Microwave

Devices

and Circuits

Third Edition

SAMUEL Y. LIAO

Professor of Electrical Engineering
California State University, Fresno

PRENTICE HALL, *Englewood Cliffs, New Jersey 07632*

Contents

Preface

This third revision has been designed, as have the first two editions, for use in a first course in microwave devices and circuits at the senior or beginning graduate level in electrical engineering. The objectives of this book are to present the basic principles, characteristics, and applications of commonly used microwave devices and to explain the techniques for designing microwave circuits. It is assumed that readers of this text have had previous courses in electromagnetics and solid-state electronics. Because this book is self-contained to a large extent, it also serves as a convenient reference for electronics engineers working in the microwave field.

The format of this edition remains the same, but there are additions and expansions as well as some corrections and deletions. The problems section has been enlarged and includes new and very practical problems. The book is reorganized into twelve chapters.

Chapter 1 discusses the interactions between electrons and fields.

Chapter 2 deals with plane-wave propagation in different media.

Chapter 3 treats transmission lines.

Chapter 4 analyzes microwave waveguides and components.

Chapter 5 describes microwave transistors and tunnel diodes, and includes heterojunction bipolar transistors (HBTs).

Chapter 6 treats microwave field-effect transistors such as JFETs, MESFETs, HEMTs, MOSFETs and the NMOS, CMOS, and the charged-coupled devices (CCDs).

Chapter 7 discusses transferred electron devices (TEDs), including the Gunn, LSA, InP, and CdTe diodes.

Chapter 8 describes avalanche transit-time devices such as the IMPATT, TRAPATT, and BARITT diodes and the parametric devices.

Chapter 9 deals with microwave linear-beam tubes including klystrons, reflex klystron, and traveling-wave tubes (TWTs).

Chapter 10 studies microwave crossed-field tubes such as magnetrons, forward-wave crossed-field amplifiers, and the backward-wave crossed-field amplifiers—Amplitron and Carcinotron.

Chapter 11 explains strip lines including microstrip, parallel, coplanar, and shielded strip lines.

Chapter 12 analyzes monolithic microwave integrated circuits (MMICs) including MMIC growth, MOSFET fabrication, thin-film formation, and hybrid integrated-circuit fabrication.

The arrangement of topics is flexible; an instructor may conveniently select or order the several topics to suit either a one-semester or a one-quarter course. Numerous problems for each chapter establish the reader's further understanding of the subjects discussed. Instructors who have adopted the book for their courses may obtain a solutions manual from the publisher.

The author is grateful to the several anonymous reviewers; their many valuable comments and constructive suggestions helped to improve this edition. The author would also like to acknowledge his appreciation to the many instructors and students who used the first two editions and who have offered comments and suggestions. All of this help was vital in improving this revision, and this continuing group effort is sincerely invited. Finally, I wish to express my deep appreciation to my wife, Lucia Hsiao Chuang Lee, and our children: Grace in bioengineering, Kathy in electrical engineering, Gary in electronics engineering, and Jeannie in teachers education, for their valuable collective contributions. Therefore, this revision is dedicated to them.

Samuel Y. Liao

Chapter 0

Introduction

The central theme of this book concerns the basic principles and applications of microwave devices and circuits. Microwave techniques have been increasingly adopted in such diverse applications as radio astronomy, long-distance communications, space navigation, radar systems, medical equipment, and missile electronic systems. As a result of the accelerating rate of growth of microwave technology in research and industry, students who are preparing themselves for, and electronics engineers who are working in, the microwave area are faced with the need to understand the theoretical and experimental design and analysis of microwave devices and circuits.

0-1 MICROWAVE FREQUENCIES

The term *microwave frequencies* is generally used for those wavelengths measured in centimeters, roughly from 30 cm to 1 mm (1 to 300 GHz). However, *microwave* really indicates the wavelengths in the micron ranges. This means microwave frequencies are up to infrared and visible-light regions. In this revision, microwave frequencies refer to those from 1 GHz up to 10^6 GHz. The microwave band designation that derived from World War II radar security considerations has never been officially sanctioned by any industrial, professional, or government organization. In August 1969 the United States Department of Defense, Office of Joint Chiefs of Staff, by message to all services, directed the use of a new frequency band breakdown as shown in Table 0-1. On May 24, 1970, the Department of Defense adopted another band designation for microwave frequencies as listed in Table 0-2. The Institute of Electrical and Electronics Engineers (IEEE) recommended new microwave band designations as shown in Table 0-3 for comparison.

1

TABLE 0-1   U.S. MILITARY MICROWAVE BANDS

Designation	Frequency range in gigahertz
P band	0.225– 0.390
L band	0.390– 1.550
S band	1.550– 3.900
C band	3.900– 6.200
X band	6.200– 10.900
K band	10.900– 36.000
Q band	36.000– 46.000
V band	46.000– 56.000
W band	56.000–100.000

TABLE 0-2   U.S. NEW MILITARY MICROWAVE BANDS

Designation	Frequency range in gigahertz	Designation	Frequency range in gigahertz
A band	0.100–0.250	H band	6.000– 8.000
B band	0.250–0.500	I band	8.000– 10.000
C band	0.500–1.000	J band	10.000– 20.000
D band	1.000–2.000	K band	20.000– 40.000
E band	2.000–3.000	L band	40.000– 60.000
F band	3.000–4.000	M band	60.000–100.000
G band	4.000–6.000		

TABLE 0-3   IEEE MICROWAVE FREQUENCY BANDS

Designation	Frequency range in gigahertz
HF	0.003– 0.030
VHF	0.030– 0.300
UHF	0.300– 1.000
L band	1.000– 2.000
S band	2.000– 4.000
C band	4.000– 8.000
X band	8.000– 12.000
Ku band	12.000– 18.000
K band	18.000– 27.000
Ka band	27.000– 40.000
Millimeter	40.000–300.000
Submillimeter	>300.000

0-2 MICROWAVE DEVICES

In the late 1930s it became evident that as the wavelength approached the physical dimensions of the vacuum tubes, the electron transit angle, interelectrode capacitance, and lead inductance appeared to limit the operation of vacuum tubes in microwave frequencies. In 1935 A. A. Heil and O. Heil suggested that microwave voltages be generated by using transit-time effects together with lumped tuned cir-

cuits. In 1939 W. C. Hahn and G. F. Metcalf proposed a theory of velocity modulation for microwave tubes. Four months later R. H. Varian and S. F. Varian described a two-cavity klystron amplifier and oscillator by using velocity modulation. In 1944 R. Kompfner invented the helix-type traveling-wave tube (TWT). Ever since then the concept of microwave tubes has deviated from that of conventional vacuum tubes as a result of the application of new principles in the amplification and generation of microwave energy.

Historically microwave generation and amplification were accomplished by means of velocity-modulation theory. In the past two decades, however, microwave solid-state devices—such as tunnel diodes, Gunn diodes, transferred electron devices (TEDs), and avalanche transit-time devices have been developed to perform these functions. The conception and subsequent development of TEDs and avalanche transit-time devices were among the outstanding technical achievements. B. K. Ridley and T. B. Watkins in 1961 and C. Hilsum in 1962 independently predicted that the transferred electron effect would occur in GaAs (gallium arsenide). In 1963 J. B. Gunn reported his "Gunn effect." The common characteristic of all microwave solid-state devices is the negative resistance that can be used for microwave oscillation and amplification. The progress of TEDs and avalanche transit-time devices has been so swift that today they are firmly established as one of the most important classes of microwave solid-state devices.

0-3 MICROWAVE SYSTEMS

A microwave system normally consists of a transmitter subsystem, including a microwave oscillator, waveguides, and a transmitting antenna, and a receiver subsystem that includes a receiving antenna, transmission line or waveguide, a microwave amplifier, and a receiver. Figure 0-1 shows a typical microwave system.

In order to design a microwave system and conduct a proper test of it, an adequate knowledge of the components involved is essential. Besides microwave devices, the text therefore describes microwave components, such as resonators, cavities, microstrip lines, hybrids, and microwave integrated circuits.
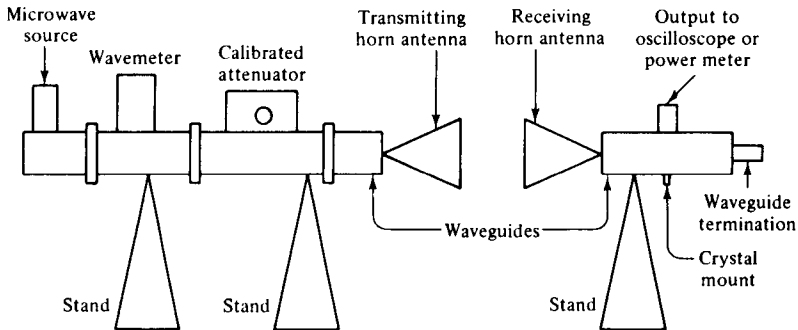
Figure 0-1 Microwave system.

0-4 MICROWAVE UNITS OF MEASURE

Microwave measures can be expressed in different units, such as the CGS (centimeter-gram-second) unit, MKS (meter-kilogram-second) unit, or another unit. The meter-kilogram-second units (the International System of Units) are used throughout unless otherwise indicated. Table 0-4 lists the most commonly used MKS units.

The prefixes tabulated in Table 0-5 are those recommended by the Interna-

TABLE 0-4   MKS UNITS

Quantity	Unit	Symbol
Capacitance	farad = coulomb per volt	F
Charge	coulomb: $A - s$	Q
Conductance	mhos	$\mho$
Current	ampere = coulomb per second	A
Energy	joule	J
Field	volt per meter	E
Flux linkage	weber = volt · second	ψ
Frequency	cycle per second	Hz
Inductance	henry = $(V - s)/A$	H
Length	meter	m
Power	watt = joule per second	W
Resistance	ohm	Ω
Time	second	s
Velocity	meter per second	v
Voltage	volt	V

Note: 1 tesla = 1 weber/m^2 = 10^4 gausses = 3×10^{-6} ESU
        1 Å (angstrom) = 10^{-10} m
        1 μm (micron) = 10^{-6} m

TABLE 0-5   PREFIXES

Prefix	Factor	Symbol
exa	10^{18}	E
peta	10^{15}	P
tera	10^{12}	T
giga	10^{9}	G
mega	10^{6}	M
kilo	10^{3}	k
hecto	10^{2}	h
deka	10	da
deci	10^{-1}	d
centi	10^{-2}	c
milli	10^{-3}	m
micro	10^{-6}	μ
nano	10^{-9}	n
pico	10^{-12}	p
femto	10^{-15}	f
atto	10^{-18}	a

tional Committee on Weights and Measures. They have been adopted by the National Bureau of Standards and are used by the Institute of Electrical and Electronics Engineers, Inc.

The physical constants commonly used in the text are listed in Table 0-6.

The temperature scales commonly used in scientific work, engineering design, and everyday life are shown in Table 0-7. Many engineering computations use the absolute temperature in degrees Kelvin, and therefore a temperature conversion is necessary to convert the temperatures from either centigrade or Fahrenheit to Kelvin scale.

TABLE 0-6 PHYSICAL CONSTANTS

Constant	Symbol	Value
Boltzmann constant	k	1.381×10^{-23} J/°K
Electronvolt	eV	1.602×10^{-19} J
Electron charge	q	1.602×10^{-19} C
Electron mass	m	9.109×10^{-31} kg
Ratio of charge to mass of an electron	e/m	1.759×10^{11} C/kg
Permeability of free space	μ_0	1.257×10^{-6} H/m or $4\pi \times 10^{-7}$ H/m
Permittivity of free space	ϵ_0	8.854×10^{-12} F/m
Planck's constant	h	6.626×10^{-34} J·s
Velocity of light in vacuum	c	2.998×10^8 m/s

TABLE 0-7 TEMPERATURE SCALES

	Rankine °R	Fahrenheit °F	Centigrade °C	Kelvin °K
Boiling point	671.4°	212°	100°	373°
Ambient point	540°	80.6°	27°	300°
Ice point	491.4°	32°	0°	273°
		0°		
Absolute zero	0°	−459.4°	−273°	0°

Chapter 1

Interactions Between
Electrons and Fields

In this chapter we are concerned with electron-field interactions. The motion of the electron beam is assumed to be in a uniform electric field, or a uniform magnetic field, or a uniform electromagnetic field because the inhomogeneous differential equations governing the motion of an electron beam in a field involve three dimensions and their solutions in a nonuniform field are, in most cases, extremely difficult to obtain and usually cannot be determined exactly. On the other hand, fortunately, all present microwave devices employ a uniform field for the electron-field interactions.

Our primary purpose here is to provide the reader with a background for understanding the electron-field interactions in microwave devices that will be discussed in later chapters.

1-1 ELECTRON MOTION IN AN ELECTRIC FIELD

In describing fields and electron-field interactions, certain experimental laws of electricity and magnetism are covered first. The fundamental force law of charges is Coulomb's law, which states that between two charges there exists either an attractive or a repulsive force, depending on whether the charges are of opposite or like sign. That is,

$$\mathbf{F} = \frac{Q_1 Q_2}{4\pi\epsilon_0 R^2}\mathbf{u}_{R_{12}} \qquad \text{newtons} \qquad (1\text{-}1\text{-}1)$$

6

where Q = charge in coulombs

$\epsilon_0 = 8.854 \times 10^{-12} \approx \dfrac{1}{36\pi} \times 10^{-9}$ F/m is the permittivity of free space

R = range between the charges in meters

$\mathbf{u}$ = unit vector

It should be noted that since the MKS system is used throughout this text, a factor of 4π appears in the preceding equation.

The electric field intensity produced by the charges is defined as the force per unit charge—that is,

$$\mathbf{E} \equiv \frac{\mathbf{F}}{Q} = \frac{Q}{4\pi\epsilon_0 R^2}\mathbf{u}_R \qquad \text{V/m} \qquad (1\text{-}1\text{-}2)$$

If there are n charges, the electric field becomes

$$\mathbf{E} = \sum_{m=1}^{n} \frac{Q_m}{4\pi\epsilon_0 R_m^2}\mathbf{u}_{R_m} \qquad (1\text{-}1\text{-}3)$$

In order to determine the path of an electron in an electric field, the force must be related to the mass and acceleration of the electron by Newton's second law of motion. So

$$\mathbf{F} = -e\mathbf{E} = ma = m\frac{dv}{dt} \qquad (1\text{-}1\text{-}4)$$

where $m = 9.109 \times 10^{-31}$ kg, mass of electron

a = acceleration in meters per second squared

v = velocity of electron in meters per second

$e = 1.602 \times 10^{-19}$ C, charge of electron that is negative

It can be seen that the force is in the opposite direction of the field because the electron has a negative charge. Thus when an electron moves in an electric field $\mathbf{E}$, it experiences a force $-e\mathbf{E}$ newtons. The differential equations of motion for an electron in an electric field in rectangular coordinates are given by

$$\frac{d^2x}{dt^2} = -\frac{e}{m}E_x \qquad (1\text{-}1\text{-}5a)$$

$$\frac{d^2y}{dt^2} = -\frac{e}{m}E_y \qquad (1\text{-}1\text{-}5b)$$

$$\frac{d^2z}{dt^2} = -\frac{e}{m}E_z \qquad (1\text{-}1\text{-}5c)$$

where $e/m = 1.759 \times 10^{11}$ C/kg is the ratio of charge to mass of electron and E_x, E_y, E_z are the components of $\mathbf{E}$ in rectangular coordinates.

In many cases, the equations of motion for electrons in an electric field in cylindrical coordinates are useful. The cylindrical coordinates (r, ϕ, z) are defined as in Fig. 1-1-1.

It can be seen that
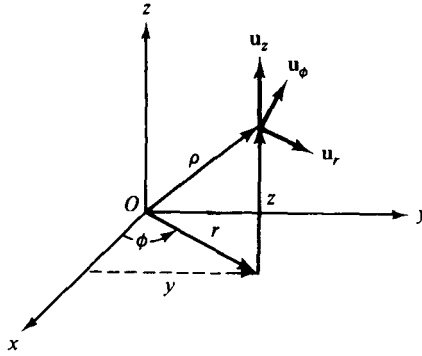
Figure 1-1-1   Cylindrical coordinates.

$$x = r \cos \phi \tag{1-1-6a}$$

$$y = r \sin \phi \tag{1-1-6b}$$

$$z = z \tag{1-1-6c}$$

and, conversely,

$$r = (x^2 + y^2)^{1/2} \tag{1-1-7a}$$

$$\phi = \tan^{-1}\left(\frac{y}{x}\right) = \sin^{-1}\frac{y}{(x^2 + y^2)^{1/2}} = \cos^{-1}\frac{x}{(x^2 + y^2)^{1/2}} \tag{1-1-7b}$$

$$z = z \tag{1-1-7c}$$

A system of unit vectors, $\mathbf{u}_r$, $\mathbf{u}_\phi$, $\mathbf{u}_z$, in the directions of increasing r, ϕ, z, respectively, is also shown in the same diagram. While $\mathbf{u}_z$ is constant, $\mathbf{u}_r$ and $\mathbf{u}_\phi$ are functions of ϕ; that is,

$$\mathbf{u}_r = \cos \phi \, \mathbf{u}_x + \sin \phi \, \mathbf{u}_y \tag{1-1-8a}$$

$$\mathbf{u}_\phi = -\sin \phi \, \mathbf{u}_x + \cos \phi \, \mathbf{u}_y \tag{1-1-8b}$$

Differentiation of Eqs. (1-1-8) with respect to ϕ yields

$$\frac{d\mathbf{u}_r}{d\phi} = \mathbf{u}_\phi \tag{1-1-9a}$$

$$\frac{d\mathbf{u}_\phi}{d\phi} = -\mathbf{u}_r \tag{1-1-9b}$$

The position vector $\boldsymbol{\rho}$ can be expressed in cylindrical coordinates in the form

$$\boldsymbol{\rho} = r\mathbf{u}_r + z\mathbf{u}_z \tag{1-1-9c}$$

Differentiation of Eq. (1-1-9c) with respect to t once for velocity and twice for acceleration yields

$$\mathbf{v} = \frac{d\boldsymbol{\rho}}{dt} = \frac{dr}{dt}\mathbf{u}_r + r\frac{d\mathbf{u}_r}{dt} + \frac{dz}{dt}\mathbf{u}_z = \frac{dr}{dt}\mathbf{u}_r + r\frac{d\phi}{dt}\cdot\frac{d\mathbf{u}_r}{d\phi} + \frac{dz}{dt}\mathbf{u}_z$$

$$= \frac{dr}{dt}\mathbf{u}_r + r\frac{d\phi}{dt}\mathbf{u}_\phi + \frac{dz}{dt}\mathbf{u}_z \tag{1-1-10}$$

$$a = \frac{d\mathbf{v}}{dt} = \left[\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2\right]\mathbf{u}_r + \left(r\frac{d^2\phi}{dt^2} + 2\frac{dr}{dt}\frac{d\phi}{dt}\right)\mathbf{u}_\phi + \frac{d^2z}{dt^2}\mathbf{u}_z$$

$$= \left[\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2\right]\mathbf{u}_r + \frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right)\mathbf{u}_\phi + \frac{d^2z}{dt^2}\mathbf{u}_z \qquad (1\text{-}1\text{-}11)$$

Therefore the equations of motion for electrons in an electric field in cylindrical coordinates are given by

$$\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 = -\frac{e}{m}E_r \qquad (1\text{-}1\text{-}12\text{a})$$

$$\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = -\frac{e}{m}E_\phi \qquad (1\text{-}1\text{-}12\text{b})$$

$$\frac{d^2z}{dt^2} = -\frac{e}{m}E_z \qquad (1\text{-}1\text{-}12\text{c})$$

where E_r, E_ϕ, and E_z are the components of $\mathbf{E}$ in cylindrical coordinates.

From Eq. (1-1-4) the work done by the field in carrying a unit positive charge from point A to point B is

$$-\int_A^B \mathbf{E} \cdot d\ell = \frac{m}{e}\int_{v_A}^{v_B} v\,dv \qquad (1\text{-}1\text{-}13)$$

However, by definition, the potential V of point B with respect to point A is the work done against the field in carrying a unit positive charge from A to B. That is,

$$V \equiv -\int_A^B \mathbf{E} \cdot d\ell \qquad (1\text{-}1\text{-}14)$$

Substitution of Eq. (1-1-14) in Eq. (1-1-13) and integration of the resultant yield

$$eV = \frac{1}{2}m(v_B^2 - v_A^2) \qquad (1\text{-}1\text{-}15)$$

The left side of Eq. (1-1-15) is the potential energy, and the right side represents the change in kinetic energy. The unit of work or energy is called the *electron volt* (eV), which means that if an electron falls through a potential of one volt, its kinetic energy will increase 1 eV. That is,

$$1 \text{ eV} = (1.60 \times 10^{-19} \text{ C})(1\text{V}) = 1.60 \times 10^{-19} \text{ J} \qquad (1\text{-}1\text{-}16)$$

If an electron starts from rest and is accelerated through a potential rise of V volts, its final velocity is

$$v = \left(\frac{2 \text{ eV}}{m}\right)^{1/2} = 0.593 \times 10^6\sqrt{V} \qquad \text{m/s} \qquad (1\text{-}1\text{-}17)$$

Since $d\ell$ is the increment of distance in the direction of an electric field E, the change in potential dV over the distance $d\ell$ can be expressed as

$$|dV| = E\,d\ell \qquad (1\text{-}1\text{-}18)$$

In vector notation it is

$$\mathbf{E} = -\nabla V \tag{1-1-19}$$

where the symbol ∇ is the vector operator in three coordinate systems. The minus sign implies that the field is directed from regions of higher potential to those of lower potential. Equation (1-1-19) is valid in regions in which there is space charge as well as in regions that are free of charge.

