,dz = \frac{\pi\epsilon_0}{4}b^2\,dJ_1^2(p_{01})$$

The corresponding voltage across the cavity at $r = 0$ is $V = E_z d = d$. Since $Q = 2\omega W_e/P_l$, the power loss in the cavity is given by

$$P_l = \frac{2\omega W_e}{Q}$$

If we define a conductance G so that $P_l = \frac{1}{2}V^2G$, we find that

$$G = \frac{2P_l}{V^2} = \frac{4\omega W_e}{V^2Q} = \frac{\pi\epsilon_0 b^2\omega_{010}J_1^2(p_{01})}{dQ} = G_0$$

since $\omega = \omega_{010}$. Thus the two definitions for the cavity conductance lead to consistent results.

If required, C and L are given by $\omega_{010}C = QG_0$ and $\omega_{010}^2LC = 1$. Thus L and C can be found from the known values of Q and G_0.

Cavity Excitation by a Velocity-Modulated Beam

In a velocity-modulated unfocused beam the ac current is predominantly the ac beam surface current, $I_s = 2\pi a J_s$, where J_s is given by (9.61) since $\omega \gg \omega_q$ for typical beams. The z dependence of the current is $\sin\beta_q z\, e^{-j\beta_0 z}$ when z is measured from the output grid of the buncher cavity. The output cavity should be located a distance $l = (n + \frac{1}{2})(\lambda_q/2)$ from the input cavity so that $\sin\beta_q z$ will equal unity and I_s will have a maximum. Because $\beta_q \ll \beta_0$, we have $\beta_q d \ll 1$ and the ac current variation with z can be taken as $e^{-j\beta_0 l}e^{-j\beta_0 z}$ in the output cavity, where we have replaced z by $l + z$, so that the new origin is at the center of the output cavity. Thus the cavity is excited by a traveling-wave current.

When the ac current has a z dependence, all the TM_{0mn} modes are excited. The solutions for A_z are thus of the form

$$C_{mn}J_0\left(\frac{p_{0m}r}{b}\right)\cos\frac{n\pi z}{d}$$

We then have

$$\sum_{m=1}^{\infty}\sum_{n=0}^{\infty} C_{mn}\left[\left(\frac{p_{0m}}{b}\right)^2 + \left(\frac{n\pi}{d}\right)^2 - k_0^2\right]J_0\left(\frac{p_{0m}r}{b}\right)\cos\frac{n\pi z}{d} = \mu_0 J_z(r,z) \tag{9.79}$$

in place of (9.70). However, if $\omega \approx \omega_{010}$, only the $m = 1$, $n = 0$, or TM_{010} mode is excited with a large amplitude. To find C_{10}, we multiply (9.79) by $rJ_0(p_{0m}r/b)\cos(n\pi z/d)$ with $m = 1$, $n = 0$ and integrate over r and z to obtain

$$C_{10} = \frac{2\mu_0}{(k_{010}^2 - k_0^2)b^2 J_1^2(p_{01})} \int_0^a J_z(r) r J_0\left(\frac{p_{01}r}{b}\right) dr \int_{-d/2}^{d/2} \frac{1}{d} e^{-j\beta_0 z}\, dz$$

This latter result shows that the excitation amplitude is modified by the factor

$$\int_{-d/2}^{d/2} \frac{1}{d} e^{-j\beta_0 z}\, dz \approx \frac{\sin(\beta_0 d/2)}{\beta_0 d/2} = M$$

which is the beam coupling factor. The integral over r may be replaced by aJ_s without the factor $e^{-j\beta_0 z}$ since $J_z = 0$ except at $r = a$ and $J_0(p_{01}a/b) \approx 1$ for $a \ll b$. Hence, for $\omega = \omega_{010} + \Delta\omega$, we obtain for E_z, at $r = 0$,

$$E_z = \frac{jM(jI_1)e^{-j\beta_0 l}}{2\pi b^2 \epsilon_0 J_1^2(p_{01})(\Delta\omega - j\omega_{010}/2Q)} \tag{9.80}$$

where I_1 is obtained from (9.62a) and is

$$I_1 = 2\pi a \frac{\eta\rho_0}{\omega_q} M I_1(\beta_0 a) V_g \approx \pi a^2 \epsilon_0 \frac{\omega_p^2}{\omega_q} \beta_0 M V_g \tag{9.81}$$

since for $\beta_0 a \ll 1$ we have $I_1(\beta_0 a) \approx \beta_0 a/2$ and are assuming that an optimum length l of drift space is employed so that $\sin \beta_q l$ equals unity. If we compare (9.80) with (9.76), we find that the only change in E_z is replacing the current I by I_1 and multiplying by the beam coupling coefficient M and an irrelevant phase factor $e^{-j\beta_0 l}$.

The voltage developed across the cavity is modified by the factor M. Consequently, the effective current that flows in the equivalent circuit to product the voltage V is

$$I_e = MI_1 \tag{9.82}$$

from which it is seen that the term beam coupling coefficient is clearly appropriate. For good coupling between the beam and the cavity a short transit time $\beta_0 d$ is required.

When an external load is coupled to the output cavity, it can be represented by an additional conductance G_L in shunt with G_0. The equivalent circuit of the output cavity is shown in Fig. 9.14a, where $I_e = MI_1$ is the equivalent current that flows in the circuit. The power supplied to the external load G_L is

$$P_0 = \frac{1}{2} \frac{|I_e|^2 G_L}{(G_0 + G_L)^2} \tag{9.83}$$

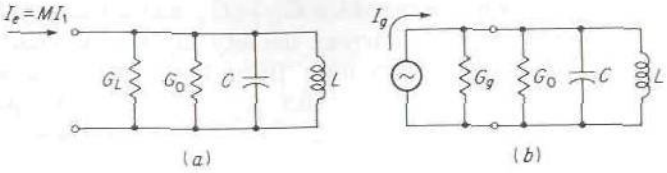


FIGURE 9.14
(a) Equivalent circuit of output cavity; (b) equivalent circuit of input cavity.

A similar equivalent circuit may be assumed for the input cavity, as in Fig. 9.14b. For identical cavities the cavity conductance is G_0 for both. If we assume that the beam produces negligible loading on the input cavity, the total conductance in the equivalent circuit of the input cavity is $G_0 + G_g$, where G_g is the equivalent conductance of the signal source.†

The input power to the buncher cavity at resonance is

$$P_{\text{in}} = \frac{1}{2} \frac{|I_g|^2 G_0}{(G_g + G_0)^2} \tag{9.84}$$

which results in a voltage

$$V_g = \frac{I_g}{G_0 + G_g} = \left(\frac{2P_{\text{in}}}{G_0} \right)^{1/2} \tag{9.85}$$

developed across G_0. This is also the velocity-modulating voltage. We can now evaluate the power gain, or amplification, of the klystron. Combining (9.81), (9.83), and (9.85), we obtain

$$\begin{aligned} A = \frac{P_0}{P_{\text{in}}} &= \frac{|I_e|^2 G_L}{|I_g|^2 G_0} \left(\frac{G_0 + G_g}{G_0 + G_L} \right)^2 \\ &= \left(\frac{\pi a^2 \epsilon_0 \beta_0 \omega_p^2}{\omega_q} \right)^2 \frac{M^4 G_L}{G_0 (G_0 + G_L)^2} \\ &= \left(\pi a^2 Y_0 \beta_0 \frac{\omega_p}{c} \frac{\omega_p}{\omega_q} \right)^2 \frac{M^4 G_L}{G_0 (G_0 + G_L)^2} \end{aligned} \tag{9.86}$$

†The conductance representing the beam loading on the input cavity is given by $G_b = 1/2(I_0/V)M[M - \cos(\beta_0 d/2)]$ and is easily taken into account by adding it to $G_0 + G_g$ when it is not negligible. Note that I_0 is the dc beam current and V the dc accelerating voltage. See Spangenberg, *op. cit.*, chap. 17.

As an example consider $G_L = G_0$ and a beam radius $a = 0.2$ cm If we also assume a beam current density of 100 mA/cm^2 and an accelerating voltage of 1,000 V, we find, from Fig. 9.10, that $\omega_p = 1.02 \times 10^9$. For $f = 10^{10}$, $\omega = 6.28 \times 10^{10}$ and $\beta_0 = \omega/v_0 = 33.6$ rad/cm. Hence $\beta_0 a = 6.72$. Since $\omega_p \ll \omega$, we have $\beta a \approx \beta_0 a$, and Fig. 9.6 then shows that $\omega_q \approx 0.7\omega_p$. Using these data, the power amplification A is found to be $0.094(Y_0/G_0)^2 M^4$. To evaluate G_0, we make use of (9.78) and (9.73). We have $p_{01} = 2.405$, and since $p_{01} = k_{010} b$, we get

$$b = \frac{2.405c}{6.28 \times 10^{10}} = 1.15 \text{ cm}$$

To keep $\beta_0 d$ small, we must choose d very small. If we take $d = 0.05$ cm, $\beta_0 d = 1.68$ and $M^4 = 0.62$. It would be desirable to make d even smaller, but then the Q, and hence G_0, become small for the type of cavity we are considering. For a copper cavity we find from (9.73) that $Q = 785$, which is not very large. If a reentrant-type cavity were used, a Q about 10 times larger could be obtained. Using (9.78) gives $G_0 \approx 0.06 Y_0$. The power amplification is thus 16.2, or 12 dB. Considerably higher gain would be obtained by using a reentrant-type cavity since d could then be made smaller and still a high unloaded Q maintained. However, even with the nonoptimum cavity that we have considered, the gain is quite good. For the particular example we have evaluated, the beam loading conductance G_b is very small compared with the cavity conductance G_0. For a more efficient cavity with a much higher unloaded Q, the cavity conductance G_0 would be much smaller and the beam loading conductance G_b might not then be negligible.

In order to obtain greater power gain than can be obtained from a two-cavity klystron, multicavity tubes are used. The gain increases exponentially with the number of cavities employed. In the multicavity klystron the first cavity is again used to provide the initial velocity modulation of the beam. The last cavity in the chain is used as the output cavity. The intermediate cavities are kept unloaded by any external circuits and used to increase the modulation and hence the ac current on the beam. Power gains of 50 to 60 dB can be achieved with multicavity klystrons.

9.8 REFLEX KLYSTRON

The reflex klystron is an oscillator tube with a built-in feedback mechanism. It uses the same cavity for bunching and for the output cavity. A sketch of the reflex klystron is given in Fig. 9.15. The operation is as follows: If we assume an initial ac field in the cavity, the beam will be velocity-modulated as it passes through the cavity. Upon entering the drift space, the beam is decelerated and reversed (reflected) by the large dc field set up by the repeller or reflector electrode at potential $-V_r$. Thus the beam is made to pass through the cavity again, but in the opposite direction. By proper choice of the reflector voltage V_r, the beam can be made to pass through the

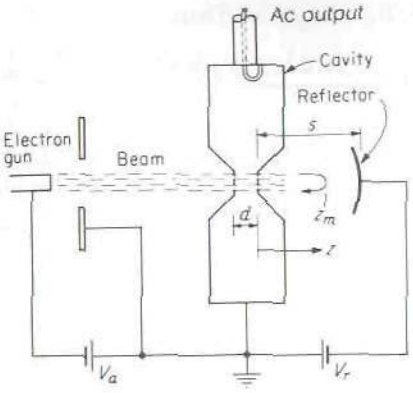


FIGURE 9.15
The reflex klystron.

cavity on its return flight when the ac current phase angle is such that the field excited in the cavity by the returning beam adds in phase with the initial modulating field. The feedback is then positive, and oscillations will build up in amplitude until the system losses and nonlinear effects prevent further buildup.

When the velocity-modulated beam enters the drift space, it is subjected to a constant decelerating field V_r/s, where s is the cavity output grid–reflector spacing. As a result, the beam propagation constant $\beta_0 = \omega/v_0$ is gradually reduced to zero, and then increased back up to $-\beta_0$. The total phase change undergone by the ac current on the beam will be given by

$$\theta = 2\int_0^{z_m} \beta_0(z)\, dz$$

where z_m is the maximum distance an electron can penetrate into the drift space. We can evaluate θ in terms of the transit time T for an electron to return to the cavity. We have, in the drift space,

$$\frac{dv_0(z)}{dt} = -\eta \frac{V_r}{s}$$

which integrates to $v_0(z) = v_0 - \eta t V_r/s$. Hence

$$v_0(z_m) = v_0 - \eta \frac{V_r}{s}\frac{T}{2} = 0$$

which gives $T/2 = v_0 s/\eta V_r$. The return time is equal to $T/2$ also, so that

$$\theta = \omega T = \frac{2v_0 s\omega}{\eta V_r} \tag{9.87}$$

If we let V_g be the accelerating-gap ac voltage, the ac beam current reflected back through the cavity is given by (9.62a) when $\beta_0 z$ is replaced

by θ and $\beta_q z$ by $\beta_q\theta/\beta_0 = \omega_q\theta/\omega$. Thus

$$I_s = \frac{j\pi a^2\eta\rho_0\beta_0 M V_g}{\omega_q}\sin\frac{\omega_q\theta}{\omega}e^{-j\theta}$$

where we have approximated $I_1(\beta_0 a)$ by $\beta_0 a/2$. In a reflex klystron θ is usually quite small, so that $\sin(\omega_q\theta/\omega)$ may be replaced by $\omega_q\theta/\omega$. The effective current for excitation of the cavity is $I_e = I_s M$, and is given by

$$I_e = \frac{\pi a^2\eta\rho_0 M^2 V_g}{v_0}\theta e^{j(\pi/2-\theta)} = \frac{I_0}{V_a}\frac{V_g}{2}M^2\theta(\sin\theta + j\cos\theta) \quad (9.88)$$

where I_0 is the total dc beam current $\pi a^2\rho_0 v_0$ and V_a is the accelerating voltage from which $v_0^2 = 2\eta V_a$.

The ac electronic admittance of the beam is defined by

$$Y_e = \frac{I_e}{V_g} = \frac{I_0}{V_a}\frac{M^2}{2}\theta(\sin\theta + j\cos\theta) \quad (9.89)$$

The equivalent circuit of the reflex klystron consists of the electronic admittance Y_e in shunt with the equivalent circuit of the loaded cavity, as in Fig. 9.16. Oscillations can take place when the net conductance is less than zero, or more specifically when

$$Y_e + (G_L + G_0)\left(1 + j\frac{2\Delta\omega}{\omega_{010}}Q_L\right) = 0 \quad (9.90)$$

where Q_L is the loaded Q of the cavity. Since θ is a function of the reflector voltage, as given by (9.87), oscillations depend on an appropriate choice of

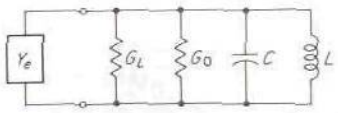


FIGURE 9.16
Equivalent circuit for a reflex klystron.

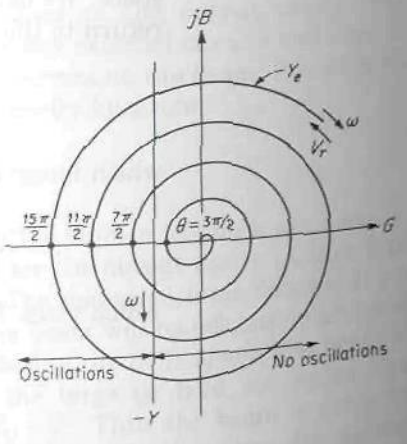


FIGURE 9.17
Admittance diagram for a reflex klystron.

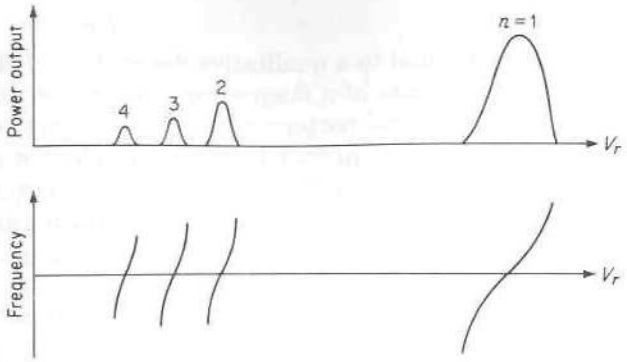


FIGURE 9.18
Reflex-klystron tuning curves.

V_r. In Fig. 9.17 we have plotted the admittance Y_e in polar form as a function of θ. Note that $|Y_e|$ increases with θ. On the same plane we have plotted the negative cavity admittance

$$-Y = -(G_0 + G_L)\left(1 + j\frac{2\,\Delta\omega}{\omega_{010}}Q_L\right)$$

which is a straight line parallel to the jB axis at $G = -(G_0 + G_L)$, provided we assume G_0, G_L, Q_L independent of ω in the vicinity of the resonant frequency ω_{010}. The construction shows that oscillations are possible for θ in the vicinity of $3\pi/2, 7\pi/2$, etc., since in this region $G_e + G_0 + G_L < 0$. In addition, oscillations will take place for a range of values of V_r about the points that make $\theta = 3\pi/2 + 2n\pi$. Each value of n gives a mode of oscillation. In typical klystrons as many as seven or more modes of oscillation can be obtained. For stable oscillations $Y_e + Y = 0$, and consequently the frequency of oscillation varies as V_r is varied to tune across a given mode. Typical tuning curves giving power output and frequency as a function of reflector voltage are shown in Fig. 9.18. Physically, the various modes arise because of the increased transmit time for electrons into the drift space when V_r is reduced. Oscillations occur when the transit time T equals $(\frac{3}{4} + n)f^{-1}$ or $\frac{3}{4} + n$ ac periods since the ac current has the proper phase under these conditions.

Commercially available reflex klystrons range from small-size units producing 100 mW of power up to units capable of delivering several watts of power under continuous operation. Klystron amplifiers employing two or more cavities are available in a size range from units capable of handling a few hundred milliwatts up to several hundred kilowatts of amplified output power.

9.9 MAGNETRON

This section is devoted to a qualitative description of the magnetron oscillator. The basic structure of a magnetron is a number of identical resonators arranged in a cylindrical pattern around a cylindrical cathode, as shown in Fig. 9.19. A permanent magnet is used to produce a strong magnetic field normal to the cross section. The anode is kept at a high positive voltage V_a relative to the cathode. Electrons emitted from the cathode are accelerated toward the anode block, but the presence of the magnetic field B_0 produces a force $-ev_r B_0$ in the azimuthal direction which causes the electron trajectory to be deflected in the same direction. If the cathode radius is a and the anode radius is b, the potential at any radius r is $V(r) = V_a[\ln(r/a)]/[\ln(b/a)]$. The velocity of an electron at this radius is given by

$$v(r) = [2\eta V(r)]^{1/2}$$

The electron can execute circular motion, at the radius r, about the cathode if the outward centrifugal force mv^2/r and the radial electric field force $-eE_r = eV_a/[r\ln(b/a)]$ are exactly balanced by the inward magnetic force $ev(r)B_0$. For circular motion at radius r, we therefore have

$$\frac{mv^2}{r} + \frac{eV_a}{r\ln(b/a)} = evB_0 \tag{9.91a}$$

or since $v = \omega_e r$, where ω_e is the electron's angular velocity,

$$\omega_e^2 - \eta B_0 \omega_e + \frac{\eta V_a}{r^2 \ln(b/a)} = 0 \tag{9.91b}$$

For later reference, we solve (9.91) for the cathode-anode accelerating

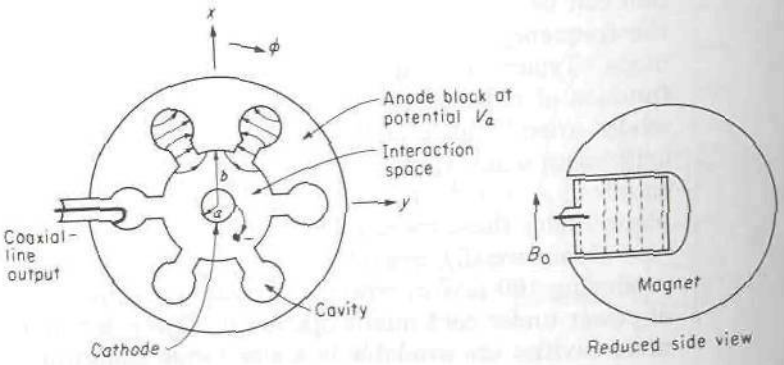


FIGURE 9.19
A multicavity magnetron.

voltage V_a:

$$V_a = \left(\omega_e r^2 \ln \frac{b}{a}\right)\left(B_0 - \frac{\omega_e}{\eta}\right) \tag{9.92}$$

This value of V_a will permit an electron to execute circular motion at a radius r and with an angular frequency ω_e. If now there is present an ac electromagnetic field that propagates in the azimuthal direction with a phase velocity equal to the electron velocity $\omega_e r$, strong interaction between the field and the circulating electron cloud can take place. The possibility of this type of electromagnetic field being present is discussed below.

The multicavity magnetron is a periodic structure in the azimuthal, or ϕ, direction. If there are N cavities, the period in ϕ is $2\pi/N$. According to Floquet's theorem, each field component can be expanded in the form

$$\begin{aligned}\psi(r,\phi) &= \sum_{n=-\infty}^{\infty} e^{-j\beta\phi - j2n\pi\phi/p}\psi_n(r) \\ &= \sum_{n=-\infty}^{\infty} \psi_n(r)e^{-j\beta\phi - jnN\phi}\end{aligned} \tag{9.93}$$

where the period $p = 2\pi/N$. But since the structure closes on itself, $\psi(r, 2\pi) = \psi(r, 0)$. The only possible values of β that will make $\beta 2\pi$ equal a multiple of 2π are

$$\beta_m = m \qquad m = 0, \pm 1, \pm 2, \ldots \tag{9.94}$$

With the value of β specified, a corresponding frequency ω becomes specified, say ω_m, which is the resonant frequency for the mth mode. In other words, when $\omega = \omega_m$, we obtain a value m for β_m. Thus a typical field component will have the form

$$\psi_m(r,\phi)e^{j\omega_m t} = \sum_{n=-\infty}^{\infty} \psi_n(r)e^{-j(m+nM)\phi + j\omega_m t}$$

The phase velocity in the azimuthal direction ϕ for the nth spatial harmonic of the mth resonant mode is

$$v_{p,nm} = \frac{\omega_m r}{\beta_{mn}} = \frac{\omega_m r}{m + nN} \tag{9.95}$$

at the radius r; that is, angular phase velocity is ω_m/β_{mn}.

The usual mode employed in a magnetron oscillator is the π mode, where the phase change between adjacent cavities is π rad, or 180°. Each cavity with its input gap acts as a short-circuited transmission line a quarter wavelength long, and hence has a maximum electric field across the gap. For the π mode the field is oppositely directed at adjacent cavities. A sketch of the electric field lines in two cavities is given in Fig. 9.19. For the π mode,

$\beta_m \phi = m\phi$ must equal π for a change in ϕ equal to one period $2\pi/N$. Hence $m = N/2$, and the phase velocity for the nth spatial harmonic becomes

$$v_{p,nN/2} = \frac{2\omega_{N/2} r}{N(1 + 2n)} \tag{9.96}$$

In order to obtain interaction between the electron cloud and one of the spatial harmonics at a particular radius r, we must choose V_a so that $\omega_e r = v(r) = v_{p,nN/2}$, or

$$\omega_e = \frac{2\omega_{N/2}}{N(1 + 2n)} \tag{9.97}$$

The required voltage V_a to obtain synchronism between the electron cloud and the ac field may be found from (9.92). If we choose a value of r midway between the cathode and anode, that is, $r = (b + a)/2$, and note that in typical magnetrons b/a is small enough so that $\ln(b/a) \approx 2(b - a)/(b + a)$, we obtain

$$V_a = \frac{\omega_m}{m + nN} \frac{b^2 - a^2}{2} \left(B_0 - \frac{\omega_m/\eta}{m + nN} \right) \tag{9.98a}$$

in general, and

$$V_a = \frac{2\omega_{N/2}}{N(1 + 2n)} \frac{b^2 - a^2}{2} \left[B_0 - \frac{2\omega_{N/2}}{\eta N(1 + 2n)} \right] \tag{9.98b}$$

for the π mode, where $m = N/2$.

From a physical viewpoint the synchronism between the electron cloud and the nth spatial harmonic and the ac field implies that those electrons located in the field where E_ϕ acts to slow down the electrons will give up energy to the field. As the electrons slow down they move radially outward [see (9.91)], and eventually are intercepted by the anode. Electrons that are accelerated by the ac field move in toward the cathode until they get into a proper phase relationship such as to give up energy to the field. When the latter happens, they begin to slow down and spiral out toward the anode. Thus the only electrons that are lost from the interaction space are those that have given up a net amount of energy to the ac field.

The ac power may be coupled out from one of the cavities by a coaxial-line loop as shown in Fig. 9.19 or by means of a waveguide.

9.10 *O*-TYPE TRAVELING-WAVE TUBE

The ordinary, or *O*-type, traveling-wave-tube employs a magnetically focused electron beam and a slow-wave structure such as a helix, discussed in Chap. 8. The electron-beam velocity is adjusted to be approximately equal to

the phase velocity for an electromagnetic wave propagating along the helix. Under these conditions a strong interaction between the beam and the field can take place. From another viewpoint we can consider the presence of the slow-wave circuit to modify the space-charge wave-propagation constant in such a manner that it becomes complex and represents a growing wave. We shall present a more satisfactory picture of the gain mechanism after we have analyzed a particular tube configuration in detail. A full appreciation of the physical principles involved is somewhat difficult to obtain without a detailed study.

For simplicity we shall use the sheath-helix model discussed in Sec. 8.10 and an axially confined flow beam model (B_0 infinite) of the type treated in Sec. 9.3. In addition, we shall assume that the beam completely fills the region interior to the helix. This assumption is not true in practice, but we make it, nevertheless, in order to simplify the analysis. The basic principle of operation of the tube is not changed by this assumption. The traveling-wave tube is operated in an axially symmetric mode; so all field quantities will be independent of the angle ϕ.

Figure 9.20 illustrates the construction of a typical traveling-wave tube. The main components are an electron gun, a helix, a solenoid to produce the focusing field B_0, and suitable input and output ac coupling to the helix. The helix is taken to have a radius a and a pitch angle ψ. It is approximated by a cylindrical sheath with infinite conductivity along the direction of the winding and zero conductivity in the perpendicular direction.

In Sec. 8.10 it was shown that both TM and TE modes were required in order to satisfy the boundary conditions at $r = a$. However, for a beam with axially confined flow, where only a z component of ac velocity is permitted, the TE modes are not affected by the beam since these have $E_z = 0$. Hence the field components for the TE mode for $n = 0$ are those given by (8.72) in Sec. 8.10. Similarly, for $r > a$, that is, outside the helix, the field components for the TM mode are those given by (8.72) in Sec. 8.10.

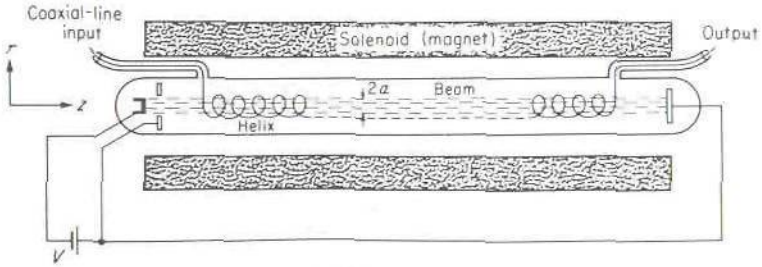


FIGURE 9.20
O-type traveling-wave tube.

Inside the helix region the TM field in the presence of the beam is that given in Sec. 9.3. However, the pertinent equations are repeated here for convenience. The vector potential A_z is a solution of (9.16a),

$$\nabla_t^2 A_z + p^2 A_z = 0$$

where

$$p^2 = (k_0^2 - \beta^2)\left[1 - \left(\frac{\omega_p}{\omega}\right)^2 \left(\frac{\beta_0}{\beta_0 - \beta}\right)^2\right] = -g^2 \tag{9.99}$$

For the present problem p^2 will turn out to be negative; so we shall replace p^2 by $-g^2$. The solution for A_z is then proportional to $I_0(gr)$. Since E_z is proportional to A_z, we can choose

$$E_z = a_0 I_0(gr) e^{-j\beta z}$$

where a_0 is an amplitude constant. The field components E_r and H_ϕ are readily found from Maxwell's equations; i.e.,

$$E_r = \frac{j\beta}{\beta^2 - k_0^2} \frac{\partial E_z}{\partial r} \qquad H_\phi = \frac{k_0}{\beta} Y_0 E_r$$

Thus we can write the following expressions for the fields in the two regions for the $n = 0$, or axially symmetric, case:

For TE modes,

$$H_z = c_0 I_0(hr) e^{-j\beta z}$$

$$H_r = \frac{j\beta}{h} c_0 I_1(hr) e^{-j\beta z} \qquad r < a$$

$$E_\phi = -\frac{j\omega\mu_0}{h} c_0 I_1(hr) e^{-j\beta z}$$

$$H_z = d_0 K_0(hr) e^{-j\beta z}$$

$$H_r = -\frac{j\beta}{h} d_0 K_1(hr) e^{-j\beta z} \qquad r > a$$

$$E_\phi = \frac{j\omega\mu_0}{h} d_0 K_1(hr) e^{-j\beta z}$$

For the TM mode,

$$E_z = a_0 I_0(gr)e^{-j\beta z}$$

$$E_r = \frac{j\beta g}{h^2} a_0 I_1(gr)e^{-j\beta z} \qquad r < a$$

$$H_\phi = \frac{j\omega\epsilon_0 g}{h^2} a_0 I_1(gr)e^{-j\beta z}$$

$$E_z = b_0 K_0(hr)e^{-j\beta z}$$

$$E_r = -\frac{j\beta}{h} b_0 K_1(hr)e^{-j\beta z} \qquad r > a$$

$$H_\phi = -\frac{j\omega\epsilon_0}{h} b_0 K_1(hr)e^{-j\beta z}$$

where $h^2 = \beta^2 - k_0^2$.

The boundary conditions at $r = a$ for the sheath helix are given by (8.69). For the present problem they yield

$$\frac{-j\omega\mu_0}{h} c_0 I_1(ha)\cos\psi + a_0 I_0(ga)\sin\psi = 0$$

$$\frac{j\omega\mu_0}{h} d_0 K_1(ha)\cos\psi + b_0 K_0(ha)\sin\psi = 0$$

$$a_0 I_0(ga)\cos\psi + \frac{j\omega\mu_0}{h} c_0 I_1(ha)\sin\psi$$

$$= b_0 K_0(ha)\cos\psi - \frac{j\omega\mu_0}{h} d_0 K_1(ha)\sin\psi$$

and

$$c_0 I_0(ha)\sin\psi + \frac{j\omega\epsilon_0 g}{h^2} a_0 I_1(ga)\cos\psi$$

$$= d_0 K_0(ha)\sin\psi - \frac{j\omega\epsilon_0}{h} b_0 K_1(ha)\cos\psi$$

If we solve for c_0 and d_0 from the first two equations and substitute into the latter two equations, we obtain two homogeneous equations for a_0 and b_0. For a nontrivial solution the determinant must vanish. Equating the determinant to zero gives

$$g\frac{I_1(ga)}{I_0(ga)} = \frac{h^3\tan^2\psi}{k_0^2}\left[\frac{I_0(ha)}{I_1(ha)} + \frac{K_0(ha)}{K_1(ha)}\right] - h\frac{K_1(ha)}{K_0(ha)} \quad (9.100)$$

For most traveling-wave tubes the parameters are such that ga and ha are large. In this case the ratio of the Bessel functions in (9.100) approaches unity, and we obtain

$$g = 2\frac{h^3}{k_0^2}\tan^2\psi - h \tag{9.101a}$$

From (9.99) we have

$$g^2 = h^2\left[1 - \left(\frac{\omega_p}{\omega}\right)^2\left(\frac{\beta_0}{\beta_0 - \beta}\right)^2\right]$$

and hence

$$\left[1 - \left(\frac{\omega_p}{\omega}\right)^2\left(\frac{\beta_0}{\beta_0 - \beta}\right)^2\right]^{1/2} = 2\frac{h^2}{k_0^2}\tan^2\psi - 1 \tag{9.101b}$$

where $h^2 = \beta^2 - k_0^2$. The above is a sixth-degree equation in β, and cannot be solved exactly.

Since we are dealing with a slow-wave system, β^2 will be large compared with k_0^2 and $h^2 \approx \beta^2$. In addition, we can equate $k_0 \cot\psi$ to β_0 since the phase velocity of the helix in the absence of a beam is chosen equal to the beam velocity v_0. That is, $k_0 \csc\psi$ is the propagation constant for the helix, and for ψ small, $\sin\psi$ can be replaced by $\tan\psi$. We thus obtain

$$1 - \left(\frac{\omega_p}{\omega}\right)^2\left(\frac{\beta_0}{\beta_0 - \beta}\right)^2 = \left(\frac{2\beta^2}{\beta_0^2} - 1\right)^2$$

We now assume that $\beta = \beta_0(1 + \delta)$, where δ is small. With this substitution we get

$$\begin{aligned}\delta^2 - \left(\frac{\omega_p}{\omega}\right)^2 &= \delta^2(1 + 4\delta + 2\delta^2)^2 \\ &= 4\delta^6 + 16\delta^5 + 20\delta^4 + 8\delta^3 + \delta^2\end{aligned} \tag{9.102}$$

For δ small, we can drop all but the term involving the lowest power of δ. This is the cubic term, and thus

$$\delta^3 = -\frac{1}{8}\left(\frac{\omega_p}{\omega}\right)^2$$

The three solutions for δ are $\frac{1}{2}(\omega_p/\omega)^{2/3}$ multiplied by the cube roots of

-1, which are -1 and $(1 \pm j\sqrt{3})/2$. Hence

$$\delta_1 = -\frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^{2/3} \tag{9.103a}$$

$$\delta_2 = \frac{1}{4}\left(\frac{\omega_p}{\omega}\right)^{2/3}(1 - j\sqrt{3}) \tag{9.103b}$$

$$\delta_3 = \frac{1}{4}\left(\frac{\omega_p}{\omega}\right)^{2/3}(1 + j\sqrt{3}) \tag{9.103c}$$

Since ω_p/ω is small, the assumption that δ was small is justified. The corresponding propagation constants are

$$j\beta_1 = j\beta_0\left[1 - \frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^{2/3}\right] \tag{9.104a}$$

$$j\beta_2 = j\beta_0\left[1 + \frac{1}{4}\left(\frac{\omega_p}{\omega}\right)^{2/3}(1 - j\sqrt{3})\right] \tag{9.104b}$$

$$j\beta_3 = j\beta_0\left[1 + \frac{1}{4}\left(\frac{\omega_p}{\omega}\right)^{2/3}(1 + j\sqrt{3})\right] \tag{9.104c}$$

The first solution corresponds to a wave with a phase velocity slightly greater than the beam velocity. The other solutions have phase velocities slightly less than the beam velocity, and in addition $j\beta_2$ corresponds to a decaying wave, whereas $j\beta_3$ corresponds to a growing wave. The growth constant α_g is

$$\alpha_g = \beta_0\frac{\sqrt{3}}{4}\left(\frac{\omega_p}{\omega}\right)^{2/3} \tag{9.105}$$

If all three waves are present at the input, only the latter wave will predominate at the output.

There are additional solutions to the eigenvalue equation (9.101*b*). We should expect a wave propagating in the $-z$ direction, with $\beta \approx -k_0 \csc\psi \approx -\beta_0$, which is not significantly perturbed by the beam. We therefore assume that $\beta = -\beta_0(1 + \delta)$ and consider δ small. Substituting into (9.101*b*) and retaining the smallest power term in δ give

$$\delta = -\frac{1}{32}\left(\frac{\omega_p}{\omega}\right)^2$$

Hence a fourth solution is

$$j\beta_4 = -j\beta_0\left[1 - \frac{1}{32}\left(\frac{\omega_p}{\omega}\right)^2\right] \approx -j\beta_0 \tag{9.106}$$

The remaining two solutions of (9.101*b*) give values for β approximately

equal to $\pm k_0$. However, the eigenvalue equation (9.101*b*) is an approximation to the true eigenvalue equation (9.100), obtained by assuming that ga and ha are large and that β is large compared with k_0. Therefore the two solutions $\beta = \pm k_0$ to the sixth-degree equation (9.101*b*) are not solutions of (9.100) and do not correspond to physical waves.

The ac current and velocity are given by (9.14) and (9.13*a*) as

$$v = \frac{j\eta E_z}{v_0(\beta_0 - \beta)} \tag{9.107}$$

$$J = -j\frac{\omega_p^2}{\omega}\frac{\beta_0^2}{(\beta_0 - \beta)^2}\epsilon_0 E_z \tag{9.108}$$

These equations show that v and J are negligible for the three waves for which β is significantly different from β_0. Thus v and J arise from the first three slow waves discussed. The fourth wave can be excited by reflection at the output end of the tube. If it is reflected at the input end also, it will be amplified and, with continued reflection and amplification, will result in oscillations. To avoid this undesirable feature, an attenuating resistive vane or an integral ferrite isolator is built into the traveling-wave tube.

At the input end we must have the total ac current and velocity associated with the three forward slow waves vanish. Thus the initial conditions at the input $z = 0$ are

$$J_1 + J_2 + J_3 = 0 \qquad v_1 + v_2 + v_3 = 0$$

When we assume that

$$E_z = I_0(\beta_0 r)\left(C_1 e^{-j\beta_1 z} + C_2 e^{-j\beta_2 z} + C_3 e^{-j\beta_3 z}\right)$$

and make use of (9.108) and the initial conditions, we find that

$$\frac{J_2}{J_1} = e^{j2\pi/3} \qquad \frac{J_3}{J_1} = e^{-j2\pi/3}$$

Consequently, all three waves at the input have equal *magnitudes*; i.e., we find that $C_1 = C_2 = C_3$. The growing wave will have an amplitude equal to one-third that of the input signal. Therefore the amplitude gain of a traveling-wave tube is

$$\frac{E_0}{E_i} = \frac{1}{3}e^{\alpha_g l}$$

where α_g is given by (9.105) and l is the tube length. The power gain in decibels is

$$\begin{aligned} A &= 20 \log 0.333 + 20\alpha_g l \log e \\ &= -9.54 + 3.75\beta_0 l\left(\frac{\omega_p}{\omega}\right)^{2/3} \end{aligned} \tag{9.109}$$

With the aid of the preceding results we can now describe the physical mechanism of the gain. We note that the growing wave has a phase velocity slightly less than the beam velocity. This growing wave is the perturbed slow space-charge wave. The ac kinetic-power density of the fast and slow space-charge waves are [see (9.49), (9.13a), and (9.14)]

$$\text{Re}\,\frac{-v_0 v_f J_f^*}{2\eta} = -\frac{\omega\omega_p^2\epsilon_0}{2v_0^2(\beta_f-\beta_0)^3}|E_{zf}|^2$$

$$\text{Re}\,\frac{-v_0 v_s J_s^*}{2\eta} = -\frac{\omega\omega_p^2\epsilon_0}{2v_0^2(\beta_s-\beta_0)^3}|E_{zs}|^2$$

The slow space-charge wave has $\beta_s > \beta_0$ and hence has a negative ac kinetic-power density, whereas the fast space-charge wave has a positive ac kinetic-power density. Since the slow wave grows, it therefore loses energy, and the conservation theorem (9.49) then requires that the electromagnetic power increase. The ac current of the slow wave will have a phase angle relative to E_z such that $\text{Re}(E_{zs}J_s^*)$ is negative and the current continually gives up energy to the field. This may be verified by substituting β_3 for β in (9.14) to obtain

$$\text{Re}(E_{zs}J_s^*) = -2\sqrt{3}\left(\frac{\omega_p}{\omega}\right)^{2/3}\beta_0^2\epsilon_0|E_{zs}|^2$$

As a further aid to the understanding of the traveling-wave tube, it may be noted that it can be viewed as a large number of closely spaced cavity gaps operating as a multicavity klystron. The adjacent turns of the helix are then considered as constituting a gap.

The main advantage of the traveling-wave tube over the klystron is its relatively broad frequency band of operation. Typical units provide gains of 30 to 50 dB over an octave or more in frequency. Power-handling capability ranges from milliwatts to megawatts.

9.11 *M*-TYPE TRAVELING-WAVE TUBE

The magnetron-type (*M*-type) traveling-wave tube is a linear version of the cylindrical magnetron. Figure 9.21 is a schematic illustration of an *M*-type tube using a corrugated, or comblike, slow-wave circuit. The electron beam is much wider than it is thick and approximates a sheet beam. A potential V_a is applied between the sole and the anode block. A large static magnetic field is applied in a direction perpendicular to the beam velocity $v_0\mathbf{a}_z$, and the static electric field $-E_a\mathbf{a}_y$ arises from the anode to sole potential V_a. The electrons moving upward from the cathode at potential V_c are deflected by the magnetic field into a beam moving in the positive z direction. The desired type of flow is the one where there is only a z-directed velocity

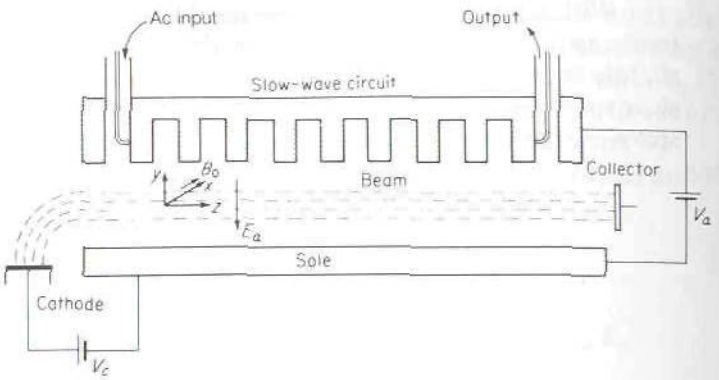


FIGURE 9.21
M-type traveling-wave tube.

$v_0(y)$, which in general is a function of y. Electron flow takes place in a crossed **E** and **B** field, which is typical of magnetron-type tubes.

For stable flow, $v_0(y)\mathbf{a}_z$ does not vary with z. If we denote by $V(y)$ the potential at an arbitrary value of y between the sole and anode block, we must have a balance between the magnetic force

$$-ev_0(y)\mathbf{a}_z \times \mathbf{a}_x B_0 = -eB_0 v_0(y)\mathbf{a}_y$$

and the electric field force $\mathbf{a}_y e\,\partial V/\partial y$. Thus

$$\frac{\partial V}{\partial y} = v_0(y)\,B_0 \tag{9.110}$$

The velocity $v_0(y)$ may be found from the energy equation

$$\tfrac{1}{2}mv_0^2(y) = e(V - V_c)$$

The derivative with respect to y gives

$$v_0(y)\frac{\partial v_0}{\partial y} = \eta\frac{\partial V}{\partial y} \tag{9.111}$$

The potential $V(y)$ arises from the applied potential V_a and from the dc space charge within the beam. Under equilibrium conditions the force $-e(\mathbf{E} + \mathbf{v} \times \mathbf{B}_0)$ acting on an electron is zero. The divergence of this equation thus gives

$$\nabla \cdot \mathbf{E} + \nabla \cdot \mathbf{v} \times \mathbf{B}_0 = 0 = -\frac{\rho_0 s}{\epsilon_0} + B_0\frac{\partial v_0}{\partial y} \tag{9.112}$$

since $\mathbf{v} = v_0\mathbf{a}_z$ and $\mathbf{B}_0 = B_0\mathbf{a}_x$. In this equation $-\rho_0$ is the dc negative space charge density and s is a factor giving the fraction of the negative space charge which is not neutralized by positive ions. For no positive ions

present, $s = 1$. If we assume that $s = 1$, the set of relations (9.110) to (9.112) can hold only if

$$\omega_c^2 = \eta^2 B_0^2 = \frac{\rho_0 \eta}{\epsilon_0} = \omega_p^2$$

or

$$\omega_c = \omega_p \tag{9.113}$$

as can be determined by eliminating $\partial V/\partial y$ and $\partial v_0/\partial y$. When this condition holds, the flow is referred to as planar Brillouin flow.

With the above model for the beam, it is possible to solve for space-charge waves that can propagate on the beam. In the presence of a slow-wave structure, the propagation constants become perturbed and a growing wave is produced similar to that in the *O*-type tube. For a detailed analysis the reader is referred to the citations given at the end of this chapter. The principles involved are not sufficiently different from those already discussed to warrant inclusion in this text.

9.12 GYROTRONS

Magnetrons and klystrons require resonant cavities to support the electromagnetic field that interacts with the electron beam. The traveling-wave tube requires a slow-wave structure. These structures have dimensions linearly proportional to the operating wavelength and become very small at millimeter wavelengths. The consequence of having to reduce the dimensions as the frequency increases is that the available area for the electron beam decreases and the power output that can be achieved decreases rapidly, approximately proportional to $1/f^2$. Thus, at frequencies of 100 GHz and above, conventional microwave tubes are not capable of producing power outputs in the kilowatt range. A relatively new tube, the gyrotron, has been developed in more recent years and does not rely on the use of resonant cavities or slow-wave structures. In the gyrotron the electromagnetic field interacts with the cyclotron motion of the electrons in a strong static magnetic field. When an electron is acted upon by the force of a steady magnetic field, its motion in the plane perpendicular to the magnetic field is a circular one. By using a sufficiently strong magnetic field, the frequency of rotation, called the cyclotron frequency, can be in the frequency range corresponding to millimeter waves. The interaction of a microwave field having the same frequency as the cyclotron frequency of the electrons results in growing waves. Thus the waveguide through which the electron beam passes and which supports the electromagnetic field is not restricted in diameter by the need to provide either a resonant structure or a slow-wave structure. As a result the fundamental size restrictions of conventional microwave tubes are not present in the gyrotron.

There are three common forms of gyrotron tubes. These are the gyromonotron oscillator; the gyro-TWT, a traveling-wave amplifier tube;

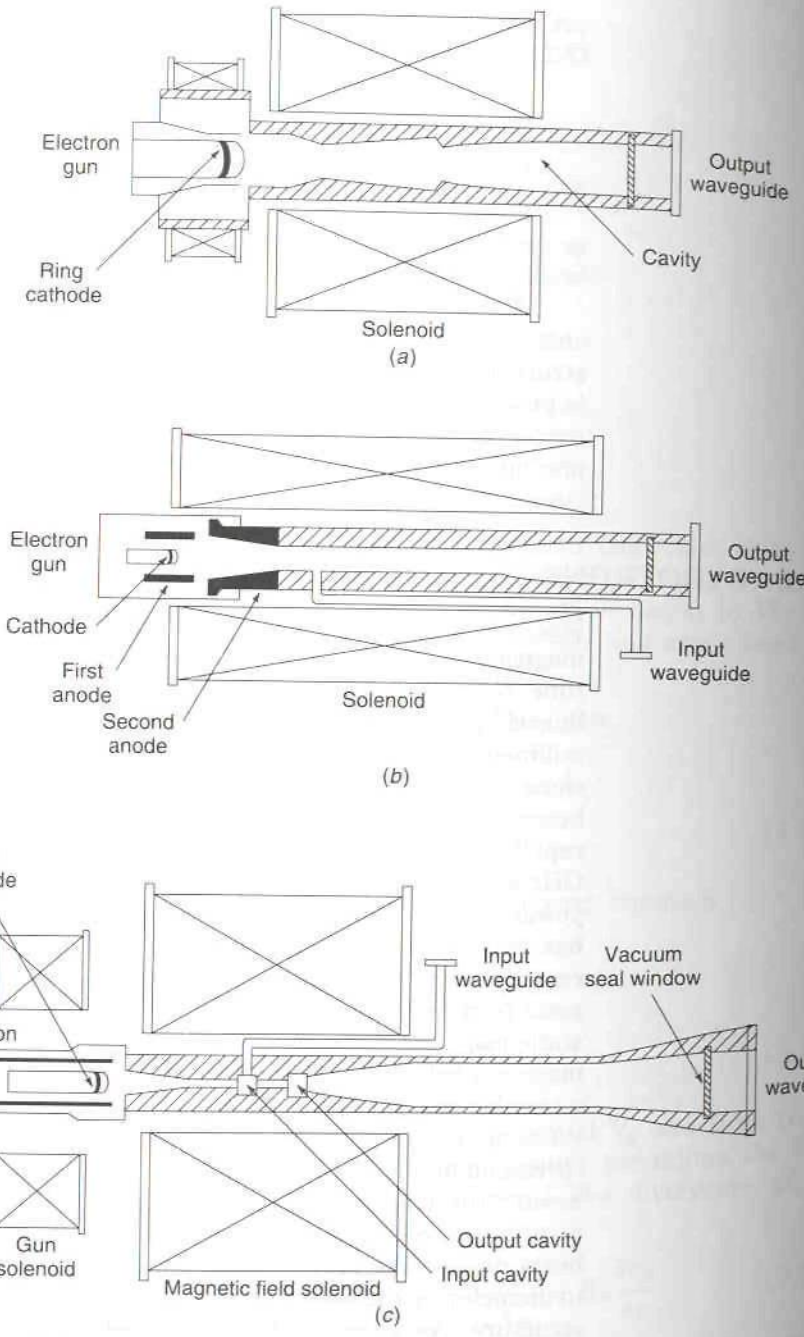


FIGURE 9.22
(*a*) The gyromonotron oscillator; (*b*) the gyro-TWT amplifier; (*c*) the gyroklystron amplifier.

and the gyroklystron, another amplifier tube. Simplified drawings of the three gyrotron tubes are shown in Fig. 9.22. Each tube has a magnetron-type electron gun which imparts a high radial velocity to the electrons before they enter the high magnetic field region. The large static magnetic field is provided by either a liquid cooled solenoid or a superconducting solenoid. In the gyromonotron shown in Fig. 9.22a, the interaction region is an enlarged circular waveguide that can support many different propagating modes. The output is taken from an output waveguide through a transparent window which is also used as a vacuum seal for the tube. In the gyro-TWT amplifier the input signal is coupled into the input of the interaction region through a waveguide as shown in Fig. 9.22*b*. The input microwave signal provides the initial bunching of the electrons in the beam. The input signal is amplified in the circular waveguide whose dimensions are large enough to support many possible propagating modes. The electron beam is in the form of a hollow beam with a radius such that it interacts strongly with only one, or at most only a few, of the circular waveguide modes. The gyroklystron uses an input and output cavity as shown in Fig. 9.22*c*. The signal to be amplified is coupled into the input cavity. The output signal is taken from the output cavity.

Gyrotron amplifiers that provide power gains of as much as 24 dB and output powers as high as 50 kW at 5 GHz have been built. Ferguson, Valier, and Symons describe a 5.2-GHz gyrotron tube producing 128 kW of output power.† This tube uses an 8-A–65 kV electron beam. Pulsed power outputs from gyrotron oscillators have been produced at levels of several hundred megawatts. It has been reported that 28-GHz gyrotrons with 200 kW of continuous-wave output power are in operation at Oak Ridge National Laboratory.‡ The high power capability of gyrotrons has been amply demonstrated; so these tubes will become more important for millimeter-wave systems in the future.

Field-Particle Interaction in a Gyrotron

In a gyrotron electron beams having a very large azimuthal velocity are used, so that the relativistic increase in the mass of the electron must be taken into account. The electron gun injects the beam into the high magnetic field region with the electrons initially having a large radial velocity component. The $\mathbf{v}_0 \times \mathbf{B}_0$ force then causes the electrons to follow a helical path with a velocity $\mathbf{v}_0 = v_{0\phi}\mathbf{a}_\phi + v_{0z}\mathbf{a}_z$ with $v_{0\phi}$ being several times larger than v_{0z}. If we treat the electrons as a charged fluid, then under the action

†P. E. Ferguson, G. Valier, and R. S. Symons, Gyrotron-TWT Operating Characteristics, *IEEE Trans.*, vol. MTT-29, pp. 794–799, 1981.

‡J. T. Coleman, "Microwave Devices," Reston Publishing Company, Inc., Reston, Va, 1982.

of a microwave field the velocity of a differential volume element of this charged fluid will have an ac component $\mathbf{v}$ in addition to the dc component $\mathbf{v}_0$; so the total velocity field will be $\mathbf{v}_t = \mathbf{v}_0 + \mathbf{v}$. The effective mass of the electron will be $m\gamma = m(1 - v_t^2/c^2)^{-1/2}$, where γ is the relativistic correction. The momentum is $m\gamma v_t$. The particle density is $N + n$, where n is the ac variation in the number density from the average value of N particles per unit volume. The charge density is given by $-e(N + n) = -\rho_0 + \rho$. If $\mathbf{E}$ and $\mathbf{B}$ represent the microwave field and $\mathbf{B}_0$ is the static magnetic field, then the equation of motion for a volume element of the electron fluid is [see (9.5*f*)]

$$(N + n)m\frac{d(\gamma \mathbf{v}_t)}{dt} = (\rho - \rho_0)(\mathbf{E} + \mathbf{v}_t \times \mathbf{B} + \mathbf{v}_t \times \mathbf{B}_0) \quad (9.114)$$

When there is no ac field present, the steady-state motion is a drift with a constant velocity v_{0z} along the direction of $\mathbf{B}_0$, which we take as the z axis, and rotation about $\mathbf{B}_0$ at the cyclotron frequency Ω given by

$$\Omega = \frac{\rho_0 B_0}{Nm\gamma_0} = \frac{eB_0}{m\gamma_0} \quad (9.115)$$

where $\gamma_0 = (1 - v_0^2/c^2)^{-1/2}$. By using a very large magnetic field, Ω can be in the microwave or millimeter-wave range of frequencies. For example, if $v_0/c = 0.8$ and $B_0 = 3$ W/m^2 (30 kG), we get $\Omega = 8.8 \times 10^{11}$, which corresponds to a frequency of 140 GHz. A magnetic field as large as 3 W/m^2 requires the use of superconducting solenoids. A gyrotron can operate at harmonics of the cyclotron frequency. The advantage of operating at a harmonic of the cyclotron frequency is that a smaller static magnetic field is required, but this is accompanied by a lower efficiency for power generation.

The possibility of field interaction with the beam at a harmonic of the cyclotron frequency is readily demonstrated by considering the current associated with a circulating electron. Consider an electron with azimuthal velocity $v_{0\phi}$ and located at $\phi = \phi_1$, $r = r_0$, at time $t = 0$. The current is in the form of an impulse $J_\phi = -ev_{0\phi}\delta(r - r_0)\delta(z)\delta(\phi - \phi_1)$, where the delta functions localize the current element at the position r_0, $z = 0$, $\phi = \phi_1$. We can make a Fourier series expansion of J_ϕ in terms of the angle ϕ; thus let

$$J_\phi = \delta(r - r_0)\delta(z) \sum_{n=-\infty}^{\infty} I_n e^{jn\phi} \quad (9.116)$$

where the I_n are given by

$$\begin{aligned} I_n &= \frac{1}{2\pi}\int_0^{2\pi} e^{-jn\phi}(-ev_{0\phi})\delta(\phi - \phi_1)\, d\phi \\ &= -\frac{ev_{0\phi}}{2\pi}e^{-jn\phi_1} \end{aligned} \quad (9.117)$$

Hence we have, at $t = 0$,

$$J_\phi = -\frac{ev_{0\phi}}{2\pi}\delta(r - r_0)\delta(z) \sum_{n=-\infty}^{\infty} e^{jn(\phi-\phi_1)} \tag{9.118}$$

At a later time the position of the electron will be at $\phi = \phi_1 + \Omega t$. A Fourier series expansion of the current at time t may also be carried out and gives

$$J_\phi(t) = -\frac{ev_{0\phi}}{2\pi}\delta(r - r_0)\delta(z) \sum_{n=-\infty}^{\infty} e^{jn(\phi-\phi_1)-jn\Omega t} \tag{9.119}$$

upon replacing ϕ_1 by $\phi_1 + \Omega t$. This expansion shows that the current associated with a single rotating electron is composed of an infinite number of equal-amplitude harmonics of the cyclotron frequency, a result due to the impulsive nature of the current.

Consider now a very large number N of electrons spaced at random around the orbit. Each electron contributes a current given by (9.119) but with ϕ_1 replaced by ϕ_i for the ith electron. The total current is obtained by averaging over all electrons and will involve the average of the following quantity over all phase angles ϕ_i:

$$J_{\phi N} = -\frac{ev_{0\phi}}{2\pi}\delta(r - r_0)\delta(z) \sum_{n=-\infty}^{\infty} e^{jn(\phi-\Omega t)} \sum_{i=1}^{N} e^{-jn\phi_i}$$

For large N the average will be zero except for the $n = 0$ term which gives a factor N. Thus the current becomes

$$J_{\phi N} = -\frac{Nev_{0\phi}}{2\pi}\delta(r - r_0)\delta(z) \tag{9.120}$$

which is a dc current. These results show that, in order to obtain an ac current at the cyclotron frequency or its harmonics, the electron distribution around the orbit must be nonuniform. We require bunching of the electrons around the orbit. If electron bunching occurs, then we will obtain an ac current with which the microwave field can interact. We will show next that the dependence of the electron mass on the velocity provides a mechanism that will cause electron bunching to occur.

If we take a scalar product of the equation of motion with $\mathbf{v}_t$, we obtain

$$(N + n)m\mathbf{v}_t \cdot \left(\frac{d}{dt}\gamma\mathbf{v}_t\right) = (\rho - \rho_0)\mathbf{v}_t \cdot \mathbf{E} \tag{9.121}$$

We now note that

$$\mathbf{v}_t \cdot \frac{d}{dt}\gamma\mathbf{v}_t = \mathbf{v}_t \cdot \mathbf{v}_t \frac{d\gamma}{dt} + \gamma\mathbf{v}_t \cdot \frac{d\mathbf{v}_t}{dt}$$

$$= v_t^2 \frac{d\gamma}{dt} + \frac{1}{2}\gamma\frac{dv_t^2}{dt}$$

since $d(\mathbf{v}_t \cdot \mathbf{v}_t)/dt = 2\mathbf{v}_t \cdot d\mathbf{v}_t/dt$. We also have

$$\frac{d\gamma}{dt} = \frac{d}{dt}\left(1 - \frac{v_t^2}{c^2}\right)^{-1/2} = \frac{1}{2c^2}\gamma^3 \frac{dv_t^2}{dt}$$

By using these results we find that

$$v_t^2 \frac{d\gamma}{dt} + \frac{\gamma}{2}\frac{dv_t^2}{dt} = \left(v_t^2 + \frac{c^2}{\gamma^2}\right)\frac{d\gamma}{dt} = c^2 \frac{d\gamma}{dt}$$

so

$$\frac{d\gamma}{dt} = \frac{\rho - \rho_0}{m(N+n)c^2}\mathbf{v}_t \cdot \mathbf{E} = -\frac{e}{mc^2}\mathbf{v}_t \cdot \mathbf{E} \tag{9.122}$$

This equation tells us that when $\mathbf{v}_t \cdot \mathbf{E}$ is positive γ will decrease and when $\mathbf{v}_t \cdot \mathbf{E}$ is negative γ will increase. From (9.115) we see that when γ decreases the cyclotron frequency will increase and when γ increases the cyclotron frequency will decrease. Consequently, those electrons that have a phase angle greater than $\pm 90°$ relative to the electric field will have their cyclotron frequency reduced, whereas those with phase angles less than $\pm 90°$ relative to that of the electric field will have an increased cyclotron frequency. This process results in bunching of the electrons in the azimuthal direction in a manner similar to the longitudinal bunching that occurs in a klystron. When the electric field adds energy to the electron, its azimuthal velocity increases. Paradoxically, this increases the value of γ but reduces the cyclotron frequency. What happens is that the radius of the orbit increases so that even though the azimuthal velocity has increased it takes longer for an electron to execute one circuit around the orbit so the cyclotron frequency is lower.

Gyrotrons generally use hollow cylindrical beams as shown in Fig. 9.23. In Fig. 9.23*a* we show a conventional beam in which the electrons revolve around individual magnetic field lines and do not have a common center of rotation (guiding center). This type of beam is best suited for field interaction at the cyclotron frequency. The beam shown in Fig. 9.23*b* is used in large-orbit gyrotrons. All of the electrons in this beam have a common center of rotation. As a typical example a beam with a cyclotron frequency of 10 GHz and an azimuthal velocity of $0.8c$ will have an orbital radius equal to $v_{0\phi}/\Omega = 2.4 \times 10^{11}/(2\pi \times 10^{10}) = 3.82$ mm. This beam is large enough to provide good field interaction with microwave fields at various harmonics of the cyclotron frequency.

The small-signal analysis of a gyrotron can be based on the linearized equation of motion. If only terms linear in the ac quantities are retained in (9.114), the equation of motion, we obtain

$$\gamma_0 Nm\left(\frac{\partial}{\partial t} + \mathbf{v}_0 \cdot \nabla\right)\left(\mathbf{v} + \frac{\gamma_0^2}{c^2}\mathbf{v}_0\mathbf{v}_0 \cdot \mathbf{v}\right)$$
$$\approx -\rho_0\mathbf{E} - \rho_0(\mathbf{v}_0 \times \mathbf{B} + \mathbf{v} \times \mathbf{B}_0) + \rho\mathbf{v}_0 \times \mathbf{B}_0 \tag{9.123}$$

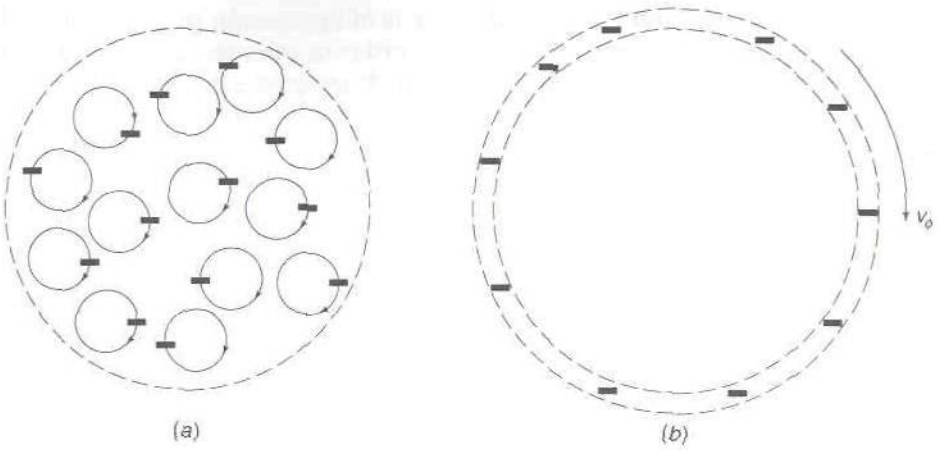


FIGURE 9.23
(*a*) Cross section of a solid electron beam with each electron having its own guiding center. (*b*) Cross section of a cylindrical sheath beam in which all electrons have a common guiding center.

In addition, the current $\mathbf{J}$ is given by

$$\mathbf{J} = -\rho_0 \mathbf{v} + \rho \mathbf{v}_0 \tag{9.124}$$

and the continuity equation

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \tag{9.125}$$

must hold. The above equations can be solved for $\mathbf{J}$ and ρ in terms of the ac fields $\mathbf{E}$ and $\mathbf{B}$. Maxwell's equations must then be solved in the circular (or other) waveguide, both within the beam region and outside, including the source terms. The results will show that growing waves are produced. All ac quantities can be assumed to have the form

$$\sum_{n=-\infty}^{\infty} C_n e^{jn\phi - j\beta z}$$

where C_n is an expansion coefficient for the quantity of interest. Since the equations are linear, the solution can be carried out for the nth term by itself. The electromagnetic field in the circular waveguide can be described in terms of left and right circularly polarized waves. Only those fields that rotate in synchronism with the electrons will interact with the beam ac currents.

For large power applications nonlinear effects must be included. The commonly used approach is to solve for the perturbed orbits of the electrons

and then evaluate the field interaction that take place. It is necessary to use numerical methods in order to solve the nonlinear equations. Typical results obtained by this method are given in the papers cited at the end of this chapter.

9.13 OTHER TYPES OF MICROWAVE TUBES

In addition to the main types of microwave tubes already discussed, there are a variety of others as well. In one form of traveling-wave tube, the resistance-wall amplifier, the helix is replaced by a circular guide lined with a resistive material. The resistive lining enables a slow wave to propagate in the guide, a wave that is highly attenuated in the absence of a beam. If an electron beam is present, amplification takes place with a growth constant α_g large enough to offset the attenuation due to the resistive lining. Thus a net overall amplification is obtained.

In another form of traveling-wave tube, the double-stream amplifier, two parallel electron beams are used. In this tube one of the beams provides the slow-wave structure, or circuit, for the other beam.

It is also possible to amplify the space-charge waves directly by passing the beam through a succession of accelerating and decelerating regions. This type of tube is called a velocity-jump amplifier because the beam velocity v_0 is periodically changed, or jumped, to new values.

For both the *O*-type and *M*-type traveling-wave tubes, it is possible to adjust the beam velocity so that it is equal to the phase velocity of any one of the spatial harmonics making up the Bloch wave that can propagate along the periodic structure used for the slow-wave circuit. In particular, interaction between the beam and one of the backward-propagating spatial harmonics is possible. Consider a Bloch wave propagating in the $-z$ direction. For this wave, E_z has the expansion

$$E_z(r, z) = \sum_{n=-\infty}^{\infty} E_n(r) e^{j\beta z + j2n\pi z/p}$$

where p is the period of the periodic structure in the z direction. If we want interaction between the beam and the $n = -1$ spatial harmonic, it is only necessary to choose

$$v_0 = v_p = \frac{\omega}{-(\beta - 2\pi/p)} = \frac{\omega}{2\pi/p - \beta}$$

If the period p is small enough, the $n = -1$ spatial harmonic has a phase velocity directed in the $+z$ direction and its group velocity is in the $-z$ direction. If the $n = -1$ spatial harmonic is amplified, all the other spatial

harmonics are also amplified, since they must all be present with very definite amplitudes in order that the boundary conditions may be satisfied. The amplification of the noninteracting spatial harmonics comes about because of the increasing surface current and charge induced on the metallic boundaries by the amplified spatial harmonic that interacts with the beam. Tubes employing interaction with a backward spatial harmonic are usually used as oscillators and are called backward-wave oscillators, or carcinotrons. They have their output coupling at the cathode end.

There are still other forms of microwave tubes, and no doubt more will be developed. For more extensive discussion the cited references at the end of this chapter should be consulted.

PROBLEMS

9.1. Consider an electron beam of radius a, velocity v_0, and space charge density ρ_0. The dc current density is then $J_0 = -\rho_0 v_0$. Show that a magnetic field

$$H_\phi = \begin{cases} -\dfrac{r\rho_0 v_0}{2} & 0 \le r \le a \\ -\dfrac{a^2 \rho_0 v_0}{2r} & r \ge a \end{cases}$$

is produced. Verify that the compression force $-e\mathbf{v}_0 \times \mathbf{B}_\phi$ is much smaller than the radial outward force due to the space-charge electric field and may therefore be neglected.

9.2. Show that an electron with velocity $\mathbf{v}$ perpendicular to $\mathbf{B}_0$ executes circular motion at the cyclotron frequency $\omega_c = e\mathbf{B}_0/m = \eta B_0$ by equating the centrifugal force to the $-e\mathbf{v} \times \mathbf{B}_0$ force.

9.3. An electron beam has a radius of 0.2 cm. The accelerating voltage is 1,000 V. The total beam current is 0.03 A. Calculate the beam perveance, the space charge density ρ_0 and velocity v_0, the number of electrons per cubic meter, and the radial electric field due to space charge. Estimate the radial displacement of an electron located at the beam boundary during the time it takes the beam to move a distance $d = 5$ cm. Use the equation $m\,d^2r/dt^2 = -eE_r$, and assume E_r to be constant and equal to its value at the beam boundary. Is the beam dispersion significant in this case if d is kept less than 5 cm?

9.4. Consider an electron beam with dc parameters ρ_0, $\mathbf{v}_0 = v_0\mathbf{a}_z$ immersed in a field $\mathbf{B}_0 = \mathbf{a}_z B_0$. Assume a time dependence $e^{j\omega t}$ and a z dependence $e^{-j\beta z}$ and solve the linearized equation of motion (9.8) for $\mathbf{v}_1 = v_x\mathbf{a}_x + v_y\mathbf{a}_y + v_z\mathbf{a}_z$ to obtain

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = -\eta \begin{bmatrix} j(\omega - \beta v_0)/\Delta & -\omega_c/\Delta & 0 \\ \omega_c/\Delta & j(\omega - \beta v_0)/\Delta & 0 \\ 0 & 0 & 1/j(\omega - \beta v_0) \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}$$

where $\Delta = \omega_c^2 - (\omega - \beta v_0)^2$.

From the continuity equation (9.5*e*) and (9.11*b*), show that

$$\mathbf{J} = \frac{\mathbf{v}_0 \nabla \cdot \mathbf{J}}{-j\omega} - \rho_0 \mathbf{v}$$

$$J_z = \frac{v_0 \nabla \cdot \mathbf{J}}{-j\omega} - \rho_0 v_z$$

$$\mathbf{J}_t = -\rho_0 \mathbf{v}_t$$

$$\nabla \cdot \mathbf{J} = -j\beta J_z - \rho_0 \nabla_t \cdot \mathbf{v}_t$$

$$J_z = \frac{jJ_0 \nabla_t \cdot \mathbf{v}_t - \omega \rho_0 v_z}{\omega - \beta v_0}$$

9.5. Using the results of Prob. 9.4, obtain solutions for β for waves in an infinite electron beam when all ac quantities are independent of x and y. Note that, for space-charge waves, $E_x = E_y = 0$ but E_z is finite. For the field waves, $E_z = 0$.

Hint: Note that $\nabla \times \mathbf{H} = -j\beta \mathbf{a}_z \times \mathbf{H}$, $\nabla \times \mathbf{E} = -j\beta \mathbf{a}_z \times \mathbf{E}$, which leads to the equation $j(\beta^2 - k_0^2)\mathbf{E}_t = \omega \mu_0 \mathbf{J}_t = -\omega \mu_0 \rho_0 \mathbf{v}_t$.

Answer: For field waves, β is a solution of

$$(\omega - \beta v_0)\frac{\omega \omega_p^2}{c^2} - (\beta^2 - k_0^2)\Delta = \pm \frac{\omega_c \omega \omega_p^2}{c^2}$$

where Δ is given in Prob. 9.4. Note that two solutions are given by $\omega - \beta v_0 = \pm \omega_c$. These are the cyclotron waves. For $\beta \approx k_0$, so that $\omega \gg \beta v_0$, four other approximate solutions are

$$\beta = \pm k_0 \left[1 - \frac{\omega_p^2}{\omega(\omega \pm \omega_c)}\right]$$

9.6. Compute the gain of a klystron amplifier of the type considered in the text where the following data apply: Beam radius = 0.3 cm, beam current density = 100 mA/cm^2. Accelerating voltage = 1,000 V, frequency = 3,000 MHz. $G_L = G_0$, cavity width $d = 0.2$ cm, cavity conductivity = 5.8×10^7 S/m. Compute the gain for $d = 0.3$ cm also, and compare with the earlier calculation.

9.7. Consider a reflex klystron employing a cylindrical cavity of the form shown in Fig. 9.12. The data of Prob. 9.6 apply, with $d = 0.2$ cm. The external loading $G_L = G_0$. The cavity grid-reflector spacing s is equal to 1 cm. Calculate and plot the electronic admittance spiral as a function of reflector voltage V_r at a frequency of 3,000 MHz. Plot also the negative cavity admittance $-Y$ on the same susceptance plane. Determine the reflector-voltage variation to tune across the $n = 2$ and $n = 3$ modes. Evaluate the change in oscillation frequency as the modes are tuned across.

9.8. Consider a cylindrical waveguide of radius a lined by a resistance sheet so that the boundary conditions at $r = a$ are $E_z = -Z_m H_\phi$, where $Z_m = (1 + j)/\sigma\delta_s$ is the surface impedance of the wall. Analyze this structure as a traveling-wave

tube when an electron beam (axially confined flow) with velocity $v_0\mathbf{a}_z$ completely fills the guide. Determine an appropriate value of Z_m in order to obtain amplification. Find the optimum value of Z_m to give a maximum gain.

Answer: β is a solution of

$$\frac{1}{g}\frac{I_0(ga)}{I_1(ga)} = -\frac{j\omega\epsilon_0 Z_m}{\beta^2 - k_0^2}$$

For ga large, so that $I_0/I_1 \approx 1$, $\beta = (1 + \delta)\beta_0$ with δ small, δ^2 is given by

$$\delta^2 = j\frac{2k_0^2(R_m/Z_0)^2(\omega_p/\omega)^2}{\beta_0^2 + j2k_0^2(R_m/Z_0)^2}$$

where $Z_m = (1 + j)R_m$.

9.9. Consider a cylindrical waveguide of radius a uniformly filled with a stationary plasma (an ionized gas with an equal number of electrons and ions per unit volume). At high frequencies the motion of the ions may be neglected because their mass is much greater than that of the electrons. Thus the guide will have the same properties as one filled with an electron beam with zero axial dc velocity. Use (9.33) to show that the guide may be considered as filled with a dielectric medium with permittivity $\epsilon = \epsilon_0(1 - \omega_0^2/\omega^2)$, where ω_0 is the plasma frequency for the plasma. Find a solution for the lowest-order circularly symmetric E mode and show that for $\omega < \omega_0$ such that ϵ is negative the wave impedance is inductive.