1-2 ELECTRON MOTION IN A MAGNETIC FIELD

A charged particle in motion in a magnetic field of flux density $\mathbf{B}$ is experimentally found to experience a force that is directly proportional to the charge Q, its velocity v, the flux density $\mathbf{B}$, and the sine of the angle between the vectors $\mathbf{v}$ and $\mathbf{B}$. The direction of the force is perpendicular to the plane of both $\mathbf{v}$ and $\mathbf{B}$. Therefore the force exerted on the charged particle by the magnetic field can be expressed in vector form as

$$\mathbf{F} = Q\mathbf{v} \times \mathbf{B} \tag{1-2-1}$$

Since the electron has negative charge, then

$$\mathbf{F} = -e\mathbf{v} \times \mathbf{B} \tag{1-2-2}$$

The motion equations of an electron in a magnetic field in rectangular coordinates can be written

$$\frac{d^2x}{dt^2} = -\frac{e}{m}\left(B_z\frac{dy}{dt} - B_y\frac{dz}{dt}\right) \tag{1-2-3a}$$

$$\frac{d^2y}{dt^2} = -\frac{e}{m}\left(B_x\frac{dz}{dt} - B_z\frac{dx}{dt}\right) \tag{1-2-3b}$$

$$\frac{d^2z}{dt^2} = -\frac{e}{m}\left(B_y\frac{dx}{dt} - B_x\frac{dy}{dt}\right) \tag{1-2-3c}$$

Since

$$\mathbf{v} \times \mathbf{B} = (B_z r v_\phi - B_\phi v_z)\mathbf{u}_r + (B_r v_z - B_z v_r)\mathbf{u}_\phi + (B_\phi v_r - B_r r v_\phi)\mathbf{u}_z \tag{1-2-4}$$

the equations of motion for electrons in magnetic field for cylindrical coordinates can be given by

$$\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 = -\frac{e}{m}\left(B_z r\frac{d\phi}{dt} - B_\phi\frac{dz}{dt}\right) \tag{1-2-5a}$$

$$\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = -\frac{e}{m}\left(B_r\frac{dz}{dt} - B_z\frac{dr}{dt}\right) \tag{1-2-5b}$$

$$\frac{d^2z}{dt^2} = -\frac{e}{m}\left(B_\phi\frac{dr}{dt} - B_r r\frac{d\phi}{dt}\right) \tag{1-2-5c}$$
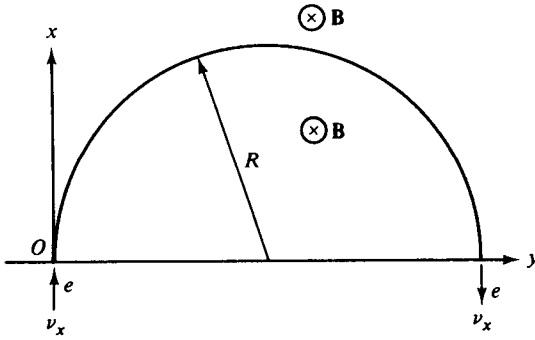
Figure 1-2-1   Circular motion of an electron in a transverse magnetic field.

Consider next an electron moving with a velocity of v_x to enter a constant uniform magnetic field that is perpendicular to v_x as shown in Fig. 1-2-1. The velocity of the electron is assumed to be

$$\mathbf{v} = v_x \mathbf{u}_x \tag{1-2-6}$$

where $\mathbf{u}_x$ is a unit vector in the x direction.

Since the force exerted on the electron by the magnetic field is normal to the motion at every instant, no work is done on the electron and its velocity remains constant. The magnetic field is assumed to be

$$\mathbf{B} = B_z \mathbf{u}_z \tag{1-2-7}$$

Then the magnetic force at the instant when the electron just enters the magnetic field is given by

$$\mathbf{F} = -e\mathbf{v} \times \mathbf{B} = evB\mathbf{u}_y \tag{1-2-8}$$

This means that the force remains constant in magnitude but changes the direction of motion because the electron is pulled by the magnetic force in a circular path. This type of magnetic force is analogous to the problem of a mass tied to a rope and twirled around with constant velocity. The force in the rope remains constant in magnitude and is always directed toward the center of the circle and thus is perpendicular to the motion. At any point on the circle the outward centrifugal force is equal to the pulling force. That is,

$$\frac{mv^2}{R} = evB \tag{1-2-9}$$

where R is the radius of the circle.

From Eq. (1-2-8) the radius of the path is given by

$$R = \frac{mv}{eB} \qquad \text{meters} \tag{1-2-10}$$

The cyclotron angular frequency of the circular motion of the electron is

$$\omega = \frac{v}{R} = \frac{eB}{m} \qquad \text{rad/s} \tag{1-2-11}$$

The period for one complete revolution is expressed by

$$T = \frac{2\pi}{\omega} = \frac{2\pi m}{eB} \qquad \text{seconds} \qquad (1\text{-}2\text{-}12)$$

It should be noted that the radius of the path is directly proportional to the velocity of the electron but that the angular frequency and the period are independent of velocity or radius. This means that faster-moving electrons or particles traverse larger circles in the same time that a slower-moving particle moves in a smaller circle. This very important result is the operating basis of such microwave devices as magnetic-focusing apparatus.

1-3 ELECTRON MOTION IN AN ELECTROMAGNETIC FIELD

If both electric and magnetic fields exist simultaneously, the motion of the electrons depends on the orientation of the two fields. If the two fields are in the same or in opposite directions, the magnetic field exerts no force on the electron, and the electron motion depends only on the electric field, which has been described in Section 1-1. Linear-beam tubes (O-type devices) use a magnetic field whose axis coincides with that of the electron beam to hold the beam together as it travels the length of the tube. In these tubes the electrons receive the full potential energy of the electric field but are not influenced by the magnetic field.

When the electric field $\mathbf{E}$ and the magnetic flux density $\mathbf{B}$ are at right angle to each other, a magnetic force is exerted on the electron beam. This type of field is called a *crossed field*. In a crossed-field tube (M-type device), electrons emitted by the cathode are accelerated by the electric field and gain velocity; but the greater their velocity, the more their path is bent by the magnetic field. The Lorentz force acting on an electron because of the presence of both the electric field $\mathbf{E}$ and the magnetic flux $\mathbf{B}$ is given by

$$\mathbf{F} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = m\frac{d\mathbf{v}}{dt} \qquad (1\text{-}3\text{-}1)$$

The equations of motion for electrons in a crossed field are expressed in rectangular coordinates and cylindrical coordinates, respectively, as

$$\frac{d^2x}{dt^2} = -\frac{e}{m}\left(E_x + B_z\frac{dy}{dt} - B_y\frac{dz}{dt}\right) \qquad (1\text{-}3\text{-}2a)$$

$$\frac{d^2y}{dt^2} = -\frac{e}{m}\left(E_y + B_x\frac{dz}{dt} - B_z\frac{dx}{dt}\right) \qquad (1\text{-}3\text{-}2b)$$

$$\frac{d^2z}{dt^2} = -\frac{e}{m}\left(E_z + B_y\frac{dx}{dt} - B_x\frac{dy}{dt}\right) \qquad (1\text{-}3\text{-}2c)$$

$$\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 = -\frac{e}{m}\left(E_r + B_z r\frac{d\phi}{dt} - B_\phi\frac{dz}{dt}\right) \qquad (1\text{-}3\text{-}3a)$$

$$\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = -\frac{e}{m}\left(E_\phi + B_r\frac{dz}{dt} - B_z\frac{dr}{dt}\right) \tag{1-3-3b}$$

$$\frac{d^2z}{dt^2} = -\frac{e}{m}\left(E_z + B_\phi\frac{dr}{dt} - B_r r\frac{d\phi}{dt}\right) \tag{1-3-3c}$$

where

$$\frac{d\phi}{dt} = \omega_c = \frac{e}{m}B \qquad \text{is the cyclotron frequency}$$

It is, of course, difficult to solve these equations for solutions in three dimensions. In microwave devices and circuits, however, only one dimension is involved in most cases. So the equations of motion become simple and can easily be solved. An example may show how to solve some of the preceding equations.

Example 1-3-1:   Electron Motion in an Electromagnetic Field

The inner cylinder of radius a is the cathode and the outer shell with radius b is the anode. A dc voltage V_0 is applied between the anode and cathode, and a magnetic flux density B is into the page as shown in Fig. 1-3-1. The problem is to adjust the applied voltage V_0 and the magnetic flux density B to such levels that the electrons emitted from the cathode will just graze the anode and travel in space between the cathode and the anode only.
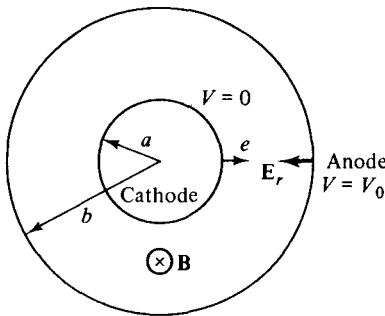
Figure 1-3-1   Electron motion in an electromagnetic field.

Solution

1. Write the equations of motion for electrons in cylindrical coordinates.

a.
$$\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 = +\frac{e}{m}E_r - \frac{e}{m}r\frac{d\phi}{dt}B_z$$

b.
$$\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = \frac{e}{m}B_z\frac{dr}{dt}$$

2. From (b)

$$\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = \frac{1}{2}\omega_c\frac{d}{dt}(r^2) \qquad \left(\text{where } \omega_c = \frac{e}{m}B_0\right)$$

$$r^2\frac{d\phi}{dt} = \frac{1}{2}\omega_c r^2 + \text{constant}$$

3. Application of the boundary conditions: At $r = a$,

$$a^2\frac{d\phi}{dt} = \frac{1}{2}\omega_c a^2 + \text{constant}$$

$$\frac{d\phi}{dt} = 0 \qquad \text{constant} = -\frac{1}{2}\omega_c a^2$$

Hence

$$r^2\frac{d\phi}{dt} = \frac{1}{2}\omega_c(r^2 - a^2)$$

4. The magnetic field does no work on the electrons:

$$\frac{1}{2}mv^2 = eV$$

$$v^2 = \frac{2e}{m}V = v_r^2 + v_\phi^2 = \left(\frac{dr}{dt}\right)^2 + \left(r\frac{d\phi}{dt}\right)^2$$

5. For grazing the anode,

$$r = b \qquad V = V_0 \qquad \frac{dr}{dt} = 0$$

$$b^2\left(\frac{d\phi}{dt}\right)^2 = \frac{2e}{m}V_0 \quad \text{and} \quad b^2\frac{d\phi}{dt} = \frac{1}{2}\omega_c(b^2 - a^2)$$

$$b^2\left[\frac{1}{2}\omega_c\left(1 - \frac{a^2}{b^2}\right)\right]^2 = \frac{2e}{m}V_0$$

6. The cutoff voltage is

$$V_{0c} = \frac{e}{8m}B_0^2 b^2\left(1 - \frac{a^2}{b^2}\right)^2 \tag{1-3-3d}$$

This means that if $V_0 < V_{0c}$ for a given B_0, the electrons will not reach the anode. Conversely, the cutoff magnetic field can be expressed in terms of V_0:

$$B_{0c} = \frac{(8V_0 m/e)^{1/2}}{b(1 - a^2/b^2)} \tag{1-3-4}$$

This implies that if $B_0 > B_{0c}$ for a given V_0, the electrons will not reach the anode.

SUGGESTED READINGS

GEWARTOWSKI, J. W., and H. A. WATSON, *Principles of Electron Tubes*. D. Van Nostrand
    Company, Princeton, N.J., 1965.

KRAUS, J. D., and K. R. CARVER, *Electromagnetics*, 3rd ed. McGraw-Hill Book Company, New York, 1984.

SKITEK, G. G., and S. V. MARSHALL, *Electromagnetic Concepts and Applications*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1982.

PROBLEMS

1-1. At time $t = t_0$ an electron is emitted from a planar diode with zero initial velocity and the anode voltage is $+15$ V. At time $t = t_1$ the electron is midway between the plates and the anode voltage changes discontinuously to -30 V.
   a. Determine which electrode the electron will strike.
   b. Compute the kinetic energy of the electron in electronvolts (eV) when it strikes the electrode.

1-2. A circular cavity is constructed of a center conductor and an outer conductor. The inner center conductor has a radius of 1 cm and is grounded. The outer conductor has a radius of 10 cm and is connected to a power supply of $+10$ kV. The electrons are emitted from the cathode at the center conductor and move toward the anode at the outer conductor.
   a. Determine the magnetic flux density **B** in webers per square meter in the axial direction so that the electrons just graze the anode and return to the cathode again.
   b. If the magnetic flux density **B** is fixed at 4 mWb/m^2, find the required supply voltage so that the electrons just graze the anode and return to the cathode again.

Chapter 2

Electromagnetic Plane Waves

2-0 INTRODUCTION

Since Maxwell's fundamental concepts of electromagnetic wave theory have been established, the electric and magnetic wave equations can readily be derived from Faraday's electromotive force law, Ampère's circuital law, and Gauss's law for the electric and magnetic fields. In Chapter 1 interactions between electron and field were discussed. Here electromagnetic plane waves are described in detail. Many topics associated with electromagnetic waves, such as Poynting theory, reflection theory, attenuation concepts, and plane-wave propagation in metallic-film coating on plastic substrates, are also analyzed, for these basic principles are used frequently in later chapters.

The principles of electromagnetic plane waves are based on the relationships between electricity and magnetism. A changing magnetic field will induce an electric field, and a changing electric field will induce a magnetic field. Also, the induced fields are not confined but ordinarily extend outward into space. The sinusoidal form of the wave causes energy to be interchanged between the magnetic and electric fields in the direction of the wave propagation.

A plane wave has a plane front, a cylindrical wave has a cylindrical front, and a spherical wave has a spherical front. The front of a wave is sometimes referred to as an *equiphase surface*. In the far field of free space, electric and magnetic waves are always perpendicular to each other, and both are normal to the direction of propagation of the wave. This type of wave is known as the *transverse electromagnetic* (TEM) wave. If only the transverse electric wave exists, the wave is called *TE-mode* wave. That means there is no component of the electric wave in the direction of propagation. In *TM modes* only the transverse magnetic wave exists.

2-1 ELECTRIC AND MAGNETIC WAVE EQUATIONS

The electric and magnetic wave equations can be basically derived from Maxwell's equations, which in time domain are expressed as

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2-1-1}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{2-1-2}$$

$$\nabla \cdot \mathbf{D} = \rho_v \tag{2-1-3}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2-1-4}$$

where the vector operator ∇ is expressed by

$$\nabla = \frac{\partial}{\partial x}\mathbf{u}_x + \frac{\partial}{\partial y}\mathbf{u}_y + \frac{\partial}{\partial z}\mathbf{u}_z \qquad \text{(cartesian)} \tag{2-1-4a}$$

$$\nabla = \frac{\partial}{\partial r}\mathbf{u}_r + \frac{\partial}{r\,\partial \phi}\mathbf{u}_\phi + \frac{\partial}{\partial z}\mathbf{u}_z \qquad \text{(cylindrical)} \tag{2-1-4b}$$

$$\nabla = \frac{\partial}{\partial r}\mathbf{u}_r + \frac{\partial}{r\,\partial \theta}\mathbf{u}_\theta + \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\mathbf{u}_\phi \qquad \text{(spherical)} \tag{2-1-4c}$$

It should be noted that boldface roman letters indicate vector quantities or complex quantities. The units of these field variables are

$\mathbf{E}$ = *electric field intensity* in volts per meter
$\mathbf{H}$ = *magnetic field intensity* in amperes per meter
$\mathbf{D}$ = *electric flux density* in coulombs per square meter
$\mathbf{B}$ = *magnetic flux density* in webers per square meter or in tesla
  (1 tesla = 1 weber/m^2 = 10^4 gausses = 3×10^{-6} ESU)
$\mathbf{J}$ = *electric current density* in amperes per square meter
ρ_v = *electric charge density* in coulombs per cubic meter

The electric current density includes two components—that is,

$$\mathbf{J} = \mathbf{J}_c + \mathbf{J}_0 \tag{2-1-5}$$

where $\mathbf{J}_c = \sigma\mathbf{E}$ is called the *conduction current density*
  $\mathbf{J}_0$ = the *impressed current density*, which is independent of the field

The current density source $\mathbf{J}_0$ may really be a current prescribed by an external agency, but more often it is simply one which we know from measurements before we start to find the fields. Alternatively, $\mathbf{J}_0$ is often the current on a surface or in a region [1]. In most cases, however, the current density source $\mathbf{J}_0$ may not exist.

In addition to Maxwell's four equations, the characteristics of the medium in which the fields exist are needed to specify the flux in terms of the fields in a specific medium. These constitutive relationships are

$$\mathbf{D} = \epsilon\mathbf{E} \tag{2-1-6}$$

$$\mathbf{B} = \mu\mathbf{H} \tag{2-1-7}$$

$$\mathbf{J}_c = \sigma\mathbf{E} \tag{2-1-8}$$

$$\epsilon = \epsilon_r\epsilon_0 \tag{2-1-9}$$

$$\mu = \mu_r\mu_0 \tag{2-1-10}$$

where ϵ = dielectric permittivity or capacitivity of the medium in farads per meter

ϵ_r = relative dielectric constant (dimensionless)

ϵ_0 = $8.854 \times 10^{-12} \approx 1/(36\pi) \times 10^{-9}$ F/m is the dielectric permittivity of vacuum or free space

μ = magnetic permeability or inductivity of the medium in henrys per meter

μ_r = the relative permeability or relative inductivity (dimensionless)

μ_0 = $4\pi \times 10^{-7}$ H/m is the permeability of vacuum or free space

σ = conductivity of the medium in mhos per meter

If a sinusoidal time function in the form of $e^{j\omega t}$ is assumed, $\partial/\partial t$ can be replaced by $j\omega$. Then Maxwell's equations in frequency domain are given by

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{2-1-11}$$

$$\nabla \times \mathbf{H} = (\sigma + j\omega\epsilon)\mathbf{E} \tag{2-1-12}$$

$$\nabla \cdot \mathbf{D} = \rho_v \tag{2-1-13}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2-1-14}$$

Taking the curl of Eq. (2-1-11) on both sides yields

$$\nabla \times \nabla \times \mathbf{E} = -j\omega\mu\nabla \times \mathbf{H} \tag{2-1-15}$$

Substitution of Eq. (2-1-12) for the right-hand side of Eq. (2-1-15) gives

$$\nabla \times \nabla \times \mathbf{E} = -j\omega\mu(\sigma + j\omega\epsilon)\mathbf{E} \tag{2-1-16}$$

The vector identity for the curl of the curl of a vector quantity $\mathbf{E}$ is expressed as

$$\nabla \times \nabla \times \mathbf{E} = -\nabla^2\mathbf{E} + \nabla(\nabla \cdot \mathbf{E}) \tag{2-1-17}$$

In free space the space-charge density is zero, and in a perfect conductor time-varying or static fields do not exist. So

$$\nabla \cdot \mathbf{D} = \rho_v = 0 \tag{2-1-18}$$

$$\nabla \cdot \mathbf{E} = 0 \tag{2-1-19}$$

Substitution of Eq. (2-1-17) for the left-hand side of Eq. (2-1-16) and replacement of Eq. (2-1-19) yield the electric wave equation as

$$\nabla^2\mathbf{E} = \gamma^2\mathbf{E} \tag{2-1-20}$$

where $\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} = \alpha + j\beta$ is called the *intrinsic propagation constant* of a medium

α = *attenuation constant* in nepers per meter

β = *phase constant* in radians per meter

Similarly, the magnetic wave equation is given by

$$\nabla^2 \mathbf{H} = \gamma^2 \mathbf{H} \qquad (2\text{-}1\text{-}21)$$

It should be noted that the "double del" or "del squared" is a scalar operator—that is,

$$\nabla \cdot \nabla = \nabla^2 \qquad (2\text{-}1\text{-}22)$$

which is a second-order operator in three different coordinate systems. In rectangular (cartesian) coordinates,

$$\nabla^2 = \frac{\partial}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \qquad (2\text{-}1\text{-}23)$$

In cylindrical (circular) coordinates,

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} \qquad (2\text{-}1\text{-}24)$$

In spherical coordinates,

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial_r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \qquad (2\text{-}1\text{-}24a)$$

Also, the solutions of Eqs. (2-1-1) and (2-1-2) solved simultaneously yield the electric and magnetic wave equations in the time domain as

$$\nabla^2 \mathbf{E} = \mu\sigma\frac{\partial \mathbf{E}}{\partial t} + \mu\epsilon\frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad (2\text{-}1\text{-}25)$$

$$\nabla^2 \mathbf{H} = \mu\sigma\frac{\partial \mathbf{H}}{\partial t} + \mu\epsilon\frac{\partial^2 \mathbf{H}}{\partial t^2} \qquad (2\text{-}1\text{-}26)$$

2-2 POYNTING THEOREM

At what rate will electromagnetic energy be transmitted through free space or any medium, be stored in the electric and magnetic fields, and be dissipated as heat? From the standpoint of complex power in terms of the complex field vectors, the time average of any two complex vectors is equal to the real part of the product of one complex vector multiplied by the complex conjugate of the other vector. Hence the time average of the instantaneous Poynting vector in steady state is given by

$$\langle P \rangle = \langle \mathbf{E} \times \mathbf{H} \rangle = \tfrac{1}{2}\text{Re}(\mathbf{E} \times \mathbf{H}^*) \qquad (2\text{-}2\text{-}1)$$

where the notation $\langle\ \rangle$ stands for the average and the factor of $1/2$ appears in the

equation for complex power when peak values are used for the complex quantities **E** and **H**. Re represents the real part of the complex power, and the asterisk indicates the complex conjugate.

It is necessary to define a complex Poynting vector as

$$\mathbf{P} = \tfrac{1}{2}(\mathbf{E} \times \mathbf{H}^*) \tag{2-2-2}$$

Maxwell's equations in frequency domain for the electric and magnetic fields are

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{2-2-3}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + j\omega\epsilon\mathbf{E} \tag{2-2-4}$$

Dot multiplication of Eq. (2-2-3) by **H*** and of the conjugate of Eq. (2-2-4) by **E** yields

$$(\nabla \times \mathbf{E}) \cdot \mathbf{H}^* = -j\omega\mu\mathbf{H} \cdot \mathbf{H}^* \tag{2-2-5}$$

$$(\nabla \times \mathbf{H}^*) \cdot \mathbf{E} = (\mathbf{J}^* - j\omega\epsilon\mathbf{E}^*) \cdot \mathbf{E} \tag{2-2-6}$$

Then subtraction of Eq. (2-2-5) from Eq. (2-2-6) results in

$$\mathbf{E} \cdot (\nabla \times \mathbf{H}^*) - \mathbf{H}^* \cdot (\nabla \times \mathbf{E}) = \mathbf{E} \cdot \mathbf{J}^* - j\omega(\epsilon|E|^2 - \mu|H|^2) \tag{2-2-7}$$

where $\mathbf{E} \cdot \mathbf{E}^*$ is replaced by $|E|^2$ and $\mathbf{H} \cdot \mathbf{H}^*$ by $|H|^2$.

The left-hand side of Eq. (2-2-7) is equal to $-\nabla \cdot (\mathbf{E} \times \mathbf{H}^*)$ by the vector identity. So

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}^*) = -\mathbf{E} \cdot \mathbf{J}^* + j\omega(\epsilon|E|^2 - \mu|H|^2) \tag{2-2-8}$$

Substituting Eqs. (2-1-5) and (2-2-2) into Eq. (2-2-8), we have

$$-\tfrac{1}{2}\mathbf{E} \cdot \mathbf{J}_0^* = \tfrac{1}{2}\sigma\mathbf{E} \cdot \mathbf{E}^* + j\omega(\tfrac{1}{2}\mu\mathbf{H} \cdot \mathbf{H}^* - \tfrac{1}{2}\epsilon\mathbf{E} \cdot \mathbf{E}^*) + \nabla \cdot \mathbf{P} \tag{2-2-9}$$

Integration of Eq. (2-2-9) over a volume and application of Gauss's theorem to the last term on the right-hand side give

$$\int_v \tfrac{1}{2}(\mathbf{E} \cdot \mathbf{J}_0^*)dv = \int_v \tfrac{1}{2}\sigma|E|^2 dv + j2\omega \int_v (w_m - w_e)dv + \oint_S \mathbf{P} \cdot ds \tag{2-2-10}$$

where $\tfrac{1}{2}\sigma|E|^2 = \sigma\langle|E|^2\rangle$ is the time-average dissipated power

$\tfrac{1}{4}\mu\mathbf{H} \cdot \mathbf{H}^* = \tfrac{1}{2}\mu\langle|H|\rangle^2 = w_m$ is the time-average magnetic stored energy

$\tfrac{1}{4}\epsilon\mathbf{E} \cdot \mathbf{E}^* = \tfrac{1}{2}\epsilon\langle|E|^2\rangle = w_e$ is the time-average electric stored energy

$-\tfrac{1}{2}\mathbf{E} \cdot \mathbf{J}_0^* =$ the complex power impressed by the source $\mathbf{J}_0$ into the field

Equation (2-2-10) is well known as the *complex Poynting theorem* or the *Poynting theorem* in frequency domain.

Furthermore, let

$P_{\text{in}} = -\displaystyle\int_v \tfrac{1}{2}(\mathbf{E} \cdot \mathbf{J}_0^*)dv$ be the total complex power suplied by a source within a region

$\langle P_d \rangle = \displaystyle\int_v \tfrac{1}{2}\sigma|E|^2 dv$ be the time-average power dissipated as heat inside the region

$$\langle W_m - W_e \rangle = \int_v (w_m - w_e)dv \qquad \text{be the difference between time-average magnetic and electric energies stored within the region}$$

$$P_{\text{tr}} = \oint \mathbf{P} \cdot d\mathbf{s} \qquad \text{be the complex power transmitted from the region}$$

The complex Poynting theorem shown in Eq. (2-2-10) can be simplified to

$$P_{\text{in}} = \langle P_d \rangle + j2\omega[\langle W_m - W_e \rangle] + P_{\text{tr}} \qquad (2\text{-}2\text{-}11)$$

This theorem states that the total complex power fed into a volume is equal to the algebraic sum of the active power dissipated as heat, plus the reactive power proportional to the difference between time-average magnetic and electric energies stored in the volume, plus the complex power transmitted across the surface enclosed by the volume.

2-3 UNIFORM PLANE WAVES AND REFLECTION

2-3-1 Uniform Plane Waves

A *plane wave* is a wave whose phase is constant over a set of planes. A *uniform* plane wave is a wave whose magnitude and phase are both constant. A spherical wave in free space is a uniform plane wave as observed at a far distance. Its equiphase surfaces are concentric spheres, expanding as the wave travels outward from the source, and its magnitude is constant.

Electromagnetic waves in free space are typical uniform plane waves. The electric and magnetic fields are mutually perpendicular to each other and to the direction of propagation of the waves. The phases of the two fields are always in time phase and their magnitudes are always constant. The stored energies are equally divided between the two fields, and the energy flow is transmitted by the two fields in the direction of propagation. Thus a uniform plane wave is a transverse electromagnetic wave or a TEM wave.

A nonuniform plane wave is a wave whose amplitude (not phase) may vary within a plane normal to the direction of propagation. Consequently, the electric and magnetic fields are no longer in time phase.

Since a uniform plane wave of electric or magnetic field has no variation of intensity in a plane normal to the direction of propagation of the wave, then

$$\frac{\partial \mathbf{E}}{\partial x} = \frac{\partial \mathbf{E}}{\partial y} = 0 \quad \text{or} \quad \frac{\partial \mathbf{H}}{\partial x} = \frac{\partial \mathbf{H}}{\partial y} = 0$$

if the direction of propagation is assumed in the positive z direction.

With the preceding assumptions and lossless dielectric—that is, $\sigma = 0$—the wave equations (2-1-25) and (2-1-26) in time domain for the electric and magnetic intensities in free space for rectangular coordinates reduce to

$$\frac{\partial^2 E_x}{\partial z^2} = \mu_0 \epsilon_0 \frac{\partial^2 E_x}{\partial t^2} \qquad (2\text{-}3\text{-}1)$$

$$\frac{\partial^2 H_y}{\partial z^2} = \mu_0 \epsilon_0 \frac{\partial^2 H_y}{\partial t^2} \tag{2-3-2}$$

in which the electric intensity is arbitrarily chosen in the x direction and the magnetic intensity in the y direction. With no loss in generality, it can be assumed that the electric intensity is given by

$$E_x = E_0 e^{j(\omega t - \beta z)} = E_0 e^{j\omega\{t - (\beta/\omega)z\}} \tag{2-3-3}$$

The magnetic intensity can be obtained by inserting Eq. (2-3-3) into the curl equation

$$\nabla \times \mathbf{E} = -j\omega\mu_0 \mathbf{H} \tag{2-3-4}$$

For the assumed conditions, the curl equation reduces to

$$\frac{\partial E_x}{\partial z} = -j\omega\mu_0 H_y \tag{2-3-5}$$

Differentiation of Eq. (2-3-3) with respect to z and substitution of the result in Eq. (2-3-5) yield

$$H_y = \sqrt{\frac{\epsilon_0}{\mu_0}} E_x \tag{2-3-6}$$

where $\dfrac{\beta}{\omega} = \sqrt{\mu_0 \epsilon_0} = \dfrac{1}{v_p}$ is accounted for in the derivation

$v_p = \dfrac{1}{\sqrt{\mu_0 \epsilon_0}} = 3 \times 10^8$ m/s is phase velocity

$\quad = c$, the velocity of light in vacuum

$\alpha = 0$

$\beta = \omega\sqrt{\mu_0 \epsilon_0}$ is phase constant

The ratio of electric to magnetic intensities is given by

$$\frac{E_x}{H_y} = \eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 377 \ \Omega \tag{2-3-7}$$

It is called the *intrinsic impedance* of free space.

Figure 2-3-1 shows uniform electric and magnetic plane waves in rectangular coordinates. In general, for a uniform plane wave propagating in a lossless dielectric medium ($\sigma = 0$), the characteristics of wave propagation would become

$$\alpha = 0 \tag{2-3-8a}$$

$$\beta = \omega\sqrt{\mu_0 \epsilon} \tag{2-3-8b}$$

$$\eta = \sqrt{\frac{\mu_0}{\epsilon}} = \frac{\eta_0}{\sqrt{\epsilon_r}} \tag{2-3-8c}$$

$$v_p = \frac{1}{\sqrt{\mu_0 \epsilon}} = \frac{c}{\sqrt{\epsilon_r}} \tag{2-3-8d}$$
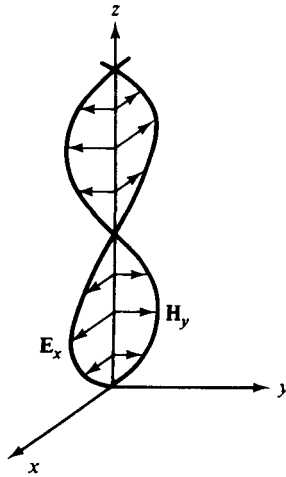
Figure 2-3-1   Uniform plane waves traveling in the z direction.

2-3-2 Boundary Conditions

Since Maxwell's equations are in the form of differential rather than algebraic equations, boundary conditions must be applied to a given problem if a specific solution is required.

There are four basic rules for boundary conditions at the surface between two different materials:

1. The tangential components of electric field intensity are continuous across the boundary.
2. The normal components of electric flux density are discontinuous at the boundary by an amount equal to the surface-charge density on the boundary.
3. The tangential components of magnetic field intensity are discontinuous at the boundary by an amount equal to the surface-current density on the boundary.
4. The normal components of magnetic flux density are continuous across the boundary.

The four statements can be proved by applying Faraday's law, Gauss's law, Ampère's law, and $\nabla \cdot \mathbf{B} = 0$ to the boundaries of Fig. 2-3-2(a) and (b).

It can be seen from the diagrams that

$$\oint_\ell \mathbf{E} \cdot d\ell = E_{t1} \, \Delta\ell - E_{t2} \, \Delta\ell = 0$$

$$\oint_s \mathbf{D} \cdot d\mathbf{s} = D_{n1} \, \Delta s - D_{n2} \, \Delta s = p_s \, \Delta s$$

$$\oint_\ell \mathbf{H} \cdot d\ell = H_{t1} \, \Delta\ell - H_{t2} \, \Delta\ell = J_s \, \Delta\ell$$

$$\oint \mathbf{B} \cdot d\mathbf{s} = B_{n1} \, \Delta s - B_{n2} \, \Delta s = 0$$
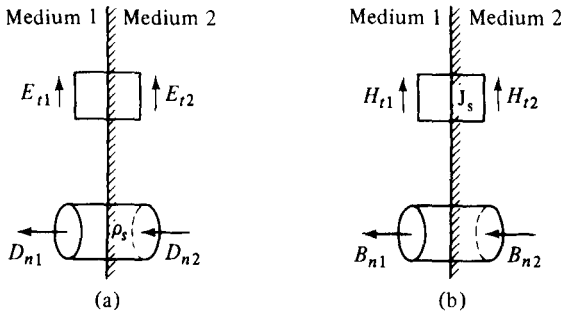
Figure 2-3-2 Boundary conditions. (a) Electric intensity. (b) Magnetic intensity.

So the boundary equations are

$$E_{t1} = E_{t2} \tag{2-3-9a}$$

$$D_{n1} = D_{n2} + p_s \tag{2-3-9b}$$

$$H_{t1} = H_{t2} + J_s \tag{2-3-9c}$$

$$B_{n1} = B_{n2} \tag{2-3-9d}$$

If medium 2 is a perfect conductor ($\sigma = \infty$, $\epsilon_r = 1$, $\mu_r = 1$) and medium 1 is a perfect dielectric (vacuum or free space, $\sigma = 0$, ϵ_0, μ_0), then

$$E_{t1} = \frac{D_{t1}}{\epsilon_0} = 0 \tag{2-3-10a}$$

$$D_{n1} = \epsilon_0 E_{n1} = p_s \tag{2-3-10b}$$

$$H_{t1} = 0 \qquad \text{if } J_s = 0 \tag{2-3-10c}$$

$$B_{n1} = 0 \tag{2-3-10d}$$

2-3-3 Uniform Plane-Wave Reflection

Normal-incidence reflection. The simplest reflection problem is that of a uniform plane wave normally incident on a plane boundary between two dielectric media with no surface-charge density and surface-current density. This situation is shown in Fig. 2-3-3.

In medium 1 the fields are the sum of an incident wave plus a reflected wave, respectively.

$$E_x^{(1)} = E_0(e^{-j\beta_1 z} + \Gamma e^{j\beta_1 z}) \tag{2-3-11}$$

$$H_y^{(1)} = \frac{E_0}{\eta_1}(e^{-j\beta_1 z} - \Gamma e^{j\beta_1 z}) \tag{2-3-12}$$

where $\beta_1 = \omega\sqrt{\mu_1\epsilon_1}$

$\eta_1 = \sqrt{\dfrac{\mu_1}{\epsilon_1}} = \dfrac{\eta_0}{\sqrt{\epsilon_{r1}}} = $ intrinsic wave impedance of medium 1

$\Gamma = $ reflection coefficient
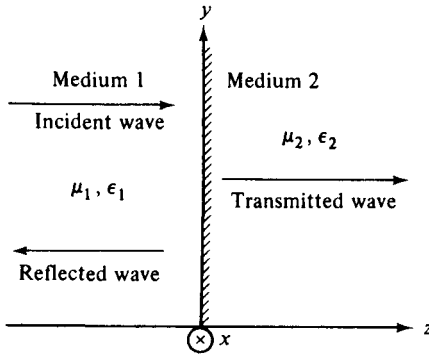
Figure 2-3-3 Uniform plane-wave reflection.

In medium 2 there are transmitted waves:

$$E_x^{(2)} = E_0 T e^{-j\beta_2 z} \qquad (2\text{-}3\text{-}13)$$

$$H_y^{(2)} = \frac{E_0}{\eta_2} T e^{-j\beta_2 z} \qquad (2\text{-}3\text{-}14)$$

where $\beta_2 = \omega \sqrt{\mu_2 \epsilon_2}$

$\eta_2 = \sqrt{\dfrac{\mu_2}{\epsilon_2}} = \dfrac{\eta_0}{\sqrt{\epsilon_{r2}}} =$ intrinsic wave impedance of medium 2

T = transmission coefficient

For continuity of wave impedance at the boundary, the wave impedance is

$$Z_z = \frac{E_x^{(1)}}{H_y^{(1)}}\bigg|_{z=0} = \eta_1 \frac{1 + \Gamma}{1 - \Gamma} = \frac{E_x^{(2)}}{H_y^{(2)}}\bigg|_{z=0} = \eta_2 \qquad (2\text{-}3\text{-}15)$$

Hence the reflection coefficient is given by

$$\Gamma = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \qquad (2\text{-}3\text{-}16)$$

From the boundary condition the tangential components of electric field intensity are continuous across the interface. Then

$$E_x^{(1)}\bigg|_{z=0} = E_0(1 + \Gamma) = E_x^{(2)}\bigg|_{z=0} = E_0 T \qquad (2\text{-}3\text{-}17)$$

Hence the transmission coefficient is expressed as

$$T = 1 + \Gamma = \frac{2\eta_2}{\eta_2 + \eta_1} \qquad (2\text{-}3\text{-}18)$$

If medium 1 is lossless dielectric (that is, $\sigma = 0$), the standing-wave ratio is defined as

$$\text{SWR} = \rho = \frac{\left| E_{\max}^{(1)} \right|}{\left| E_{\min}^{(1)} \right|} = \frac{1 + \Gamma}{1 - \Gamma} \qquad (2\text{-}3\text{-}19)$$

The power density transmitted across the boundary is

$$p_{tr} = \frac{1}{2}(\mathbf{E} \times \mathbf{H}^*)\Big|_{z=0} \cdot \mathbf{u}_z = \frac{E_0^2}{2\eta_1}(1 - \Gamma^2) \tag{2-3-20}$$

Then

$$p_{tr} = p_{inc}(1 - \Gamma^2) \tag{2-3-21}$$

where p_{inc} is the incident power density.

The incident power density minus the transmitted power density would yield the reflected power density as

$$p_{ref} = p_{inc}\Gamma^2 \tag{2-3-22}$$

Oblique-incidence reflection.

E Is in the Plane of Incidence.   The plane of incidence is defined by the direction of propagation and the line normal to the boundary. The linearly polarized uniform plane waves with **E** lying in and **H** normal to the plane of incidence are impinging obliquely on a boundary between two lossless dielectric materials as shown in Fig. 2-3-4.

Whenever a wave is incident obliquely on the boundary surface between two media, the polarization of the wave is vertical or horizontal if the electric field is normal to or parallel to the boundary surface. The terms *horizontal* and *vertical polarization* refer to the phenomenon of waves from horizontal and vertical antennas, respectively, producing the corresponding orientations of wave polarization when
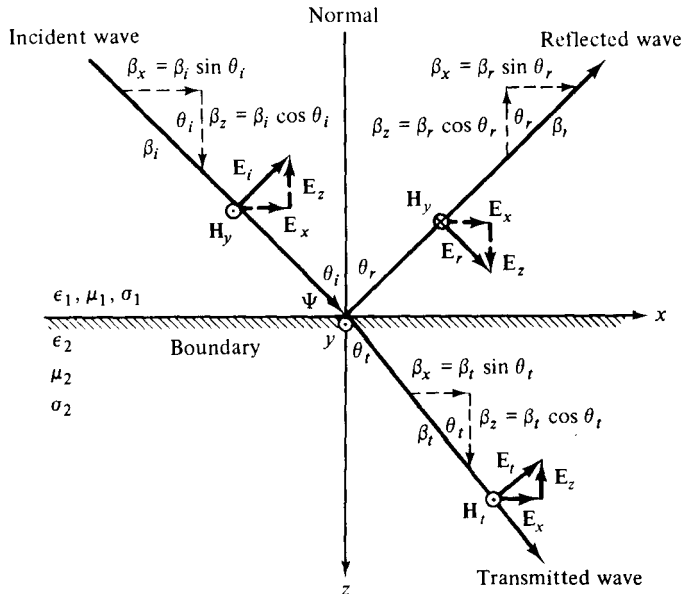
Figure 2-3-4   Reflection and transmission of oblique incidence.

the waves strike the surface of the earth. For guided waves in waveguides, the terms *transverse electric* (TE) and *transverse magnetic* (TM) are used to designate the fact that either the electric or the magnetic field is normal to the direction of propagation. The polarization of a wave is an extremely useful concept for computing electromagnetic power flow. For example, a Poynting vector indicates that the power-flow density is the cross product of an electric and a magnetic field with the specific direction determined by the polarizations of the two fields.

As Fig. 2-3-4 shows, for a lossless dielectric medium, the phase constants (or propagation constants) of the two media in the x direction on the interface are equal as required by the continuity of tangential $\mathbf{E}$ and $\mathbf{H}$ on the boundary. Thus

$$\beta_i \sin \theta_i = \beta_r \sin \theta_r \tag{2-3-23a}$$

$$\beta_i \sin \theta_i = \beta_t \sin \theta_t \tag{2-3-23b}$$

From Eq. (2-3-23a), since $\beta_i = \beta_r = \beta_1$, the angle of reflection is equal to the angle of incidence. This is

$$\theta_1 = \theta_r \tag{2-3-24}$$

From Eq. (2-3-23b)

$$\frac{\sin \theta_t}{\sin \theta_i} = \frac{\beta_i}{\beta_t} = \frac{v_2}{v_1} = \sqrt{\frac{\mu_1 \epsilon_1}{\mu_2 \epsilon_2}} \tag{2-3-25}$$

where v represents the phase velocity. This is well known as Snell's law. In general, all low-loss dielectrics have equal permeability—that is, $\mu_1 = \mu_2 = \mu_0$. If medium 2 is free space and medium 1 is a nonmagnetic dielectric, the right-hand side of Eq. (2-3-25) becomes $\sqrt{\epsilon_r}$, which is called the *index of refraction* of the dielectric.

The components of electric intensity $\mathbf{E}$ are

$$E_x = E_0 \cos \theta_i e^{-j\beta_1(x \sin \theta_i + z \cos \theta_i)} \tag{2-3-26}$$

$$E_y = 0 \tag{2-3-27}$$

$$E_z = -E_0 \sin \theta_i e^{-j\beta_1(x \sin \theta_i + z \cos \theta_i)} \tag{2-3-28}$$

The components of magnetic intensity $\mathbf{H}$ are

$$H_x = 0 \tag{2-3-29}$$

$$H_y = \frac{E_0}{\eta_1} e^{-j\beta_1(x \sin \theta_i + z \cos \theta_i)} \tag{2-3-30}$$

$$H_z = 0 \tag{2-3-31}$$

The wave impedance in the z direction is given by

$$Z_z = \frac{E_x}{H_y} = \eta \cos \theta \tag{2-3-32}$$

It should be noted that the subscripts of η and θ have dropped because the wave impedances of the two regions in the z direction are the same.

The wave impedance can be expressed in terms of the reflection coefficient of

the normal components at the boundary $z = 0$. In medium 1

$$E_x^{(1)} = E_0 \cos \theta_i [e^{-j\beta_1 z \cos \theta_i} + \Gamma e^{j\beta_1 z \cos \theta_r}] \qquad (2\text{-}3\text{-}33)$$

$$H_y^{(1)} = \frac{E_0}{\eta_1} [e^{-j\beta_1 z \cos \theta_i} - \Gamma e^{j\beta_1 z \cos \theta_r}] \qquad (2\text{-}3\text{-}34)$$

$$Z_z = \frac{E_x^{(1)}}{H_y^{(1)}} \bigg|_{z=0} = \eta_1 \cos \theta_i \frac{1 + \Gamma}{1 - \Gamma} \qquad (2\text{-}3\text{-}35)$$

The impedance must be equal to the z-directed wave impedance in region 2 at the boundary. Substitution of $Z_z = \eta_2 \cos \theta_t$ in Eq. (2-3-35) yields

$$\Gamma = \frac{\eta_2 \cos \theta_t - \eta_1 \cos \theta_i}{\eta_2 \cos \theta_t + \eta_1 \cos \theta_i} \qquad (2\text{-}3\text{-}36)$$

Then the transmission coefficient is given by

$$T = \frac{2\eta_2 \cos \theta_t}{\eta_2 \cos \theta_t + \eta_1 \cos \theta_i} \qquad (2\text{-}3\text{-}37)$$

The preceding two equations are known as Fresnel's formulas for **E** in the plane of incidence.