9.10. The results of Prob. 9.9 may be used to analyze the beam-plasma amplifier. Consider an electron beam passing through the plasma-filled guide. Use a confined-flow model to describe the beam and show that for a beam completely filling the guide the equations of Sec. 9.3 are valid provided ϵ_0 is replaced by $\epsilon = \epsilon_0(1 - \omega_0^2/\omega^2)$ throughout, where ω_0 is the plasma frequency for the plasma. In particular, (9.24) to (9.26) hold. Thus in (9.26), if $\omega_p = (\epsilon\rho_0/m\epsilon_0)^{1/2}$ is replaced by $(e\rho_0/m\epsilon)^{1/2} = (e\rho_0/m\epsilon_0)^{1/2}[\omega/(\omega^2 - \omega_0^2)^{1/2}]$, it is seen that β becomes complex for $\omega < \omega_0$ and a growing and decaying pair of waves are obtained. Show that the gain constant is given by

$$\text{Re}\,\beta_0\left[\frac{m\epsilon_0}{e\rho_0}\left(1 + \frac{p_{01}^2}{a^2\beta_0^2}\right)(\omega^2 - \omega_0^2)\right]^{1/2}$$

and is very large when ω is close to ω_0. Note also that for a finite-radius beam with unconfined flow passing through an unbounded plasma medium the equations of Sec. 9.4 apply with ϵ_0 again replaced by ϵ. For this model of the beam-plasma amplifier (9.45) may be used in place of (9.26) and will predict a gain constant of the same order of magnitude as does the confined-flow model.

9.11. Show that when both the plasma electrons and the beam electrons are subjected to the confined-flow condition the only change which occurs in the result given in Prob. 9.10 is the replacement of ω_0 by $F\omega_0$, where $F = (1 + p_{0m}^2/\beta_0^2 a^2)^{-1/2}$ is the plasma-frequency reduction factor.

REFERENCES

1. Slater, J. C.: "Microwave Electronics," D. Van Nostrand Company, Inc., Princeton, N.J., 1950.
2. Pierce, J. R.: "Traveling Wave Tubes," D. Van Nostrand Company, Inc., Princeton, N.J., 1950.
3. Kleen, W. J.: "Electronics of Microwave Tubes," Academic Press, Inc., New York, 1958.
4. Beck, A. H. W.: "Space Charge Waves," Pergamon Press, New York, 1958.
5. Hutter, R. G.: "Beam and Wave Electronics in Microwave Tubes," D. Van Nostrand Company, Inc., Princeton, N.J., 1960.
6. Spangenberg, K. R.: "Vacuum Tubes," McGraw-Hill Book Company, New York, 1948.
7. Hamilton, D. R., J. K. Knipp, and J. B. Horner Kuper: "Klystrons and Microwave Triodes," McGraw-Hill Book Company, New York, 1948.
8. Collins, G. B.: "Microwave Magnetrons," McGraw-Hill Book Company, New York, 1948.
9. Chodorow, M., and C. Susskind: "Fundamentals of Microwave Electronics," McGraw-Hill Book Company, New York, 1964.
10. Reich, H. J., P. F. Ordung, H. L. Krauss, and J. K. Skalnik: "Microwave Theory and Techniques," D. Van Nostrand Company, Inc., Princeton, N.J., 1953.
11. Reich, H. J., J. K. Skalnik, P. F. Ordung, and H. L. Krauss: "Microwave Principles," D. Van Nostrand Company, Inc., Princeton, N.J., 1957.

Space-charge wave theory

12. Ramo, S.: The Electronic Wave Theory of Velocity Modulated Tubes, *Proc. IRE*, vol. 27, p. 757, 1939.
13. Ramo, S.: Space-Charge and Field Waves in an Electron Beam, *Phys. Rev.*, vol. 56, p. 276, 1939.
14. Hahn, W. C.: Small Signal Theory of Velocity Modulated Electron Beams, *Gen. Elec. Rev.*, vol. 42, p. 258, 1939.
15. Chodorow, M., and L. Zitelli: The Radio Frequency Current Distribution in Brillouin Flow, *IRE Trans.*, vol. ED-6, p. 352, 1959.
16. Rigrod, W., and J. Lewis: Wave Propagation along a Magnetically Focused Cylindrical Electron Beam, *Bell System Tech. J.*, vol. 33, p. 399, 1954.
17. Brewer, G. R.: Some Effects of Magnetic Field Strength on Space-Charge Wave Propagation, *Proc. IRE*, vol. 44, p. 896, 1956.

Gyrotrons

18. Sprangle, P. and A. T. Drobot: The Linear and Self-Consistent Nonlinear Theory of the Electron Cyclotron Maser Instability, *IEEE Trans.*, vol. MTT-25, pp. 528–544, 1977.
19. Li, Q. F., S. Y. Park, and J. L. Hirshfield: Theory of Gyrotron Traveling-Wave Amplifiers, *IEEE Trans.*, vol. MTT-34, pp. 1044–1058, 1986.
20. Silverstein, J. D., M. E. Read, K. R. Chu, and A. T. Drobot: Practical Considerations in the Design of a High-Power 1 mm Gyromonotron, *IEEE Trans.*, vol. MTT-28, pp. 962–966, 1980.
21. Vitello, P., W. Miner, and A. Drobot: Theory and Numerical Modeling of a Compact Low-Field High-Frequency Gyrotron, *IEEE Trans.*, vol. MTT-32, pp. 373–386, 1984.

CHAPTER 10

SOLID-STATE AMPLIFIERS

The first solid-state amplifiers for microwave applications were negative-resistance diodes such as the tunnel diode. This was followed by the development of parametric amplifiers that used a variable-capacitance diode (varactor) and an oscillator (pump source) to vary the junction capacitance at the pump frequency. An outstanding feature of parametric amplifiers was the low noise that could be achieved by cooling the diode to liquid-nitrogen temperatures. The theory and design of parametric amplifiers is described in Chap. 11.

Parametric amplifiers became the prominent and most widely used solid-state amplifiers during the period 1958 to about 1970. By 1970, improvements in materials preparation and processing technology had resulted in development of *npn* silicon bipolar transistors with a maximum frequency of oscillation greater than 10 GHz. During the next two decades further progress in the design and manufacture of high-frequency microwave bipolar transistors and field-effect transistors was dramatic. The key to successful microwave transistor design is miniaturization which is a necessity in order to reduce device and package parasitic capacitances and lead inductances and to overcome the finite transit time of the charge carriers. An appreciation for the need to reduce parasitic capacitance and inductance can be obtained by referring to Table 10.1 where representative values of reactances are given at several frequencies. For example, an inductance of 0.1 nH at 10 GHz represents a reactance of 6.28 Ω which is not a negligible value in a 50-Ω system. A capacitance of 0.1 pF at 10 GHz has a reactance of 159 Ω and would be a significant shunt reactance across a 50-Ω transmission line.

Transit times are dependent on the electron mobility and saturation velocity in the semiconductor material. In this regard gallium arsenide

TABLE 10.1
Reactance as a function of frequency

Frequency (GHz)	1	10	100
Reactance			
$L = 0.1$ nH	0.628	6.28	62.8
$L = 1$ nH	6.28	62.8	628
$C = 0.1$ pF	1592	159	15.9
$C = 1$ pF	159	15.9	1.6

(GaAs) is significantly better than silicon for high-frequency devices. By 1980, the design and fabrication of metal-semiconductor field-effect transistors (MESFETs) were well established. In the frequency range above 5 GHz, MESFET devices are widely used.

In order to achieve the high-frequency performance in transistors, it was necessary to develop the technology that would enable key device dimensions to be less than 1 μm, e.g., gate lengths with submicron dimensions. By means of molecular beam epitaxy (MBE), it has been possible to grow high-quality epitaxial layers and controlled doping profiles in highly localized regions. MBE techniques also led to the development of heterostructures which, in turn, led to the development of the high-electron-mobility transistor (HEMT) which can operate at frequencies as high as 100 GHz.

Microwave amplifiers are usually constructed either as hybrid microwave integrated circuits (MIC) or as monolithic microwave integrated circuits (MMIC). In hybrid construction the transmission lines and matching networks are usually realized as microstrip circuit elements on a suitable substrate material and then the discrete components such as chip capacitors, resistors, and transistors are connected in place by soldering or using wire-bonding techniques. Discrete devices are available with beam leads for easy insertion into the hybrid circuit.

The word monolithic is derived from the two Greek words *monos* meaning single and *lithos* meaning stone. Thus a monolithic microwave integrated circuit is a circuit where all active devices, e.g., transistors, and passive circuit elements such as transmission lines, capacitors, resistors, and spiral inductors are fabricated on a single semiconductor crystal. The substrate material used has typically been gallium arsenide because of its high resistivity in the undoped state and because of its superiority for high-frequency field-effect device construction. A number of processes such as ion implantation for active devices, metal deposition and evaporation to form ohmic contacts, electrode pads, and transmission lines, via hole etching and plating, dielectric deposition, etc., is involved in monolithic circuit construction. The overall design and mask making is facilitated by the use

of computer-aided design (CAD) programs. Electron-beam lithography and plasma-enhanced etching and deposition techniques are used for fabrication of submicron device elements.

The cost of a monolithic microwave integrated circuit is related directly to how many circuits can be built on a single wafer since the processing of a single wafer is generally a fixed-cost operation. Consequently, in the frequency range below 10 GHz, where distributed circuit elements are relatively large, the hybrid form of construction is often less costly than monolithic construction. However, in the frequency range of 0.1 to 10 GHz, the ability to produce miniature inductors and capacitors has led to the development and production of many MMIC systems using lumped circuit elements instead of distributed circuit elements. In the millimeter-wavelength band monolithic microwave integrated circuit construction promises to be more cost effective and to yield circuits with greater reliability and uniformity.

In this chapter we are primarily concerned with the design of microwave amplifiers from the engineer's point of view. That is, starting from the measured or manufacturer's given two-port parameters of the device, we want to design an amplifier that meets a set of given system requirements such as gain, noise figure, bandwidth, and input and output VSWR. For this reason we only give a short discussion of the main characteristics of bipolar and field-effect transistors. Bias requirements and some typical bias circuits are also described.

The design methodology that we will develop is based on the use of the scattering-matrix parameters for the device. At high frequencies a transistor will have some intrinsic feedback from the output to the input, usually caused by a finite capacitance from collector to base and emitter lead inductance or in an FET the capacitance from drain to gate. Thus the device may be potentially unstable, and unless the amplifier is properly designed, it may oscillate, in which case it would not be useful as an amplifier. Thus, after the basic equations for amplifier gain have been derived, we examine the conditions under which the device is unconditionally stable or potentially unstable. If it is potentially unstable we will find that only for a certain range of values for the source and load impedances will the amplifier be stable. The available source and load impedances are easily displayed by constructing the source and load stability circles on a Smith chart. The equations for these stability circles will be derived and their interpretation and use will be examined.

The Smith chart is an indispensable aid in the visualization of the different constraints that the engineer must take into account in the design of a microwave amplifier. In addition to the input and output port stability circles already mentioned, there are a number of other useful circles that aid the design process and can be plotted on the Smith chart. The most important of these are circles of constant gain, circles of constant noise

figure, and circles of constant input and output mismatch. The equations for these other circles will be derived and their interpretation and use in amplifier design will be discussed.

The chapter will conclude with the description of a design strategy for single- and two-stage amplifiers that are subjected to various system requirements. A semiinteractive computer program that implements the design strategy will also be described. This computer program removes the drudgery of carrying out all the computations that are required and will give the user valuable experience in a design process where a number of constraints are imposed and tradeoffs among conflicting requirements must be made.

Most of the relationships involved in amplifier design are of the form of a bilinear transformation from one complex variable to another complex variable. If Z and W are two complex variables, then an equation of the form

$$W = \frac{AZ + B}{CZ + D}$$

where A, B, C, and D are complex constants, is a bilinear transformation of Z into W. This transformation has the property that circles in the Z plane will map into circles in the W plane, with straight lines as limiting forms of circles, with infinite radii, and some points as circles with zero radius. The relationship between impedance and reflection coefficient, namely,

$$\bar{Z} = \frac{1 + \Gamma}{1 - \Gamma}$$

is a bilinear transformation. For this transformation the straight lines $\bar{R}$ = constant and $\bar{X}$ = constant in the $\bar{Z}$ plane map into circles in the Γ plane. These mapped circles make up the Smith chart. Since the bilinear transformation occurs over and over again in amplifier design, we discuss its circle-mapping properties before we take up the design theory. This will provide results that enable us to readily identify circle mappings, in particular, the center and radii of mapped circles, from the particular bilinear transformation involved.

10.1 BIPOLAR TRANSISTORS

The basic principle of operation of a bipolar transistor designed for microwave applications is the same as that of low-frequency transistors. The device must be biased to set the operating point. In a common emitter amplifier circuit, an input network must be designed so that a signal voltage can be applied to the base. A suitable load impedance must be connected to the collector and the output signal is developed across this impedance. The main difference in the analysis of the operation of a transistor amplifier at

microwave frequencies relative to low-frequency operation is due to the fact that the intrinsic device, along with package parasitic capacitances, resistances, and lead inductances, requires a much more complex equivalent-circuit model. In addition, there is often sufficient capacitive feedback from the collector to the base so that the device is potentially unstable and will be prone to oscillate unless the input and output circuits are designed to prevent oscillations from occurring.

In the frequency range below 5 GHz, silicon bipolar transistors are generally preferred over GaAs FETs except for very low noise amplifier designs. Silicon bipolar technology is more mature and manufacturing costs are less. The gain obtained from a bipolar transistor is inherently greater than that from a field-effect transistor because of a much higher transconductance g_m. Bipolar transistors are suitable for oscillator and power amplifier applications in addition to small-signal amplifiers. Power gains of 15 to 20 dB can be obtained at 2 GHz with noise figures of around 2 dB. At 10 GHz the power gain for many presently available bipolar transistors is around 5 dB and the GaAs FET is then a better alternative. It is expected that, by perfecting the technology for making the critical dimensions of the emitter structure smaller, silicon bipolar transistors with a maximum frequency of oscillation approaching 100 GHz can be achieved. In recent years the heterojunction technology, originally applied to the construction of field-effect transistors, has also been used to improve the high-frequency performance of bipolar transistors. These transistors are called heterojunction bipolar transistors (HBTs) and exhibit very low base resistance, high current gain, and a speed increase by a factor of 2 to 3. An AlGaAs/GaAs HBT with a cutoff frequency of 105 GHz and maximum frequency of oscillation of 175 GHz has been reported.† Thus the future application of bipolar transistors can be expected to extend well into the millimeter-wavelength region.

The basic construction used in a microwave bipolar transistor involves a multifinger interdigitated emitter-base construction. A simplified drawing of the cross section and top view of a bipolar transistor is shown in Fig. 10.1. The use of this particular design is for the purpose of overcoming transit-time limitations and yet maintain a sufficient emitter area. An equivalent-circuit model of the intrinsic bipolar transistor and the additional parasitic elements added by the package is shown in Fig. 10.2. A drawing of a packaged microwave transistor with beam leads is shown in Fig. 10.3.

---

†N. H. Sheng, *et al.*, High Power GaAlAs/GaAs HBT's for Microwave Applications, *1987 IEEE Int. Electron Devices Meeting Digest*, pp. 619–622, 1987.

C. H. Liechti, High Speed Transitors: Directions for the 1990's *Microwave* pp. 165–177, September, 1989.

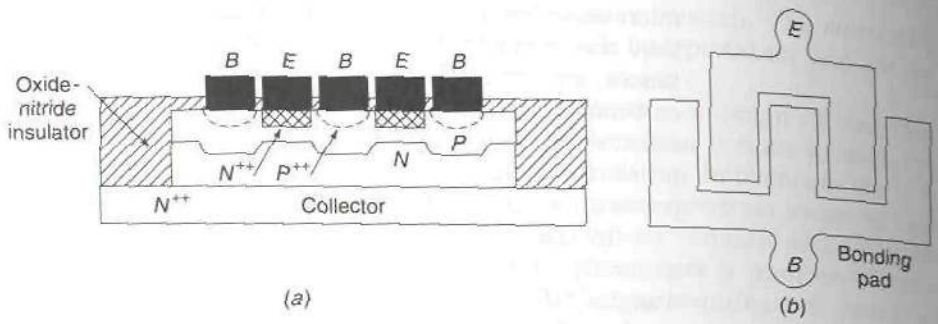


FIGURE 10.1
(a) Cross section of a microwave silicon bipolar transistor; (b) top view showing interdigitated emitter-base construction.

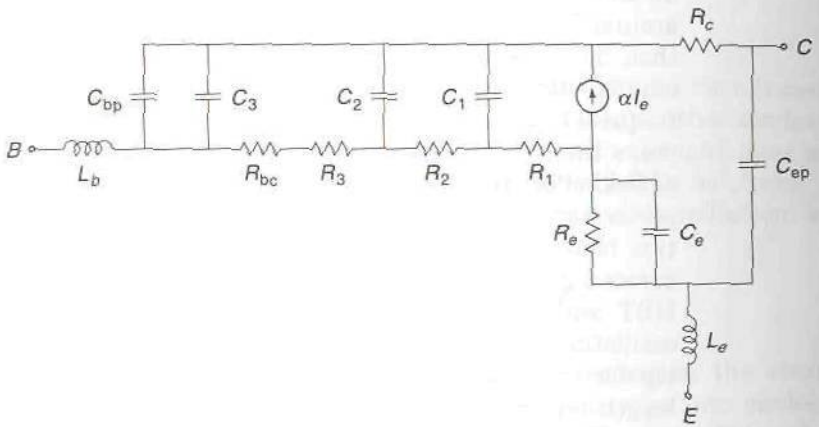


FIGURE 10.2
Equivalent-circuit model of a silicon bipolar transistor.

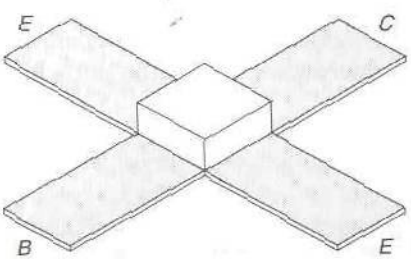


FIGURE 10.3
A packaged transistor with beam leads.

The equivalent-circuit model shown in Fig. 10.2 is based on that given by Vendelin, Pavio, and Rohde.† In this circuit model the various circuit elements are identified as follows:

C_{bp}—base bond pad capacitance
C_{ep}—emitter bond pad capacitance
R_{bc}—base contact resistance
R_{ec}—emitter contact resistance
R_1, R_2, R_3—base distributed resistance
C_1, C_2, C_3—collector-base distributed capacitance
R_e—dynamic emitter-base diode resistance
C_e—emitter-base diode junction capacitance
R_c—collector resistance
L_b, L_e—base and emitter bond wire inductances

For an Avantek AT-60500 silicon bipolar transistor operated with a collector current of 2 mA and a collector-emitter voltage of 8 V, typical values for these parameters are:

$$C_{bp} + C_3 = 0.055 \text{ pF} \qquad C_1 = 0.01 \text{ pF} \qquad C_2 = 0.039 \text{ pF}$$
$$C_{ep} = 0.026 \text{ pF} \qquad C_e = 0.75 \text{ pF}$$
$$R_{bc} + R_3 = 4.2\ \Omega, \qquad R_{ec} = 0.66\ \Omega, \qquad R_1 = 7.5\ \Omega$$
$$R_2 = 10.3\ \Omega \qquad R_e = 12.9\ \Omega \qquad R_c = 5\ \Omega$$
$$L_b = 0.5 \text{ nH} \qquad L_e = 0.2 \text{ nH}$$

This transistor has a base cutoff frequency of 22.7 GHz. The common base current gain is given by

$$\alpha = \frac{\alpha_0 e^{-j\omega\tau_d}}{1 + jf/f_b}$$

where f_b is the base cutoff frequency and τ_d is the collector depletion region delay time (6.9 ps).

The only device parameters that can be easily measured at microwave frequencies are the scattering-matrix parameters S_{11}, S_{12}, S_{21}, and S_{22}. These can be measured by embedding the transistor in a 50-Ω microstrip transmission line and using a network analyzer for the measurements. It is also possible to measure the S_{ij} parameters with on-wafer probes. By measuring the S_{ij} parameters over a range of frequencies, the equivalent-

†G. D. Vendelin, A. M. Pavio, and U. L. Rohde, "Microwave Circuit Design Using Linear and Nonlinear Techniques," chap. 3, John Wiley and Sons, Inc., New York, 1990.

circuit parameters can be adjusted, by using a computer program, so that they produce a circuit model with calculated scattering parameters that correlate with the measured ones. The equivalent-circuit model provides a better physical understanding of the various circuit parameters that will affect the operation of the transistor. However, for amplifier design it is easier to use the scattering-matrix parameters instead of the equivalent-circuit model.

Transistor Biasing

There are two main considerations involved in the design of a biasing circuit: (1) The biasing circuit must provide a stable operating point that is insensitive to variations in the device parameters and temperature changes and (2) the biasing circuit must be isolated from the high-frequency circuit so that high-frequency signal currents do not flow in the dc biasing circuit. The first requirement is met by incorporating dc feedback in the biasing circuit. The second requirement is met by inserting high-impedance high-frequency circuit elements in series with the dc components and by using low-impedance capacitive bypass circuits to shunt high-frequency currents around the dc circuit elements. The overall bias circuit and RF matching circuits must provide stable terminations for each active device outside the frequency band of interest in order to ensure that oscillations do not occur at any frequency.

The bias circuit shown in Fig. 10.4*a* provides a stable operating point. It is commonly used in low-frequency electronic circuits and can also be used in microwave amplifier circuits but with more difficulty because of parasitic inductance associated with the capacitor leads. The bias circuit is shown isolated from the transistor by incorporating series inductors (RF chokes) between the device terminals and the bias-circuit resistors. The

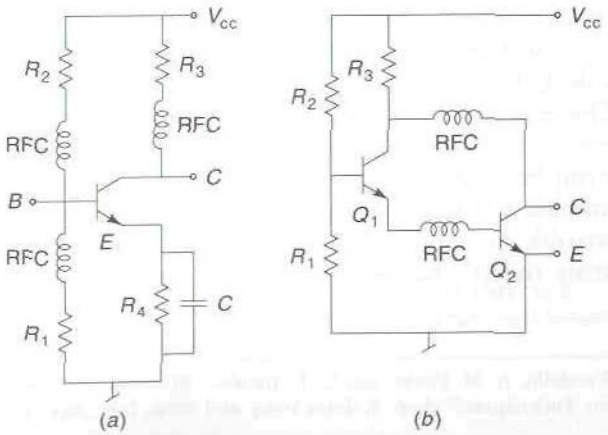


FIGURE 10.4
(*a*) A passive bias circuit; (*b*) an active bias circuit.

emitter terminal is maintained at RF ground by means of the bypass capacitor C.

An active bias circuit is shown in Fig. 10.4*b*. In this circuit the collector current in transistor Q_1 is established by means of the resistors R_1, R_2, and R_3. The base current in the microwave transistor Q_2 is the collector current of Q_1. Since the bias circuit for Q_1 is a stable one, the collector current of Q_1, and hence the base current of Q_2, is maintained at a value that is essentially independent of the transistor parameters. This circuit has the advantages that it consumes less dc power and requires only two RF chokes for isolation as compared with the circuit shown in Fig. 10.4*a* which is shown with three RF chokes and a bypass capacitor for isolation purposes.

In a microwave amplifier the RF chokes are often replaced by a quarter-wave high-impedance transmission line or a combination of transmission-line sections.

10.2 FIELD-EFFECT TRANSISTORS

There are two main characteristics of field-effect transistors that make them superior to bipolar transistors in microwave amplifiers. These are the lower noise characteristics and the higher frequency of operation. The higher operating frequency is due to the higher electron mobility in gallium arsenide, which is the material used in field-effect transistors, compared to that in silicon, the standard bipolar transistor material. The higher electron mobility and the absence of shot noise are important features that result in low noise. The first microwave field-effect transistors were metal-semiconductor field-effect transistors (MESFETs). High-frequency operation required a very short gate length, typically less than 1 μm. Thus it was only after the processing technology had advanced to the stage that submicron device features could be reliably made that microwave solid-state device development advanced rapidly. Currently produced MESFETs have gate lengths of the order 0.3 to 0.5 μm. The frequency at which the short-circuit current gain becomes equal to unity is given approximately by the relationship $f_T = v_s/2\pi L_g$, where v_s is the electron saturation velocity and L_g is the gate length. In gallium arsenide (GaAs) the maximum drift velocity is about 2×10^7 cm/s, so that for a gate length of 0.5 μm, $f_T = 60$ GHz. Clearly, very short gate lengths are essential for high-frequency operation.

Beginning around 1980, a new technology involving heterojunctions began to find applications in device construction. A heterojunction is a junction formed at the interface of say an aluminum-gallium-arsenide (AlGaAs) doped alloy and an undoped GaAs layer. The use of a heterojunction enables a channel with a very high electron mobility to be obtained. The field-effect transistor which is made using a heterojunction is called a high-electron-mobility transistor (HEMT). Since high mobility is achieved by doping only the large bandgap material, the name modulation-doped

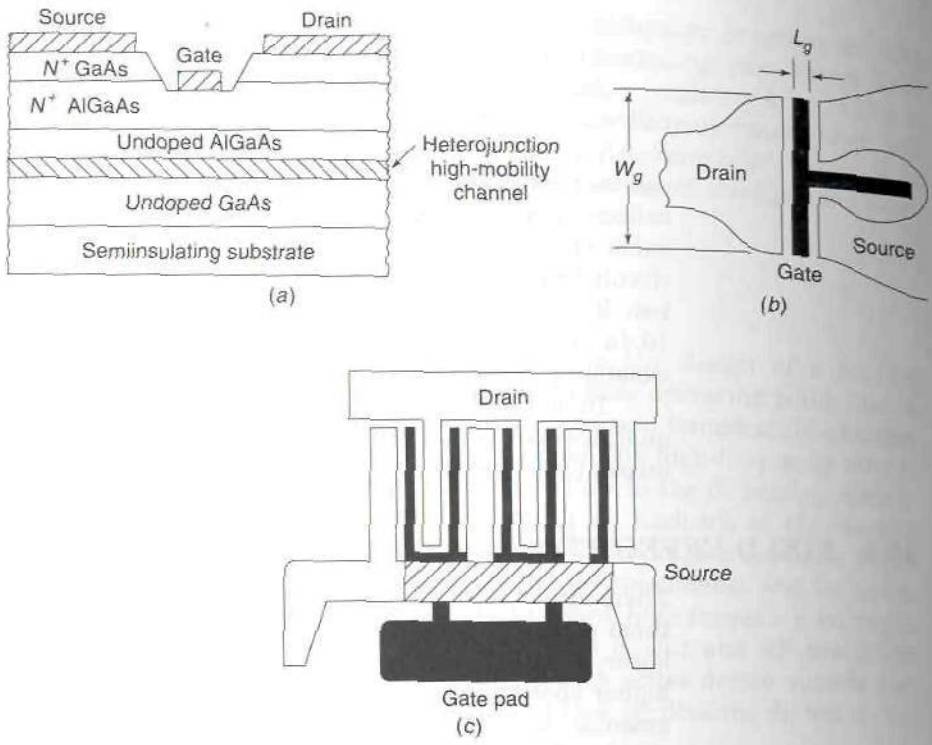


FIGURE 10.5
(*a*) Cross section of a high-electron-mobility transistor (HEMT); (*b*) source, drain, and gate structure for a low-power FET; (*c*) interdigitated construction used for a power FET. The source metalization passes above the gate structure and is insulated from it.

field-effect transistor (MODFET) is also used. The HEMT device has a higher frequency of operation and a lower noise figure than the standard MESFET device. A good discussion of the technology and fabrication of HEMT devices and the theory of operation can be found in the books by Chang and Pengelly.†

Figure 10.5*a* shows the cross section of a typical HEMT device, while Fig. 10.5*b* shows a top view of the source, drain, and gate structure. The gate width W_g is much greater than the gate length L_g. In a power FET as many as 10 source and drain fingers adjacent to gate fingers are used to obtain large drain currents and hence large output powers. The interdigi-

†K. Chang (ed.), "Handbook of Microwave and Optical Components. Microwave Solid-State Components," *vol. 2, John Wiley & Sons Inc., New York, 1990.*
R. S. Pengelly, "Microwave Field-Effect Transistors—Theory, Design and Applications," *2nd ed., Research Studies Press, Letchworth, England, 1986.*

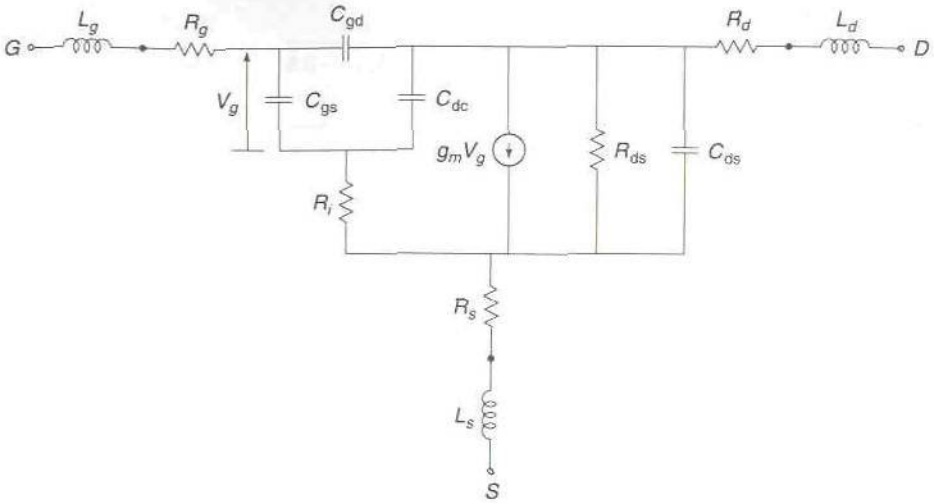


FIGURE 10.6
Simplified small-signal equivalent-circuit model of a microwave GaAs FET.

tated construction is shown in Fig. 10.5*c*. The transconductance g_m of the FET is increased by using a large gate effective width and this is needed for large power output. However, with a larger gate width the input capacitance is increased.

A simplified small-signal equivalent-circuit model of a microwave GaAs FET is shown in Fig. 10.6. Typical values for the circuit parameters for a device with a 1-μm gate length and a 300-μm gate width are listed below:

$$\text{Gate-to-source capacitance } C_{gs} = 0.4 \text{ pF}$$
$$\text{Gate-to-drain capacitance } C_{gd} = 0.01 \text{ pF}$$
$$\text{Channel resistance } R_{ds} = 500\ \Omega$$
$$\text{Transconductance } g_m = 30 \text{ mS}$$
$$R_i = 3\ \Omega$$
$$C_{dc} = 0.015 \text{ pF}$$

The circuit parameters R_g, R_d, R_s, and C_{ds} are extrinsic elements. The gate, drain, and source ohmic contact resistances R_g, R_d, and R_s are typically a few ohms. The drain-to-substrate capacitance $C_{ds} = 0.07$ pF. The inductors L_g, L_s, and L_d have inductances in the range 0.05 to 0.3 nH. The equivalent current source is $V_g g_m$, where V_g is the signal voltage across C_{gs}. The microwave FET can also be described in terms of measured scattering-matrix parameters.

Gallium-arsenide MESFET devices can give a single-stage gain of 8 to 15 dB at 2 GHz with noise figures below 1 dB. For HEMT devices a

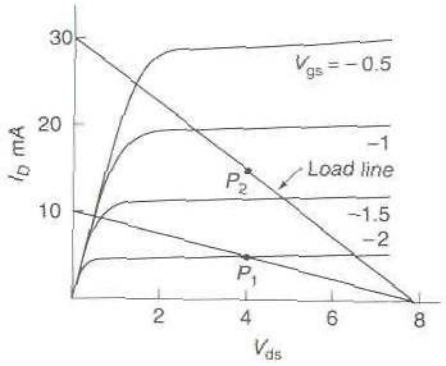


FIGURE 10.7
Output characteristics for a GaAs MESFET showing suitable operating points P_1 and P_2 for small- and large-signal inputs.

single-stage gain of 15 dB at 8 GHz and 6 dB at 50 GHz can be achieved with corresponding noise figures of 0.4 and 1.8 dB, respectively.† For power applications output powers of several watts from a single device can be obtained. Several devices may be operated in parallel using power-combining techniques to achieve higher output powers.

FET Biasing

The output drain current versus drain-to-source voltage V_{ds} is shown in Fig. 10.7 as a function of gate-to-source voltage for a typical microwave GaAs MESFET. For small-signal application an operating point in the vicinity of the point P_1 would be suitable. For maximum dynamic range the operating point should be placed in the central region of the output characteristics which is depicted by point P_2. In either case the dc voltage of the gate must be negative with respect to that of the source for a depletion mode device.

This bias condition can be achieved by grounding the gate through an RF choke or a high-impedance quarter-wavelength transmission line and obtaining the desired bias voltage from the voltage drop across the source resistance R_s as shown in Fig. 10.8. The required value of R_s is given by $R_s = -V_{gs}/I_{ds}$. The source resistance R_s should be bypassed to ground for RF signals by means of a capacitor C_s. In small-signal low-noise applications, the best noise figure is obtained for a dc drain current equal to about 20 percent of the drain saturation current at zero gate-to-source bias voltage. At low values of drain current, the transconductance is reduced, so that gain must be sacrificed to achieve a low noise figure.

†K. Chang, *loc. cit.*, p. 465.

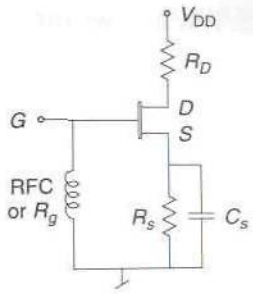


FIGURE 10.8
FET bias circuit.

10.3 CIRCLE-MAPPING PROPERTIES OF BILINEAR TRANSFORMATIONS

A circle with center at x_0, y_0 and with radius R is described by $(x - x_0)^2 + (y - y_0)^2 - R^2 = 0$ or

$$x^2 + y^2 - 2xx_0 - 2yy_0 + (x_0^2 + y_0^2 - R^2) = 0$$

Now let $Z = x + jy$, $Z_0 = x_0 + jy_0$. The circle equation can be written as $|Z - Z_0|^2 - R^2 = 0$ or $(Z - Z_0)(Z^* - Z_0^*) - R^2 = 0$ which gives

$$ZZ^* - ZZ_0^* - Z^*Z_0 + (Z_0Z_0^* - R^2) = 0 \tag{10.1}$$

Consider now the bilinear transformation from the complex Z plane to the complex W plane given by

$$W = \frac{AZ + B}{CZ + D}$$

This transformation will map circles in the Z plane into circles in the W plane (straight lines are limiting cases). Consider the circle $|W|^2 = \rho^2$ or $WW^* - \rho^2 = 0$. Using the transformation, we get

$$\frac{AZ + B}{CZ + D}\frac{A^*Z^* + B^*}{C^*Z^* + D^*} - \rho^2 = 0$$

By expanding this equation we get

$$ZZ^*(AA^* - \rho^2CC^*) - Z(\rho^2CD^* - AB^*) - Z^*(\rho^2C^*D - A^*B)$$
$$+ BB^* - \rho^2DD^* = 0 \tag{10.2}$$

By comparison with (10.1) we see that this is a circle with center at (coefficient of $-Z^*$)

$$Z_0 = \frac{\rho^2C^*D - A^*B}{AA^* - \rho^2CC^*} = \frac{\rho^2C^*D - A^*B}{|A|^2 - \rho^2|C|^2} \tag{10.3}$$

The constant term equals $|Z_0|^2 - R^2$, so that we can identify the radius R^2 as given by

$$R^2 = Z_0 Z_0^* - \frac{|B|^2 - \rho^2|D|^2}{|A|^2 - \rho^2|C|^2}$$

from which we get

$$R = \rho \frac{|AD - BC|}{\left||A|^2 - \rho^2|C|^2\right|} \tag{10.4}$$

If the circle in the W plane is $|W - W_0| = \rho$, then we note that

$$W - W_0 = \frac{AZ + B}{CZ + D} - W_0 = \frac{(A - CW_0)Z + (B - DW_0)}{CZ + D}$$

If we define $A' = A - CW_0$, $B' = B - DW_0$, then the above formulas (10.3) and (10.4) apply with A, B replaced by A', B'.

Many of the relationships that occur in amplifier design involve bilinear transformations and their circle-mapping properties can easily be identified by comparison with the above equations.

10.4 MICROWAVE AMPLIFIER DESIGN USING S_{ij} PARAMETERS

At microwave frequencies impedance and admittance parameters of a transistor cannot be directly measured. The scattering-matrix parameters can be measured and therefore a design methodology based on using the S_{ij} parameters is widely used. The S_{ij} parameters are measured by inserting the transistor into a test circuit with 50-Ω input and output lines, applying appropriate bias voltages and currents, and measuring the S_{ij} parameters. In any design using S_{ij} parameters, it should be kept in mind that these vary with bias conditions, temperature, and from transistor to transistor even if it is the same device number. Thus the design should leave some margin for S_{ij} variations.

The following are the usual microwave amplifier design goals:

1. Maximum power gain.
2. Minimum noise figure for the first stage. This requires a specific source impedance $\bar{Z}_s$ for the input stage. The optimum $\bar{Z}_s$ giving the lowest noise figure is generally given by the manufacturer of the transistor.
3. Stable gain, i.e., no oscillations.
4. Input and output VSWR as close to unity as possible.
5. Adequate gain and uniformity of gain over a specified frequency band.

6. Phase response that is a linear function of ω (no distortion, only group delay).
7. Insensitivity to nominal changes or variations in the device S_{ij} parameters.

These objectives cannot all be realized at the same time; so the design procedure must trade off one objective against another one, e.g., gain must be sacrificed for stability. Input VSWR must be sacrificed for a low noise figure since the normalized source impedance $\overline{Z}_s$ is fixed for minimum noise.

Designing a microwave amplifier using potentially unstable transistors is like designing a bridge using interconnected beams whose lengths can be telescoped and with joints that are free to rotate with insufficient external constraints to define a rigid structure. It is a "loose-jointed" problem without a unique solution. The many design specifications are all interrelated which makes the problem almost unmanageable without some computer optimization strategy. In the next several sections we will examine in detail the constraints that are imposed by stability requirements, by the need for large power gain, low noise, and low input and output VSWRs. We can obtain a good physical understanding of these often conflicting requirements from a study of the stability circles, the constant power-gain circles, the constant impedance-mismatch circles, and the constant noise-figure circles plotted on a Smith chart. From the insight obtained from such a study, we will be able to formulate a design strategy using some computer optimization that will lead to satisfactory designs. The reader will need to bear with us as we work our way through the maze of details but will in the end be rewarded with the satisfaction of having obtained the necessary insight and understanding to be able to implement the theory for practical amplifier design.

All the various low-frequency amplifier circuits such as balanced push-pull amplifiers, cascode amplifiers, and traveling-wave amplifiers can also be used at microwave frequencies. A device that is potentially unstable can be stabilized by resistive loading at the input, output, or both input and output, with a resultant reduction in power gain and increase in noise figure. Stabilization can also be achieved by using negative feedback. Space limitations do not allow a detailed discussion of the variety of amplifier circuits and configurations that are used in practice. Many of these circuits are described in the book by Pengelly and the one by Vendelin, Pavio, and Rohde already cited. The reader is referred to these texts for a detailed discussion. In this text we will limit our attention to amplifier design based on the use of scattering-matrix parameters and linear two-port design.

For high-power amplifier design, it is necessary to consider the nonlinear characteristics of transistors and to pay more attention to device power

dissipation and the design of adequate heat sinks. Small-signal amplifier design can often provide a first approximation to the design of large-signal amplifiers, particularly in connection with stability. Thus the linear two-port design methodology developed in this chapter will also provide useful insight into the large-signal design problem even though we do not cover the latter topic.

10.5 AMPLIFIER POWER GAIN

We will begin our discussion by deriving expressions for the power gain of an amplifier. There are several definitions used for the gain of an amplifier and they are given below.

$$\text{Power gain } G_p = \frac{\text{power delivered to load}}{\text{input power to amplifier}} \tag{10.5a}$$

$$\text{Transducer gain } G = \frac{\text{power delivered to load}}{\text{available input power from source}} \tag{10.5b}$$

$$\text{Available power gain } G_a = \frac{\text{available load power}}{\text{available input power from source}} \tag{10.5c}$$

If the device is unconditionally stable, then conjugate impedance matching can be used at both the input and output. If this is done then $G_p = G = G_a = G_{\max}$ = maximum gain. For a device that is only conditionally stable, conjugate impedance matching at both the input and output cannot be used when the stability parameter $K < 1$. The power gain achieved in this case is G_p. Power gain is the most useful definition in practice since it applies to any actual amplifier independent of whether conjugate impedance matching is used.

For the two-stage amplifier shown in Fig. 10.9, the incident power is given by

$$P_{\text{inc}} = \tfrac{1}{2}|V^+|^2 Y_c$$

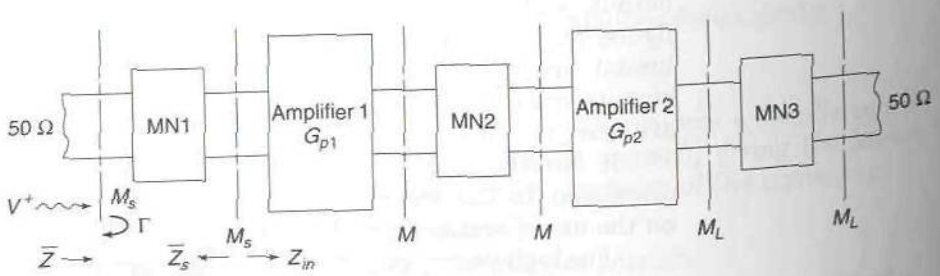


FIGURE 10.9
A two-stage amplifier with matching networks.

and the input power is

$$P_{\text{in}} = (1 - |\Gamma|^2)P_{\text{inc}}$$

where Y_c is the characteristic admittance of the input line and Γ is the input reflection coefficient. The power P_L delivered to the output 50-Ω line is given by $G_{p1}G_{p2}P_{\text{in}}$. Hence the two-stage power gain is

$$G_p = \frac{P_L}{P_{\text{in}}} = G_{p1}G_{p2} \tag{10.6}$$

and the corresponding transducer gain is

$$G = \frac{P_L}{P_{\text{inc}}} = (1 - |\Gamma|^2)G_{p1}G_{p2} \tag{10.7}$$

For lossless matching networks the impedance mismatch is the same on the input side as on the output side as shown in Sec. 5.7. If M_s is the impedance mismatch between the first amplifier input and its source impedance $\overline{Z}_s$ as seen looking into the output side of the first matching network, then

$$M_s = \frac{4\overline{R}_s\overline{R}_{\text{in}}}{|\overline{Z}_s + \overline{Z}_{\text{in}}|^2}$$

where $\overline{Z}_{\text{in}} = \overline{R}_{\text{in}} + j\overline{X}_{\text{in}}$ is the amplifier input impedance. On the input line

$$M = \frac{4Z_cR}{|Z_c + Z|^2} = \frac{4\overline{R}}{|1 + \overline{Z}|^2}$$

where $\overline{Z} = (1 - \Gamma)/(1 + \Gamma)$. When we use the relations

$$2\overline{R} = \overline{Z} + \overline{Z}^* = \frac{1 + \Gamma}{1 - \Gamma} + \frac{1 + \Gamma^*}{1 - \Gamma^*}$$

and

$$1 + \overline{Z} = \frac{2}{1 + \Gamma}$$

we find that

$$M = 1 - \Gamma\Gamma^* = 1 - |\Gamma|^2 = M_s \tag{10.8}$$

upon substituting for $\overline{R}$ and $1 + \overline{Z}$ in terms of Γ and simplifying the resultant expression. The input voltage standing-wave ratio is given by

$$\text{VSWR}_1 = \frac{1 + |\Gamma|}{1 - |\Gamma|} = \frac{1 + \sqrt{1 - M}}{1 - \sqrt{1 - M}} = \frac{1 + \sqrt{1 - M_s}}{1 - \sqrt{1 - M_s}} \tag{10.9}$$

Thus the degree of mismatch between $\overline{Z}_s$ and the input to stage 1 determines the input VSWR. If we have a constraint on VSWR_1 in our design and we want to use an optimum source impedance $\overline{Z}_s$ for minimum noise, then we must terminate stage 1 in a load that will produce an impedance mismatch M_s that will keep the input VSWR_1 at the specified value (or

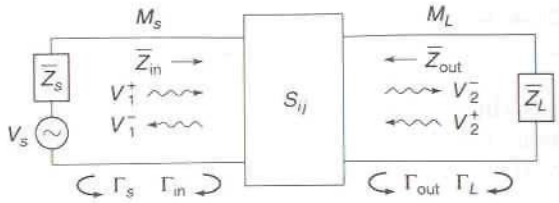


FIGURE 10.10
A basic amplifier circuit.

lower). The required load termination must, of course, not lead to an unstable (oscillating) amplifier and must also yield good power gain. A power gain of 10 or more is desirable, since this will make the effect of following stages on the noise figure small. Sometimes the input VSWR constraint has to be relaxed because it is in conflict with other requirements.

⋆Derivation of Expressions for Gain

Consider the basic amplifier circuit shown in Fig. 10.10. The source and load are viewed as connected to the amplifier by means of transmission lines with characteristic impedance Z_c and having negligible lengths. For this circuit the source and load reflection coefficients are given by

$$\Gamma_L = \frac{\overline{Z}_L - 1}{\overline{Z}_L + 1} \qquad \text{and} \qquad \Gamma_s = \frac{\overline{Z}_s - 1}{\overline{Z}_s + 1}$$

For the amplifier we have

$$V_1^- = S_{11}V_1^+ + S_{12}V_2^+$$
$$V_2^- = S_{21}V_1^+ + S_{22}V_2^+$$

But $V_2^+ = \Gamma_L V_2^-$ so $V_2^- = S_{21}V_1^+ + S_{22}\Gamma_L V_2^-$ or $V_2^- = S_{21}V_1^+/(1 - S_{22}\Gamma_L)$. We use this in the first equation to get

$$V_1^- = V_1^+\left(S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}\right)$$

and hence

$$\frac{V_1^-}{V_1^+} = \Gamma_{in} = \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L} \tag{10.10}$$

where $\Delta = S_{11}S_{22} - S_{12}S_{21}$.

In a similar way we find that

$$\Gamma_{out} = \frac{S_{22} - \Delta\Gamma_s}{1 - S_{11}\Gamma_s} \tag{10.11}$$

The input power to the amplifier is given by

$$P_{\text{in}} = \frac{|V_s|^2}{8\overline{R}_s} M_s = \frac{|V_s|^2 \overline{R}_{\text{in}}}{2|\overline{Z}_s + \overline{Z}_{\text{in}}|^2}$$

where $|V_s|^2/8\overline{R}_s$ is the available power from the source. First, we will express M_s in terms of Γ_s and Γ_{in}. To do this, we use

$$\overline{Z}_s = \frac{1 + \Gamma_s}{1 - \Gamma_s}$$

and

$$2\overline{R}_s = \overline{Z}_s + \overline{Z}_s^* = \frac{1 + \Gamma_s}{1 - \Gamma_s} + \frac{1 + \Gamma_s^*}{1 - \Gamma_s^*} = \frac{2(1 - |\Gamma_s|^2)}{|1 - \Gamma_s|^2}$$

since $(1 - \Gamma_s)(1 - \Gamma_s^*) = |1 - \Gamma_s|^2$. Similarly, we get

$$2\overline{R}_{\text{in}} = \frac{2(1 - |\Gamma_{\text{in}}|^2)}{|1 - \Gamma_{\text{in}}|^2}$$

Hence we obtain

$$M_s = \frac{4\overline{R}_{\text{in}}\overline{R}_s}{|\overline{Z}_s + \overline{Z}_{\text{in}}|^2} = \frac{4(1 - |\Gamma_{\text{in}}|^2)(1 - |\Gamma_s|^2)}{|1 - \Gamma_{\text{in}}|^2|1 - \Gamma_s|^2} \frac{1}{\left|\dfrac{1 + \Gamma_s}{1 - \Gamma_s} + \dfrac{1 + \Gamma_{\text{in}}}{1 - \Gamma_{in}}\right|^2}$$

$$= \frac{4(1 - |\Gamma_{\text{in}}|^2)(1 - |\Gamma_s|^2)}{|1 - \Gamma_{\text{in}}|^2|1 - \Gamma_s|^2} \frac{|1 - \Gamma_s|^2|1 - \Gamma_{\text{in}}|^2}{|(1 + \Gamma_s)(1 - \Gamma_{\text{in}}) + (1 - \Gamma_s)(1 + \Gamma_{\text{in}})|^2}$$

$$= \frac{(1 - |\Gamma_{\text{in}}|^2)(1 - |\Gamma_s|^2)}{|1 - \Gamma_s\Gamma_{\text{in}}|^2} \tag{10.12a}$$

Similarly, at the output

$$M_L = \frac{(1 - |\Gamma_{\text{out}}|^2)(1 - |\Gamma_L|^2)}{|1 - \Gamma_L\Gamma_{\text{out}}|^2} \tag{10.12b}$$

The load voltage $V_L = V_2^- + V_2^+ = V_2^-(1 + \Gamma_L)$. The load current is $I_L = Y_c(V_2^- - V_2^+)$, so that the power delivered to the load is given by

$$P_L = \tfrac{1}{2}\,\text{Re}\, V_L I_L^* = \tfrac{1}{2}Y_c|V_2^-|^2(1 + \Gamma_L)(1 - \Gamma_L^*)$$

Now

$$(1 + \Gamma_L)(1 - \Gamma_L^*) = 1 - \Gamma_L\Gamma_L^* + \Gamma_L - \Gamma_L^* = 1 - |\Gamma_L|^2 + 2\,\text{Im}\,\Gamma_L$$

so

$$P_L = \tfrac{1}{2}|V_2^-|^2(1 - |\Gamma_L|^2)Y_c$$

as expected. We now use the expression for V_2^- given above (10.10) to obtain

$$P_L = \frac{1}{2}(1 - |\Gamma_L|^2)\frac{|S_{21}|^2}{|1 - S_{22}\Gamma_L|^2}|V_1^+|^2 Y_c$$

The input power is $\frac{1}{2}|V_1^+|^2(1 - |\Gamma_{in}|^2)\, Y_c$. Hence

$$G_p = \frac{P_L}{P_{in}} = \frac{(1 - |\Gamma_L|^2)|S_{21}|^2}{(1 - |\Gamma_{in}|^2)|1 - S_{22}\Gamma_L|^2} \tag{10.13}$$

The transducer gain is given by

$$G = \frac{P_L}{P_{ava}} = \frac{P_L}{|V_s|^2/8\overline{R}_s}$$

The available power P_{ava} is related to the input power by $P_{in} = MP_{ava} = M_s P_{ava}$ and thus the expression for transducer gain becomes

$$G = \frac{P_L}{P_{in}}M_s = \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_s|^2)|S_{21}|^2}{|1 - S_{22}\Gamma_L|^2|1 - \Gamma_{in}\Gamma_s|^2} \tag{10.14}$$

We now eliminate Γ_{in} from the gain expressions by using (10.10) for Γ_{in}, that is,

$$1 - |\Gamma_{in}|^2 = 1 - \frac{|S_{11} - \Delta\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} = \frac{|1 - S_{22}\Gamma_L|^2 - |S_{11} - \Delta\Gamma_L|}{|1 - S_{22}\Gamma_L|^2}$$

The elimination of Γ_{in} gives us

$$G_p = \frac{(1 - |\Gamma_L|^2)|S_{21}|^2}{|1 - S_{22}\Gamma_L|^2 - |S_{11} - \Delta\Gamma_L|^2}$$

Next we use $|1 - S_{22}\Gamma_L|^2 = (1 - S_{22}\Gamma_L)(1 - S_{22}^*\Gamma_L^*)$ and similarly for $|S_{11} - \Delta\Gamma_L|^2$ to get for the denominator

$$\begin{aligned} &1 + S_{22}S_{22}^*\Gamma_L\Gamma_L^* - S_{22}\Gamma_L - S_{22}^*\Gamma_L^* - S_{11}S_{11}^* + S_{11}\Delta^*\Gamma_L^* \\ &\quad + S_{11}^*\Delta\Gamma_L - \Delta\Delta^*\Gamma_L\Gamma_L^* \\ &\quad = 1 - |S_{11}|^2 + |\Gamma_L|^2(|S_{22}|^2 - |\Delta|^2) - 2\,\mathrm{Re}\,\Gamma_L(S_{22} - \Delta S_{11}^*) \end{aligned}$$

Hence

$$G_p = \frac{(1 - |\Gamma_L|^2)|S_{21}|^2}{1 - |S_{11}|^2 + |\Gamma_L|^2(|S_{22}|^2 - |\Delta|^2) - 2\,\mathrm{Re}\,\Gamma_L(S_{22} - \Delta S_{11}^*)} \tag{10.15}$$

The same substitution for Γ_{in} reduces the expression for G to

$$G = \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_s|^2)|S_{21}|^2}{|1 - S_{22}\Gamma_L - S_{11}\Gamma_s + \Delta\Gamma_s\Gamma_L|^2}$$

$$= \frac{(1 - |\Gamma_L|^2)(1 - |\Gamma_s|^2)|S_{21}|^2}{|(1 - S_{22}\Gamma_L)(1 - S_{11}\Gamma_s) - S_{12}S_{21}\Gamma_s\Gamma_L|^2} \qquad (10.16)$$

These are the final forms for the power gain and transducer gain.

When the device (transistor) is absolutely stable, we can use conjugate impedance matching, i.e., choose $\bar{Z}_s = \bar{Z}_{in}^*$, $\bar{Z}_L = \bar{Z}_{out}^*$. In this case $\Gamma_s = \Gamma_{in}^*$, $\Gamma_L = \Gamma_{out}^*$, and $M_s = M_L = 1$. Clearly, when $M_s = 1$, the power gain G_p equals the transducer gain G. For conjugate impedance matching we require

$$\Gamma_s^* = \Gamma_{in} = \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L} \qquad \text{and} \qquad \Gamma_L = \Gamma_{out}^* = \frac{S_{22}^* - \Delta^*\Gamma_s^*}{1 - S_{11}^*\Gamma_s^*}$$

We can substitute the second equation into the first one and solve for Γ_s which gives

$$\Gamma_s = \Gamma_{sM} = \frac{1}{2B_1}\left[A_1 \pm \left(A_1^2 - 4|B_1|^2\right)^{1/2}\right] \qquad (10.17a)$$

Similarly,

$$\Gamma_L = \Gamma_{LM} = \frac{1}{2B_2}\left[A_2 \pm \left(A_2^2 - 4|B_2|^2\right)^{1/2}\right] \qquad (10.17b)$$

where

$$A_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2$$

$$A_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2$$

$$B_1 = S_{11} - \Delta S_{22}^*$$

$$B_2 = S_{22} - \Delta S_{11}^*$$

The minus sign is used when $A_i > 0$ and the plus sign is used when $A_i < 0$ in order to get $|\Gamma_{sM}| \leq 1$ and $|\Gamma_{LM}| \leq 1$. We will show in the next section that for an absolutely stable device $A_1 > 0$ and $A_2 > 0$, so the minus sign is the appropriate one to use in (10.17). When $A_2^2 < 4|B_2|^2$ the solution for Γ_{LM} can be expressed as $[A_2 \pm j(4|B_2|^2 - A_2^2)^{1/2}]/2B_2$. For this case the magnitude of Γ_{LM} is equal to unity which corresponds to a pure reactive load termination and zero power gain. Thus, for an absolutely stable transistor, we must have $|A_2| > 2|B_2|$. It is this condition that leads to the choice of sign to use in (10.17).

In the next section we will show that a device is absolutely stable if the stability parameter K is greater than one, i.e.,

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} > 1 \tag{10.18}$$

In terms of the parameter K, the power gain for an absolutely stable device, using conjugate impedance matching, is given by

$$G_p = G_{p,\max} = G = G_{\max} = \left|\frac{S_{21}}{S_{12}}\right|\left(K - \sqrt{K^2 - 1}\right) \tag{10.19}$$

The parameter $|S_{21}/S_{12}|$ is called the "Figure of Merit" for the transistor. When $K = 1$ it gives the maximum stable gain $G_{\text{MSG}} = |S_{21}/S_{12}|$. The expression for $G_{\max}$ for an amplifier with conjugate impedance matching can be derived following the steps outlined below. By direct expansion we can show that

$$A_2^2 - 4|B_2|^2 = 4|S_{12}S_{21}|^2(K^2 - 1)$$

From the equations for A_2 and K, we can show that

$$1 - |S_{11}|^2 = A_2 + |\Delta|^2 - |S_{22}|^2 = 2K|S_{12}S_{21}| - \left(|\Delta|^2 - |S_{22}|^2\right)$$

so by addition and subtraction we get

$$1 - |S_{11}|^2 = K|S_{12}S_{21}| + \frac{A_2}{2}$$

$$|\Delta|^2 - |S_{22}|^2 = K|S_{12}S_{21}| - \frac{A_2}{2}$$

We also take note of the relationship

$$2\Gamma_{LM}(S_{22} - S_{11}^*\Delta) = A_2 \pm \left(A_2^2 - 4|B_2|^2\right)^{1/2}$$

which is a real quantity. By using these expressions the denominator in (10.15), which is the equation for G_p, can be written as

$$K|S_{12}S_{21}| + \frac{A_2}{2} + |\Gamma_L|^2\left(\frac{A_2}{2} - K|S_{12}S_{21}|\right) - A_2 \mp 2|S_{12}S_{21}|\sqrt{K^2 - 1}$$

$$= \left(1 - |\Gamma_L|^2\right)\left[\frac{\mp 2|S_{12}S_{21}|\sqrt{K^2 - 1}}{1 - |\Gamma_L|^2} - \left(\frac{A_2}{2} - K|S_{12}S_{21}|\right)\right]$$

The last step is to use the solution Γ_{LM} for Γ_L to get

$$1 - |\Gamma_L|^2 = \frac{2|S_{12}S_{21}|\sqrt{K^2 - 1}}{|B_2|^2}\left(\mp\frac{A_2}{2} - |S_{12}S_{21}|\sqrt{K^2 - 1}\right)$$

With this substitution and a few more algebraic steps, we get

$$G_p = \frac{|S_{21}|^2}{|S_{12}S_{21}|\left(K \mp \sqrt{K^2 - 1}\right)} = \left|\frac{S_{21}}{S_{12}}\right|\left(K \pm \sqrt{K^2 - 1}\right)$$

There are two possible solutions for G_p which correspond to the two possible solutions for Γ_{LM}. Since G_p should not keep increasing for $K \gg 1$, only the solution given by (10.19) corresponds to the gain obtained with a passive load. We will have a further comment on these two possible solutions at a later point. For now we note that the chosen solution corresponds to using the minus sign in (10.17*b*) which implies that $A_2 > 0$.

10.6 AMPLIFIER STABILITY CRITERIA

When the transistor is potentially unstable, which occurs when the stability factor K given by (10.18) is less than one, a stable amplifier can still be designed but only for restricted values of source and load impedances. Furthermore, it will not be possible to use conjugate impedance matching at both the input and output ports. In this section we will derive the expressions for the allowed terminating impedances in terms of input and output reflection-coefficient stability circles. These stability circles can be plotted on a Smith chart and will show what values of source and load impedances can be used in order to achieve a stable (nonoscillating) amplifier.

The conditions for amplifier stability are established by requiring that the reflected power from the amplifier ports be smaller than the incident power. This means that the reflection coefficients looking into the amplifier ports must have a magnitude less than one for all passive source and load impedances. If a reflection coefficient has a magnitude greater than unity, the amplifier input or output impedance would have a negative real part, e.g., if $Z_{in} = -R + jX$ then

$$|\Gamma_{in}| = \left|\frac{-R + jX - Z_c}{-R + jX + Z_c}\right| = \left[\frac{(R + Z_c)^2 + X^2}{(Z_c - R)^2 + X^2}\right]^{1/2} > 1$$

If $Z_{in} = -R_{in} + jX_{in}$ then the input current is

$$I = \frac{V_s}{R_s - R_{in} + j(X_{in} + X_s)}$$

If $R_s = R_{in}$ and $X_{in} + X_s = 0$ which can occur at some frequency, then I becomes infinite. We can set $V_s = 0$ and thermal noise in the input can produce self-sustained oscillations at the frequency where the total loop impedance in the input equals zero. Oscillation at any frequency generally makes the amplifier unusable.

The conditions for stability are

$$|\Gamma_{in}| = \left|\frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L}\right| < 1 \qquad \text{for all } |\Gamma_L| < 1 \tag{10.20a}$$

$$|\Gamma_{out}| = \left|\frac{S_{22} - \Delta\Gamma_s}{1 - S_{11}\Gamma_s}\right| < 1 \qquad \text{for all } |\Gamma_s| < 1 \tag{10.20b}$$

When these conditions hold conjugate impedance matching can be used. In general, we find that $|\Gamma_{in}| < 1$ for only a restricted set of values for Γ_L (i.e., Z_L) and $|\Gamma_{out}| < 1$ for only a restricted set of values for the source impedance Z_s. In this circumstance the device is said to be conditionally stable. If $Z_L = Z_c$, then $\Gamma_L = 0$ and $|\Gamma_{in}| < 1$ only if $|S_{11}| < 1$. Similarly, if $Z_s = Z_c$, $|\Gamma_{out}| < 1$ only if $|S_{22}| < 1$. Hence two necessary conditions for absolute stability are

$$|S_{11}| < 1 \qquad \text{and} \qquad |S_{22}| < 1$$

Values of Γ_L that result in $|\Gamma_{in}| < 1$ are called stable ones. The corresponding region of the Smith chart is the stable region. The boundary between stable values of Γ_L and unstable ones is the circle in the Γ_L plane that corresponds to the mapping of the circle $|\Gamma_{in}| = 1$ *in the Γ_{in} plane.* From the bilinear transformation

$$\Gamma_{in} = \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L} = \frac{\Delta\Gamma_L - S_{11}}{S_{22}\Gamma_L - 1}$$

we find that the center of the mapped circle is at ($\Delta = A$, $-S_{11} = B$, $S_{22} = C$, $-1 = D$)

$$\Gamma_{LC} = \frac{S_{11}\Delta^* - S_{22}^*}{|\Delta|^2 - |S_{22}|^2} \tag{10.21a}$$

and the radius of the circle is

$$R_{LC} = \frac{|S_{12}S_{21}|}{\left|(|\Delta|^2 - |S_{22}|^2)\right|} \tag{10.21b}$$

This load stability circle may or may not include the origin $\Gamma_L = 0$ as shown in Figs. 10.11*a* and *b*. When $\Gamma_L = 0$, $\Gamma_{in} = S_{11}$ and $|S_{11}| < 1$ for a stable circuit. Hence, if the stability circle encloses the origin, then all values of Γ_L inside the circle will give values for Γ_{in} such that $|\Gamma_{in}| < 1$. These values of Γ_L are stable ones. If $\Gamma_L = 0$ lies outside the circle, then all values of Γ_L outside the circle are stable ones. The origin is included only if the radius R_{LC} is greater than the distance $|\Gamma_{LC}|$ to the center. If this situation prevails then we require $R_{LC} > 1 + |\Gamma_{LC}|$, so that all values of $|\Gamma_L| \leq 1$ lie inside the circle and represent stable values for all possible passive load terminations. If the origin is not included, then we require that $|\Gamma_{LC}| > 1 + R_{LC}$, so that

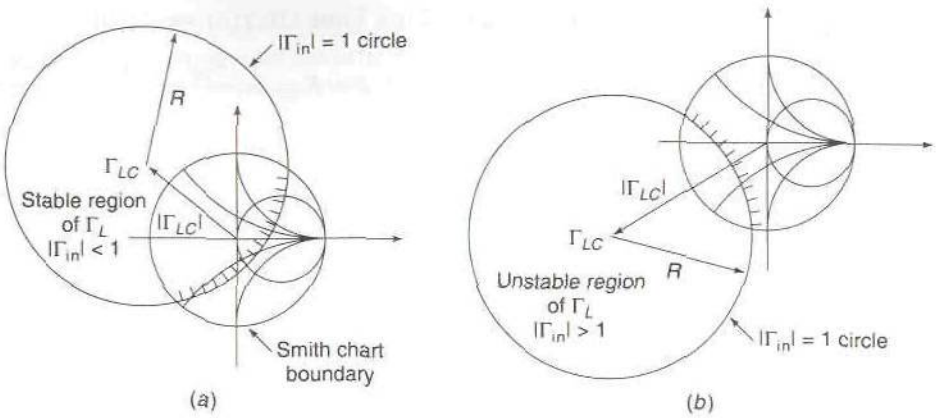


FIGURE 10.11
Load stability circles plotted on the Smith chart. (*a*) Origin is included within stable region; (*b*) origin is excluded so load impedances outside stability circle are stable values.

the entire stability circle $|\Gamma_{in}| = 1$ maps into a circle outside the Smith chart boundary $|\Gamma_L| = 1$. Then again all values of $|\Gamma_L| < 1$ will be stable ones.

We need to find an expression that will enable us to state, in terms of the scattering-matrix parameters S_{11}, S_{12}, S_{21}, and S_{22} of the device, whether or not it is an absolutely stable device. If the amplifier will be stable for all passive source and load impedances, then the device is absolutely stable. If it is stable only for a limited set of source and load terminations, then the device is only conditionally stable. There are two cases to consider and they both lead to the same statement or criterion for stability.

Case 1. Origin ($\Gamma_L = 0$) lies outside the circles of Γ_L values that make $|\Gamma_{in}| = 1$. As noted earlier this case corresponds to all values of Γ_L outside the load stability circle $|\Gamma_{in}| = 1$ being stable values. All passive values of Z_L will be acceptable if the $|\Gamma_{in}| = 1$ circle lies outside the Smith chart boundary $|\Gamma_L| = 1$. This requires that the distance $|\Gamma_{LC}|$ to the center of the circle be greater than the radius plus one, that is, $|\Gamma_{LC}| > 1 + R_{LC}$. In order to have $|\Gamma_{LC}| > 1 + R_{LC}$, we must have $|\Gamma_{LC}|^2 > R_{LC}^2$. From (10.21) we see that this requires $|S_{11}\Delta^* - S_{22}^*|^2 > |S_{12}S_{21}|^2$. By direct expansion we can show that

$$|S_{11}\Delta^* - S_{22}^*|^2 = |S_{12}S_{21}|^2 + \left(1 - |S_{11}|^2\right)\left(|S_{22}|^2 - |\Delta|^2\right) \qquad (10.22)$$

Hence our stated condition is equivalent to

$$\left(1 - |S_{11}|^2\right)\left(|S_{22}|^2 - |\Delta|^2\right) > 0$$

which is possible only if $|S_{22}| > |\Delta|$ since $|S_{11}| < 1$. Thus Case 1 can occur only when this condition is true. When the condition is true, then $|(|S_{22}|^2 - |\Delta|^2)| = |S_{22}|^2 - |\Delta|^2$. The stability condition $1 + R_{LC} < |\Gamma_{LC}|$ can be stated in the form

$$|S_{11}\Delta^* - S_{22}^*|^2 > (1 + R_{LC})^2\left(|\Delta|^2 - |S_{22}|^2\right)^2$$

upon using (10.21a). From (10.21b) we obtain

$$1 + R_{LC} = \frac{|S_{22}|^2 - |\Delta|^2 + |S_{12}S_{21}|}{|S_{22}|^2 - |\Delta|^2}$$

and hence

$$|S_{11}\Delta^* - S_{22}^*|^2 > \left(|S_{22}|^2 - |\Delta|^2 + |S_{12}S_{21}|\right)^2$$

Upon using the expansion (10.22), we can restate the condition for stability as

$$|S_{12}S_{21}|^2 + \left(1 - |S_{11}|^2\right)\left(|S_{22}|^2 - |\Delta|^2\right)$$
$$> \left(|S_{22}|^2 - |\Delta|^2 + |S_{12}S_{21}|\right)^2$$

After multiplying out the term on the right, canceling the common factor $|S_{12}S_{21}|^2$ on both sides, and also canceling a common factor $|S_{22}|^2 - |\Delta|^2$, we get

$$\left(|S_{22}|^2 - |\Delta|^2\right) + \left(2|S_{12}S_{21}|\right) < 1 - |S_{11}|^2$$

We now introduce the stability parameter K and express the final result in the form

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} > 1 \qquad (10.23a)$$

$$|S_{11}| < 1 \qquad (10.23b)$$

Case 2. It may happen that the circle of Γ_L values giving $|\Gamma_{in}| = 1$ encloses the origin ($\Gamma_L = 0$, $\Gamma_{in} = S_{11}$). In this case all Γ_L values inside the circle are stable ones. For absolutely stable devices the circle must then be large enough to enclose the entire Smith chart $|\Gamma_L| \leq 1$, so that any passive load impedance Z_L can be used. This requires that the radius R_{LC} be greater than one. With reference to Fig. 10.12 it is seen that we now require $d > 1$ or $R_{LC} - |\Gamma_{LC}| > 1$ or equivalently $|\Gamma_{LC}| < R_{LC} - 1$. This case can occur only if $|\Gamma_{LC}| < R_{LC}$. By

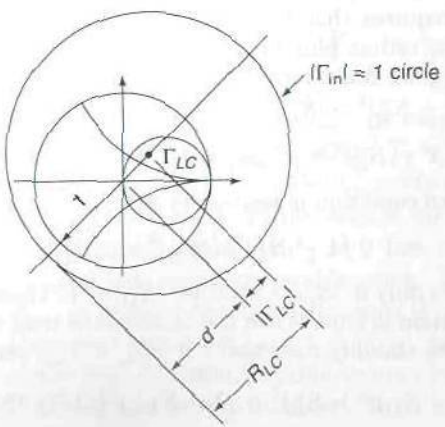


FIGURE 10.12
Illustration for deriving stability condition when the stability circle encloses the origin.

using (10.21) we find that we require $|S_{22}| < |\Delta|$. We now use (10.21*a*) and (10.21*b*) along with the expansion for $|S_{11}\Delta^* - S_{22}^*|^2$ given by (10.22) to state the stability criterion in the form

$$\left[\left|\left(|\Delta|^2 - |S_{22}|^2\right)\right| - |S_{12}S_{21}|\right]^2 > |S_{12}S_{21}|^2 + \left(1 - |S_{11}|^2\right)\left(|S_{22}|^2 - |\Delta|^2\right)$$

or

$$\left(|\Delta|^2 - |S_{22}|^2\right)^2 + |S_{12}S_{21}|^2 - 2|S_{12}S_{21}|\left(|\Delta|^2 - |S_{22}|^2\right) > |S_{12}S_{21}|^2 - \left(1 - |S_{11}|^2\right)\left(|\Delta|^2 - |S_{22}|^2\right)$$

When we cancel the term $|S_{12}S_{21}|^2$ on both sides and a common factor $|\Delta|^2 - |S_{22}|^2$ in the remaining terms, the final result is the same as obtained earlier for Case 1.

The requirement that R_{LC} be greater than one places a further restriction on the condition for absolute stability. From (10.21*b*) we readily see that, when $|\Delta| > |S_{22}|$ in order for $R_{LC} > 1$, we must have

$$|\Delta|^2 - |S_{22}|^2 < |S_{12}S_{21}|$$

From the expression for K we get

$$\frac{1 - |S_{11}|^2}{|S_{12}S_{21}|} + \frac{|\Delta|^2 - |S_{22}|^2}{|S_{12}S_{21}|} = 2K$$

Let $(|\Delta|^2 - |S_{22}|^2)/|S_{12}S_{21}| = 1 - \delta$, where δ is a positive quantity since this term is less than one. Thus we get

$$\frac{1 - |S_{11}|^2}{|S_{12}S_{21}|} = 2K - 1 + \delta > 1$$

since $K > 1$. Hence we require that

$$|S_{12}S_{21}| < 1 - |S_{11}|^2$$

in order that we can have $R_{LC} > 1$. The condition $K > 1$ is a necessary condition for absolute stability but may not always be sufficient since the condition that $R_{LC} > 1$ when $|\Delta| > |S_{22}|$ may be a more stringent condition as we will shortly see.

The source stability circle is the circle of source reflection-coefficient Γ_s values that make $|\Gamma_{out}| = 1$. By direct analogy with the derivation of (10.21), we find that the center for the source stability circle, in the Γ_s plane, is given by

$$\Gamma_{SC} = \frac{S_{22}\Delta^* - S_{11}^*}{|\Delta|^2 - |S_{11}|^2} \qquad (10.24a)$$

and its radius is given by

$$R_{SC} = \frac{|S_{12}S_{21}|}{\left|\left(|\Delta|^2 - |S_{11}|^2\right)\right|} \qquad (10.24b)$$

These equations are the same as (10.21) with S_{11} and S_{22} interchanged.

Since K is symmetrical in the variables S_{11} and S_{22}, it can correctly be inferred that the same condition $K > 1$ for absolute stability is obtained from the requirement $|\Gamma_{out}| < 1$, along with the two conditions $|S_{22}| < 1$ and $|S_{12}S_{21}| < 1 - |S_{22}|^2$.

The necessary and sufficient conditions for absolute stability are thus

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} > 1 \tag{10.25a}$$

$$|S_{11}| < 1 \tag{10.25b}$$

$$|S_{22}| < 1 \tag{10.25c}$$

$$|S_{12}S_{21}| < 1 - |S_{11}|^2 \tag{10.25d}$$

$$|S_{12}S_{21}| < 1 - |S_{22}|^2 \tag{10.25e}$$

By adding (10.25d) and (10.25e) and using (10.25a), we get

$$2|S_{12}S_{21}| < 1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2 + (1 - |\Delta|^2)$$

$$< 2|S_{12}S_{21}|K + (1 - |\Delta|^2)$$

Thus

$$K > 1 + \frac{|\Delta|^2 - 1}{2|S_{12}S_{21}|} \tag{10.25f}$$

When $|\Delta| < 1$ then clearly this condition holds whenever $K > 1$. However, when $|\Delta| > 1$ the condition (10.25*f*) is more stringent than (10.25*a*). For most devices $|\Delta| < 1$, so that (10.25*a*) to (10.25*c*) are sufficient to guarantee absolute stability.