H Is in the Plane of Incidence.   If **H** is in the plane of incidence, the components of **H** are

$$H_x = H_0 \cos \theta_i e^{-j\beta_1(x \sin \theta_i + z \cos \theta_i)} \qquad (2\text{-}3\text{-}38)$$

$$H_y = 0 \qquad (2\text{-}3\text{-}39)$$

$$H_z = -H_0 \sin \theta_i e^{-j\beta_1(x \sin \theta_i + z \cos \theta_i)} \qquad (2\text{-}3\text{-}40)$$

The components of electric intensity **E** normal to the plane of incidence are

$$E_x = 0 \qquad (2\text{-}3\text{-}41)$$

$$E_y = -\eta_1 H_0 e^{-j\beta_1(x \sin \theta_i + z \cos \theta_i)} \qquad (2\text{-}3\text{-}42)$$

$$E_z = 0 \qquad (2\text{-}3\text{-}43)$$

The wave impedance in the z direction is given by

$$Z_z = -\frac{E_y}{H_x} = \frac{\eta}{\cos \theta} = \eta \sec \theta \qquad (2\text{-}3\text{-}44)$$

It should be noted that the subscripts of η and θ have been dropped for the same reason stated previously.

Fresnel's formulas for **H** in the plane of incidence are

$$\Gamma = \frac{\eta_2 \sec \theta_t - \eta_1 \sec \theta_i}{\eta_2 \sec \theta_t + \eta_1 \sec \theta_i} \qquad (2\text{-}3\text{-}45)$$

$$T = \frac{2\eta_2 \sec \theta_t}{\eta_2 \sec \theta_t + \eta_1 \sec \theta_i} \qquad (2\text{-}3\text{-}46)$$

2-4 PLANE-WAVE PROPAGATION IN FREE SPACE AND LOSSLESS DIELECTRIC

2-4-1 Plane-Wave Propagation In Free Space

The electromagnetic wave being propagated in free space near the surface of the earth is divided into two parts: the ground wave and the sky wave or ionosphere wave. The ground wave is further divided into a direct wave, an earth-reflected wave, and a surface wave. Figure 2-4-1 shows the wave components of electromagnetic wave from a nondirectional antenna to a receiving station.
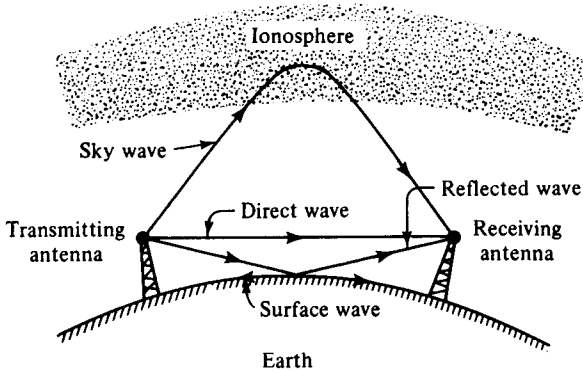
Figure 2-4-1   Wave components near the surface of the earth.

The ionosphere is that region of the earth's atmosphere in which the constituent gases are ionized by solar radiation. This region extends from about 50 km above the earth to several earth radii and has different layers designated as C, D, E, and F layers in order of height. The electron-density distribution of each layer varies with the time of day, season, year, and the geographical location. During the day the electron density N is approximately 10^{12} electrons per cubic meter at an altitude between 90 and 1000 km. The E and F layers have a permanent existence, but the D layer is present only during the day. The electron density determines the reflection and refraction of microwaves. For vertical incidence, the critical frequency is given by

$$F_{cr} = 9\sqrt{N_{max}} \qquad Hz \qquad (2\text{-}4\text{-}1)$$

This means that a microwave of frequency F_{cr} will be reflected back to the earth if the electron density is equal to or higher than the required maximum electron density N_{max} (electrons per cubic meter).

The sky wave reaches the receiving station after reflection from the ionosphere. Although important in many communication systems, the sky wave need not be considered in most microwave applications because a wavelength shorter than about 4 m will not return to the earth from the ionosphere. The reflected wave is reflected from the earth in such a way as to reach the receiver. Energy radiated from the nondirectional antenna of Fig. 2-4-1 strikes the earth at all points between the base of the antenna and the horizon, but only that wave which leaves the antenna in the direction shown reaches the receiver. The surface wave is a wave diffracted

around the surface of the earth or guided by the ground-air interface. This component is important at broadcast frequencies; at microwave frequencies, however, the surface wave is rapidly attenuated, and at a distance of 2 km from the antenna it has an amplitude of only a fraction of 1% of the direct wave. This component must be considered in blind-landing systems in which ranges of less than 2 km are important. The direct wave travels a nearly straight path from the transmitting antenna to the receiving station. It is the only wave considered in this book. The term *free space* will be used to denote vacuum or any other medium having essentially the same characteristics as vacuum, such as open air, anechoic chamber, and shielded enclosure. When power radiates from the transmitting antenna, the power density carried by the spherical wave will decrease with distance as the energy in the wave spreads out over an ever-increasing surface area as the wave progresses.

The power density is given by

$$p_d = \frac{p_t g_t}{4\pi R^2} \qquad \text{W/m}^2 \qquad (2\text{-}4\text{-}2)$$

where p_t = transmitting power in watts

g_t = transmitting antenna gain (numerical)

R = distance between antenna and field point in meters

The power received by the receiving antenna is given as

$$p_r = p_d A_e = \left(\frac{p_t g_t}{4\pi R^2}\right)\left(\frac{\lambda_2}{4\pi} g_r\right) \qquad \text{watts} \qquad (2\text{-}4\text{-}3)$$

where $A_e = \dfrac{\lambda^2}{4\pi} g_r$ = effective antenna aperture in square meters

$\dfrac{\lambda^2}{4\pi} = A_a$ = isotropic antenna aperture in square meters

g_r = receiving antenna gain (numerical)

Figure 2-4-2 shows the relationships of electromagnetic energy transmission in free space between two antennas.
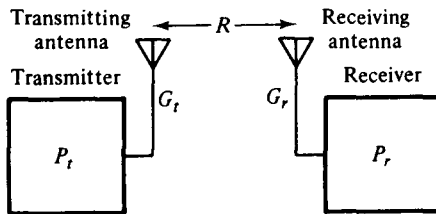
Figure 2-4-2 Electromagnetic energy transmission between two antennas.

If the received power is expressed in terms of decibels, Eq. (2-4-3) becomes

$$P_r = P_t + G_t + G_r - 20 \log\left(\frac{4\pi R}{\lambda}\right) \qquad \text{dBW} \qquad (2\text{-}4\text{-}4)$$

where P_t is in dBW, G_t and G_r are in decibels (dB). The term $20 \log (4\pi R/\lambda)$ is well known as the free-space attenuation in decibels. It can easily be found from the stan-

dard nomograph shown in Fig. 2-4-3. For example, if the wavelength of a signal is 0.03 m and the range is 20 m, the free-space attenuation is about 79 dB.

It should be noted that the free-space attenuation is entirely different from the dissipative attenuation of a medium such as atmosphere that absorbs energy from the wave. The factor $(4\pi R^2)$ in Eq. (2-4-3) simply accounts for the fact that the power density is inversely proportionally decreasing with the squared distance when the energy spreads out over free space. The factor $(\lambda^2/4\pi)$ is the isotropic aperture of a receiving antenna. It does not imply that a higher-frequency wave decreases in magnitude more rapidly than a lower-frequency wave. It is simply a consequence of the fact that, for a given antenna gain, the aperture of a higher-frequency antenna is smaller than that of a lower-frequency antenna so that it intercepts a smaller amount of power from the wave.
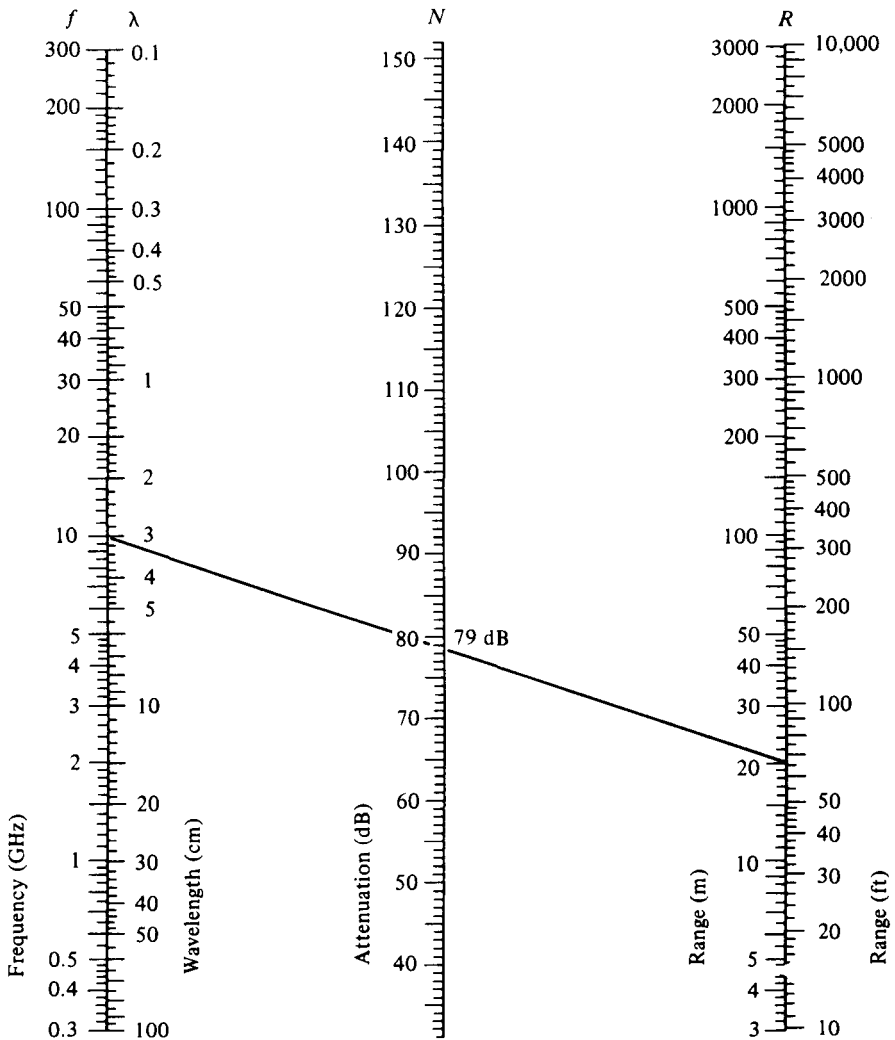
Figure 2-4-3   Nomograph of free-space attentuation.

2-4-2 Plane-Wave Propagation in Lossless Dielectric

The lossless dielectric, which is often called the good or perfect dielectric, is characterized by $\sigma = 0$. Hence the intrinsic impedance for a lossless dielectric can be expressed in terms of air. This is

$$\eta = \sqrt{\frac{\mu}{\epsilon}} = \sqrt{\frac{\mu_0}{\epsilon_r \epsilon_0}} = \frac{377}{\sqrt{\epsilon_r}} \qquad \text{ohms} \qquad (2\text{-}4\text{-}5)$$

The attenuation constant α is zero, and the phase constant β is given by

$$\beta = \omega \sqrt{\mu\epsilon} \qquad (2\text{-}4\text{-}6)$$

The phase velocity is expressed by

$$v_p = \frac{1}{\sqrt{\mu\epsilon}} \qquad (2\text{-}4\text{-}7)$$

2-5 PLANE-WAVE PROPAGATION IN LOSSY MEDIA

The lossy media are characterized by $\sigma \neq 0$. There are three types of lossy media: good conductor, poor conductor, and lossy dielectric, which are discussed in this section. The presence of a loss in the medium introduces wave dispersion by conductivity. Dispersion makes a general solution in the time domain impossible except by Fourier expansion methods. Thus only solutions for the frequency domain (or steady state) will be given.

The electric and magnetic wave equations in the frequency domain as shown in Eqs. (2-1-20) and (2-1-21) are repeated here:

$$\nabla^2 \mathbf{E} = j\omega\mu(\sigma + j\omega\epsilon)\mathbf{E} \qquad (2\text{-}5\text{-}1)$$

$$\nabla^2 \mathbf{H} = j\omega\mu(\sigma + j\omega\epsilon)\mathbf{H} \qquad (2\text{-}5\text{-}2)$$

For one dimension in the positive z direction, they become

$$\frac{\partial^2 E_x}{\partial z^2} = j\omega\mu(\sigma + j\omega\epsilon)E_x \qquad (2\text{-}5\text{-}3\text{a})$$

$$\frac{\partial^2 H_y}{\partial z^2} = j\omega\mu(\sigma + j\omega\epsilon)H_y \qquad (2\text{-}5\text{-}3\text{b})$$

The complex-frequency solutions would be given by

$$E_x = E_0 e^{-\alpha z} \cos(\omega t - \beta z) \qquad (2\text{-}5\text{-}4)$$

$$H_y = \frac{E_0}{\eta} e^{-\alpha z} \cos(\omega t - \beta z) \qquad (2\text{-}5\text{-}5)$$

where $\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} = \alpha + j\beta$

$$\eta = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}}$$

2-5-1 Plane Wave in Good Conductor

A good conductor is defined as one having a very high conductivity; consequently, the conduction current is much larger than the displacement current. The energy transmitted by the wave traveling through the medium will decrease continuously as the wave propagates because ohmic losses are present. Expressed mathematically, a good conductor requires the criterion

$$\sigma \gg \omega\epsilon \qquad (2\text{-}5\text{-}6)$$

The propagation constant γ is expressed as

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} = j\omega\sqrt{\mu\epsilon}\sqrt{1 - j\frac{\sigma}{\omega\epsilon}}$$

$$= j\omega\sqrt{\mu\epsilon}\sqrt{-j\frac{\sigma}{\omega\epsilon}} \qquad \text{for } \frac{\sigma}{\omega\epsilon} \gg 1$$

$$= j\sqrt{\omega\mu\sigma}\sqrt{-j} = j\sqrt{\omega\mu\sigma}\left(\frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}\right) \qquad (2\text{-}5\text{-}7)$$

$$= (1 + j)\sqrt{\pi f\mu\sigma}$$

Hence

$$\alpha = \beta = \sqrt{\pi f\mu\sigma} \qquad (2\text{-}5\text{-}8)$$

The exponential factor $e^{-\alpha z}$ of the traveling wave becomes $e^{-1} = 0.368$ when

$$z = \frac{1}{\sqrt{\pi f\mu\sigma}} \qquad (2\text{-}5\text{-}9)$$

This distance is called the *skin depth* and is denoted by

$$\delta = \frac{1}{\sqrt{\pi f\mu\sigma}} = \frac{1}{\alpha} = \frac{1}{\beta} \qquad (2\text{-}5\text{-}10)$$

Interestingly, at microwave frequencies the skin depth is extremely short and a piece of glass with an evaporated silver coating $5.40\text{-}\mu\text{m}$ thick is an excellent conductor at these frequencies. Table 2-5-1 lists the conductivities of materials.

The intrinsic impedance of a good conductor is given as

$$\eta = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}} = \sqrt{\frac{j\omega\mu}{\sigma}} \qquad \text{for } \sigma \gg \omega\epsilon$$

$$= \sqrt{\frac{\omega\mu}{\sigma}}\underline{/45°} = (1 + j)\sqrt{\frac{\omega\mu}{2\sigma}} \qquad (2\text{-}5\text{-}11)$$

$$= (1 + j)\frac{1}{\sigma\delta} = (1 + j)R_s$$

in which $R_s = \sqrt{\omega\mu/(2\sigma)}$ is known as the *skin effect* and the magnitude of the con-

TABLE 2-5-1 TABLE OF CONDUCTIVITIES

Substance	Type	Conductivity σ (mhos/m)
Quartz, fused	insulator	10^{-17} approx.
Ceresin, wax	insulator	10^{-17} approx.
Sulfur	insulator	10^{-15} approx.
Mica	insulator	10^{-15} approx.
Paraffin	insulator	10^{-15} approx.
Rubber, hard	insulator	10^{-15} approx.
Glass	insulator	10^{-12} approx.
Bakelite	insulator	10^{-9} approx.
Distilled water	insulator	10^{-4} approx.
Seawater	conductor	4 approx.
Tellurium	conductor	5×10^2 approx.
Carbon	conductor	3×10^4 approx.
Graphite	conductor	10^5 approx.
Cast iron	conductor	10^6 approx.
Mercury	conductor	10^6
Nichrome	conductor	10^6
Constantan	conductor	2×10^6
Silicon steel	conductor	2×10^6
German silver	conductor	3×10^6
Lead	conductor	5×10^6
Tin	conductor	9×10^6
Phosphor bronze	conductor	10^7
Brass	conductor	1.1×10^7
Zinc	conductor	1.7×10^7
Tungsten	conductor	1.8×10^7
Duralumin	conductor	3×10^7
Aluminum, hard-drawn	conductor	3.5×10^7
Gold	conductor	4.1×10^7
Copper	conductor	5.8×10^7
Silver	conductor	6.1×10^7

ductor surface resistance. The average power density for a good conductor is given by

$$p = \tfrac{1}{2}|H|^2 R_s \qquad (2\text{-}5\text{-}12)$$

and the phase velocity within a good conductor is

$$v = \omega\delta \qquad (2\text{-}5\text{-}13a)$$

The reflectivity and transmittance of a good conductor in vertical and horizontal polarizations are usually measured in terms of the grazing angle. The grazing angle ψ is defined as the angle between the incident ray and the media boundary.

Vertical polarization. From Fig. 2-3-4 it can be seen that $\psi = 90° - \theta_i$; then $\sin \psi = \cos \theta_i$, $\sin \theta_i = \cos \psi$, $\sin^2 \theta_t + \cos^2 \theta_t = 1$, and $v_1 \sin \theta_t = v_2 \sin \theta_i$. The vertical reflectivity of a good conductor for the tangential components of electric intensity as shown in Eq. (2-3-36) is simplified to

$$\Gamma_v = \frac{\eta_2[1 - (v_2/v_1 \cos \psi)^2]^{1/2} - \eta_1 \sin \psi}{\eta_2[1 - (v_2/v_1 \cos \psi)^2]^{1/2} + \eta_1 \sin \psi} \qquad (2\text{-}5\text{-}13b)$$

For vertical polarization, the normal components of the electric fields are generally used to determine the reflection coefficient. From Fig. 2-3-4 it can be seen that the vertical components of the incident and reflected electric fields are in opposite directions. Therefore the reflectivity of a good conductor in vertical polarization is

$$\Gamma_v = \frac{\eta_1 \sin \psi - \eta_2[1 - (v_2/v_1 \cos \psi)^2]^{1/2}}{\eta_1 \sin \psi + \eta_2[1 - (v_2/v_1 \cos \psi)^2]^{1/2}} \qquad (2\text{-}5\text{-}13c)$$

Similarly, the vertical transmittance of a good conductor for electric fields as shown in Eq. (2-3-37) is given by

$$T_v = \frac{2\eta_2[1 - (v_2/v_1 \cos \psi)^2]^{1/2}}{\eta_2[1 - (v_2/v_1 \cos \psi)^2]^{1/2} + \eta_1 \sin \psi} \qquad (2\text{-}5\text{-}13d)$$

Horizontal polarization.   The reflectivity of a good conductor for electric fields in horizontal polarization as shown in Eq. (2-3-45) is simplified in terms of ψ as

$$\Gamma_h = \frac{\eta_2 \sin \psi[1 - (v_2/v_1 \cos \psi)^2]^{-1/2} - \eta_1}{\eta_2 \sin \psi[1 - (v_2/v_1 \cos \psi)^2]^{-1/2} + \eta_1} \qquad (2\text{-}5\text{-}13e)$$

Similarly, the transmittance of a good conductor for electric fields in horizontal polarization as shown in Eq. (2-3-46) can be expressed as

$$T_h = \frac{2\eta_2[1 - (v_2/v_1 \cos \psi)^2]^{-1/2}}{\eta_2 \sin \psi[1 - (v_2/v_1 \cos \psi)^2]^{-1/2} + \eta_1} \qquad (2\text{-}5\text{-}13f)$$

In Fig. 2-3-4 it is assumed that medium 1 is free space or air and that medium 2 is copper; then

$$\eta_1 = 377 \ \Omega \qquad\qquad \eta_2 = (1 + j) \sqrt{\frac{\omega\mu}{2\sigma}}$$

$$v_1 = 3 \times 10^8 \text{ m/s} \qquad v_2 = \omega\delta = \sqrt{\frac{2\omega}{\mu\sigma}}$$

The conductivity σ of copper is 5.8×10^7 mhos/m and its relative permeability is unity. The magnitudes of reflectivity of copper for vertical and horizontal polarizations are computed by Eqs. (2-5-13c) and (2-5-13e) against the grazing angle ψ of 0 to 90° at a frequency range of 0.1 to 40 GHz. This result indicates that copper is a perfect reflector for electromagnetic waves.

2-5-2 Plane Wave in Poor Conductor

Some conducting materials with low conductivity normally cannot be considered either good conductors or good dielectrics. Seawater is a good example. It has a conductivity of 4 mhos/m and a dielectric constant of 20. At some low frequencies the

conduction current is greater than the displacement current, whereas at some high frequencies the reverse is true.