Conditionally Stable Devices

For GaAs MESFETs in the common source connection and bipolar transistors in the common emitter connection, one generally finds that $|S_{11}|$ and $|S_{22}|$ are both less than unity. For bipolar transistors in the common base connection, $|S_{11}|$ and $|S_{22}|$ are usually greater than one. Likewise, for MESFETs in the common gate connection, one often finds that $|S_{11}|$ and $|S_{22}|$ will be greater than one. The same is true for the common drain connection.

For an unstable device there are four possible load stability circle configurations and a similar number for the source stability circles. These four cases are described below, along with the necessary conditions for them to occur.

Case 1. The load stability circle may lie entirely outside the Smith chart and $|S_{11}| > 1$. In this case all values of Γ_L on the Smith chart are unstable values since the origin is an unstable point. The device in this case is absolutely

unstable and would be of no interest or use for an amplifier. The necessary conditions for this case to occur are

$$K < -1$$
$$|S_{22}| < |\Delta|$$
$$|S_{11}|^2 > 1$$

Case 2. The load stability circle may enclose the Smith chart, and when $|S_{11}| > 1$ all values of Γ_L on the Smith chart will be unstable values. Again, the device will be an absolutely unstable one. The necessary conditions for this case to occur are

$$K < -1$$
$$|S_{22}| > |\Delta|$$
$$|S_{11}|^2 > 1 - (2K + 1)|S_{12}S_{21}|$$

Case 3. The load stability circle may lie entirely inside the Smith chart. In this case the device is conditionally stable since a region of the Smith chart will represent stable values of Γ_L. In order for the stability circle to lie inside the Smith chart, we require that $|\Gamma_{LC}| + R_{LC} < 1$ and $R_{LC} < 1$. When $|S_{22}| < |\Delta|$ these two conditions lead to the following necessary conditions for this case to occur:

$$K > 1$$
$$|S_{22}| < |\Delta|$$
$$|S_{11}|^2 > 1 - (2K - 1)|S_{12}S_{21}|$$

The last condition is equivalent to $R_{LC} < 1$. When $|S_{22}| > |\Delta|$ the corresponding necessary conditions for the load stability circle to lie entirely inside the Smith chart are

$$K < -1$$
$$|S_{22}| > |\Delta|$$
$$|S_{11}|^2 < 1 - (2K + 1)|S_{12}S_{21}|$$

These conditions are obtained by requiring that $|\Gamma_{LC}|^2 < (1 - R_{LC})^2$ and using (10.21) and (10.22).

When $K > 1$, $|S_{22}| < |\Delta|$, and $|S_{11}|^2 < 1 - (2K - 1)|S_{12}S_{21}|$, the device *is absolutely stable. Devices for which* $K > 1$ *are almost always stable, and* since devices with $K < -1$ are not likely to occur, Case 3 is not likely to be encountered in practice.

Case 4. The usual situation that occurs in practice when a device is only conditionally stable is the one for which the load stability circle intersects the boundary of the Smith chart. This case is the one on which we will focus our attention. The load stability circle $|\Gamma_{in}| = 1$ then maps into a circle of $|\Gamma_L|$ values that intersects the Smith chart boundary $|\Gamma_L| = 1$ at two points. The load stability circle may or may not enclose the origin. In order to establish

whether values of Γ_L inside or outside the $|\Gamma_{in}| = 1$ circle are stable values, the following rules apply:

(*a*)	$\lvert\Gamma_{in}\rvert = 1$	circle encloses the origin
	$\lvert S_{11}\rvert < 1$	all interior values are stable values
	$\lvert S_{11}\rvert > 1$	all exterior values are stable values
(*b*)	$\lvert\Gamma_{in}\rvert = 1$	circle does not enclose the origin
	$\lvert S_{11}\rvert < 1$	all exterior values are stable values
	$\lvert S_{11}\rvert > 1$	all interior values are stable values

These rules are established by simply noting that the origin $\Gamma_L = 0$, which gives $\Gamma_{in} = S_{11}$ is a stable point when $|S_{11}| < 1$ and is an unstable point when $|S_{11}| > 1$.

A similar set of rules applies to the source stability circle which is a plot of all values of the source reflection coefficient Γ_s that gives $|\Gamma_{out}| = 1$. For example, when $|S_{22}| < 1$ the origin $\Gamma_s = 0$ giving $\Gamma_{out} = S_{22}$ is a stable point. Thus, if the circle $|\Gamma_{out}| = 1$ encloses the origin, all values of Γ_s inside the circle are stable values. They are unstable values when $|S_{22}| > 1$.

By plotting the load stability circle whose center and radius is given by (10.21), it is easy to determine what values of Γ_L, and hence load impedances Z_L, can be used to ensure that $|\Gamma_{in}| < 1$. A similar plot of the source stability circle $|\Gamma_{out}| = 1$ shows what values of source reflection coefficient Γ_s can be used and will ensure a stable circuit.

It is useful to have an analytical test of when the load or source stability circle will enclose the origin. Such a test is easily established by examining the ratio $|\Gamma_{LC}|^2/R_{LC}^2$. From (10.21) and using the expansion (10.22), we obtain

$$\frac{|\Gamma_{LC}|^2}{R_{LC}^2} = 1 + \frac{\left(1 - |S_{11}|^2\right)\left(|S_{22}|^2 - |\Delta|^2\right)}{|S_{12}S_{21}|^2}$$

In order for the origin to be enclosed, we must have $R_{LC} > |\Gamma_{LC}|$ and this requires that the last term on the right-hand side of the above equation be negative, since the ratio must be less than one. Hence we can state the following rule:

> The load stability circle encloses the origin when $|S_{11}| < 1$ and $|\Delta| > |S_{22}|$ or when $|S_{11}| > 1$ and $|\Delta| < |S_{22}|$.

When these conditions do not hold, the origin is not enclosed. A similar test applies to the source stability circle. The origin is enclosed only when $|S_{22}| < 1$ and $|\Delta| > |S_{11}|$ or when $|S_{22}| > 1$ and $|\Delta| < |S_{11}|$.

After the derivation of the equations for the load and source reflection coefficients needed for conjugate impedance matching at both ports, i.e., (10.17), we promised to show that for an absolutely stable device A_1 and A_2

are positive. The expression for A_2 is

$$A_2 = \left(1 - |S_{11}|^2\right) - \left(|\Delta|^2 - |S_{22}|^2\right)$$

Since $|S_{11}|^2 < 1$ for an absolutely stable device, the first term is positive. For $|S_{22}| > |\Delta|$ the second term is also positive and A_2 is positive. When $|S_{22}| < |\Delta|$ then we must have $|\Delta|^2 - |S_{22}|^2 < |S_{12}S_{21}|$ in order to get $R_{LC} > 1$. This condition comes directly from (10.21*b*) which gives R_{LC}. In order for A_2 to be positive, we require

$$1 - |S_{11}|^2 > |\Delta|^2 - |S_{22}|^2$$

But our stability requirements specify that $1 - |S_{11}|^2 > |S_{12}S_{21}| > |\Delta|^2 - |S_{22}|^2$ and hence A_2 is positive. A similar proof shows that A_1 is positive.

Example 10.1 Stability circles. A MESFET has the following scattering-matrix parameters at 5 GHz:

$$S_{11} = 0.75\angle -120^\circ \qquad S_{12} = 0.08\angle 50^\circ$$
$$S_{21} = 3.9\angle 90^\circ \qquad S_{22} = 0.4\angle -25^\circ$$

and at 10 GHz:

$$S_{11} = 0.72\angle 170^\circ \qquad S_{12} = 0.1\angle 40^\circ$$
$$S_{21} = 2.3\angle 45^\circ \qquad S_{22} = 0.1\angle -55^\circ$$

We wish to find the stability parameter K, the maximum stable gain, and the load and source stability circles. At 5 GHz we find from (10.23*a*) that $K = 0.64234$; so the transistor is not absolutely stable at 5 GHz. By using (10.21) we obtain the load stability circle parameters which are

$$\Gamma_{LC} = 3.7678 + j8.3712 \qquad R_{LC} = 8.5057$$

From (10.24) we obtain the source stability circle parameters which are

$$\Gamma_{SC} = -0.9971 + j1.19296 \qquad R_{SC} = 0.7104$$

These circles are shown plotted in Fig. 10.13. Neither circle encloses the origin, and since $|S_{11}| < 1$ and $|S_{22}| < 1$, all values of Γ_L and Γ_s outside their respective stability circles are stable values and may be used in an amplifier design.

At 10 GHz we find that for the given parameters $K = 1.0891$; so at this frequency the transistor is absolutely stable. The stability circle parameters are

$$\Gamma_{LC} = -2.1677 - j10.5364 \qquad R_{LC} = 11.8549$$
$$\Gamma_{SC} = -1.47156 - j0.28336 \qquad R_{SC} = 0.47035$$

For the load stability circle $|\Gamma_{LC}| = 10.757$, and since $R_{LC} > |\Gamma_{LC}| + 1$, the circle encloses the Smith chart. Since $|S_{11}| < 1$ all values of Γ_L inside this circle are stable values. The circle is a very large one, so that only a small portion of it is shown in Fig. 10.13 (dashed line). The source stability circle is also shown in Fig. 10.13. This circle lies entirely outside the Smith chart, and since $|S_{22}| < 1$ all values of Γ_s outside this circle are stable values.

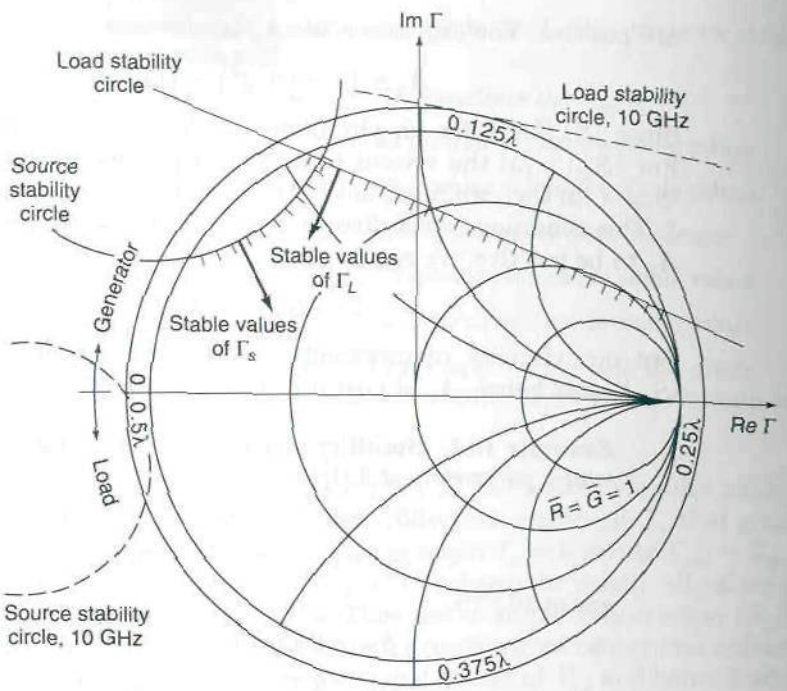


FIGURE 10.13
Load and source stability circles for a MESFET at 5 GHz (solid circles) and 10 GHz (dashed circles). The device is conditionally stable at 5 GHz and absolutely stable at 10 GHz.

10.7 CONSTANT POWER-GAIN CIRCLES

It is convenient to introduce the normalized power gain g_p given by

$$g_p = \frac{G_p}{|S_{21}|^2} \tag{10.26}$$

From (10.15) we obtain

$$g_p = \frac{1 - \Gamma_L \Gamma_L^*}{1 - |S_{11}|^2 + \Gamma_L \Gamma_L^* \left(|S_{22}|^2 - |\Delta|^2\right) - \Gamma_L (S_{22} - \Delta S_{11}^*) - \Gamma_L^* (S_{22}^* - \Delta^* S_{11})} \tag{10.27a}$$

upon writing $|\Gamma_L|^2 = \Gamma_L \Gamma_L^*$ and using the relationship $2 \operatorname{Re} Z = Z + Z^*$. We can rewrite the expression for normalized power gain in the form

$$\Gamma_L \Gamma_L^* - \left[\frac{g_p (S_{22} - \Delta S_{11}^*) \Gamma_L + g_p (S_{22}^* - \Delta^* S_{11}) \Gamma_L^* - \left(1 - |S_{11}|^2\right) g_p + 1}{\left(|S_{22}|^2 - |\Delta|^2\right) g_p + 1} \right] = 0 \tag{10.27b}$$

by multiplying g_p by the denominator on the right-hand side and bringing the term $1 - |\Gamma_L|^2$ over to the left-hand side and then dividing by the coefficient of $\Gamma_L\Gamma_L^*$. When we compare the above expression with (10.1), we see that it describes a circle in the Γ_L plane. The center of the circle is at $\Gamma_L = \Gamma_{Lg}$, where Γ_{Lg} is the coefficient multiplying Γ_L^*. The radius R_{Lg} of the circle is obtained by noting that the constant term is equal to $|\Gamma_{Lg}|^2 - R_{Lg}^2$. Thus we readily find that the constant normalized power-gain circles have a center at

$$\Gamma_{Lg} = \frac{(S_{22}^* - \Delta^* S_{11})g_p}{(|S_{22}|^2 - |\Delta|^2)g_p + 1} \tag{10.28a}$$

and a radius given by

$$R_{Lg} = \frac{\left(1 - 2Kg_p|S_{12}S_{21}| + g_p^2|S_{12}S_{21}|^2\right)^{1/2}}{\left|(|S_{22}|^2 - |\Delta|^2)g_p + 1\right|} \tag{10.28b}$$

where we have used the expansion (10.22) to replace $|S_{22}^* - \Delta^* S_{11}|^2$ and expressed $1 - |S_{11}|^2$ as $|S_{22}|^2 - |\Delta|^2 + 2K|S_{12}S_{21}|$ in order to simplify the expression for R_{Lg}. When g_p approaches infinity the equations above become those for the load stability circle. Thus, on the load stability circle, the power gain is infinite. When $g_p = 0$ we obtain $\Gamma_{Lg} = 0$, $R_{Lg} = 1$, which is the boundary of the Smith chart. On the boundary of the Smith chart, $|\Gamma_L| = 1$, which corresponds to a pure reactive load impedance and consequently no power is delivered to the load and $g_p = 0$, as the above equations verify.

It is common practice to plot the normalized constant power-gain circles that correspond to gains 1 dB, 2 dB, 3 dB, etc., less than the maximum normalized power gain $(K - \sqrt{K^2 - 1})/|S_{12}S_{21}|$ [see (10.19)] for an absolutely stable device. For a potentially unstable device, constant power-gain circles for normalized gains 1 dB, 2 dB, etc., less than the normalized "Figure of Merit" gain $(|S_{21}|/|S_{12}|)/|S_{21}|^2 = 1/|S_{12}S_{21}|$ are usually plotted.

The equation for the center of a constant power-gain circle lies on the same ray from the origin as the center of the load stability circle does, since both Γ_{LC} and Γ_{Lg} have the same complex numerator except for the factor g_p in Γ_{Lg}.

By plotting the constant power-gain circles and the load stability circle, we can easily determine those values of load reflection coefficients Γ_L that will give the largest power gain and yet result in a stable amplifier design. Those values of Γ_L in a stable region and lying on a given g_p = constant circle will give a power gain $G_p = |S_{21}|^2 g_p$. Later on we will find that there will be additional constraints imposed on Γ_L because of low-noise and low input and output VSWR requirements.

Properties of the Constant Gain Circles

By examining the geometric properties of the constant gain circles, we will obtain considerable insight into the operational characteristics of a microwave amplifier. The characteristics of the constant power-gain circles are quite different for absolutely stable devices and potentially unstable devices; so we examine each case separately.

Stable Devices

For a stable device one set of constant gain circles lies entirely inside the Smith chart. The $g_p = 0$ circle coincides with the boundary of the Smith chart. As g_p increases from zero, the radius R_{Lg} decreases and becomes zero when

$$g_p = \frac{K \pm \sqrt{K^2 - 1}}{|S_{12}S_{21}|}$$

For circles inside the Smith chart, $R_{Lg} = 0$ when

$$g_p = \frac{K - \sqrt{K^2 - 1}}{|S_{12}S_{21}|}$$

When $K = 1$, $R_{Lg} = 0$ when $g_p = 1/|S_{12}S_{21}|$, which gives the maximum gain. A second set of circles exists outside the Smith chart and gives values of g_p for values of $|\Gamma_L| > 1$, which implies the use of an active load. For passive loads these circles are not relevant. In the region

$$g_p = \frac{K - \sqrt{K^2 - 1}}{|S_{12}S_{21}|} \quad \text{to} \quad \frac{K + \sqrt{K^2 - 1}}{|S_{12}S_{21}|}$$

there are no real solution for R_{Lg}, when $g_p > 0$ as illustrated in Fig. 10.14. *Typical constant power-gain circles for a stable device* are shown in Fig. 10.15 for the case when the load stability circle lies outside the Smith chart.

Inside the Smith chart boundary the maximum normalized power gain

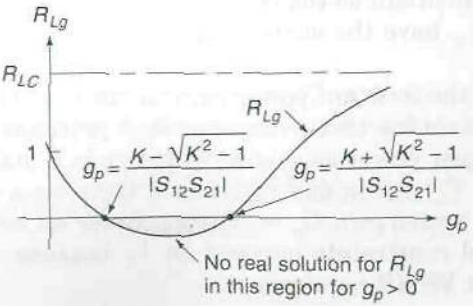


FIGURE 10.14
Illustration of region where there is a solution for the radius of the constant power-gain circles.

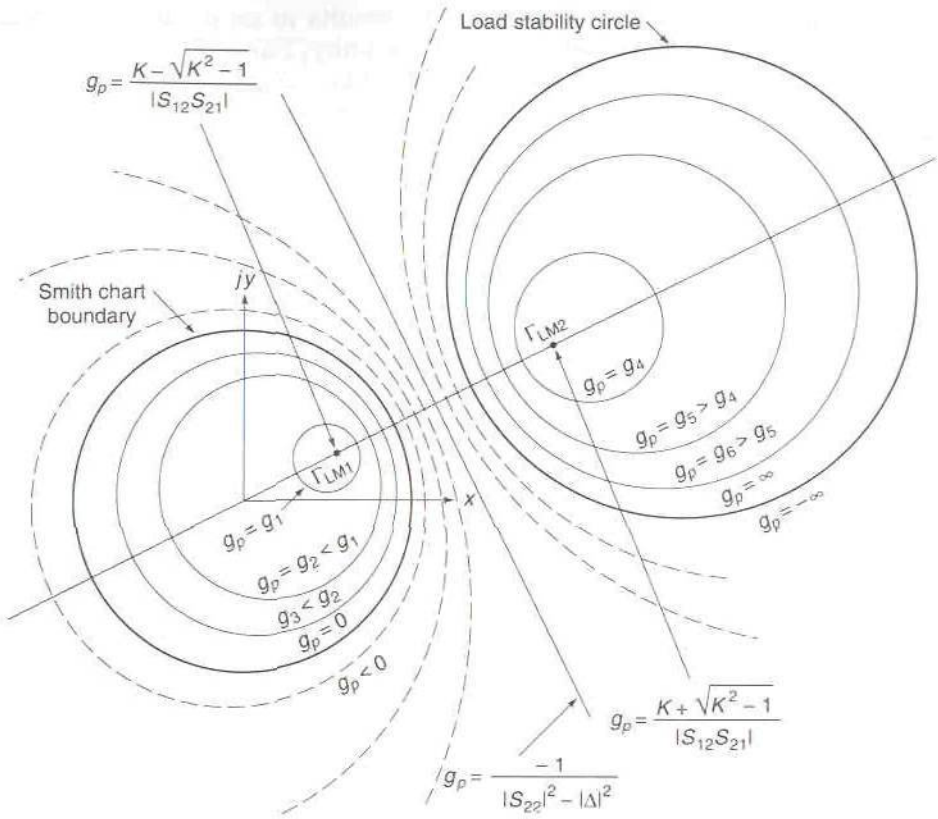


FIGURE 10.15
Circles of constant normalized power gain for an absolutely stable transistor when the load stability circle lies outside the Smith chart. The circles of constant negative power gain are shown by the dashed circles.

$(K - \sqrt{K^2 - 1})/|S_{12}S_{21}|$ is obtained at the single point where $\Gamma_L = \Gamma_{LM1}$ as given by (10.17*b*). This is the gain for the conjugate-impedance-matching condition. Away from this point the constant gain circles have increasing radii and decreasing gain values. The limiting circle coincides with the Smith chart boundary on which $g_p = 0$.

A second set of similar constant power-gain circles exists inside the load stability circle. In this region the minimum normalized gain is $(K + \sqrt{K^2 - 1})/|S_{12}S_{21}|$ and occurs for $\Gamma_L = \Gamma_{LM2}$, where now Γ_{LM2} is the other solution given by (10.17*b*) for conjugate impedance matching, i.e., using the positive sign in front of the square-root term. As we move away from this point, the constant gain circles have increasing radii and increasing gain values. The limiting circle is the load stability circle on which the gain is infinite. Inside the load stability circle, which is the unstable region

of Γ_L values, each Γ_L results in an input reflection coefficient Γ_{in} with a magnitude greater than unity. Since $|\Gamma_L| > 1$ also in this region, it implies that both Z_{in} and Z_L have negative resistive parts. This type of load impedance could arise by connecting an unstable microwave amplifier to the output of the amplifier under discussion. When $|\Gamma_{in}| > 1$ and $|\Gamma_L| > 1$, more power is reflected from the amplifier input and load termination than is incident on either one. The input power and load power are thus both negative but the ratio is positive, so that the power gain is positive. As the boundary of the load stability circle is approached, which is the circle of Γ_L values that give $|\Gamma_{in}| = 1$, the gain becomes infinite since, when $|\Gamma_{in}| = 1$, there is zero input power to the amplifier.

In the region between the Smith chart and the load stability circle, $|\Gamma_L| > 1$, but since this is a stable region each Γ_L produces a Γ_{in} with $|\Gamma_{in}| < 1$. We now have a situation where there is a finite positive input power to the amplifier but a negative load power since, with $|\Gamma_L| > 1$, the load termination reflects more power than is incident upon it. Thus the power gain is negative. Indeed, if we assume a negative power gain, we will get another system of constant negative power-gain circles that fills the region between the Smith chart boundary and the load stability circle. As the load stability circle is approached from this region, the gain approaches minus infinity. These negative power-gain circles are shown as dashed circles in Fig. 10.15. The circle with infinite radius occurs when

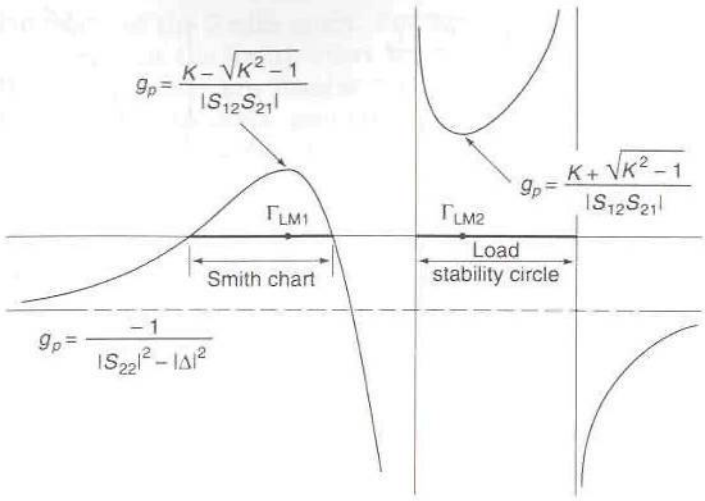
$$g_p = \frac{1}{|\Delta|^2 - |S_{22}|^2} \tag{10.29}$$

i.e., when the denominator in the expression (10.28*b*) for R_{L_g} vanishes. On this straight line the gain is constant and is given by (10.29). In order to obtain finite values for Γ_L from (10.27*b*) when g_p is given by (10.29), the numerator of the second term in (10.27*b*) must vanish. If we let $\Gamma_L = x + jy$, we find that setting the numerator equal to zero gives the following equation for the straight line which is the constant power-gain circle having an infinite radius

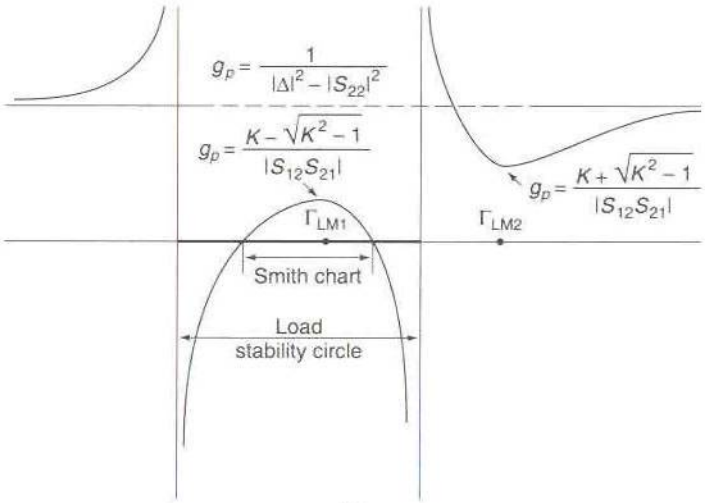
$$y = x\frac{\mathrm{Re}(S_{22} - \Delta S_{11}^*)}{\mathrm{Im}(S_{22} - \Delta S_{11}^*)} + |\Delta|^2 - |S_{22}|^2 + |S_{11}|^2 - 1 \tag{10.30}$$

In Fig. 10.16*a* we show the gain profile along the line connecting the origin of the Smith chart to the center of the load stability circle. Note the regions in which the power gain is negative.

When $|\Delta| > |S_{22}|$ the load stability circle encloses the Smith chart. For this case the constant power-gain circles have the configuration shown in Fig. 10.17. For this case the circles again begin at the point $\Gamma_L = \Gamma_{LM1}$ inside the Smith chart and increase in radii and decrease in gain value. One circle is $|\Gamma_L| = 1$ on which $g_p = 0$. Beyond the Smith chart boundary but inside the load stability circle, $|\Gamma_L| > 1$, $|\Gamma_{in}| < 1$, and the power gain in-



(*a*)



(*b*)

FIGURE 10.16
Gain profile along the line joining the Smith chart center and the load stability circle center. (*a*) Gain profile when the load stability circle lies outside the Smith chart; (*b*) gain profile when the load stability circle encloses the Smith chart.

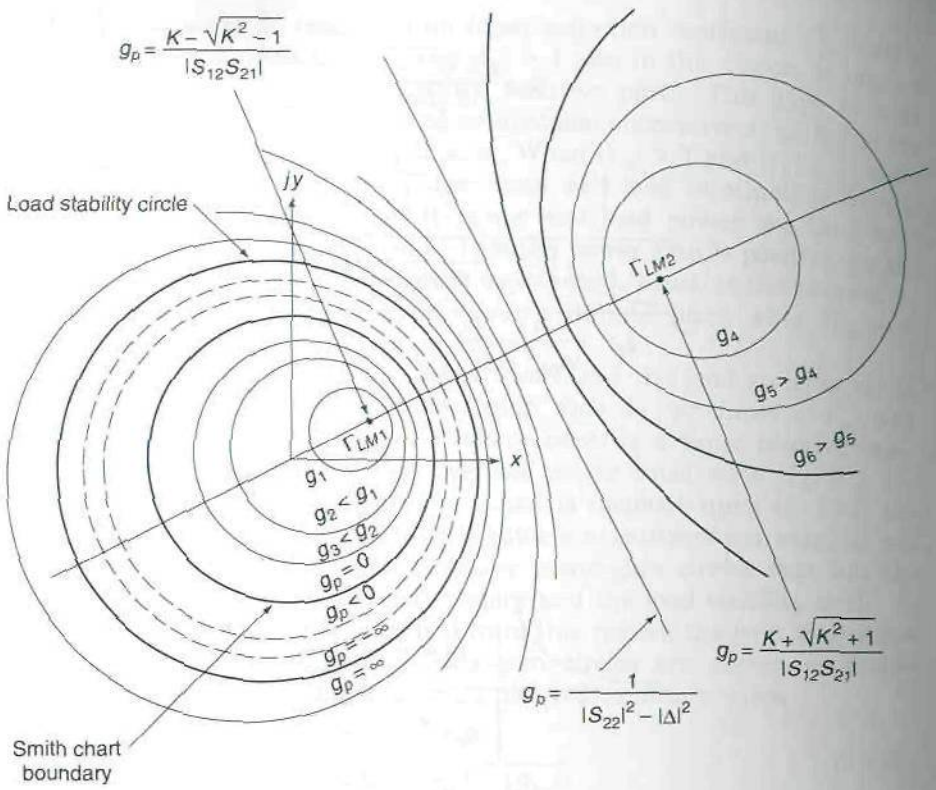


FIGURE 10.17
Circles of constant normalized power gain for an absolutely stable transistor when the load stability circle encloses the Smith chart. The dashed circles are circles of constant negative power gain.

creases in a negative direction and approaches minus infinity on the load stability circle. Beyond this circle the radii keep increasing until the straight line given by (10.30) is reached. On this line g_p is given by (10.29) but is now positive since $|\Delta| > |S_{22}|$. The radii then begin to decrease until the limit point at $\Gamma_L = \Gamma_{LM2}$, the second solution given by (10.17*b*), is reached and at which the normalized gain is $(K + \sqrt{K^2 - 1})/|S_{12}S_{21}|$. A gain profile along the ray from the origin to the center of the load stability circle for this case is shown in Fig. 10.16*b*. When $|\Gamma_L|$ approaches infinity the gain approaches the value given by (10.29).

Unstable Devices

For an unstable device the load stability circle will usually intersect the Smith chart and may enclose the origin. For unstable devices with $|K| < 1$, no value of g_p that is real will make $R_{L_g} = 0$. Again the $g_p = 0$ circle is the

boundary of the Smith chart. For most unstable devices the load stability circle will cut the Smith chart boundary at two points. We can show that these two points are invariant points that all constant gain circles pass through. The constant gain circle on which the gain is given by (10.29) is the circle with infinite radius. This circle is the straight line given by (10.30).

We can prove that all constant gain circles intersect the Smith chart boundary at the same two points that the load stability circle does as follows: The load stability circle is given by

$$|\Gamma_L - \Gamma_{LC}|^2 = R_{LC}^2 = |\Gamma_L|^2 + |\Gamma_{LC}|^2 - \Gamma_L \Gamma_{LC}^* - \Gamma_L^* \Gamma_{LC}$$

We now let $\Gamma_L = x + jy$. The stability circle intersects the Smith chart when $|\Gamma_L| = 1$. By setting $|\Gamma_L|^2 = 1$ we then find that the above equation is that of a straight line given by

$$y = -\frac{\text{Re}\,\Gamma_{LC}}{\text{Im}\,\Gamma_{LC}} x + \frac{|\Gamma_{LC}|^2 - R_{LC}^2 + 1}{2\,\text{Im}\,\Gamma_{LC}}$$

This line intersects the circle $|\Gamma_L| = 1$ at two points. In a similar derivation we find that the constant gain circles will intersect the circle $|\Gamma_L| = 1$ at the two points where the line

$$y = -\frac{\text{Re}\,\Gamma_{Lg}}{\text{Im}\,\Gamma_{Lg}} x + \frac{|\Gamma_{Lg}|^2 - R_{Lg}^2 + 1}{2\,\text{Im}\,\Gamma_{Lg}}$$

intersects the $|\Gamma_L| = 1$ circle. By substituting for Γ_{LC}, R_{LC} from (10.21) into the first equation and for Γ_{Lg}, R_{Lg} from (10.28) into the second equation, we find that both lines have the same slope and intercept. We omit the detailed algebra involved but it is straightforward. As a consequence of the above property, the constant normalized power-gain circles for an unstable device have the configuration shown in Fig. 10.18. When Γ_L approaches either one of the two invariant points, both the numerator and denominator in the expression (10.27*a*) for the power gain approach zero. The ratio is thus indeterminate and must be evaluated as a limit with the result that the limit depends on the direction along which the invariant points are approached. Thus we can have many constant gain circles with different gains at the two invariant points, as is clearly evident from an examination of Fig. 10.18.

We will now examine the constant gain circles for the four special cases that can occur in practice. The usual situation is the one where $|S_{11}| < 1$. When this is true the origin, that is, $\Gamma_L = 0$, is a stable point. The load stability circle may or may not enclose the origin. When the origin is not enclosed, the constant gain circles within the Smith chart but outside the load stability circle are in the stable region and Γ_L may be chosen to lie on one of these gain circles as shown in Fig. 10.19*a*. In the region common to both the Smith chart and the load stability circle, the power gain is

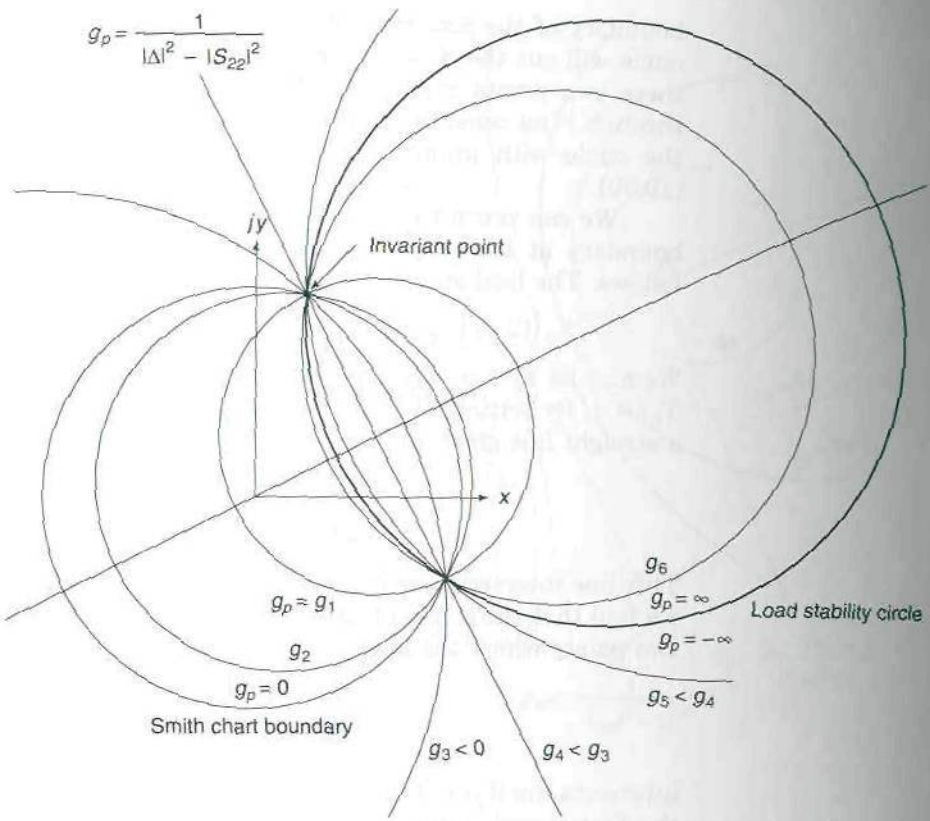


FIGURE 10.18
Constant normalized power-gain circles for an unstable device.

negative. In the region outside both circles, the power gain is also negative; in the region inside the stability circle but outside the Smith chart, the gain is again positive. The gain always changes sign when the circle $|\Gamma_L| = 1$ on which $g_p = 0$ is crossed and also when the load stability circle on which the gain is infinite is crossed. On one side of the load stability circle boundary, $g_p = \infty$, and on the other side of the load stability circle boundary, $g_p = -\infty$.

For this case the centers of the constant gain circles move from zero to Γ_{LC} as g_p increases from zero to infinity, as shown in Fig. 10.19*a*. For other locations of the centers on the ray joining the origin to the stability circle center, the constant gain circles have negative gain values.

When the stability circle encloses the origin, then the stable region is the region inside the load stability circle as shown in Fig. 10.19*b*. For a stable amplifier design Γ_L must be chosen to lie on a constant gain circle in this region. For this case the centers of the positive constant gain circles move from zero to infinity in the direction away from Γ_{LC} and return from

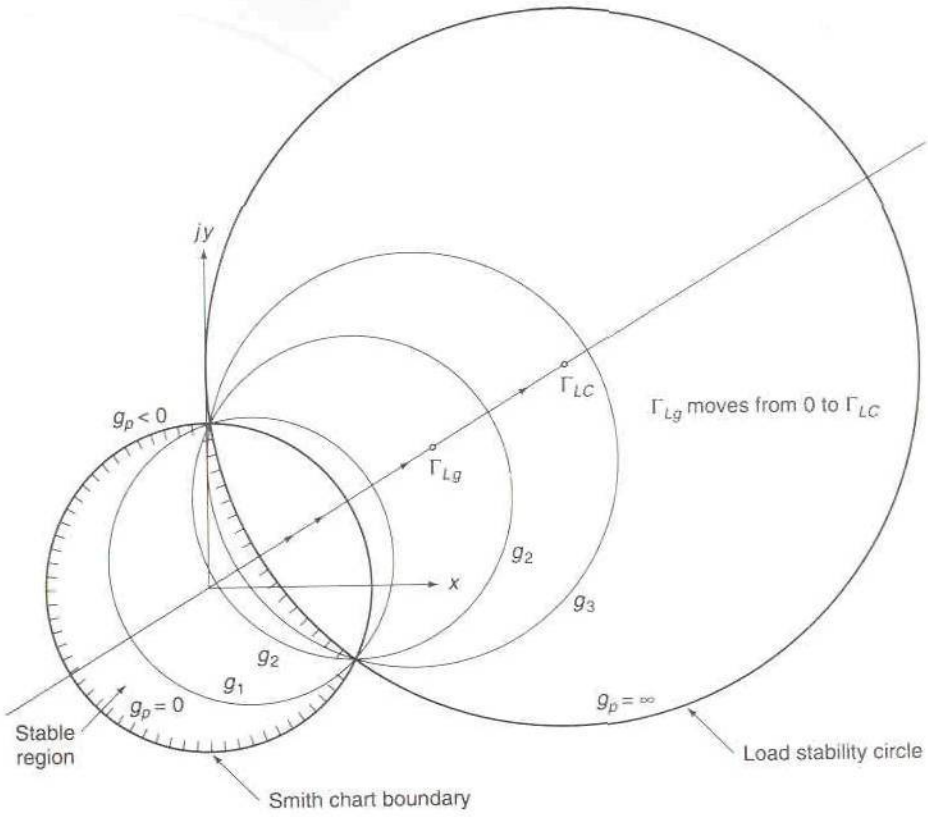


FIGURE 10.19
Constant normalized power-gain circles when $|S_{11}| < 1$ so that the origin is a stable point. (*a*) Constant gain circles and stable region when the origin is not enclosed; (*b*) constant gain circles and stable region when the origin is enclosed.

infinity toward Γ_{LC} as the gain increases from zero to infinity, as shown in Fig. 10.19*b*.

The other two cases occur when $|S_{11}| > 1$ for which the origin is an unstable point. The constant gain circles for these two cases are shown in Fig. 10.20. The stable and unstable regions are interchanged from those shown in Fig. 10.19. The values of Γ_L must be chosen to lie on a constant gain circle in that part of the Smith chart in which the origin is not included. Hence, when the stability circle encloses the origin, Γ_L must be outside the stability circle for a stable design. When the stability circle does not enclose the origin, then Γ_L should be chosen as a point inside the stability circle.

For an absolutely unstable device, $K < -1$, and there are two values of g_p that are negative and correspond to circles of constant gain that have

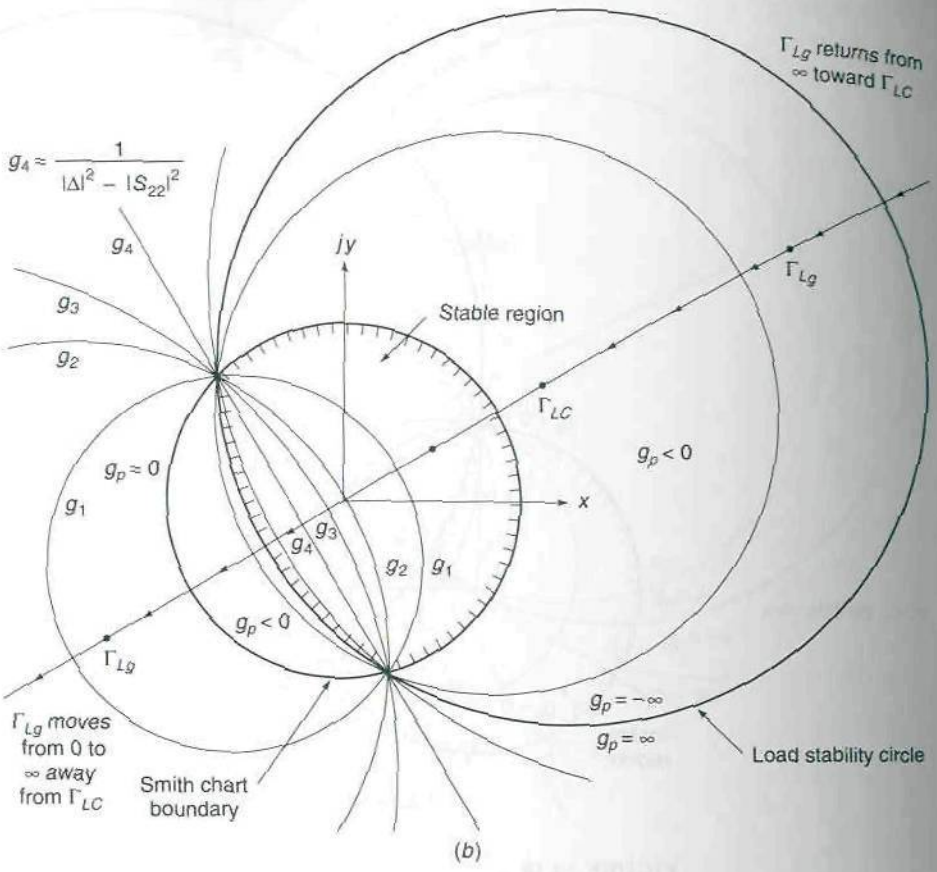


FIGURE 10.19 *Continued.*

zero radii. These gain values are given by $g_p = -(|K| \pm \sqrt{K^2 - 1})/|S_{12}S_{21}|$. The constant gain circles are similar to those shown in Figs. 10.15 and 10.17 but g_p is negative inside the Smith chart. For the case when the load stability circle lies inside the Smith chart, K may be greater than 1 or less than -1. In both cases there are two points corresponding to constant gain circles with zero radii. The constant gain circles are again similar to those shown in Figs. 10.15 and 10.17. On the load stability circle g_p equals plus or minus infinity and on the Smith chart boundary $g_p = 0$. If the interior of the load stability circle is a stable region, g_p will be positive in this region. Between the load stability circle and the boundary of the Smith chart, g_p will be negative, and outside the Smith chart it will be positive. If the interior of the load stability circle is an unstable region, the sign of g_p will be opposite to that described above.

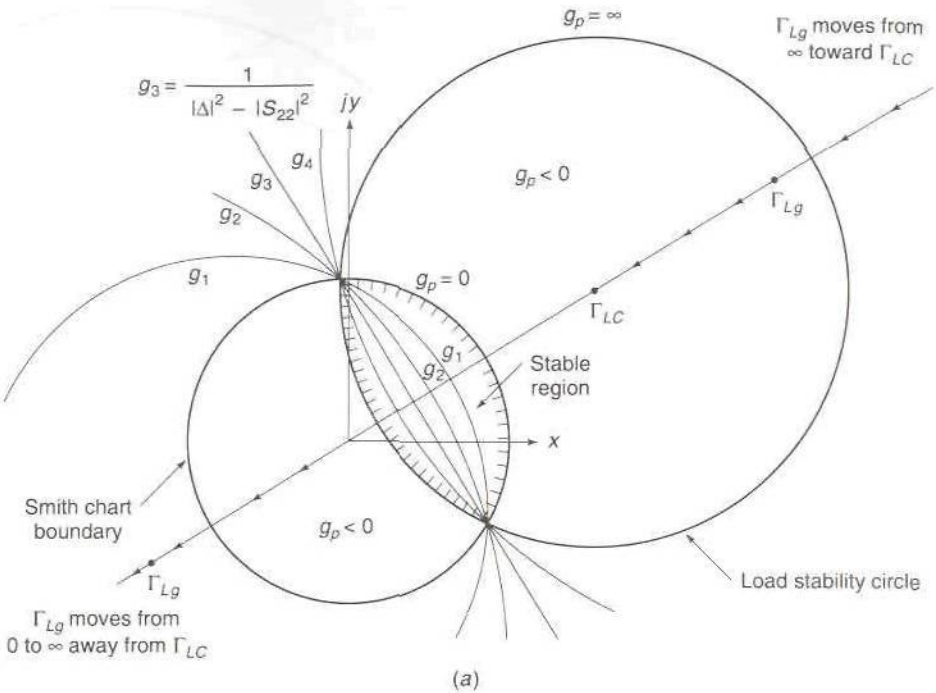


FIGURE 10.20
Constant normalized power-gain circles when $|S_{11}| > 1$ so that the origin is an unstable point. (*a*) Constant gain circles and stable region when the origin is not enclosed; (*b*) constant gain circles and stable region when the origin is enclosed.

Example 10.2 Preliminary amplifier design. A given microwave transistor has the following scattering-matrix parameters: $S_{11} = 0.9\angle 60°$, $S_{12} = 0.06\angle 60°$, $S_{21} = 3\angle 120°$, $S_{22} = 0.82\angle -30°$. We will use this device in the design of a single-stage amplifier having 50-Ω input and output transmission lines.

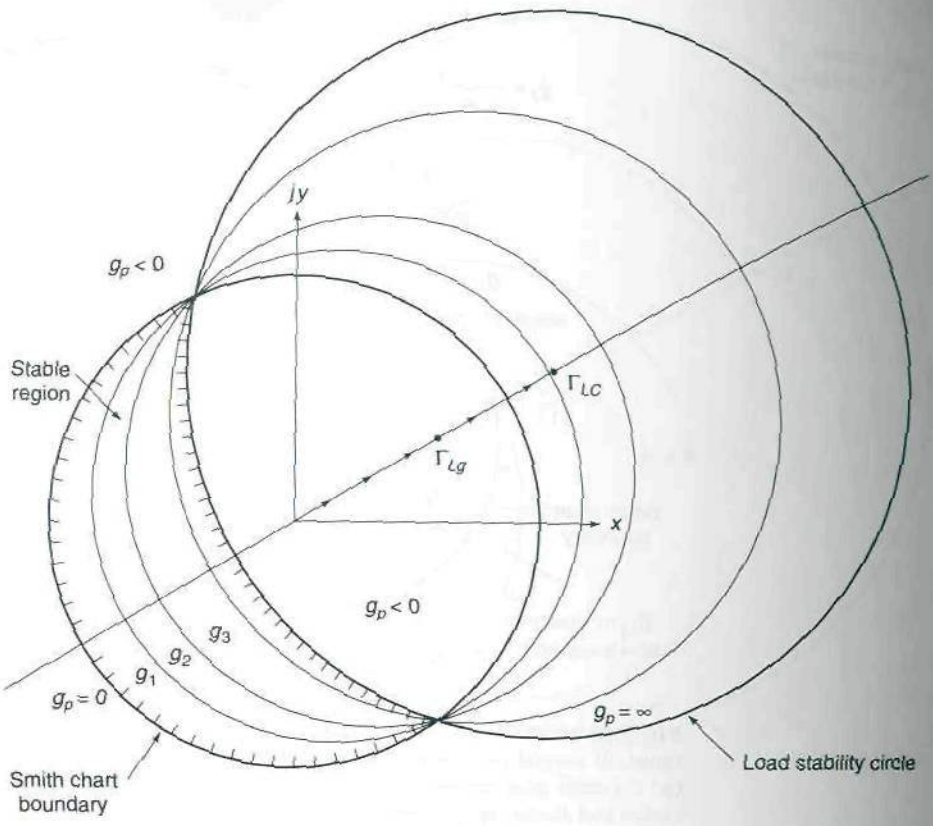
For the given transistor we find that

$$\text{Stability parameter } K = \frac{1 - 0.9^2 - 0.82^2 + |\Delta|^2}{2 \times 3 \times 0.06} = 0.902026$$

where $|\Delta| = |S_{11}S_{22} - S_{12}S_{21}| = 0.898404$. The transistor is potentially unstable since $K < 1$. The "Figure of Merit" gain is $|S_{21}/S_{12}| = 50$ or 17 dB. The normalized value is $50/|S_{21}|^2 = 5.555$. The parameters for the stability circles are obtained using (10.21) and (10.24) and are

$$\Gamma_{LC} = -0.40026 + j0.46312 \qquad R_{LC} = 1.336$$
$$\Gamma_{SC} = 6.8262 - j63.243 \qquad R_{SC} = 62.7068$$

The centers and radii of the constant normalized power-gain circles for $g_p = 5$,



(b)

FIGURE 10.20 *Continued.*

4, and 3 are

$$g_p = 5 \quad \Gamma_{Lg} = 0.8262 - j0.95596 \quad R_{Lg} = 1.3227$$
$$g_p = 4 \quad \Gamma_{Lg} = 0.4678 - j0.5413 \quad R_{Lg} = 1.01606$$
$$g_p = 3 \quad \Gamma_{Lg} = 0.2715 - j0.31417 \quad R_{Lg} = 0.9456$$

The stability circles and constant gain circles are shown in Fig. 10.21. The load stability circle encloses the origin. The source stability circle intersects only a small portion of the Smith chart, so that almost all values of source reflection coefficients Γ_s will be stable values. Since $|S_{11}| < 1$ the origin represents a stable point, so that all values of Γ_L inside the load stability circle are stable values. The source stability circle appears as a straight line in Fig. 10.21 since its radius is very large, that is, $R_{SC} = 62.7$.

From the figure we see that $\Gamma_L = 0$ is a stable point and would give a normalized gain somewhat greater than 5. From (10.27*a*) we see that the normalized gain, when $\Gamma_L = 0$, is $1/(1 - |S_{11}|^2)$ which equals 5.26. The advantage gained by using $\Gamma_L = 0$ is that no output matching network is

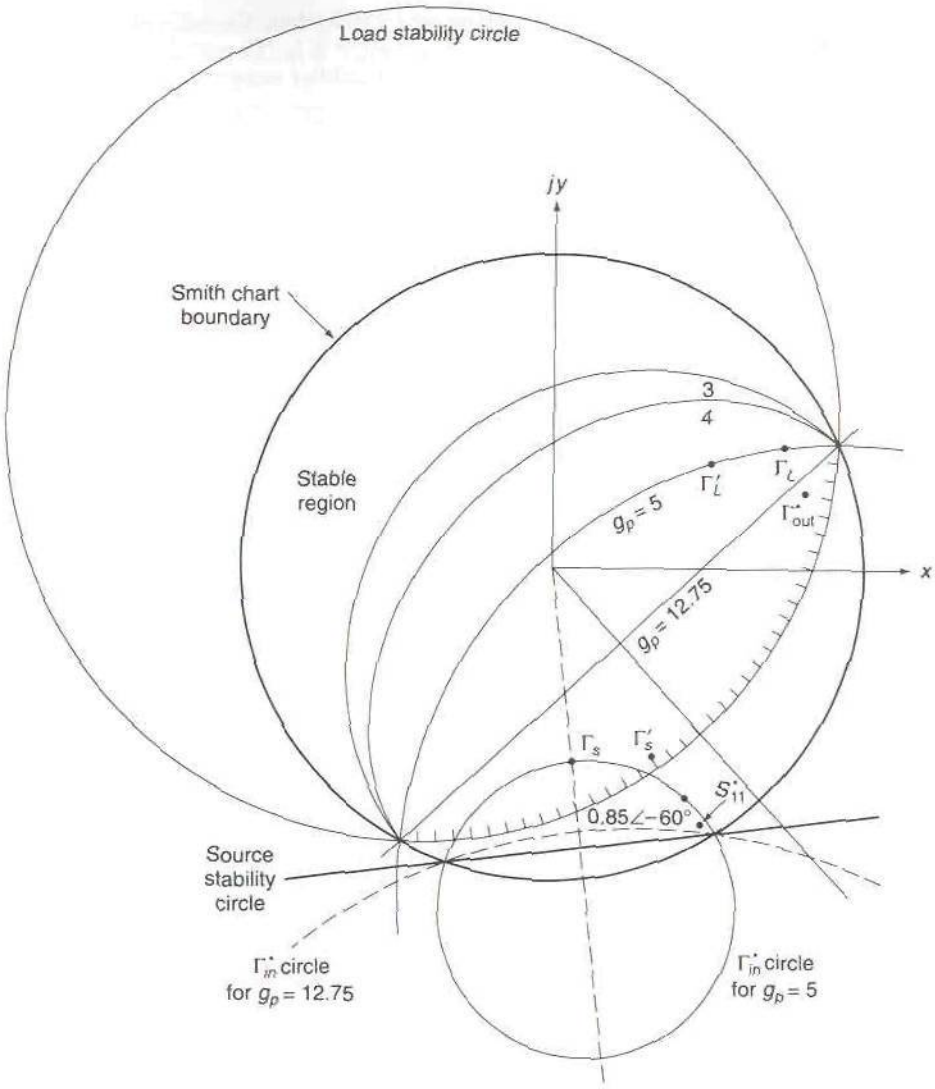


FIGURE 10.21
Load and source stability circles and constant normalized power-gain circles for the amplifier designed in Example 10.2.

required. On the straight line joining the two invariant points, $g_p = 1/(|\Delta|^2 - |S_{22}|^2) = 12.75$ which is considerably larger than the "Figure of Merit" gain. We could choose Γ_L to lie on this constant gain circle and still have $|\Gamma_{in}| < 1$ with a sufficient stability margin. However, as we will see, the choice of too large a gain will result in poor stability at the output of the amplifier or a very high input and output voltage standing-wave ratio.

If we choose $\Gamma_L = 0$, then $\Gamma_{in} = S_{11}$. For conjugate impedance matching at the input, $\Gamma_s = \Gamma_{in}^* = S_{11}^*$. If we choose this for the source reflection coefficient, then the amplifier output reflection coefficient becomes

$$\Gamma_{out} = \frac{S_{22} - \Delta\Gamma_s}{1 - S_{11}\Gamma_s} = \frac{S_{22} - \Delta S_{11}^*}{1 - |S_{11}|^2} = 0.2838 + j0.3384$$

The magnitude of the output reflection coefficient is 0.4416. Since $\Gamma_L = 0$ the output VSWR is $(1.4416)/(1 - 0.4416) = 2.58$.

The point $\Gamma_s = S_{11}^*$ is shown in Fig. 10.21 and is seen to lie quite close to the source stability boundary. A relatively small increase in Γ_s could lead to oscillations in the output circuit of the amplifier since $|\Gamma_{out}| > 1$ when Γ_s lies inside the source stability circle. A more conservative choice for Γ_s would be $0.85\angle -60°$. This point is also shown in Fig. 10.21. For this choice the input is not conjugate matched. The mismatch at the input is, from (10.12*a*),

$$M = \frac{(1 - |\Gamma_{in}|^2)(1 - |\Gamma_s|^2)}{|1 - \Gamma_s\Gamma_{in}|^2} = \frac{0.19 \times 0.2775}{0.05523} = 0.9547$$

The input VSWR is now

$$\text{VSWR}_1 = \frac{1 + \sqrt{1 - M}}{1 - \sqrt{1 - M}} = 1.54$$

With the above choice for Γ_s, we find that $\Gamma_{out} = 0.3846 + j0.1615$ and the output VSWR is 2.43. For this choice we improved the stability margin but at the expense of having to accept an input VSWR of 1.54.

Before we leave this example, we will introduce a useful method that helps one to choose a value for Γ_s that gives a good stability margin, a low input VSWR, and a given normalized power gain. The input reflection coefficient is given by

$$\Gamma_{in} = \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L}$$

The complex conjugate of this equation is

$$\Gamma_{in}^* = \frac{\Delta^*\Gamma_L^* - S_{11}^*}{S_{22}^*\Gamma_L^* - 1}$$

This is a bilinear transformation. Thus all values of Γ_L that lie on a constant gain circle will map into a circle of Γ_{in}^* values. If we choose Γ_s to lie on the Γ_{in}^* circle, then the input will be matched. We will discuss the Γ_{in}^* circles in more detail in Sec. 10.9. For now we note that the centers of these circles lie on the ray from the origin to the center of the source stability circle. The Γ_{in}^* circles also have the interesting property that all circles cut the Smith chart boundary at the same two points that the source stability circle does. This property is similar to the corresponding property of constant gain circles for unstable devices. In Fig. 10.21 we show the Γ_{in}^* circle for all values of Γ_L on the $g_p = 5$ circle. Some of the Γ_{in}^* values will correspond to Γ_L values outside the Smith chart. The best stability margin is obtained by choosing Γ_s as the point shown as Γ_s in Fig. 10.21, since this is farthest away from the source stability circle.

For this example the Γ_{in}^* circle has a center at $0.11833 - j1.09625$ and a radius of 0.46922. The value of Γ_s at the illustrated point is $0.06797 - j0.6297$. With this choice for $\Gamma_s = \Gamma_{in}^*$, the input is matched and we get a normalized power gain of 5 or a real power gain of $5 \times |S_{21}|^2 = 45$ or 16.53 dB. However, we must check if the corresponding value of Γ_L is satisfactory. With the above choice of Γ_s, we find that $\Gamma_{out} = 0.7819 - j0.2077$. For the chosen value of Γ_{in}^*, we can solve for Γ_L using the equation

$$\Gamma_L = \frac{S_{11} - \Gamma_{in}}{\Delta - S_{22}\Gamma_{in}} = 0.7097 + j0.36165$$

The value of Γ_L for the chosen Γ_{in} is well within the load stability circle. In Fig. 10.21 we show the points Γ_L and Γ_{out}^*. The output VSWR is easily calculated and is 2.52. Thus this design gives a matched input, a good power gain of 45, good stability, and an output VSWR equal to 2.52. However, we do need an output matching network that will transform the line impedance of 50 Ω into a load impedance given by

$$Z_L = \frac{1 + \Gamma_L}{1 - \Gamma_L} Z_c = 85 + j168.2$$

A better output VSWR could be obtained by relaxing the input VSWR. By trial and error we find that, by using $\Gamma_s = 0.35 - j0.6062$ and $\Gamma_L = 0.47237 + j0.31859$, we can achieve a normalized power gain of 5, an input VSWR of 1.366, and an output VSWR of 1.8515. These final values of Γ_s and Γ_L are shown in Fig. 10.21 as the points Γ_s' and Γ_L'. This last design results in good gain, low input and output VSWR, and good stability margins.

The last item that we will examine is what happens if we try to design the amplifier for a gain significantly larger than the "Figure of Merit" gain. A very large gain means that the amplifier is very close to oscillation. Hence it can be anticipated that either Γ_s, Γ_L, or both will be close to their respective stability circle boundaries. The Γ_{in}^* circle for all values of Γ_L on the $g_p = 12.75$ gain circle is shown as the dashed circle in Fig. 10.21. This Γ_{in}^* circle lies very close to the source stability circle. For a matched input, Γ_s would have to be chosen to lie on this Γ_{in}^* circle and clearly this would not give an adequate stability margin. If we choose Γ_s sufficiently far away from this circle to obtain a good stability margin, then we will end up with a very poor impedance match at the amplifier input. The consequence of a poor impedance match at the input is that only a small fraction of the incident power on the input transmission line will be delivered to the amplifier input. Thus, even though the power gain will be high, the transducer gain will be low. In practice, microwave amplifiers are generally designed for power gains no greater than the "Figure of Merit" gain because of the poor stability and high input and output VSWR that results for larger gain designs.

As is apparent there is no direct way to control all design objectives individually, since improvements in some areas generally are accompanied by a deterioration in some other characteristic.

In Sec. 10.9 we will discuss low-noise amplifier design and will find that this will result in an additional constraint on Γ_s, since the best noise figure is obtained only for a particular value of Γ_s.

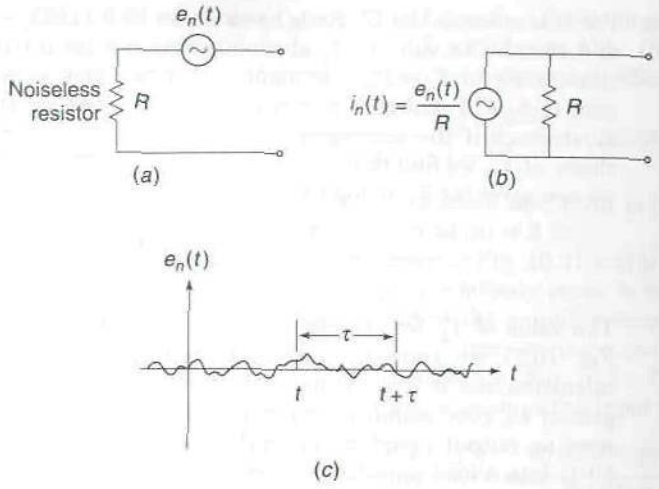


FIGURE 10.22
(*a*) Thévenin equivalent circuit which uses a noise voltage generator; (*b*) equivalent circuit for a noisy resistor in which a noise current source is used; (*c*) typical noise voltage waveform.

10.8 BASIC NOISE THEORY

As a result of thermal agitation, the electrons in a resistor have an inherent random motion which results in a random voltage appearing across the resistor terminals. This random voltage is referred to as noise. There is no analytical way to describe the exact voltage waveform; so we must be content with a description of certain average characteristics of the noise. We can model a noisy resistor as a noise-free resistor in series with a noise voltage generator $e_n(t)$ or in shunt with a noise current source $i_n(t)$ as shown in Figs. 10.22*a* and *b*. In Fig. 10.22*c* we show a typical noise voltage waveform that might be produced.

Noise is a random process and its effects in a linear system are analyzed using statistical methods. For this purpose we construct an ensemble of macroscopically identical systems, e.g., we consider an infinite number of resistors with each one producing its own noise voltage. Averages of various products of the noise voltages at different times, such as $e_n(t_1)$, $e_n(t_1)e_n(t_2)$, etc., are obtained by averaging over the ensemble of noise waveforms.

Thermal noise is generally regarded as a stationary ergodic noise process which is a random process for which ensemble averages can be replaced by time averages. Thus, in our brief summary of basic noise theory, we will use time averages. Since our objectives are only to obtain the basic results needed to derive the equations required for low-noise amplifier design, our discussion will be brief. Much more complete treatments of noise are widely available in texts on statistical communication theory and

the reader should consult some of these for the details that are missing in our treatment.

The time-average value of the noise voltage, given by

$$\langle e_n(t)\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e_n(t)\,dt = 0 \tag{10.31}$$

is zero. The correlation function for the noise voltage is the average value of the product of the noise voltage at time t and that at a later time $t + \tau$; thus

$$\begin{aligned} C(\tau) &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e_n(t)e_n(t+\tau)\,dt \\ &= \langle e_n(t)e_n(t+\tau)\rangle \end{aligned} \tag{10.32}$$

where $C(\tau)$ is the correlation function. If $\tau = 0$ we obtain the average power $\langle e_n^2\rangle$ associated with the noise (the noise voltage is thought of as being applied to a 1-Ω resistor, so that the dimensions are those of power). The average power in noise is distributed over a broad band of frequencies because noise voltage waveforms contain a broad spectrum of frequencies. The power spectral density $S_n(\omega)$ of noise is given by the Fourier transform of the correlation function; thus

$$S_n(\omega) = \int_{-\infty}^{\infty} C(\tau)e^{-j\omega\tau}\,d\tau \tag{10.33a}$$

The inverse transform relationship is

$$C(\tau) = \int_{-\infty}^{\infty} S_n(\omega)e^{j\omega\tau}\frac{d\omega}{2\pi} \tag{10.33b}$$

The power spectral density represents the noise power in the spectral domain; so $S_n(\omega)\,\Delta f$ is the noise power in a frequency increment Δf.

At room temperatures the power spectral density of thermal noise is constant up to frequencies of the order of 1,000 GHz and decreases at higher frequencies. Thus, at microwave frequencies and below, we can assume that the spectral density is a constant or is flat. This is equivalent to having a correlation function that is a constant multiplying the delta function $\delta(\tau)$, that is,

$$C(\tau) = C_0\,\delta(\tau)$$

since the Fourier transform of $\delta(\tau)$ is a constant equal to unity. Noise with a constant power spectral density is called white noise and is uncorrelated noise.

The power spectral density is an even function of ω; so we can choose the spectral density such that only positive values of ω need to be considered. For thermal noise in a resistor, the power spectral density for the noise voltage is given by Nyquist's formula

$$S_e(\omega) = 4kTR \qquad \omega \geq 0 \tag{10.34}$$

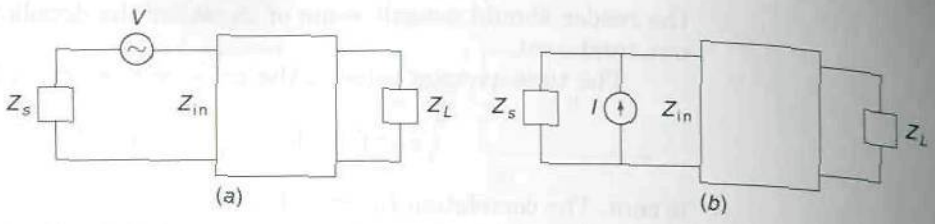


FIGURE 10.23
A two-port network connected to (*a*) a voltage source; (*b*) a current source.

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant and T is the absolute temperature of the resistor R. Thus the amount of noise power P_n in a frequency interval Δf is given by

$$P_n = 4kTR\,\Delta f \tag{10.35}$$

Thermal noise in a resistor is also called Johnson noise, after one of the early investigators of noise.

If we use the equivalent current source model shown in Fig. 10.22*b*, then the average power, if the current $i_n(t)$ flows in a 1-Ω resistor, is given by $\langle i_n^2(t)\rangle$ and has a power spectral density given by

$$S_i(\omega) = \frac{4kT}{R} \qquad \omega \geq 0 \tag{10.36}$$

for thermal noise.

Filtered Noise

Consider a sinusoidal voltage generator with a complex rms voltage V that is connected in series with a source impedance Z_s and a two-port network as shown in Fig. 10.23*a*. The input current produced by V is $V/(Z_s + Z_{in})$, where Z_{in} is the input impedance to the network. The input power produced by V is given by

$$P_{in,1} = \left|\frac{V}{Z_s + Z_{in}}\right|^2 R_{in} = \frac{|V|^2}{4R_s}\frac{4R_s R_{in}}{|Z_s + Z_{in}|^2}$$

$$= \frac{|V|^2}{4R_s} M \tag{10.37}$$

where $|V|^2/4R_s$ is the available power and M is the impedance-mismatch factor. The power transfer function is $M/4R_s$ and is a function of ω since Z_s and Z_{in} are functions of ω. If the voltage generator is replaced by a noise voltage source $e_n(t)$ with a power spectral density $S_e(\omega)$, then the input noise power in a frequency band Δf centered on ω is given by multiplying the source power spectral density by the power transfer function and the

factor Δf; thus

$$P_{n,e}(\omega) = S_e(\omega)\frac{M(\omega)}{4R_s}\Delta f \tag{10.38}$$

The total input noise power is obtained by integrating over all frequencies; thus

$$P_{nT} = \int_0^\infty S_e(\omega)\frac{M(\omega)}{4R_s}\frac{d\omega}{2\pi} \tag{10.39}$$

Consider next the circuit shown in Fig. 10.23*b*. The input current from the current source is $Z_s I/(Z_s + Z_{\text{in}})$ and the input power is

$$P_{\text{in},2} = \left|\frac{IZ_s}{Z_s + Z_{\text{in}}}\right|^2 R_{\text{in}} = \frac{|I|^2|Z_s|^2}{4R_s}\frac{4R_s R_{\text{in}}}{|Z_s + Z_{\text{in}}|^2}$$

$$= \frac{|I|^2}{4G_s}M \tag{10.40}$$

where $G_s = \text{Re}(1/Z_s) = R_s/|Z_s|^2$. If the sinusoidal current source is replaced by a noise current source $i_n(t)$ with power spectral density $S_i(\omega)$, then the input noise power in a frequency band Δf is given by

$$P_{n,i}(\omega) = S_i(\omega)\frac{M(\omega)}{4G_s}\Delta f \tag{10.41}$$

The power spectral density of the noise power delivered to Z_L would, by analogy, be given by the product of the power spectral density of the source and the power transfer function from the source to the output load impedance. Since we can describe the power delivered to Z_L by the product of P_{in} with the power gain $G_p(\omega)$, we see that the noise current source will produce an output power in Z_L with a spectral density given by $G_p(\omega)M(\omega)S_i(\omega)/4G_s$. A similar expression holds for the output power spectral density produced by the noise voltage source $e_n(t)$ acting alone.

When both sources V and I are acting, the input current will be

$$I_{\text{in}} = \frac{V + IZ_s}{Z_s + Z_{\text{in}}}$$

and the input power will be given by

$$P_{\text{in},3} = \frac{(V + IZ_s)(V^* + I^*Z_s^*)}{|Z_s + Z_{\text{in}}|^2}R_{\text{in}}$$

$$= \frac{|V|^2}{4R_s}M + \frac{|I|^2}{4G_s}M + 2\,\text{Re}\,\frac{VI^*Z_s^*R_{\text{in}}}{|Z_s + Z_{\text{in}}|^2} \tag{10.42}$$

Because of interaction between the two sources, the input power is not simply the sum of that from each source acting independently.

When we have two noise sources $e_n(t)$ and $i_n(t)$ acting simultaneously, there is no input power caused by the interaction between $e_n(t)$ and $i_n(t)$ when $e_n(t)$ and $i_n(t)$ are uncorrelated or statistically independent noise sources. For this case the input noise power in a frequency band Δf is the sum of that given by (10.38) and (10.41). When there is a degree of correlation between $e_n(t)$ and $i_n(t)$, there will be some input noise power due to the source interaction.

The cross-correlation between the current source $i_n(t)$ and voltage source $e_n(t)$ is given by

$$C_x(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} i_n(t)e_n(t+\tau)\,dt \qquad (10.43a)$$

The Fourier transform of $C_x(\tau)$ gives the cross-power spectral density $S_x(\omega)$, that is,

$$S_x(\omega) = S_{xr}(\omega) + jS_{xi}(\omega) = \int_{-\infty}^{\infty} C_x(\tau)e^{-j\omega\tau}\,d\tau \qquad (10.43b)$$

If we replace ω by $-\omega$, we see that $S_x(-\omega) = S_x^*(\omega)$ since $C_x(\tau)$ is real. From this result we find that $S_{xr}(\omega)$ is an even function of ω and $S_{xi}(\omega)$ is an odd function of ω. For input noise power calculations, we replace $|V|^2$ by the noise voltage source power spectral density $S_e(\omega)$, replace $|I|^2$ by the power spectral density $S_i(\omega)$ of $i_n(t)$, and replace VI^* by the cross-power spectral density $S_x(\omega)$ in (10.42). Thus, for partially correlated noise sources, the total input noise in a frequency band Δf is given by

$$P_n = \Delta f\left\{S_e(\omega)\frac{M}{4R_s} + S_i(\omega)\frac{M}{4G_s} + \frac{4[S_{xr}(\omega)R_s + S_{xi}(\omega)X_s]}{|Z_s + Z_{in}|^2}R_{in}\right\} \qquad (10.44)$$

The extra factor of 2 in the last term is due to the fact that we have combined the contribution from negative values of ω with those from positive values of ω using the fact that $S_{xr}R_s$ and $S_{xi}X_s$ are even functions of ω. Also, we have defined the spectral densities $S_e(\omega)$ and $S_i(\omega)$ so that only positive values of ω are to be integrated over to get the total input noise power.

The power spectral density of the noise produced in a network differs from that of the noise source because the network response depends on frequency. The source noise spectrum is filtered by the network. If the source noise spectrum is flat (white noise), the noise spectrum produced at some point in the network is not flat. Noise with a nonconstant power spectral density is called colored noise.

From (10.44) we find that the power spectral density of the noise power delivered to Z_L is

$$S(\omega) = \left\{ \frac{M}{4R_s} S_e(\omega) + \frac{M}{4G_s} S_i(\omega) + \left[S_{xr}(\omega) + \frac{X_s}{R_s} S_{xi}(\omega) \right] M \right\} G_p(\omega) \tag{10.45}$$

after multiplying by the power gain G_p of the two-port network. The total output noise power delivered to Z_L is

$$P_{n,\text{out}} = \int_0^\infty S(\omega) \frac{d\omega}{2\pi} \tag{10.46}$$

Noise in Active Devices

In an active device such as a transistor, there are three main contributions to the noise produced by the device: (1) thermal noise in the resistive elements that are present, (2) shot noise due to the discrete nature of the charge carriers that constitute the current flow across p-n junctions, and (3) flicker noise that has a power spectral density proportional to $1/f$ in the frequency domain. The current that flows across a p-n junction is not a smooth stream of charge. It consists of discrete charges that cross the junction in a more or less random manner like raindrops falling on a tin roof. The noise associated with this current is called shot noise and is directly proportional to the dc bias currents. Shot noise, like thermal noise, has a flat power spectral density. For low noise an active device is operated with low dc bias currents. However, since the transconductance, and hence the gain of the device, decreases with a reduction of the bias current, there is an optimum lower value of bias current that gives the best compromise between low noise and gain and produces the optimum noise figure for the device. In Fig. 10.24 we show the typical behavior of noise figure versus drain current for a MESFET.