In general, the propagation constant and intrinsic impedance for a poor conductor are given by

$$\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} \qquad (2\text{-}5\text{-}14)$$

$$\eta = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}} \qquad (2\text{-}5\text{-}15)$$

2-5-3 Plane Wave in Lossy Dielectric

All dielectric materials have some conductivity, but the conductivity is very small ($\sigma \ll \omega\epsilon$). When the conductivity cannot be neglected, the electric and magnetic fields in the dielectric are no longer in time phase. This fact can be seen from the intrinsic impedance of the dielectric as

$$\eta = \sqrt{\frac{j\omega\mu}{\sigma + j\omega\epsilon}} = \sqrt{\frac{\mu}{\epsilon}}\left(1 - j\frac{\sigma}{\omega\epsilon}\right)^{-1/2} \qquad (2\text{-}5\text{-}16)$$

The term $\sigma/(\omega\epsilon)$ is referred to as the *loss tangent* and is defined by

$$\tan\theta = \frac{\sigma}{\omega\epsilon} \qquad (2\text{-}5\text{-}17)$$

This relationship is the result of displacement current density leading conduction current density by 90°, just as the current flowing through a capacitor leads the current through a resistor in parallel with it by 90° in an ordinary electric circuit. This phase relationship is shown in Fig. 2-5-1.
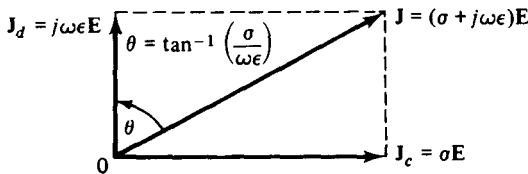
Figure 2-5-1 Loss tangent for lossy dielectric.

If the loss tangent is very small—that is, $\sigma/(\omega\epsilon) \ll 1$—the propagation constant and intrinsic impedance can be calculated approximately by a binomial expansion. Since

$$\gamma = j\omega\sqrt{\mu\epsilon}\sqrt{1 - j\frac{\sigma}{\omega\epsilon}} \qquad (2\text{-}5\text{-}18)$$

then

$$\gamma = j\omega\sqrt{\mu\epsilon}\left[1 - j\frac{\sigma}{2\omega\epsilon} + \frac{1}{8}\left(\frac{\sigma}{\omega\epsilon}\right)^2 + \cdots\right] \qquad (2\text{-}5\text{-}19)$$

where $\epsilon_r = \dfrac{\epsilon}{\epsilon_0}$

$$x = \frac{\sigma}{\omega\epsilon_0} = \frac{18\sigma}{f_{\text{GHz}}}$$

$\psi = 90° - \theta_i$ is called the *pseudo-Brewster angle*

Equation (2-5-25) is applicable only for the tangential components of incident and reflected fields that are in the same directions as shown in Fig. 2-3-4. For total reflection ($\Gamma_v = 1$) we set $\theta_t = 90°$ in Eq. (2-3-25), and the incident angle is given by

$$\theta_i = \theta_c = \arcsin \sqrt{\frac{\mu_2\epsilon_2}{\mu_1\epsilon_1}} \qquad (2\text{-}5\text{-}24c)$$

The angle specified by Eq. (2-5-24c) is called the *critical incident angle* for total reflection. A wave incident on the boundary at an angle equal to or greater than the critical angle will be totally reflected. There is a real critical angle only if $\mu_1\epsilon_1 > \mu_2\epsilon_2$, or, in the nonmagnetic material, if $\epsilon_1 > \epsilon_2$. Hence the total reflection occurs only if the wave propagates from a dense medium into a less dense medium. This is because the value of $\sin \theta_c$ must be equal to or less than unity.

For vertical polarization, the normal components of electric fields are usually used to determine the reflection coefficient. Therefore the reflectivity of a lossy dielectric in vertical polarization is given by

$$\Gamma_v = \frac{(\epsilon_r - jx) \sin \psi - \sqrt{(\epsilon_r - jx) - \cos^2 \psi}}{(\epsilon_r - jx) \sin \psi + \sqrt{(\epsilon_r - jx) - \cos^2 \psi}} \qquad (2\text{-}5\text{-}26)$$

Similarly, the vertical transmittance of a lossy dielectric for electric fields as shown in Eq. (2-3-37) is

$$T_v = \frac{2\sqrt{(\epsilon_r - jx) - \cos^2 \psi}}{\sqrt{(\epsilon_r - jx) - \cos^2 \psi} + (\epsilon_r - jx) \sin \psi} \qquad (2\text{-}5\text{-}27)$$

Horizontal polarization.   The reflectivity of a lossy dielectric for electric fields in horizontal polarization as shown in Eq. (2-3-45) becomes

$$\Gamma_h = \frac{\sin \psi - \sqrt{(\epsilon_r - jx) - \cos^2 \psi}}{\sin \psi + \sqrt{(\epsilon_r - jx) - \cos^2 \psi}} \qquad (2\text{-}5\text{-}28)$$

Similarly, the transmittance of a lossy dielectric for electric fields in horizontal polarization as shown in Eq. (2-3-46) is expressed as

$$T_h = \frac{2 \sin \psi}{\sin \psi + \sqrt{(\epsilon_r - jx) - \cos^2 \psi}} \qquad (2\text{-}5\text{-}29)$$

The reflections of electromagnetic waves by such lossy dielectric materials as seawater, dry sand, and concrete cement are often of concern to many electronics engi-

neers. The conductivities σ and relative dielectric constants ϵ_r of seawater, dry sand, and concrete cement are tabulated in Table 2-5-2.

Figures 2-5-2 to 2-5-5 show, respectively, the magnitudes of reflectivity of seawater, dry sand, and concrete cement for vertical and horizontal polarizations against the grazing angle ψ of 0 to 90° at a frequency range of 0.1 to 40 GHz [2].

TABLE 2-5-2 CONDUCTIVITIES AND RELATIVE DIELECTRIC
CONSTANTS OF SEAWATER, DRY SAND, AND CEMENT

	Seawater	Dry sand	Concrete cement
σ (mhos/m)	4	2×10^{-4}	2×10^{-5}
ϵ_r	20	4	3
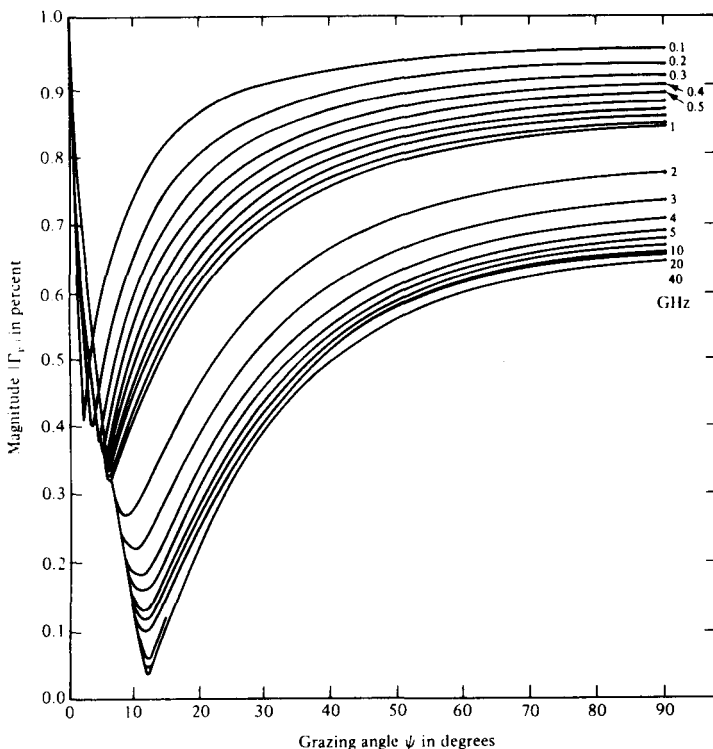
Figure 2-5-2 Magnitude of the reflectivity in vertical polarization versus grazing angle for seawater.
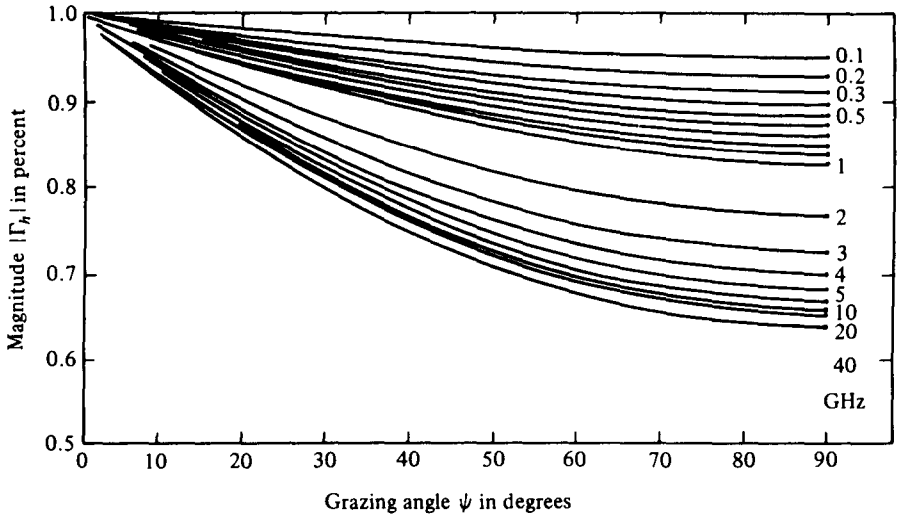
Figure 2-5-3 Magnitude of the reflectivity in horizontal polarization versus grazing angle for seawater.
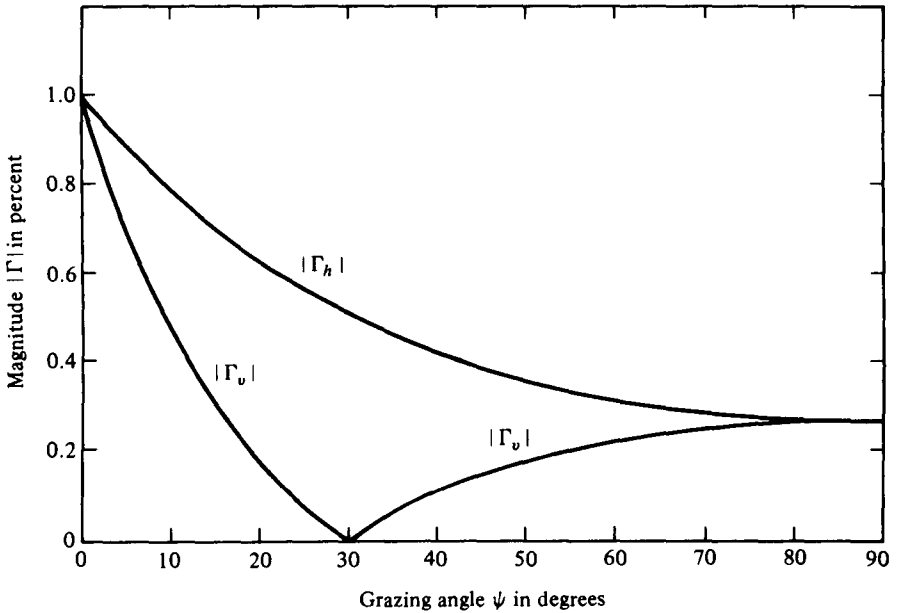
Figure 2-5-4 Magnitude of the reflectivity in vertical and horizontal polarizations versus grazing angle for dry sand.
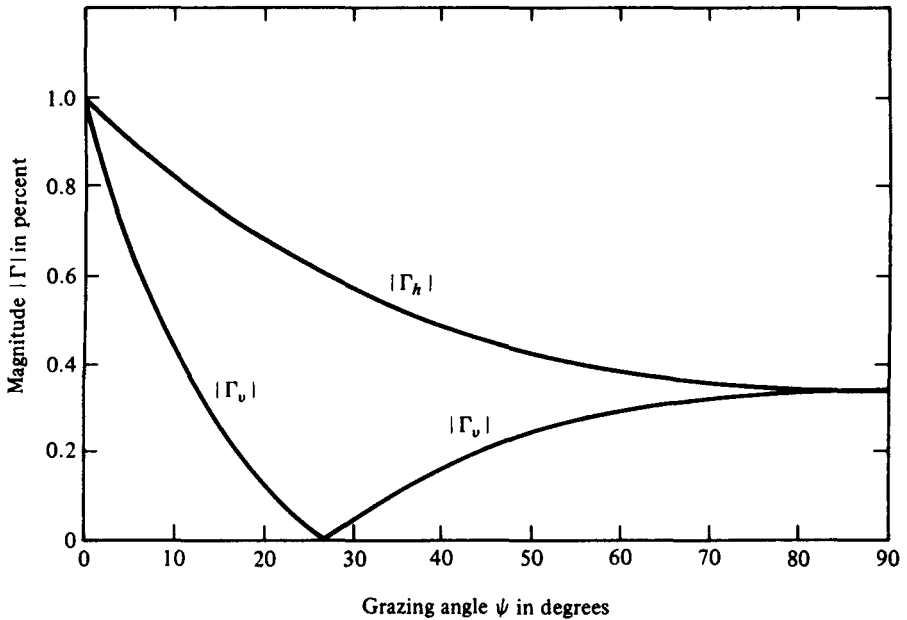
Figure 2-5-5 Magnitude of the reflectivity in vertical and horizontal polarizations versus grazing angle for concrete cement.

2-6 PLANE-WAVE PROPAGATION IN METALLIC-FILM COATING ON PLASTIC SUBSTRATE*

In certain engineering applications, it is often desirable to use a metallic-film-coated glass to attenuate optimum electromagnetic radiation at microwave frequencies and also to transmit as much light intensity as possible at visible-light frequencies. Generally the coated metallic film should have a high melting point, a high electrical conductivity, high adhesion to glass, high resistance to oxidation, and insensitivity to light and water, as well as the capability of dissipating some power for deicing, defogging, or maintaining certain temperature levels. The metallic-film coatings on a plastic substrate are used in such applications as windshields on airplanes or automobiles, medical equipment in hospitals, and on dome windows of space vehicles, missiles, and other military devices.

2-6-1 Surface Resistances of Metallic Films

Very thin metallic films have a much higher resistivity than a bulk metal because of electron scattering from the film surface. If the film thickness is very large compared to the electron mean-free-path, the resistivity is expected to be nearly the same as that of a bulk metal. When the film thickness is on the order of the electron mean-free-path, then the role of electron scattering becomes dominant. Fuchs [4] and Sondheimer [5] considered the general form of the solution of the Boltzmann equation for the case of a conducting film and found the film conductivity σ_f in terms of the bulk conductivity σ, the film thickness t, and the electron mean-free-path p:

$$\sigma_f = \frac{3t\sigma}{4p}\left[\ell n\left(\frac{p}{t}\right) + 0.4228\right] \qquad \text{for } t \ll p \qquad (2\text{-}6\text{-}1)$$

The surface resistance of conducting films is generally quoted in units of ohms per square because in the equation for resistance

$$R = \frac{\text{specific resistivity} \times \text{length}}{\text{thickness} \times \text{width}} = \frac{\rho\ell}{tw} \qquad (2\text{-}6\text{-}2)$$

when units of length ℓ and width w are chosen to have equal magnitude (that is, resulting in a square), the resistance R in ohms per square is independent of the dimensions of the square and equals

$$R_s = \frac{\rho_f}{t} = \frac{1}{t\sigma_f} \qquad \Omega/\text{square} \qquad (2\text{-}6\text{-}3)$$

According to the Fuchs–Sondheimer theory, the surface resistance of a metallic film is decreased as the thickness of the film is increased.

2-6-2 Optical Constants of Plastic Substrates and Metallic Films

The optical properties of materials are usually characterized by two constants, the refractive index n and the extinction index k. The refractive index is defined as the

TABLE 2-6-1 SUBSTRATE MATERIALS

Substrate material	Refractive index n
Corning Vycor	1.458
Crystal quartz	1.540
Fused silica	1.458
Plexiglass	1.490
Polycyclohexyl methacrylate	1.504
Polyester glass	1.500
Polymethyl methacrylate	1.491
Zinc crown glass	1.508

ratio of the phase velocities of light in vacuum and in a medium. The extinction index is related to the exponential decay of the wave as it passes through a medium. Most optical plastics are suitable as substrate materials for a dome window and for metallic-film applications. Table 2-6-1 lists the values of the refractive index n of several nonabsorbing plastic substrate materials in common use [6].

The measured values of the refractive index n and the extinction index k of thin metallic-film coatings deposited in a vacuum [6] are tabulated in Table 2-6-2.

TABLE 2-6-2 REFRACTIVE INDEX n AND EXTINCTION INDEX k OF THIN METALLIC FILMS

Wavelength (Å)	Copper film		Gold film		Silver film	
	n	k	n	k	n	k
2000			1.427	1.215	1.13	1.23
2200					1.32	1.29
2300					1.38	1.31
2400					1.37	1.33
2500					1.39	1.34
2600					1.45	1.35
2700					1.51	1.33
2800					1.57	1.27
2900					1.60	1.17
3000					1.67	0.96
3100					1.54	0.54
3200					1.07	0.32
3300					0.30	0.55
3400					0.16	1.14
3500					0.12	1.35
3600					0.09	1.52
3700					0.06	1.70
3800						
4000						
4500	0.870	2.200	1.400	1.880		
4920						
5000	0.880	2.420	0.840	1.840		
5460						
5500	0.756	2.462	0.331	2.324		
6000	0.186	2.980	0.200	2.897		
6500	0.142	3.570	0.142	3.374		
7000	0.150	4.049	0.131	3.842		
7500	0.157	4.463	0.140	4.266		
8000	0.170	4.840	0.149	4.654		
8500	0.182	5.222	0.157	4.993		
9000	0.190	5.569	0.166	5.335		
9500	0.197	5.900	0.174	5.691		
10,000	0.197	6.272	0.179	6.044		

2-6-3 Microwave Radiation Attenuation of
Metallic-Film Coating on Plastic Substrate

A conductor of high conductivity and low permeability has low intrinsic impedance. When a radio wave propagates from a medium of high intrinsic impedance into a medium of low intrinsic impedance, the reflection coefficient is high. From electromagnetic plane wave theory in the far field, high attenuation occurs in a medium made of material having high conductivity and low permeability. Good conductors, such as gold, silver, and copper, have high conductivity and are often used as the material for attenuating electromagnetic energy. Microwave radiation attenuation by a metallic-film coating on substrate consists of three parts [7]:

$$\text{Attenuation} = A + R + C \qquad \text{dB} \qquad (2\text{-}6\text{-}4)$$

where A = absorption or attenuation loss in decibels inside the metallic-film coating while the substrate is assumed to be nonabsorbing plastic glass

R = reflection loss in decibels from the multiple boundaries of a metallic-film coating on substrate

C = correction term in decibels required to account for multiple internal reflections when the absorption loss A is much less than 10 dB for electrically thin film

Figure 2-6-1 shows the absorption and reflection of a metallic-film coating on a plastic substrate.
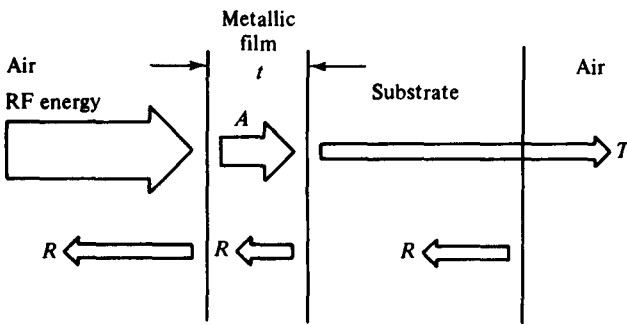
Figure 2-6-1 Absorption and reflection of film coating on plastic substrate.

Absorption loss A.   As described in Section 2-5-1, the propagation constant γ for a uniform plane wave in a good conducting material is given in Eq. (2-5-7) as

$$\gamma = \alpha + j\beta = (1 + j)\sqrt{\pi f \mu \sigma_f} \qquad \text{for } \sigma_f \gg \omega \epsilon \qquad (2\text{-}6\text{-}5)$$

If the plastic substrate is assumed to be a nonabsorbing material, the absorption loss A of the metallic-film coating on a substrate is related only to the thickness t of the coated film and the attenuation α as shown:

$$A = 20 \log_{10} e^{\alpha t} = 20(\alpha t) \log_{10} e = 20(0.4343)(\alpha t)$$

$$= 8.686 t \sqrt{\pi f \mu \sigma_f} \qquad \text{dB} \qquad (2\text{-}6\text{-}6)$$

where t = thickness of the film coating in meters

μ = permeability of the film in henrys per meter
f = frequency in hertz
σ_f = conductivity of the coated film in mhos per meter

Since the thickness of the coated film is very thin—for example, 100 angstroms at most ($\text{Å} = 10^{-10}$ m)—the absorption loss A is very small and can be ignored.