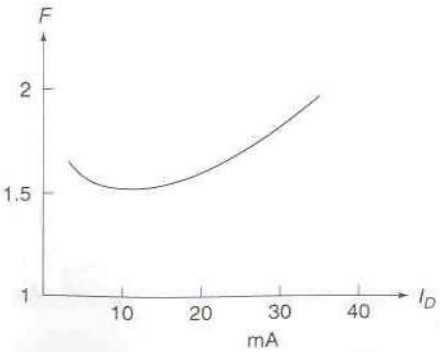


FIGURE 10.24
Variation of the noise figure F of a MESFET as a function of drain current.

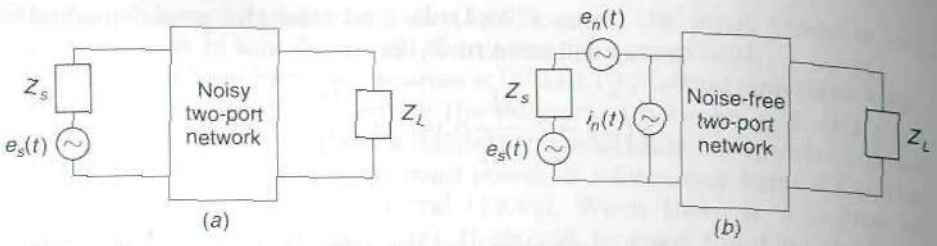


FIGURE 10.25
Equivalent input noise sources for a noisy linear two-port network.

Flicker noise is usually negligible relative to thermal noise and shot noise at frequencies greater than a few kilohertz for bipolar transistors but may be important in MESFETs up to frequencies as high as 100 MHz.

Noisy Two-Port Networks

In analyzing the noise produced at the output of a linear two-port network due to the internal noise sources, we can replace all of the internal noise sources by a series noise voltage generator $e_n(t)$ and a shunt noise current generator $i_n(t)$ at the input as shown in Fig. 10.25. The total noise power at the output can be found by evaluating the noise output produced by $e_n(t), i_n(t)$, and the thermal noise in the resistive component R_s of the source impedance. Two equivalent noise sources are needed at the input because if the input is short-circuited, that is, $Z_s = 0$, the source $i_n(t)$ does not produce any output noise, yet the noisy two-port does have a noise output under short-circuit conditions at the input so a noise voltage source $e_n(t)$ is required. Similarly, under open-circuit conditions $e_n(t)$ does not produce any output noise; so a noise source $i_n(t)$ is needed to represent the equivalent input noise source under open-circuit conditions. The two noise sources $e_n(t)$ and $i_n(t)$ are not completely independent since a part of $e_n(t)$ and $i_n(t)$ may arise from the same basic noise-producing mechanism within the two-port network. Thus, in general, there is some cross-correlation between $e_n(t)$ and $i_n(t)$ with a resultant nonzero cross-power spectral density.

It is common practice to express the power spectral density associated with the two noise sources $e_n(t)$ and $i_n(t)$ in a form similar to that given by (10.34) and (10.36) for thermal noise. When this is done, flicker noise, which is low-frequency noise with a $1/f$ spectrum, is not included. Since linear microwave amplifiers do not produce an output for low-frequency input signals, the neglect of flicker noise can be justified. We will thus specify the

spectral densities as follows:

$$\text{For } e_n(t) \qquad S_e(\omega) = 4kTR_e$$

$$\text{For } i_n(t) \qquad S_i(\omega) = 4kTG_i$$

and

$$2[S_{xr}(\omega) + jS_{xi}(\omega)] = 4kT(\gamma_r + j\gamma_i)$$

where R_e is an equivalent noise resistance, G_i is an equivalent noise conductance, and $\gamma_r + j\gamma_i$ is a complex equivalent noise impedance. A total of four parameters, $R_e, G_i, \gamma_r, \gamma_i$, are needed to describe the noise properties of a noisy two-port network.

In terms of the above spectral densities, we can express the total noise input to the noise-free two-port network, in a frequency band Δf, as follows by using (10.44):

$$P_{n,\text{in}} = kT\,\Delta f M + kT\,\Delta f \frac{R_e}{R_s} M + kT\,\Delta f \frac{G_i}{G_s} M + 2kT\,\Delta f \frac{R_s\gamma_r + X_s\gamma_i}{R_s} M \tag{10.47}$$

In this equation the first term on the right, $kT\,\Delta f M$, is the input thermal noise from the source resistance R_s. The output noise in Z_L in a frequency band Δf is obtained by multiplying by the power gain $G_p(\omega)$ of the network. The noise produced at the output termination by the equivalent sources $e_n(t)$ and $i_n(t)$ placed at the input of the network is fully equivalent to that produced by the internal noise mechanisms in the real noisy two-port network.

10.9 LOW-NOISE AMPLIFIER DESIGN

In a typical microwave communication system, the information to be communicated is modulated onto a microwave carrier and radiated into space by means of an antenna. At the receiving site an antenna is also used to intercept a small portion of the radiated signal energy. The receiving antenna will also pick up a certain amount of noiselike radiation from atmospheric disturbances, radio stars, the sun, and other celestial bodies. The received signal, along with some noise, is very weak and must be amplified to a level where it can be used to produce the desired video, audio, or digital output information that was transmitted. The function of the first amplifier stage is to amplify the signal with the addition of a minimum amount of extra noise. Thus the first amplifier stage should be designed for minimum noise. If the power gain of the first stage is around 10 or more, the signal will be sufficiently large at the output of the first stage, so that additional noise contributed by the following amplifier stages will have a

negligible degrading effect on the overall signal-to-noise power ratio, provided that the noise contribution of the second stage is moderate. In the design of the first stage, the minimum noise requirement is more important than maximum power gain or output VSWR, provided a power gain of 10 or more can be achieved. Hence we can relax the gain and output VSWR requirements in order to achieve the objective of a minimum noise contribution from the first stage.

In this section we will discuss noise figure and the design of an amplifier for minimum noise. We will show that there is an optimum source impedance Z_s (or source reflection coefficient Γ_s) that will result in the lowest noise figure. We will also introduce constant noise-figure circles that can be plotted on the Γ_s plane and which will show in a pictorial way the increase in noise figure that occurs when the optimum source reflection coefficient cannot be used. For transistors that are not absolutely stable, the use of the optimum source reflection coefficient for minimum noise could result in an unstable amplifier, in which case a noise figure somewhat larger than the minimum one will have to be accepted.

Noise Figure

With reference to the circuit shown in Fig. 10.25*a*, the definition of noise figure F (also called noise factor) is

$$F = \frac{\text{signal-to-noise ratio at input}}{\text{signal-to-noise ratio at output}} \tag{10.48}$$

The output noise is the amplified thermal noise from the source resistance plus the noise produced by the amplifier. The standard definition of F requires the source to be conjugate impedance matched to the network, that is, $Z_s = Z_{\text{in}}^*$, and the source resistance R_s to be at the standard temperature $T_0 = 290$ K. Very often in practice the source resistance is at a different temperature and the source is not matched to the network. In this case the definition (10.48) gives the operating noise figure. If F is given at a single frequency and is based on the noise power in a small frequency band Δf, then the noise figure is called the spot noise figure. When all of the noise sources are referred to the input as equivalent noise sources, then the spot noise figure can be defined as follows:

$$F = \frac{\text{total input noise power to network}}{\text{thermal noise input power from source resistance}} \tag{10.49}$$

where the noise powers are those in a narrow frequency band Δf. Based on this latter definition, the spot noise figure for the system shown in Fig.

10.25 may be obtained from (10.47) by dividing by $kT\,\Delta f M$. Thus we get

$$F = 1 + \frac{R_e}{R_s} + \frac{G_i}{G_s} + 2\frac{R_s\gamma_r + X_s\gamma_i}{R_s} \tag{10.50}$$

This noise figure is seen to depend on the source impedance as well as on the noise parameters R_e, G_i, γ_r, and γ_i. In (10.50), $G_s = R_s/(R_s^2 + X_s^2)$. The noise figure does not depend on the frequency bandwidth Δf. However, the input signal-to-noise ratio will deteriorate if the amplifier bandwidth is greater than that required to accommodate the signal.

The optimum source impedance that will minimize the noise figure is obtained by setting $\partial F/\partial R_s = 0$, $\partial F/\partial X_s = 0$, and solving for R_s and X_s. It is readily found that

$$X_s = X_m = -\frac{\gamma_i}{G_i} \tag{10.51a}$$

and $R_s^2 + X_s^2 = R_e/G_i$, which gives

$$R_s = R_m = \sqrt{\frac{R_e}{G_i} - \frac{\gamma_i^2}{G_i^2}} \tag{10.51b}$$

When these values for R_s and X_s are used in (10.50), we obtain the minimum noise figure F_m.

We now replace γ_i by $-X_m G_i$ in (10.50) and consider $F - F_m$ which is given by

$$\begin{aligned}
F - F_m &= \frac{1}{R_s}\left[R_e + (R_s^2 + X_s^2)G_i - 2X_m X_s G_i\right] \\
&\quad - \frac{1}{R_m}\left[R_e + (R_m^2 + X_m^2)G_i - 2X_m^2 G_i\right] \\
&= \frac{1}{R_s}\left[R_e + (R_s - R_m)^2 G_i + (X_s - X_m)^2 G_i + 2R_s R_m G_i\right. \\
&\qquad \left. - R_m^2 G_i - X_m^2 G_i\right] \\
&\quad - \frac{1}{R_m}\left[R_e + R_m^2 G_i - X_m^2 G_i\right] \\
&= \frac{G_i}{R_s}\left[(R_s - R_m)^2 + (X_s - X_m)^2\right] \\
&\quad + \left(\frac{1}{R_s} - \frac{1}{R_m}\right)(R_e - R_m^2 G_i - X_m^2 G_i)
\end{aligned}$$

upon using $2R_s R_m G_i/R_s = 2R_m^2 G_i/R_m$ and putting this term with the

factor $-1/R_m$. The factor $R_e - (R_m^2 + X_m^2)G_i$ is zero as may be seen from (10.51). Thus we obtain

$$F = F_m + \frac{G_i}{R_s}\left[(R_s - R_m)^2 + (X_s - X_m)^2\right] \tag{10.52}$$

This is a very useful relationship in practice, since it determines the noise figure in terms of the minimum value F_m obtained when the optimum source impedance $R_m + jX_m$ is used along with only one additional parameter G_i. By adjusting the source impedance, both the optimum source impedance and the minimum noise figure can be determined experimentally. Transistor manufacturers will often give the minimum noise figure F_m, the optimum source impedance $R_m + jX_m$ or source reflection coefficient $\Gamma_m = (R_m + jX_m - Z_c)/(R_m + jX_m + Z_c)$, and the noise conductance G_i or noise resistance R_e. If R_e is given the equation $R_s^2 + X_s^2 = R_e/G_i$, which gave (10.51*b*), can be used to find G_i; thus

$$G_i = \frac{R_e}{R_m^2 + X_m^2} \tag{10.53}$$

The given data can be used in (10.52) to find the noise figure F.

Noise Figure for Cascaded Stages

Figure 10.26 shows a two-stage microwave amplifier. The power gains of the two stages are G_{P1} and G_{P2}. The input impedance mismatch is M_1 for the first stage and M_2 for the second stage. The equivalent noise sources for stage 2 are designated as $e_n'(t)$ and $i_n'(t)$. The source impedance for stage 2 is the output impedance of stage 1. The noise figure for stage 1 will be called F_1 and that for stage 2 will be called F_2. The noise figure for the system will be called F.

From the definition (10.49) for noise figure, we see that the total input noise to the first stage is $F_1(kT\Delta f M_1)$. At the output of stage 2, this noise

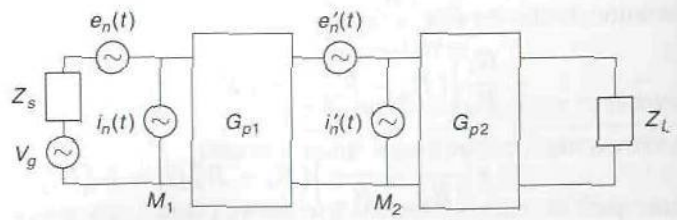


FIGURE 10.26
A two-stage microwave amplifier showing equivalent noise sources.

has been amplified by the factor $G_{P1}G_{P2}$ to give an output noise contribution

$$P_{n1} = G_{P1}G_{P2}F_1kT\,\Delta f M_1$$

A similar analysis shows that the noise sources $e'_n(t)$ and $i'_n(t)$ produce an input noise power equal to $kT\,\Delta f M_2(F_2 - 1)$ to stage 2. Note that we do not include the thermal noise from the source resistance for stage 2, since this comes from the resistive part of the output impedance of stage 1. The latter is an internal noise source for stage 1 and its effects have already been included in the equivalent noise sources for stage 1. The total output noise power in a frequency band Δf is readily seen to be

$$P_{n,\text{out}} = G_{P1}G_{P2}F_1kT\,\Delta f M_1 + G_{P2}(F_2 - 1)kT\,\Delta f M_2$$

where the last term is the amplified noise produced by $e'_n(t)$ and $i'_n(t)$. The output noise can be considered to be the amplified thermal noise in R_s multiplied by the two-stage amplifier noise factor F; thus

$$P_{n,\text{out}} = G_{P1}G_{P2}FkT\,\Delta f M_1$$

By equating these two expressions for output noise power, we get

$$F = F_1 + (F_2 - 1)\frac{M_2}{M_1G_{P1}} \tag{10.54}$$

This result shows that the second-stage noise does not produce a large change in the noise figure above that of stage 1 alone, since the contribution from the second stage is divided by G_{P1}, provided G_{P1} is of order 10 or more, and the second-stage noise figure is on the order of that of the first stage.

For a cascade of three or more amplifier stages, a similar analysis shows that the overall noise figure is given by

$$F = F_1 + (F_2 - 1)\frac{M_2}{M_1G_{P1}} + (F_3 - 1)\frac{M_3}{M_1G_{P1}G_{P2}} + \cdots \tag{10.55}$$

Clearly, successive stages do not degrade the overall noise figure significantly.

The noise in an amplifier can be accounted for by imagining that it comes from the thermal noise in the source resistance by assigning an equivalent noise temperature T_e to the source resistance. Thus we can write

$$\begin{aligned} P_{n,\text{out}} &= G_{P1}G_{P2} \cdots M_1kT_e\,\Delta f \\ &= G_{P1}G_{P2} \cdots FM_1kT\,\Delta f \end{aligned}$$

and hence $T_e = FT$. The excess temperature $T_e - T = (F - 1)T$ is called the noise temperature of the amplifier. A low-noise amplifier with a noise figure of 1.4 would have a noise temperature of $0.4 \times 290 = 116$ K.

If all sources of noise in a system, which includes amplifier noise, thermal noise, and radiation noise picked up by the antenna, are regarded

as thermal noise in the source resistance R_s, then the temperature that must be assigned to R_s so as to give the same amount of total noise is T_s and is called *the system noise temperature.*

Constant Noise-Figure Circles

For the purpose of low-noise amplifier design, it is useful to plot constant noise-figure circles on the source reflection-coefficient plane. Since

$$\Gamma_s = \frac{Z_s - Z_c}{Z_s + Z_c} \qquad \Gamma_m = \frac{Z_m - Z_c}{Z_m + Z_c},$$

$$(R_s - R_m)^2 + (X_s - X_m)^2 = |Z_s - Z_m|^2$$

the expression (10.52) for noise figure can be written as

$$F - F_m = \frac{G_i}{R_s} Z_c^2 \left(\left| \frac{1 + \Gamma_s}{1 - \Gamma_s} - \frac{1 + \Gamma_m}{1 - \Gamma_m} \right|^2 \right)$$

$$= \frac{G_i}{R_s} Z_c^2 \frac{4|\Gamma_s - \Gamma_m|^2}{|1 - \Gamma_s|^2 |1 - \Gamma_m|^2}$$

Next we use

$$\frac{2R_s}{Z_c} = \frac{Z_s + Z_s^*}{Z_c} = 2\,\mathrm{Re}\,\frac{1 + \Gamma_s}{1 - \Gamma_s}$$

$$= 2\,\mathrm{Re}\,\frac{(1 - \Gamma_s)(1 - \Gamma_s^*)}{|1 - \Gamma_s|^2} = \frac{2\left(1 - |\Gamma_s|^2\right)}{|1 - \Gamma_s|^2}$$

to get

$$F - F_m = 4G_I \frac{|\Gamma_s - \Gamma_m|^2}{|1 - \Gamma_m|^2 \left(1 - |\Gamma_s|^2\right)} \qquad (10.56a)$$

where $G_I = G_i Z_c$ is the normalized value of G_i. A similar expression can be derived involving $R_N = R_e/Z_c$, namely,

$$F - F_m = 4R_N \frac{|\Gamma_s - \Gamma_m|^2}{|1 + \Gamma_m|^2 \left(1 - |\Gamma_s|^2\right)} \qquad (10.56b)$$

We now introduce a parameter N_i defined as

$$N_i = \frac{(F_i - F_m)|1 + \Gamma_m|^2}{4R_N} = \frac{(F_i - F_m)|1 - \Gamma_m|^2}{4G_I} \qquad (10.57)$$

where N_i corresponds to a chosen value F_i for F. From (10.56) we now

obtain

$$N_i = \frac{|\Gamma_s - \Gamma_m|^2}{1 - |\Gamma_s|^2} = \frac{\Gamma_s\Gamma_s^* - \Gamma_s\Gamma_m^* - \Gamma_s^*\Gamma_m + \Gamma_m\Gamma_m^*}{1 - \Gamma_s\Gamma_s^*}$$

which can be written in a form such as (10.1) and identified as describing a circle in the Γ_s plane. The center of the circle is located at

$$\Gamma_{sf} = \frac{\Gamma_m}{1 + N_i} \tag{10.58a}$$

and the radius of the circle is given by

$$R_f = \frac{\sqrt{N_i^2 + N_i(1 - |\Gamma_m|^2)}}{1 + N_i} \tag{10.58b}$$

We can plot the circles for various values of N_i, determined by (10.57) for chosen values of F_i. Each circle shows the values of Γ_s that can be used in order to get a noise figure equal to F_i. For $N_i = 0$ the circle degenerates to a single point at Γ_m giving the minimum noise figure F_m. If $\Gamma_s = \Gamma_m$ is a stable point, then the amplifier can be designed to give a minimum noise figure F_m. Usually, we can choose the load Z_L so that $\Gamma_s = \Gamma_m$ gives a stable design. However, sometimes the input VSWR is too high if we use $\Gamma_s = \Gamma_m$. In such a case a choice for Γ_s on a constant noise-figure circle with $F_i > F_m$ would be used in order to obtain a better input VSWR.

Example 10.3 Low-noise amplifier design. A silicon bipolar transistor has the following parameters at 4 GHz:

$$S_{11} = 0.36\angle 148^\circ \qquad S_{12} = 0.11\angle 42^\circ \qquad S_{21} = 1.57\angle 27^\circ$$

$$S_{22} = 0.67\angle -64^\circ \qquad \Gamma_m = 0.38\angle -153^\circ \qquad R_N = 0.4$$

$$F_m = 1.905\ (2.8\ \text{dB})$$

By using (10.25a) we find $K = 1.2421 > 1$; so the transistor is an absolutely stable device. From (10.19) we find that the maximum power gain is 7.2123 for which the maximum normalized gain is $7.2123/|S_{21}|^2 = 2.926$. Since the device is absolutely stable, we can use conjugate impedance matching. From (10.17) we find that this requires $\Gamma_s = -0.53287 - j0.40911$ and $\Gamma_L = 0.34159 + j0.74723$. With conjugate impedance matching, the input and output VSWR equals unity. If we use the above value of Γ_s in (10.56b), we find that the noise figure is 2.49 or 3.96 dB. This is 1.16 dB greater than the minimum value. It is desirable to have a lower noise figure and yet not sacrifice any gain which already is on the low side. We can accept an increase in VSWR at both the input and the output of the amplifier. A design using $g_p = 2.9$, which is very close to the maximum value, will be attempted. For an aid in the design process, the two constant normalized power-gain circles $g_p = 2.9$ and $g_p = 2.3$

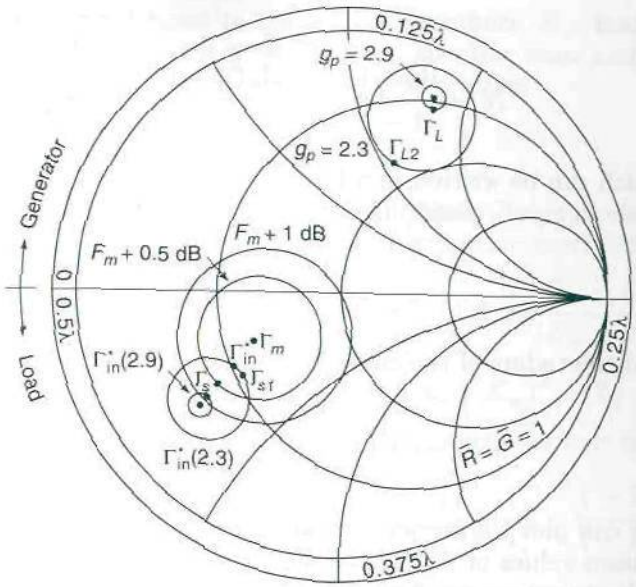


FIGURE 10.27
Constant normalized power-gain circles, constant noise-figure circles, and Γ_{in}^* circles used in the low-noise amplifier design example.

have been constructed and are shown in Fig. 10.27. The centers and radii of these circles were found using (10.28) and are

$$\Gamma_{Lg} = 0.34026 + j0.7443 \qquad R_{Lg} = 0.03574 \qquad \text{for } g_p = 2.9$$

$$\Gamma_{Lg} = 0.30532 + j0.6679 \qquad R_{Lg} = 0.2052 \qquad \text{for } g_p = 2.3$$

As explained in Example 10.2 the values of Γ_L on a g_p = constant circle generate a set of values for Γ_{in} and hence a circle of Γ_{in}^* values. The Γ_{in}^* circles for $g_p = 2.9$ and 2.3 are also shown in Fig. 10.27. The centers and radii of these two circles [the equations for these circles are given in Sec. 10.10 as (10.59a) and (10.59b)] are

$$\Gamma_{in,c}^* = -0.53198 - j0.40843 \qquad R_{in} = 0.03029 \qquad \text{for } g_p = 2.9$$

$$\Gamma_{in,c}^* = -0.51225 - j0.39328 \qquad R_{in} = 0.148 \qquad \text{for } g_p = 2.3$$

Also shown in Fig. 10.27 are the constant noise-figure circles for $F = 2.138$ (3.3 dB) and 2.4 (3.8 dB), which correspond to 0.5 dB and 1 dB greater than F_m. The centers and radii for these circles are found using (10.57) and (10.58) and are

$$\Gamma_{sf} = -0.31706 - j0.16155 \qquad R_f = 0.23444 \qquad \text{for } F_m + 0.5 \text{ dB}$$

$$\Gamma_{sf} = -0.2959 - j0.1508 \qquad R_f = 0.3317 \qquad \text{for } F_m + 1 \text{ dB}$$

If we want an input VSWR of unity, then we must choose Γ_s to lie on the Γ_{in}^* circle. The dot shown inside the smallest Γ_{in}^* circle is the value of Γ_{in}^* when

conjugate impedance matching is used. The figure clearly shows that if we choose Γ_s to lie on this point, the noise figure will be more than 1 dB greater than F_m. Our earlier calculation gave a value of 1.16 dB greater. If we are willing to relax the gain requirement to $g_p = 2.3$, then the figure shows that we can obtain a unity input VSWR and a noise figure somewhat better than $F_m + 0.5$ dB by choosing Γ_s as the point Γ_{s1} shown in Fig. 10.27. If we choose $\Gamma_s = \Gamma_m$ and $g_p = 2.3$, then the best input VSWR is obtained by choosing a Γ_L that will produce a Γ_{in}^* that lies as close as possible to $\Gamma_s = \Gamma_m$. This is the point $\Gamma_{in}^* = -0.3386 - j0.1725$ shown in Fig. 10.27. The corresponding value of Γ_L is $\Gamma_{L2} = 0.1818 + j0.504$ and is also shown in the figure. These choices result in a minimum noise figure, an input VSWR equal to 1.392, and an output VSWR equal to 2.069. The noise figure is a minimum value and the VSWRs are acceptable but the gain is only $2.3 \times |S_{21}|^2 = 5.67$.

If we insist on having $g_p = 2.9$ which gives a power gain of 7.15 and we also want a good input VSWR, then we have to accept some increase in noise figure. If we accept a noise figure 0.5 dB greater than F_m, then for the best input VSWR we choose Γ_s to lie on the $F = F_m + 0.5$ dB constant noise-figure circle and as close as possible to the Γ_{in}^* circle for $g_p = 2.9$. This point is $\Gamma_s = -0.471 - j0.338$ and is shown in Fig. 10.27 as the point Γ_s. The required value of Γ_L that will place Γ_{in}^* as close as possible to Γ_s is $\Gamma_L = 0.3217 + j0.7137$ and is shown in Fig. 10.27. These choices result in an input VSWR equal to 1.22 and an output VSWR equal to 1.065. This last design can be considered to be acceptable.

The input and output impedance-matching networks using parallel open-circuited transmission-line stubs can be designed using the method described in Sec. 5.6 [i.e., (5.29*a*) and (5.29*b*)]. Since we use parallel stubs each stub is required to produce only one-half of the susceptance given by (5.29*a*). From the chosen values of Γ_s and Γ_L given above, we readily find that the required normalized source and load admittances are

$$\bar{Y}_s = 1.683 + j1.716 \qquad \bar{Y}_L = 0.1715 - j0.6327$$

The matching networks along with the stub positions and lengths are shown in Fig. 10.28. This figure does not show the dc bias circuit which is also required.

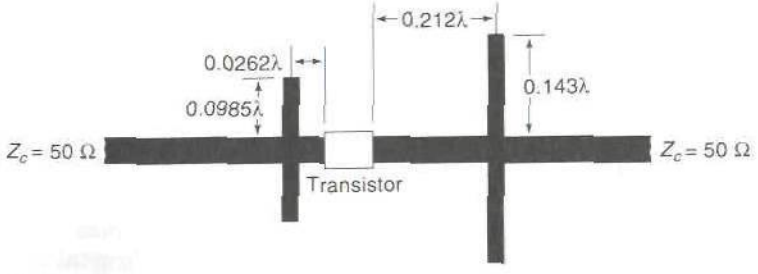


FIGURE 10.28
Matching networks for the low-noise amplifier.

All of the calculations for the above amplifier design and the determination of the values of Γ_L and Γ_s that give the best input and output VSWR for the chosen g_p and F were carried out using the computer program MICROAMP. The matching networks were also designed using this program.

10.10 CONSTANT MISMATCH CIRCLES

In Examples 10.2 and 10.3 we used the Γ_{in}^* circle as an aid in the design of a microwave amplifier with a low input VSWR. The parameters that describe the Γ_{in}^* circle will be derived here. For an amplifier terminated in a load having a load reflection coefficient Γ_L, the input reflection coefficient is given by

$$\Gamma_{in} = \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L}$$

The complex conjugate of this equation is

$$\Gamma_{in}^* = \frac{\Delta^*\Gamma_L^* - S_{11}^*}{S_{22}^*\Gamma_L^* - 1}$$

This is a bilinear transformation, so that all values of Γ_L that lie on a *constant g_p circle will map into a circle of* Γ_{in} values. The center Γ_{Lg} and radius R_{Lg} for a constant normalized power-gain circle are given by (10.28). By using these circle parameters, we can obtain the corresponding parameters for the circle of Γ_{in} and circle of Γ_{in}^* values as described in Sec. 10.7. The center $\Gamma_{in,c}^*$ for the circle of Γ_{in}^* values is given by

$$\Gamma_{in,c}^* = \frac{R_{Lg}^2 S_{22}\Delta^* + (S_{22}\Gamma_{Lg} - 1)(S_{11}^* - \Delta^*\Gamma_{Lg}^*)}{|R_{Lg}S_{22}|^2 - |S_{22}\Gamma_{Lg} - 1|^2} \qquad (10.59a)$$

and the radius is given by

$$R_{in} = \frac{|S_{12}S_{21}|R_{Lg}}{\left||R_{Lg}S_{22}|^2 - |S_{22}\Gamma_{Lg} - 1|^2\right|} \qquad (10.59b)$$

In the design of a microwave amplifier, the choice for the source reflection coefficient is constrained by the requirements that are necessary to obtain a stable amplifier with a low noise figure. Unity input VSWR will be obtained if Γ_s can be chosen to lie on the Γ_{in}^* circle for the chosen g_p circle. If the constraints do not allow this choice or, as sometimes happens, the Γ_{in}^* values inside the Smith chart are unstable values for Γ_s, then Γ_s should be located as close as possible to the Γ_{in}^* circle to obtain the lowest input VSWR. When the choice for Γ_s has been made, then, if Γ_s lies on the

Γ_{in}^* circle, we will require $\Gamma_{in} = \Gamma_s^*$ and we can then find the required value of Γ_L using

$$\Gamma_L = \frac{S_{11} - \Gamma_{in}}{\Delta - S_{22}\Gamma_{in}} \tag{10.60}$$

If Γ_s cannot be placed on the Γ_{in}^* circle, then the best value for Γ_{in}^* is the one that is closest to Γ_s but on the specified Γ_{in}^* circle. The complex conjugate value of this should be used for Γ_{in} in (10.60) to find Γ_L.

For an unstable device a part of the g_p = constant circle will lie outside the Smith chart boundary. The values of Γ_L on the g_p = constant circle outside the Smith chart produce values of Γ_{in}^* that lie outside the Smith chart. As explained in Example 10.2 the Γ_{in}^* circles have two invariant points for an unstable device. These invariant points coincide with the two points at which the source stability circle intersects the Smith chart boundary as illustrated in Fig. 10.21. The proof that these points are invariant points is as follows: For an unstable device we have shown that the g_p = constant circles have two invariant points that coincide with the two points at which the load stability circle intersects the Smith chart boundary (see Sec. 10.7). Since Γ_{in}^* is given by

$$\Gamma_{in}^* = \frac{\Delta^*\Gamma_L^* - S_{11}^*}{S_{22}^*\Gamma_L^* - 1}$$

it is clear that these two particular values of Γ_L, which we will call Γ_{L1} and Γ_{L2}, that are invariant points for the g_p = constant circles will map into two fixed values for Γ_{in}^* that are common to all Γ_{in}^* circles. Thus the Γ_{in}^* circles have two invariant points. Since Γ_{L1} and Γ_{L2} also lie on the load stability circle that makes $|\Gamma_{in}| = 1$, the two points, which we will label as $\Gamma_{in,1}$ and $\Gamma_{in,2}$, must lie on the Smith chart boundary. Thus the two invariant points for the Γ_{in}^* circles lie on the Smith chart boundary. It remains to be shown that these points coincide with the two points at which the source stability circle intersects the Smith chart boundary.

The source stability circle is the circle of Γ_s values that make $|\Gamma_{out}| = 1$, where

$$\Gamma_{out} = \frac{\Delta\Gamma_s - S_{22}}{S_{11}\Gamma_s - 1} \tag{10.61}$$

We can rewrite (10.60) in the following form:

$$\frac{1}{\Gamma_L} = \frac{\Delta\dfrac{1}{\Gamma_{in}} - S_{22}}{S_{11}\dfrac{1}{\Gamma_{in}} - 1} \tag{10.62}$$

which is a bilinear transformation of the same form as in (10.61). Let us choose $\Gamma_{s1} = \Gamma_{in,1}^*$. Since $\Gamma_{in,1}\Gamma_{in,1}^* = 1$ we have $\Gamma_{s1} = 1/\Gamma_{in,1}$. From (10.62),

$1/\Gamma_{\text{in},1}$ maps into the point $1/\Gamma_{L1}$ for which $|\Gamma_{L1}| = 1$. The bilinear transformation in (10.61) is the same, so it follows that $\Gamma_{s1} = 1/\Gamma_{\text{in},1}$ maps into a point $\Gamma_{\text{out},1}$ for which $|\Gamma_{\text{out},1}| = 1$. Thus Γ_{s1} is a point on the source stability circle. For the same reasons Γ_{s2} is a point on the source stability circle. Consequently, $\Gamma^*_{\text{in},1} = \Gamma_{s1}$ and $\Gamma^*_{\text{in},2} = \Gamma_{s2}$ must be on the source stability circle. Hence the invariant points for the Γ^*_{in} circles coincide with the two points at which the source stability circle intersects the Smith chart boundary.

Constant Input Mismatch Circle

If we want to design an amplifier with a specified input VSWR, then, if the load reflection coefficient Γ_L has been chosen, there will be a circle of Γ_s values that can be used which will produce the specified input VSWR. Let the required input VSWR be VSWR_1. The reflection-coefficient magnitude is given by

$$\rho = \frac{\text{VSWR}_1 - 1}{\text{VSWR}_1 + 1}$$

and the input impedance mismatch M_1 is given by $M_1 = 1 - \rho^2$. From (10.12*a*) the input mismatch is given by

$$M_1 = \frac{4R_sR_{\text{in}}}{|Z_s + Z_{\text{in}}|^2} = \frac{(1 - |\Gamma_{\text{in}}|^2)(1 - |\Gamma_s|^2)}{|1 - \Gamma_{\text{in}}\Gamma_s|^2} \tag{10.63}$$

Let Γ'_L be the chosen value of Γ_L. From (10.60) we can calculate the corresponding value of Γ_{in}, say Γ'_{in}. By using this particular value of Γ_{in} in (10.63), we can express (10.63) in the form of an equation for a circle in the Γ_s plane. The center of this circle is located at

$$\Gamma_{sM} = \frac{M_1(\Gamma'_{\text{in}})^*}{1 - (1 - M_1)|\Gamma'_{\text{in}}|^2} \tag{10.64a}$$

and the radius of the circle is given by

$$R_{sM} = \frac{\sqrt{1 - M_1}\,(1 - |\Gamma'_{\text{in}}|^2)}{1 - (1 - M_1)|\Gamma'_{\text{in}}|^2} \tag{10.64b}$$

The following example will illustrate the application of the constant input mismatch circle in amplifier design.

Example 10.4 Application of constant input mismatch circle in amplifier design. An FET has the following parameters:

$$S_{11} = 0.8\angle -140^\circ \qquad S_{12} = 0.2\angle 30^\circ \qquad S_{21} = 2.8\angle 60^\circ$$
$$S_{22} = 0.2\angle 150^\circ \qquad \Gamma_m = 0.7\angle 100^\circ \qquad R_N = 0.4$$
$$F_m = 1.5\ (1.76\ \text{dB})$$

We want to design an amplifier with low noise and an input VSWR equal to

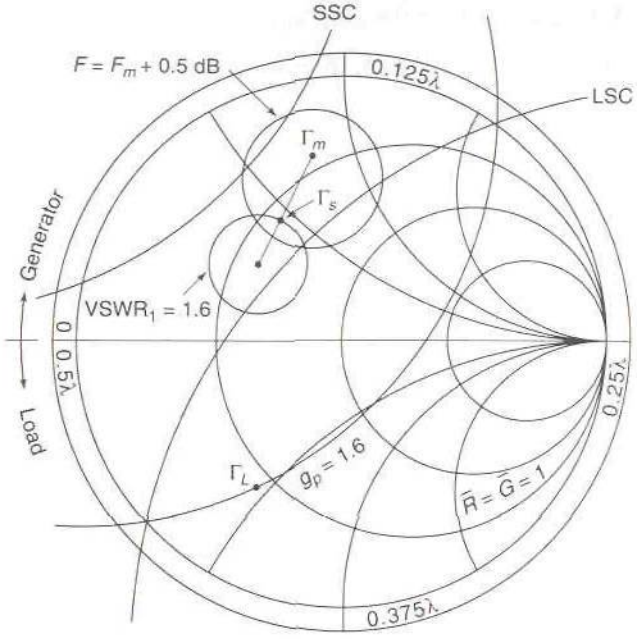


FIGURE 10.29
Constant gain, constant noise-figure, and constant input impedance-mismatch circles used for the amplifier design in Example 10.4.

1.6. The required value of the input impedance mismatch is found to be $M_1 = 0.94674$.

In Fig. 10.29 we have plotted the $g_p = 1.6$ constant normalized power-gain circle, the load stability circle, the source stability circle, and the optimum value Γ_m for Γ_s to obtain a noise figure equal to the minimum value F_m. The $F = 1.683$ constant noise-figure circle is also plotted. This noise figure is 0.5 dB greater than F_m. For this example we have chosen $\Gamma_L = 0.307 - j0.55$, which is a point on the $g_p = 1.6$ constant gain circle and not too close to the load stability circle. For this choice of $\Gamma_L = \Gamma_L'$, the input reflection coefficient $\Gamma_{in}' = -0.3302 - j0.3062$. The input mismatch circle that will give $VSWR_1 = 1.6$ has a center and radius given by

$$\Gamma_{sM} = -0.316 + j0.293 \qquad R_{sM} = 0.186$$

and is also plotted in Fig. 10.29. All values of Γ_s on this circle will give an input VSWR equal to 1.6. In order to obtain the best possible noise figure, we should choose Γ_s to lie on this circle and on the line that joins the center of this circle to the optimum point Γ_m. The best choice for Γ_s is shown in Fig. 10.29 and lies inside the $F = F_m + 0.5$ dB constant noise-figure circle, so that the input VSWR requirement can be met with a noise figure somewhat less than 1.683. The power gain obtained is $1.6|S_{21}|^2 = 12.54$, which is quite close to the "Figure of Merit" gain of 14.

Output Impedance-Mismatch Circle

If it is required to design an amplifier with a specified output VSWR_2, then, for each chosen value of $\Gamma_s = \Gamma_s'$, we can plot a circle of Γ_L values that will ensure that the specified output VSWR is obtained. The equations for the output impedance-mismatch circle are of the same form as (10.64). The center of the circle is located at

$$\Gamma_{LM} = \frac{M_2(\Gamma_{out}')^*}{1 - (1 - M_2)|\Gamma_{out}'|^2} \tag{10.65a}$$

and the radius of the circle is given by

$$R_{LM} = \frac{\sqrt{1 - M_2}\,(1 - |\Gamma_{out}'|^2)}{1 - (1 - M_2)|\Gamma_{out}'|^2} \tag{10.65b}$$

where

$$M_2 = 1 - \left(\frac{\text{VSWR}_2 - 1}{\text{VSWR}_2 + 1}\right)^2$$

$$\Gamma_{out}' = \frac{\Delta\Gamma_s' - S_{22}}{S_{11}\Gamma_s' - 1}$$

In the design of a two-stage amplifier, the design of stage 1 leads to a specified output impedance mismatch for stage 1. Consequently, the design of stage 2 is constrained by the requirement that the input mismatch to stage 2 be equal to the output mismatch of stage 1, since the impedance mismatch is conserved in the lossless matching network that is used to couple the first and second amplifier stages. For this reason the constant impedance-mismatch circles described above are useful aids in the design of a two-stage amplifier. The application of the constant impedance-mismatch circles in two-stage amplifier design is described more completely in the next section.

10.11 MICROWAVE AMPLIFIER DESIGN

In this section we will present a design strategy for designing narrowband one- and two-stage amplifiers. The first stage can be designed for a low noise figure. The method to be described can be used with both stable and potentially unstable transistors. It is assumed that the scattering-matrix parameters S_{ij}, the optimum source reflection coefficient Γ_m for minimum noise, the minimum noise figure F_m, and the normalized noise resistance R_N or noise conductance G_l are all known. The design specifications are assumed to be a power gain greater than some minimum value, a noise figure no greater than a specified maximum value, and input and output VSWRs that do not exceed specified maximum values.

There is no unique design for an amplifier that meets the design specifications. Also, there is no unique method for carrying out the design. The method described in this section works quite well for achieving a satisfactory design, but many other systematic approaches could also be developed. In general, we have to examine a range of possible load and source reflection coefficients in order to obtain an optimum design. It would be very tedious to carry out the required optimization procedures using hand calculations. Consequently, in practice a suitable computer program is used. The design strategy that is described in this section has been implemented as the computer program MICROAMP.

Single-Stage Amplifier Design

The first stage of a multistage amplifier or a single-stage amplifier is normally designed for a minimum noise figure, maximum power gain, and a chosen maximum input and output VSWR. In a multistage amplifier the output VSWR of the first stage is usually not a critical parameter. When the constraint on the output VSWR is relaxed, there is a greater degree of freedom available that makes it easier to achieve the other design requirements. The design of the second stage is also easier to carry out when the first-stage output VSWR is relatively large. We will describe the design of the first stage as a series of steps or procedures to be carried out.

1. The first step is to evaluate the stability parameter K given by (10.25a), i.e.,

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|}$$

and also to check if (10.25b) to (10.25f) are satisfied. When these conditions hold and $K > 1$, the transistor is absolutely stable and steps 2 to 4 should be followed. If $K < 1$ the device is potentially unstable and steps 5 and 6 should be followed.

2. For $K > 1$ conjugate impedance matching can be used. The required values for the source and load reflection coefficients are given by (10.17). Let Γ_{s1} be the solution for Γ_s given by (10.17a). From this value of Γ_s the noise figure F can be calculated by using (10.56). The power gain with conjugate impedance matching is given by

$$G_p = G_{p,\max} = \left(K - \sqrt{K^2 - 1}\right)\left|\frac{S_{21}}{S_{12}}\right|$$

If the noise figure is acceptable, then the design is finished except for the design of the input and output matching networks. In practice, it usually turns out that the noise figure obtained using conjugate impedance matching is not satisfactory. In order to obtain a better noise figure, it will be

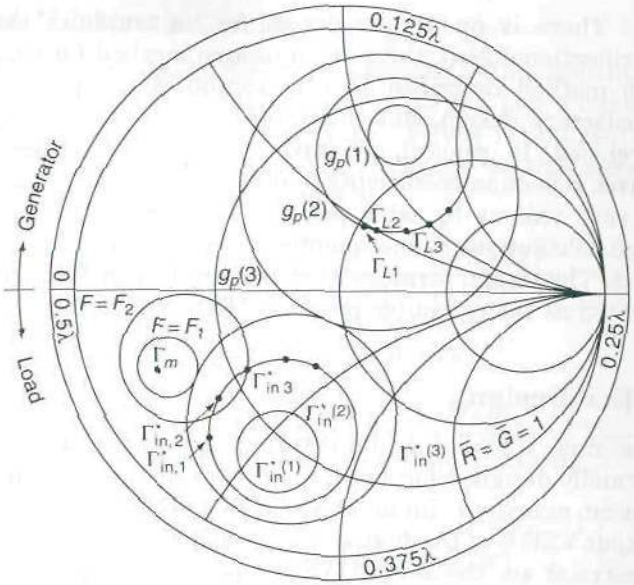


FIGURE 10.30
Constant gain, Γ^*_{in}, and constant noise-figure circles used for amplifier design.

necessary to design for a lower gain and some mismatch at the input and output ports, as described in the following steps.

3. In order to visualize the design procedure, we have plotted three constant normalized power-gain circles, the corresponding three Γ^*_{in} circles, and two constant noise-figure circles for a hypothetical device in Fig. 10.30.

We first try to obtain a satisfactory design using $\Gamma_s = \Gamma_m$ for minimum noise and a chosen normalized power gain. For example, we can choose $g_p = g_p(2)$ in Fig. 10.30. We now construct an objective function

$$\text{OF} = W_1 M_1 + W_2 M_2$$

where M_1 is the input mismatch and M_2 is the output mismatch and W_1 and W_2 are weights that can be chosen to place different levels of importance on the input and output VSWRs. A value of Γ_L, say Γ_{L1}, is chosen on *the $g_p(2)$ constant gain circle. From this value of Γ_L we can calculate* Γ_{in} using

$$\Gamma_{in} = \frac{\Delta\Gamma_{L1} - S_{11}}{S_{22}\Gamma_{L1} - 1}$$

We can also calculate Γ_{out} using $\Gamma_s = \Gamma_m$; thus

$$\Gamma_{out} = \frac{\Delta\Gamma_m - S_{22}}{S_{11}\Gamma_m - 1}$$

From these we calculate M_1 and M_2 using

$$M_1 = \frac{\left(1 - |\Gamma_{\text{in}}|^2\right)\left(1 - |\Gamma_m|^2\right)}{|1 - \Gamma_{\text{in}}\Gamma_m|^2}$$

$$M_2 = \frac{\left(1 - |\Gamma_{\text{out}}|^2\right)\left(1 - |\Gamma_{L1}|^2\right)}{|1 - \Gamma_{\text{out}}\Gamma_{L1}|^2}$$

The objective function for this value of Γ_L is now evaluated. The best input and output VSWRs are obtained by maximizing the objective function. Thus we search the $g_p(2)$ circle for the value of Γ_L that maximizes OF. This search is carried out by incrementing Γ_L in the direction that increases OF, that is, Γ_L is set equal to $\Gamma_{L2}, \Gamma_{L3}, \ldots$, as shown in Fig. 10.30. If satisfactory values for the input VSWR_1 given by

$$\text{VSWR}_1 = \frac{1 + \sqrt{1 - M_1}}{1 - \sqrt{1 - M_1}}$$

and the output VSWR_2 given by

$$\text{VSWR}_2 = \frac{1 + \sqrt{1 - M_2}}{1 - \sqrt{1 - M_2}}$$

are obtained, then the design process is terminated.

If we obtain a good value for the output VSWR_2 but an unacceptable value for the input VSWR_1, then we can set the weight $W_2 = 0$ and $W_1 = 1$ so as to place all of the emphasis on achieving a good input VSWR_1. With all of the emphasis placed on M_1, the optimum value of Γ_L will be that value which produces a Γ_{in}^* that is as close as possible to Γ_m, since this produces the best input impedance match. However, M_1 and M_2 have the same dependence on Γ_L so the largest values of M_1 and M_2 occur for the same value of Γ_L, so changing the weights will not improve the results. It will then be necessary to search for an optimum value of Γ_L on a lower constant gain curve, say $g_p(3)$. The corresponding Γ_{in}^* circle is larger; so clearly Γ_{in}^* can be brought closer to Γ_m, thereby improving the input VSWR_1.

The above process is repeated until the lowest acceptable constant gain circle has been searched. If this does not result in acceptable input and output VSWRs, the only alternative left is to accept a noise figure greater than F_m. In this case it is helpful to compile a table of best possible input VSWR_1 values for a given noise figure and normalized power gain, as described in step 4.

4. Figure 10.31 shows the $F = F_1$ constant noise-figure circle and the $\Gamma_{\text{in}}^*(2)$ circle for $g_p = g_p(2)$. If we are designing the amplifier for this noise figure and power gain, then the optimum choice for Γ_s and Γ_{in}^* that will result in the best input VSWR_1 is the value of Γ_s on the $F = F_1$ circle and the value of Γ_{in}^* on the $\Gamma_{\text{in}}^*(2)$ circle that are as close together as possible.

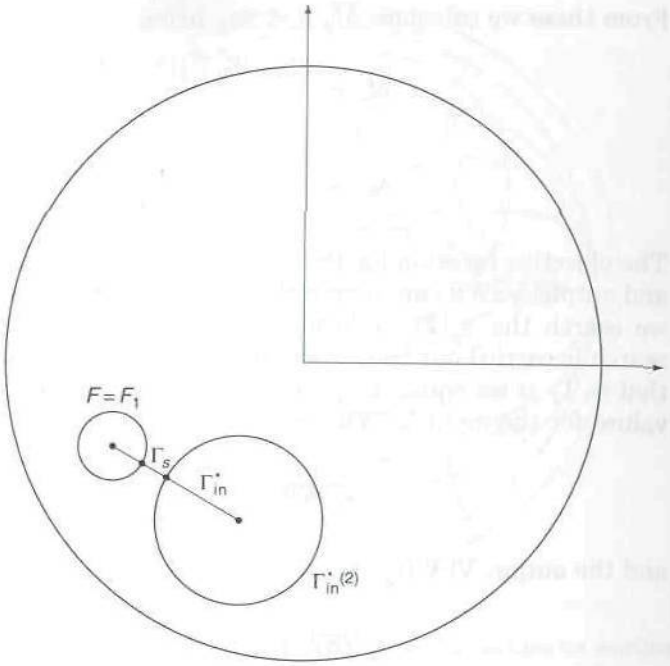


FIGURE 10.31
A constant noise-figure circle and a Γ_{in}^* circle and the optimum choice for Γ_s and Γ_{in}^* that will give the lowest input VSWR.

These points lie on the line that joins the center of the $F = F_1$ circle to the center of the $\Gamma_{in}^*(2)$ circle as shown in Fig. 10.31.

The center for the noise-figure circle is Γ_{sf} given by (10.58*a*) and the radius R_f of the circle is given by (10.58*b*). The center $\Gamma_{in,c}^*$ and radius R_{in} for the Γ_{in}^* circle are given by (10.59). The vector from the center of the $F = F_1$ circle to the center of the $\Gamma_{in}^*(2)$ circle is given by

$$\mathbf{r} = \Gamma_{in,c}^* - \Gamma_{sf}$$

A unit vector pointing from Γ_{sf} to $\Gamma_{in,c}^*$ is $\mathbf{r}/|\mathbf{r}|$. The point Γ_s lies a distance R_f from Γ_{sf} and in the direction of $\mathbf{r}$; hence

$$\Gamma_s = \frac{\mathbf{r}}{|\mathbf{r}|}R_f = \frac{\Gamma_{in,c}^* - \Gamma_{sf}}{|\Gamma_{in,c}^* - \Gamma_{sf}|}R_f$$

The optimum value of Γ_{in}^* is similarly given by

$$\Gamma_{in}^* = -\frac{\mathbf{r}}{|\mathbf{r}|}R_{in} = \frac{\Gamma_{sf} - \Gamma_{in,c}^*}{|\Gamma_{sf} - \Gamma_{in,c}^*|}R_{in}$$

Thus we can calculate the optimum choice of Γ_s and Γ_{in}^* and from these

evaluate the input mismatch M_1 and VSWR_1. The value of Γ_L that produces Γ_{in}^* is given by (10.60). The above calculation can be repeated for various choices of g_p and F and allows us to compile a table of best VSWR_1 values as a function of g_p and noise figure F. By consulting such a table we can easily see the tradeoffs between power gain, noise figure, and input VSWR_1. We now choose Γ_s for the lowest noise figure consistent with the lowest value of acceptable power gain and the largest acceptable value of input VSWR_1. The next step is to repeat the optimization procedure described in step 3 by searching the identified g_p = constant gain circle so as to optimize the objective function. The maximum value of the objective function might not correspond to the optimum choice for Γ_{in}^* that maximizes M_1. Also, the output VSWR_2 might be higher than specified. If this is the case, then either the design requirements have to be relaxed or a different transistor must be used.

5. If the transistor is potentially unstable, then we cannot carry out a design with conjugate impedance matching. For this case a design for minimum noise using $\Gamma_s = \Gamma_m$ should be explored first. This requires that we determine the source stability circle and check that Γ_m lies in a stable region of the Smith chart and not too close to the boundary of the source stability circle. When Γ_m is a stable value, the design procedure is the same as described in step 3, i.e., a chosen g_p = constant gain circle is searched for the best value of Γ_L that will maximize the objective function OF given in step 3. The "Figure of Merit" gain is $|S_{21}/S_{12}|$ with a normalized value $|S_{12}S_{21}|^{-1}$. It is good practice to design an amplifier for a normalized power gain that does not exceed this value. Thus constant power-gain circles with $g_p = |S_{21}S_{12}|^{-1}$ and smaller are searched. If acceptable values of g_p, VSWR_1, and VSWR_2 are obtained, then the design process is terminated. The resultant amplifier will have a minimum noise figure F_m. If a satisfactory design cannot be obtained using $\Gamma_s = \Gamma_m$ or if Γ_m is an unstable value, then the procedure outlined in step 6 should be followed. If Γ_m is an unstable value, it would be advisable to use another transistor.

6. In order to minimize the amount of searching for the optimum value of Γ_L subject to the constraints on g_p, F, VSWR_1, and VSWR_2, it is again helpful to compile a table of best VSWR_1 values as a function of g_p and F. By consulting such a table we can determine if the design objectives can be met and/or the best power gain, noise figure, and input VSWR_1 that can be obtained using the chosen transistor. When the Γ_{in}^* values inside the Smith chart represent stable values of Γ_s, then the input VSWR_1 for the optimum choice of Γ_{in}^* and Γ_s on a chosen F = constant circle are calculated the same way as for a stable device. The procedure is described in step 4. The optimization of the input VSWR_1 and output VSWR_2 for the chosen normalized power gain and noise figure can be carried out in the same way as described in step 4.

A bipolar transistor used in a common base configuration or an FET used in a common gate configuration often have $|S_{11}| > 1$ and $|S_{22}| > 1$.

When this latter set of conditions holds true, it is not possible to design an amplifier with a low input or output VSWR. The reason is that stable values of Γ_L produce values of Γ_{in}^* that lie in the unstable region of the Γ_s plane; so we cannot choose Γ_s equal to Γ_{in}^*. Similarly, stable values of Γ_s produce Γ_{out}^* values that lie in the unstable region of the Γ_L plane so the output port cannot be matched. Since the common base and common gate connections usually have poorer noise performance as well, the common base or common gate circuits are not used in low noise amplifiers. In Example 10.5 we illustrate the impossibility of matching the input port of a common base amplifier.

Example 10.5 Common base amplifier. A bipolar transistor in the common base connection has the following parameters at 5 GHz:

$$S_{11} = 1.3\angle 140^\circ \qquad S_{12} = 0.2\angle 130^\circ \qquad S_{21} = 2\angle -85^\circ$$

$$S_{22} = 1.15\angle -50^\circ \qquad \Gamma_m = 0.7\angle 135^\circ \qquad F_m = 2.5 \qquad R_N = 0.4$$

For this transistor $K = -0.579$ so it is potentially unstable. In Fig. 10.32 we have plotted the load stability circle, the source stability circle, the $g_p = 1.5$

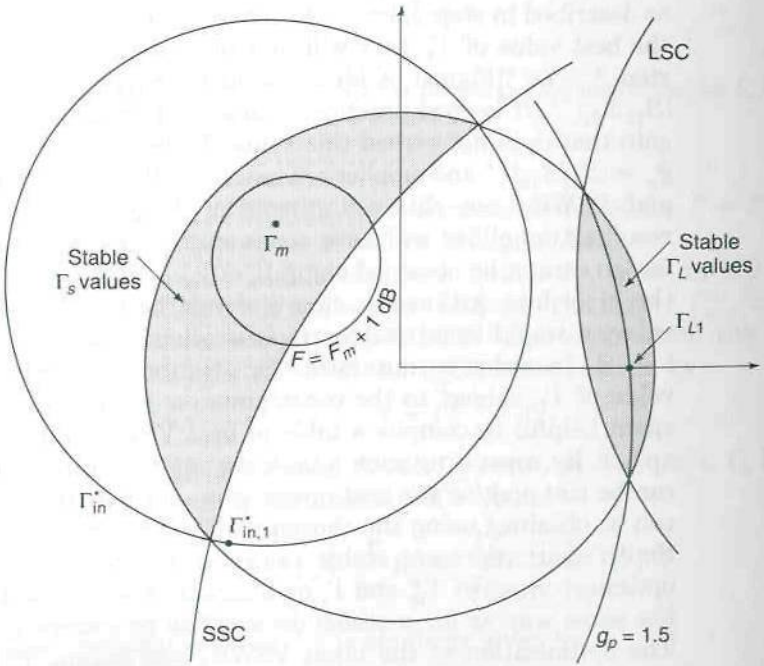


FIGURE 10.32
Load stability circle, source stability circle, $g_p = 1.5$ circle and corresponding Γ_{in}^* circle, and $F = F_m + 1$ dB constant noise-figure circle for the bipolar transistor, in a common base configuration as used in Example 10.5.

circle and corresponding Γ_{in}^* circle, and the $F = F_m + 1$ dB = 3.147 circle. Since $|S_{11}| > 1$ and $|S_{22}| > 1$, the origin is an unstable point for both Γ_s and Γ_L. The regions of the Smith chart where stable values of Γ_s and Γ_L are located are shown cross-hatched. The point Γ_{L1} on the $g_p = 1.5$ circle maps into the point $\Gamma_{in,1}^*$ which lies in the unstable part of the Γ_s plane. All values of Γ_L in the stable region map into Γ_{in}^* values that lie in the unstable part of the Γ_s plane. Hence we cannot choose $\Gamma_s = \Gamma_{in}^*$ and consequently the input port cannot be matched.

If we try to design a low-noise amplifier by choosing $\Gamma_s = \Gamma_m$ and with a normalized power gain of 2, the resultant input $\text{VSWR}_1 = 15.42$ and the output $\text{VSWR}_2 = 19.03$. If we reduce the gain requirement G_p to 5, for which $g_p = 1.25$, we obtain $\text{VSWR}_1 = 12.76$ and $\text{VSWR}_2 = 26.56$ for a design with minimum noise. There is a small improvement in the input VSWR_1 but the output VSWR_2 is increased.

From Fig. 10.32 it is quite clear that any value of Γ_s on the $F = F_m + 1$ dB noise-figure circle and in the stable region will still be far away from all Γ_{in}^* values inside the Smith chart. Thus a low-noise amplifier with acceptable gain and input and output VSWRs cannot be designed using the above transistor in a common base circuit.

Example 10.6 Low-noise amplifier design. A GaAs FET has the following parameters at 10 GHz:

$$S_{11} = 0.73\angle 42^\circ \qquad S_{12} = 0.2\angle -58^\circ \qquad S_{21} = 1.52\angle -66^\circ$$

$$S_{22} = 0.5\angle 34^\circ \qquad \Gamma_m = 0.52\angle -70^\circ \qquad F_m = 1.25 \qquad R_N = 0.75$$

We want to use this device in a low-noise amplifier design which meets the following specifications:

Noise figure $F \leq 1.5$
Input $\text{VSWR}_1 \leq 1.5$
Output $\text{VSWR}_2 \leq 1.5$
Power gain G_p as large as possible

For this device $K = 1.071$ so that the FET is absolutely stable. For a conjugate-impedance-matched design, we get $G_p = G_{p,\max} = 5.222$, $\text{VSWR}_1 = \text{VSWR}_2 = 1$, and $F = 1.7$. The noise figure is too large; so we must consider a design that is not matched at the input and output. The power gain which can be obtained is no greater than 5.222; so we do not want to sacrifice much power gain for an improved noise figure. The best VSWR_1 that can be achieved using $g_p = 2.25$, which gives $G_p = 5.1984$, for $F = F_m + 0.5$ dB is 1.3644 and for $F = F_m + 1$ dB it is 1.026. This shows that the design objectives can be met by allowing a noise figure equal to $F_m + 0.5$ dB = 1.4. By searching the $g_p = 2.25$ circle for the optimum value of Γ_L and using the optimum value of Γ_s, which was determined so as to lie on the $F = F_m + 0.5$ dB noise circle and give $\text{VSWR}_1 = 1.3644$, we obtain $G_p = 5.1984$, $\text{VSWR}_1 = 1.3643$, $\text{VSWR}_2 = 1.1179$, and $F = 1.4$. The required values of Γ_s and Γ_L are $\Gamma_s = 0.389 - j0.534$ and $\Gamma_L = -0.0028 - j0.323$. This design meets all of the specifications.

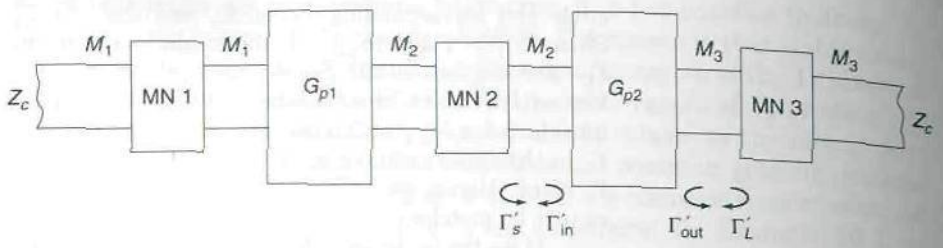


FIGURE 10.33
A block diagram of a two-stage amplifier.

Design of Second Stage for a Two-Stage Amplifier

The design specifications for the second stage of a two-stage amplifier emphasizes power gain and output VSWR. The noise figure of stage 2 is not very critical since the noise contribution of the second stage is reduced by the power gain of the first stage as shown by (10.54). However, at the higher microwave frequencies, the power gain of the first stage is often not very large, so that some consideration of the noise figure of stage 2 is necessary. If we assume that the matching network used between the output of stage 1 and the input to stage 2 is a lossless network, then, since the impedance mismatch is constant throughout a chain of lossless networks, the impedance mismatch at the input to stage 2 is the same as that at the output of stage 1. This places a constraint on the design of stage 2, namely, that the input impedance mismatch must equal M_2, where M_2 is the output mismatch for stage 1 and was determined in the design of stage 1. In order that the interstage matching network be physically realizable, this constraint cannot be violated.

In Fig. 10.33 we show a block diagram for a two-stage amplifier. The source reflection coefficient, input reflection coefficient, output reflection coefficient, and load reflection coefficient for stage 2 are identified by a superscript prime. The output impedance mismatch of stage 2 is M_3 and the corresponding output VSWR will be called VSWR_3. It is assumed that the same type of transistor that was used in stage 1 is also used in stage 2.

The design of stage 2 will be based on the optimization of the noise figure and output VSWR or mismatch M_3 for a chosen power gain g_p. This optimization is carried out subject to the constraint that the input impedance mismatch equals M_2. The following objective function is used for the design of stage 2:

$$\text{OF} = W_1 M_3 + W_2 \frac{F_m}{F}$$

where W_1 and W_2 are suitable weights. We choose F_m/F for the second term, since this quantity is of the same order of magnitude as M_3 and it becomes larger for smaller values of F. Thus our goal is to find the best

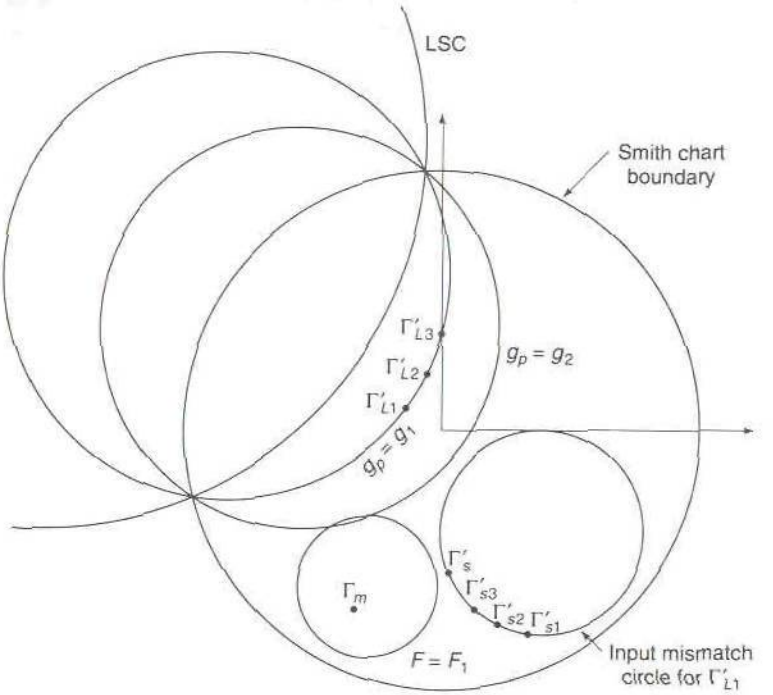


FIGURE 10.34
A constant power-gain circle and a constant input impedance-mismatch circle used in the design of the second stage of a two-stage amplifier.

value of Γ_L on a chosen g_p = constant gain circle so as to maximize OF. The design procedure is the same for both stable and unstable devices.

A visualization of the optimization process can be obtained by referring to Fig. 10.34. In this figure we show the load stability circle, two constant power-gain circles, and a constant noise-figure circle for a hypothetical transistor. Let us assume that we will choose $g_p = g_1$. We then pick an initial value of Γ'_L, say Γ'_{L1}, on the g_1 circle. From this value of Γ'_L we can calculate Γ'_{in} using

$$\Gamma'_{\text{in}} = \frac{\Delta\Gamma'_{L1} - S_{11}}{S_{22}\Gamma'_{L1} - 1}$$

The input mismatch M_2 is given by

$$M_2 = \frac{\left(1 - |\Gamma'_{\text{in}}|^2\right)\left(1 - |\Gamma'_s|^2\right)}{|1 - \Gamma'_{\text{in}}\Gamma'_s|^2}$$

Since M_2 is fixed by the stage 1 design, this equation determines a circle of

Γ'_s values that can be used. The center and radius of this input mismatch circle are given by (10.64) and are

$$\Gamma'_{sM} = \frac{M_2(\Gamma'_{\text{in}})^*}{1 - (1 - M_2)|\Gamma'_{\text{in}}|^2}$$

$$R'_{sM} = \frac{\sqrt{1 - M_2}\left(1 - |\Gamma'_{\text{in}}|^2\right)}{1 - (1 - M_2)|\Gamma'_{\text{in}}|^2}$$

Each value of Γ'_s such as $\Gamma'_{s1}, \Gamma'_{s2}$, etc., shown in Fig. 10.34 enables us to calculate a corresponding noise figure using (10.56). Also for each Γ'_{si} we can calculate an output reflection coefficient $\Gamma'_{\text{out},i}$. From $\Gamma'_{\text{out},i}$ and the chosen value Γ'_{L1} for Γ'_L, we can evaluate the output mismatch M_3. Thus we can evaluate the objective function. Our procedure is now to search the input mismatch circle for the value of Γ'_s that maximizes the objective function. The maximum value is recorded. We now increment Γ'_L to a new value Γ'_{L2}. This results in a new input mismatch circle which is searched for the value of Γ'_s that maximizes OF. This value of OF is compared with the previous one, and if it is greater then Γ'_L is incremented to a new value Γ'_{L3} in the same direction on the $g_p = g_1$ circle. If the second value of OF is smaller than the first one, then Γ'_L is incremented in the opposite direction. The search process is continued until the optimum values of Γ'_L and Γ'_s are found. The optimization routine would be very time consuming to carry out with hand calculations but can be done very quickly on a computer. The optimization carried out on the input mismatch circle will result in a choice for Γ'_s that lies as close as possible to the optimum point Γ_m if the weight W_1 is set equal to zero (see Fig. 10.34). When W_1 is not zero, Γ'_s will generally deviate from this point in order to get a lower output VSWR_3.

If satisfactory values of noise figure and output VSWR_3 are not obtained from a value of Γ'_L on the chosen g_p = constant circle, then the procedure outlined above must be repeated on a constant gain circle having a lower gain. If, after searching the lowest acceptable power gain circle, an acceptable design is not obtained, then the design specifications will have to be relaxed, a different transistor used, or a third stage added. A third stage can be designed using the same approach as used for the second stage.

Some of the Γ'_s values on an input mismatch circle may lie in an unstable region of the Smith chart. If Γ'_{sj} is an unstable point, then it produces an output reflection coefficient $\Gamma'_{\text{out},j}$ with a magnitude greater than unity. When $|\Gamma'_{\text{out},j}| > 1$ the output mismatch M_3 and output VSWR_3 will be negative. Consequently, M_3 will contribute a negative quantity to the objective function. Since the objective function is being maximized, unstable values of Γ'_s are not selected since they tend to minimize the objective function.

When the design of the second stage has been completed, the two-stage amplifier will have a power gain

$$G_p = G_{p1}G_{p2}$$

and a noise figure

$$F = F_1 + (F_2 - 1)\frac{M_2}{M_1 G_{p1}}$$

The interstage matching network must transform the output admittance Y_{out} of stage 1 into the required source admittance Y_s' for stage 2. At the same time it must transform the input admittance Y_{in}' of stage 2 into the required load admittance Y_L for stage 1. This matching network is physically realizable because of the constraint that was placed on the stage 2 input impedance mismatch when it was designed. The design of an interstage matching network is described in Sec. 5.7.

Example 10.7 Two-stage amplifier design. At 6 GHz an FET has the following parameters:

$S_{11} = 0.8\angle -130°$ $S_{12} = 0.2\angle 30°$ $S_{21} = 3\angle 60°$

$S_{22} = 0.3\angle 140°$ $\Gamma_m = 0.6\angle 160°$ $F_m = 1.5$ $R_N = 0.6$

The design specifications are:

Two-stage power gain $G_p \geq 120$
Total noise figure $F \leq 1.8$
Input $\text{VSWR}_1 \leq 1.5$
Output $\text{VSWR}_3 \leq 1.5$

For this device $K = 0.513$, so that the transistor is not absolutely stable. The "Figure of Merit" gain equals $|S_{21}/S_{12}| = 15$. The normalized value is $g_p = 1.666$.

For the design of the first stage, the following table of minimum VSWR_1 values as a function of normalized power gain and noise figure was compiled:

g_p	F	VSWR_1	Γ_s
1.6	F_m	2.39	$-0.564 + j0.205$
1.4	F_m	2.084	$-0.564 + j0.205$
1.2	F_m	1.809	$-0.564 + j0.205$
1.6	1.68	1.71	$-0.581 + j0.304$
1.4	1.68	1.49	$-0.576 + j0.306$
1.2	1.68	1.294	$-0.569 + j0.307$

From the above table we see that we can obtain a noise figure of 1.68, a normalized power gain of 1.4, and an input $\text{VSWR}_1 = 1.49$ or, with a reduced normalized gain of 1.2, an input VSWR_1 of 1.294. We will choose a power gain $g_p = 1.2$ and a value of Γ_s midway between those giving VSWR_1 equal to 1.294

and 1.809, so as to obtain a noise figure somewhat smaller than 1.68. Thus we choose $\Gamma_s = -0.568 + j0.27$. We now use this value and search the $g_p = 1.2$ gain circle for the value of Γ_L that will give the smallest values for VSWR_1 and VSWR_2. The result is $\Gamma_L = 0.458 + j0.14$ and an amplifier having a noise figure 1.572 and an input $\text{VSWR}_1 = 1.484$ and a power gain $G_{p1} = 10.8$. The output $\text{VSWR}_2 = 6.5015$ which gives $M_2 = 0.4621$.

The design of stage 2 is carried out using the procedure described earlier. The objective function $W_1 M_3 + W_2 F_m / F$ is optimized by searching for the value of Γ'_L on a chosen g_p = constant gain circle and the value of Γ'_s on the associated input impedance-mismatch circle using M_2 determined above. The search on the $g_p = 1.6$ circle resulted in a design with $\text{VSWR}_3 = 1.65$ and $F_2 = 3.693$ using $W_1 = 1$, $W_2 = 0$. For this example a search on lower gain circles gave higher values of VSWR_3. If we relax the design specifications to allow the somewhat larger output VSWR_3 value, then for the two-stage amplifier the following performance is obtained:

$$G_p = 10.8 \times 14.4 = 155.5 \qquad \text{VSWR}_1 = 1.484 \qquad \text{VSWR}_3 = 1.65$$

$$F = F_1 + M_2 \frac{F_2 - 1}{M_1 G_{p1}} = 1.572 + 0.4621 \times \frac{2.693}{0.962 \times 10.8} = 1.692$$

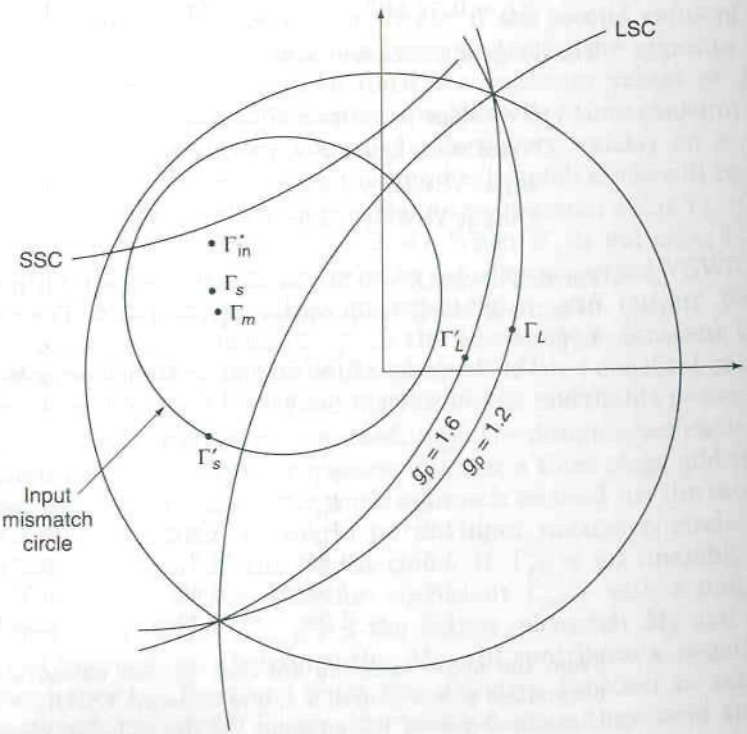


FIGURE 10.35
Illustration for the two-stage amplifier designed in Example 10.7.

For all specifications except the output $VSWR_3$, this design meets the stated criteria. For practical applications an output $VSWR_3$ equal to 1.65 instead of 1.5 is acceptable.

In Fig. 10.35 we show the load and source stability circles, the $g_p = 1.6$ and 1.2 constant gain circles, the values of Γ_s and Γ_L for stage 1, the values of Γ_s' and Γ_L' for stage 2, and the point $\Gamma_s = \Gamma_m$. Also shown is the point Γ_{in}^* for Γ_L' and the input mismatch circle for Γ_L'. Note that Γ_s' lies on this circle but is not as close as possible to the optimum point Γ_m. The reason for this is that Γ_s' was chosen to obtain the best output $VSWR_3$, not the best noise figure. The figure clearly shows that Γ_s, Γ_L, Γ_s', and Γ_L' are sufficiently far away from the stability circle boundaries so the design has an adequate stability margin.

A second design was carried out for which the input stage was designed for a power gain of 7.2 and a minimum noise figure and a resultant input $VSWR_1 = 1.358$ was obtained. For the second stage we were then able to obtain $G_{p2} = 14.4$, $VSWR_3 = 1.517$, and $F = 4.28$. For the two-stage amplifier we obtained $G_p = 103.7$, $VSWR_1 = 1.358$, $VSWR_3 = 1.517$, and $F = 1.685$. This design has an input $VSWR_1$ lower than required, a noise figure essentially the same as for the first design, an output $VSWR_3$ nearly equal to 1.5, but a significantly lower gain. The small increase in output $VSWR_3$ in the first design is only a small price to pay for the much larger power gain that was obtained, so the first design is a better one.

10.12 OTHER ASPECTS OF MICROWAVE AMPLIFIER DESIGN

Microwave amplifier design as described in the preceding section is only a small part of the overall design problem. Once it has been verified that the specified performance can be obtained, a decision has to be made as regards to whether hybrid construction or monolithic integration will be used in the fabrication of the physical amplifier. A decision of whether to use microstrip circuits or coplanar-waveguide circuits must also be made, as well as a decision of whether to use lumped elements or distributed elements for the impedance-matching networks. The layout of the circuit must be designed and the circuit must incorporate both the bias circuit and the RF elements with suitable decoupling of the RF circuit from the dc bias circuit. The physical dimensions of all transmission lines and other printed-circuit elements must be calculated and will depend on the substrate material used. Suitable masks must be prepared for use in the fabrication of the amplifier. After the amplifier has been built, it must be tested to determine if it meets the design specifications.

After the RF circuit and dc bias circuit have been designed, a theoretical evaluation or computer simulation of the amplifier should be carried out to verify that it is stable at all frequencies. This check should be performed before the amplifier construction is undertaken.

For a broadband amplifier the design of the matching networks is considerably more complex than that for narrowband amplifiers. The

matching networks must be designed so as to provide adequate stability, flat gain throughout the passband, and constant group delay. The latter requires that the phase function $\phi(\omega)$ in the overall amplifier transfer function $H(\omega) = |H(\omega)|e^{j\phi(\omega)}$ be a linear function of ω (see Sec. 3.19).

For a power amplifier, dynamic range, nonlinear distortion, and intermodulation characteristics must be taken into account. Also, suitable provision must be made to remove the heat produced in the active devices.

Companies that manufacture microwave amplifiers use a number of computer-aided design (CAD) software packages to facilitate the overall design. The effort expended in carrying out a thorough design before construction is undertaken pays large dividends since there is very little that can be changed in either the circuit component values or the circuit layout once the amplifier has been built, particularly so for monolithic microwave integrated circuits (MMICs).

Hewlett-Packard manufactures modular microcircuit packages that are very useful in the early stages of microwave amplifier development. These packages provide a ready-made miniature box with input and output miniature 3-mm coaxial-line connectors and dc bias terminals that allow for easy mounting of a prototype circuit so that it can be tested. A photograph of these package modules is shown in Fig. 10.36.

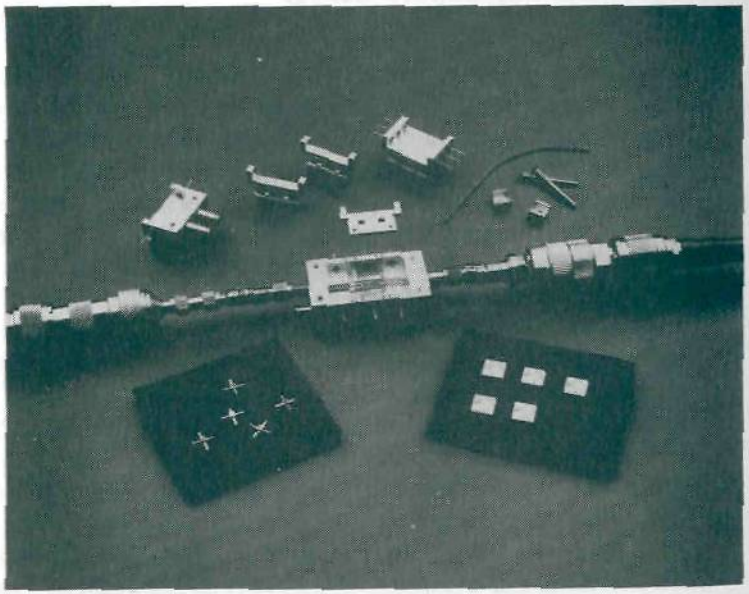


FIGURE 10.36
Modular microcircuit package for prototype circuit design and testing. (Photograph courtesy of Ray Moskaluk, Hewlett-Packard Company.)

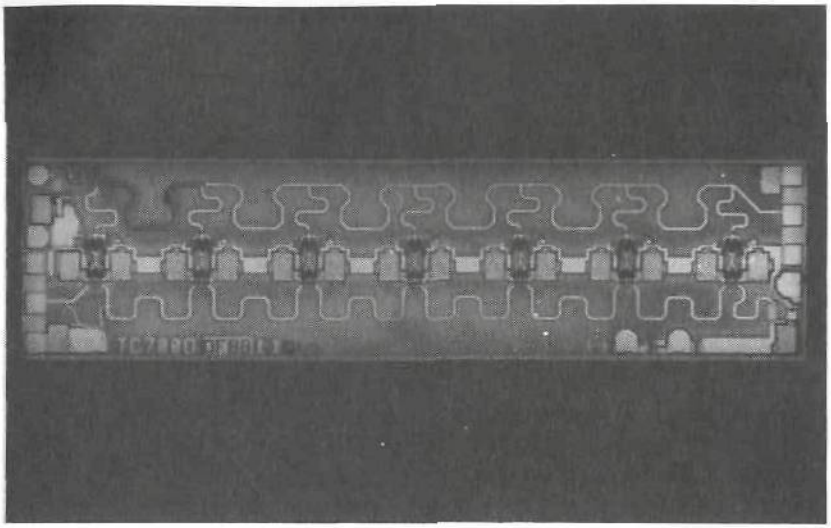


FIGURE 10.37
A broadband traveling-wave amplifier (MMIC) circuit. All components including matching networks are built on a single chip. (Photograph courtesy of Ray Moskaluk, Hewlett-Packard Company.)

A typical MMIC circuit is shown in Fig. 10.37. This is a broadband general-purpose traveling-wave amplifier. It is a GaAs MMIC chip using seven MESFET gain stages and has a flat gain of 8.5 ± 1 dB over the frequency range 2 to 26.5 GHz. A large number of gain stages is required since for any amplifier the gain-bandwidth product tends to remain constant, so that the gain per stage is necessarily low in a very broadband amplifier. The noise figure ranges from around 5 dB at the low-frequency end to 7 to 8 dB at the high-frequency end. The input and output VSWRs are less than 1.5.

PROBLEMS

10.1. A bipolar transistor has the following scattering-matrix parameters at 2 GHz:

$$S_{11} = 0.56\angle 170^\circ \qquad S_{12} = 0.06\angle 75^\circ$$

$$S_{21} = 4.04\angle 76^\circ \qquad S_{22} = 0.41\angle -23^\circ$$

Find the stability parameter K, and if $K > 1$ also find the maximum stable gain. Find the load and source stability circle parameters and plot these on a Smith chart.

10.2. A silicon bipolar transistor is used in a common base amplifier circuit at 5 GHz. Its scattering-matrix parameters are:

$$S_{11} = 1.3\angle 140^\circ \qquad S_{12} = 0.2\angle 130^\circ$$
$$S_{21} = 2\angle -85^\circ \qquad S_{22} = 1.15\angle -55^\circ$$

Find the stability parameter K, the maximum stable gain, and the load and source stability circle parameters. Plot the stability circles on a Smith chart.

10.3. A GaAs FET has the following scattering-matrix parameters at 2 GHz:

$$S_{11} = 0.91\angle -42^\circ \qquad S_{12} = 0.05\angle 33^\circ$$
$$S_{21} = 6\angle 105^\circ \qquad S_{22} = 0.62\angle -95^\circ$$

Find the stability parameter K and the load and source stability circles. Plot the stability circles and show the regions of the Smith chart where stable values of the load and source impedances lie.

10.4. A bipolar transistor is used as a common collector (source follower) amplifier at 5 GHz. Its scattering-matrix parameters are:

$$S_{11} = 0.63\angle -96^\circ \qquad S_{12} = 0.8\angle 15^\circ$$
$$S_{21} = 1.3\angle -53^\circ \qquad S_{22} = 0.62\angle 98^\circ$$

Evaluate the stability parameter K. Find and plot the stability circles and show the regions of the Smith chart where stable values of Γ_L and Γ_s occur.

Answer: $K = 0.3993$, $\Gamma_{LC} = -1.298 + j4.0718$, $R_{LC} = 4.5735$, $\Gamma_{SC} = 2.347 - j3.8792$, $R_{SC} = 4.8395$. Stable values lie inside the stability circles.

10.5. An engineer adds an external network to a bipolar transistor and finds that in a common base connection its scattering parameters at 5 GHz are

$$S_{11} = 1.3\angle 140^\circ \qquad S_{12} = 0.2\angle 130^\circ$$
$$S_{21} = 2\angle 85^\circ \qquad S_{22} = 1.15\angle -55^\circ$$

Find the stability parameter K and the stability circles. Show in what regions of the Smith chart stable values of Γ_L and Γ_s occur.

Answer: $K = 1.439$ but the device is only conditionally stable since $|S_{11}| > 1$ and $|S_{22}| > 1$; $\Gamma_{LC} = 0.174 + j0.6258$, $R_{LC} = 0.2172$, $\Gamma_{SC} = -0.2179 - j0.495$, $R_{SC} = 0.2714$. Origin is unstable.

10.6. For the amplifier discussed in Example 10.2 and using the design that requires $Z_L = 85 + j168.2$, design a matching network consisting of an open-circuited stub located a distance d from the output that will transform the 50-Ω line impedance into the required load impedance (see Sec. 5.6).

10.7. Redesign the amplifier discussed in Example 10.3 so as to get a noise figure of 2 and a normalized power gain of 2.9. Find the required values of Γ_s, Γ_L, and the resultant input and output VSWRs.

Hint: Construct the $F = 2$ noise-figure circle. The best choice for Γ_s is a point on this circle lying on the line joining the center of the $F = 2$ circle with the center of the Γ_{in}^* circle for $g_p = 2.9$. This Γ_s can be found since the centers and radii of the circles are known. The corresponding Γ_{in}^* on this line can also be found. From Γ_s calculate Γ_{out} and from Γ_{in}^* calculate Γ_L. The input and output impedance mismatches can now be found using (10.12*a*) and (10.12*b*). From these the input and output VSWRs can be found.

10.8. A MESFET has the following parameters at 8 GHz:

$$S_{11} = 0.65\angle -150^\circ \qquad S_{12} = 0.12\angle 32^\circ$$
$$S_{21} = 2.2\angle 67^\circ \qquad S_{22} = 0.1\angle 150^\circ$$
$$\Gamma_m = 0.45\angle 130^\circ \qquad R_N = 0.32 \qquad F_m = 1.3$$

Use the computer program MICROAMP to design a single-stage low-noise amplifier using the MESFET described above. The design requirements are: power gain of 9 or more, noise figure equal to 1.46 or less, input VSWR no greater than 1.5. What noise figure is obtained if conjugate impedance matching is used?

Answer: $G_p = 9.196$, $F = 1.4586$, input VSWR = 1.346, output VSWR = 1.166, $\Gamma_s = -0.5043 + j0.3715$, $\Gamma_L = -0.2188 + j0.1157$. For conjugate impedance matching $F = 1.87$.

10.9. Show that for a microwave amplifier the available power gain G_a can be expressed as $G_a = M_1 G_p / M_2$, where G_p is the power gain and M_1 and M_2 are the input and output impedance-mismatch factors. By using relations of this type, show that the noise figure for a cascade connection of amplifier stages can be expressed as

$$F = F_1 + \frac{F_2 - 1}{G_{a1}} + \frac{F_3 - 1}{G_{a1}G_{a2}} + \cdots$$

10.10. Verify that the matching networks shown in Fig. 10.28 will transform the line characteristic admittance $Y_c = 0.02$ S into the required source and load admittances needed for the amplifier discussed in Example 10.3.

10.11. Design a single-stage low-noise amplifier using an FET having the following parameters:

$$S_{11} = 0.74\angle -115^\circ \qquad S_{12} = 0.14\angle 40^\circ \qquad S_{21} = 2.7\angle 87^\circ$$
$$S_{22} = 0.13\angle -60^\circ \qquad \Gamma_m = 0.5\angle 100^\circ \qquad F_m = 1.3 \qquad R_N = 0.24$$

The design specifications are: $G_p \geq 15$, $\text{VSWR}_1 \leq 2$, $\text{VSWR}_2 \leq 2$, and $F \leq 1.5$. If you cannot meet the design specifications, relax one or more of the requirements.

Answer: A design with $G_p = 18.95$, $F = 1.33$, $\text{VSWR}_1 = 2.11$, and $\text{VSWR}_2 = 2.265$ can be achieved using $\Gamma_s = -0.17 + j0.61$.

10.12. The transistor whose parameters are given in Prob. 10.11 is to be used in a two-stage amplifier. The design calls for $G_p \geq 300$, $F \leq 1.5$, $\text{VSWR}_1 \leq 1.5$, $\text{VSWR}_3 \leq 1.5$. Plot the load and source stability circles, the $g_p = 2.6$ and 2.4 constant gain circles, and the $F = F_m + 0.5$ dB constant noise-figure circle. On this figure show your final design values for Γ_s and Γ_L for the first stage and Γ_s' and Γ_L' for the second stage. Does your final design have a good stability margin?

10.13. A transistor has the following parameters:

$$S_{11} = 0.5\angle 160^\circ \qquad S_{12} = 0.06\angle 50^\circ \qquad S_{21} = 3.6\angle 60^\circ$$
$$S_{22} = 0.5\angle -45^\circ \qquad \Gamma_m = 0.4\angle 145^\circ \qquad R_N = 0.4 \qquad F_m = 1.6$$

Design a single-stage amplifier with the best possible noise figure subject to the constraints $G_p \geq 10$, $\text{VSWR}_1 \leq 2$, $\text{VSWR}_2 \leq 2$.

10.14. A transistor with the following parameters is to be used in a two-stage amplifier:

$$S_{11} = 0.85\angle 159^\circ \qquad S_{12} = 0.06\angle -74^\circ \qquad S_{21} = 1.8\angle -34^\circ$$
$$S_{22} = 0.66\angle -162^\circ \qquad \Gamma_m = 0.8\angle -160^\circ \qquad F_m = 1.74 \qquad R_N = 0.3$$

The design objectives are: $G_p = 250$, $\text{VSWR}_1 \le 2$, $\text{VSWR}_3 \le 2$, $F \le 2.2$. For your final design construct the load and source stability circles, the $g_p = 6, 7, 8, 9$ gain circles, and the $F = F_m + 0.5$ dB and $F = F_m + 1$ dB constant noise-figure circles. On this figure show the design values of Γ_s, Γ_L and Γ_s', Γ_L' at which you arrived.

REFERENCES

1. Vendelin, G. D., A. M. Pavio, and U. L. Rohde: "Microwave Circuit Design Using Linear and Nonlinear Techniques," John Wiley & Sons, Inc., New York, 1990. This is a very comprehensive text covering small-signal and low noise amplifier design, power amplifier design, as well as oscillator and mixer design.
2. Pengelly, R. S.: "Microwave Field-Effect Transistors—Theory, Design, and Applications," Research Studies Press, Hertfordshire, England, 1986, distributed by John Wiley & Sons, Inc., New York.
3. Gonzalez, G.: "Microwave Transistor Amplifiers, Analysis and Design," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1984.
4. Vendelin, G. D.: "Design of Amplifiers and Oscillators by the *S*-Parameter Method," John Wiley & Sons, Inc., New York, 1982.
5. Ha, T. T.: "Solid-State Microwave Amplifier Design," John Wiley & Sons, Inc., New York, 1981.
6. Gentile, C.: "Microwave Amplifiers and Oscillators," McGraw-Hill Book Company, New York, 1987.
7. Carson, R.: "High Frequency Amplifiers," 2nd ed., John Wiley & Sons, Inc., New York, 1982.
8. Chang, K. (ed.): "Handbook of Microwave and Optical Components. Microwave and Solid-State Components," vol. 2, John Wiley & Sons, Inc., New York, 1990. This is a very good reference source on semiconductor theory, device design and fabrication, device modeling, and applications, for microwave transistors and diodes.

CHAPTER 11

PARAMETRIC AMPLIFIERS

A parametric amplifier is an amplifier utilizing a nonlinear reactance, or a reactance that can be varied as a function of time by applying a suitable pump signal. The time variation of a reactive parameter can be used to produce amplification. This is the origin of the term parametric amplifier. The possibility of parametric amplification of signals was shown theoretically, as long ago as 1831, by Lord Rayleigh. The first analysis of the nonlinear capacitance was given by van der Ziel in 1948.† He pointed out that this device could also be useful as a low-noise amplifier since it was essentially a reactive device in which no thermal noise is generated. The first realization of a microwave parametric amplifier was made by Weiss, following the earlier proposal (1957) by Suhl, suggesting the use of the nonlinear effect in ferrites (Sec. 6.7). In the following few years, the semiconductor-diode (sometimes called a varactor, for variable reactance) parametric amplifier was developed through the combined efforts of many workers. At the present time the semiconductor junction diode is the most widely used parametric amplifier. For this reason we limit the discussion in this text to this particular type of parametric amplifier. The p-n junction diode has a nonlinear capacitance. If a pumping signal at frequency ω_p and a small-amplitude signal at frequency ω_s are applied simultaneously, the device behaves like a time-varying linear capacitance at the signal frequency ω_s. As we show in later sections, a time-varying capacitance or a nonlinear

†A. van der Ziel, On the Mixing Properties of Nonlinear Capacitances, *J. Appl. Phys.*, vol. 19, pp. 999–1006, November, 1948.

capacitance can be incorporated into a circuit to give linear amplification of a small-amplitude signal. Before presenting this analysis a brief description of some of the properties of junction diodes is given, followed by a presentation of the Manley-Rowe relations. The latter are a set of power-conservation relations, of considerable value in determining the maximum gain and other performance features of parametric amplifiers.

11.1 *p-n* JUNCTION DIODES

The diodes used for parametric amplifiers consist of a junction of n-type and p-type semiconductor material. An n-type semiconductor has an excess supply of electrons, which is why it is called n-, or negative-, type. An example of an n-type material is a pure semiconductor such as germanium, with a small amount (about 1 part in 10^5) of impurity doping with an element such as arsenic or antimony. Germanium has a valence of 4, whereas arsenic or antimony has a valence of 5. Thus, at each site in the germanium or host crystal where an arsenic or antimony atom replaces a germanium atom, four of the valence electrons are used up to form the bond, and this leaves one excess valence electron. The valence electrons left over are relatively free to move around in the crystal under the influence of applied electric fields and make the material a donor of electrons, or n-type.

In p-, or positive-, type material, the impurity atoms are chosen to have a valence less than that of the host atoms. For example, gallium, with a valence of 3, may be used in a germanium crystal. When a gallium atom replaces a germanium atom, there are only three available valence electrons to form the required bond. A stable bond requires four valence electrons, and consequently, at each site where a gallium atom is located, a hole is created which can be filled by an electron that may be passing by. If an electron from some other bond moves over to fill the hole, the result is the creation of a new hole at some other point. The overall effect is as though the holes were positive carriers of electricity, i.e., equivalent positive electrons, that can move through the crystal. The holes do, in fact, behave as equivalent positive carriers, and thus p-type material can be considered as essentially the same as n-type material except that the signs of the charge carriers are opposite.

Consider now a linearly graded junction of n-type and p-type material, as shown in Fig. 11.1a. In the linearly graded junction, the n-type material changes gradually and in a linear fashion over to p-type material in a distance d. This variation is obtained by gradually reducing the doping, or concentration of donor atoms, down to zero in the region $x = d/2$ down to $x = 0$ and then linearly increasing the concentration of acceptor atoms in the region $x = 0$ to $x = -d/2$. If the number density of acceptor atoms is N_a and the density of donor atoms is N_d, the difference will vary linearly

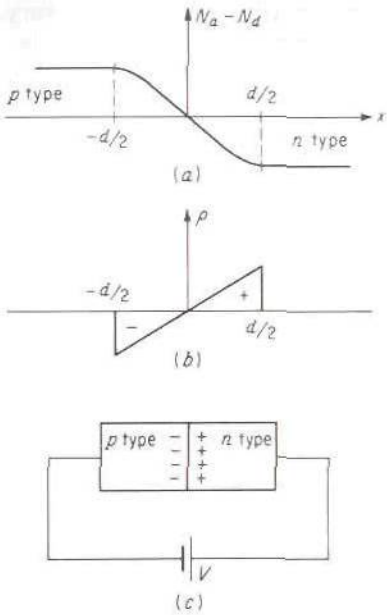


FIGURE 11.1
The linearly graded junction.

with x across the junction; thus

$$N_a - N_d = kx \tag{11.1}$$

where k is a suitable constant.

When there is a gradient in the impurity-concentration densities, electrons will diffuse from a region of high concentration to one of low concentration. Holes will diffuse in a similar manner. Thus the electrons will diffuse into the p-type side of the junction and holes will diffuse into the n-type side. This diffusion continues until a space-charge distribution, together with a resultant electric field, is set up of sufficient strength to produce a force that is equal and opposite to the effective force created by the concentration gradients. When equilibrium has been reached, a small region, called the depletion region, substantially free of charge carriers, is produced at the junction. The space charge built up on either side of the depletion region, together with the electric field existing across the depletion region, constitutes an equivalent capacitor. If a reversed-bias voltage is applied across the junction, the electron distribution and hole distribution will be forced farther apart. This widening of the depletion layer results in a decrease in the junction capacitance. It is now apparent that if an ac pumping voltage is superimposed on the bias voltage, the equivalent junction capacitance can be varied as a function of time.

In the graded junction the space-charge density will vary linearly across the junction so that a depletion layer completely free of carriers is not

produced. The effect of having the space charge vary linearly across the junction instead of being concentrated at $x = \pm d$ is, however, much the same. If we consider a linearly varying space-charge density (per unit cross-sectional area) $\rho = qx$, where q is a suitable constant, Poisson's equation gives

$$\frac{d^2\Phi}{dx^2} = -\frac{\rho}{\epsilon} = -\frac{q}{\epsilon}x$$

for the potential Φ. Integrating and using the boundary conditions that the electric field, and hence $d\Phi/dx$, is zero for $|x| > d/2$ and that $\Phi = 0$ at $x = 0$ from symmetry considerations, we get

$$\Phi = -\frac{qx}{2\epsilon}\left(\frac{x^2}{3} - \frac{d^2}{4}\right) \tag{11.2}$$

The potential difference across the junction is

$$\Phi\left(\frac{d}{2}\right) - \Phi\left(-\frac{d}{2}\right) = \frac{qd^3}{12\epsilon}$$

Under equilibrium conditions this potential difference must equal the contact potential Φ_c plus the negative applied bias voltage $-V$; thus

$$\Phi_c - V = \frac{qd^3}{12\epsilon} \tag{11.3}$$

The total stored charge per unit area is given by

$$Q = \int_0^{d/2} qx\,dx = \frac{qd^2}{8}$$

Eliminating d by means of (11.3) gives

$$Q = \frac{q}{8}\left[\frac{12\epsilon(\Phi_c - V)}{q}\right]^{2/3} \tag{11.4}$$

Since the capacitance is a function of the voltage, it must be defined as the ratio of an incremental change in Q to incremental change in $\Phi_c - V$. Thus the capacitance per unit cross-sectional area is

$$C = \frac{dQ}{d(\Phi_c - V)} = \epsilon\left[\frac{q}{12\epsilon(\Phi_c - V)}\right]^{1/3} \tag{11.5}$$

As seen from this equation, the junction capacitance C is nonlinear since it depends on the voltage V. If C is a linear element, $Q = CV$. In an abrupt junction diode C is proportional to $(\Phi_c - V)^{-1/2}$.

If we denote $\Phi_c - V$ by V_0 and apply in addition a pumping voltage $v_p = V_p \cos \omega_p t$, the capacitance becomes a function of time:

$$C(t) = \epsilon \left(\frac{q}{12\epsilon V_0} \right)^{1/2} \left(1 + \frac{V_p}{V_0} \cos \omega_p t \right)^{-1/3} \tag{11.6}$$

We now have a nonlinear capacitance that is also a function of time. The capacitance is a periodic function of time, and can be represented by a Fourier series expansion of the form

$$C(t) = \sum_{n=0}^{\infty} C_n \cos n\omega_p t \tag{11.7}$$

The coefficients are given by

$$C_0 = \frac{\epsilon}{2\pi} \left(\frac{q}{12\epsilon V_0} \right)^{1/3} \int_{-\pi}^{\pi} \left(1 + \frac{V_p}{V_0} \cos \theta \right)^{-1/3} d\theta$$

$$C_n = \frac{\epsilon}{\pi} \left(\frac{q}{12\epsilon V_0} \right)^{1/3} \int_{-\pi}^{\pi} \left(1 + \frac{V_p}{V_0} \cos \theta \right)^{-1/3} \cos n\theta \, d\theta$$

where $\theta = \omega_p t$. To evaluate the coefficients would require a numerical procedure. However, we do not need to know the values of the C_n in order to analyze the general properties of a parametric amplifier. The important feature brought out in the foregoing analysis is that $C(t)$ is a function of time that can be represented by a Fourier series involving all harmonics of the pumping frequency f_p. It is important to note that the coefficients are not, in general, linear functions of the ac voltage V_p or the voltage V_0. Thus, since the junction capacitance $C(t)$ is a nonlinear capacitance, the principle of superposition *does not* hold for arbitrary ac signal amplitudes. Under small-signal conditions, A Taylor series expansion of $C(t)$ about an operating point may be used and only the linear term in the signal amplitude retained. In this case superposition *does* hold. The situation here is no different from that in any other amplifying device since all are nonlinear for sufficiently large applied signals.

In addition to the capacitance associated with the diode junction, there is a shunt conductance arising from the bulk resistance of the material in the depletion layer. This shunt conductance is proportional to the area, and since C is also proportional to the area, the ratio is independent of the cross-sectional area of the diode. The shunt conductance of the depletion layer is small, and can often be neglected. Of more importance is the series resistance of the n- and p-type semiconductor material outside the depletion layer. When the p-n junction is encapsulated and connecting leads are put on, an additional shunt capacitance C_p due to the cartridge and a series inductance arising from the leads are also present. The overall equivalent circuit is thus of the form shown in Fig. 11.2. Typical values of C_p and L_s are somewhat less than 1 pF and 1 nH, respectively, at microwave frequen-

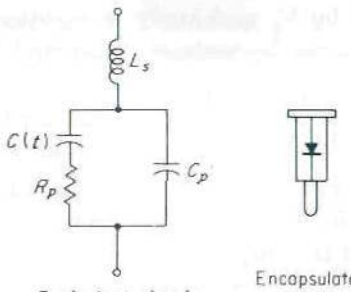


FIGURE 11.2
Equivalent circuit of a parametric diode.

cies. The junction capacitance C is also about 1 pF, and typical values of R_p are a few ohms.

11.2 MANLEY-ROWE RELATIONS

Manley and Rowe have derived a set of power-conservation relations that are extremely useful in evaluating the performance which can be achieved from a parametric device consisting of a nonlinear reactance.† These relations are also derived below.

The circuit considered by Manley and Rowe is shown in Fig. 11.3. It consists of resistive loads in series with ideal bandpass filters connected in shunt with a lossless nonlinear capacitance. Two sinusoidal signals at frequencies f_1 and f_2 are applied. The nonlinear capacitance causes frequencies at the harmonics of f_1 and f_2 to be generated. Each bandpass filter is considered to pass only one harmonic component $nf_1 + mf_2$. The overall circuit thus isolates all the harmonics and dissipates their power in separate resistive loads. The Manley-Rowe relations establish two constraints governing the conversion of input power at the frequencies f_1 and f_2 into power at other frequencies.

Let the charge Q on C be a single-valued function of the ac voltage $v = v_1 + v_2 = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t$ applied across it. Thus $Q = Q(v)$. We may expand Q in a Taylor series in v to obtain

$$Q = Q(0) + \frac{\partial Q}{\partial v} v + \frac{1}{2} \frac{\partial^2 Q}{\partial v^2} v^2 + \cdots \tag{11.8}$$

where all derivatives are evaluated at $v = 0$. Since all powers of v occur, it is clear that, because $v = (V_1/2)(e^{j\omega_1 t} + e^{-j\omega_1 t}) + (V_2/2)(e^{j\omega_2 t} + e^{-j\omega_2 t})$, the charge Q will have frequencies at all harmonics of f_1 and f_2. If currents at

†J. M. Manley and H. E. Rowe, Some General Properties of Nonlinear Elements, Part I, General Energy Relations, *Proc. IRE*, vol. 44, pp. 904–913, July, 1956. See also *Proc. IRE*, vol. 47, pp. 2115–2116, December, 1959.

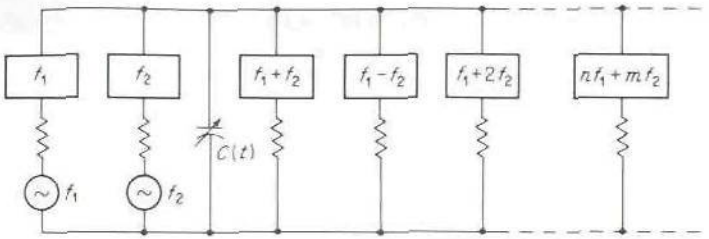


FIGURE 11.3
Circuit for illustration of the Manley-Rowe relations.

the various harmonics are permitted to pass through C, the voltage developed across C will also contain all possible harmonics. In this case Q is a function of all voltages present. However, the expansion (11.8) will still be valid, except that now the coefficients will have different values. Consequently, the general expansion of Q has the form

$$Q = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} Q_{nm} e^{j(n\omega_1 + m\omega_2)t} \tag{11.9}$$

The charge is a real function of time; so we must have $Q_{-n-m} = Q_{nm}^*$ in order that the n, m and $-n, -m$ terms will combine to form a real function of time with frequency $n\omega_1 + m\omega_2$.

The total voltage V can be expressed as a function $V(Q)$ of the charge. A similar Taylor series expansion of $V(Q)$ then shows that V can be expressed in a form similar to (11.9); thus

$$V = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} V_{nm} e^{j(n\omega_1 + m\omega_2)t} \tag{11.10}$$

For V to be real, we must have $V_{-n-m} = V_{nm}^*$.

The current through C is the total rate of change of Q with time, and is given by

$$I = \frac{dQ}{dt} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} j(n\omega_1 + m\omega_2) Q_{nm} e^{j(n\omega_1 + m\omega_2)t}$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} I_{nm} e^{j(n\omega_1 + m\omega_2)t} \tag{11.11}$$

where $I_{nm} = j(n\omega_1 + m\omega_2)Q_{nm}$.

Since C is a pure reactive element, there can be no net power into or out of C. If we assume that ω_1 and ω_2 are incommensurable, there will be no time-average power due to interacting harmonics. The average power at

the frequencies $\pm|n\omega_1 + m\omega_2|$ is

$$\begin{aligned} P_{nm} &= (V_{nm}I_{nm}^* + V_{nm}^*I_{nm}) \\ &= (V_{nm}I_{nm}^* + V_{-n-m}I_{-n-m}^*) = P_{-n-m} \end{aligned} \tag{11.12}$$

since the time average of

$$\begin{aligned} &\left(I_{nm}e^{j(n\omega_1+m\omega_2)t} + I_{-n-m}e^{-j(n\omega_1+m\omega_2)t}\right) \\ &\times\left(V_{nm}e^{j(n\omega_1+m\omega_2)t} + V_{-n-m}e^{-j(n\omega_1+m\omega_2t)}\right) \end{aligned}$$

is

$$V_{nm}I_{-n-m} + V_{-n-m}I_{nm} = V_{nm}I_{nm}^* + V_{nm}^*I_{nm} = V_{nm}I_{nm}^* + V_{-n-m}I_{-n-m}^*$$

Conservation of power can therefore be expressed as

$$\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} P_{nm} = 0 \tag{11.13}$$

since $P_{nm} = P_{-n-m}$ from (11.12). To obtain the Manley-Rowe relations, we multiply each term by $(n\omega_1 + m\omega_2)/(n\omega_1 + m\omega_2)$ and split the sum into two parts; thus

$$\omega_1\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\frac{nP_{nm}}{n\omega_1 + m\omega_2} + \omega_2\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\frac{mP_{nm}}{n\omega_1 + m\omega_2} = 0 \tag{11.14}$$

We can now show that each double sum must vanish separately. We can replace each $I_{nm}/(n\omega_1 + m\omega_2)$ by jQ_{nm}, and then $P_{nm}/(n\omega_1 + m\omega_2)$ becomes $-jV_{nm}Q_{nm}^* - jV_{-n-m}Q_{-n-m}^*$ and does not depend explicitly on ω_1 or ω_2. For any choice of ω_1 and ω_2, we can always adjust the network external to C so that the currents which result keep all the voltage amplitudes V_{nm} unchanged. The Q_{nm} are then also unchanged since they depend only on the V_{nm}. When this is done we see from (11.14) that it is possible to change ω_1 and ω_2 arbitrarily but keep the two double sums involving $P_{nm}/(n\omega_1 + m\omega_2) = -jV_{nm}Q_{nm}^* - jV_{-n-m}Q_{-n-m}^*$ unchanged. Consequently, (11.14) can hold for arbitrary ω_1 and ω_2 only if

$$\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\frac{nP_{nm}}{n\omega_1 + m\omega_2} = 0$$

$$\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\frac{mP_{nm}}{n\omega_1 + m\omega_2} = 0$$

That is, the coefficients of ω_1 and ω_2 must vanish separately. The above two relations are the Manley-Rowe relations. They are usually written in a somewhat different form, however. We may write the first sum as

$$\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}\frac{nP_{nm}}{n\omega_1 + m\omega_2} + \sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}\frac{-nP_{-n-m}}{-n\omega_1 - m\omega_2}$$

where n and m have been replaced by $-n$ and $-m$ in the second term.

Since $P_{-n-m} = P_{nm}$, the two parts are equal; so we obtain

$$\sum_{n=0}^{\infty} \sum_{m=-\infty}^{\infty} \frac{nP_{nm}}{n\omega_1 + m\omega_2} = 0 \qquad (11.15a)$$

Similarly, we can obtain

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{nm}}{n\omega_1 + m\omega_2} = 0 \qquad (11.15b)$$

These are the standard forms for the Manley-Rowe relations. The Manley-Rowe relations are general power-conservation relations, and do not depend on any specific circuit such as that in Fig. 11.3. This is apparent since no reference to an external circuit was made in the derivation.

For an example of the application of the Manley-Rowe relations, consider a circuit similar to that in Fig. 11.3 with generators at frequencies f_1 and f_2. Let all harmonics be open-circuited except $f_1 + f_2$. Thus currents at the three frequencies f_1, f_2, and $f_1 + f_2$ are the only ones that can exist. The $n = \pm 1$, $m = 0$ and $n = 0$, $m = \pm 1$ and $n = m = \pm 1$ terms in (11.15) are the only ones present. Thus we get

$$\frac{P_{10}}{\omega_1} + \frac{P_{11}}{\omega_1 + \omega_2} = 0 \qquad (11.16a)$$

$$\frac{P_{01}}{\omega_2} + \frac{P_{11}}{\omega_1 + \omega_2} = 0 \qquad (11.16b)$$

Since power is supplied at the frequencies ω_1 and ω_2, we must have P_{10} and P_{01} positive. Therefore P_{11} is negative, and power is delivered from the nonlinear capacitor C at the frequency $\omega_1 + \omega_2$. If ω_1 is the input-signal frequency and ω_2 is the pump frequency, then $\omega_3 = \omega_1 + \omega_2$ is the output frequency. The maximum signal gain is

$$-\frac{P_{11}}{P_{10}} = \frac{\omega_1 + \omega_2}{\omega_1} = \frac{\omega_3}{\omega_1} = 1 + \frac{\omega_2}{\omega_1} \qquad (11.17)$$

as obtained from (11.16a). A parametric amplifier of this type is called an up-converter. Because of losses that are always present in a practical amplifier, the gain will be less than ω_3/ω_1. The Manley-Rowe relations give the maximum gain possible and hence provide a criterion by which a practical up-converter can be judged.

11.3 LINEARIZED EQUATIONS FOR PARAMETRIC AMPLIFIERS

Consider a linear capacitance C for which the charge-voltage relationship is $Q = Cv$. The current flowing through C is given by

$$i = \frac{dQ}{dt} = \frac{d}{dt}(Cv) = C\frac{dv}{dt}$$

If C is made a function of time, for example, a parallel-plate capacitor with a plate separation that is varied with time, the current will be given by

$$i = \frac{d}{dt}(Cv) = v\frac{dC}{dt} + C\frac{dv}{dt} \tag{11.18}$$

If instead of a time-varying linear capacitance we have a nonlinear capacitance, where the charge Q is a nonlinear function $Q(v)$ of the voltage, the current is given by

$$i = \frac{dQ}{dt} = \frac{\partial Q}{\partial v}\frac{dv}{dt}$$

If the voltage v is the sum of a pump voltage v_p at frequency ω_p and a signal voltage v_s at frequency ω_s and $|v_s| \ll |v_p|$, we can expand

$$Q(v) = Q(v_p + v_s)$$

in a Taylor series about the point $v_s = 0$. Thus we obtain

$$Q(v_p + v_s) = Q(v_p) + \left.\frac{\partial Q}{\partial v}\right|_{v_s=0} v_s + \frac{1}{2}\left.\frac{\partial^2 Q}{\partial v^2}\right|_{v_s=0} v_s^2 + \cdots$$

For v_s sufficiently small, we can obtain satisfactory accuracy by retaining the first two terms only. The current is then given by

$$i = \frac{dQ(v_p)}{dt} + \frac{d}{dt}\left(\left.\frac{\partial Q}{\partial v}\right|_{v_s=0} v_s\right) \tag{11.19a}$$

Let the quantity $\partial Q/\partial v$ for $v_s = 0$ be denoted by $C(t)$, in which case we can write

$$i = \frac{dQ(v_p)}{dt} + \frac{d}{dt}[C(t)v_s] \tag{11.19b}$$

If we compare this result with (11.18), we see that the nonlinear capacitance behaves like a *time-varying linear capacitance for signals* with amplitudes that are small compared with the pump signal amplitude. The first term, $dQ(v_p)/dt$, in (11.19) gives a current at the pumping frequency, and is not related to the signal current. If the pumping voltage is also small compared with the dc bias voltage in a junction diode, we can assume $C(t)$ to have the form [see (11.6)]

$$C(t) = C_0(1 + 2M\cos\omega_p t) \tag{11.20}$$

since $(V_0 + V_p\cos\omega_p t)^{-1/3} \approx V_0^{-1/3}[1 - (V_p/3V_0)\cos\omega_p t]$ for $V_p \ll V_0$. The linearized equations (11.19) and (11.20) are the ones we shall use in the *analysis of parametric amplifiers*.

The equivalent circuit of a p-n junction diode is illustrated in Fig. 11.2. For the purpose of analysis it is more convenient to use an equivalent series circuit of the form shown in Fig. 11.4. The two circuits are equivalent

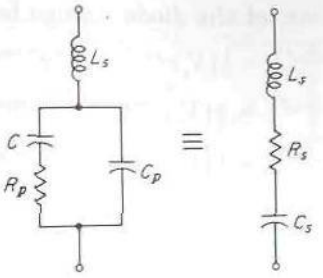


FIGURE 11.4
Equivalent series circuit for a junction diode.

if we choose

$$R_s = \frac{R_p C^2}{(\omega C C_p R_p)^2 + (C + C_p)^2} \qquad C_s = \frac{(\omega C C_p R_p)^2 + (C + C_p)^2}{(\omega C R_p)^2 C_p + C + C_p}$$

which makes the input impedance the same for both circuits at the frequency ω. In most diodes the resistance R_p is small compared with the cartridge reactance $1/\omega C_p$; so $\omega C_p R_p \ll 1$. In this case we find that

$$R_s \approx \left(\frac{C}{C + C_p}\right)^2 R_p \qquad C_s \approx C + C_p$$

When these approximations are valid, the two circuits are equivalent, independently of frequency. This is a necessary requirement if the series circuit is to be useful for analysis purposes, since in a parametric amplifier currents and voltages at several different frequencies are simultaneously present. When C is a function of the voltage, R_s will also be voltage-dependent. The effect of a voltage-dependent resistance R_s will, for simplicity, be neglected, since it is not too important. In other words, we shall consider R_s to be a constant resistance.

11.4 PARAMETRIC UP-CONVERTER

In the up-converter, a pump voltage at frequency ω_p and a signal at frequency ω_s are applied to the diode, and the output signal is taken at the higher frequency $\omega_s + \omega_p$. Mixing effects take place that give rise to all possible harmonics of ω_p and ω_s. However, in the up-converter, the circuit external to the diode is chosen so as to permit currents only at the signal frequency ω_s, the pump frequency ω_p, and the output frequency ω_0, which is chosen as the sum of the pump and signal frequencies, that is, at $\omega_0 = \omega_s + \omega_p$. There will, consequently, be a voltage across the diode at the

three possible frequencies. If we let the diode voltage be represented as

$$v_s = \mathrm{Re}\left(V_s e^{j\omega_s t}\right) = \tfrac{1}{2}\left(V_s e^{j\omega_s t} + V_s^* e^{-j\omega_s t}\right)$$

$$v_p = \mathrm{Re}\left(V_p e^{j\omega_p t}\right) = \tfrac{1}{2}\left(V_p e^{j\omega_p t} + V_p^* e^{-j\omega_p t}\right)$$

$$v_0 = \mathrm{Re}\left(V_0 e^{j\omega_0 t}\right) = \tfrac{1}{2}\left(V_0 e^{j\omega_0 t} + V_0^* e^{-j\omega_0 t}\right)$$

we may generalize (11.19) and (11.20) to give

$$i = \frac{dQ(v_p)}{dt} + \frac{1}{2}\frac{d}{dt}C_0(1 + 2M\cos\omega_p t)$$
$$\times\left(V_s e^{j\omega_s t} + V_s^* e^{-j\omega_s t} + V_0 e^{j\omega_0 t} + V_0^* e^{-j\omega_0 t}\right) \tag{11.21}$$

Let the current at the frequencies ω_s and ω_0 be expressed as

$$i_s = \tfrac{1}{2}\left(I_s e^{j\omega_s t} + I_s^* e^{-j\omega_s t}\right) \tag{11.22}$$

$$i_0 = \tfrac{1}{2}\left(I_0 e^{j\omega_0 t} + I_0^* e^{-j\omega_0 t}\right) \tag{11.23}$$

When the time derivative of (11.21) is taken and the terms at frequencies ω_s and $\pm\omega_0$ only are retained, we obtain (a knowledge of the pump current at frequency ω_p will not be required; so we do not need to evaluate it)

$$i_s + i_0 = \frac{C_0}{2}\left(j\omega_s V_s e^{j\omega_s t} - j\omega_s V_s^* e^{-j\omega_s t} + j\omega_0 V_0 e^{j\omega_0 t} - j\omega_0 V_0^* e^{-j\omega_0 t}\right.$$
$$\left. + j\omega_0 M V_s e^{j\omega_0 t} - j\omega_s M V_0^* e^{-j\omega_s t} - j\omega_0 M V_s^* e^{-j\omega_0 t} + j\omega_s M V_0 e^{j\omega_s t}\right)$$

Using (11.22) and (11.23) gives

$$I_s = j\omega_s C_0 V_s + j\omega_s C_0 M V_0 \tag{11.24a}$$

$$I_0 = j\omega_0 C_0 V_0 + j\omega_0 C_0 M V_s \tag{11.24b}$$

These two equations show that, for the input signal current i_s and output signal current i_0, the junction capacitance may be represented by an admittance matrix such that

$$\begin{bmatrix} I_s \\ I_0 \end{bmatrix} = \begin{bmatrix} j\omega_s C_0 & j\omega_s C_0 M \\ j\omega_0 C_0 M & j\omega_0 C_0 \end{bmatrix}\begin{bmatrix} V_s \\ V_0 \end{bmatrix} \tag{11.24c}$$

The parameter M is proportional to the pump voltage and gives the coupling between the voltages at the two frequencies ω_s and ω_0.

Figure 11.5 illustrates an equivalent-circuit model of an up-converter. The series tuned circuits are chosen so that the three circuit loops have resonant frequencies of ω_s, ω_0, and ω_p and only currents with these respective frequencies can exist in each loop. Thus, in the input circuit loop, only I_s is present. The three circuit loops are coupled together through the time-varying part C of C_s only. Therefore, for the two frequencies ω_s and ω_0, the equivalent circuit can be reduced to that illustrated in Fig. 11.6. The box labeled $C(t)$ in this circuit is an equivalent impedance network that maintains the relationship given by (11.24) between the terminal currents

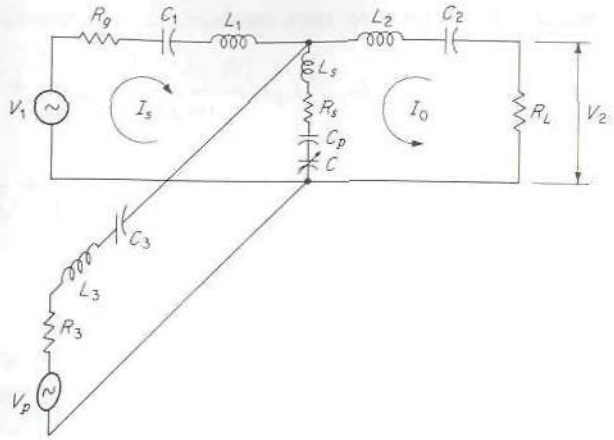


FIGURE 11.5
Equivalent-circuit model of an up-converter.

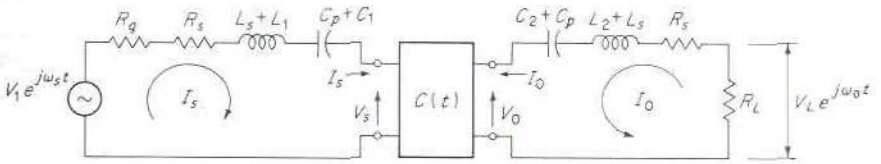


FIGURE 11.6
Reduced equivalent circuit for an up-converter.

and voltages. The analysis of the parametric amplifier is a conventional network-analysis problem since the diode has been replaced by an equivalent linear two-port network with terminal relations described by (11.24).† Each loop in the circuit is assumed to provide a very high impedance to currents at all frequencies present except the resonant frequency for that loop. The resonant circuits in Fig. 11.5 have been assumed to have zero loss. Circuit losses can be considered included in R_g and R_L. At the end of the analysis, R_g and R_L can then be split into two parts so as to separate the circuit losses from the generator and load impedances. In practice, circuit losses are small compared with the loss arising from the diode resistance R_s and the external loading represented by R_g and R_L.

We may solve (11.24c) for V_s and V_0 in terms of I_s and I_0 to obtain

$$\begin{bmatrix} V_s \\ V_0 \end{bmatrix} = \frac{1}{1-M^2} \begin{bmatrix} 1/j\omega_s C_0 & -M/j\omega_0 C_0 \\ -M/j\omega_s C_0 & 1/j\omega_0 C_0 \end{bmatrix} \begin{bmatrix} I_s \\ I_0 \end{bmatrix} \tag{11.25}$$

†Care must be exercised in the analysis since currents and voltages at several different frequencies are simultaneously present.

For the input circuit we may now write

$$V_1 = I_s\left[R_g + R_s + j\omega_s(L_s + L_1) + \frac{1}{j\omega_s(C_p + C_1)}\right] + V_s$$

$$= I_s\left[R_g + R_s + j\omega_s(L_s + L_1) + \frac{1}{j\omega_s(C_p + C_1)} + \frac{1}{j\omega_s(1 - M^2)C_0}\right]$$

$$- \frac{MI_0}{j\omega_0(1 - M^2)C_0}$$

whereas for the output circuit

$$0 = I_0\left[R_L + R_s + j\omega_0(L_s + L_2) + \frac{1}{j\omega_0(C_p + C_2)}\right] + V_0$$

$$= I_0\left[R_L + R_s + j\omega_0(L_s + L_2) + \frac{1}{j\omega_0(C_p + C_2)} + \frac{1}{j\omega_0(1 - M^2)C_0}\right]$$

$$- \frac{MI_s}{j\omega_s(1 - M^2)C_0}$$

Let us now assume that the circuits are tuned so that the following conditions hold:

$$\omega_s^2 = \frac{1}{L_s + L_1}\left[\frac{1}{C_p + C_1} + \frac{1}{(1 - M^2)C_0}\right]$$

$$\omega_0^2 = \frac{1}{L_s + L_2}\left[\frac{1}{C_p + C_2} + \frac{1}{(1 - M^2)C_0}\right]$$

We then obtain

$$V_1 = I_s(R_g + R_s) + \frac{jMI_0}{\omega_0(1 - M^2)C_0}$$

$$0 = I_0(R_L + R_s) + \frac{jMI_s}{\omega_s(1 - M^2)C_0}$$

We may solve for I_0 to obtain

$$I_0 = \frac{-jV_1M\omega_0C_0(1 - M^2)}{M^2 + (R_g + R_s)(R_L + R_s)\omega_0\omega_s(1 - M^2)^2C_0^2} \tag{11.26}$$

The maximum available input power from the generator is $\frac{1}{2}(V_1^2/4R_g)$, and the output power developed in R_L is $\frac{1}{2}|I_0|^2R_L$. The midband transducer

power gain is thus given by

$$G_0 = \frac{4|I_0|^2 R_L R_g}{V_1^2}$$

$$= \frac{4R_L R_g M^2}{\omega_s^2(1-M^2)^2 C_0^2\left[(R_g+R_s)(R_L+R_s)+\dfrac{M^2}{\omega_0\omega_s(1-M^2)^2 C_0^2}\right]^2} \tag{11.27}$$

when (11.26) is used to express I_0 in terms of V_1. If desired, circuit losses can be included at this point by replacing R_g and R_L in the denominator in (11.27) by $R_g + R_{1l}$ and $R_L + R_{2l}$, where R_{1l} and R_{2l} represent the loss resistance in the input and output circuits. For simplicity we shall take $R_{1l} = R_{2l} = 0$.

To achieve maximum gain requires adjustment of R_g and R_L. Since R_g and R_L occur symmetrically in the expression for G_0, the optimum values of R_L and R_g are equal. Hence we need to maximize

$$G_0 = \frac{4R_L^2 M^2}{\omega_s^2(1-M^2)^2 C_0^2\left[(R_L+R_s)^2+\dfrac{M^2}{\omega_0\omega_s(1-M^2)^2 C_0^2}\right]^2}$$

Equating dG_0/dR_L to zero and solving for R_L give

$$R_L = R_s\left[1+\frac{M^2}{\omega_0\omega_s R_s^2(1-M^2)^2 C_0^2}\right]^{1/2} \tag{11.28}$$

The effective Q of the diode may be defined as

$$Q = \frac{1}{R_s\omega_s(1-M^2)C_0} \tag{11.29}$$

We then find that

$$R_L = R_s\left[1+\frac{\omega_s}{\omega_0}(MQ)^2\right]^{1/2} \tag{11.30}$$

and the maximum gain is

$$G_0 = \frac{\omega_0}{\omega_s}\frac{\delta}{(1+\sqrt{1+\delta})^2} \tag{11.31}$$

where $\delta = (\omega_s/\omega_0)(MQ)^2$. According to the Manley-Rowe relations discussed in Sec. 11.2, the maximum gain of an up-converter is ω_0/ω_s. The quantity $\delta/(1+\sqrt{1+\delta})^2$ may therefore be regarded as a gain-degradation

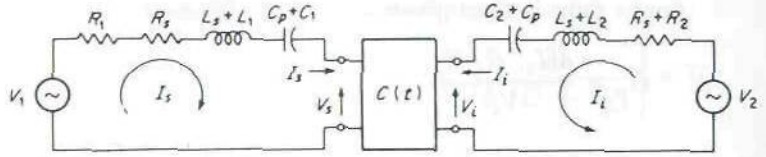


FIGURE 11.7
Equivalent circuit for a negative-resistance parametric amplifier.

factor. As the diode Q approaches infinity, that is, as R_s goes to zero, δ approaches infinity, and the gain-degradation factor becomes equal to unity. Hence, for a lossless diode, the gain becomes equal to ω_0/ω_s, as predicted by the Manley-Rowe relations. In a typical microwave diode, MQ could be equal to 10. If $\omega_0/\omega_s = 10$ also, the maximum gain as given by (11.31) is 7.3 dB.

To achieve high gain with an up-converter requires a large ratio ω_0/ω_s of output-to-input frequency. At the higher microwave frequencies this is not a very practical requirement, and for this reason up-converters are usually restricted to operation at signal frequencies f_s below 1,000 MHz. Higher gain can be obtained from the negative-resistance parametric amplifier, which is discussed in the next section.

11.5 NEGATIVE-RESISTANCE PARAMETRIC AMPLIFIER

In the negative-resistance parametric amplifier currents are permitted to exist at the signal frequency ω_s, the pump frequency ω_p, and the idler frequency $\omega_i = \omega_p - \omega_s$. The equivalent-circuit model that will be analyzed is shown in Fig. 11.7.

When we replace the voltage v_0 in (11.21) by $v_i = \frac{1}{2}(V_i e^{j\omega_i t} + V_i^* e^{-j\omega_i t})$ and introduce the idler current $i_i = \frac{1}{2}(I_i e^{j\omega_i t} + I_i^* e^{-j\omega_i t})$, we may solve for I_s and I_i in terms of V_s and V_i in the same manner that was used to treat the up-converter. It is readily found that, for $\omega_i = \omega_p - \omega_s$,

$$\begin{bmatrix} I_i \\ I_s^* \end{bmatrix} = \begin{bmatrix} j\omega_i C_0 & j\omega_i C_0 M \\ -j\omega_s C_0 M & -j\omega_s C_0 \end{bmatrix} \begin{bmatrix} V_i \\ V_s^* \end{bmatrix} \tag{11.32a}$$

and

$$\begin{bmatrix} V_i \\ V_s^* \end{bmatrix} = \frac{1}{1 - M^2} \begin{bmatrix} \dfrac{1}{j\omega_i C_0} & \dfrac{M}{j\omega_s C_0} \\ \dfrac{-M}{j\omega_i C_0} & \dfrac{-1}{j\omega_s C_0} \end{bmatrix} \begin{bmatrix} I_i \\ I_s^* \end{bmatrix} \tag{11.32b}$$

For the circuit of Fig. 11.7, we can write the following equations:

$$V_1 = I_s\left[R_1 + R_s + j\omega_s(L_1 + L_s) + \frac{1}{j\omega_s(C_p + C_1)} + \frac{1}{j\omega_s(1 - M^2)C_0}\right] + \frac{M}{(1 - M^2)j\omega_i C_0}I_i^* \tag{11.33a}$$

$$V_2 = I_i\left[R_2 + R_s + j\omega_i(L_2 + L_s) + \frac{1}{j\omega_i(C_p + C_2)} + \frac{1}{j\omega_i(1 - M^2)C_0}\right] + \frac{M}{(1 - M^2)j\omega_s C_0}I_s^* \tag{11.33b}$$

If we impose the tuning conditions

$$\omega_s^2 = \frac{1}{L_1 + L_s}\left[\frac{1}{C_p + C_1} + \frac{1}{(1 - M^2)C_0}\right]$$

$$\omega_i^2 = \frac{1}{L_2 + L_s}\left[\frac{1}{C_p + C_2} + \frac{1}{(1 - M^2)C_0}\right]$$

we obtain

$$V_1 = (R_1 + R_s)I_s - \frac{jM}{\omega_i(1 - M^2)C_0}I_i^* \tag{11.34a}$$

$$V_2 = (R_2 + R_s)I_i - \frac{jM}{\omega_s(1 - M^2)C_0}I_s^* \tag{11.34b}$$

Let us now assume that $V_2 = 0$. We can then determine I_i from (11.34) and evaluate the gain $G_0 = 4R_1R_2|I_i|^2/V_1^2$. We readily find that

$$G_0 = \frac{4R_1R_2M^2}{\left[R_1 + R_s - \dfrac{M^2}{(R_2 + R_s)\omega_i\omega_s(1 - M^2)^2C_0^2}\right]^2(R_2 + R_s)^2\omega_s^2C_0^2(1 - M^2)^2} \tag{11.35}$$

The term $-M^2/[(R_2 + R_s)\omega_i\omega_s(1 - M^2)^2C_0^2]$ may be interpreted as an equivalent negative resistance $-R$. Introducing R, we may express G_0 as

$$G_0 = \frac{4R_1R_2\left[\omega_i(1 - M^2)C_0R\right]^2}{M^2(R_1 + R_s - R)^2} \tag{11.36}$$

It is clear that a very large gain can be obtained if R is made almost equal to $R_1 + R_s$. However, care must be taken not to make R too close to $R_1 + R_s$ because a small change in parameters will then cause large changes

in the gain and will cause oscillations to occur if R becomes equal to $R_1 + R_s$.

The parametric amplifier discussed above is called a negative-resistance converter. It is possible to take the output at the same frequency ω_s as the input. If we split R_1 into a generator internal resistance R_g plus a load resistance R_L, the power delivered to R_L is $\frac{1}{2}R_L|I_s|^2$. The transducer power gain will be

$$G_0 = \frac{4R_gR_L|I_s|^2}{V_1^2}$$

We may evaluate I_s from (11.34) to obtain

$$G_0 = \frac{4R_gR_L}{(R_g + R_L + R_s - R)^2} \tag{11.37}$$

where, as before,

$$R = \frac{M^2}{(R_2 + R_s)\omega_i\omega_s(1 - M^2)^2C_0^2} \tag{11.38}$$

The effective negative resistance $-R$ arises in the following manner: The application of signal plus pump power to the nonlinear capacitance causes frequency mixing to occur. When current is permitted to exist at the idler frequency $\omega_p - \omega_s$, further frequency mixing of power at the pump and idler frequencies occurs. This latter mixing causes harmonics of ω_p and $\omega_i = \omega_p - \omega_s$ to be generated; in particular, power at the frequency ω_s is generated. When the power generated through frequency mixing exceeds that being supplied at the signal frequency ω_s, the diode appears to have a negative resistance. If idler current is not permitted to exist, the negative resistance vanishes, as reference to (11.38) shows when R_2 is made infinite (open circuit for the idler signal).

The negative-resistance parametric amplifier with input and output at the same frequency is not very stable. The reason is that in a microwave system R_g and R_L are the impedances seen looking into the input and output transmission-line ports. If the loads connected to these transmission lines are not matched, reflected waves occur. Reflected waves in the output line return to the amplifier and are amplified and fed into both the input and output lines. The result is that the gain becomes a sensitive function of the external generator and load impedances. The stability of the amplifier is greatly improved by the use of a circulator, as illustrated in Fig. 11.8. The use of a circulator makes the load termination R_L for the amplifier equal to the characteristic impedance of the transmission line independently of the external generator or load impedances Z_g and Z_L. The available power from the generator is still given by $\frac{1}{2}V_1^2/4R_g$. However, the amplifier power is

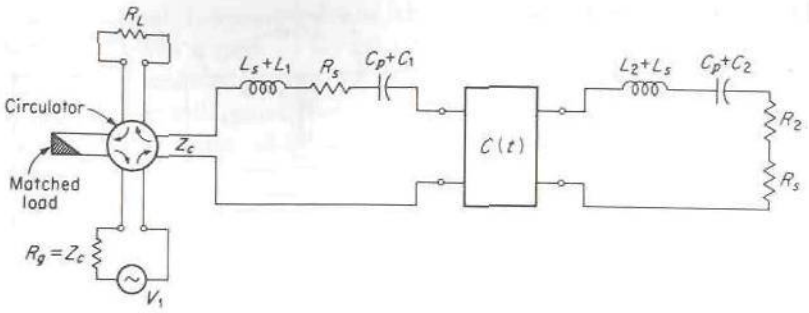


FIGURE 11.8
A negative-resistance parametric amplifier using a circulator.

now all delivered to the load $R_L = Z_c$, and none of it is dissipated in the internal generator impedance R_g. Consequently, the power gain is nearly four times greater since I_s is nearly twice as large, because the series resistance in the input circuit is now $R_L + R_s$ instead of $R_L + R_g + R_s \approx 2R_L$, when $R_L = R_g \gg R_s$. The power gain is the square of the voltage reflection coefficient, and is given by

$$G_0 = \left| \frac{Z_{\text{in}} - Z_c}{Z_{\text{in}} + Z_c} \right|^2 = \frac{(R_s - R - R_L)^2}{(R_s + R_L - R)^2}$$

since $Z_{\text{in}} = R_s - R$ at resonance and we are taking $R_L = Z_c$. For high gain, $R \approx R_L + R_s$ and is large compared with R_s. Consequently, the gain can be expressed as

$$G_0 = \frac{4R_L^2}{(R_L + R_s - R)^2} \tag{11.39}$$

The maximum value of negative resistance that can be obtained is determined by the diode that is used. If we make $R_2 = 0$ in (11.38), we see that maximum R is R_m, where

$$R_m = \frac{\omega_s}{\omega_i}(MQ)^2 R_s \tag{11.40}$$

(11.29) having been used to introduce the diode Q. If we regard R as fixed, we see that, for the amplifier without a circulator, we must make $R_g + R_L + R_s \approx R$, as (11.37) shows. But with $R_g = R_L = Z_c$ large compared with R_s, we get $R_g = R_L \approx R/2$, and (11.37) gives

$$G_0 = \frac{R^2}{(2R_L + R_s - R)^2} \tag{11.41a}$$

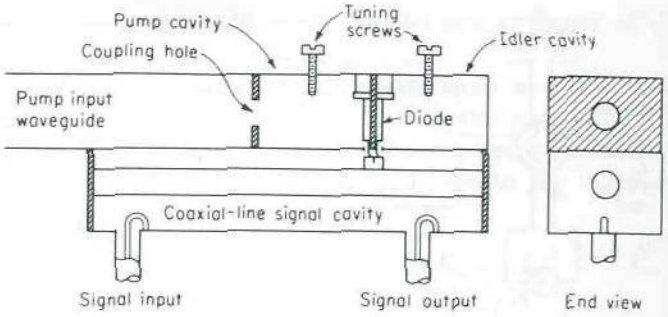


FIGURE 11.9
A microwave negative-resistance parametric amplifier.

whereas for the amplifier with a circulator, $R_L \approx R$ and

$$G_0 = \frac{4R^2}{(R_L + R_s - R)^2} \qquad (11.41b)$$

Note that when a circulator is used, the optimum value of R_L is twice what it is when a circulator is not used. Thus the denominators in (11.41*a*) and (11.41*b*) are the same. These relations then show that the use of a circulator gives 6 dB more gain for the same amount of diode loading, and hence will have a gain-bandwidth product twice as great.

If the pump frequency ω_p is chosen equal to twice the signal frequency ω_s, the idler frequency $\omega_i = \omega_p - \omega_s = \omega_s$ is equal to the signal frequency. In this case the amplifier is called a degenerate negative-resistance amplifier. For the degenerate amplifier the signal and idler circuits would be a single resonant circuit. The analysis of the degenerate amplifier is similar to that already carried out, and so will not be presented here (some results on noise properties are given in Sec. 11.6).

There are many different ways of building microwave parametric amplifiers. Transmission lines, waveguides, or a combination of the two may be used to construct suitable cavities to use as resonant circuits. A typical microwave negative-resistance amplifier is illustrated in Fig. 11.9.† The pump and idler cavities are formed in an X-band rectangular waveguide. The signal cavity is a coaxial-transmission-line cavity. The varactor diode is mounted in the center of an inductive diaphragm located between the pump and idler cavities. Coupling to the signal cavity is achieved by having the diode terminate in the center conductor of the coaxial-line signal cavity. The pump is coupled to its cavity by an aperture. The pump frequency is chosen around 9,200 MHz, and the idler frequency is 7,900 MHz. The input and

†W. O. Troetschel and H. J. Heuer, A Parametric Amplifier for 1296 Mc, *QST*, January, 1961.

output signal frequencies are the same and equal to 1,300 MHz. This amplifier gives a gain of 25 dB or more with a bandwidth of about 5 MHz.

The bandwidth over which high gain can be obtained in a negative-resistance amplifier of the type discussed above is small. The negative resistance has the effect of increasing the loaded Q, which results in a high-Q resonant circuit with a narrow bandwidth. To analyze the bandwidth properties, we shall assume that the circuit model given in Fig. 11.7 is valid. For high-Q circuits the impedance of the signal and idler circuits may be expressed in the form

$$Z_s = (R_1 + R_s)\left(1 + 2j\frac{\Delta\omega_s}{\omega_s}Q_1\right) \tag{11.42a}$$

$$Z_i = (R_2 + R_s)\left(1 + 2j\frac{\Delta\omega_i}{\omega_i}Q_2\right) \tag{11.42b}$$

where

$$Q_1 = \omega_s\frac{L_1 + L_s}{R_1 + R_s} \qquad Q_2 = \omega_i\frac{L_2 + L_s}{R_2 + R_s}$$

The derivation of these expressions is as follows: From (11.33a) the impedance of the signal circuit at a frequency $\omega_s + \Delta\omega_s$ is

$$\begin{aligned} Z_s &= R_1 + R_s + j(\omega_s + \Delta\omega_s)\left[(L_1 + L_s)\right. \\ &\qquad \left. - \frac{1}{(\omega_s + \Delta\omega_s)^2}\left(\frac{1}{C_p + C_1} + \frac{1}{C_0 - M^2C_0}\right)\right] \\ &= R_1 + R_s + j(\omega_s + \Delta\omega_s)(L_1 + L_s)\left[1 - \frac{\omega_s^2}{(\omega_s + \Delta\omega_s)^2}\right] \\ &= R_1 + R_s + j\frac{L_1 + L_s}{\omega_s + \Delta\omega_s}(2\omega_s\,\Delta\omega_s + \Delta\omega_s^2) \\ &\approx (R_1 + R_s)\left(1 + j2Q_1\frac{\Delta\omega_s}{\omega_s}\right) \end{aligned}$$

A similar derivation holds for Z_i.

From (11.33) we obtain

$$V_1 = I_s(R_1 + R_s)\left(1 + 2jQ_1\frac{\Delta\omega_s}{\omega_s}\right) + \frac{MI_i^*}{(1 - M^2)j(\omega_i + \Delta\omega_i)C_0} \tag{11.43a}$$

$$V_2 = I_i(R_2 + R_s)\left(1 + 2jQ_2\frac{\Delta\omega_i}{\omega_i}\right) + \frac{MI_s^*}{(1 - M^2)j(\omega_s + \Delta\omega_s)C_0} \tag{11.43b}$$

If $R_1 = R_g + R_L$ and the output is taken at the frequency ω_s, the gain is given by

$$G = \frac{4R_g R_L |I_s|^2}{V_1^2}$$

When we solve (11.43) for I_s, we find that (note that $V_2 = 0$ in this case)

$$G = \frac{4R_g R_L}{\left| Z_s - \dfrac{M^2}{(\omega_i + \Delta\omega_i)(\omega_s + \Delta\omega_s)(1 - M^2)^2 C_0^2 Z_i^*} \right|^2} \tag{11.44}$$

Since the pump frequency is fixed, $\Delta\omega_i = -\Delta\omega_s$. The midband gain G_0 is given by (11.37). To determine the bandwidth, we equate G given by (11.44) to $G_0/2$ and solve for $\Delta\omega_s$. When the gain G_0 is high, we find that, to a good approximation,

$$4\left(\frac{\Delta\omega_s}{\omega_s}\right)^2 = \frac{(R_1 + R_s - R)^2}{(R_1 + R_s)^2 (Q_1 + \omega_s Q_2/\omega_i)^2} \tag{11.45}$$

Thus the gain-bandwidth product becomes

$$2\frac{\Delta\omega_s}{\omega_s}\sqrt{G_0} = \frac{2\sqrt{R_g R_L}}{(R_1 + R_s)(Q_1 + Q_2\omega_s/\omega_i)} \tag{11.46}$$

If we assume as a typical case $R_g + R_L = R_1 + R_s$ and $R_g = R_L$ and note that, for high gain, $R_1 + R_s \approx R$, we get

$$2\sqrt{G_0}\frac{\Delta\omega_s}{\omega_s} = \frac{1}{Q_1 + (\omega_s/\omega_i)Q_2}$$

The smallest possible value of Q_1 occurs if $C_p + C_1$ equals zero. In this case

$$Q_1 = \frac{\omega_s(L_1 + L_s)}{R_1 + R_s} \approx \frac{1}{\omega_s(1 - M^2)C_0(R_1 + R_s)} \approx \frac{1}{\omega_s(1 - M^2)C_0 R}$$

Similarly, the smallest possible value of Q_2 is obtained if $C_2 + C_p = 0$, and is

$$Q_2 = \frac{1}{\omega_i(1 - M^2)C_0(R_2 + R_s)}$$

If we refer to (11.38) for R, we now find that $M^2 Q_1 Q_2 = 1$. Thus

$$2\sqrt{G_0}\frac{\Delta\omega_s}{\omega_s} = \frac{1}{Q_1 + \omega_s/M^2\omega_i Q_1}$$

This expression has a maximum value for

$$Q_1 = \frac{1}{M}\sqrt{\frac{\omega_s}{\omega_i}}$$

Hence the maximum gain-bandwidth product is

$$\left(2\sqrt{G_0}\,\frac{\Delta\omega_s}{\omega_s}\right)_{\max} = \frac{M}{2}\sqrt{\frac{\omega_i}{\omega_s}} \tag{11.47}$$

For 20-dB gain we obtain a bandwidth of

$$\frac{2\,\Delta\omega_s}{\omega_s} = \frac{M}{20}\sqrt{\frac{\omega_i}{\omega_s}}$$

Usually M is no greater than about 0.2, and consequently the fractional bandwidth in percent is approximately equal to $\sqrt{\omega_i/\omega_s}$. In a practical amplifier the gain-bandwidth product will be less since the capacitances $C_1 + C_p$ and $C_2 + C_p$ cannot be reduced to zero.

The parametric amplifier may, of course, be broadbanded by using broadband circuits at the signal and idler frequencies. An alternative scheme for obtaining broadband operation is the traveling-wave parametric amplifier, where resonant circuits are avoided entirely. In the traveling-wave amplifier a waveguiding system is loaded periodically with varactor diodes. With the application of pump power and signal power, mixing occurs, with resultant power generation at the signal frequency. For a detailed analysis the references cited at the end of this chapter may be consulted.

11.6 NOISE PROPERTIES OF PARAMETRIC AMPLIFIERS

The noise produced by parametric amplifiers is the thermal noise in the resistances in the equivalent circuit. For the up-converter illustrated in Fig. 11.6, the input thermal noise at frequency ω_s is that arising from the generator resistance R_g. In analyzing the noise power in a circuit, it is useful to consider noise as consisting of a spectrum of noise waveforms with an effective root-mean-square (rms) voltage $e(\omega)$ equal to the square root of the noise power spectral density. Thus, for the generator resistance R_g, the mean-square noise voltage across R_g is $e_1^2 = 4kTR_g\,\Delta f$. If we neglect circuit losses compared with the diode resistance R_s (this is a reasonably good approximation), the only other noise originating in the signal circuit is that generated in R_s across which a mean-square noise voltage $e_2^2 = 4kT_sR_s\,\Delta f$ at frequency ω_s exists. We denote the equivalent noise temperature of the diode at frequency ω_s by T_s. The noise at frequency ω_s in the signal circuit is amplified and converted to noise at the output frequency ω_0.

The amplified output noise at midband is given by (replace V_1 by $\sqrt{e_1^2 + e_2^2}$ in Fig. 11.6)

$$P_1 = |I_1|^2 R_L = \frac{R_L(e_1^2 + e_2^2)M^2\omega_0^2 C_0^2(1 - M^2)^2}{\left[M^2 + (R_g + R_s)(R_L + R_s)\omega_0\omega_s(1 - M^2)^2 C_0^2\right]^2}$$

where the noise current I_1 is obtained from (11.26), with V_1 replaced by $(e_1^2 + e_2^2)^{1/2}$. If we introduce the midband gain G_0, we obtain [see (11.27)]

$$P_1 = \frac{e_1^2 + e_2^2}{4R_g} G_0 = \frac{k\,\Delta f(R_g T + R_s T_s)}{R_g} G_0 \tag{11.48}$$

There is also noise generated in the output circuit at frequency ω_0 in the resistance R_s. If the diode noise temperature at the frequency ω_0 is T_0, an equivalent noise voltage $e_3^2 = 4kT_0 R_s \Delta f$ appears across R_s in the output circuit. When we solve for the noise current I_2 that results and compute $P_2 = |I_2|^2 R_L$, we obtain

$$P_2 = \frac{\omega_s^2(1 - M^2)^2 C_0^2(R_s + R_g)^2 G_0 e_3^2}{4R_g M^2}$$

$$= \frac{\omega_s^2(1 - M^2)C_0^2(R_s + R_g)^2 kT_0 R_s \Delta f G_0}{R_g M^2} \tag{11.49}$$

The total output noise power at frequency ω_0 is $P_n = P_1 + P_2$. The available input noise power from R_g is $kT\Delta f$. Hence the noise figure is given by ($T = 290°$ in the definition for F)

$$F = \frac{P_1 + P_2}{G_0 kT\Delta f} = 1 + \frac{R_s T_s}{R_g T} + \frac{\omega_s^2(1 - M^2)^2 C_0^2(R_s + R_g)^2 R_s T_0}{R_g M^2 T} \tag{11.50}$$

Note that the gain G_0 was defined as output power divided by the available power from the input signal source, so that F as evaluated above conforms to the accepted definition of noise figure. If we take the diode noise temperature at the two frequencies ω_s and ω_0 equal to T_d, we obtain

$$F = 1 + \frac{R_s}{R_g}\frac{T_d}{T}\left[1 + \frac{(R_s + R_g)^2}{M^2 Q^2 R_s^2}\right] \tag{11.51}$$

after introducing the diode Q from (11.29). For maximum gain,

$$R_g = R_L = R_s\left(1 + \frac{\omega_s}{\omega_0} M^2 Q^2\right)^{1/2}$$

from (11.30). Thus the noise figure under maximum-gain conditions is given by

$$F = 1 + \frac{T_d}{T}\left(1 + \frac{\omega_s}{\omega_0}M^2Q^2\right)^{-1/2}\left(1 + \frac{\left\{1 + \left[1 + (\omega_s/\omega_0)M^2Q^2\right]^{1/2}\right\}^2}{M^2Q^2}\right)$$

$$\approx 1 + \frac{T_d}{T}\left(1 + \frac{\omega_s}{\omega_0}M^2Q^2\right)^{-1/2} \tag{11.52}$$

As a typical example consider $T_d = T$, $\omega_0 = 10\omega_s$, and $MQ = 10$. We then find that $F = 1.36$, or 1.3 dB. This example clearly demonstrates the low-noise property of the parametric amplifier.