Reflection loss R.   The reflection loss R due to the multiple boundaries of the substrate glass coated with a metallic film can be analyzed by means of the energy-transmission theory (see Eq. 2-3-18 in Section 2-3-3), and it is expressed as

$$R = -20 \log \frac{2|\eta_f|}{|\eta_a + \eta_f|} - 20 \log \frac{2|\eta_g|}{|\eta_f + \eta_g|} - 20 \log \frac{2|\eta_a|}{|\eta_g + \eta_a|}$$

$$= 20 \log \frac{|\eta_a + \eta_f||\eta_f + \eta_g||\eta_g + \eta_a|}{8|\eta_f||\eta_g||\eta_a|} \qquad \text{dB} \qquad (2\text{-}6\text{-}7)$$

where η_f = intrinsic impedance of the coated metallic film
η_g = intrinsic impedance of the glass substrate
η_a = intrinsic impedance of air or free space = 377 Ω

The intrinsic impedance of a metallic film is given by Eq. (2-5-11) as

$$|\eta_f| = \left|(1 + j)\sqrt{\frac{\mu\omega}{2\sigma_f}}\right| = \sqrt{\frac{\mu\omega}{\sigma_f}} \qquad (2\text{-}6\text{-}8)$$

and the intrinsic impedance of a glass substrate is expressed in Eq. (2-5-16) as

$$\eta_g \approx \frac{\eta_a}{\sqrt{\epsilon_r}} = \frac{377}{\sqrt{3.78}} = 194 \ \Omega \qquad \text{for } \sigma_g \ll \omega\epsilon_q \qquad (2\text{-}6\text{-}9)$$

where σ_g = about 10^{-12} mho/m is the conductivity of the glass substrate
$\epsilon_g = 4.77 \times 10^{-11}$ F/m is the permittivity of the glass substrate
$\epsilon_r = 3.78$ is the relative permittivity of the glass substrate

Substituting the values of the intrinsic impedances η_f, η_g, and η_a in Eq. (2-6-7) yields the reflection loss as [12]

$$R \approx 20 \log \left[28.33 \sqrt{\frac{\sigma_f}{\mu f}}\right] = 88 + 10 \log \left(\frac{\sigma_f}{f}\right) \qquad \text{dB} \qquad (2\text{-}6\text{-}10)$$

Correction term C.   For very electrically thin film, the value of the absorption loss A is much less than 10 dB and the correction term is given by [8]

$$C = 20 \log |1 - \rho \, 10^{-A/10}(\cos\theta - j\sin\theta)| \qquad (2\text{-}6\text{-}11)$$

where $\rho = \left(\dfrac{\eta_f - \eta_a}{\eta_f + \eta_a}\right)^2 \approx 1 \qquad$ for $\eta_a \gg \eta_f$

$\theta = 3.54t \sqrt{f\mu\sigma_f}$

Over the frequency range of 100 MHz to 40 GHz, the angle θ is much smaller than 1° so that $\cos\theta \approx 1$ and $\sin\theta \approx \theta$. Thus the correction term of Eq. (2-6-11)

can be simplified to

$$C \approx 20 \log \left[3.54t \sqrt{f\mu\sigma_f} \right] = -48 + 20 \log \left[t \sqrt{f\sigma_f} \right] \quad \text{dB} \quad (2\text{-}6\text{-}12)$$

Finally, the total microwave radiation attenuation by a metallic-film coating on a glass substrate, defined in Eq. (2-6-4) in the far field, becomes

$$\text{Attenuation} = 40 - 20 \log (R_s) \quad \text{dB} \quad (2\text{-}6\text{-}13)$$

It is interesting to note that the microwave radiation attenuation due to the coated metallic film on a glass substrate in the far field is independent of frequency and is related only to the surface resistance of the coated metallic film [12].

2-6-4 Light Transmittance of Metallic-Film Coating on Plastic Substrate

The complex refractive index of an optical material is given by [6] as

$$N = n - jk \quad (2\text{-}6\text{-}14)$$

It is assumed that light in air is normally incident on a thin absorbing film N_1 of thickness t_1 and that it is transmitted through an absorbing substrate of complex refractive index N_2 and then emerges into air. The incidence and the emergence media are dielectrics of refractive index n_0. The reflection loss between the substrate and the air is small and, for convenience, is taken as zero. Figure 2-6-2 shows light transmittance, reflection, and absorption through a thin absorbing metallic film and a plastic substrate.
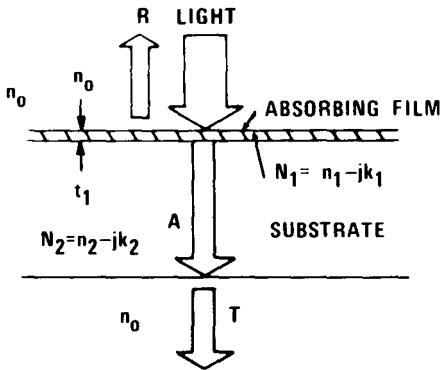
Figure 2-6-2 Light transmittance, reflection, and absorption through a thin metallic film coated on plastic substrate.

Using the multireflection and transmission theory, the reflection loss is expressed by

$$R = \frac{a_1 e^{\alpha} + a_2 e^{-\alpha} + a_3 \cos v + a_4 \sin v}{b_1 e^{\alpha} + b_2 e^{-\alpha} + b_3 \cos v + b_4 \sin v} \quad (2\text{-}6\text{-}15)$$

where $a_1 = [(n_0 - n_1)^2 + k_1^2][(n_1 + n_2)^2 + (k_1 + k_2)^2]$
$a_2 = [(n_0 + n_1)^2 + k_1^2][(n_1 - n_2)^2 + (k_1 - k_2)^2]$
$a_3 = 2\{[n_0^2 - (n_1^2 + k_1^2)][(n_1^2 + k_1^2) - (n_2^2 + k_2^2)] + 4n_0 k_1 (n_1 k_2 - n_2 k_1)\}$
$a_4 = 4\{[n_0^2 - (n_1^2 + k_1^2)](n_1 k_2 - n_2 k_1) - n_0 k_1 [(n_1^2 + k_1^2) - (n_2^2 + k_2^2)]\}$

$$\alpha = \frac{4\pi k_1 t_1}{\lambda_0}$$

$\lambda_0 = \dfrac{c}{f}$ is the wavelength in a vacuum

$c = 3 \times 10^8$ m/s is the velocity of light in a vacuum; f is the frequency in hertz

$$\nu = \frac{4\pi n_1 t_1}{\lambda_0}$$

$b_1 = [(n_0 + n_1)^2 + k_1^2][(n_1 + n_2)^2 + (k_1 + k_2)^2]$

$b_2 = [(n_0 - n_1)^2 + k_1^2][(n_1 - n_2)^2 + (k_1 - k_2)^2]$

$b_3 = 2\{[n_0^2 - (n_1^2 + k_1^2)][(n_1^2 + k_1^2) - (n_2^2 + k_2^2)] - 4n_0 k_1(n_1 k_2 - n_2 k_1)\}$

$b_4 = 4\{[n_0^2 - (n_1^2 + k_1^2)](n_1 k_2 - n_2 k_1) + n_0 k_1[(n_1^2 + k_1^2) - (n_2^2 + k_2^2)]\}$

Transmittance T is given by [6] as

$$T = \frac{16 n_0 n_2 (n_1^2 + k_1^2)}{b_1 e^{\alpha} + b_2 e^{-\alpha} + b_3 \cos \nu + b_4 \sin \nu} \tag{2-6-16}$$

Absorption loss A is given by

$$A = 1 - R - T \tag{2-6-17}$$

and the total attenuation loss L is

$$L = A + R \tag{2-6-18}$$

When the concave surface of a plastic dome is uniformly coated with an electromagnetic interference shield of metallic film, however, the light is normally incident on the plastic substrate N_2, transmits through the thin metallic film N_1, and then emerges into the air n_0. From the electromagnetic theory of luminous transmission in transparent media, the light transmittance is the same regardless of whether light is normally incident on the substrate medium N_2 or on the absorbing film N_1. Thus the total attenuation loss is the same in both cases.

2-6-5 Plane Wave in Gold-Film Coating on Plastic Glass

Metallic-film coatings on plastic glasses have many engineering applications [9]. A gold film, for example, is coated on the concave surface of a plastic-glass dome so that an optimum amount of microwave radiation is attenuated by the gold film while, at the same time, a sufficient light intensity is transmitted through the gold film.

Surface resistance.   At room temperature the properties of bulk gold are

Conductivity: $\sigma = 4.10 \times 10^7$ mhos/m
Resistivity: $\rho = 2.44 \times 10^{-8}$ Ω-m
Electron mean-free-path: $p = 570$ Å

It is assumed that the thickness t of the gold film varies from 10 to 100 Å. Its surface

TABLE 2-6-3 SURFACE RESISTANCE OF GOLD FILM

Thickness t (Å)	Conductivity σ_f (mho/m $\times 10^7$)	Resistivity ρ_f (Ω-m $\times 10^{-7}$)	Surface resistance R_s (Ω/square)
100	1.17	0.86	8.60
90	1.11	0.90	10.00
80	1.03	0.97	12.13
70	0.96	1.04	14.86
60	0.85	1.17	19.50
50	0.77	1.30	26.00
40	0.68	1.48	37.00
30	0.54	1.86	62.00
20	0.42	2.41	120.00
10	0.22	4.48	448.00

resistances are computed by using Eqs. (2-6-1) and (2-6-3) and are tabulated in Table 2-6-3.

Figure 2-6-3 shows surface resistances of gold film in ohms per square against the thicknesses of gold film from 10 to 100 Å. According to the Fuchs–Sondheimer theory, gold films have a typical surface resistance at about 10 to 30 Ω/square for a thickness of about 90 to 45 Å. The surface resistance is decreased as the thickness of the gold film is increased.
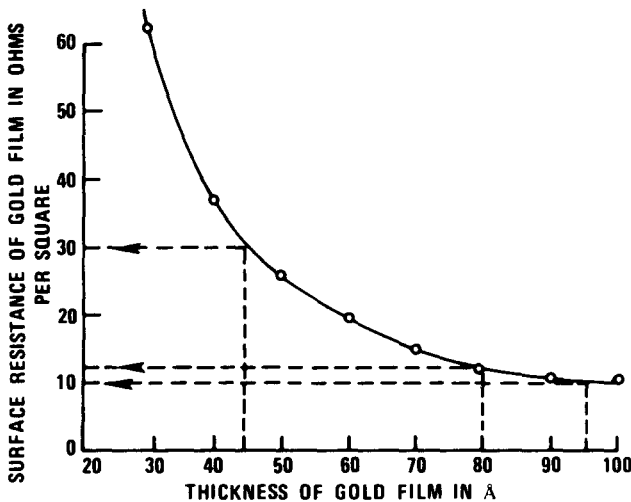
Figure 2-6-3 Surface resistance of gold film versus thickness of gold film.

Microwave radiation attenuation. Substituting the values of the surface resistances for gold films in Eq. (2-6-13) yields the microwave radiation attenuation in decibels by the gold-film coating on a plastic glass. Figure 2-6-4 shows graphically the microwave radiation attenuation versus the surface resistance of the gold-film coating. For a coated gold film having a surface resistance of 12 Ω/square, the
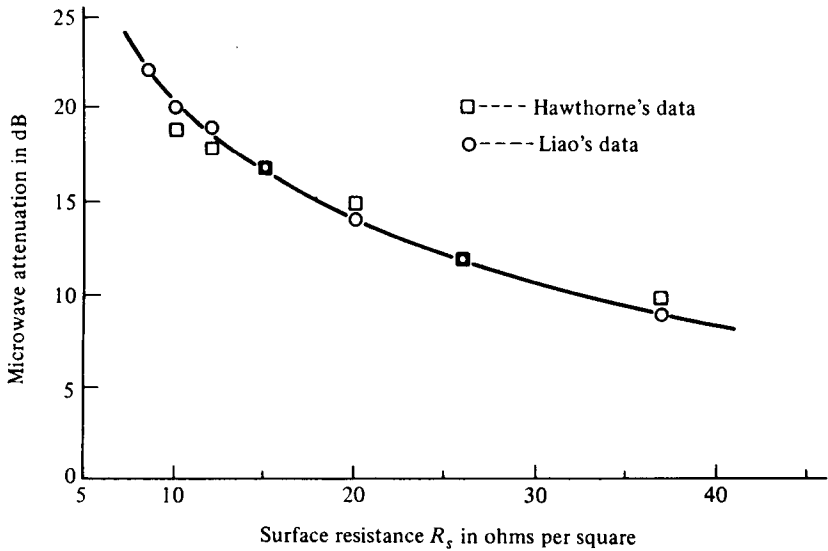
Figure 2-6-4 Microwave radiation attenuation versus surface resistance of gold film.

microwave radiation attenuation is about 19 dB. The data agree with Hawthorne's conclusion [10,12].

Light transmittance. For the visible-light region, the values of the refractive index n and the extinction index k of a gold-film coating on a plastic glass deposited in a vacuum are taken from Table 2-6-2. The refractive index n_0 of air or vacuum is unity. The refractive index n_2 of the nonabsorbing plastic glass is taken as 1.50. Light transmittance T and light reflection loss R of a gold-film coating on a plastic glass are computed by using Eqs. (2-6-16) and (2-6-15), respectively. Then from the values of T and R absorption loss A and total attenuation L are calculated. The results are presented graphically in Fig. 2-6-5. It can be seen that for a light transmittance of 80% the thickness of the gold-film coating is about 80 Å. When the absorption loss in the substrate material is considered, however, the light transmittance may be a little less than 80%.

Optimum condition. The surface resistance of a metallic film decreases as the thickness of the film coating increases. However, the luminous transmittance is decreased as the surface resistance of the metallic film is decreased. This relationship for the visible-light region is shown in Fig. 2-6-6.

Figure 2-6-7 illustrates the relationship of light transmittance versus wavelength for a given surface resistance of gold film. If a power dissipation of 5 W/square is allowed for deicing and defogging or keeping warm by the gold-film coating on a plastic substrate and if the effective area of the coated film is 13 square inches in a missile, the surface resistance of the coated film must be 12 Ω/square.
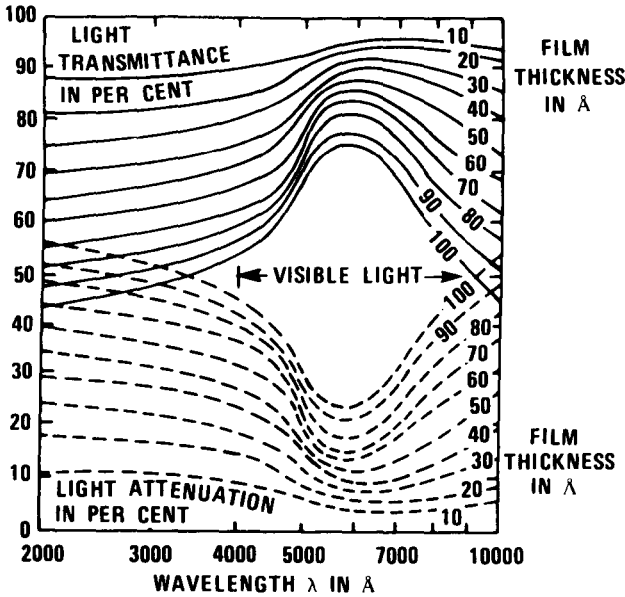
Figure 2-6-5 Light transmittance T and light attenuation loss L versus wavelength λ with film thickness t as parameter for gold film.
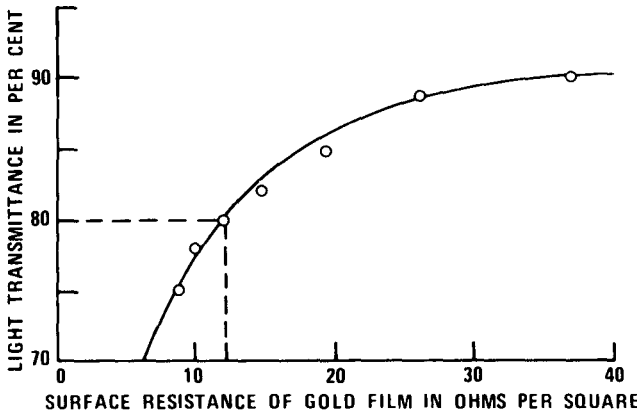
- LIGHT TRANSMITTANCE DECREASES, AS SURFACE RESISTANCE DECREASES
- FOR 80 PER CENT OF LIGHT TRANSMITTANCE, SURFACE RESISTANCE IS ABOUT 12 OHMS PER SQUARE

Figure 2-6-6 Light transmittance versus surface resistance of gold film.

The power dissipation can be expressed as

$$P = \frac{V^2}{R} = \frac{(28)^2}{12 \times 13} = 5 \text{ W/square}$$

in which the voltage applied to the film-coating terminations is 28 V. The optimum condition occurs at 19 dB of microwave radiation attenuation and 80% of light transmittance.

Figure 2-6-7   Light transmittance versus wavelength with surface resistance R_s as parameter for gold film.

Example 2-6-5:   Calculation of a Gold-Film Coating

A gold film of 80 Å is coated on a plastic substrate with a refractive index of 1.50. Determine:

    a. The gold-film surface resistance in ohms per square
    b. The microwave attenuation in decibels
    c. The light transmittance T for red light of 0.69 μm
    d. The light reflection loss R at $\lambda = 0.69$ μm

Solution

    a. From Eq. (2-6-1) the gold-film conductivity is

$$\sigma_f = \frac{3t\sigma}{4p}\left[\ln\left(\frac{p}{t}\right) + 0.4228\right]$$

$$= \frac{3 \times 80 \times 10^{-10} \times 4.1 \times 10^7}{4 \times 570 \times 10^{-10}}\left[\ln\left(\frac{570}{80}\right) + 0.4228\right]$$

$$= 1.03 \times 10^7 \text{ mhos/m}$$

    From Eq. (2-6-3) the gold-film surface resistance is

$$R_s = \frac{1}{t\sigma_f} = \frac{1}{80 \times 10^{-10} \times 1.03 \times 10^7}$$

$$= 12.12 \ \Omega/\text{squares}$$

b. From Eq. (2-6-13) the microwave attenuation is

$$\text{Attenuation} = 40 - 20 \log (R_s)$$

$$= 40 - 20 \log (12.12)$$

$$= 18 \ \text{dB}$$

c. From Fig. 2-6-5 the light transmittance T is estimated to be 75%.

d. From the same figure the light reflection loss R is about 25%.

2-6-6 Plane Wave in Silver-Film or Copper-Film Coating on Plastic Substrate*

Silver-film or copper-film coating on a plastic substrate has many uses in engineering [11]. The surface resistance, microwave radiation attenuation, light transmittance, and optimum condition of both silver-film coating and copper-film coating can be described in the same way as for fold-film coating.

Surface resistance.   At room temperature the properties of bulk silver and bulk copper are

Silver

Conductivity:	$\sigma = 6.170 \times 10^7$ mhos/m
Resistivity:	$\rho = 1.620 \times 10^{-8}$ Ω-m
Electron mean-free-path:	$p = 570$ Å

Copper

Conductivity:	$\sigma = 5.800 \times 10^7$ mhos/m
Resistivity:	$\rho = 1.724 \times 10^{-8}$ Ω-m
Electron mean-free-path:	$p = 420$ Å

It is assumed that the thickness t of the silver and copper films varies from 10 to 100 Å. The surface resistances of silver and copper films are computed by using Eqs. (2-6-1) and (2-6-3) and are tabulated in Tables 2-6-4 and 2-6-5, respectively.

Figure 2-6-8 plots the surface resistances of silver and copper films, respectively, in ohms per square versus the thickness of the silver and copper films from 10 to 100 Å.

TABLE 2-6-4   SURFACE RESISTANCE R_s OF SILVER FILM

Thickness t (Å)	Conductivity σ_f (mho/m $\times$ 10^7)	Resistivity ρ_f (Ω-m $\times$ 10^7)	Surface resistance R_s (Ω/square)
100	1.78	0.571	5.71
90	1.66	0.602	6.69
80	1.55	0.645	8.06
70	1.44	0.695	9.93
60	1.31	0.763	12.72
50	1.16	0.862	17.24
40	0.99	1.010	25.25
30	0.81	1.230	41.00
20	0.61	1.640	82.00
10	0.36	2.760	276.00

TABLE 2-6-5   SURFACE RESISTANCE R_s OF COPPER FILM

Thickness t (Å)	Conductivity σ_f (mho/m $\times$ 10^7)	Resistivity ρ_f (Ω-m $\times$ 10^{-7})	Surface resistance R_s (Ω/square)
100	1.93	0.52	5.20
90	1.85	0.54	6.00
80	1.73	0.58	7.25
70	1.62	0.62	8.86
60	1.47	0.68	11.33
50	1.33	0.75	15.00
40	1.17	0.86	21.50
30	0.95	1.05	35.00
20	0.73	1.37	68.50
10	0.43	2.31	231.00
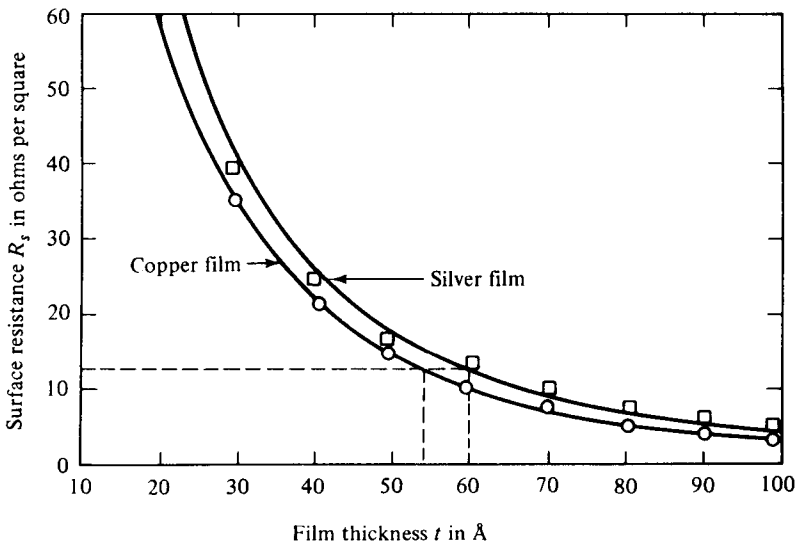
Figure 2-6-8   Surface resistance of silver and copper film versus thickness of film.

RF radiation attenuation. Substitution of the values of the surface resistances of silver or copper films in Eq. (2-6-13) yields the microwave radiation attenuation in decibels by the silver-film or copper-film coating on a plastic substrate. Figure 2-6-9 shows the microwave radiation attenuation versus the surface resistance of silver-film or copper-film coating, respectively.
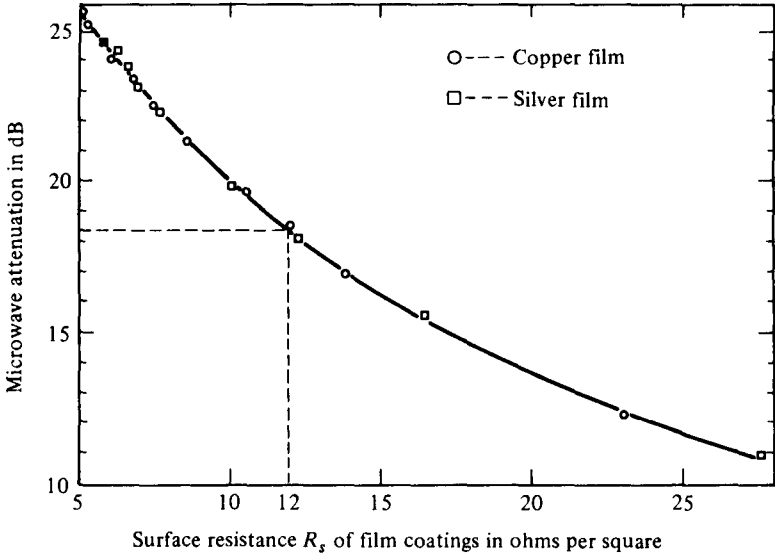
Figure 2-6-9 Microwave radiation attenuation versus surface resistance of silver and copper film.

Light transmittance. Light transmittance T and light reflection loss R of silver-film and copper-film coatings are computed by using Eqs. (2-6-16) and 2-6-15), respectively. The values of the refractive index n and the extinction index k of the silver-film and copper-film coatings deposited in a vacuum for the light-frequency range are taken from Table 2-6-2. The refractive index n_0 of air or vacuum is unity. The refractive index n_2 of the nonabsorbing plastic glass is taken as 1.5. From the values of light transmittance T and light reflection loss R, absorption loss A and total attenuation L are calculated. The results are illustrated in Figs. 2-6-10 and 2-6-11 for silver-film and copper-film coatings, respectively.

Optimum Condition. The light transmittance is increased as the surface resistance is increased. The relationship is illustrated in Fig. 2-6-12 for silver film and copper film, respectively. The optimum condition occurs at 18 dB of microwave radiation attenuation and 94% of light transmittance with a surface resistance of about 12 Ω/square.
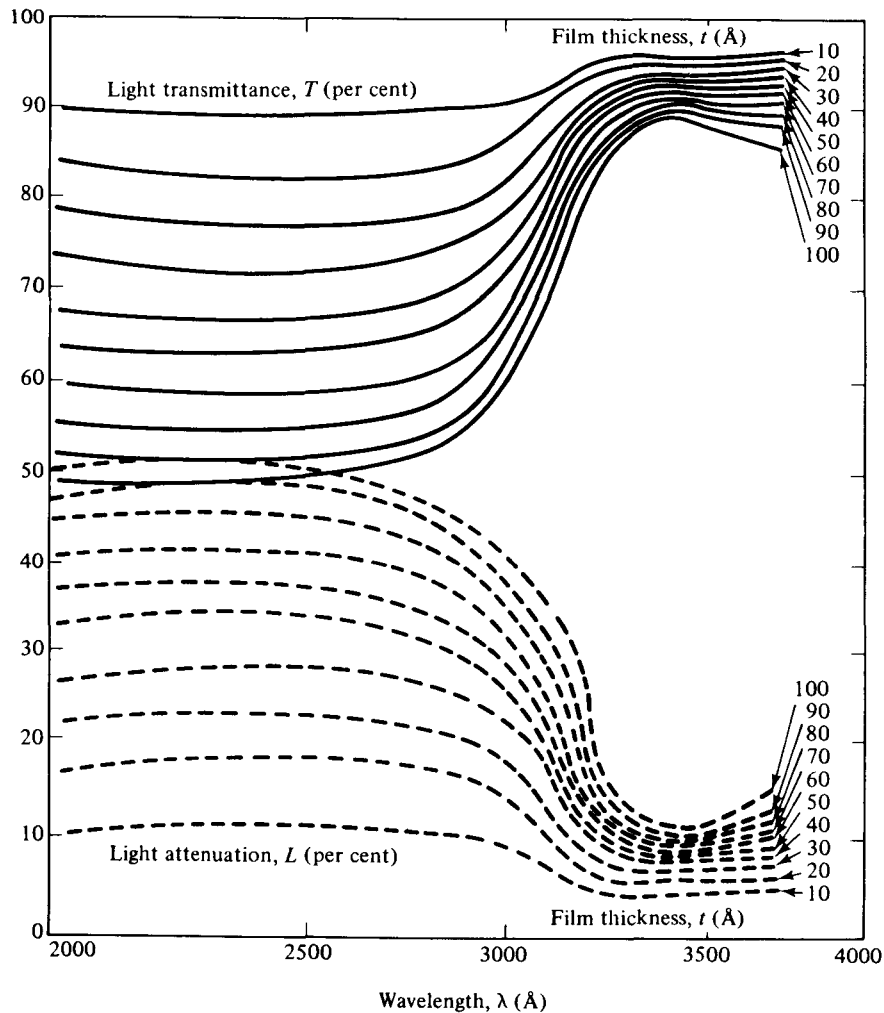
Figure 2-6-10 Light transmittance T and light attenuation loss L of silver film versus wavelength λ with film thickness t as parameter.

Figure 2-6-11 Light transmittance T and light attenuation loss L of copper film versus wavelength with film thickness t as parameter.
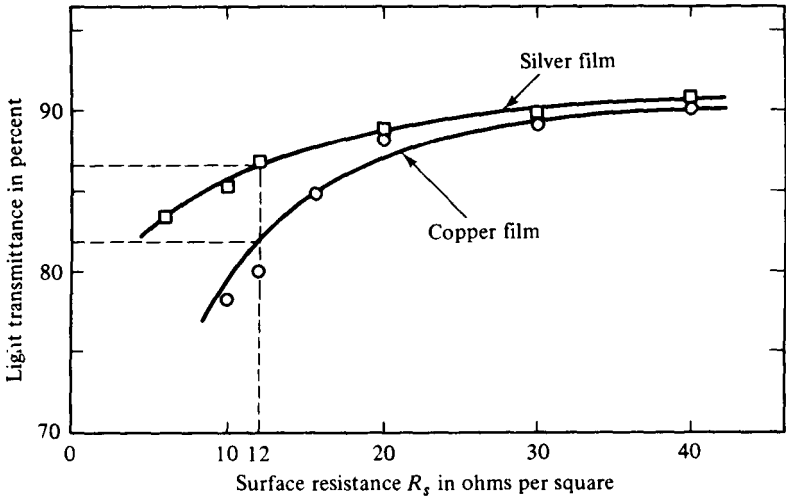
Figure 2-6-12 Light transmittance versus surface resistance of silver and copper films.

Example 2-6-6:   Computation of a Copper-Film Coating

A copper film of 60 Å is coated on a plastic substrate with a refractive index of 1.50. Compute:

a. The copper-film surface resistance in ohms per square
b. The microwave attenuation in decibels
c. The light transmittance T for red light of 0.69 μm
d. The light reflection loss R at $\lambda = 0.69$ μm

Solution

a. From Eq. (2-6-1) the copper-film conductivity is

$$\sigma_f = \frac{3t\sigma}{4p}\left[\ln\left(\frac{p}{t}\right) + 0.4228\right]$$

$$= \frac{3 \times 60 \times 10^{-10} \times 5.8 \times 10^7}{4 \times 420 \times 10^{-10}}\left[\ln\left(\frac{420}{60}\right) + 0.4228\right]$$

$$= 1.47 \times 10^7 \text{ mhos/m}$$

From Eq. (2-6-3) the copper-film surface resistance is

$$R_s = \frac{1}{t\sigma_f} = \frac{1}{60 \times 10^{-10} \times 1.47 \times 10^7}$$

$$= 11.34 \text{ }\Omega/\text{square}$$

b. From Eq. (2-6-13) the microwave attenuation is

$$\text{Attenuation} = 40 - 20 \log (R_s)$$

$$= 40 - 20 \log (11.34)$$
$$= 19 \text{ dB}$$

c. The light transmittance T is estimated from Fig. 2-6-11 to be 82%

d. From the same figure the light reflection loss R is about 18%.

REFERENCES

[1] ADLER, R. B., et al., *Electromagnetic Energy Transmission and Radiation*. P. 8. MIT Press, Cambridge, Mass., 1969.

[2] LIAO, S. Y., Reflectivities of electromagnetic waves by seawater, dry sand, concrete cement, and dry ground. A report for the Naval Weapons Center, Department of the Navy, China Lake, Calif., August 1976.

[3] LIAO, S. Y., Light transmittance and microwave attenuation of a gold-film coating on a plastic substrate. *IEEE Trans. on Microwave Theory and Techniques,* **MTT-23,** No. 10, October 1975.

[4] FUCHS, K., The conductivity of thin metallic films according to the electron theory of metals. *Proc. Camb. Phil. Soc.,* **30,** 100 (1938).

[5] SONDHEIMER, E. H., The mean-free-path of electrons in metals. *Advances in Physics,* **1,** 1 (1952).

[6] *American Institute of Physics Handbook,* Sec. 6–12, 6–119 to 6–121, and 6–138. McGraw-Hill Book Company, New York, 1972.

[7] SCHULZ, R. B., et al., Shielding theory and practice. *Proc. 9th Tri-Service Conf. on Electromagnetic Compatibility,* October 1963.

[8] VASAKA, C. S., Problems in shielding electrical and electronic equipment. U.S. Naval Air Development Center. Johnsville, Pa., *Rep. No. NACD-EL-N5507,* June 1955.

[9] LIAO, S. Y., Design of a gold film on a glass substrate for maximum light transmittance and RF shielding effectiveness. *IEEE Electromagnetic Compatibility Symposium Records,* San Antonio, Texas, October 1975.

[10] HAWTHORNE, E. I., Electromagnetic shielding with transparent coated glass. *Proc. IRE.,* **42,** 548–553, March 1954.

[11] LIAO, S. Y., Light transmittance and RF shielding effectiveness of a gold film on a glass substrate. *IEEE Trans. on Electromagnetic Compatibility,* **EMC-17,** No. 4, November 1975.

[12] LIAO, S. Y., RF shielding effectiveness and light transmittance of copper or silver coating on plastic substrate. *IEEE Trans. on Electromagnetic Compatibility,* **EMC-18,** No.4, November 1976.

SUGGESTED READINGS

COLLIN, R. E., *Foundations for Microwave Engineering,* Chapter 2. McGraw-Hill Book Company, New York, 1966.

HAYT, W. H., *Engineering Electromagnetics,* 4th ed., Chapters 10 and 11. McGraw-Hill Book Company, New York, 1981.

JORDAN, E. and K. G. BALMAIN, *Electromagnetic Waves and Radiating Systems,* 2nd ed., Chapters 4, 5, and 6. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1968.

SEELY, S., and A. D. POULARIKAS, *Electromagnetics, Classical and Modern Theory and Applications.* Marcel Dekker, Inc., New York, 1979.

PROBLEMS

2-1. In a certain homogeneous medium the group velocity as measured by the propagation time of a pulse is found to be proportional to the square root of the frequency ($v_g = \sqrt{A\omega}$ where A is a constant) over a particular frequency range. It is assumed the medium is a nonmagnetic insulator.
   a. Determine the relationship between the phase and group velocities.
   b. Derive an expression for the relative dielectric constant of this medium.

2-2. Show that $\partial/\partial z = \beta_g$ and $\partial/\partial t = j\omega$ for a sinusoidal wave propagating in the z direction.

2-3. The electric field of a plane wave propagating in free space is given in complex notation by

$$\mathbf{E} = 10^{-4}e^{j(\omega t + 20\pi z)}\mathbf{U}_x + 10^{-4}e^{j(\omega t + \pi/2 + 20\pi z)}\mathbf{U}_y$$

Where $\mathbf{U}_x$ and $\mathbf{U}_y$ are unit vectors in the x and y directions of a right-hand coordinate system.
   a. In which direction is the wave propagating?
   b. Find the frequency of the propagating signal.
   c. Determine the type of wave polarization (linear, circular, or elliptical).
   d. Express the magnetic field intensity $\mathbf{H}$ of the propagating wave.
   e. Calculate the average power flow per square meter in the direction of the propagation.

2-4. Determine the permittivity of a slab of dielectric material that will reflect 20% of the energy in a plane wave. The wave is striking normally to and propagating through the slab. Neglect the reflections from the back face of the slab.

2-5. The reflection and refraction of microwave propagating in the ionosphere are determined by the electron density in the ionosphere. If the electron density is assumed to be 10^{14} electrons per cubic meter, detemine the critical frequency for vertical incidence so that the signal wave will be reflected back to the earth.

2-6. The conductivity σ of copper is 5.8×10^7 mhos/m and its relative permeability is unity. Calculate the magnitudes of reflectivity of copper for vertical and horizontal polarizations against the grazing angle ψ of 0 to 90° at a frequency range of 1 to 40 GHz. The increment of the angle ψ is 10° each step and the increment of the frequency is 10 GHz each step.

2-7. At the Brewster angle there is no reflected wave when the incident wave is vertically polarized. If the incident wave is not entirely vertically polarized there will be some reflection but the reflected wave will be entirely of horizontal polarization. Verify Eq. (2-5-24B) for the Brewster angle in terms of the relative dielectrics.

2-8. Determine the pseudo-Brewster angle ψ in terms of v_1, v_2, η_1, and η_2 for a good conductor. [*Hint*: Start from Eq. (2-5-13b).]

2-9. Calculate the pseudo-Brewster angles for seawater, dry sand, concrete cement, and dry ground.

2-10. Determine the pseudo-Brewster angle ψ in terms of ϵ_r and x for a lossy dielectric. [*Hint:* Start from Eq. (2-5-26).]

2-11. Bulk gold has a conductivity of 4.1×10^7 mhos/m, a resistivity of 2.44×10^{-8} Ω-m, and an electron mean-free-path of 570 Å. Calculate the surface conductivity, surface resistivity, and surface resistance of gold film for thicknesses of 10 to 100 Å with an increment of 10 Å for each step.

2-12. Silver has a conductivity of 0.617×10^8 mhos/m, a resistivity of 1.620×10^{-8} Ω-m, and an electron mean-free-path of 570 Å. Repeat Problem 2-11 for silver film.

2-13. Seawater has a conductivity of 4 mhos/m and a relative dielectric constant of 20 at a frequency of 4 GHz. Compute:
a. The intrinsic impedance
b. The propagation constant
c. The phase velocity

2-14. Repeat Problem 2-13 for dry sand ($\sigma = 2 \times 10^{-4}$ mho/m and $\epsilon_r = 4$) and copper ($\sigma = 5.8 \times 10^7$ mhos/m).

2-15. A uniform plane wave is incident normally from air onto the surface of seawater. The electric intensity of the incident wave is 100×10^{-3} V/m at a frequency of 5 GHz in the vertical polarization. Calculate:
a. The electric intensity of the reflected wave
b. The electric intensity of the transmitted wave

2-16. Repeat Problem 2-15 for an angle of incidence of 30°.

2-17. Dry ground has a conductivity of 5×10^{-4} mhos/m and a relative dielectric constant of 10 at a frequency of 500 MHz. Compute:
a. The intrinsic impedance
b. The propagation constant
c. The phase velocity

2-18. Copper has a conductivity of 5.8×10^7 mhos/m and is considered an ideal material for shielding. A shield is made of copper with a thickness of 1 mm. If a uniform plane wave is normally incident upon the copper shield, compute the absorption loss in decibels by the copper at $f = $ MHz.

2-19. A radar transmitter has an output power of 100 kW average. Calculate the power density in dBW/m² at a range of 3000 m and the free-space attenuation in decibels at $f = 10$ GHz.

2-20. Write a complete FORTRAN program to compute the magnitudes of reflectivity in vertical polarization against a grazing angle of seawater. The frequency varies from 0.1 to 40 GHz with an increment of 0.1 GHz between 0.1 to 1 GHz, 1 GHz between 1 to 10 GHz, and 5 GHz between 10 to 40 GHz. Use F10.5 format for numerical outputs and Hollerith format for character outputs. Print the outputs in three columns such as frequency (GHz), grazing angle (degrees), and gamma (vertical reflectivity).

2-21. Write a complete FORTRAN program to compute the magnitudes of reflectivity in horizontal polarization against a grazing angle for seawater. (Refer to Problem 2-20 for specifications.)

2-22. Write a complete FORTRAN program to compute the light transmittance and light reflection of a gold-film coating on a nonabsorbing plastic glass for thicknesses of 10 to 100 Å with an increment of 10 Å each step. The wavelength varies from 2000 to

10,000 Å with an increment of 500 Å each step. The values of the refractive index n and the extinction index k of a gold-film coating on a nonabsorbing plastic glass deposited in a vacuum are listed in Table 2-6-2. The refractive index n of the nonabsorbing plastic glass is 1.5. Use F10.5 format for numerical outputs, Hollerith format for character outputs, and Data statements to read in the input values.

2-23. Write a complete FORTRAN program to compute the light transmittance and light reflection for an aluminum-film coating on a nonabsorbing plastic substrate for thicknesses of 10 to 100 Å and print out the results in percentages. Use F10.5 format for numerical outputs, Hollerith format for character outputs, and Data statements to read in the input values. Print the outputs in column form with proper head-letters and units. The refractive index n of the nonabsorbing plastic glass is 1.50. The refractive index n and extinction index k for aluminum film are tabulated in Table P2-23. (Refer to Problem 2-22 for specifications.)

2-24. Repeat Problem 2-22 for a silver-film coating on a nonabsorbing plastic glass for the wavelengths from 2000 to 3700 Å with an increment of 100 Å each step. (Refer to Table 2-6-2 for the values of n and k.)

TABLE P2-23

Wavelength (Å)	Refractive index n	Extinction index k	Wavelength (Å)	Refractive index n	Extinction index k
0000	0.00	0.00	0000	0.00	0.00
2000	0.11	2.20	4500	0.51	5.00
2200	0.13	2.40	5000	0.64	5.50
2400	0.16	2.54	5500	0.82	5.99
2600	0.19	2.85	6000	1.00	6.50
2800	0.22	3.13	6500	1.30	7.11
3000	0.25	3.33	7000	1.55	7.00
3200	0.28	3.56	7500	1.80	7.12
3400	0.31	3.80	8000	1.99	7.05
3600	0.34	4.01	8500	2.08	7.15
3800	0.37	4.25	9000	1.96	7.70
4000	0.40	4.45	9500	1.75	8.50

2-25. Repeat Problem 2-22 for a copper-film coating on a nonabsorbing plastic glass for the wavelengths from 4500 to 10,000 Å with an increment of 500 Å each step.

2-26. Repeat Problems 2-20 and 2-21 for dry sand.

2-27. Repeat Problems 2-20 and 2-21 for concrete cement.

2-28. Repeat Problems 2-20 and 2-21 for dry ground.

2-29. Start from Eqs. (2-3-25) and (2-3-36) and verify Eq. (2-5-24a).

2-30. Derive Eq. (2-5-25) from Eq. (2-5-24a) by assuming the loss tangent to be much less than unity.

Chapter 3

Microwave Transmission Lines

3-0 INTRODUCTION

Conventional two-conductor transmission lines are commonly used for transmitting microwave energy. If a line is properly matched to its characteristic impedance at each terminal, its efficiency can reach a maximum.

In ordinary circuit theory it is assumed that all impedance elements are lumped constants. This is not true for a long transmission line over a wide range of frequencies. Frequencies of operation are so high that inductances of short lengths of conductors and capacitances between short conductors and their surroundings cannot be neglected. These inductances and capacitances are distributed along the length of a conductor, and their effects combine at each point of the conductor. Since the wavelength is short in comparison to the physical length of the line, distributed parameters cannot be represented accurately by means of a lumped-parameter equivalent circuit. Thus microwave transmission lines can be analyzed in terms of voltage, current, and impedance only by the distributed-circuit theory. If the spacing between the lines is smaller than the wavelength of the transmitted signal, the transmission line must be analyzed as a waveguide.

3-1 TRANSMISSION-LINE EQUATIONS AND SOLUTIONS

3-1-1 Transmission-Line Equations

A transmission line can be analyzed either by the solution of Maxwell's field equations or by the methods of distributed-circuit theory. The solution of Maxwell's equations involves three space variables in addition to the time variable. The distributed-circuit method, however, involves only one space variable in addition to

the time variable. In this section the latter method is used to analyze a transmission line in terms of the voltage, current, impedance, and power along the line.

Based on uniformly distributed-circuit theory, the schematic circuit of a conventional two-conductor transmission line with constant parameters R, L, G, and C is shown in Fig. 3-1-1. The parameters are expressed in their respective names per unit length, and the wave propagation is assumed in the positive z direction.
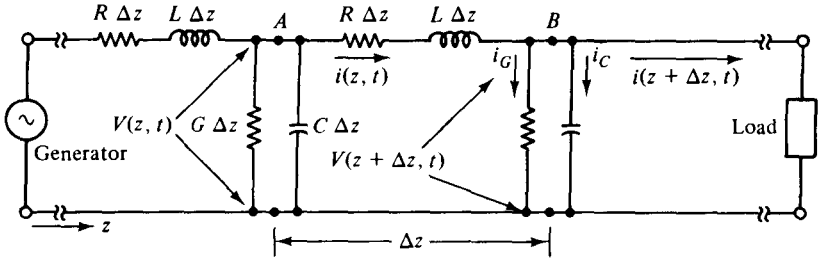
Figure 3-1-1   Elementary section of a transmission line.