The diode noise temperature T_d is somewhat greater than the ambient temperature T because of shot noise that arises from random motion of the carriers across the junction. In addition, the resistance R_s is usually not exactly the same at the two frequencies. However, this difference can be taken into account by choosing appropriate values of T_0 and T_s. The thermal noise arising from circuit losses can be included in (11.50) very simply by replacing R_sT_s by $R_{1l}T + R_sT_s$ and R_sT_0 by $R_{2l}T + R_sT_0$, where R_{1l} and R_{2l} represent the equivalent-circuit resistances. However, since the effective Q of most diodes, that is, $[(1 - M^2)\omega_sC_0R_s]^{-1}$, would rarely exceed 50 whereas circuit Q's would normally be 1,000 or more, the resistances R_{1l} and R_{2l} are negligible compared with the diode resistance R_s.

The noise properties of negative-resistance parametric amplifiers have been analyzed and measured by Uenohara† and others. The noise theory for a negative-resistance parametric amplifier employing a circulator, and with the output signal taken at the same frequency ω_s as the input, is presented below.

With reference to Fig. 11.10, the following sources of noise are the main ones that need to be considered: (1) input noise from $R_g = Z_c$ at temperature T_1 and frequency ω_s; (2) input noise at the idler frequency $\omega_i = \omega_p - \omega_s$ that arises in R_2; (3) noise arising in the diode resistance R_s at frequency ω_s (equivalent noise temperature T_s); and (4) noise arising in R_s at the idler frequency ω_i (equivalent noise temperature T_i). It is important to consider noise sources at the idler frequency because these noise signals are converted into noise at the frequency ω_s by frequency mixing that takes place in the diode. The equivalent circuit and noise voltage sources are shown in Fig. 11.10.

†M. Uenohara, Noise Considerations of the Variable Capacitance Parametric Amplifier, *Proc. IRE*, vol. 48, pp. 169–179, February, 1960.

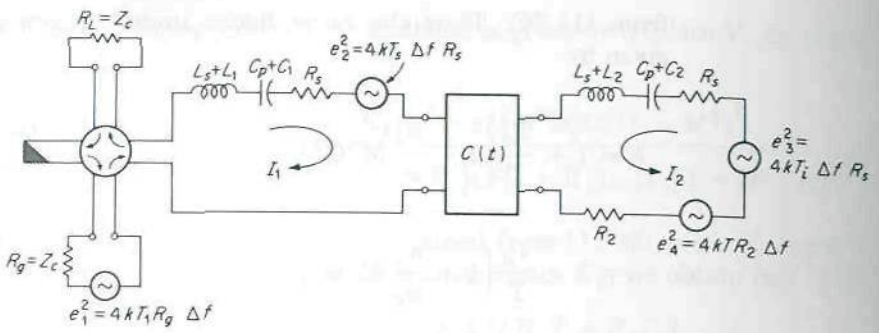


FIGURE 11.10
Noise circuit for a negative-resistance parametric amplifier.

The noise-power input from R_g is $kT_1\,\Delta f$. This will appear at the output load R_L as amplified noise of amount

$$P_1 = G_0 kT_1\,\Delta f = \frac{4R_L^2 kT_1\,\Delta f}{(R_L + R_s - R)^2} \tag{11.53}$$

where G_0 was obtained from (11.39) for the high-gain case.

The noise contributed by the amplifier is represented by the voltage sources e_2, e_3, and e_4. Since the three noise sources are uncorrelated, the noise powers add. Hence the effective noise voltage in the idler circuit is $e_i = \sqrt{e_3^2 + e_4^2}$. The equations describing the noise currents I_1 and I_2 at midband are obtained from (11.34). Thus

$$e_2 = (R_L + R_s)I_1 - \frac{jMI_2^*}{\omega_i(1 - M^2)C_0} \tag{11.54a}$$

$$e_i = (R_2 + R_s)I_2 - \frac{jMI_1^*}{\omega_s(1 - M^2)C_0} \tag{11.54b}$$

since the loading R_1 of the signal circuit as seen by the source e_2 is $Z_c = R_L$.

When we solve (11.54) for I_1, we can evaluate the noise power

$$P_2 = |I_1|^2 R_L$$

delivered to the external load R_L. Giving the final results, we find that

$$I_1 = \frac{e_2}{R_L + R_s - R} + \frac{jMe_i^*}{\omega_i(1 - M^2)C_0(R_L + R_s - R)(R_2 + R_s)}$$

We must find the noise power contributed by e_2 and e_i separately, since

noise voltages do not add. From e_2 we obtain

$$P(e_2) = \frac{e_2^2 R_L}{(R_L + R_s - R)^2}$$

and from e_i we obtain

$$P(e_i) = \frac{e_i^2 R_L M^2}{\left[\omega_i(1 - M^2)C_0(R_L + R_s - R)(R_2 + R_s)\right]^2}$$

$$= \frac{e_i^2 R_L R \omega_s}{\omega_i (R_L + R_s - R)^2 (R_2 + R_s)}$$

Hence

$$P_2 = P(e_2) + P(e_i) = \frac{G_0}{4R_L}\left(e_2^2 + \frac{\omega_s}{\omega_i}\frac{e_i^2 R}{R_2 + R_s}\right)$$

$$= \frac{G_0 k\,\Delta f}{R_L}\left[R_s T_s + \frac{\omega_s}{\omega_i}\frac{R}{R_2 + R_s}(R_s T_i + R_2 T)\right] \qquad (11.55)$$

If we assume that $T_s = T_i = T = T_d$, we obtain the simplified expression

$$P_2 = G_0 k\,\Delta f T_d \frac{R_s}{R_L}\left(1 + \frac{\omega_s}{\omega_i}\frac{R}{R_s}\right) \qquad (11.56)$$

The corresponding noise figure F is given by

$$F = \frac{P_1 + P_2}{kT_1\,\Delta f G_0} = 1 + \frac{T_d}{T_1}\frac{R_s}{R_L}\left(1 + \frac{\omega_s}{\omega_i}\frac{R}{R_s}\right) \qquad (11.57)$$

where T_1 must be taken to be 290 K for the standard definition of F. In (11.57) R may be replaced by $R_L + R_s$ since this is the requirement for high gain. We then obtain

$$F = 1 + \frac{T_d}{T_1}\left(\frac{\omega_p}{\omega_i}\frac{R_L + R_s}{R_L} - 1\right) \qquad (11.58)$$

As an example, assume $T_d = T_1$, choose $\omega_p = 1.5\omega_i$, and let R_L be much greater than R_s. We then obtain a noise figure of 1.5, or 1.76 dB.

The foregoing analysis is valid only when the two frequencies ω_s and ω_i are spaced by an amount greater than the passband of the signal and idler circuits. When this is not the case, $\omega_s \approx \omega_i$, and the amplifier is a degenerate negative-resistance amplifier. Noise current from the input source resistance R_g at the frequency ω_i will now exist in the signal circuit. Similarly, noise currents at the frequencies ω_i and ω_s arising from R_s will be present in both the signal and idler circuits. In the degenerate amplifier the signal and idler cavities are identical and the equivalent circuit is that shown in Fig. 11.11. If the amplifier passband is symmetrical about $\omega_p/2$,

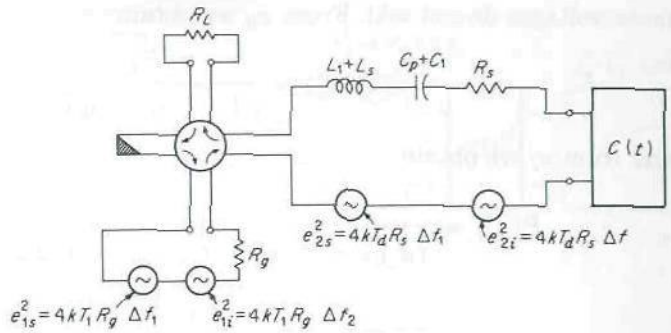


FIGURE 11.11
Noise circuit for a degenerate parametric amplifier.

noise at the frequency $\omega_s < \omega_p/2$ is amplified to give noise output at the frequency ω_s, and also converted and amplified to give noise-power output at the frequency $\omega_i = \omega_p - \omega_s$. The two frequencies ω_s and ω_i are symmetrically located about $\omega_p/2$ and fall within the amplifier passband. Therefore we must consider the noise from R_g and R_s at the two frequencies ω_s and ω_i as separate uncorrelated noise in evaluating the total output noise power. The present situation is illustrated schematically in Fig. 11.12. The passband is split into two equal parts Δf_1 and Δf_2, located symmetrically on either side of $\omega_p/2$. Noise in the band Δf_1 is amplified to give noise in the same band Δf_1 at the output. In addition, noise in the band Δf_1 is amplified and converted into noise in the band Δf_2 at the output. The lower sideband Δf_1 will be regarded as the signal sideband.

Since $\omega_s \approx \omega_i$, the noise power in the band Δf_1 arising from R_g and R_s produces the same amount of noise power in the output bands Δf_1 and Δf_2 as does the noise power arising from R_g and R_s in the band Δf_2. Hence we need to consider only the noise in the lower band Δf_1 in detail.

The noise from R_g in the band Δf_1 is represented by an equivalent voltage source e_{1s} in Fig. 11.11. To evaluate the noise power delivered to R_L from e_{1s}, we shall make the approximation of taking the series impedance of the circuit as $R_s + R_L$ throughout the passband. We denote the noise currents at frequencies ω_s and ω_i by I_1 and I_2.

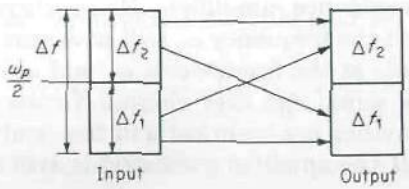


FIGURE 11.12
Illustration of noise conversion.

At the signal frequency ω_s the available gain is $G_0 = |\Gamma|^2$. Under high-gain conditions $R_L + R_s \approx R$; so (11.39) may be used for G_0. The available input noise power from e_{1s} is $e_{1s}^2/4R_g = e_{1s}^2/4R_L$, since we assume that $R_g = R_L = Z_c$. The noise power P_{1s} delivered to R_L in the band Δf_1 is thus

$$P_{1s} = G_0 \frac{e_{1s}^2}{4R_L} = G_0 kT_1 \Delta f_1 \tag{11.59}$$

When there is no impressed voltage at the frequency ω_i, the currents I_1 and I_2 are related by [see (11.34) with $R_2 = R_L$]

$$(R_L + R_s) I_2 = \frac{jMI_1^*}{\omega_s(1 - M^2)C_0}$$

The power P_{1i} delivered to R_L in the band Δf_2 is given by

$$P_{1i} = |I_2|^2 R_L = \frac{|I_2|^2}{|I_1|^2} P_{1s} = \frac{R}{R_L + R_s} P_{1s} \tag{11.60}$$

since $R = M^2/[\omega_s^2(1 - M^2)^2 C_0^2(R_L + R_s)]$ when $\omega_s \approx \omega_i$.

The source e_{1i} delivers an amount of power equal to P_{1s} to R_L in the band Δf_2 and an amount equal to P_{1i} into the band Δf_1. Hence the total noise power delivered to R_L from R_g is the same in the two bands Δf_1 and Δf_2, and is given by

$$P_1 = P_{1s} + P_{1i} = G_0 kT_1 \Delta f_1 \left(1 + \frac{R}{R_L + R_s}\right) = 2G_0 kT_1 \Delta f_1 = G_0 kT_1 \Delta f \tag{11.61}$$

since $R \approx R_L + R_s$ and $2\Delta f_1 = \Delta f$.

To evaluate the noise power delivered to R_L from R_s, we consider the noise in the band Δf_1 first. The equivalent voltage source is e_{2s}, and the circuit equations are (we can put $\omega_i = \omega_s$)

$$e_{2s} = (R_L + R_s) I_1 - \frac{jMI_2^*}{\omega_i(1 - M^2)C_0} \tag{11.62a}$$

$$0 = (R_L + R_s) I_2 - \frac{jMI_1^*}{\omega_s(1 - M^2)C_0} \tag{11.62b}$$

The output noise in the two bands Δf_1 and Δf_2 is given by

$$P_{2s} = |I_1|^2 R_L \qquad P_{2i} = |I_2|^2 R_L$$

When we solve for I_1 and I_2, we obtain

$$P_{2s} = \frac{e_{2s}^2 R_L}{(R_L + R_s - R)^2} = \frac{e_{2s}^2 G_0}{4R_L} = \frac{kT_d \Delta f_1 R_s G_0}{R_L} \tag{11.63}$$

The output noise from e_{2s} in the band Δf_2 is given by

$$P_{2i} = \frac{R}{R_L + R_s} P_{2s} \approx P_{2s} \tag{11.64}$$

The noise source e_{2i} contributes a noise power P_{2s} in the band Δf_2 and P_{2i} in the band Δf_1. Hence the total noise power delivered to R_L from the internal amplifier noise sources is (in each band Δf_1 and Δf_2)

$$P_2 = P_{2s} + P_{2i} = \frac{2kT_d \Delta f_1 R_s G_0}{R_L} = \frac{kT_d R_s G_0 \Delta f}{R_L} \tag{11.65}$$

The total output noise from both R_g and R_s is

$$P_n = P_1 + P_2 = k \Delta f G_0 \left(T_1 + \frac{R_s}{R_L} T_d \right) \tag{11.66}$$

If the degenerate amplifier is used as a single-sideband amplifier (signal input in the lower band Δf_1 only), the single-sideband noise figure F_{SSB} is defined by the ratio of the total output noise power in the band $\Delta f = 2 \Delta f_1$ divided by the available input noise power in the signal band Δf_1. Thus

$$F_{\text{SSB}} = \frac{k \Delta f G_0 T_1}{k \Delta f_1 G_0 T_1} \left(1 + \frac{T_d}{T_1} \frac{R_s}{R_L} \right) = 2 \left(1 + \frac{T_d}{T_1} \frac{R_s}{R_L} \right) \tag{11.67}$$

It is seen that, for single-sideband operation, the noise figure cannot be less than 2, or 3 dB. The signal-to-noise power ratio in the signal band Δf_1 at the input is $P_s/(k \Delta f_1 T_1)$. The signal-to-noise power ratio in the band Δf_1 at the output, given by $G_0 P_s/P_n$, is worse by a factor equal to the single-sideband noise figure F_{SSB}. The noise degradation is due to noise entering in the idler band Δf_2, in which no signal is present.

For double-sideband operation input signal power is present in both bands Δf_1 and Δf_2. In this case the available input noise power is taken to be that in the band Δf. Hence the double-sideband noise figure F_{DSB} is

$$F_{\text{DSB}} = \frac{k \Delta f G_0 T_1}{k \Delta f G_0 T_1} \left(1 + \frac{T_d}{T_1} \frac{R_s}{R_L} \right) = 1 + \frac{T_d}{T_1} \frac{R_s}{R_L} \tag{11.68}$$

and is a factor of 2 (3 dB) better than for the single-sideband case.

The double-sideband noise figure has been measured by Uenohara† for a number of different diodes. It is found that the theory given above is reasonably accurate. Typical noise figures that were measured ranged from 0.9 to 4.5 dB. For diodes with $\omega_s C_0 R_s$ less than 0.1, the noise figure was 2 dB or better. There was a strong correlation between the measured noise figure and the diode quality factor Q. This is predicted by the theory as well.

†*Ibid.*

Since $R_L + R_s \approx R$ for high gain and R can be expressed by

$$R = \frac{R}{R_L + R_s} M^2 Q^2 R_s \gg R_s$$

from (11.29) and (11.38), we see that the optimum value of $R_L + R_s$ is MQR_s. The factor R_s/R_L in the expression for noise figure may now be replaced by $(MQ - 1)^{-1}$ to give

$$F_{\text{DSB}} = 1 + \frac{T_d}{T_1} \frac{1}{MQ - 1}$$

showing that the noise figure improves with diode quality factor Q.

Parametric amplifier noise is primarily thermal noise in the diode resistance R_s. The equivalent amplifier noise temperature is given by $T_A = (F - 1)T_1$, where $T_1 = 290$ K [for the degenerate amplifier with single-sideband operation $T_A = (F - 2)T_1$]. By cooling the amplifier to liquid-nitrogen temperature, noise temperatures below 100 K have been obtained.

PROBLEMS

11.1. Consider a square-law mixer for which the output current $i(t) = k[v(t)]^2$, where $v(t)$ is the applied voltage and k is a constant. Let a local-oscillator signal $V_p \cos \omega_p t$ and a signal $V_s \cos \omega_s t$ be applied, with $V_s \ll V_p$. Show that the output current at the sum or difference frequencies $\omega_p \pm \omega_s$ is a linear function of V_s when $V_s \ll V_p$. Thus the square-law mixer is a linear converter for small-signal amplitudes.

11.2. Consider a parallel-plate capacitor with capacitance C_0. Let a voltage $V = V_s \cos \omega_s t$ be applied. At time $t = 0$ the plate separation is suddenly increased to change the capacitance from C_0 to $C = C_0 - \Delta C$. Since the charge cannot change instantaneously, the voltage must increase. At time $t = (4f_s)^{-1}$, when $V = 0$ and $Q = 0$, let the plate separation be brought back to its original value. There is no change in V produced since $V = 0$ at this time. When $t = (2f_s)^{-1}$, let the capacitance be suddenly decreased to a value C again. When this process is continued, the resultant voltage across C is amplified and will have the waveform illustrated. This is an example of a linear capacitance varied at a rate twice that of the signal frequency. Evaluate the incremental change in voltage that occurs every half cycle and the power supplied by the pump. To evaluate the latter, determine the change in stored energy that occurs every time C is suddenly decreased.

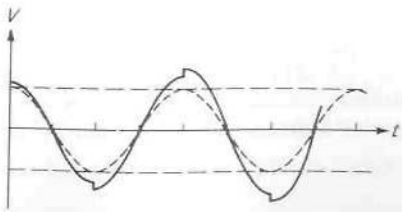


FIGURE P11.2

11.3. A down-converter is a parametric amplifier with an input signal at the frequency $\omega_0 = \omega_s + \omega_p$ and the output signal taken at frequency ω_s. Using the circuit of Fig. 11.6, show that the down-converter gain (actually a loss) is given by (11.27), with ω_0 and ω_s interchanged.

11.4. Derive (11.32).

11.5. Derive the expression (11.35) for the gain of the negative-resistance parametric amplifier.

11.6. Derive (11.49) for the noise power P_2.

11.7. Obtain an expression for the equivalent noise temperature of a parametric up-converter.

11.8. Derive an expression for the gain of the negative-resistance degenerate parametric amplifier illustrated in Fig. 11.11.

11.9. Consider the degenerate parametric amplifier with circulator shown in Fig. 11.11. Assume an input generator voltage V_s at frequency ω_s in place of e_{1s} and e_{1i}. The generator sends a wave with current I_1^+ into the amplifier, where I_1^+ must be equal to $V_s/(R_g + Z_c) = V_s/2Z_c$, when $R_g = Z_c$, since V_s sees a matched load. A reflected wave is set up with a current $I_1^- = -\Gamma I_1^+$. The load current in R_L is I_1^-, apart from a phase angle. The total amplifier current at frequency ω_s is $I_1 = I_1^- + I_1^+$. With this information determine the appropriate circuit equations, analogous to (11.34), for I_1 and I_2.

11.10. A parametric diode has the following parameter values: $C_0 = 2$ pF, $R_s = 1\ \Omega$. The modulation index $M = 0.25$. The frequency $f_s = 5{,}000$ MHz and $f_p = 12{,}000$ MHz. Evaluate the diode effective Q. Determine the load resistance R_L to give a gain of 20 dB for a negative-resistance amplifier of the form shown in Fig. 11.8. Assume $R_g = R_L = R_2$. Calculate R for 20-dB gain.

REFERENCES

1. Blackwell, L. A., and K. L. Kotzebue: "Semiconductor-Diode Parametric Amplifiers," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1961.
2. Penfield, P., and R. P. Rafuse: "Varactor Applications," The M.I.T. Press, Cambridge, Mass., 1962.
3. Chang, K. K. N.: "Parametric and Tunnel Diodes," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

Traveling-wave parametric amplifiers

4. Cullen, A. L.: A Traveling Wave Parametric Amplifier, *Nature*, vol. 181, February, 1958.
5. Honey, R. C., and E. M. T. Jones: A Wide-Band UHF Traveling Wave Variable Reactance Amplifier, *IRE Trans.*, vol. MTT-8, pp. 351–361, May, 1960.
6. Heilmeier, G. H.: An Analysis of Parametric Amplification in Periodically Loaded Transmission Lines, *RCA Rev.*, vol. 20, pp. 442–454, September, 1959.

Broadbanding techniques

7. Matthaei, G. L.: A Study of the Optimum Design of Wideband Parametric Amplifiers and Up-Converters, *IRE Trans.*, vol. MTT-9, pp. 23–28, January, 1961.
8. Gilden, M., and G. L. Matthaei: Practical Design and Performance of Nearly Optimum Wide Band Degenerate Parametric Amplifiers, *IRE Trans.*, vol. MTT-9, pp. 484–490, November 1961.

CHAPTER 12

OSCILLATORS AND MIXERS

In this chapter we will examine the operating characteristics of two types of negative-resistance solid-state devices, namely, Gunn devices and IMPATT diodes. These devices are widely used for low-power oscillators in microwave and millimeter-wave transmitters. The Gunn device is useful as a local oscillator in receiver front ends. We will also discuss the use of bipolar and FET transistors in oscillators.

The last part of the chapter provides an introductory treatment of mixers. A mixer is a nonlinear device, very often a diode or several diodes in a bridge arrangement, that will cause the microwave signal and local-oscillator (LO) signal to mix to produce a translation of the signal spectrum to a lower frequency called the intermediate frequency (IF). All superheterodyne receivers use a mixer for this purpose. The advantage gained is that the amplification of the signal, before demodulation, is more easily accomplished at the lower fixed IF frequency. The IF amplifier also establishes the bandwidth of the system since selectivity in the high-frequency RF amplifier is usually low. The main purpose of the RF amplifier is to increase the signal amplitude to a level such that the mixer noise will not produce a significant degradation of the signal-to-noise ratio.

Apart from economic factors the requirements of an oscillator include:

1. Excellent frequency stability, i.e., negligible variation in the frequency of oscillations due to variations in temperature, power-supply voltage, and oscillator loading (load pulling).
2. Adequate power output for the intended use.
3. Low amplitude-, phase-, and frequency-modulation noise.

4. Variable tuning, including mechanical tuning and voltage control.
5. Capability to be modulated in amplitude (AM), frequency (FM), or phase (PM).
6. Simple circuit requirements.

A variable-frequency oscillator (VFO) whose frequency is varied by means of an applied control voltage is called a voltage-controlled oscillator (VCO).

Good frequency stability and low noise are obtained by employing a high-Q resonator in the oscillator circuit. The resonator should have a resonant frequency that is insensitive to variations in the ambient temperature, and this usually translates into low thermal expansion of the resonator. For critical applications the oscillator may be placed in a temperature-controlled oven.

The frequency of an oscillator will vary with the dc bias voltages that are applied. This effect is called oscillator pushing and can sometimes be used to advantage to fine-tune an oscillator over a narrow band of frequencies.

The frequency of oscillation is determined by the resonant frequency of the input and output networks. Consequently, any change in the impedance of the load connected to the oscillator will result in a change in the oscillator frequency. This effect, which is referred to as oscillator load pulling, is usually undesirable. The pulling effect can be minimized by using loose coupling between the oscillator and the load, i.e., the external Q should be large. A disadvantage of using a loosely coupled load impedance is that the output power will be smaller and the oscillator efficiency will be reduced. Load pulling can also be minimized by using a very high Q resonator as the main frequency-determining element in the oscillator circuit.

Oscillator pulling can be determined by measuring the change in oscillator frequency as a function of the phase angle of the load reflection coefficient.

12.1 GUNN OSCILLATORS

Some bulk semiconductor materials, such as gallium arsenide (GaAs), indium phosphide (InP), and cadmium telluride (CdTe), have two closely spaced energy bands in the conduction band. A typical energy versus momentum band structure is shown in Fig. 12.1. At low electric field strengths in the material, most of the electrons will be located in the lower-energy band. At high electric field strengths, most of the electrons will be transferred into the high-energy band. In the high-energy band the effective electron mass is larger and hence the electron mobility is lower than what it is in the low-energy band. Since the conductivity is directly proportional to the mobility, there is an intermediate range of electric field

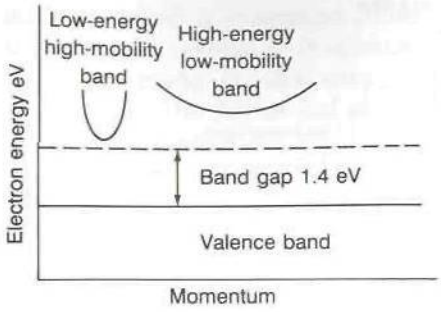


FIGURE 12.1
Typical double-energy conduction band for a Gunn material such as GaAs.

strengths for which the fraction of electrons that are transferred into the high-energy low-mobility conduction band is such that the average mobility, and hence conductivity, decreases with an increase in electric field strength. Thus there is a range of applied voltages over which the current decreases with increasing voltage and a negative incremental resistance is displayed by the device. A typical current-voltage characteristic for a Gunn device is shown in Fig. 12.2. A Gunn device is also called a transferred-electron device since the negative resistance arises from the transfer of electrons *from the low- to the high-energy band.*

The oscillations that occur in materials with the energy band structure noted above was discovered by J. B. Gunn. The possibility of obtaining negative differential resistance had been predicted earlier by Ridley and Watkins.

There are two principal modes of operation that result in oscillations for a Gunn device. When the applied voltage exceeds the threshold value, a dipole domain (a region of electron concentration and depletion) forms near the cathode end with most of the voltage drop appearing across the high-resistance part of the device. A short section of the input region is in the low-energy high-mobility state and electrons leave the cathode with a large

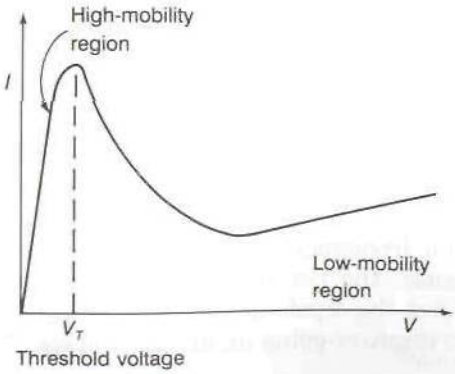


FIGURE 12.2
Current-voltage characteristics for a Gunn device. Note the negative-resistance region.

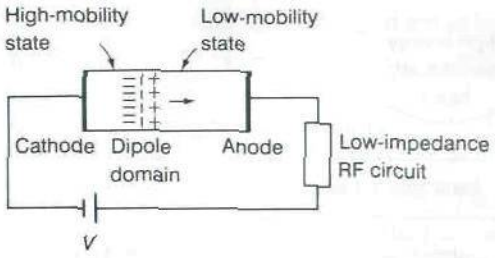


FIGURE 12.3
A simple Gunn Oscillator circuit using the transit-time mode of oscillation.

velocity. At the point in the material that separates the high-mobility and low-mobility states, electrons accumulate on the left side and are depleted on the right side by virtue of the different mobilities. This dipole arrangement of charge is shown pictorially in Fig. 12.3. This dipole domain sweeps across the device, and when it arrives at the anode, the device is in a high-mobility state and a new dipole domain forms at the cathode end and moves toward the anode. This mechanism is self-repeating and represents an oscillation with a period equal to the transit time. This mode of oscillation has a low efficiency (a few percent) of power generation and a frequency that cannot be controlled by the external circuit. This mode of oscillation is called the transit-time mode or Gunn mode.

The second mode of oscillation is the limited-space-charge (LSA) mode. Operation of a Gunn oscillator in the LSA mode can produce several watts of power with efficiencies of around 20 percent or more. The power outputs that have been obtained decrease with frequency and are below 1 W at frequencies greater than 10 GHz. Output power of several milliwatts can be obtained at 100 GHz.

In the LSA mode the Gunn device is incorporated as part of a resonant circuit as shown in Fig. 12.4*a*. The frequency of the resonant circuit is adjusted so that it is several times greater than that of the transit-time mode. As a consequence, dipole domains do not have sufficient time to form and the device operates essentially as a negative-resistance device. The dc bias is adjusted to a value somewhat greater than the threshold voltage. The RF voltage of the oscillations will build up to a peak-peak value approximately equal to the voltage increment over which the device resistance is negative as shown in Fig. 12.4*b*. The resonator loading, represented by the resistor R, is adjusted to a value about 20 percent greater than the maximum negative resistance of the device. This will ensure that oscillations will start. The amplitude of the oscillations will build up until the average negative resistance of the Gunn device becomes equal to the resonator resistance R.

If the resonator frequency is adjusted to a value slightly above that of the transit-time mode, the Gunn device will operate very much like the basic Gunn mode, but the dipole domain will be quenched before it arrives at the anode by the negative-going oscillation voltage. This type of operating

mode is called a quenched-domain mode. Oscillations can also occur by adjusting the resonator frequency, so that it is lower than the frequency of the Gunn mode. In this case the dipole domains have sufficient time to sweep across the device and arrive at the anode. However, the initiation of a new dipole domain is delayed until the oscillation voltage rises above the threshold value. This mode of operation is called the inhibited or delayed mode.

Gunn Oscillator Circuits

The equivalent circuit of a Gunn device operating in the LSA mode is a negative resistance $-R_d$ in parallel with a capacitance C_d as shown in Fig. 12.5. The negative resistance has a value that typically lies in the range -5 to $-20\ \Omega$. The required resistive loading from the cavity and the external load should be around 20 percent higher than the Gunn device resistance so that the parallel combination $-RR_d/(R - R_d)$ will be negative. The cavity used for the resonator must generally have an impedance-transforming property in order to reduce the high impedance of the output waveguide to

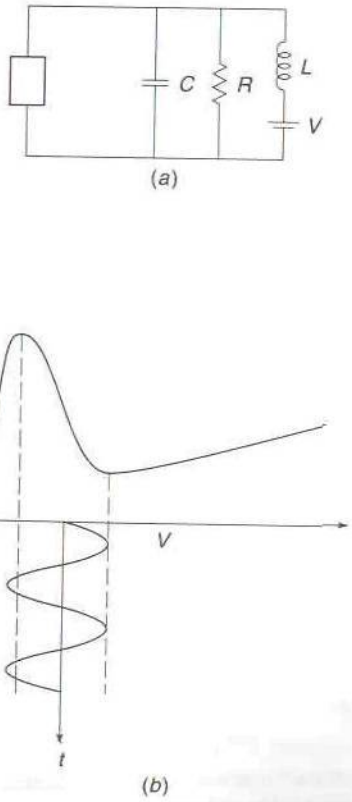


FIGURE 12.4
(*a*) A basic Gunn oscillator operating in the LSA mode; (*b*) RF oscillating voltage across a Gunn device operating in the LSA mode.

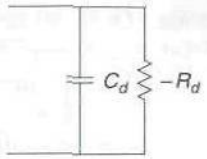


FIGURE 12.5
Equivalent circuit of a Gunn device operating in the LSA mode.

the appropriate low value required by the Gunn device. A simple cavity structure is shown in Fig. 12.6. The Gunn device is located under a post in a rectangular waveguide. The cavity is resonated at the desired frequency by adjusting the short-circuit position. The degree of coupling to the external waveguide is adjusted by changing the window opening in the inductive diaphragm located at the front of the cavity. The top of the post is insulated from the waveguide. The dc bias voltage (typically around 12 V) is applied to the post. There is sufficient capacity between the post and the surrounding waveguide to provide an adequate low-impedance RF bypass capacitance and thus RF currents do not flow through the bias voltage supply. Fine tuning of the cavity can be obtained by means of a tuning screw.

Another simple cavity arrangement for a Gunn oscillator is shown in Fig. 12.7. In this cavity the high impedance of the waveguide is transformed into a low impedance at the location of the Gunn device by means of quarter-wave transformers. The cavity resonant frequency can be adjusted by changing the location of the short circuit. A tuning screw can be used for fine tuning of the cavity.

The cavity shown in Fig. 12.6 is easily modified to have two posts, one for mounting the Gunn device and a second one for mounting a varactor diode. The capacitance of the varactor diode is a function of the control voltage V_c. By varying V_c the resonant frequency of the cavity can be varied.

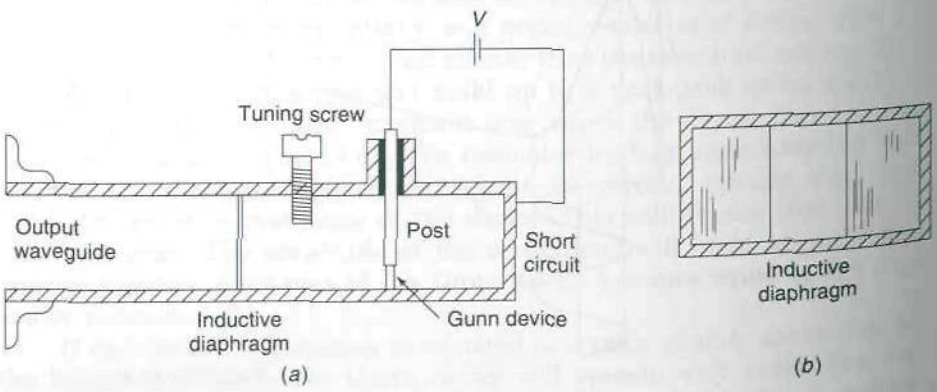


FIGURE 12.6
(*a*) A simple waveguide cavity for a Gunn oscillator; (*b*) inductive diaphragm used to adjust the coupling between the output waveguide and cavity.

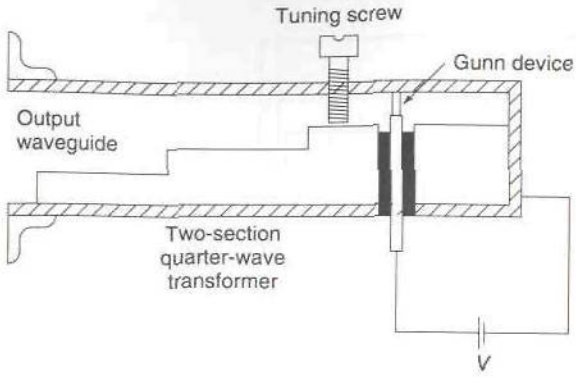


FIGURE 12.7
A Gunn oscillator cavity which uses a two-section quarter-wave transformer to transform the high impedance of the waveguide to a low impedance at the Gunn device.

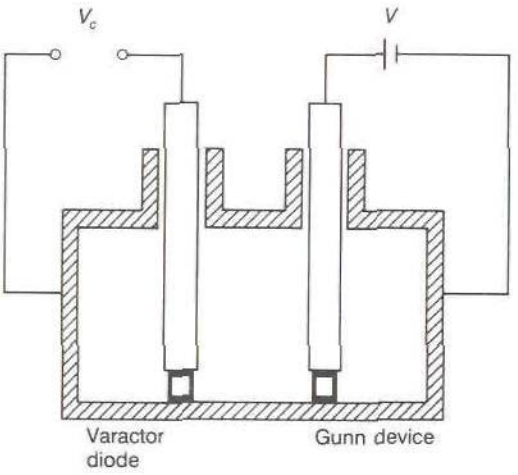


FIGURE 12.8
A Gunn oscillator cavity which has a post-mounted varactor diode for cavity tuning. A sawtooth sweep voltage applied to the varactor diode will produce frequency modulation of the oscillator.

If a sawtooth sweep voltage is applied to the varactor diode, the Gunn oscillator will be frequency-modulated. The cross section of the cavity is shown in Fig. 12.8.

The Gunn device can be operated as a pulsed oscillator by applying the dc bias voltage in the form of a pulse train of short rectangular pulses. If the duty cycle is low enough and very short bias pulses are used, the peak power output will be limited only by the peak current, since thermal heating of the device will be negligible during the short on time. For pulsed oscillator applications, the IMPATT diode, or variations of it, is preferred because of higher output power. The Gunn device can also be, and has been, used as a negative-resistance amplifier.

12.2 IMPATT DIODES

The acronym IMPATT stands for *IMP*act Ionization *A*valanche *T*ransit *T*ime and describes the phenomenon associated with reverse voltage breakdown in a *p-n* junction diode and the transport or transit of charge carriers

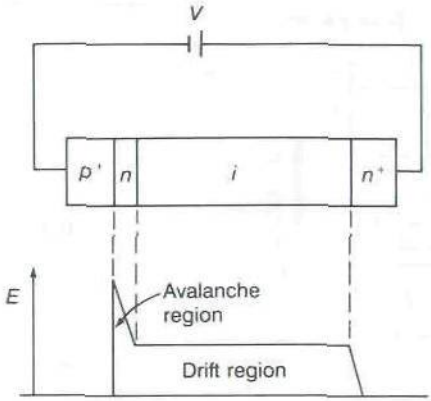


FIGURE 12.9
Structure of a Read (IMPATT) diode and the electric field profile across the diode.

through a drift region. W. T. Read had proposed in 1958 that there should be a phase delay of more than 90° between an applied RF voltage and the avalanching current if the RF voltage caused the total voltage to exceed the reverse breakdown voltage in a diode. In 1965, R. L. Johnson verified Read's prediction. When the current lags the RF voltage by more than 90°, the diode will exhibit a negative resistance and can be used as a source of microwave power in an oscillator circuit.

The Read diode, designated as a p^+nin^+ *type, consists of a heavily* doped p^+ region, a normally doped n region, an undoped or intrinsic semiconductor section, and an n^+ region as shown in Fig. 12.9. The p^+n diode junction will break down when the applied reverse bias voltage exceeds a threshold value. The current-voltage characteristic is shown in Fig. 12.10 and is similar to that in the familiar zener diode. The rapid increase in current at the breakdown voltage is caused by avalanche multiplication of the density of the holes and electrons. If a Read diode is placed in a cavity and a reverse bias somewhat smaller than the breakdown voltage is applied, along with a small RF voltage, then breakdown will occur when the RF voltage becomes positive. When breakdown is initiated a large number of holes and electrons are created at the p^+n junction. The electrons are swept across the n region into the intrinsic semiconductor drift region. After a transit-time delay, the electrons are collected at the n^+ terminal. The current pulse moves through the diode from right to left. *When the time for avalanche charge* buildup plus that for charge transit through the drift region exceeds one-half RF period, the output current will lag the RF voltage by more than 90°. With these conditions the diode will exhibit a negative resistance for RF currents. In an oscillator circuit the initial RF voltage comes from the cavity resonant-frequency component of the noise that excites the cavity. Once the oscillations start they grow in amplitude until the average negative resistance of the diode becomes equal to the total equivalent resistance of the cavity and the external load.

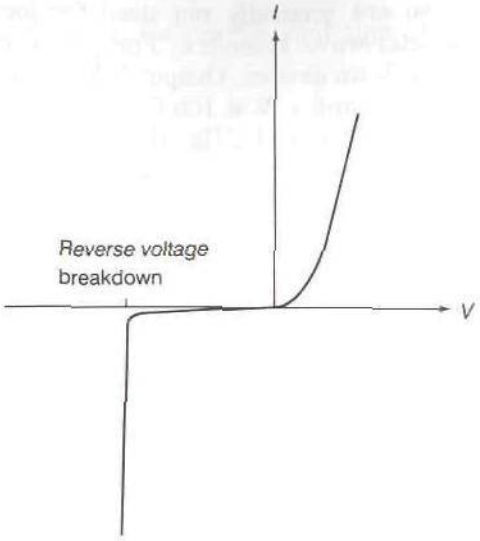


FIGURE 12.10
Current-voltage characteristic of a diode. Note the reverse voltage breakdown region.

Since the introduction of the Read diode as a generator of microwave power, a number of other diode structures have been developed that will also produce microwave oscillations. The two most common variations of the Read diode are the BARRITT (*BAR*rier *I*njection *T*ransit-*T*ime) diode and the TRAPATT (*TRA*pped *P*lasma *A*valanche *T*riggered *T*ransit) diode. The BARRITT diode is a p^+np^+ or back-to-back diode. The charge carriers that traverse the drift region in a BARRITT device consist of minority carriers that are injected from the forward-biased p^+n junctions. Since the BARRITT diode does not involve an avalanche breakdown, it produces less noise than an IMPATT diode does. However, the power output and efficiency is less.

The TRAPATT diode is a p^+nn^+ diode and is driven by a large repetitive pulse of current. Breakdown will occur at one of the p^+n diode junctions, and since the current drive is very large, a large collection of electrons and holes (a plasma) is generated. The violent breakdown creates a high electric field shock front that moves across the n-type drift region. After passage of the shock front, the plasma is located in a low-field region and is said to be trapped because it takes a long time to clear the drift region of charge carriers. When the plasma has been cleared from the drift region, the cycle will repeat. In a TRAPATT diode, oscillations start with the diode operating as an IMPATT device. When the amplitude of oscillation becomes large enough, the TRAPATT mode of oscillation is established. The TRAPATT diode will not operate at as high a frequency as the IMPATT diode does. It is also noisier, so its use is decreasing.

IMPATT diode oscillators are used with the same cavity structures that are used with Gunn devices. IMPATT diodes are much noisier than

Gunn devices and so are generally not used for local oscillators in microwave and millimeter-wave receivers. For power generation, IMPATT diodes are superior to Gunn devices. Output CW powers of as much as 10 W at a few gigahertz and around 1 W at 100 GHz can be obtained from a single device. At frequencies above 100 GHz, the output power from currently produced IMPATT diodes decreases approximately as $1/f^3$.

12.3 TRANSISTOR OSCILLATORS

Silicon bipolar transistors are a good choice for oscillators at frequencies up to 5 GHz. From 5 GHz up to about 40-GHz, MESFET devices can be used in oscillator circuits. In the frequency range 30 GHz up to around 100 GHz, the high-electron-mobility transistor (HEMT) would be used because of its higher frequency of oscillation.

In Chap. 10 we were concerned with the general problem of designing amplifiers that would not oscillate. We noted that many microwave transistors were only conditionally stable and only a restricted range of load and source impedances would ensure amplifier stability. The stable and unstable regions were shown graphically by plotting the input and output stability circles on the Smith chart. In the design of an oscillator, we choose the input and output port terminations in the unstable regions. In the unstable regions both the input and output impedances of the transistor circuit will have a negative resistance and oscillations will occur at a frequency at which the total reactance in the input and output circuits vanish (resonance condition). Transistor oscillators can thus be viewed as negative-resistance oscillators. If a transistor is absolutely stable, it can be made unstable by using feedback from the output to the input of the device. Common base or common gate circuit configurations tend to have the greatest amount of instability as Example 10.5 showed. One suitable feedback arrangement is the use of a reactance in series with the emitter or source lead. A series reactance can also be used in the common base or common gate circuits.

The criteria for a transistor terminated in an impedance $\bar{Z}_s$ at the input port and $\bar{Z}_L$ at the output port to oscillate are readily established. Initially, assume that a voltage source V_g acts in series with $\bar{Z}_s + \bar{Z}_{in}$. The input current to the transistor will be $V_g/(\bar{Z}_s + \bar{Z}_{in}) = I_{in}$. If we now reduce V_g to zero but at the same time make $\bar{Z}_s + \bar{Z}_{in}$ tend to zero, we can maintain the current I_{in} and a finite output power. Thus, for oscillations to occur when $V_g = 0$, we must have

$$\bar{Z}_s = \bar{R}_s + j\bar{X}_s = -\bar{R}_{in} - j\bar{X}_{in}$$

or

$$\bar{R}_{in} = -\bar{R}_s \tag{12.1a}$$

$$j\bar{X}_{in} = -j\bar{X}_s \tag{12.1b}$$

Thus the transistor must have a negative input resistance and the input must be tuned to resonance. The frequency of oscillation is determined by (12.1*b*). From the relations $\Gamma_s = (\bar{Z}_s - 1)/(\bar{Z}_s + 1)$ and $\Gamma_{in} = (\bar{Z}_{in} - 1)/$

$(\bar{Z}_{in} + 1)$, we readily find that the condition for oscillations can also be stated as

$$\Gamma_s \Gamma_{in} = 1 \tag{12.2}$$

Since $\bar{R}_{in}$ is negative $|\Gamma_{in}| > 1$, which is in accordance with our requirement for instability as given in Chap. 10.

We will now show that when (12.1) or (12.2) hold, at the output $\bar{Z}_L = -\bar{Z}_{out}$ or

$$\bar{R}_{out} = -\bar{R}_L \tag{12.3a}$$

$$j\bar{X}_{out} = -j\bar{X}_L \tag{12.3b}$$

$$\Gamma_L \Gamma_{out} = 1 \tag{12.3c}$$

Thus the conditions for oscillations are satisfied at both ports if they are satisfied at one port. The proof of this property is as follows: The following relations hold

$$\Gamma_{out} = \frac{S_{22} - \Delta\Gamma_s}{1 - \Gamma_s S_{11}} = \frac{S_{22} - \Delta/\Gamma_{in}}{1 - S_{11}/\Gamma_{in}} = \frac{S_{22}\Gamma_{in} - \Delta}{\Gamma_{in} - S_{11}}$$

upon using (12.2). We also have

$$\Gamma_{in} = \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L}$$

which can be solved for Γ_L to give

$$\Gamma_L = \frac{\Gamma_{in} - S_{11}}{S_{22}\Gamma_{in} - \Delta} = \frac{1}{\Gamma_{out}}$$

which is the relationship we wanted to prove.

Since the impedance looking into both ports has a negative real part, the transistor will deliver power to the external circuit at both ports. Since it is operating in the unstable region, the power gain is negative as explained in Sec. 10.7. An absolutely stable transistor with a suitable external feedback network can be viewed as a new modified potentially unstable device for which the above relations will hold true. We normally view an oscillator as an amplifier with a feedback loop that feeds a fraction of the output power to the transistor input and the input circuit. Part of the power fed back is absorbed in the resistance associated with the input circuit. When an unstable transistor is used as an oscillator, the feedback takes place internally and is described in terms of the reverse transmission coefficient S_{12}. Thus the power dissipated by the input circuit as well as the input power that drives the oscillator is provided by internal feedback of power from the output to the input of the oscillator. The only significant difference in the two oscillator types is the feedback path which can be either external or internal.

In order for oscillations to start, it is necessary to choose $\bar{R}_s < |\bar{R}_{in}|$ using the small-signal scattering-matrix parameters to evaluate $\bar{R}_{in}$. Oscil-

lations will then build up until the nonlinear characteristics of the transistor cause the power gain to saturate. Thus steady-state oscillations will cause the transistor to operate under large-signal conditions. For large signals the nonlinear behavior means that the large-signal scattering-matrix parameters will be different from those that apply for small-signal conditions. Furthermore, the nonlinear behavior will cause harmonics of the fundamental frequency to be generated. Usually the presence of harmonics in the output is undesirable unless the oscillator is specifically designed for an output at one of the harmonics. The basic problem in oscillator design is choosing the port terminations so that the desired output power is obtained, the harmonics are adequately suppressed, and the desired frequency of oscillation with good stability against variations due to load, temperature, and bias conditions is obtained.

Some general guidelines that should be followed in order to achieve the above objectives are:

1. The loaded Q of the output circuit should be at least 10 to give good harmonic suppression.
2. The input reactance $j\overline{X}_s$ should have a large-frequency derivative or slope $d\overline{X}_s/d\omega$. This will mean that a small change in $j\overline{X}_{in}$ can then be matched by a small change in ω to bring $j\overline{X}_s$ back to $-j\overline{X}_{in}$. Usually this requirement is met by incorporating a high-Q temperature-stable resonator in the input circuit or in the feedback path.
3. For minimum harmonic generation the oscillator should operate as a class A oscillator.
4. For best efficiency and largest amount of output power, the oscillator should be designed for class B or class C operation. This will require a self-bias circuit, so that initially the circuit operates as a class A oscillator in order for oscillations to start.

In class A operation the collector (drain) current flows continuously over a complete RF cycle. In a class B oscillator the current flows for one-half of the RF cycle, while in class C operation the current flows for less than half of one period in each RF cycle. In class C operation the maximum collector (drain) current flows when the RF voltage at the collector (drain) is negative. Thus the power dissipation in the device is minimized and the efficiency (RF output power/dc input power) can be high. At lower frequencies class C efficiencies as high as 80 percent or more can be achieved, but at microwave frequencies the efficiency is much lower because of circuit losses, relatively low power gain, and limited supply voltages to avoid damaging the transistor. The efficiency of microwave oscillators and amplifiers is often described in terms of the power-added efficiency. The power-added efficiency is defined as follows:

$$\textit{Power-added efficiency} = \frac{\text{RF output power} - \text{RF input power}}{\text{dc input power}} \tag{12.4}$$

12.4 THREE-PORT DESCRIPTION OF A TRANSISTOR

In order to facilitate the analysis of an oscillator circuit, when an impedance is connected between one of the transistor terminals and the ground plane, it is useful to have a scattering-matrix description of a transistor viewed as a three-port circuit. In Fig. 12.11 we show a transistor with a microstrip line connected to each terminal. For clarity we have labeled the base terminal as port 1, the collector terminal as port 2, and the emitter terminal as port 3. Any other labeling could be used equally well. The normalized amplitudes of the incident and reflected voltage waves will be denoted by a_1, a_2, a_3 and b_1, b_2, b_3, respectively. The three-port circuit is described by the scattering-matrix relation

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \hat{S}_{11} & \hat{S}_{12} & \hat{S}_{13} \\ \hat{S}_{21} & \hat{S}_{22} & \hat{S}_{23} \\ \hat{S}_{31} & \hat{S}_{32} & \hat{S}_{33} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \tag{12.5}$$

where $\hat{S}_{ij}$ are the three-port scattering-matrix parameters. We will show that the three-port parameters have the property that the sum of all elements in any row or in any column equals unity; thus

$$\sum_{i=1}^{3} \hat{S}_{ij} = 1 \qquad j = 1, 2, 3 \tag{12.6a}$$

$$\sum_{j=1}^{3} \hat{S}_{ij} = 1 \qquad i = 1, 2, 3 \tag{12.6b}$$

Consequently, the three-port scattering-matrix elements are not all independent. In fact, they can be expressed in terms of the two-port scattering-matrix parameters that describe the transistor when one of the terminals is grounded, e.g., common emitter parameters when the emitter is grounded.

In order to prove (12.6*a*) we note that, provided there is negligible stray capacitance between each transistor terminal and the ground plane, *the sum of all currents entering the three-transistor terminals must be* zero. Thus we have

$$\sum_{i=1}^{3} (a_i - b_i) = 0 = \sum_{i=1}^{3} \left(a_i - \sum_{j=1}^{3} \hat{S}_{ij} a_j \right) = 0$$

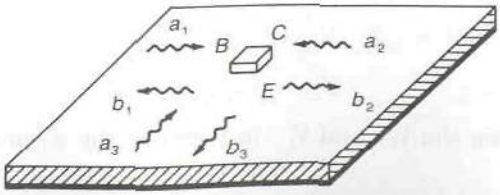


FIGURE 12.11
A transistor viewed as a three-port network.

We can choose the incident-wave amplitudes independently; so if we choose $a_2 = a_3 = 0$, we obtain

$$a_1 - \sum_{i=1}^{3} \hat{S}_{i1} a_1 = 0$$

or

$$\sum_{i=1}^{3} \hat{S}_{i1} = 1$$

which is one of the relations in (12.6a). By choosing in turn a_2 and a_3 as the nonzero amplitude, the other two relations are obtained.

The relations given by (12.6b) are obtained by noting that, when all port (terminal) voltages are equal, the input current at each port will be zero, provided again that there is negligible capacitance between each terminal and the ground plane. Thus we must have $b_i = a_i$ when $a_1 = a_2 = a_3 = a$, and consequently, $b_1 = b_2 = b_3 = a$ and

$$b_i = \sum_{j=1}^{3} \hat{S}_{ij} a_j = \sum_{j=1}^{3} \hat{S}_{ij} a = a$$

which gives the desired result.

The total normalized voltage at port 3 is $V_3 = a_3 + b_3$. Let us define new voltage-wave amplitudes at ports 1 and 2 as follows:

$$V_1^+ = a_1 - \frac{V_3}{2} = a_1 - \frac{a_3 + b_3}{2} \tag{12.7a}$$

$$V_1^- = b_1 - \frac{V_3}{2} = b_1 - \frac{a_3 + b_3}{2} \tag{12.7b}$$

$$V_2^+ = a_2 - \frac{V_3}{2} = a_2 - \frac{a_3 + b_3}{2} \tag{12.7c}$$

$$V_2^- = b_2 - \frac{V_3}{2} = b_2 - \frac{a_3 + b_3}{2} \tag{12.7d}$$

The total new voltages for ports 1 and 2 are now $V_1 = V_1^+ + V_1^- = a_1 + b_1 - V_3$ and $V_2 = a_2 + b_2 - V_3$ and are thus referenced to the port 3 voltage. The above definitions for the new incident- and scattered-wave amplitudes were chosen so as to leave the port currents unchanged, that is, $V_1^+ - V_1^- = a_1 - b_1$, $V_2^+ - V_2^- = a_2 - b_2$. The common emitter (common port 3 terminal) two-port scattering matrix is the scattering matrix that relates V_1^- and V_2^- to V_1^+ and V_2^+; thus

$$V_1^- = S_{11} V_1^+ + S_{12} V_2^+ \tag{12.8a}$$

$$V_2^- = S_{21} V_1^+ + S_{22} V_2^+ \tag{12.8b}$$

Upon expressing the V_i^+ and V_i^- in terms of the a_i and b_i using (12.7), we

obtain

$$b_1 - \frac{a_3 + b_3}{2} = S_{11}\left(a_1 - \frac{a_3 + b_3}{2}\right) + S_{12}\left(a_2 - \frac{a_3 + b_3}{2}\right)$$

$$b_2 - \frac{a_3 + b_3}{2} = S_{21}\left(a_1 - \frac{a_3 + b_3}{2}\right) + S_{22}\left(a_2 - \frac{a_3 + b_3}{2}\right)$$

Since the sum of all currents flowing into the three transistor terminals is zero (Kirchhoff's law), we must also have

$$a_1 - b_1 + a_2 - b_2 + a_3 - b_3 = 0$$

The above three equations are easily rearranged to give three equations expressing the b_i in terms of the a_i. The first two equations can be written in the form

$$b_1 = S_{11}a_1 + S_{12}a_2 + \sigma_{11}\frac{V_3}{2}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 + \sigma_{22}\frac{V_3}{2}$$

where $\sigma_{11} = 1 - S_{11} - S_{12}$ and $\sigma_{22} = 1 - S_{22} - S_{21}$. The third equation can be solved for b_3 to give $b_3 = a_1 + a_2 + a_3 - b_1 - b_2$. When we substitute for b_1 and b_2, we readily find that

$$b_3 = \frac{2\sigma_{12}}{4 - \sigma}a_1 + \frac{2\sigma_{21}}{4 - \sigma}a_2 + \frac{\sigma}{4 - \sigma}a_3 \qquad (12.9a)$$

where $\sigma_{12} = 1 - S_{11} - S_{21}$, $\sigma_{21} = 1 - S_{22} - S_{12}$, $\sigma = S_{11} + S_{12} + S_{21} + S_{22} = 2 - \sigma_{12} - \sigma_{21} = 2 - \sigma_{11} - \sigma_{22}$. By using this expression in the equations for b_1 and b_2, we obtain

$$b_1 = \left(S_{11} + \frac{\sigma_{11}\sigma_{12}}{4 - \sigma}\right)a_1 + \left(S_{12} + \frac{\sigma_{11}\sigma_{21}}{4 - \sigma}\right)a_2 + \frac{2\sigma_{11}}{4 - \sigma}a_3 \qquad (12.9b)$$

$$b_2 = \left(S_{21} + \frac{\sigma_{22}\sigma_{12}}{4 - \sigma}\right)a_1 + \left(S_{22} + \frac{\sigma_{22}\sigma_{21}}{4 - \sigma}\right)a_2 + \frac{2\sigma_{22}}{4 - \sigma}a_3 \qquad (12.9c)$$

Equations (12.9a) to (12.9c) provide the three-port scattering-matrix description of the transistor in terms of the two-port scattering-matrix parameters. The $\hat{S}_{ij}$ are given by

$$\hat{S}_{11} = S_{11} + \frac{\sigma_{11}\sigma_{12}}{4 - \sigma} \qquad \hat{S}_{12} = S_{12} + \frac{\sigma_{11}\sigma_{21}}{4 - \sigma} \qquad \hat{S}_{13} = \frac{2\sigma_{11}}{4 - \sigma}$$

$$\hat{S}_{21} = S_{21} + \frac{\sigma_{22}\sigma_{12}}{4 - \sigma} \qquad \hat{S}_{22} = S_{22} + \frac{\sigma_{22}\sigma_{21}}{4 - \sigma} \qquad \hat{S}_{23} = \frac{2\sigma_{22}}{4 - \sigma}$$

$$\hat{S}_{31} = \frac{2\sigma_{12}}{4 - \sigma} \qquad \hat{S}_{32} = \frac{2\sigma_{21}}{4 - \sigma} \qquad \hat{S}_{33} = \frac{\sigma}{4 - \sigma} \qquad (12.10)$$

where

$$\sigma = S_{11} + S_{12} + S_{21} + S_{22} \qquad \sigma_{11} = 1 - S_{11} - S_{12}$$

$$\sigma_{12} = 1 - S_{11} - S_{21} \qquad \sigma_{22} = 1 - S_{22} - S_{21}$$

$$\sigma_{21} = 1 - S_{22} - S_{12}$$

The reader can readily verify that the above scattering-matrix parameters satisfy the relations (12.6).

If the three-port scattering-matrix parameters have been measured, it is easy to obtain expressions for the two-port scattering-matrix parameters. Let us assume that we want to find the common emitter two-port scattering parameters. This requires that we make $V_3 = 0$ or $b_3 = -a_3$. By setting $b_3 = -a_3$ in the last equation in the set given by (12.5), we can solve for a_3 in terms of a_1 and a_2. By using this solution in the first two equations, we obtain the solutions for b_1 and b_2 in terms of a_1 and a_2 from which the two-port scattering-matrix parameters are readily identified. It is found that

$$S_{11} = \hat{S}_{11} - \frac{\hat{S}_{13}\hat{S}_{31}}{1 + \hat{S}_{33}} \qquad S_{12} = \hat{S}_{12} - \frac{\hat{S}_{13}\hat{S}_{32}}{1 + \hat{S}_{33}}$$

$$S_{21} = \hat{S}_{21} - \frac{\hat{S}_{23}\hat{S}_{31}}{1 + \hat{S}_{33}} \qquad S_{22} = \hat{S}_{22} - \frac{\hat{S}_{23}\hat{S}_{32}}{1 + \hat{S}_{33}} \tag{12.11}$$

Consider now the case when a series normalized impedance $\bar{Z}_s$ is connected between the emitter and the ground plane. For this situation, as shown in Fig. 12.12, $a_3 = \Gamma b_3$ where $\Gamma = (\bar{Z}_s - 1)/(\bar{Z}_s + 1)$. The same procedure used to obtain (12.11) can be used to find the two-port scattering matrix for the transistor with a series feedback impedance $\bar{Z}_s$ connected in series with the common emitter lead. It is readily found that the new scattering-matrix parameters are given by (12.11) upon replacing $\hat{S}_{33} + 1$ by $\hat{S}_{33} - \Gamma^{-1}$. As a final note we point out that, if the relations (12.10) are used to express the $\hat{S}_{ij}$ in terms of the S_{ij}, (12.11) is satisfied identically, i.e., it gives $S_{ij} \equiv S_{ij}$.

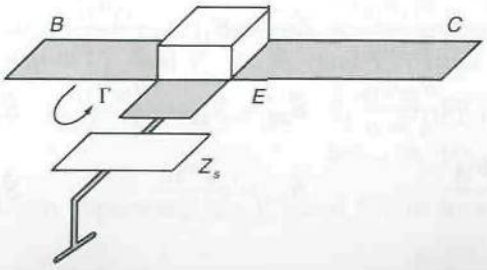


FIGURE 12.12
A transistor with a series feedback impedance in the common emitter lead.

The scattering parameters S_{11} and S_{22} are given by

$$S_{11} = \frac{\Delta_1 \Gamma - \hat{S}_{11}}{\hat{S}_{33}\Gamma - 1} \tag{12.12a}$$

$$S_{22} = \frac{\Delta_2 \Gamma - \hat{S}_{22}}{\hat{S}_{33}\Gamma - 1} \tag{12.12b}$$

where $\Delta_1 = S_{11}S_{33} - S_{13}S_{31}$ and $\Delta_2 = S_{22}S_{33} - S_{23}S_{32}$. The above equations can be solved for Γ to give

$$\Gamma = \frac{S_{11} - \hat{S}_{11}}{\hat{S}_{33}S_{11} - \Delta_1} \tag{12.13a}$$

$$\Gamma = \frac{S_{22} - \hat{S}_{22}}{\hat{S}_{33}S_{22} - \Delta_2} \tag{12.13b}$$

If we restrict the series impedance to be a pure reactive element, then $|\Gamma| = 1$. The circle of $|\Gamma| = 1$ values maps into circles in the S_{11} and S_{22} planes in accordance with the bilinear transformations given by (12.13). The center and radius of the S_{11} circle are

$$\text{Center} = \frac{\hat{S}_{11} - \Delta_1 \hat{S}_{33}^*}{1 - |\hat{S}_{33}|^2} \tag{12.14a}$$

$$\text{Radius} = \frac{|\hat{S}_{13}\hat{S}_{31}|}{|1 - |\hat{S}_{33}|^2|} \tag{12.14b}$$

while those for the S_{22} circle are

$$\text{Center} = \frac{\hat{S}_{22} - \Delta_2 \hat{S}_{33}^*}{1 - |\hat{S}_{33}|^2} \tag{12.15a}$$

$$\text{Radius} = \frac{|\hat{S}_{23}\hat{S}_{32}|}{|1 - |\hat{S}_{33}|^2|} \tag{12.15b}$$

By plotting these circles it is possible to visually see the range of values that can be obtained for S_{11} and S_{22} by varying the series feedback reactance. For each value of $j\bar{X}$ the stability parameter K for the equivalent two-port network can also be evaluated [see (10.18)]. This parameter is a useful measure of the degree of instability a series feedback impedance can produce. Some impedances, particularly resistive ones, will actually improve the stability of the device. Such impedances would be useful in amplifier design where instability is undesirable. Some caution should be exercised in using the above two-port to three-port relations since they are based on the assumption that the stray capacitance from each transistor terminal to the ground plane is negligible.

The above results show that a knowledge of the three-port scattering-matrix parameters is very useful in determining the equivalent two-port scattering-matrix parameters with a series feedback impedance connected in one common lead. The following example illustrates an application of the above results.

Example 12.1 Transistor with common source feedback impedance. A MESFET device has the following common source two-port scattering-matrix parameters at 10 GHz:

$$S_{11} = 0.73\angle 172^\circ \qquad S_{12} = 0.093\angle 39^\circ$$
$$S_{21} = 2.31\angle 44^\circ \qquad S_{22} = 0.09\angle -55^\circ$$

This device has a stability parameter $K = 1.13 > 1$ and is absolutely stable. By using (12.10) the following three-port scattering-matrix parameters were computed:

$$\hat{S}_{11} = 0.735\angle 241.5^\circ \qquad \hat{S}_{12} = 0.52\angle 27.8^\circ$$
$$\hat{S}_{13} = 0.98\angle 24.4^\circ \qquad \hat{S}_{21} = 1.71\angle 61.4^\circ$$
$$\hat{S}_{22} = 0.517\angle -79.2^\circ \qquad \hat{S}_{23} = 1\angle -85^\circ$$
$$\hat{S}_{31} = 1\angle -58^\circ \qquad \hat{S}_{32} = 0.517\angle 30.9^\circ$$
$$\hat{S}_{33} = 0.59\angle 87.8^\circ$$

If a normalized series reactance $j\bar{X}$ is inserted into the common source lead, a potentially unstable equivalent transistor or two-port network can be obtained. The possible range of S_{11} and S_{22} values that can be obtained are shown by the circles plotted in Fig. 12.13. Also shown on these circles are the values of S_{11} and S_{22} obtained using a normalized inductive reactance of $j0.5$ and $j1$ and the values obtained using normalized series capacitive reactance of $-j0.5$ and $-j1$. The computed values of the equivalent two-port network and the stability parameter K for the four reactances considered are tabulated below.

$$j\bar{X} = j0.5 \qquad K = 0.921$$
$$S_{11} = 0.244\angle 181.7^\circ \qquad S_{12} = 0.33\angle 66.3^\circ$$
$$S_{21} = 1.83\angle 40.3^\circ \qquad S_{22} = 0.342\angle -38.9^\circ$$
$$j\bar{X} = j1 \qquad K = 0.915$$
$$S_{11} = 0.126\angle 263.3^\circ \qquad S_{12} = 0.463\angle 65.1^\circ$$
$$S_{21} = 1.57\angle 39.7^\circ \qquad S_{22} = 0.475\angle -41.37^\circ$$
$$j\bar{X} = -j0.5 \qquad K = -0.49$$
$$S_{11} = 1.818\angle 188.2^\circ \qquad S_{12} = 0.556\angle -63.5^\circ$$
$$S_{21} = 3.23\angle 58.7^\circ \qquad S_{22} = 0.545\angle 185.7^\circ$$
$$j\bar{X} = -j1 \qquad K = -0.453$$
$$S_{11} = 3.12\angle 235.2^\circ \qquad S_{12} = 1.52\angle -19.7^\circ$$
$$S_{21} = 3.57\angle 98.7^\circ \qquad S_{22} = 1.52\angle 231.1^\circ$$

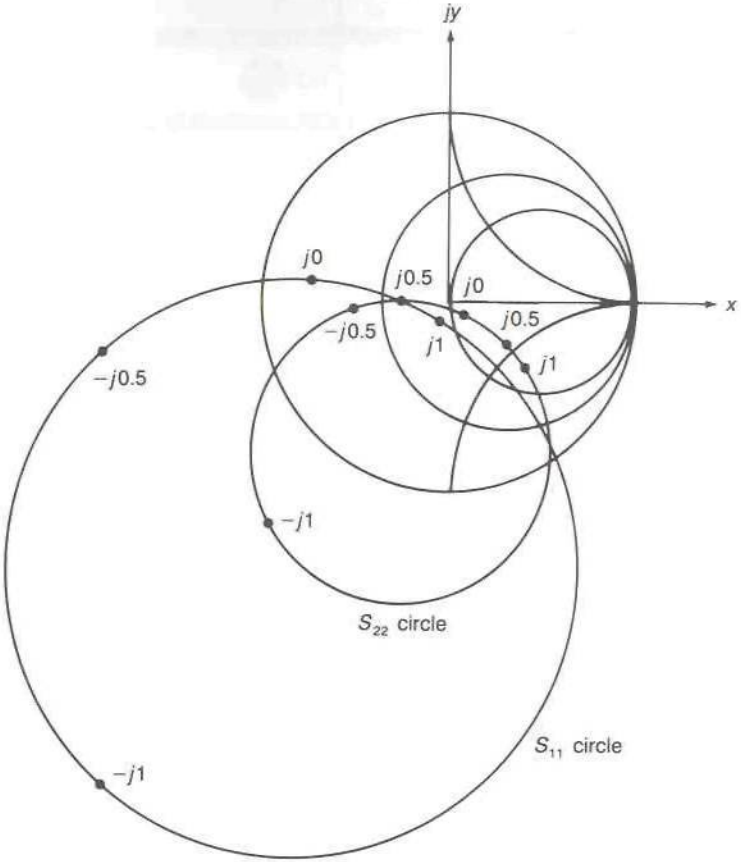


FIGURE 12.13
Circles showing the values of the scattering-matrix parameters S_{11} and S_{22} for a MESFET with a series reactance in the common source lead.

For the above reactance values $K < 1$, so the MESFET with common source series feedback is unstable (will oscillate). Note that a series capacitive reactance produces a highly unstable device, in particular, for $j\overline{X} = -j1$ the reverse transmission coefficient S_{12} is very large relative to normal values and both $|S_{11}|$ and $|S_{22}|$ are greater than unity.

The results given above were obtained using the computer program TRIPORT.

12.5 OSCILLATOR CIRCUITS

A microwave oscillator can be designed using any of the standard low-frequency oscillator circuits such as the Hartley, Colpitts, or Clapp circuits. Various variations of these circuits can also be used. The frequency stability of the oscillator is generally achieved by incorporating a resonator in either

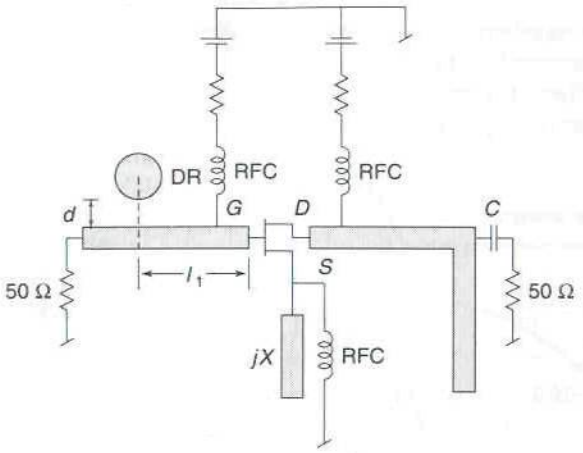


FIGURE 12.14
A 5-GHz FET oscillator using a dielectric resonator DR in the input circuit for frequency stabilization. The feedback is obtained using a series capacitive reactance in the common source lead.

the input or output circuits or as part of the feedback loop. A disk resonator can be used but its Q is relatively low; so the resultant frequency stability will not be very high. A high-Q metallic cavity can be used, but because of its large size the compact high-Q, temperature-stable, dielectric resonator is often the preferred choice.

In Fig. 12.14 we show a 5-GHz FET oscillator that is stabilized by using a dielectric resonator in the input circuit. The magnitude of the source reflection coefficient is controlled by the coupling to the resonator which can be varied by changing the spacing d between the resonator and the microstrip line. The phase angle of the source reflection coefficient is controlled by the length l_1 of the input line. The output circuit is a standard stub-matched circuit that transforms the 50-Ω load impedance to the required value for the oscillator. The FET is made to oscillate by using a series capacitive reactance in the common source lead. This feedback arrangement makes the equivalent transistor two-port circuit unstable as was shown in Example 12.1. The dc bias currents are applied through RF chokes. The output load is isolated from the oscillator dc voltages by the low-impedance dc blocking capacitor C.†

In Fig. 12.15 we show an FET oscillator using a dielectric resonator in the feedback path from the drain to the gate. The amount of feedback can be adjusted by the coupling to the dielectric resonator. The resonator is located a distance $\lambda/4$ from the open-circuited ends of the coupling lines since the standing wave of current and magnetic field is greatest at this location from the open ends and thus provides the strongest coupling to the resonator. The correct phase of the feedback voltage is controlled by the line

†V. Rizzoli, A. Neri, and A. Costanzo, Analysis and Optimization of DROs Using a General Purpose CAD Program, *Alta Frequenza*, vol. 57, pp. 389–398, 1988.

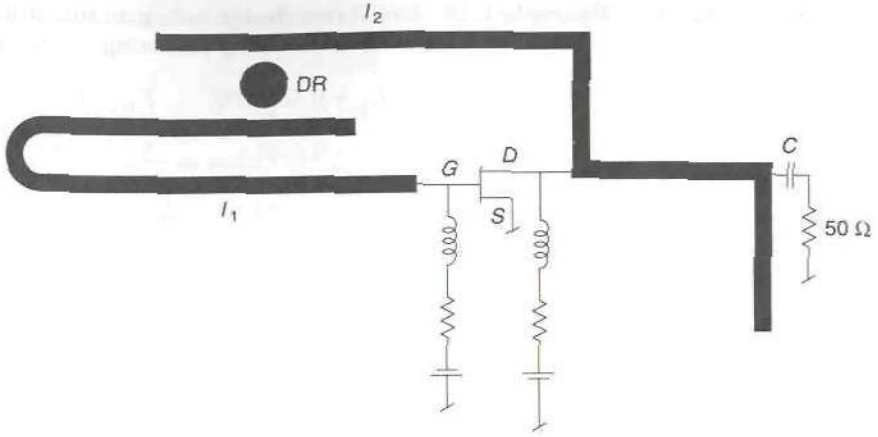


FIGURE 12.15
An FET oscillator using a dielectric resonator in the feedback path from the drain to the gate.

lengths l_1 and l_2, as is the phase of the source reflection coefficient seen at the gate terminal.

Electronic control of the oscillator frequency can be obtained using a varactor diode as part of the input circuit. Variable-frequency oscillators are also built using a yttrium garnet (YIG) ferrite sphere whose resonant frequency is controlled by the dc magnetic biasing field.

12.6 OSCILLATOR DESIGN

When maximum power output and efficiency are not of prime importance, a satisfactory oscillator design can be achieved using the small-signal scattering-matrix parameters. The amount of power generated can be varied by adjusting the dc bias voltages. The major shortcoming of small-signal oscillator design is that it does not provide any way of predicting the steady-state oscillating signal level. Oscillator design based on large-signal scattering-matrix parameters is much more difficult because of the difficulty of obtaining large-signal parameters. Two approaches are possible, namely, measuring the scattering-matrix parameters under large-signal conditions or obtaining these from computer simulations using a theoretical nonlinear model of the transistor. Neither method is easy to carry out so as to achieve a high accuracy. Space limitations will not allow us to discuss the large-signal approach. However, the references at the end of this chapter provide information on methods that have been developed and used successfully. We will only consider the small-signal design approach, and even this in a limited way, by means of two examples.

Example 12.2 Oscillator design using an unstable transistor. A silicon bipolar transistor has the following scattering-matrix parameters at 6 GHz:

$$S_{11} = 0.65\angle 130^\circ \qquad S_{12} = 0.2\angle 80^\circ$$

$$S_{21} = 2\angle 42^\circ \qquad S_{22} = 0.4\angle -60^\circ$$

The stability parameter $K = 0.646$, which is less than 1, so the transistor is potentially unstable. In Fig. 12.16 we have plotted the input and output stability circles. One of the primary effects of large-signal operation is a reduction of gain because of gain saturation. For simplicity, we will assume that under large-signal conditions S_{21} changes to a value $1.5\angle 42^\circ$ and all other scattering-matrix parameters stay the same. Thus, under large-signal conditions, we find that the stability parameter K has increased to 0.798. As a result the stability circles move. The large-signal stability circles are shown by the dashed circles in Fig. 12.16.

Let us choose a source reflection coefficient $\Gamma_s = 1\angle 210^\circ$ which corresponds to a pure capacitive reactance load at the base. For this value of Γ_s

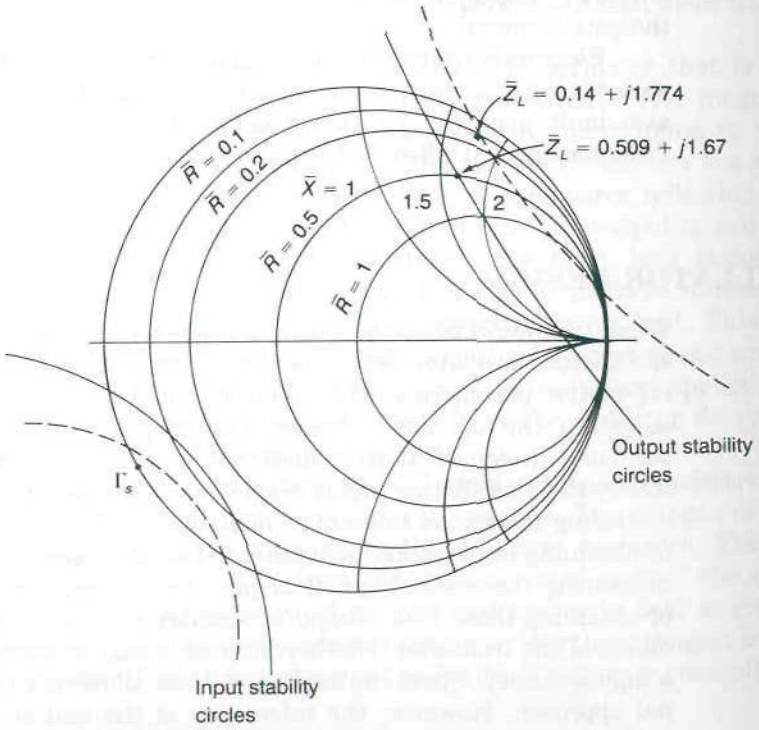


FIGURE 12.16
Stability circles for small- and large-signal conditions for the oscillator discussed in Example 12.2. The source reflection coefficient and two values of the load impedance are also shown.

we find that the output normalized impedance of the transistor is

$$\bar{Z}_{\text{out}} = -0.509 - j1.67 \qquad \text{for small signal conditions}$$

and

$$\bar{Z}_{\text{out}} = -0.139 - j1.774 \qquad \text{for large signal conditions}$$

In order to satisfy the conditions (12.3) for oscillations, we must choose a load termination $\bar{Z}_L$ at the collector where $\bar{Z}_L = 0.509 + j1.67$ for low-level oscillations and $\bar{Z}_L = 0.139 + j1.774$ for large-signal oscillations. The reactive parts are almost of the same value, but the resistive part for large-signal conditions is only 0.27 of the required value for small-signal conditions. In the small-signal design approach, it is usual practice to choose the load resistance around a factor of 3 smaller than what is required by the condition (12.3*a*) for steady-state oscillations. This will allow the oscillations to build up in amplitude until gain saturation makes $-\bar{R}_{\text{out}}$ equal to $\bar{R}_L$. There will be some shift in the resonant frequency as the oscillations build up since $\bar{X}_{\text{out}}$ changes and thus the frequency must change until $\bar{X}_{\text{out}} = -\bar{X}_L$. When the input network produces a very rapid change in the phase of Γ_s with frequency, only a small frequency shift will occur. Some means of tuning the oscillator is normally used so as to establish the desired frequency of oscillation.

In the example we are discussing, we will choose $\bar{Z}_L = 0.14 + j1.774$. This impedance point is shown in Fig. 12.16, as is the point $\bar{Z}_L = 0.509 + j1.67$. The former point lies on the stable side of the assumed large-signal stability circle and the oscillator would not oscillate under these conditions. Our choice for $\bar{Z}_L$ will, in actual practice, limit the oscillation amplitude at a value for which $|S_{21}|$ is somewhat greater than 1.5 so as to keep the point $\bar{Z}_L$ in the unstable region. Since we have used a pure reactive termination at the input, there is no power delivered to the input circuit. Thus the power gain of the circuit is negative and infinite, so that the stable point of oscillation occurs when the stability circle moves outwards so as to make $\bar{Z}_L$ lie on the circle since the output stability circle coincides with the infinite gain circle. When the input termination has a resistive part, the power gain must be negative and finite, so that $\bar{Z}_L$ must lie on the appropriate negative gain circle and inside the unstable region (see Chap. 10 for a discussion of negative gain circles in the unstable region).

The circuit for the oscillator is shown in Fig. 12.17. An open-circuited transmission line is used to produce the input reactance. Since $\Gamma_s = e^{-2j\beta l}$ the minimum length l is equal to $5\lambda/24$ in order to make the phase angle of Γ_s equal to $-150°$, which is equivalent to $210°$. The frequency sensitivity of the phase angle can be increased by using a transmission line $\lambda/2$ longer, that is $l = 17\lambda/24$. A 1 percent change in frequency or β will then change the phase angle of Γ_s by $(-150 - 180)/100 = -3.3°$. The computed values of $\bar{Z}_{\text{out}}$ for a $\pm 2°$ change in the phase angle of Γ_s are $\bar{Z}_{\text{out}} = -0.083 - j1.697$ for $\Gamma_s = 1\angle 218°$ and $\bar{Z}_{\text{out}} = -0.202 - j1.863$ for $\Gamma_s = 1\angle 222°$ under large-signal conditions. It is apparent that $\bar{Z}_{\text{out}}$ is quite sensitive to small changes in the phase angle of Γ_s.

The output matching network consists of a 50-Ω transmission line of length 0.1975λ and an open-circuit 50-Ω stub of length 0.22λ. This circuit transforms the 50-Ω load to a normalized impedance of $0.14 + j1.774$ at the collector. The bias voltages are applied through 150-Ω quarter-wave

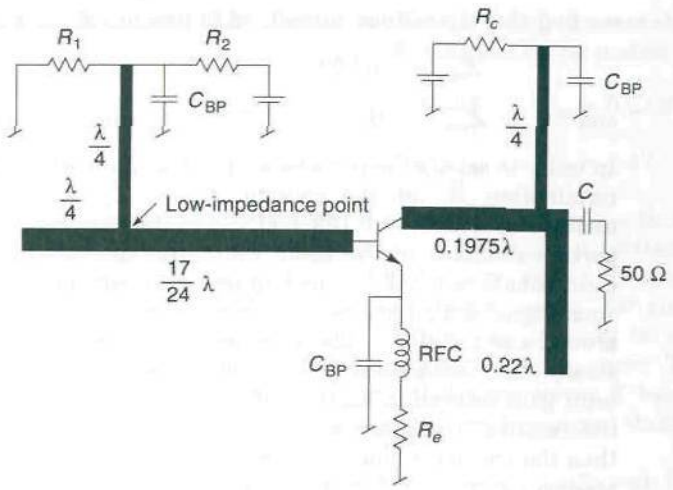


FIGURE 12.17
The oscillator circuit designed in Example 12.2.

transmission lines that are bypassed to the ground plane by capacitors C_{BP}. These lines are connected at low-impedance points on the input and output circuits. The 50-Ω output is isolated from the transistor by the dc blocking capacitor C.

The design carried out above is a hypothetical one since we do not know how the scattering-matrix parameters change with signal level. The example does show the basic physical principles involved in determining the steady-state operation of the oscillator. The small-signal design approach that uses $\overline{Z}_L = -(\overline{R}_{out}/3 + j\overline{X}_{out})$ usually leads to a reasonably satisfactory design.

Example 12.3 Oscillator design using a dielectric resonator. In this example we will use the MESFET described in Example 12.1 with a series reactance $-j0.5$ in the common source lead. The equivalent two-port scattering-matrix parameters are $S_{11} = 1.818\angle 188.2°$, $S_{12} = 0.556\angle -63.5°$, $S_{21} = 3.23\angle 58.7°$, $S_{22} = 0.545\angle 185.7°$, and the stability parameter $K = -0.49$. The input and output stability circles for the equivalent two-port network are shown in Fig. 12.18. Since $|S_{22}| < 1$ the origin is a stable point for Γ_s, so that values of Γ_s inside the source or input stability circle are unstable ones. The values of $\overline{Z}_L$ outside the load stability circle are unstable ones since $|S_{11}| > 1$.

We will choose $\Gamma_s = -0.8$ for the initial design of the oscillator. This value of Γ_s is shown in Fig. 12.18 and is in the unstable region. For this value of Γ_s we find that $\overline{Z}_{out} = -1.6 - j0.906$. For $\overline{Z}_L$ we will choose $\overline{Z}_L = 0.6 + j0.906$ for which $\overline{R}_L = -0.375\overline{R}_{out}$. The oscillator circuit used is shown in Fig. 12.19. The resonator equivalent circuit is a parallel combination of R, L, and C which is series-coupled to the microstrip line by an ideal transformer with turns ratio $n:1$ as shown in Fig. 7.23. We can choose R equal to unity by

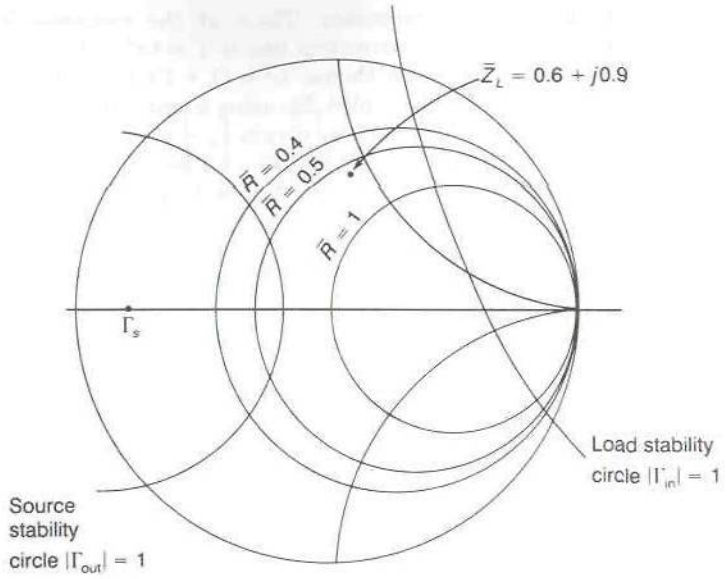


FIGURE 12.18
The input and output stability circles for the oscillator in Example 12.3. The design values of Γ_s and $\bar{Z}_L$ are also shown.

choosing an appropriate value for the turns ratio $n:1$. The resonator Q is given by $R/\omega_0 L$, which determines the inductance in terms of the resonator Q. The capacitance in the equivalent circuit is given by the resonance condition $\omega_0^2 LC = 1$. At $\omega = \omega_0$ the impedance coupled into the microstrip line is a series resistance equal to $n^2 R = n^2$, so n^2 represents the coupling coefficient. The transmission line of length $\lambda/4$ can be replaced by a short circuit at the

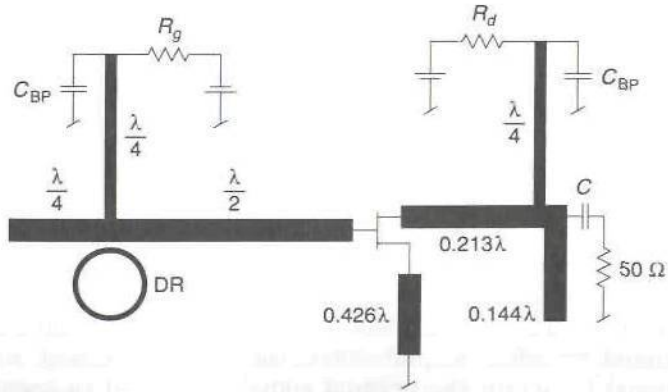


FIGURE 12.19
The oscillator circuit designed in Example 12.3.

location of the resonator. Thus, at the resonator location, the reflection coefficient on the microstrip line is $\Gamma = (n^2 - 1)/(n^2 + 1)$. In order to make $\Gamma = -0.8$, we must choose $n^2 = (1 + \Gamma)/(1 - \Gamma) = 0.2/1.8 = 0.111$; so the resonator is undercoupled. By using a transmission line $\lambda/2$ long between the resonator and the gate, we obtain $\Gamma_s = \Gamma = -0.8$.

The output circuit consists of a 50-Ω transmission line 0.213λ long and an open-circuited 50-Ω stub 0.144λ long. This network transforms the 50-Ω load impedance to the required normalized value $0.6 + j0.906$ at the drain terminal. The capacitive feedback normalized reactance of $-j0.5$ is obtained by using a short-circuited 50-Ω stub of length equal to 0.4262λ in the common source lead.

In general, the resonator impedance coupled into the input microstrip line is given by

$$\bar{Z} = \frac{n^2}{1 + 2jQ\dfrac{\omega - \omega_0}{\omega_0}}$$

in the vicinity of the resonant frequency ω_0. If we assume that the resonator Q equals 500, then a 0.1 percent change in frequency will change Γ from -0.8 to

$$\Gamma = \frac{n^2\bar{Z} - 1}{n^2\bar{Z} + 1} = \frac{n^2 - 1 - j}{n^2 + 1 + j} = \frac{-8 - j9}{10 + j9} = 0.895\angle 186.38°$$

The new value of $\bar{Z}_{\text{out}}$ becomes $-1.37 - j1.41$, where we have assumed, for a 0.1 percent change in frequency, that there is a negligible change in the S_{ij} parameters and the transmission-line length. If we used stronger coupling to the resonator, we would obtain a larger change in Γ and this would improve the frequency stability of the oscillator, since the resonator would have a stronger control of the oscillation frequency. A better choice for Γ_s would be -0.4 which gives $n^2 = 3/7$ and $\bar{Z}_{\text{out}} = -0.494 - j0.089$. A 0.1 percent change in frequency would now give

$$\Gamma = \frac{3 - 7 - j7}{3 + 7 + j7} = \frac{-4 - j7}{10 + j7} = 0.6605\angle 205.3°$$

and $\bar{Z}_{\text{out}} = -0.523 - j0.905$. The change in the phase angle of Γ has been increased by a factor of about 4 by using the larger coupling. The frequency stability is also increased by about the same amount. The design of the oscillator circuit using $\Gamma_s = -0.4$ is left as a problem to be solved (Prob. 12.5).

12.7 MIXERS

In Fig. 12.20 we show a block diagram of a microwave superheterodyne receiver. The signal from the antenna is first amplified by a low-noise amplifier. After amplification, the signal is mixed with a local-oscillator signal to obtain the original signal translated to a much lower frequency called the intermediate (IF) frequency. The mixer is a nonlinear device such as a diode or dual-gate FET. If the microwave carrier frequency is ω_c and

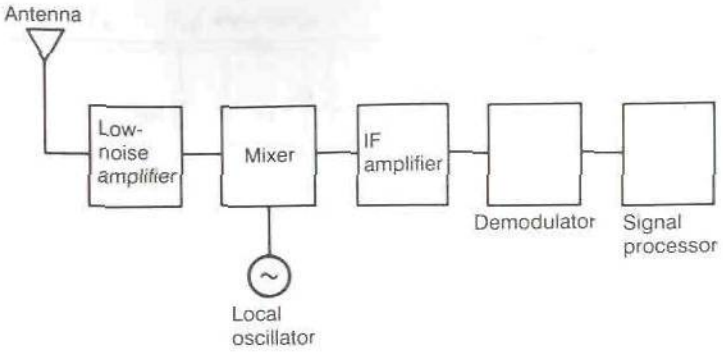


FIGURE 12.20
A block diagram of a microwave receiver.

the oscillator frequency is ω_0, the nonlinear mixer device will produce signals at the IF frequency $\omega_{IF} = \omega_0 - \omega_c$, at the frequency $\omega_0 + \omega_c$, and, in general, at many harmonic frequencies $n\omega_0 \pm m\omega_c$ where n and m are integers. The signal at the IF frequency is further amplified, then demodulated, and finally processed for the intended output application.

A single-tone AM signal with modulation frequency ω_m has the form

$$(1 + M\cos\omega_m t)\cos\omega_c t = \cos\omega_c t + \tfrac{1}{2}M[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$$

After mixing, the sideband spectrum becomes

$$\cos[(\omega_c - \omega_0)t + \omega_m t] + \cos[(\omega_c - \omega_0)t - \omega_m t]$$

when the carrier frequency ω_c is greater than the local-oscillator frequency ω_0. When the carrier frequency is less than the local-oscillator frequency, the sideband spectrum, after mixing, remains unchanged because an AM signal has symmetrical sidebands. Thus a local-oscillator frequency greater or smaller than the carrier frequency can be used.

For an FM or phase-modulated signal of the form $\cos[\omega_c t + \phi(t)]$ the spectrum, after mixing, is of the form $\cos[(\omega_c - \omega_0)t + \phi(t)]$ when $\omega_c > \omega_0$ and $\cos[(\omega_0 - \omega_c)t - \phi(t)]$ when $\omega_c < \omega_0$. For the case when ω_0 is greater than the carrier frequency ω_c, the sideband spectrum is reversed with the high frequencies becoming low frequencies, and vice versa. This phenomenon occurs because FM and phase-modulated signals do not have symmetrical sidebands. The subtraction of the spectrum of $\phi(t)$ from the IF frequency reverses the high- and low-frequency components. In order to avoid this inversion of the signal spectrum, the local-oscillator frequency must be less than the carrier frequency in FM and phase-modulated systems.

In this section we will discuss those mixer characteristics that are important from a systems point of view. For the purpose of this discussion,

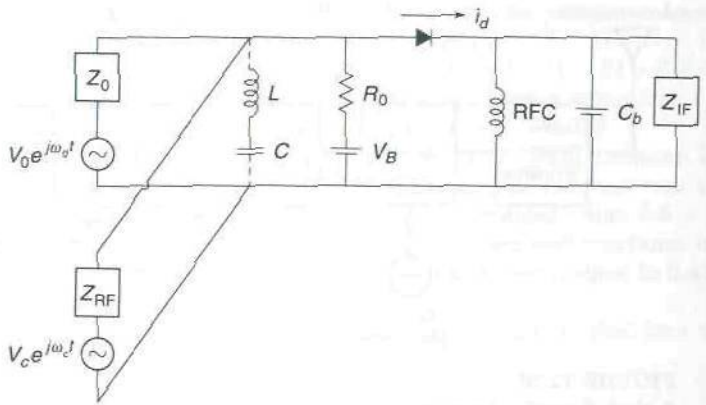


FIGURE 12.21
A single diode mixer circuit.

we will use a simple single-diode-mixer circuit to help clarify some of the operational characteristics of diode mixers. In the following section we will examine some of the more complex mixer circuits that enhance the overall mixer performance.

In Fig. 12.21 we show a simplified circuit for a single diode mixer. The RF signal with carrier frequency ω_c is connected to the diode through a filter network with impedance Z_{RF}. Similarly, the local-oscillator signal at frequency ω_0 is applied to the diode through the impedance Z_0. The IF signal appears across Z_{IF} which represents the input to the IF amplifier. If a point contact diode is used, a dc biasing circuit is not used but the diode must have a dc current return path to ground which is through the RF choke RFC. When a Schottky barrier diode is used, a small forward bias is normally applied to the diode. This serves to overcome the barrier potential and increases the sensitivity of the diode. In a conventional p-n diode, the junction capacitance is quite large and will shunt the RF and LO signals across the junction, thus making these diodes ineffective as mixer diodes at frequencies greater than 1 GHz. The minority carriers also limit the diode recovery time. The Schottky diode, consisting of a metal-semiconductor junction does not have a depletion layer and also has very little stored charge at the junction. Hence it has a very small junction capacitance. It is usable as a mixer diode for frequencies in the microwave and millimeter-wave range and beyond.

The IF frequency is much lower than the LO and RF frequencies. For example, if the RF frequency $f_c = 10$ GHz and a typical IF frequency of 50 MHz is assumed, then the local-oscillator frequency must be (10 ± 0.05) GHz. In this case the IF frequency is a factor of 200 smaller than the RF and LO frequencies. Hence the capacitor C_b in the circuit shown in Fig. 12.21 can be chosen large enough to short-circuit the high-frequency currents at the IF amplifier input. Consequently, the IF port is isolated from

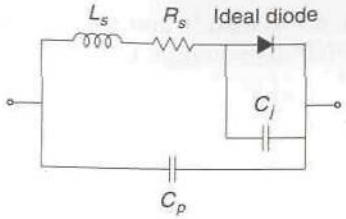


FIGURE 12.22
Equivalent circuit for a microwave diode.

the RF and LO ports. On the other hand, the RF and LO frequencies are almost equal, so it is not practical to use filter networks Z_0 and Z_{RF} that provide good isolation between the RF and LO ports. We can, however, assume that Z_0 and Z_{RF} are essentially zero-impedance elements at the IF frequency. We could, for example, insert a series resonant circuit across the RF and LO ports (shown by dashed lines in Fig. 12.21) with L and C chosen so that $LC\omega_{IF}^2 = 1$, where ω_{IF} is the IF frequency. This circuit provides a zero-impedance path across the RF and LO ports at the IF frequency. At the RF and LO frequencies, $j\omega L$ represents a very high reactance that shunts the RF and LO ports and can be neglected.

The equivalent circuit of the diode is shown in Fig. 12.22. In this circuit we have an ideal diode shunted by the diode junction capacitance C_j along with a series resistance R_s, a series inductance L_s, and a package shunting capacitance C_p. The ideal diode is described by the equation

$$i_d = I_s(e^{\gamma v_d} - 1) \tag{12.16}$$

where i_d is the diode current, I_s is the reverse saturation current, v_d is the voltage across the diode; $\gamma = e/kTn$, where e is the electron charge, k is Boltzmann's constant, T is the absolute temperature, and n is a diode-dependent parameter having a value between 1 and 1.5. At room temperatures γ has the approximate value of 40.

For the purpose of the discussion in this section, we are going to neglect all of the parasitic elements in the equivalent circuit of the diode. This can be justified only if we assume that the RF and LO frequencies are quite low. We will carry out a more careful analysis of the diode mixer in a later section. With the assumptions we have made, the mixer equivalent circuit for RF, LO, and IF signals reduce to those shown in Fig. 12.23. The network that the diode is embedded in has been assumed to have a zero impedance at all frequencies except those in the vicinity of the RF, LO, and IF frequencies.† Thus the only voltages that can exist across the diode are those at the RF, LO, and IF frequencies since all other frequency components are short-circuited by the embedding network.

†When diode biasing is used, the RF and LO circuits must have nonzero dc resistance so as not to short-circuit the dc voltage applied to the diode.

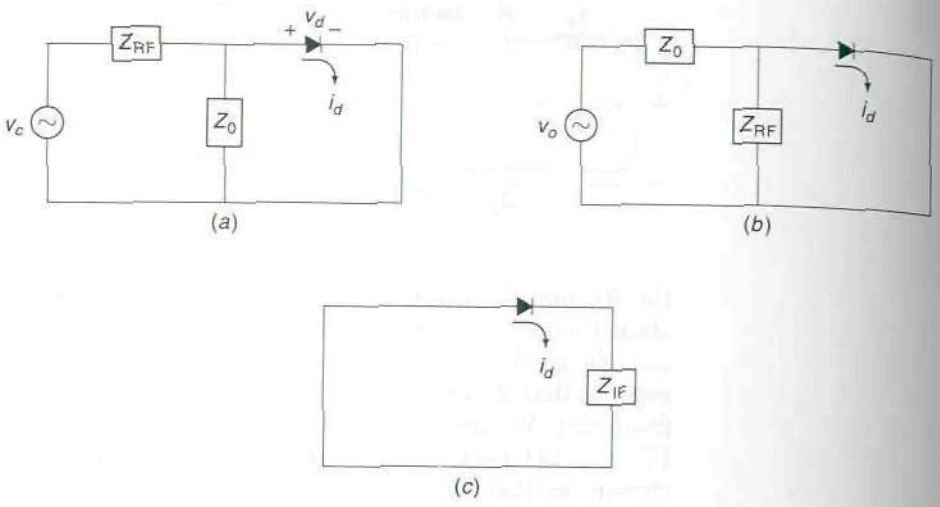


FIGURE 12.23
(*a*) Equivalent circuit for RF signals in the mixer; (*b*) equivalent circuit for LO signals; (*c*) equivalent circuit for IF signals.

The current i_d through the diode can be expanded in a Taylor series about the dc operating point. This is equivalent to assuming a power-series expansion of the form

$$i_d = I_0 + a_1 v + a_2 v^2 + a_3 v^3 + \cdots \tag{12.17}$$

where I_0 is the dc biasing current, v is the voltage excursion about the operating point, and $a_1, a_2, a_3, \ldots$ are suitable coefficients. We will assume that the voltage v consists of a local-oscillator signal $v_0 = V_0 \cos \omega_0 t$, an RF signal $v_1 = V_1 \cos \omega_1 t$, and an IF signal $-v_{IF}$ at the IF frequency $\omega_0 - \omega_1$. The IF signal has the form $v_{IF} = |V_{IF}| \cos(\omega_{IF} t + \phi)$, where ϕ is the phase angle. The complex-phasor IF voltage is $V_{IF} = |V_{IF}| e^{j\phi}$. When we expand (12.17) we obtain

$$\begin{aligned} i_d = I_0 &+ a_1(v_0 + v_1 - v_{IF}) + a_2(v_0^2 + v_1^2 + v_{IF}^2 + 2v_0 v_1 - 2v_0 v_{IF} - 2v_1 v_{IF}) \\ &+ a_3(v_0^3 + v_1^3 - v_{IF}^3 + 3v_0^2 v_1 - 3v_0^2 v_{IF} + 3v_0 v_1^2 + 3v_0 v_{IF}^2 \\ &\qquad - 3v_1^2 v_{IF} - 6v_0 v_1 v_{IF}) \end{aligned} \tag{12.18}$$

In a number of published analyses of mixers using a power series such as that in (12.17), it is assumed that the voltage across the diode consists only of the RF and LO signals. This assumption would imply that the embedding network that the diode is connected to has a zero impedance at the IF frequency, which is an unrealistic assumption. If power is to be delivered to one port of the embedding network at the IF frequency, then the network must have a nonzero impedance at the IF frequency at this

port. The impedance of the embedding network at the various harmonics of the applied signals determines the harmonic voltages across the diode.

Linear Mixer Operation

In practice, the local-oscillator signal has an amplitude much larger than the RF signal and the IF signal. When the RF signal amplitude is small, we can neglect terms that are proportional to higher-order powers of the RF and IF signals. Thus, when we retain only those terms that are linear in v_1 and v_{IF}, the diode current is given by

$$i_d = I_0 + a_1(v_0 + v_1 - v_{IF}) + a_2(v_0^2 + 2v_0v_1 - 2v_0v_{IF}) + a_3(v_0^3 + 3v_0^2v_1 - 3v_0^2v_{IF}) \tag{12.19}$$

A term such as v_0^2 equals $V_0^2\cos^2\omega_0 t = \frac{1}{2}V_0^2 + \frac{1}{2}v_0^2\cos 2\omega_0 t$. The product term $2v_0v_1$ equals $V_0V_1[\cos(\omega_0 - \omega_1)t + \cos(\omega_0 + \omega_1)t]$ and has an IF frequency component. The product term $-2v_0v_{IF}$ equals $-V_0|V_{IF}|\{\cos[(\omega_0 - \omega_{IF})t - \phi] + \cos[(\omega_0 + \omega_{IF})t + \phi]\}$ and does not contain frequency components falling within the IF amplifier passband. The term $-3a_3v_0^2v_{IF}$ contributes an IF current component equal to $-\frac{3}{2}a_3V_0^2v_{IF}$. The total IF current is given by

$$i_{IF} = -a_1v_{IF} + a_2V_0V_1\cos\omega_{IF}t - \tfrac{3}{2}a_3V_0^2v_{IF}$$

We now let $i_{IF} = \text{Re}\, I_{IF}e^{j\omega_{IF}t}$ and use phasor analysis to obtain

$$I_{IF} = -a_1V_{IF} + a_2V_0V_1 - \tfrac{3}{2}a_3V_0^2V_{IF}$$

At the IF frequency the circuit equation or constraint imposed by the embedding network shown in Fig. 12.23*c* gives $V_{IF} = I_{IF}Z_{IF}$. Thus we find that

$$V_{IF} = \frac{a_2V_0Z_{IF}}{1 + (a_1 + 1.5a_3V_0^2)Z_{IF}}V_1 \tag{12.20}$$

The power-series expansion up to terms in a_3 is valid only for small values of V_0. Consequently, the term $1.5a_3V_0^2$ is usually small relative to a_1. The above equation shows that under the assumptions made, the voltage at the IF amplifier input is linearly proportional to the RF voltage amplitude V_1. In this operating range the mixer functions as a linear mixer.

The conversion loss, in decibels, for the mixer is given by

$$L_c = \text{conversion loss} = 10\log\frac{\text{available RF power}}{\text{IF input power}} \tag{12.21}$$

Typical values for conversion loss for a single-diode mixer are 6 to 10 dB.

Nonlinear Mixer Operation

For larger values of the RF voltage, we must take additional terms into account. From the term multiplied by a_3, we have a term

$$-3a_3v_1^2v_{IF} = -1.5a_3V_1^2v_{IF} - 1.5a_3V_1^2v_{IF}\cos 2\omega_1 t$$

and a term

$$-a_3v_{IF}^3 = -a_3|V_{IF}|^3\cos^3(\omega_{IF}t + \phi) = -0.75a_3|V_{IF}|^3\cos(\omega_{IF}t + \phi) - 0.25a_3|V_{IF}|^3\cos(3\omega_{IF}t + 3\phi)$$

The IF frequency component of this latter term can be expressed as $-0.75a_3|V_{IF}|^2v_{IF}$. We can use (12.20) to obtain an approximate solution for $|V_{IF}|^2$ and then find that

$$V_{IF} = \frac{a_2V_0Z_{IF}V_1}{1 + \left[a_1 + 1.5a_3(V_0^2 + V_1^2) + 0.75a_3|V_{IF}|^2\right]Z_{IF}}$$

$$= \frac{a_2V_0V_1}{Y_{IF} + a_1 + 1.5a_3V_0^2\left[1 + \dfrac{V_1^2}{V_0^2} + \dfrac{a_2^2V_1^2|Z_{IF}|^2}{2|1 + (a_1 + 1.5a_3V_0^2)Z_{IF}|^2}\right]} \tag{12.22}$$

where $Y_{IF} = Z_{IF}^{-1}$. This equation shows that the effect of the nonlinear terms $-a_3v_{IF}^3$ and $-3a_3v_1^2v_{IF}$ is to reduce the IF voltage. Thus a mixer will exhibit nonlinear saturation and the range of allowed input RF voltage amplitudes (dynamic range) must be limited in order to avoid nonlinear distortion of the down-converted RF signal. The dynamic range of a mixer is specified by giving the RF power level at which a compression of 3 dB occurs in the IF power.† In Fig. 12.24 we show a plot of IF power versus RF input power and the 3-dB compression point. The units are dBm or decibels relative to 1 mW.

The nonlinear behavior of a mixer also results in intermodulation distortion. Consider an input signal consisting of two closely spaced sinusoidal signals at the frequencies ω_1 and ω_2, where $|\omega_1 - \omega_2| \ll \omega_{IF}$. The presence of two closely spaced RF signals results in the generation of a larger number of frequency components. For linear operation we obtain two IF signals $|V_{IF1}|\cos[(\omega_0 - \omega_1)t + \phi_1]$ and $|V_{IF2}|\cos[(\omega_0 - \omega_2)t + \phi_2]$. The cubic term $-a_3v_{IF}^3$ will result in new additional frequency components that come from the products $v_{IF1}^2v_{IF2}$ and $v_{IF1}v_{IF2}^2$. The new frequencies generated, which fall within the IF amplifier passband, are $\omega_0 - 2\omega_1 + \omega_2$ and

†Sometimes the 1-dB compression point is used.

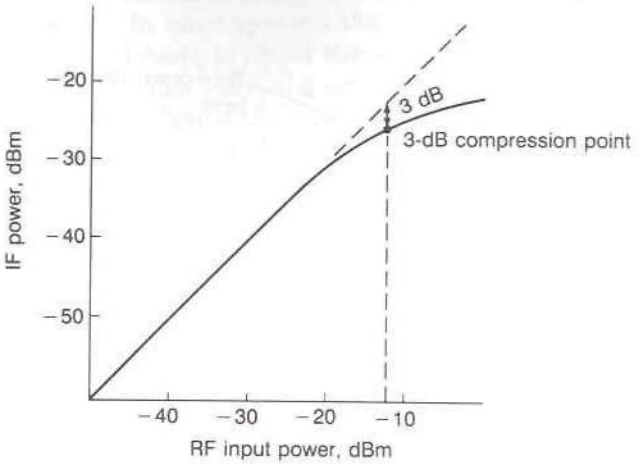


FIGURE 12.24
Illustration of nonlinear mixer response and the 3-dB compression point.

$\omega_0 - 2\omega_2 + \omega_1$. We will let $\omega_{IF3} = (\omega_0 - \omega_1) + (\omega_2 - \omega_1)$ and $\omega_{IF4} = (\omega_0 - \omega_2) + (\omega_1 - \omega_2)$. For a self-consistent solution for (12.18), we see that we must assume the presence of at least four IF signals, namely,

$$v_{IF} = V_{IF1}\cos(\omega_{IF1}t + \phi_1) + V_{IF2}\cos(\omega_{IF2}t + \phi_2) + V_{IF3}\cos(\omega_{IF3}t + \phi_3) + V_{IF4}\cos(\omega_{IF4}t + \phi_4) \quad (12.23)$$

When v_{IF3} and v_{IF4} are included, then the term multiplied by a_3 will also have product terms of the form $v_{IF1}^2 v_{IF3}$, $v_{IF3}^2 v_{IF4}$, etc. These product terms result in additional new frequency components that fall within the IF passband. Since v_{IF1} and v_{IF2} are proportional to V_1 and V_2, we see that v_{IF3} and v_{IF4} will be proportional to $V_1^2 V_2$ and $V_2^2 V_1$, respectively. A term such as $v_{IF1}^2 v_{IF3}$ is thus proportional to $V_1^4 V_2$ and can be neglected since its amplitude will be much smaller.

The production of new frequency terms that arise from the mixing of v_{IF1} and v_{IF2} is called two-tone intermodulation distortion. If $V_1 = V_2 = V$ these terms are proportional to the third power of the RF signal amplitude. The IF power associated with the intermodulation terms will increase proportional to the third power of the input RF power. The intermodulation distortion of a mixer is specified by giving the RF power level at which the intermodulation IF power becomes equal to the IF power associated with the desired IF signals v_{IF1} and v_{IF2}. This is a theoretical value that is obtained by extrapolating the linear mixer response or IF power output at the desired frequencies ω_{IF1} and ω_{IF2} and that of the IF power output for the intermodulation terms until they meet. On a decibel scale the slope of the curve giving the intermodulation IF power is three times that of the

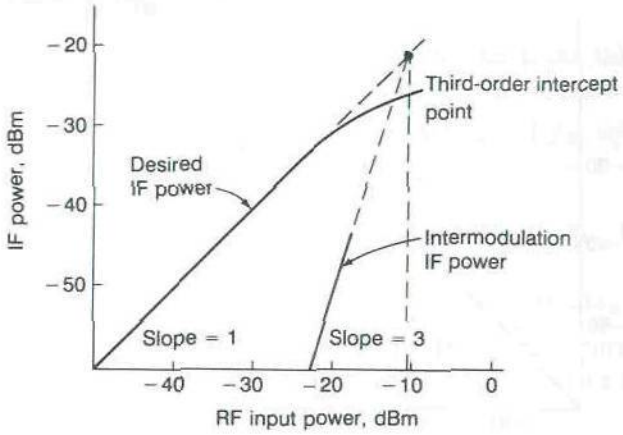


FIGURE 12.25
Illustration of diode mixer response for the desired IF power and the IF power at the intermodulation frequencies. The intersection of the two extrapolated response curves gives the third-order intercept point.

curve giving the desired IF signals. The intercept point is called the third-order intercept and is illustrated in Fig. 12.25. In practice, the RF input power to a mixer must be kept below the 3-dB saturation point and also below the third-order intercept point to avoid signal distortion.

When the amplitude V_2 of the RF signal at the frequency ω_2 is held constant, but V_1 is increased, then the slope of the line giving the intermodulation power in the IF passband will be 2.

12.8 MIXER NOISE FIGURE

The noise figure of a mixer is another important parameter that specifies the performance of a mixer. Thermal input noise at frequencies around $\omega_0 - \omega_{IF}$ will be down-converted into noise in the IF passband. Thermal noise at the image frequencies around $\omega_0 + \omega_{IF}$ when mixed with the LO signal will also fall within the IF passband. The mixer will also introduce additional noise with frequency components in the IF passband. This additional noise arises primarily from thermal noise in the resistive elements of the mixer and from shot noise.

The noise figure of a mixture is defined by the relation

$$F = \frac{S_i/N_i}{S_o/N_o} \tag{12.24}$$

where S_i/N_i is the input signal-to-noise power ratio and S_o/N_o is the output signal-to-noise ratio. If the local-oscillator frequency is ω_0, then RF signals at $\omega_0 + \omega_{IF}$ and $\omega_0 - \omega_{IF}$ are converted to IF signals. The frequency $\omega_0 + \omega_{IF}$ is called the upper sideband and the frequency $\omega_0 - \omega_{IF}$ is

the lower sideband. Noise at both sidebands is converted into noise in the IF passband. In most systems the signal is present in one sideband only. The unused sideband is called the image frequency. When the signal is present in one sideband only, the noise figure is designated as the single-sideband (SSB) noise figure F_{SSB}. If the signal is present in both sidebands, the corresponding noise figure is called the double-sideband (DSB) noise figure and denoted by F_{DSB}. A double-sideband signal would be of the form $A[1 + m(t)][\cos(\omega_0 + \omega_{IF})t + \cos(\omega_0 - \omega_{IF})t]$, where $m(t)$ is the modulation. When down-converted the IF signal voltage will be proportional to $2A[1 + m(t)]\cos \omega_{IF}t$, and consequently, the IF power is increased by a factor of 4. Thus for double-sideband signals, the ratio S_i/S_o is one-half of that for a single-sideband signal since the output IF power is increased by a factor of 4 but the input RF power is increased by a factor of 2 only. This assumes that the mixer conversion loss is the same for both sidebands. As a consequence of the above, the single-sideband noise figure is twice as large as the double-sideband noise figure, that is $F_{SSB} = 2F_{DSB}$. The double-sideband noise figure is easiest to measure since noise sources are usually very broadband and would provide noise signal input in both sidebands.

Schottky diode mixers have typical single-sideband noise figures in the range of 4 to 8 dB. The IF amplifier noise figure will also increase the noise figure of the mixer–IF amplifier combination because of the mixer conversion loss. The noise figure of the mixer–IF amplifier combination is given by

$$F = F_m + L_c(F_{IF} - 1) \tag{12.25}$$

where F_m is the mixer noise figure, F_{IF} is the IF amplifier noise figure, and L_c is the numerical value of the mixer conversion loss. As an example assumes that the conversion loss is 4 (6 dB), the IF amplifier noise figure is 2 (3 dB), and the mixer noise figure is 4. The noise figure of the combination is thus 8 which is a 3-dB increase over that of the mixer by itself. Since the signal level at the output of a mixer needs additional amplification, the mixer conversion loss has a significant effect on the overall noise figure. For this reason it is important to have sufficient signal amplification ahead of the mixer. The noise figure of the preamplifier will then govern the system noise figure.

12.9 BALANCED MIXERS

There are several disadvantages associated with a single-diode mixer. These are: no isolation between the RF and LO ports, poor isolation between the IF port and the RF and LO ports, a high level of oscillator noise input to the IF amplifier, and the generation of many spurious signals. Balanced mixers can improve the characteristics of a mixer and alleviate many of the listed shortcomings.

The basic circuit for a single balanced mixer is shown in Fig. 12.26. The two diodes need to be well matched in their electrical characteristics in

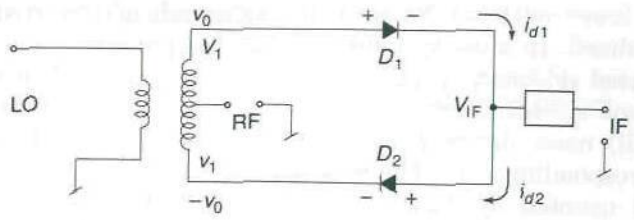


FIGURE 12.26
Basic circuit for a single balanced diode mixer.

order to achieve a well-balanced mixer. From the symmetry properties of the circuit, we can see that the voltage across diode D_1 is $v_0 + v_1 - v_{IF}$, while that acting across diode D_2 is $v_0 - v_1 + v_{IF}$. If we retain terms up to a_2 in the expression (12.18) for the diode current, we find that for diode D_1,

$$i_{d1} = I_0 + a_1(v_0 + v_1 - v_{IF}) + a_2(v_0^2 + v_1^2 + v_{IF}^2 + 2v_0v_1 - 2v_0v_{IF} - 2v_1v_{IF})$$

while for diode D_2,

$$i_{d2} = I_0 + a_1(v_0 - v_1 + v_{IF}) + a_2(v_0^2 + v_1^2 + v_{IF}^2 - 2v_0v_1 + 2v_0v_{IF} - 2v_1v_{IF})$$

The input current to the IF low-pass filter is

$$i_{d1} - i_{d2} = 2a_1(v_1 - v_{IF}) + 4a_2(v_0v_1 - v_0v_{IF})$$

The IF frequency component is given by

$$I_{IF} = -2a_1V_{IF} + 2a_2V_0V_1 \tag{12.26}$$

From the above equations we see that there is no local-oscillator voltage at the IF port so the LO and IF ports are isolated. Amplitude-modulation noise from the oscillator, as well as up-converted (or down-converted) thermal noise in the oscillator circuit, will consequently not be present at the IF input. From an examination of the circuit in Fig. 12.26, it is also apparent that the LO and RF ports are isolated.

For broadband mixers operating up to 1 GHz or so, the single balanced mixer can be built using transformers consisting of windings on a ferrite toroidal core. At microwave frequencies the single balanced mixer can be constructed using a 180° 3-dB hybrid junction such as the magic T or hybrid ring described in Chap. 6. A single balanced mixer using a magic T is shown in Fig. 12.27. A 90° hybrid junction can also be used, but the mixer will then not be fully balanced with respect to LO and IF port interactions. In a 90° hybrid junction reflected waves at the output port appear at both input ports with a phase that makes the reflected waves add out of phase. Thus

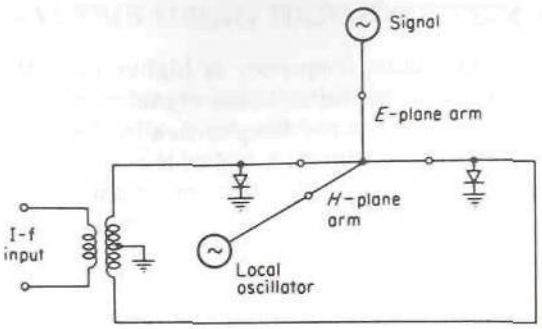


FIGURE 12.27
Single balanced mixer using a waveguide magic T.

the LO and RF port input VSWRs tend to be small. For a 180° hybrid junction, the reflected waves add in phase at each input port so the resultant input VSWRs are higher for the same degree of output impedance matching. A VSWR of 1.5 using a 90° hybrid junction can be readily achieved, whereas the VSWR for a 180° hybrid junction is around 2. It is relatively difficult to provide a good impedance match between a diode and a transmission line over a wide band of frequencies.

By using the four diodes in a bridge network, a double balanced mixer can be built. The basic circuit for a double balanced mixer is shown in Fig. 12.28 and provides isolation between the RF and LO ports, as well as between the IF and the RF and LO ports. In addition, many more spurious signals cancel at the IF port. The double balanced mixer suppresses the even harmonics of both the RF and LO signals and hence has a lower level of intermodulation distortion than that of the single balanced mixer. At the lower frequencies double balanced mixers are readily constructed using transformer hybrids. At microwave frequencies transmission-line baluns are used and the overall mixer configuration can become quite complex.

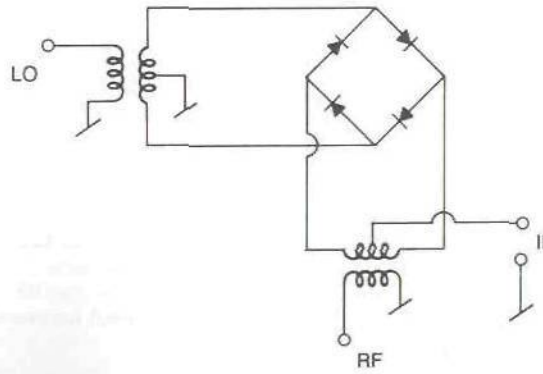


FIGURE 12.28
Basic circuit for a double balanced mixer.

12.10 OTHER TYPES OF MIXERS

When the local-oscillator frequency is higher than the RF signal frequencies, the IF spectrum contains those signal components in the frequency band $\omega_0 - \omega_{IF}$. The undesired frequencies in the image band $\omega_0 + \omega_{IF}$ are suppressed by means of a filter ahead of the mixer. It is possible to arrange two mixers in such a manner that IF signals corresponding to signals in the upper and lower sidebands are down-converted and appear at separate output ports. The unwanted sideband signals can be dissipated in a resistive termination. A mixer of this type is called an image-rejection mixer and is particularly useful in wideband signals where it is difficult to eliminate signals in the undesired sideband by filtering.

The up-converted signal proportional to $V_0V_1 \cos(\omega_0 + \omega_1)t$ represents a transfer of RF power to an undesired frequency. By reactively terminating this frequency component, it can be reflected back to the mixer and mixed with the second harmonic of the LO signal to produce a desired IF signal. With a proper adjustment of the circuit parameters, the conversion loss can be reduced by 1 to 2 dB. This type of mixer is called an image-recovery or image-enhanced mixer.

At millimeter wavelengths it is not always possible to build a suitable oscillator at the very high required frequency. In such circumstances a lower-frequency oscillator can be used and the desired IF frequency signals can be obtained by mixing the RF signal with one of the harmonics of the oscillator. These mixers are referred to as subharmonic mixers. By using two well-matched diodes connected in antiparallel as shown in Fig. 12.29, a second harmonic mixer is obtained. This particular diode arrangement results in an absence of all mixing products that involve odd harmonics of the local-oscillator signal.

FET mixers can also be configured to operate as balanced mixers, although current practice is to use an FET in an unbalanced circuit configuration. Discussions of FET mixers can be found in the references given at the end of this chapter.

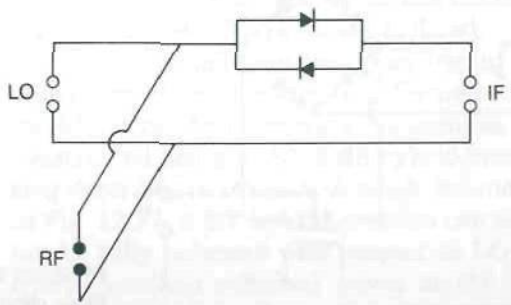


FIGURE 12.29
Basic circuit of a subharmonic mixer where the RF signal is mixed with the second harmonic of the oscillator signal.

12.11 MIXER ANALYSIS USING HARMONIC BALANCING

A power-series expansion of the diode current, such as that given by (12.17), is valid only when the voltage across the diode is of the order of a few millivolts. In practice the local-oscillator-signal voltage across the diode is much larger in order to obtain a relatively low conversion loss. The harmonic-balance method of mixer analysis is based on first analyzing the nonlinear mixer circuit with only a local-oscillator signal applied. This leads to a description of the diode in terms of a time-varying conductance and capacitance. In this section we will develop the basic concepts associated with the harmonic-balancing method, which are similar to those used in describing the operation of parametric amplifiers in Chap. 11.

The equivalent circuit for the diode is shown in Fig. 12.22. The junction capacitance C_j is a function of the voltage across the ideal diode junction. The other circuit elements, namely R_s, L_s, and C_p, can be considered to be part of the embedding network. We can replace the network as seen from the terminals of the ideal diode by a Thévenin equivalent circuit which includes an equivalent local-oscillator source as shown in Fig. 12.30*a*. The impedance $Z(\omega)$ of this network has a different

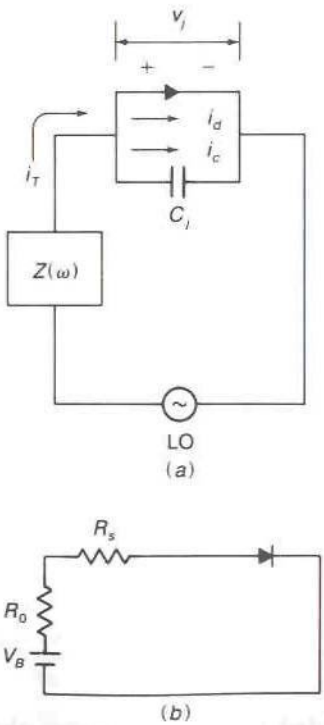


FIGURE 12.30
(*a*) Thévenin equivalent circuit of a diode mixer as seen from the terminals of the ideal diode; (*b*) Thévenin equivalent circuit of a diode mixer at $\omega = 0$.

value at each harmonic $n\omega_0$ of the local-oscillator signal. The total current i_T flowing through the diode and junction capacitor is $i_T = i_d + i_c$. The current through the capacitor is given by

$$i_c = \frac{dQ(v_j)}{dt} = \frac{dQ}{dv_j}\frac{dv_j}{dt}$$

where $Q(v_j)$ is the charge stored at the junction. The incremental capacitance $dQ/dv_j = C_j(V_j)$ is a voltage-dependent capacitance since, in general, Q is not a linear function of v_j. The current i_d through the ideal diode is given by

$$i_d = I_s(e^{\gamma v_j} - 1)$$

When only the local-oscillator signal is applied to the circuit, the voltage v_j and the currents i_T, i_d, i_c will contain all harmonics $n\omega_0$ of the fundamental frequency ω_0. Hence we can write

$$i_d = \sum_{n=-\infty}^{\infty} I_n e^{jn\omega_0 t}$$

$$v_j = \sum_{n=-\infty}^{\infty} V_n e^{jn\omega_0 t}$$

$$i_c = \sum_{n=-\infty}^{\infty} I_{cn} e^{jn\omega_0 t}$$

Since i_d and v_j are real, $I_{-n} = I_n^*$ and $V_{-n} = V_n^*$. The voltage v_j is periodic, so $C_j(v_j)$ is a periodic function and has a Fourier series representation of the form

$$C_j = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_0 t}$$

where the C_n will be dependent on v_j. C_j is a time-varying circuit element. The presence of a dc bias voltage simply changes the values of I_0, V_0 of the $n = 0$ terms.

Consider now the presence of an additional small-signal voltage v across the junction, with $|v| \ll |v_j|$. The current through the diode is now given by

$$i_T + i = I_s(e^{\gamma(v_j+v)} - 1) + \frac{dQ(v_j + v)}{dt} \qquad (12.27)$$

This equation can be expanded in a Taylor series about v_j, the dynamic

operating point; thus

$$i_T + i = I_s(e^{\gamma v_j} - 1) + \frac{dQ(v_j)}{dt} + \left[\gamma I_s e^{\gamma v_j} v + \frac{d}{dt}\left(\frac{dQ(v_j)}{dv_j} v\right)\right] + \cdots$$

$$= i_T + \gamma I_s e^{\gamma v_j} v + \frac{d}{dt}(C_j v)$$

The current i is given by

$$i = \left(\gamma I_s e^{\gamma v_j} + \frac{dC_j}{dt}\right) v + C_j(v_j)\frac{dv}{dt} \tag{12.28}$$

This is a *linear* differential equation of first order with time-dependent coefficients. Under small RF signal conditions the diode behaves as a time-varying linear circuit element and superposition holds for the RF signal.

If an RF signal at frequency ω_1 is applied to the circuit, the time-varying diode-operating characteristics will cause currents and voltages at all the frequencies $\omega_1 + n\omega_0$, $n = 0, \pm 1, \pm 2, \ldots$, to be generated. Hence i_1 and v_1 will have the expansions

$$i = \sum_{n=-\infty}^{\infty} i_n e^{j(\omega_1 + n\omega_0)t} \qquad v = \sum_{n=-\infty}^{\infty} v_n e^{j(\omega_1 + n\omega_0)t}$$

There are complex-conjugate terms at $-\omega_1$. The term $\gamma I_s e^{\gamma v_j} = \gamma(i_d + I_s) \approx \gamma i_d$ since I_s is very small. Thus $\gamma I_s e^{\gamma v_j}$ can be replaced by γi_d. Each term γI_n has the dimensions of a conductance and will be called g_n, that is, $\gamma I_n = g_n$. From the basic equation describing the diode small-signal behavior

$$\begin{aligned} i &= \sum_{n=-\infty}^{\infty} i_n e^{j(\omega_1 + n\omega_0)t} \\ &= \sum_{s=-\infty}^{\infty} (g_s + js\omega_0 C_s) \sum_{n=-\infty}^{\infty} v_n e^{js\omega_0 t} e^{j(\omega_1 + n\omega_0)t} \\ &\quad + \sum_{s=-\infty}^{\infty} C_s \sum_{n=-\infty}^{\infty} v_n j(\omega_1 + n\omega_0) e^{js\omega_0 t} e^{j(\omega_1 + n\omega_0)t} \end{aligned} \tag{12.29}$$

The harmonic terms on each side of this equation must be equal (principle of harmonic balancing); so upon putting $\omega_1 + (s + n)\omega_0 = \omega_1 + m\omega_0$, where $m = s + n$, we can write

$$\begin{aligned} i_m &= \sum_{n=-\infty}^{\infty} \{v_n[g_{m-n} + j(m - n)\omega_0 C_{m-n}] + j(\omega_1 + n\omega_0)C_{m-n}v_n\} \\ &= \sum_{n=-\infty}^{\infty} [g_{m-n} + j(\omega_1 + m\omega_0)C_{m-n}]v_n \end{aligned} \tag{12.30}$$

In matrix form we get

$$[i_n] = [Y_{nm}][v_m] \tag{12.31}$$

where $Y_{nm} = g_{n-m} + j(\omega_1 + n\omega_0)C_{n-m}$. The $[Y_{nm}]$ matrix describes the mixing action of the diode. It relates the harmonic amplitudes of the RF signal current through the diode to the harmonic amplitudes of the RF signal voltage across the diode junction. Each harmonic current $i_n e^{j(\omega_1 + n\omega_0)t}$ has a constraint imposed on it by the external circuit (embedding network) in that the voltage around any closed loop, including the diode, must sum to zero (Kirchhoff's law) at each frequency $\omega_1 + n\omega_0$. These circuit constraints will be given later.

In order to determine the elements in the mixer conversion matrix $[Y_{nm}]$, it is necessary to know the function that describes the junction capacitance $C_j(v_j)$ in terms of the junction voltage. The solutions for i_T, i_d, i_c, and v_j are constrained by circuit relations imposed by the embedding network. The Thévenin equivalent circuit of the mixer for dc currents is shown in Fig. 12.30*b*. At $\omega = 0$ we must have

$$V_B = I_0(R_0 + R_s) + V_0 \tag{12.32a}$$

At the local-oscillator frequency, we have

$$\frac{V_0}{2} = I_{T1}Z(\omega_0) + V_1 \tag{12.32b}$$

$$\frac{V_0}{2} = I_{T-1}Z(-\omega_0) + V_{-1} = I_{T1}^*Z^*(\omega_0) + V_1^* \tag{12.32c}$$

At all other harmonics of the local-oscillator frequency

$$0 = I_{Tn}Z(n\omega_0) + V_n \tag{12.32d}$$

These equations along with the known function $C_j(v_j)$ and the ideal diode equation must be solved in order to find the current amplitudes I_n, the voltage amplitudes V_n, and the capacitance elements C_n. Numerical methods are required because of the nonlinear behavior of the diode.

After the elements of the mixer conversion matrix have been found, we can solve the small RF signal problem by solving a linear network problem. The embedding network can be replaced by a Norton equivalent network at each frequency $\omega_1 + m\omega_0$ as shown in Fig. 12.31. The RF signal source can be chosen as $I_{eo}e^{j\omega_1 t}$. The circuit equation at the frequency ω_1 is

$$I_{eo} = \frac{v_o}{Z(\omega_1)} + i_0 = \frac{v_o}{Z(\omega_1)} + \sum_{m=-\infty}^{\infty} Y_{0m}v_m \tag{12.33a}$$

At the frequency $\omega_1 + n\omega_0$ the circuit equation is

$$\frac{v_n}{Z(\omega_1 + n\omega_0)} + \sum_{m=-\infty}^{\infty} Y_{nm}v_m = 0 \tag{12.33b}$$

At the frequency $\omega_1 + n\omega_0$ the current through the impedance $Z(\omega_1 + n\omega_0)$

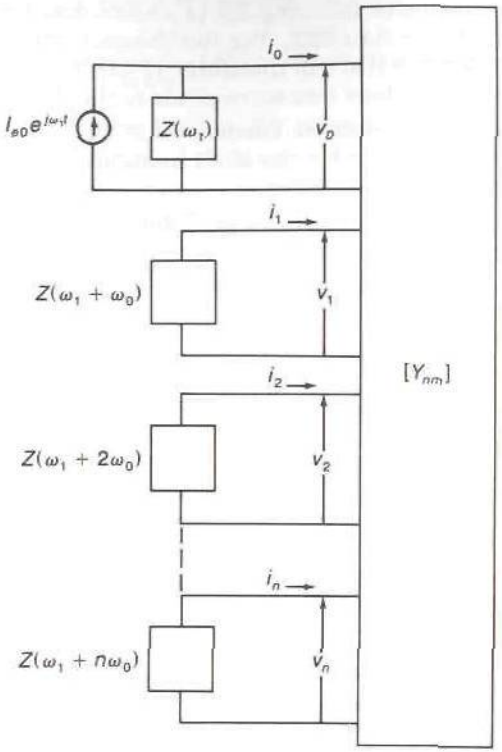


FIGURE 12.31
Equivalent circuit for a mixer for linear RF signal response.

is given by the first term on the left-hand side in (12.33*b*). The current i_n is given by the second term in (12.33*b*). The equation states that the sum of these two currents must be zero.

In principle, the solution of the mixer problem using harmonic balancing is straightforward. In practice, it is a complex problem that requires numerical evaluation using a computer program. Space limitations do not allow us to develop the method in greater detail. However, some of the references given at the end of this chapter provide more details.

PROBLEMS

12.1. The three-port scattering-matrix parameters are given by (12.10). Verify that the sum of the elements in any row or column equals one.

12.2. In a transistor a series impedance producing a reflection coefficient Γ is connected in the common lead. Show that the equivalent two-port network scattering-matrix parameters are given by (12.11) upon replacing $\hat{S}_{33} + 1$ by $\hat{S}_{33} - \Gamma^{-1}$.

12.3. For the unstable bipolar transistor in Example 12.2, find the smallest series resistance to be added in the common emitter lead to obtain an absolutely stable two-port network. Use the computer program TRIPORT.

12.4. Design an oscillator following the procedure described in Example 12.2 but assume that $\Gamma_s = 0.9\angle 225°$. For this design it will be necessary to design a matching network that will transform a 50-Ω input termination into a source impedance at the base that corresponds to the chosen value of Γ_s.

12.5. Redesign the oscillator in Example 12.3 using a resonator coupling with $n^2 = 2/7$. Specify the lengths of all transmission lines used in the oscillator circuit.

12.6. Design a 10-GHz common gate oscillator using a GaAs MESFET having the following scattering-matrix parameters:

$$S_{11} = 1.1\angle 160° \qquad S_{12} = 0.16\angle 130°$$

$$S_{21} = 2.7\angle -62° \qquad S_{22} = 1.4\angle -57°$$

Use a pure reactive termination for the input (source terminal) and a circuit topology similar to that in Fig. 12.17. The parameters of the stability circles can be found by using the MICROAMP program (arbitrary values for Γ_m and noise resistance can be entered. Arbitrary gain circles can also be chosen since this information is not used). The output impedance for the chosen value of Γ_s can be found using the TWOPORT program but the scattering-matrix parameters must be entered in rectangular coordinate form (real and imaginary parts).

12.7. Design an 8-GHz MESFET oscillator similar to that in Example 12.3 using a device with the following common source scattering-matrix parameters:

$$S_{11} = 0.25\angle 8° \qquad S_{12} = 0.11\angle 145°$$

$$S_{21} = 3.5\angle 168° \qquad S_{22} = 0.43\angle -60°$$

This device is absolutely stable so a series reactance must be connected into the common source lead. Choose this reactance to obtain a value for the stability parameter K that is less than 0.6.

12.8. For the oscillator in Fig. 12.19 the dc gate voltage is applied through a high-impedance line $\lambda/4$ long. The feed line is connected at a point where $\overline{Z}_{in} = \overline{R}_{in}$. Find the distance from the load end at which the feed line should be connected.

12.9. Show that a term such as $V^3 \cos^3 \omega_1 t$ has signal components at the two frequencies ω_1 and $3\omega_1$, while $V^4 \cos^4 \omega_1 t$ has components at the frequencies 0, $2\omega_1$, and $4\omega_1$.

12.10. Show that a term such as $V_1^2 V_2^2 \cos^2 \omega_1 t \cos^2 \omega_2 t$ has signal components at $\omega = 0$, $2\omega_1$, $2\omega_2$, and $2(\omega_1 \pm \omega_2)$.

12.11. Find all of the frequencies generated from the term multiplied by a_4 in (12.18).

12.12. When $a_3|V_{IF}|^3 \ll a_2|V_{IF}|$, show that, by using (12.20) to evaluate the term $-\frac{3}{4}a_3|V_{IF}|^2$ in the expression for the diode current, in place of (12.20) the IF input voltage is given by (12.22).

12.13. When (12.19) with terms up to a_2 is used to evaluate i_{IF} and v_{IF1}^2 and v_{IF2}^2, show that the term $-3a_3(v_{IF1}^2 v_{IF2} + v_{IF2}^2 v_{IF1})$ results in the following contribution to the intermodulation component of the diode current in the IF

amplifier passband:

$$\frac{-3a_3a_2^2V_0^2}{2|1+a_1Z_{IF}|^2}[V_1^2V_2\cos(\omega_0-2\omega_1+\omega_2)t+V_1V_2^2\cos(\omega_0-2\omega_2+\omega_1)t]\frac{a_2V_0|Z_{IF}|}{2|1+a_1Z_{IF}}$$

Assume that Z_{IF} is constant in the passband.

12.14. A mixer has a noise figure of 8 dB and a conversion loss of 10 dB. The IF amplifier noise figure is 3 dB. Find the noise figure of the combination.

12.15. Find the noise figure of a mixer and IF amplifier combination if the conversion loss is 5 dB, the mixer noise figure is 6 dB, and the IF amplifier noise figure is 4 dB.

REFERENCES

1. Chang, K.: "Handbook of Microwave and Optical Components. Microwave Solid State Components," vol. 2, John Wiley & Sons, Inc., New York, 1990.
2. Vendelin, G. D., A. M. Pavio, and U. L. Rohde: "Microwave Circuit Design Using Linear and Non-Linear Techniques," John Wiley & Sons, Inc., New York, 1990.
3. Bahl I., and P. Bhartia: "Microwave Solid-State Circuit Design," John Wiley & Sons, Inc., New York, 1988.
4. Maas, S. A.: "Microwave Mixers," Artech House Books, Norwood, Mass., 1986.
5. Maas, S. A.: Two-Tone Intermodulation in Diode Mixers, *IEEE Trans.*, vol. MTT-35, pp. 307–314, 1987.
6. Kollberg, E. L. (ed.): "Microwave and Millimeter Wave Mixers," IEEE Press, Piscataway, N.J., 1984.
7. Kerr, A. R.: A Technique for Determining the Local Oscillator Waveforms in a Microwave Mixer, *IEEE Trans.*, vol. MTT-23, pp. 828–831, 1975.
8. Held, D. N., and A. R. Kerr: Conversion Loss and Noise of Microwave and Millimeter-Wave Mixers. I. Theory, *IEEE Trans.*, vol. MTT-26, pp. 49–55, 1978.

APPENDIX I

USEFUL RELATIONS FROM VECTOR ANALYSIS

I.1 VECTOR ALGEBRA

Let vectors **A** and **B** be expressed as components along unit vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ in a right-hand orthogonal coordinate system. Then

$$\mathbf{A} \pm \mathbf{B} = (A_1 \pm B_1)\mathbf{a}_1 + (A_2 \pm B_2)\mathbf{a}_2 + (A_3 \pm B_3)\mathbf{a}_3 \tag{I.1}$$

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos\theta = A_1B_1 + A_2B_2 + A_3B_3 \tag{I.2}$$

where θ is the angle between **A** and **B**.

$$\mathbf{A} \times \mathbf{B} = \mathbf{a}_1(A_2B_3 - A_3B_2) + \mathbf{a}_2(A_3B_1 - A_1B_3) + \mathbf{a}_3(A_1B_2 - A_2B_1) \tag{I.3a}$$

$$|\mathbf{A} \times \mathbf{B}| = |\mathbf{A}||\mathbf{B}|\sin\theta \tag{I.3b}$$

$$\mathbf{A} \cdot \mathbf{B} \times \mathbf{C} = \mathbf{A} \times \mathbf{B} \cdot \mathbf{C} = \mathbf{C} \times \mathbf{A} \cdot \mathbf{B} \tag{I.4}$$

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \tag{I.5}$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C} \tag{I.6}$$

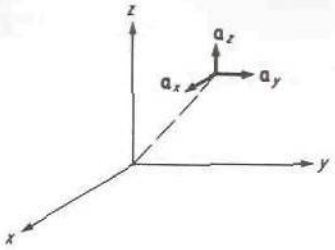


FIGURE I.1
Rectangular coordinates.

I.2 VECTOR OPERATIONS IN COMMON COORDINATE SYSTEMS

Rectangular Coordinates

$$\nabla\Phi = \mathbf{a}_x\frac{\partial\Phi}{\partial x} + \mathbf{a}_y\frac{\partial\Phi}{\partial y} + \mathbf{a}_z\frac{\partial\Phi}{\partial z} \tag{I.7}$$

$$\operatorname{div}\mathbf{A} = \nabla\cdot\mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{I.8}$$

$$\operatorname{curl}\mathbf{A} = \nabla\times\mathbf{A} = \mathbf{a}_x\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + \mathbf{a}_y\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + \mathbf{a}_z\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \tag{I.9}$$

$$\nabla^2\Phi = \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} \tag{I.10}$$

$$\nabla^2\mathbf{A} = \mathbf{a}_x\nabla^2A_x + \mathbf{a}_y\nabla^2A_y + \mathbf{a}_z\nabla^2A_z \tag{I.11}$$

Cylindrical Coordinates

$$\nabla\Phi = \mathbf{a}_r\frac{\partial\Phi}{\partial r} + \mathbf{a}_\phi\frac{1}{r}\frac{\partial\Phi}{\partial\phi} + \mathbf{a}_z\frac{\partial\Phi}{\partial z} \tag{I.12}$$

$$\nabla\cdot\mathbf{A} = \frac{1}{r}\frac{\partial}{\partial r}(rA_r) + \frac{1}{r}\frac{\partial A_\phi}{\partial\phi} + \frac{\partial A_z}{\partial z} \tag{I.13}$$

$$\nabla\times\mathbf{A} = \mathbf{a}_r\left(\frac{1}{r}\frac{\partial A_z}{\partial\phi} - \frac{\partial A_\phi}{\partial z}\right) + \mathbf{a}_\phi\left(\frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}\right) + \mathbf{a}_z\left[\frac{1}{r}\frac{\partial(rA_\phi)}{\partial r} - \frac{1}{r}\frac{\partial A_r}{\partial\phi}\right] \tag{I.14}$$

$$\nabla^2\Phi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\Phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\phi^2} + \frac{\partial^2\Phi}{\partial z^2} \tag{I1.5}$$

$$\nabla^2\mathbf{A} = \nabla\nabla\cdot\mathbf{A} - \nabla\times\nabla\times\mathbf{A} \tag{I.16}$$

Note that $\nabla^2\mathbf{A} \neq \mathbf{a}_r\nabla^2A_r + \mathbf{a}_\phi\nabla^2A_\phi + \mathbf{a}_z\nabla^2A_z$ since $\nabla^2\mathbf{a}_r\,A_r \neq \mathbf{a}_r\nabla^2A_r$, etc., because the orientation of the unit vectors $\mathbf{a}_r, \mathbf{a}_\phi$ varies with the coordinates r, ϕ.

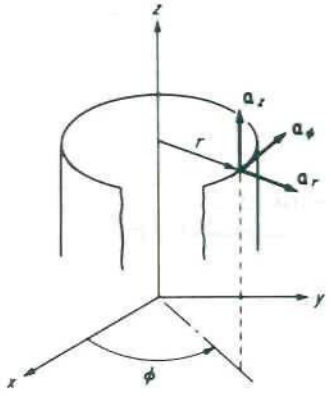


FIGURE I.2
Cylindrical coordinates.

Spherical Coordinates

$$\nabla\Phi = \mathbf{a}_r\frac{\partial\Phi}{\partial r} + \mathbf{a}_\theta\frac{1}{r}\frac{\partial\Phi}{\partial\theta} + \frac{\mathbf{a}_\phi}{r\sin\theta}\frac{\partial\Phi}{\partial\phi} \tag{I.17}$$

$$\nabla\cdot\mathbf{A} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2A_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\,A_\theta) + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial\phi} \tag{I.18}$$

$$\nabla\times\mathbf{A} = \frac{\mathbf{a}_r}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(A_\phi\sin\theta) - \frac{\partial A_\theta}{\partial\phi}\right] + \frac{\mathbf{a}_\theta}{r}\left[\frac{1}{\sin\theta}\frac{\partial A_r}{\partial\phi} - \frac{\partial}{\partial r}(rA_\phi)\right] + \frac{\mathbf{a}_\phi}{r}\left[\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right] \tag{I.19}$$

$$\nabla^2\Phi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Phi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Phi}{\partial\phi^2} \tag{I.20}$$

$$\nabla^2\mathbf{A} = \nabla\nabla\cdot\mathbf{A} - \nabla\times\nabla\times\mathbf{A} \tag{I.21}$$

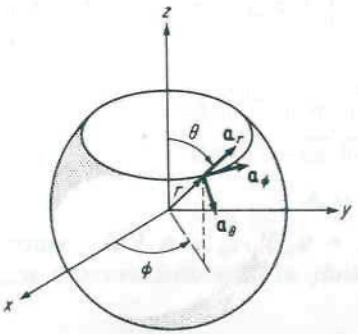


FIGURE I.3
Spherical coordinates.

I.3 VECTOR IDENTITIES

$$\nabla(\Phi\psi) = \psi\nabla\Phi + \Phi\nabla\psi \tag{I.22}$$

$$\nabla\cdot(\psi\mathbf{A}) = \mathbf{A}\cdot\nabla\psi + \psi\nabla\cdot\mathbf{A} \tag{I.23}$$

$$\nabla\cdot(\mathbf{A}\times\mathbf{B}) = (\nabla\times\mathbf{A})\cdot\mathbf{B} - (\nabla\times\mathbf{B})\cdot\mathbf{A} \tag{I.24}$$

$$\nabla\times(\psi\mathbf{A}) = (\nabla\psi)\times\mathbf{A} + \psi\nabla\times\mathbf{A} \tag{I.25}$$

$$\nabla\times(\mathbf{A}\times\mathbf{B}) = \mathbf{A}\nabla\cdot\mathbf{B} - \mathbf{B}\nabla\cdot\mathbf{A} + (\mathbf{B}\cdot\nabla)\mathbf{A} - (\mathbf{A}\cdot\nabla)\mathbf{B} \tag{I.26}$$

$$\nabla(\mathbf{A}\cdot\mathbf{B}) = (\mathbf{A}\cdot\nabla)\mathbf{B} + (\mathbf{B}\cdot\nabla)\mathbf{A} + \mathbf{A}\times(\nabla\times\mathbf{B}) + \mathbf{B}\times(\nabla\times\mathbf{A}) \tag{I.27}$$

$$\nabla\cdot\nabla\Phi = \nabla^2\Phi \tag{I.28}$$

$$\nabla\cdot\nabla\times\mathbf{A} = 0 \tag{I.29}$$

$$\nabla\times\nabla\Phi = 0 \tag{I.30}$$

$$\nabla\times\nabla\times\mathbf{A} = \nabla\nabla\cdot\mathbf{A} - \nabla^2\mathbf{A} \tag{I.31}$$

If $\mathbf{A}$ and Φ are continuous functions with at least piecewise continuous first derivatives in V and on S (or on S and the contour C bounding S),

$$\int_V \nabla\Phi\, dV = \oint_S \Phi\, d\mathbf{S} \tag{I.32}$$

$$\int_V \nabla\cdot\mathbf{A}\, dV = \oint_S \mathbf{A}\cdot dS \qquad \text{(divergence theorem)} \tag{I.33}$$

$$\int_V \nabla\times\mathbf{A}\, dV = \oint_S \mathbf{n}\times\mathbf{A}\, dS \qquad d\mathbf{S} = \mathbf{n}\, dS \tag{I.34}$$

$$\int_S \mathbf{n}\times\nabla\Phi\, dS = \oint_C \Phi\, d\mathbf{l} \tag{I.35}$$

$$\int_S \nabla\times\mathbf{A}\cdot d\mathbf{S} = \oint_C \mathbf{A}\cdot d\mathbf{l} \qquad \text{(Stokes' theorem)} \tag{I.36}$$

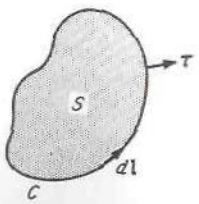


FIGURE I.4
Surface S bounded by contour C.

I.4 GREEN'S IDENTITIES

If $\mathbf{A}$, $\mathbf{B}$, Φ, and ψ are continuous with piecewise continuous first derivatives,

$$\int_V (\nabla\Phi \cdot \nabla\psi + \psi \nabla^2\Phi)\, dV = \oint_S \psi \nabla\Phi \cdot d\mathbf{S} \tag{I.37}$$

which is Green's first identity. Green's second identity is

$$\int_V (\psi \nabla^2\Phi - \Phi \nabla^2\psi)\, dV = \oint (\psi \nabla\Phi - \Phi \nabla\psi) \cdot d\mathbf{S} \tag{I.38}$$

In two dimensions (I.37) becomes

$$\int_S (\nabla_t\Phi \cdot \nabla_t\psi + \psi \nabla_t^2\Phi)\, dS = \oint_C \psi \nabla_t\Phi \cdot \boldsymbol{\tau}\, dl \tag{I.39}$$

where ∇_t is the del operator in two dimensions and $\boldsymbol{\tau}$ is a unit vector normal to C and in the plane of S. The two-dimensional form of (I.38) is similar.

The vector forms of Green's identities are

$$\int_V \nabla \cdot (\mathbf{A} \times \nabla \times \mathbf{B})\, dV = \int_V [(\nabla \times \mathbf{A}) \cdot (\nabla \times \mathbf{B}) - \mathbf{A} \cdot \nabla \times \nabla \times \mathbf{B}]\, dV$$

$$= \oint_S \mathbf{A} \times (\nabla \times \mathbf{B}) \cdot d\mathbf{S} \tag{I.40}$$

$$\int_V (\mathbf{B} \cdot \nabla \times \nabla \times \mathbf{A} - \mathbf{A} \cdot \nabla \times \nabla \times \mathbf{B})\, dV$$

$$= \oint_S [\mathbf{A} \times (\nabla \times \mathbf{B}) - \mathbf{B} \times (\nabla \times \mathbf{A})] \cdot d\mathbf{S} \tag{I.41}$$

APPENDIX II

BESSEL FUNCTIONS

II.1 ORDINARY BESSEL FUNCTIONS

The wave equation and Helmholtz's and Laplace's equations are separable in cylindrical coordinates. The differential equation describing the radial dependence of the solution is Bessel's differential equation. Bessel's equation is

$$\frac{1}{r}\frac{d}{dr}r\frac{df}{dr} + \left(k^2 - \frac{n^2}{r^2}\right)f = 0 \tag{II.1}$$

When k^2 is real and positive, the two independent solutions of Bessel's equation are called Bessel functions of the first and second kind, denoted by $J_n(kr)$ and $Y_n(kr)$, respectively. These solutions may be expressed as power series as follows:

$$J_n(kr) = \sum_{m=0}^{\infty} \frac{(-1)^m (kr/2)^{n+2m}}{m!(n+m)!} \tag{II.2}$$

$$\begin{aligned} Y_n(kr) = {} & \frac{2}{\pi}\left(\gamma + \ln\frac{kr}{2}\right)J_n(kr) - \frac{1}{\pi}\sum_{m=0}^{n-1}\frac{(n-m-1)!}{m!}\left(\frac{2}{kr}\right)^{n-2m} \\ & - \frac{1}{\pi}\sum_{m=0}^{\infty}\frac{(-1)^m (kr/2)^{n+2m}}{m!(n+m)!} \\ & \times\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m} + 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n+m}\right) \end{aligned} \tag{II.3}$$

where $\gamma = 0.5772$ is Euler's constant. The subscript n denotes the order of the function and is usually an integer in physical problems. The Y_n functions become infinite at $r = 0$. For large values of kr, the Bessel functions approach damped sinusoids:

$$\lim_{r\to\infty} J_n(kr) = \sqrt{\frac{2}{\pi kr}} \cos\left(kr - \frac{\pi}{4} - \frac{n\pi}{2}\right) \tag{II.4a}$$

$$\lim_{r\to\infty} Y_n(kr) = \sqrt{\frac{2}{\pi kr}} \sin\left(kr - \frac{\pi}{4} - \frac{n\pi}{2}\right) \tag{II.4b}$$

A few of the lowest-order Bessel functions are plotted in Fig. II.1.

To represent radially propagating waves, linear combinations of the J_n and Y_n are formed, called Hankel functions of the first and second kind. Thus the Hankel function of the first kind is

$$H_n^1(kr) = J_n(kr) + jY_n(kr) \tag{II.5a}$$

and the Hankel function of the second kind is given by

$$H_n^2(kr) = J_n(kr) - jY_n(kr) \tag{II.5b}$$

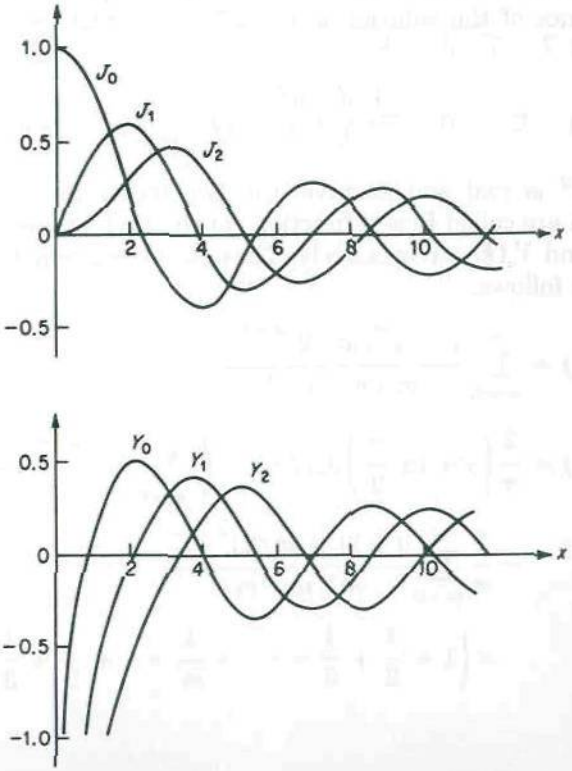


FIGURE II.1
Ordinary Bessel functions.

For large values of kr, the Hankel functions are given by the following expressions:

$$H_n^1(kr) \sim \sqrt{\frac{2}{kr\pi}}\, e^{j(kr-\pi/4-n\pi/2)} \tag{II.6a}$$

$$H_n^2(kr) \sim \sqrt{\frac{2}{kr\pi}}\, e^{-j(kr-\pi/4-n\pi/2)} \tag{II.6b}$$

Some useful relations that hold for any of the Bessel functions J_n, Y_n, or H_n are given below, where Z_n denotes J_n, Y_n, or H_n.

$$xZ_n'(x) = nZ_n(x) - xZ_{n+1}(x) = -nZ_n(x) + xZ_{n-1}(x) \tag{II.7}$$

where the prime denotes differentiation with respect to x.

$$\int^x Z_n(kx)Z_n(lx)x\,dx = \frac{x}{k^2-l^2}\left[kZ_n(lx)Z_{n+1}(kx) - lZ_n(kx)Z_{n+1}(lx)\right] \tag{II.8}$$

$$\int^x Z_n^2(kx)x\,dx = \frac{x^2}{2}\left[Z_n'^2(kx) + \left(1 - \frac{n^2}{k^2x^2}\right)Z_n^2(kx)\right] \tag{II.9}$$

II.2 MODIFIED BESSEL FUNCTIONS

When k^2 is negative, k is pure imaginary. If we let $k = jh$, the solutions are given by $J_n(jhr)$ and $Y_n(jhr)$. However, for convenience, new modified Bessel functions are introduced and denoted by $I_n(hr)$ and $K_n(hr)$. The modified Bessel function of the first kind is $I_n(hr)$, and is given by

$$I_n(hr) = j^{-n}J_n(jhr) = j^nJ_n(-jhr) \tag{II.10}$$

and the modified Bessel function of the second kind is given by

$$K_n(hr) = \frac{\pi}{2}j^{n+1}\left[J_n(jhr) + jY_n(jhr)\right] = \frac{\pi}{2}j^{n+1}H_n^1(jhr) \tag{II.11}$$

For large values of hr we have

$$I_n(hr) \sim \frac{e^{hr}}{\sqrt{2\pi hr}} \tag{II.12a}$$

$$K_n(hr) \sim \sqrt{\frac{\pi}{2hr}}\, e^{-hr} \tag{II.12b}$$

The first few modified Bessel functions are plotted in Fig. II.2.

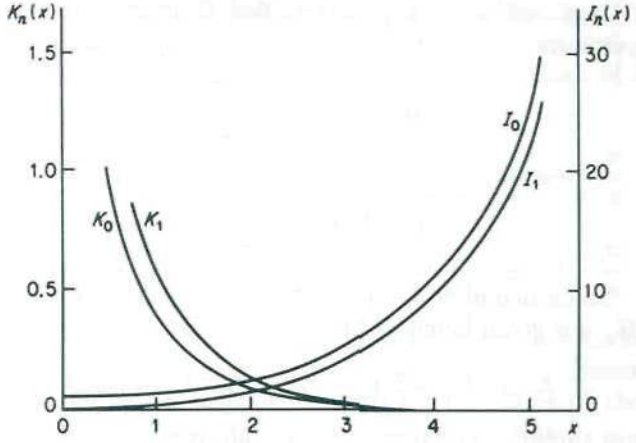


FIGURE II.2
Modified Bessel functions.

A number of useful relations that hold for the modified Bessel functions are given:

$$xI_n'(x) = nI_n(x) + xI_{n+1}(x) = -nI_n(x) + xI_{n-1}(x) \tag{II.13a}$$

$$I_0'(x) = I_1(x) \tag{II.13b}$$

$$\int^x x^{-n} I_{n+1}(x)\,dx = x^{-n} I_n(x) \tag{II.14a}$$

$$\int^x x^{n} I_{n-1}(x)\,dx = x^{n} I_n(x) \tag{II.14b}$$

When $n > -1$, we have

$$\int_0^x I_n(kx) I_n(lx) x\,dx = \frac{x}{k^2 - l^2}\left[kI_n(lx) I_{n+1}(kx) - lI_n(kx) I_{n+1}(lx)\right] \tag{II.15}$$

$$\int_0^x I_n^2(kx) x\,dx = -\frac{x^2}{2}\left[I_n'^2(kx) - \left(1 + \frac{n^2}{k^2x^2}\right) I_n^2(kx)\right] \tag{II.16}$$

$$xK_n'(x) = nK_n(x) - xK_{n+1}(x) = -nK_n(x) - xK_{n-1}(x) \tag{II.17a}$$

$$K_0'(x) = -K_1(x) \tag{II.17b}$$

$$\int^x x^{-n} K_{n+1}(x)\,dx = -x^{-n} K_n(x) \tag{II.18a}$$

$$\int^x x^{n} K_{n-1}(x)\,dx = -x^{n} K_n(x) \tag{II.18b}$$

When $\mathrm{Re}(k+1) > 0$, we have

$$\int_x^\infty K_n(kx)K_n(lx)x\,dx = \frac{x}{k^2-l^2}\left[kK_n(lx)K_{n+1}(kx) - lK_n(kx)K_{n+1}(lx)\right] \tag{II.19}$$

For $\mathrm{Re}\,k > 0$

$$\int_x^\infty xK_n^2(kx)\,dx = \frac{x^2}{2}\left[K_n'^2(x) - \left(1+\frac{n^2}{k^2x^2}\right)K_n^2(kx)\right] \tag{II.20}$$

REFERENCES

1. Watson, G. N.: "Theory of Bessel Functions," Cambridge University Press, New York, 1922.
2. McLachlan, N. W.: "Bessel Functions for Engineers," 2d ed., Oxford University Press, Fair Lawn, N.J., 1948.
3. Bowman, F.: "Introduction to Bessel Functions," Dover Publications, Inc., New York, 1958.

APPENDIX III

CONFORMAL MAPPING TECHNIQUES

The determination of the distributed capacitance and inductance of a transmission line requires a solution of Laplace's equation in two dimensions. For transmission-line structures like those used in many planar transmission lines, it is difficult to construct solutions for Laplace's equation if one used the system of coordinates that is the natural one for describing the transmission-line configuration. A powerful method for solving two-dimensional potential problems is to use conformal mapping to map the boundaries into a simpler configuration for which solutions to Laplace's equation are easily found. This conformal mapping technique is equivalent to a coordinate transformation and its application to planar transmission lines is described in this appendix.

III.1 CONFORMAL MAPPING

Let $\zeta = \sigma + j\eta$ be a complex variable and consider the function

$$\sin \zeta = \sin(\sigma + j\eta) = \sin \sigma \cosh \eta + j \cos \sigma \sinh \eta$$

This function is periodic along the real σ axis with a period of 2π. The function takes on all its possible values in a strip extending from $-\pi/2$ to $\pi/2$ along σ as shown in Fig. III.1. Along the contour labeled *A-B-C-D* shown in the figure, the $\sin \zeta$ function goes from $-\infty$ to -1 and then to $+1$ and finally to $+\infty$. If we let a new complex variable W be defined by

$$W = u + jv = \sin \zeta \tag{III.1}$$

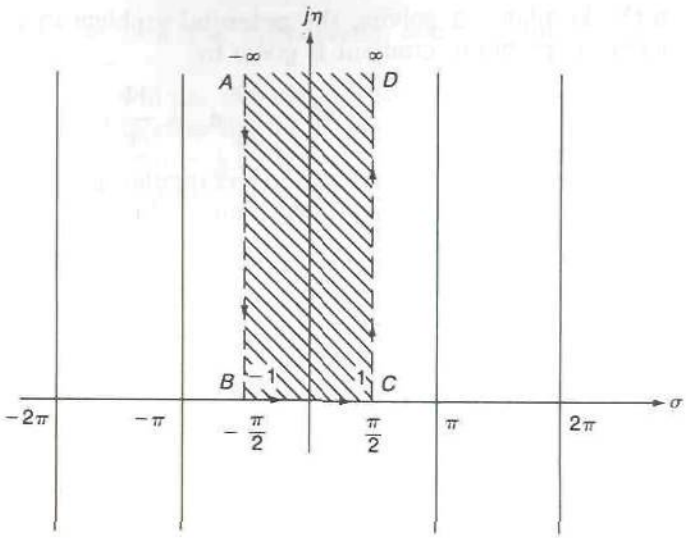


FIGURE III.1
The sin ζ function.

then the chosen contour in the ζ plane maps into the real axis in the W plane as shown in Fig. III.2. All values of ζ in the cross-hatched region shown in Fig. III.1 map into the upper half of the W plane.

In the W plane, which we will regard as our real physical space, let a total charge Q per meter be placed on each side of the strip extending from $u = -1$ to $u = 1$. Furthermore, let us regard the strip as a conducting strip at constant potential. By symmetry the two boundaries $-\infty < u < -1$ and $1 < u < \infty$ are magnetic walls on which $\partial\Phi/\partial v = 0$. We can find the solution for Laplace's equation

$$\frac{\partial^2 \Phi}{\partial u^2} + \frac{\partial^2 \Phi}{\partial v^2} = 0$$

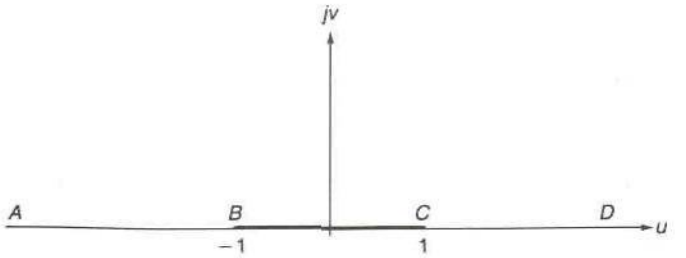


FIGURE III.2
The mapping $W = \sin \zeta$.

in the W plane by solving the potential problem in the ζ plane. In the W plane the potential gradient is given by

$$\nabla\Phi = \frac{\partial\Phi}{\partial u}\mathbf{a}_u + \frac{\partial\Phi}{\partial v}\mathbf{a}_v$$

If we regard u and v as rectangular coordinates, then σ and η represent new curvilinear coordinates. In the σ, η coordinate system, Laplace's equation has the form

$$\nabla^2\Phi = \frac{\partial}{\partial\sigma}\frac{h_\eta}{h_\sigma}\frac{\partial\Phi}{\partial\sigma} + \frac{\partial}{\partial\eta}\frac{h_\sigma}{h_\eta}\frac{\partial\Phi}{\partial\eta} = 0$$

where h_η and h_σ are the metric coefficients. For a coordinate transformation that is obtained through a conformal mapping, the metric coefficients are equal and given by†

$$h_\sigma = h_\eta = \left|\frac{dW}{d\zeta}\right|$$

Consequently, Laplace's equation reduces to the same form as it has in a rectangular coordinate system. Thus σ and η can be treated as rectangular coordinates when solving for the potential field. However, the gradient of Φ must be evaluated using

$$\nabla\Phi = \frac{1}{h_\sigma}\frac{\partial\Phi}{\partial\sigma}\mathbf{a}_\sigma + \frac{1}{h_\eta}\frac{\partial\Phi}{\partial\eta}\mathbf{a}_\eta = \nabla\Phi\Big|_\zeta\left|\frac{d\zeta}{dW}\right| = \nabla\Phi\Big|_W \qquad \text{(III.2)}$$

where $\nabla\Phi|_\zeta$ represents the gradient in the ζ plane obtained by treating σ and η as rectangular coordinates. The conformal mapping of the boundary of a polygon in the ζ plane into the real axis in the W plane is called a Schwarz-Christoffel transformation.

For a differential element dW along a contour in the W plane, the corresponding differential element along the mapped contour in the ζ plane is $d\zeta = (d\zeta/dW)\,dW$. The angle that $d\zeta$ makes with the real axis is the sum of the angles of $d\zeta/dW$ and dW. When the contour in the W plane is the real axis, the angle of dW is zero. When $d\zeta/dW$ has the form $\sqrt{W - W_1}\,/\sqrt{W - W_2}$, then the angle of $\sqrt{W - W_1}$ is $\pi/2$ for $W < W_1$ and zero for $W > W_1$, while that for $1/\sqrt{W - W_2}$ is $-\pi/2$ for $W < W_2$ and zero for $W > W_2$. Hence, as W moves along the real axis, there will be a change of $-\pi/2$ in the angle of $d\zeta$ as W moves past the point W_1 and a change of $\pi/2$ when W moves past the point W_2. This causes the contour in the ζ plane to change direction in a step-like fashion by $\mp 90°$. This property is

†R. E. Collin, "Field Theory of Guided Waves," 2nd ed., Chap. 4, IEEE Press, Pisacataway, N.J., 1991.

used to pick the appropriate mapping functions that are used in this appendix.

We will use the above results to find the charge distribution along the conducting strip in the W plane. In the ζ plane the presence of magnetic walls along $\sigma = \pm\pi/2$, $\eta \geq 0$ requires that the electric field be uniform and in the η direction. Hence the charge is uniformly distributed on the lower boundary $-\pi/2 \leq \sigma \leq \pi/2$ with density Q/π. Thus the potential gradient is given by

$$-\epsilon_0 \frac{\partial \Phi}{\partial \eta} = \frac{Q}{\pi}$$

so $\nabla\Phi|_\zeta = -Q/\pi\epsilon_0$. From (III.2) we find that in the W plane

$$\rho_s = -\epsilon_0 \nabla\Phi|_W = -\epsilon_0 \nabla\Phi|_\zeta \left| \frac{d\zeta}{dW} \right|$$

Upon using $dW/d\zeta = \cos\zeta = (1 - \sin^2\zeta)^{1/2} = (1 - W^2)^{1/2}$ for $W = u$, we obtain

$$\rho_s = \frac{Q}{\pi\sqrt{1 - u^2}} \tag{III.3}$$

for the charge density on one side of an isolated conducting strip two units wide and located in air or completely surrounded by dielectric. This example clearly shows the power of the conformal mapping method of solving potential problems. However, it does depend on our ability to find a mapping function that will transform the boundaries of the physical problem into a configuration for which the solution to Laplace's equation is easily obtained. In the transformed plane the new coordinates may be treated as rectangular coordinates provided we relate the gradients in the two planes by the relation (III.2). Furthermore, the capacitance between conductors remains invariant under a conformal mapping, so that it may be found for the transformed configuration.

III.2 ELLIPTIC SINE FUNCTION

For the function $W = \sin\zeta$, we have $dW/d\zeta = \cos\zeta = (1 - W^2)^{1/2}$, so

$$\frac{d\zeta}{dW} = \frac{1}{\sqrt{1 - W^2}} \tag{III.4}$$

which gives the inverse function

$$\zeta = \sin^{-1} W = \int_0^W \frac{dW}{\sqrt{1 - W^2}} \tag{III.5}$$

The period along the σ axis is obtained from the integral from 0 to 1 which

gives one quarter of the period, i.e.,

$$\frac{\pi}{2} = \int_0^1 \frac{dW}{\sqrt{1 - W^2}} \tag{III.6}$$

A function that is much more useful in the solution of a number of planar-transmission-line problems is a function that is periodic along both the σ and η axis in the complex ζ plane because it would take on all of its possible values inside a rectangle. A doubly periodic function is the elliptic sine function, which is expressed as

$$W = \mathrm{sn}(\zeta, k) \tag{III.7}$$

The parameter k is called the modulus and determines the two periods. The elliptic sine function has a period of 4K along σ and $2K'$ along η. The inverse function is given by

$$\zeta = \mathrm{sn}^{-1}(W, k) = \int_0^W \frac{dW}{\sqrt{(1 - W^2)(1 - k^2W^2)}} \tag{III.8}$$

The two quantities K and K' are given by

$$K = \int_0^1 \frac{dW}{\sqrt{(1 - W^2)(1 - k^2W^2)}} \tag{III.9a}$$

$$K + jK' = \int_0^{1/k} \frac{dW}{\sqrt{(1 - W^2)(1 - k^2W^2)}} \tag{III.9b}$$

or

$$K' = \int_1^{1/k} \frac{dW}{\sqrt{(W^2 - 1)(1 - k^2W^2)}} \tag{III.9c}$$

By using the substitution $W = \sin\theta$, the expression for K becomes

$$K = K(k) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2\theta}} = F\left(k, \frac{\pi}{2}\right) \tag{III.10}$$

which is the complete elliptic integral of the first kind. In the expression for K' we can let $k^2(W^2 - 1) = (1 - k^2)\cos^2\theta$, which reduces the expression to the form

$$K' = K'(k) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k'^2 \sin^2\theta}} = F\left(k', \frac{\pi}{2}\right) = K(k') \tag{III.11}$$

where the complementary modulus $k' = \sqrt{1 - k^2}$. From (III.8) and (III.9) we also see that

$$\mathrm{sn}(0, k) = 0 \tag{III.12a}$$

$$\mathrm{sn}(K, k) = 1 \tag{III.12b}$$

$$\mathrm{sn}(K + jK', k) = \frac{1}{k} \tag{III.12c}$$

A further useful value is

$$\text{sn}(jK', k) = \pm\infty \tag{III.12d}$$

The elliptic sine function is a generalization of the sine function and reduces to the latter when $k = 0$.

Two useful approximations for the K and K' parameters for the cases $k \ll 1$ and $k' \ll 1$ are

$$K(k) = K'(k') = \frac{\pi}{2}\left(1 + \frac{k^2}{4} + \frac{9}{64}k^4\right) \qquad k \le 0.4 \tag{III.13a}$$

$$K'(k') = K(k) = \left(1 + \frac{k'^2}{4} + \frac{9}{64}k'^4\right)\ln\frac{4}{k'} - \frac{k'^2}{4} - \frac{21}{168}k'^4 \qquad k \ge 0.65 \tag{III.13b}$$

Note that k and k' can be interchanged. The ratio of $K(k)/K'(k) = K(k)/K(k')$ can be evaluated to an accuracy of one part in 10^5 using the following expression:†

$$\frac{K}{K'} = \frac{1}{\pi}\ln\left(2\frac{1+\sqrt{k}}{1-\sqrt{k}}\right) \qquad 0.7 \le k \le 1$$

$$= \left[\frac{1}{\pi}\ln\left(2\frac{1+\sqrt{k'}}{1-\sqrt{k'}}\right)\right]^{-1} \qquad 0 \le k \le 0.7 \tag{III.13c}$$

For intermediate values of k and k', the following formula can be used:

$$K' = K(k') = \frac{2}{1+k}K\left(\frac{1-k}{1+k}\right) \tag{III.13d}$$

where k and k' can be interchanged. When k' is less than 0.707, then $(1-k)/(1+k)$ is always less than 0.172 and (III.13a) can be applied. When k' is greater than 0.707, we can use (III.13b).

Consider now the rectangle shown in the ζ plane in Fig. III.3. Also shown are the values of $\text{sn}(\zeta, k)$ at the points labeled A, B, C, D, E, F. The mapping $W = \text{sn}(\zeta, k)$ maps the rectangle into the real W axis and all interior values into the upper half of the W plane. If the segments AB, CDE, and FG are conducting metal boundaries, then the configuration in the W plane is a coplanar transmission line in air. Its distributed capacitance is twice that between the plates CDE and BAF for the ideal parallel-

†T. S. Gradshteyn and T. M. Ryzhik, "Table of Integrals, Series, and Products," Academic Press, Inc., New York, 1965, p. 925, formula 8.198 and using $q = \exp(-\pi K'/K)$.

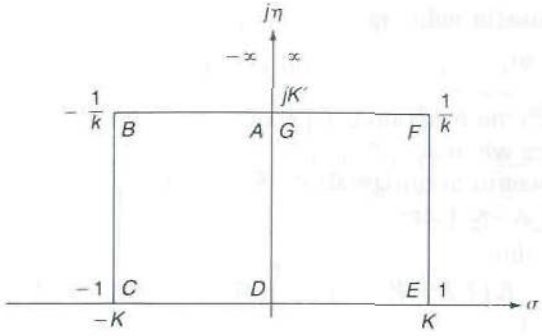


FIGURE III.3
The sn(ζ, k) function.

plate capacitor in the ζ plane and hence is given by

$$C = 2\epsilon_0 \frac{2K}{K'} = 4\epsilon_0 \frac{K}{K'} \tag{III.14}$$

The relevant parameter for the coplanar line is the ratio of the width of the center strip to the spacing between the two ground planes which is given by $2/2u_1 = k$, the modulus of the elliptic sine function. Note that in the ζ plane the boundaries BC and EF are magnetic walls so there is no fringing capacitance.

If we choose the segments BC and EF to be conducting strips, we obtain a coplanar strip transmission line. Its distributed capacitance is that associated with the ideal parallel-plate capacitor with the boundary BC at potential V and EF at potential $-V$; thus

$$C = 2\epsilon_0 \frac{K'}{2K} = \epsilon_0 \frac{K'}{K} \tag{III.15}$$

When we let k tend toward zero, the points $\pm u_1$ move out to infinity and we obtain a slot line. However, we cannot let $k = 0$ because the capacitance between two semiinfinite planes separated by a slot becomes infinite.

III.3 CAPACITANCE BETWEEN TWO PARALLEL STRIPS

We will now consider the problem of two parallel strips in air as shown in Fig. III.5. The plate separation is $2H$ and the plates have a width $2w$. The following mapping will map the boundary shown by the dashed line in Fig. III.5 into the real axis in the W plane shown in Fig. III.4:

$$Z = \int_0^W \frac{(1 - k_0^2 W^2)}{\sqrt{(1 - W^2)(1 - k^2 W^2)}}\, dW \tag{III.16}$$

where $k_0 = 1/u_0$ and $k = 1/u_1$. When $Z = H$ on the interior side of the

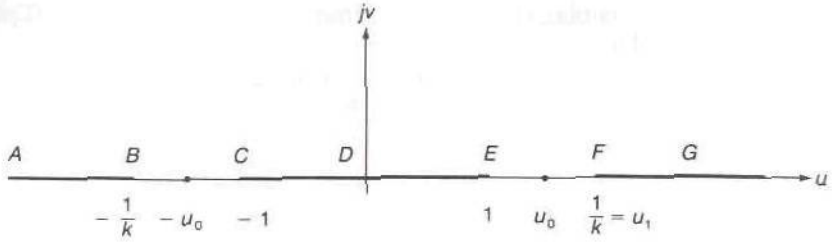


FIGURE III.4
The mapping $W = \mathrm{sn}(\zeta, k)$.

right-hand strip, $W = 1$; so

$$H = \int_0^1 \frac{1 - k_0^2 W^2}{\sqrt{(1 - W^2)(1 - k^2 W^2)}}\, dW \qquad \text{(III.17}a\text{)}$$

At the outer edge of this plate, $Z = H + jw$, and we make this correspond to the unknown point u_0 lying between 1 and u_1; thus

$$H + jw = \int_0^{u_0} \frac{1 - k_0^2 W^2}{\sqrt{(1 - W^2)(1 - k^2 W^2)}}\, dW \qquad \text{(III.17}b\text{)}$$

At the point $Z = H$ on the outer side of the strip, we set $W = u_1$; thus

$$H = \int_0^{u_1} \frac{1 - k_0^2 W^2}{\sqrt{(1 - W^2)(1 - k^2 W^2)}}\, dW \qquad \text{(III.17}c\text{)}$$

The distributed capacitance between the parallel strips is equal to that for

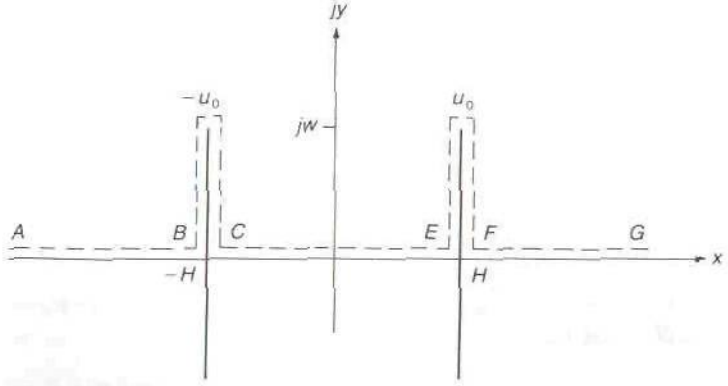


FIGURE III.5
Two parallel strips and the boundary to be mapped.

the coplanar strip line that results from the mapping and is given by (III.15).

For the parallel strips the capacitance depends only on the ratio w/H. When we compare (III.17b) and (III.17c), we see that we require

$$jw = -\int_{u_0}^{u_1} \frac{1 - k_0^2 W^2}{\sqrt{(1 - W^2)(1 - k^2 W^2)}}\, dW$$

A similar comparison of (IV.17a) and (IV.17b) shows that

$$jw = \int_{1}^{u_0} \frac{1 - k_0^2 W^2}{\sqrt{(1 - W^2)(1 - k^2 W^2)}}\, dW$$

By combining these two expressions, we obtain

$$\int_{1}^{u_1} \frac{1 - k_0^2 W^2}{\sqrt{(1 - W^2)(1 - k^2 W^2)}}\, dW = 0 \qquad \text{(III.18)}$$

which is the equation that will determine k_0 and hence u_0. The integral in (III.18) can be separated into two integrals, and by using (III.9c) to replace the integral that goes from 0 to 1, we find that (III.18) can be expressed in the form

$$K' = k_0^2 \int_{1}^{1/k} \frac{W^2}{\sqrt{(W^2 - 1)(1 - k^2 W^2)}}\, dW$$

We can express this integral in terms of complete elliptic integrals. However, in practice it is more expedient to evaluate the integral numerically. By using the substitution $k^2(W^2 - 1) = (1 - k^2)\cos^2\theta$ as before we obtain

$$K' = \frac{k_0^2}{k^2} \int_0^{\pi/2} \sqrt{1 - k'^2 \sin^2\theta}\, d\theta = \frac{k_0^2}{k^2} E\left(k', \frac{\pi}{2}\right) \qquad \text{(III.19)}$$

where E is the complete elliptic integral of the second kind.

From the above we get

$$k_0^2 = \frac{k^2 K'}{E(k', \pi/2)} \qquad \text{(III.20)}$$

The expression for H given by (III.17a) can be reduced to the form (we use $W = \sin\theta$)

$$H = K - k_0^2 \int_0^{\pi/2} \frac{\sin^2\theta}{\sqrt{1 - k^2 \sin^2\theta}}\, d\theta$$

which gives

$$H = K + \frac{k_0^2}{k^2}\left[E\left(k, \frac{\pi}{2}\right) - F\left(k, \frac{\pi}{2}\right)\right]$$

$$= \frac{K'[E(k, \pi/2) - K] + KE(k', \pi/2)}{E(k', \pi/2)}$$

For convenience, the complete elliptic integrals of the second kind will be written in the compact form

$$E\left(k, \frac{\pi}{2}\right) = E(k) = E \qquad E\left(k', \frac{\pi}{2}\right) = E(k') = E'(k) = E'$$

We now use $E(k', \pi/2) = E'(k, \pi/2)$ and the following identity:

$$KE' + K'E - KK' = \frac{\pi}{2}$$

which allows us to express H in the simpler form

$$H = \frac{\pi}{2E(k', \pi/2)} = \frac{\pi}{2E'} \tag{III.21}$$

The last reduction is the expression for w to the form

$$w = \frac{1}{k^2}\int_0^\theta \frac{k_0^2(1 - k'^2 \sin^2\theta) - k^2}{\sqrt{1 - k'^2 \sin^2\theta}}\, d\theta$$

$$= \frac{k_0^2}{k^2}E(k', \theta) - F(k', \theta) \tag{III.22}$$

where $k_0' = \sqrt{1 - k_0^2}$ and

$$\theta = \cos^{-1}\frac{kk_0'}{k'k_0} \tag{III.23}$$

This equation gives w in terms of incomplete elliptic integrals. Values for the complete and incomplete elliptic integrals are available in the book by Jahnke and Emde.†

The easiest way to use the above formulas is to choose a value for k, solve (III.20) for k_0, (III.21) for H, and (III.22) for w. The integrals are readily evaluated numerically. By this means values of w/H and the capacitance C given by (III.15) can be compiled as a function of k. The capacitance between a strip of width $2w$ and spaced a distance H above a ground plane is twice as large as that given by (III.15). Table III.1 gives

†E. Jahnke and F. Emde, "Tables of Functions," Dover Publications, Inc. New York, 1945.

TABLE III.1
Capacitance between a strip of width $2w$ at a height H above a ground plane

$\frac{2w}{H}$	$\frac{C}{\epsilon_0}$	$\frac{2w}{H}$	$\frac{C}{\epsilon_0}$
0.05	1.238	1	2.981
0.1	1.436	1.5	3.621
0.2	1.703	2	4.232
0.3	1.912	2.5	4.822
0.4	2.094	3	5.399
0.5	2.26	4	6.529
0.6	2.415	5	7.63
0.7	2.563	6	8.72
0.8	2.706	8	10.86
0.9	2.845	10	12.98

computed values of C/ϵ_0 as a function of $2w/H$. The integrals were evaluated numerically using Simpson's quadrature formula.

III.4 STRIP TRANSMISSION LINE

The strip transmission line is shown in Fig. III.6. The boundary *A-B-C-D-E* is mapped into the real axis in the W plane by means of the mapping function

$$W = \sin Z = \sin x \cosh y + j \cos x \sinh y \tag{III.24}$$

In the Z plane H is π units, so we make the point B be $j\pi w/H - \pi/2$ so as to maintain the same width-to-spacing ratio w/H as in the physical strip line. The real axis in the W plane is mapped into a rectangle in the ζ plane, shown in Fig. III.6*d*, by the mapping function

$$W = 1 + 2\,\mathrm{sn}(\zeta, k)$$

The point B corresponds to $W = -\cosh \pi w/H$ and to $\mathrm{sn}(\zeta, k) = \mathrm{sn}(-K + jK', k) = -1/k$. Consequently, we must have

$$\frac{2}{k} = 1 + \cosh \frac{\pi w}{H} = 2 \cosh^2 \frac{\pi w}{2H}$$

which gives

$$k = \frac{1}{\cosh \pi w/2H} \tag{III.25}$$

In the ζ plane the capacitance is given by $\epsilon_0 K'/K$ and represents one-quarter of the capacitance of the strip-line configuration. Hence for the strip

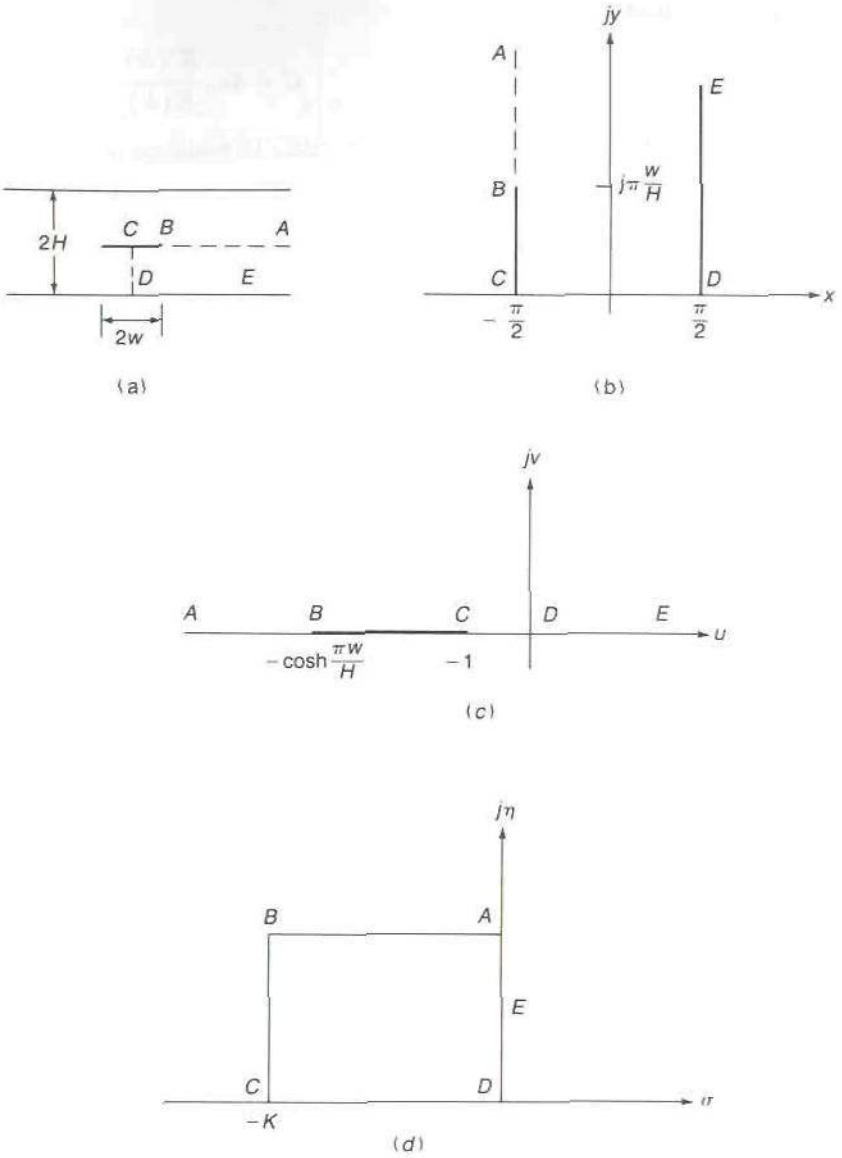


FIGURE III.6
The strip transmission line and its mapping.

line

$$C = 4\epsilon_0 \frac{K'(k)}{K(k)} \tag{III.26}$$

The corresponding characteristic impedance is

$$Z_c = \frac{\sqrt{\mu_0 \epsilon_0}}{C} = 30\pi \frac{K(k)}{K'(k)} \tag{III.27}$$

III.5 CONDUCTOR LOSS

Since the current density and charge density are related by the continuity equation [see (3.138*b*)], the current density has the same functional form as the charge density. If we know the current density J_z, then the power loss is given by

$$P_l = \frac{R_m}{2} \int |J_z|^2 \, dl$$

where R_m is the skin-effect surface resistance and the integral is taken around all the contours along the conductor surfaces. If we have infinitely thin conductors with edges, then the current density becomes infinite at the edge inversely proportional to the square root of the distance from the edge. For example, for an isolated strip of width $2w$ the current density, according to (III.3), would be

$$J_z = \frac{I}{2\pi w \sqrt{1 - x^2/w^2}}$$

on each side of the strip. This current density is too singular to allow a finite value for the power loss to be evaluated. Consequently, it will be necessary to take into account the finite thickness of the conductors.

The simplest case to consider is that of an isolated strip of width $2w$ and thickness $2t$ as shown in Fig. III.7. The labeled boundary in the Z plane can be mapped into the real u axis in the W plane using

$$\frac{dZ}{dW} = A \sqrt{\frac{1 - W^2}{1 - k^2 W^2}} \tag{III.28}$$

which upon integration gives

$$Z = A \int_0^W \sqrt{\frac{1 - W^2}{1 - k^2 W^2}} \, dW + jt \tag{III.29}$$

The constant jt is added since $W = 0$ is made to correspond to $Z = jt$. When $W = 1$ we require $Z = w + jt$ and when $W = 1/k$ we want $Z = w$. These two conditions provide a solution for A and the modulus k. By following a

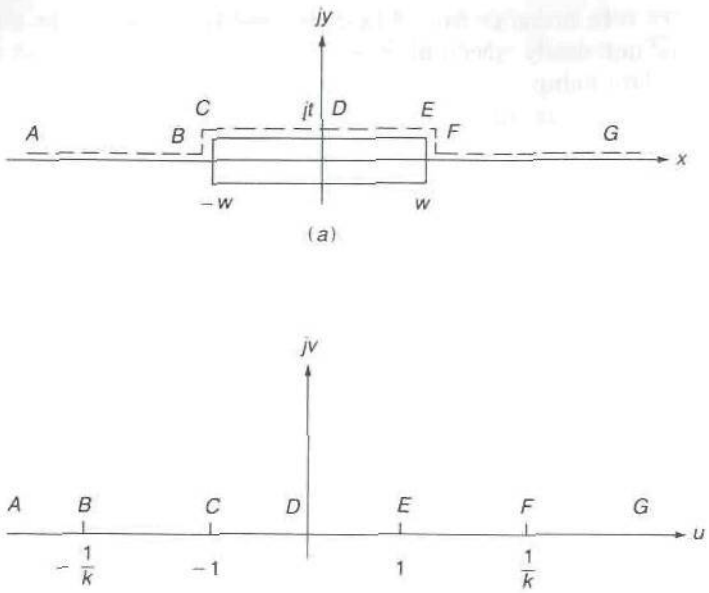


FIGURE III.7
An isolated conducting strip with finite thickness and the mapping of one-half of the strip boundary into the W plane.

procedure similar to that for the parallel strips, we find that

$$A = \frac{wk^2}{E(k) - k'^2K(k)} \tag{III.30a}$$

$$\frac{t}{w} = \frac{E(k') - k^2K(k')}{E(k) - k'^2K(k)} \tag{III.30b}$$

In the W plane the conducting strip extends from $-1/k$ to $1/k$ along u. Since it is an isolated infinitely thin strip, the current density in the W plane is of the form

$$J_u = \frac{A}{\sqrt{1 - k^2u^2}}$$

where the factor A has been inserted arbitrarily. In the Z plane the current density is proportional to

$$\nabla\Phi|_Z = \nabla\Phi|_W \left|\frac{dW}{dZ}\right|$$

or

$$J_z = \frac{A}{\sqrt{1 - k^2u^2}} \left|\frac{\sqrt{1 - k^2u^2}}{A\sqrt{1 - u^2}}\right| = \frac{1}{|\sqrt{1 - u^2}|}$$

In order to find $J_z(x)$, we need to express u as a function of x, which is not easily accomplished. Fortunately, we will not need this functional relationship.

Let us consider the integration of (III.28) very close to the 90° corner, where $Z = w + jt$, of the conducting strip. When $x = w$, $u = 1$; hence

$$\int_x^w dx = A\int_u^1 \sqrt{\frac{1-u^2}{1-k^2u^2}}\, du$$

For u very close to one, we can use the approximations $1 - u^2 = (1 + u)(1 - u) \approx 2(1 - u)$ and $1 - k^2u^2 \approx 1 - k^2$. We then obtain

$$w - x = \sqrt{\frac{2}{1-k^2}}\, A\int_u^1 \sqrt{1-u}\, du = \frac{3}{2}\sqrt{\frac{2}{1-k^2}}\, A(1-u)^{3/2}$$

which gives

$$J_z = \frac{1}{\sqrt{2}\sqrt{1-u}} = \frac{(3A)^{1/3}2^{-2/3}}{(1-k^2)^{1/6}(w-x)^{1/3}} \qquad \text{(III.31)}$$

This fundamental result shows that close to a 90° corner the current density has a weaker singularity, namely, it is inversely proportional to the cube root of the distance from the corner. This current density, when squared, can be integrated to give a finite result for the power loss. The edge behavior is a local field phenomenon, so that the result obtained is true near any 90° corner.

At some distance away from the corner, the expression $[(1 - u^2)/(1 - k^2u^2)]^{1/2}$ can be approximated by unity since k is close to one for a thin strip, i.e., for t/w small. Thus we have

$$\int_0^x dx \approx A\int_0^u du$$

or $x \approx Au$. Hence away from the corner the current density is given by

$$J_z = \frac{1}{\sqrt{1-u^2}} \approx \frac{1}{\sqrt{1-(x/A)^2}} \approx \frac{1}{\sqrt{1-(x/w)^2}}$$

since $A \approx w$ then t/w is small. We can summarize the above result with the statement that for a thin conducting strip the current density over most of the strip is the same as for an infinitely thin strip, but as the corner is approached the one over the square root of the distance from the corner behavior changes over to a less singular behavior going as one over the cube root of the distance from the corner.

For the isolated finite thickness strip, we can evaluate the power loss per meter by doing the integration in the W plane. Thus

$$P_l = 4\frac{R_m}{2}\int_0^{1/k}\frac{1}{|1-u^2|}\left|\frac{dx}{du}\right|du$$

$$= 2R_m A\int_0^{1/k}\frac{du}{\left|\sqrt{(1-u^2)(1-k^2u^2)}\right|}$$

$$= 2R_m A[K(k) + K'(k)] \qquad \text{(III.32)}$$

The integral gives the loss for one quarter section, so we multiplied by a factor of 4 to get the total loss. The total current on the strip is given by

$$I = 4\int_0^{1/k}\frac{1}{\sqrt{1-u^2}}\left|\frac{dx}{du}\right|du = 4A\int_0^{1/k}\frac{du}{\sqrt{1-k^2u^2}}$$

$$= \frac{2\pi A}{k} \qquad \text{(III.33)}$$

The integral was evaluated by using the substitution $ku = \sin\theta$.

The series-distributed resistance R per meter for the conductor is defined by the relationship

$$\tfrac{1}{2}RI^2 = P_l$$

From the derived expressions we obtain

$$R = \frac{k^2R_m}{\pi^2 A}(K + K') \qquad \text{(III.34)}$$

In practice, the ratio t/w is small. For example, a board plated with 1-oz copper has a metalization thickness of 0.036 mm; so a microstrip of width 1 mm will have $t/w = 1/28$. When $t/w < 0.05$ the following approximations are valid since $k \approx 1$:

$$E(k') \approx \frac{\pi}{2}\left(1 - \frac{k'^2}{4}\right) \qquad E(k) \approx 1$$

$$K(k') \approx \frac{\pi}{2}\left(1 + \frac{k'^2}{4}\right) \qquad K(k) \approx \ln\frac{4}{k'}$$

By using these approximations in (III.30), we obtain

$$A \approx k^2 w \qquad \text{and} \qquad k'^2 \approx \frac{4t}{\pi w}$$

We now find that the expression for R reduces to the simple form

$$R = \frac{R_m}{2\pi^2 w}\left(\pi + \ln\frac{4\pi w}{t}\right) \qquad \text{(III.35)}$$

In this expression w is one-half of the strip width and t is one-half of the strip thickness. The part $R_m/2\pi w$ is the contribution from the currents that flow on the two end faces of the strip. From the expression used for J_z, we find that the current density at the center of the broad face where $u = 0$ is unity. At the center of the end face where $u = 1/k$, the current density is $k/k' \approx \sqrt{\pi w/2t}$ and hence is much larger. It is for this reason that there is a significant contribution to the power loss from the current on the end faces even though t is very small.

If the conductor cross section was approximated by an ellipse with major axis equal to $2w$ and minor axis equal to $2t$, it would be found that the series resistance is

$$R = \frac{R_m}{\pi^2 w} \ln \frac{4w}{t}$$

when t/w is small. For very small values of t/w, this resistance is twice as large as that for a rectangular bar of width $2w$ and thickness $2t$. The reason is that for the elliptical cross section the current density at the narrow ends is a factor w/t greater than that at the center of the broad face and thus exhibits a more singular behavior than that for the current at the center of the narrow face for a rectangular bar.

A useful approximate way to find the series resistance of a thin conducting strip is to use the current density for an infinitely thin strip but to integrate the square of the current density up to a point a distance d from the edge. The distance d is chosen so that the same resistance as given by (III.35) is obtained. Thus for the isolated strip we use

$$P_l = 2R_m \int_0^{w-d} \frac{dx}{1 - x^2/w^2} = R_m w \ln \frac{w + x}{w - x}\bigg|_0^{w-d}$$

$$\approx R_m w \ln \frac{2w}{d}$$

The total current on the strip is $2\pi w = I$; so by equating $I^2R/2$ to P_l we obtain

$$R = \frac{R_m}{2\pi^2 w} \ln \frac{2w}{d}$$

We now equate this expression to (III.35) and solve for d, which gives

$$d = \frac{2t}{4\pi e^{\pi}} = \frac{2t}{290.8}$$

In the derivation we expressed π as $\ln e^{\pi}$. This method of finding the loss resistance was proposed by Lewin.† This method is based on the fact that

†L. Lewin, A Method of Avoiding the Edge Current Divergence in Perturbation Loss Calculations, *IEEE Trans.*, vol. MTT-32, pp. 717–719, July, 1984.

the effect of finite thickness is to change the local behavior of the current near the edge but leaving the current density on the major portion of the conducting strip relatively unchanged.

III.6 CONDUCTOR LOSSES FOR A MICROSTRIP TRANSMISSION LINE

Consider a microstrip line with a strip of width $2w$, *thickness* $2t$, and spaced a distance H above a ground plane. In order to apply the rule established above, we need to know the division of the total current on the upper and lower faces of the strip, which is affected by the presence of the ground plane. Apart from an unequal division of the total current, the density of current is very nearly the same as on an isolated strip. From the mapping function (III.16), we find that the current density is proportional to

$$J_z = \nabla\Phi|_W \left|\frac{dW}{dZ}\right| = \nabla\Phi|_\zeta \left|\frac{d\zeta}{dW}\right| = \frac{1}{|1 - k_0^2 W^2|}$$

At the center of the upper face $u = u_1 = 1/k$ and at the center of the lower face $u = 1$. Hence the ratio of the currents on the two faces is

$$\frac{J_{z1}}{J_{z2}} = \frac{1 - k_0^2}{|1 - k_0^2/k^2|} = \frac{k^2(1 - k_0^2)}{k_0^2 - k^2}$$

The fraction of the total current on the upper face is p, where p is given by

$$p = \frac{J_{z1}}{J_{z1} + J_{z2}} = \frac{k^2(1 - k_0^2)}{k_0^2(1 - k^2)} = \frac{E' - k^2 K'}{(1 - k^2)K'} \qquad \text{(III.36}$$

where (III.20) has been used to eliminate k_0^2. The power loss on the upper face is proportional to $(1 - p)^2$. The total power loss is proportional to $p^2 + (1 - p)^2$. For an isolated strip $p = 1/2$, so that the series resistance R_1 of the microstrip in the presence of the ground plane is given by multiplying R by the ratio $[p^2 + (1 - p)^2]/0.5$; thus

$$R_1 = 2(2p^2 - 2p + 1)R \qquad \text{(III.37}$$

where p depends only on the ratio $2w/H$ and R is the same as in (III.35). When $p = 0.25$ we find that $R_1 = 1.25R$ which shows that an unequal division of the total current increases the resistance.

The ground plane corresponds to the y axis in Fig. III.5, which maps into the v axis in the W plane. Thus the current density will be proportional to the gradient along the v axis in the W plane which is given by

$$J_z \propto \frac{1}{1 - k_0^2(jv)^2} = \frac{1}{1 + k_0^2 v^2}$$

Let J_z be chosen as

$$J_z = \frac{I_0}{1 + k_0^2 v^2}$$

where I_0 is to be found so that the total current will equal I. The total current on the ground plane is given by

$$I = 2I_0 \int_0^\infty \frac{1}{1 + k_0^2 v^2} \left| \frac{dZ}{dW} \right| dv$$

$$= 2I_0 \int_0^\infty \frac{dv}{\sqrt{(1 + k^2 v^2)(1 + v^2)}}$$

The integral is easily done in the ζ plane by using

$$dv = \left| \frac{dW}{d\zeta} \right| d\eta = \sqrt{(1 + k^2 v^2)(1 + v^2)}\, d\eta$$

and the limits 0 to $K'(k)$ for η. Thus we find that $I = 2I_0 K'(k)$.

The power loss in the ground plane is given by

$$P_l = R_m \frac{I^2}{4K'^2} \int_0^\infty \frac{1}{(1 + k_0^2 v^2)^2} \left| \frac{dZ}{dW} \right| dv$$

$$= \frac{R_m I^2}{4K'^2} \int_0^\infty \frac{dv}{(1 + k_0^2 v^2)\sqrt{(1 + k^2 v^2)(1 + v^2)}}$$

The series resistance R_2 of the ground plane is obtained by equating the power loss to $I^2 R_2/2$. The series resistance will depend on the absolute dimensions of the microstrip line. Thus we will multiply P_l by $(2w/H)$ and divide by this same factor but use (III.21) for H. We thereby obtain

$$R_2 = \frac{R_m}{2w} \left(\frac{\pi}{4} \frac{2w}{H} \frac{1}{E'K'^2} \right) \int_0^\infty \frac{dv}{(1 + k_0^2 v^2)\sqrt{(1 + k^2 v^2)(1 + v^2)}} \qquad \text{(III.38)}$$

The normalized resistance $2wR_2/R_m$ is a function only of the ratio $2w/H$ and thus only needs to be evaluated once. The relevant expressions for the loss ratio

$$\text{LR} = 2(2p^2 - 2p + 1) \qquad \text{(III.39)}$$

and the normalized resistance $2wR_2/R_m$ have been evaluated numerically as a function of $2w/H$. We have fitted simple polynomial functions to the resultant data, so that for application purposes the loss ratio and normalized resistance can be computed from the following expressions to an

accuracy of 2 percent or better:

$$\text{LR} = 1 \qquad \frac{2w}{H} < 0.5 \tag{III.40a}$$

$$\text{LR} = 0.94 + 0.134\frac{2w}{H} - 0.0062\left(\frac{2w}{H}\right)^2 \qquad 0.5 \le \frac{2w}{H} \le 10 \tag{III.40b}$$

$$\frac{2w}{R_m}R_2 = \frac{2w/H}{2w/H + 5.8 + 0.03H/2w} \qquad 0.1 \le \frac{2w}{H} \le 10 \tag{III.41}$$

This formula states that the resistance of the ground plane is that of a strip of width $2w + 5.8H$ and with a uniform current density. It is remarkable that this relationship holds for such a wide range of $2w/H$ ratios. In practice, surface roughness increases the resistance by 10 to 50 percent so an accuracy greater than a few percent in the theoretical formulas is not needed.

III.7 ATTENUATION FOR A COPLANAR LINE

We consider a coplanar line as shown in Fig. III.8*a*. The ground planes and center conductor have a thickness $2t$. The strip width is $2a$ and the ground plane separation is $2b$. The upper half of the Z plane will be mapped into the upper half of the W plane with the contour labeled 1-2---8 mapping into

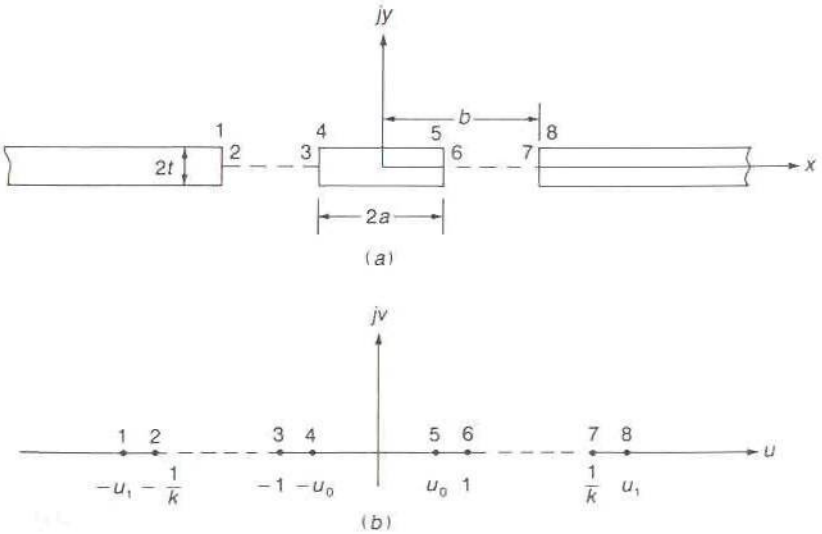


FIGURE III.8
Conformal mapping of the cross section of a coplanar transmission line with finite thickness conductors.

the real axis in the W plane as shown in Fig. III.8*b*. The required mapping function is

$$Z = A\int_0^W \sqrt{\frac{(1-k_0^2W^2)(1-k_1^2W^2)}{(1-W^2)(1-k^2W^2)}}\,dW + jt \tag{III.42}$$

where $k_0 = 1/u_0$ and $k_1 = 1/u_1$. The parameters A, k_0, k, and k_1 are determined by requiring that the four points $Z = a + jt$, a, b, and $b + jt$ map into the points $W = u_0$, 1, $1/k$, and u_1. These requirements lead to the following four equations:

$$a = A\int_0^{u_0} F(W)\,dW \tag{III.43a}$$

$$t = jA\int_{u_0}^{1} F(W)\,dW \tag{III.43b}$$

$$b = A\int_0^{u_1} F(W)\,dW \tag{III.43c}$$

$$t = -jA\int_{1/k}^{u_1} F(W)\,dW \tag{III.43d}$$

where $F(W)$ is the integrand shown in (III.42). The real axis in the W plane can be mapped into the rectangle in the ζ plane using (III.8). The boundary-value problem in the ζ plane is that of an ideal parallel-plate capacitor. Hence $\nabla\Phi|_\zeta$ is a constant which we set equal to one.

The current density J_z is thus given by

$$J_z = \nabla\Phi|_\zeta \left|\frac{d\zeta}{dZ}\right| = \left|\frac{d\zeta}{dZ}\right| = \left|\frac{d\zeta}{dW}\right|\left|\frac{dW}{dZ}\right|$$

Consequently,

$$J_z\,dZ = \left|\frac{d\zeta}{dW}\right|\left|\frac{dW}{dZ}\right|dZ = \left|\frac{d\zeta}{dW}\right|dW$$

The total current on the center conductor is

$$I = 4\int_0^1 \left|\frac{d\zeta}{dW}\right|dW = 4\int_0^1 \frac{dW}{\sqrt{(1-W^2)(1-k^2W^2)}} = 4K(k)$$

The power loss on the center conductor is given by

$$P_{l1} = 4\frac{R_m}{2}\int_0^a |J_z|^2\,dZ = 2R_m\int_0^1 \left|\frac{d\zeta}{dW}\right|^2\left|\frac{dW}{dZ}\right|^2\left|\frac{dZ}{dW}\right|dW$$

$$= \frac{2}{a}R_m\int_0^1 \frac{du}{\left|\sqrt{(1-u^2)(1-k_0^2u^2)(1-k^2u^2)(1-k_1^2u^2)}\right|} \tag{III.44a}$$

while that on the two ground planes is given by

$$P_{l2} = \frac{2}{a} R_m \int_{1/k}^{u_1} \frac{du}{\left|\sqrt{(1-u^2)(1-k_0^2u^2)(1-k^2u^2)(1-k_1^2u^2)}\right|} \quad \text{(III.44}$$

The integrals in (III.43) and (III.44) can be evaluated numerically but iterative technique is necessary in order to find the required values of k_0, and k_1. We will only consider the case when the thickness t is very sm say $t < 0.05a$. For this important special case, a number of approximatic can be made that will lead to relatively simple expressions for the pow loss.

When t/a is small k_0 is approximately equal to one and k_1 approximately equal to k. Thus the intervals u_0 to 1 and $1/k$ to u_1 small. We also note that if $t = 0$, then $k_0 = 1$, $k_1 = k$ and (III.42) giv $Z = AW$ from which we obtain $a = A$ and $b = A/k$ or $k = a/b$. We will u these values for A and k for the case of small but finite values of t. (III.43b) we can set u equal to 1 in all but the critical factors $1 - u$ a $1 - k_0u$, since the interval of integration is very small. Thus we find th

$$t = a\int_{u_0}^{1} \sqrt{\frac{(1-k_0)(1-k_1^2)(k_0u-1)}{2(1-k^2)(1-u)}}\, du$$

$$\approx a\int_{u_0}^{1} \sqrt{\frac{k_0u-1}{1-u}}\, du$$

This integral can be evaluated in terms of elementary functions and gives

$$t = \frac{\pi}{2} a \frac{k_0 - 1}{\sqrt{k_0}} \approx \frac{\pi a}{2}(k_0 - 1)$$

Hence we obtain

$$k_0 = 1 + \frac{2t}{\pi a} \quad \text{(III.4}$$

which verifies that k_0 is close to one.

In (III.43d) we put $u = 1/k$ in all factors except the critical on $1 - ku$ and $1 - k_1u$, since the interval of integration in this equation is al small. An integral of the same type as found above is obtained and readi leads to the result

$$k_1 = \frac{a}{b}\left(1 - \frac{2t}{\pi b}\right) \quad \text{(III.4}$$

which shows that k_1 is close to $k = a/b$.

The power-loss terms involve the following four integrals:

$$I_1 = \int_0^{u_0} G(u)\,du \qquad I_2 = \int_{u_0}^{1} G(u)\,du$$

$$I_3 = \int_{u_1}^{\infty} G(u)\,du \qquad I_4 = \int_{1/k}^{u_1} G(u)\,du$$

where $G(u) = |[(1-u^2)(1-k_0^2u^2)(1-k^2u^2)(1-k_1^2u^2)]^{-1/2}|$. The integrals I_2 and I_4 will give us the power loss on the conductor edges and will be considered first. For the evaluation of I_2 we use the approximation $[(1-k^2u^2)(1-k_1^2u^2)]^{-1/2} \approx (1-k^2)^{-1}$ and make the substitution $k_0u = \lambda$, $k_a = 1/k_0$ to obtain

$$I_2 \approx \frac{1}{k_0(1-k^2)} \int_1^{1/k_a} \frac{d\lambda}{\sqrt{(\lambda^2-1)(1-k_a^2\lambda^2)}} = \frac{K'(k_a)}{k_0(1-k^2)}$$

Since k_a is close to one, we can use $K'(k_a) \approx \pi/2$ and $k_0 = 1$ to get

$$I_2 = \frac{\pi}{2(1-a^2/b^2)} \tag{III.47a}$$

The same approach is used to evaluate I_4 to obtain

$$I_4 \approx \frac{k}{1-k^2} \int_1^{k_b} \frac{d\lambda}{\sqrt{(\lambda^2-1)(1-k_b^2\lambda^2)}} = \frac{k}{1-k^2} K'(k_b)$$

$$\approx \frac{\pi a}{2b(1-a^2/b^2)} \tag{III.47b}$$

since $k_b = k_1/k$ is close to one.

In order to evaluate I_1 we first make the substitution $k_0u = \cos\theta$ to obtain

$$I_1 = \int_0^{\pi/2} \frac{k_0^2\,d\theta}{\sqrt{(k_0^2-\cos^2\theta)(k_0^2-k^2\cos^2\theta)(k_0^2-k_1^2\cos^2\theta)}}$$

We now split the integral into one over a small interval 0 to θ_1 plus an integral from θ_1 to $\pi/2$. Over the first interval the expression under the square-root sign is approximated by

$$(k_0+1)(k_0^2-k^2)(k_0^2-k_1^2)(k_0-\cos\theta)$$

$$\approx (1-k^2)^2[(k_0-1)2+\theta^2]$$

The resultant integral is an elementary one and easily integrated to give

$$I_{1a} = \int_0^{\theta_1} \frac{k_0^2}{1-k^2} \frac{d\theta}{\sqrt{2(k_0-1)+\theta^2}} = \frac{k_0^2}{1-k^2} \ln\left[\theta + \sqrt{2(k_0-1)+\theta^2}\right]\Bigg|_0^{\theta_1}$$

$$= \frac{1}{1-k^2} \ln \frac{\theta_1 + \sqrt{\theta_1^2 + 4t/\pi a}}{\sqrt{4t/\pi a}} \approx \frac{1}{1-k^2} \ln \frac{2\theta_1}{\sqrt{4t/\pi a}}$$

Even if $4t/\pi a$ is of the same order as θ_1^2, the error in the last approximation is small. For $4t/\pi a = \theta_1^2$ it is $\ln(1.207) \approx 0.207$, which is small relative to other terms that occur for P_{l1}.

In the integral from θ_1 to $\pi/2$, we can assume that $t = 0$ so $k_0 = 1$, $k_1 = k$, and we then obtain

$$I_{1b} = \int_{\theta_1}^{\pi/2} \frac{d\theta}{\sin\theta(1 - k^2\cos^2\theta)}$$

By using $\lambda = \cos\theta$ this integral becomes

$$I_{1b} = \int_0^{\cos\theta_1} \frac{d\lambda}{(1-\lambda^2)(1-k^2\lambda^2)}$$

and is readily evaluated. We obtain

$$I_{1b} = \frac{1}{2(1-k^2)}\left[\ln\frac{4-\theta_1^2}{\theta_1^2} - k\ln\frac{2(1+k)-k\theta_1^2}{2(1-k)+k\theta_1^2}\right]$$

upon using $\cos\theta_1 \approx 1 - \theta_1^2/2$. If we choose $\theta_1 = 0.25$, then a number of terms involving θ_1^2 can be dropped. When we combine I_{1a} and I_{1b}, we get

$$I_1 = \frac{1}{2(1-k^2)}\left[\ln\frac{4\pi a}{t} - k\ln\frac{1+k}{1-k}\right] \qquad \text{(III.48}a\text{)}$$

which has the nice feature that it does not depend on θ_1.

For the final integral I_3, we make the substitution $k_1 u = \cosh\theta$ and again split the integral into one over the interval 0 to θ_1 plus one over the interval θ_1 to ∞. In the first interval $\cosh\theta$ is replaced by unity in the noncritical terms and by $1 + \theta^2/2$ in the $1 - k^2u^2 = 1 - (k^2/k_1^2)\cosh^2\theta$ term. In the second integral we use $k_0 = 1$, $k_1 = k$. We then find that

$$I_3 = \frac{k}{2(1-k^2)}\left[\ln\frac{4\pi b}{t} - \frac{1}{k}\ln\frac{1+k}{1-k}\right] \qquad \text{(III.48}b\text{)}$$

The series resistance of the center conductor is R_1 and is defined by the relationship

$$\tfrac{1}{2}I^2R_1 = P_{l1} = \frac{2}{a}R_m(I_1 + I_2)$$

The series resistance R_2 of the ground planes is defined by a similar

relationship, namely,

$$\frac{1}{2}I^2R_2 = P_{l2} = \frac{2}{a}R_m(I_3 + I_4)$$

The total current I equals $4K(k)$, so we find that

$$R_1 = \frac{R_m}{8aK^2(k)(1-k^2)}\left[\pi + \ln\frac{4\pi a}{t} - k\ln\frac{1+k}{1-k}\right] \quad \text{(III.49}a\text{)}$$

$$R_2 = \frac{R_m k}{8aK^2(k)(1-k^2)}\left[\pi + \ln\frac{4\pi b}{t} - \frac{1}{k}\ln\frac{1+k}{1-k}\right] \quad \text{(III.49}b\text{)}$$

where $k = a/b$. The attenuation is given by

$$\alpha_c = \frac{R_1 + R_2}{2Z_c} \quad \text{(III.50)}$$

where Z_c is the characteristic impedance of the coplanar line.

When t is very small and the ground planes are widely separated, the expression for R_1 reduces to that for an isolated strip, as one would expect. Apart from the factor $1/(1-k^2)$ the ground-plane losses are approximately what one would find for an isolated conductor of width $2b$. The formulas derived above are estimated to be accurate to within 10 percent for $t < 0.05a$ and $k < 0.8$ More accurate approximations can be made but the resultant formulas would be more complex. The expressions obtained above can also be derived with fewer steps by using Lewin's method. The alternative derivation was chosen so as to provide an example of a more complete derivation and to more fully show the approximations involved.

APPENDIX IV

PHYSICAL CONSTANTS AND OTHER DATA

IV.1 PHYSICAL CONSTANTS

Permittivity of vacuum = $\epsilon_0 = 8.854 \times 10^{-12} \approx (1/36\pi) \times 10^{-9}$ F/m
Permeability of vacuum = $\mu_0 = 4\pi \times 10^{-7}$ H/m
Impedance of free space = $Z_0 = 376.7 \approx 120\pi\ \Omega$
Velocity of light = $c = 2.998 \times 10^8$ m/s
Charge of electron = $e = 1.602 \times 10^{-19}$ C
Mass of electron = $m = 9.107 \times 10^{-31}$ kg
$\eta = e/m = 1.76 \times 10^{11}$ C/kg
Mass of proton = $M = 1.67 \times 10^{-27}$ kg
Boltzmann's constant = $k = 1.380 \times 10^{-23}$ J/K
Planck's constant = $h = 6.547 \times 10^{-34}$ J · s
10^7 ergs = 1 J
1 J = 0.6285×10^{19} eV
1 eV = energy gained by an electron in accelerating through a potential of 1 V
Energy of 1 eV = equivalent electron temperature of 1.15×10^4 K
Electron plasma frequency $f_p = \dfrac{e}{2\pi}\left(\dfrac{N}{m\epsilon_0}\right)^{1/2} = 8.97N^{1/2}$ Hz, where N is the number of electrons per cubic meter
Electron cyclotron frequency $f_c = eB/2\pi m = 28{,}000B$ MHz for B in webers per square meter; $f_c = 2.8B$ MHz for B in gauss
10^4 G = 1 Wb/m^2

IV.2 CONDUCTIVITIES OF MATERIALS

Material	Conductivity, S/m	Material	Conductivity, S/m
Copper (annealed)	5.8×10^7	Steel	$0.5–1.0 \times 10^7$
Aluminum	3.54×10^7	Water (distilled)	2×10^{-4}
Silver	6.14×10^7	Sea water	3–5
Nickel	1.28×10^7	Quartz (fused)	$< 2 \times 10^{-17}$

IV.3 DIELECTRIC CONSTANTS OF MATERIALS

Material	*Frequency, MHz*	ϵ'/ϵ_0	*Loss tangent* ϵ''/ϵ'
Polystyrene	3,000	2.54	0.00025–0.0016
Polystyrene (foam)	3,000	1.05	0.00003
Lucite	10,000	2.56	0.005
Teflon	10,000	2.08	0.00037
Fused quartz	10,000	3.78	0.0001
Ruby mica	3,000	5.4	0.0003
Titanium dioxide	10,000	90	0.002
Mahogany wood	10,000	1.7	0.021

IV.4 SKIN DEPTH IN COPPER

Frequency, Hz	10	60	10^2	10^3	10^4	10^8
Skin depth δ_s, cm	2.08	0.85	0.66	0.208	6.6×10^{-2}	6.6×10^{-4}

$\delta_s = \sqrt{2/\omega\mu\sigma} = 6.6 f^{-1/2}$ cm for copper ($\sigma = 5.8 \times 10^7$ S/m).

AUTHOR INDEX

SUBJECT INDEX