By Kirchhoff's voltage law, the summation of the voltage drops around the central loop is given by

$$v(z, t) = i(z, t)R\,\Delta z + L\Delta z \frac{\partial i(z, t)}{\partial t} + v(z, t) + \frac{\partial v(z, t)}{dz}\,\Delta z \qquad (3\text{-}1\text{-}1)$$

Rearranging this equation, dividing it by Δz, and then omitting the argument (z, t), which is understood, we obtain

$$-\frac{\partial v}{\partial z} = Ri + L\frac{\partial i}{\partial t} \qquad (3\text{-}1\text{-}2)$$

Using Kirchhoff's current law, the summation of the currents at point B in Fig. 3-1-1 can be expressed as

$$i(z, t) = v(z + \Delta z, t)G\,\Delta z + C\,\Delta z\,\frac{\partial v(z + \Delta z, t)}{\partial t} + i(z + \Delta z, t)$$

$$= \left[v(z, t) + \frac{\partial v(z, t)}{\partial z}\,\Delta z\right]G\,\Delta z \qquad (3\text{-}1\text{-}3)$$

$$+\, C\,\Delta z\,\frac{\partial}{\partial t}\left[v(z, t) + \frac{\partial v(z, t)}{\partial z}\,\Delta z\right] + i(z, t) + \frac{\partial i(z, t)}{\partial z}\,\Delta z$$

By rearranging the preceding equation, dividing it by Δz, omitting (z, t), and assuming Δz equal to zero, we have

$$-\frac{\partial i}{\partial z} = Gv + C\frac{\partial v}{\partial t} \qquad (3\text{-}1\text{-}4)$$

Then by differentiating Eq. (3-1-2) with respect to z and Eq. (3-1-4) with respect to t and combining the results, the final transmission-line equation in voltage form is

found to be

$$\frac{\partial^2 v}{\partial z^2} = RGv + (RC + LG)\frac{\partial v}{\partial t} + LC\frac{\partial^2 v}{\partial t^2} \qquad (3\text{-}1\text{-}5)$$

Also, by differentiating Eq. (3-1-2) with respect to t and Eq. (3-1-4) with respect to z and combining the results, the final transmission-line equation in current form is

$$\frac{\partial^2 i}{\partial z^2} = RGi + (RC + LG)\frac{\partial i}{\partial t} + LC\frac{\partial^2 i}{\partial t^2} \qquad (3\text{-}1\text{-}6)$$

All these transmission-line equations are applicable to the general transient solution. The voltage and current on the line are the functions of both position z and time t. The instantaneous line voltage and current can be expressed as

$$v(z, t) = \text{Re } \mathbf{V}(z)e^{j\omega t} \qquad (3\text{-}1\text{-}7)$$

$$i(z, t) = \text{Re } \mathbf{I}(z)e^{j\omega t} \qquad (3\text{-}1\text{-}8)$$

where Re stands for "real part of." The factors $\mathbf{V}(z)$ aand $\mathbf{I}(z)$ are complex quantities of the sinusoidal functions of position z on the line and are known as *phasors*. The phasors give the magnitudes and phases of the sinusoidal function at each position of z, and they can be expressed as

$$\mathbf{V}(z) = \mathbf{V}_+ e^{-\gamma z} + \mathbf{V}_- e^{\gamma z} \qquad (3\text{-}1\text{-}9)$$

$$\mathbf{I}(z) = \mathbf{I}_+ e^{-\gamma z} + \mathbf{I}_- e^{\gamma z} \qquad (3\text{-}1\text{-}10)$$

$$\gamma = \alpha + j\beta \qquad \text{(propagation constant)} \qquad (3\text{-}1\text{-}11)$$

where $\mathbf{V}_+$ and $\mathbf{I}_+$ indicate complex amplitudes in the positive z direction, $\mathbf{V}_-$ and $\mathbf{I}_-$ signify complex amplitudes in the negative z direction, α is the attenuation constant in nepers per unit length, and β is the phase constant in radians per unit length.

If we substitute $j\omega$ for $\partial/\partial t$ in Eqs. (3-1-2), (3-1-4), (3-1-5), and (3-1-6) and divide each equation by $e^{j\omega t}$, the transmission-line equations in phasor form of the frequency domain become

$$\frac{d\mathbf{V}}{dz} = -\mathbf{ZI} \qquad (3\text{-}1\text{-}12)$$

$$\frac{d\mathbf{I}}{dz} = -\mathbf{YV} \qquad (3\text{-}1\text{-}13)$$

$$\frac{d^2\mathbf{V}}{dz^2} = \gamma^2 \mathbf{V} \qquad (3\text{-}1\text{-}14)$$

$$\frac{d^2\mathbf{I}}{dz^2} = \gamma^2 \mathbf{I} \qquad (3\text{-}1\text{-}15)$$

in which the following substitutions have been made:

$$\mathbf{Z} = R + j\omega L \qquad \text{(ohms per unit length)} \qquad (3\text{-}1\text{-}16)$$

$$\mathbf{Y} = G + j\omega C \qquad \text{(mhos per unit length)} \qquad (3\text{-}1\text{-}17)$$

$$\gamma = \sqrt{ZY} = \alpha + j\beta \qquad \text{(propagation constant)} \qquad (3\text{-}1\text{-}18)$$

For a lossless line, $R = G = 0$, and the transmission-line equations are expressed as

$$\frac{d\mathbf{V}}{dz} = -j\omega L\, \mathbf{I} \tag{3-1-19}$$

$$\frac{d\mathbf{I}}{dz} = -j\omega C\, \mathbf{V} \tag{3-1-20}$$

$$\frac{d^2\mathbf{V}}{dz^2} = -\omega^2 LC\, \mathbf{V} \tag{3-1-21}$$

$$\frac{d^2\mathbf{I}}{dz^2} = -\omega^2 LC\, \mathbf{I} \tag{3-1-22}$$

It is interesting to note that Eqs. (3-1-14) and (3-1-15) for a transmission line are similar to equations of the electric and magnetic waves, respectively. The only difference is that the transmission-line equations are one-dimensional.

3-1-2 Solutions of Transmission-Line Equations

The one possible solution for Eq. (3-1-14) is

$$\mathbf{V} = \mathbf{V}_+ e^{-\gamma z} + \mathbf{V}_- e^{\gamma z} = \mathbf{V}_+ e^{-\alpha z} e^{-j\beta z} + \mathbf{V}_- e^{\alpha z} e^{j\beta z} \tag{3-1-23}$$

The factors $\mathbf{V}_+$ and $\mathbf{V}_-$ represents complex quantities. The term involving $e^{-j\beta z}$ shows a wave traveling in the positive z direction, and the term with the factor $e^{j\beta z}$ is a wave going in the negative z direction. The quantity βz is called the *electrical length of the line* and is measured in radians.

Similarly, the one possible solution for Eq. (3-1-15) is

$$\mathbf{I} = \mathbf{Y}_0(\mathbf{V}_+ e^{-\gamma z} - \mathbf{V}_- e^{\gamma z}) = \mathbf{Y}_0(\mathbf{V}_+ e^{-\alpha z} e^{-j\beta z} - \mathbf{V}_- e^{\alpha z} e^{j\beta z}) \tag{3-1-24}$$

In Eq. (3-1-24) the characteristic impedance of the line is defined as

$$\mathbf{Z}_0 = \frac{1}{\mathbf{Y}_0} \equiv \sqrt{\frac{\mathbf{Z}}{\mathbf{Y}}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = R_0 \pm jX_0 \tag{3-1-25}$$

The magnitude of both voltage and current waves on the line is shown in Fig. 3-1-2.

(a) Voltage wave   (b) Current wave
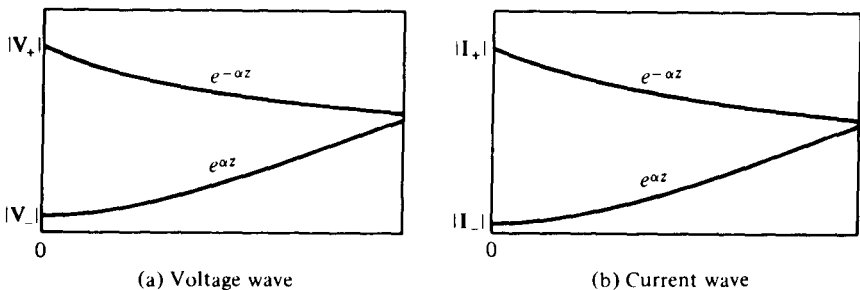
Figure 3-1-2   Magnitude of voltage and current traveling waves.

At microwave frequencies it can be seen that

$$R \ll \omega L \quad \text{and} \quad G \ll \omega C \tag{3-1-26}$$

By using the binomial expansion, the propagation constant can be expressed as

$$
\begin{aligned}
\gamma &= \sqrt{(R + j\omega L)(G + j\omega C)} \\
&= \sqrt{(j\omega)^2 LC} \sqrt{\left(1 + \frac{R}{j\omega L}\right)\left(1 + \frac{G}{j\omega C}\right)} \\
&\approx j\omega \sqrt{LC}\left[\left(1 + \frac{1}{2}\frac{R}{j\omega L}\right)\left(1 + \frac{1}{2}\frac{G}{j\omega C}\right)\right] \\
&\approx j\omega \sqrt{LC}\left[1 + \frac{1}{2}\left(\frac{R}{j\omega L} + \frac{G}{j\omega C}\right)\right] \\
&= \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) + j\omega\sqrt{LC}
\end{aligned} \tag{3-1-27}
$$

Therefore the attenuation and phase constants are, respectively, given by

$$\alpha = \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \tag{3-1-28}$$

$$\beta = \omega\sqrt{LC} \tag{3-1-29}$$

Similarly, the characteristic impedance is found to be

$$
\begin{aligned}
\mathbf{Z}_0 &= \sqrt{\frac{R + j\omega L}{G + j\omega C}} \\
&= \sqrt{\frac{L}{C}}\left(1 + \frac{R}{j\omega L}\right)^{1/2}\left(1 + \frac{G}{j\omega C}\right)^{-1/2} \\
&\approx \sqrt{\frac{L}{C}}\left(1 + \frac{1}{2}\frac{R}{j\omega L}\right)\left(1 - \frac{1}{2}\frac{G}{j\omega C}\right) \\
&\approx \sqrt{\frac{L}{C}}\left[1 + \frac{1}{2}\left(\frac{R}{j\omega L} - \frac{G}{j\omega C}\right)\right] \\
&\approx \sqrt{\frac{L}{C}}
\end{aligned} \tag{3-1-30}
$$

From Eq. (3-1-29) the phase velocity is

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{3-1-31}$$

The product of LC is independent of the size and separation of the conductors and depends only on the permeability μ and permittivity of ϵ of the insulating medium. If a lossless transmission line used for microwave frequencies has an air dielectric and contains no ferromagnetic materials, free-space parameters can be assumed.

Thus the numerical value of $1/\sqrt{LC}$ for air-insulated conductors is approximately equal to the velocity of light in vacuum. That is,

$$v_p = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = c = 3 \times 10^8 \text{ m/s} \qquad (3\text{-}1\text{-}32)$$

When the dielectric of a lossy microwave transmission line is not air, the phase velocity is smaller than the velocity of light in vacuum and is given by

$$v_\epsilon = \frac{1}{\sqrt{\mu \epsilon}} = \frac{c}{\sqrt{\mu_r \epsilon_r}} \qquad (3\text{-}1\text{-}33)$$

In general, the relative phase velocity factory can be defined as

$$\text{Velocity factor} = \frac{\text{actual phase velocity}}{\text{velocity of light in vacuum}}$$

$$v_r = \frac{v_\epsilon}{c} = \frac{1}{\sqrt{\mu_r \epsilon_r}} \qquad (3\text{-}1\text{-}34)$$

A low-loss transmission line filled only with dielectric medium, such as a coaxial line with solid dielectric between conductors, has a velocity factor on the order of about 0.65.

Example 3-1-1:   Line Characteristic Impedance and Propagation Constant

A transmission line has the following parameters:

$$R = 2 \text{ } \Omega/\text{m} \qquad G = 0.5 \text{ mmho/m} \qquad f = 1 \text{ GHz}$$

$$L = 8 \text{ nH/m} \qquad C = 0.23 \text{ pF}$$

Calculate: **(a)** the characteristic impedance; **(b)** the propagation constant.

Solution

a. From Eq. (3-1-25) the line characteristic impedance is

$$\mathbf{Z}_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = \sqrt{\frac{2 + j2\pi \times 10^9 \times 8 \times 10^{-9}}{0.5 \times 10^{-3} + j2\pi \times 10^9 \times 0.23 \times 10^{-12}}}$$

$$= \sqrt{\frac{50.31 \underline{/87.72°}}{15.29 \times 10^{-4} \underline{/70.91°}}} = 181.39 \underline{/8.40°} = 179.44 + j26.50$$

b. From Eq. (3-1-18) the propagation constant is

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \sqrt{(50.31 \underline{/87.72°})(15.29 \times 10^{-4} \underline{/70.91°})}$$

$$= \sqrt{769.24 \times 10^{-4} \underline{/158.63°}}$$

$$= 0.2774 \underline{/79.31°} = 0.051 + j0.273$$

3-2 REFLECTION COEFFICIENT AND TRANSMISSION COEFFICIENT

3-2-1 Reflection Coefficient

In the analysis of the solutions of transmission-line equations in Section 3-1, the traveling wave along the line contains two components: one traveling in the positive z direction and the other traveling the negative z direction. If the load impedance is equal to the line characteristic impedance, however, the reflected traveling wave does not exist.

Figure 3-2-1 shows a transmission line terminated in an impedance $\mathbf{Z}_\ell$. It is usually more convenient to start solving the transmission-line problem from the receiving rather than the sending end, since the voltage-to-current relationship at the load point is fixed by the load impedance. The incident voltage and current waves traveling along the transmission line are given by

$$\mathbf{V} = \mathbf{V}_+ e^{-\gamma z} + \mathbf{V}_- e^{+\gamma z} \tag{3-2-1}$$

$$\mathbf{I} = \mathbf{I}_+ e^{-\gamma z} + \mathbf{I}_- e^{+\gamma z} \tag{3-2-2}$$

in which the current wave can be expressed in terms of the voltage by

$$\mathbf{I} = \frac{\mathbf{V}_+}{\mathbf{Z}_0} e^{-\gamma z} - \frac{\mathbf{V}^-}{\mathbf{Z}_0} e^{\gamma z} \tag{3-2-3}$$

If the line has a length of ℓ, the voltage and current at the receiving end become

$$\mathbf{V}_\ell = \mathbf{V}_+ e^{-\gamma \ell} + \mathbf{V}_- e^{\gamma \ell} \tag{3-2-4}$$

$$\mathbf{I}_\ell = \frac{1}{\mathbf{Z}_0} (\mathbf{V}_+ e^{-\gamma \ell} - \mathbf{V}_- e^{\gamma \ell}) \tag{3-2-5}$$

The ratio of the voltage to the current at the receiving end is the load impedance. That is,

$$\mathbf{Z}_\ell = \frac{\mathbf{V}_\ell}{\mathbf{I}_\ell} = \mathbf{Z}_0 \frac{\mathbf{V}_+ e^{-\gamma \ell} + \mathbf{V}_- e^{\gamma \ell}}{\mathbf{V}_+ e^{-\gamma \ell} - \mathbf{V}_- e^{\gamma \ell}} \tag{3-2-6}$$
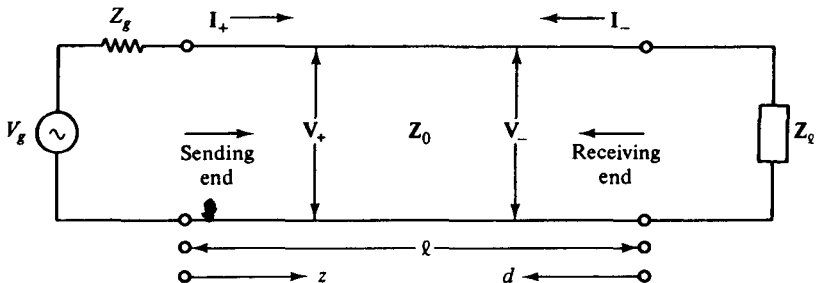
Figure 3-2-1   Transmission line terminated in a load impedance.

The reflection coefficient, which is designated by Γ (gamma), is defined as

$$\text{Reflection coefficient} \equiv \frac{\text{reflected voltage or current}}{\text{incident voltage or current}}$$

$$\Gamma \equiv \frac{V_{\text{ref}}}{V_{\text{inc}}} = \frac{-I_{\text{ref}}}{I_{\text{inc}}} \tag{3-2-7}$$

If Eq. (3-2-6) is solved for the ratio of the reflected voltage at the receiving end, which is $V_- e^{\gamma \ell}$, to the incident voltage at the receiving end, which is $V_+ e^{\gamma \ell}$, the result is the reflection coefficient at the receiving end:

$$\Gamma_\ell = \frac{V_- e^{\gamma \ell}}{V_+ e^{-\gamma \ell}} = \frac{Z_\ell - Z_0}{Z_\ell + Z_0} \tag{3-2-8}$$

If the load impedance and/or the characteristic impedance are complex quantities, as is usually the case, the reflection coefficient is generally a complex quantity that can be expressed as

$$\Gamma_\ell = |\Gamma_\ell| e^{j\theta_\ell} \tag{3-2-9}$$

where $|\Gamma_\ell|$ is the magnitude and never greater than unity—that is, $|\Gamma_\ell| \le 1$. Note that θ_ℓ is the phase angle between the incident and reflected voltages at the receiving end. It is usually called the phase angle of the reflection coefficient.

The general solution of the reflection coefficient at any point on the line, then, corresponds to the incident and reflected waves at that point, each attenuated in the direction of its own progress along the line. The generalized reflection coefficient is defined as

$$\Gamma \equiv \frac{V_- e^{\gamma z}}{V_+ e^{-\gamma z}} \tag{3-2-10}$$

From Fig. 3-2-1 let $z = \ell - d$. Then the reflection coefficient at some point located a distance d from the receiving end is

$$\Gamma_d = \frac{V_- e^{\gamma(\ell-d)}}{V_+ e^{-\gamma(\ell-d)}} = \frac{V_- e^{\gamma \ell}}{V_+ e^{-\gamma \ell}} e^{-2\gamma d} = \Gamma_\ell e^{-2\gamma d} \tag{3-2-11}$$

Next, the reflection coefficient at that point can be expressed in terms of the reflection coefficient at the receiving end as

$$\Gamma_d = \Gamma_\ell e^{-2\alpha d} e^{-j2\beta d} = |\Gamma_\ell| e^{-2\alpha d} e^{j(\theta_\ell - 2\beta d)} \tag{3-2-12}$$

This is a very useful equation for determining the reflection coefficient at any point along the line. For a lossy line, both the magnitude and phase of the reflection coefficient are changing in an inward-spiral way as shown in Fig. 3-2-2. For a lossless line, $\alpha = 0$, the magnitude of the reflection coefficient remains constant, and only the phase of Γ is changing circularly toward the generator with an angle of $-2\beta d$ as shown in Fig. 3-2-3.

It is evident that Γ_ℓ will be zero and there will be no reflection from the receiving end when the terminating impedance is equal to the characteristic impedance
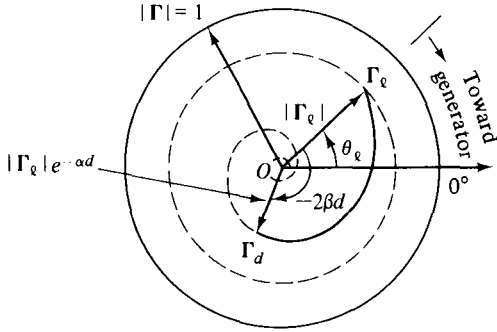
Figure 3-2-2   Reflection coefficient for lossy line.
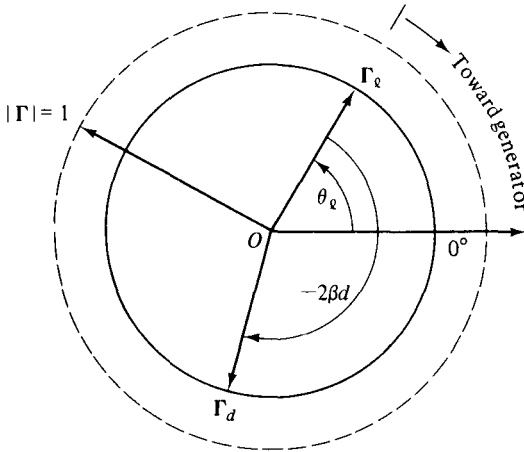
Figure 3-2-3   Reflection coefficient for lossless line.

of the line. Thus a terminating impedance that differs from the characteristic impedance will create a reflected wave traveling toward the source from the termination. The reflection, upon reaching the sending end, will itself be reflected if the source impedance is different from the line characteristic impedance at the sending end.

3-2-2 Transmission Coefficient

A transmission line terminated in its characteristic impedance Z_0 is called a *properly terminated* line. Otherwise it is called an *improperly terminated* line. As described earlier, there is a reflection coefficient Γ at any point along an improperly terminated line. According to the principle of conservation of energy, the incident power minus the reflected power must be equal to the power transmitted to the load. This can be expressed as

$$1 - \Gamma_\ell^2 = \frac{Z_0}{Z_\ell} T^2 \tag{3-2-13}$$

Equation (3-2-13) will be verified later. The letter T represents the transmission coefficient, which is defined as

$$\mathbf{T} \equiv \frac{\text{transmitted voltage or current}}{\text{incident voltage or current}} = \frac{\mathbf{V}_{tr}}{\mathbf{V}_{inc}} = \frac{\mathbf{I}_{tr}}{\mathbf{I}_{inc}} \tag{3-2-14}$$

Figure 3-2-4 shows the transmission of power along a transmission line where P_{inc} is the incident power, P_{ref} the reflected power, and P_{tr} the transmitted power.
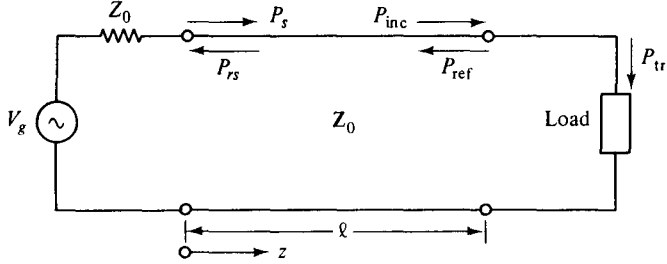
Figure 3-2-4   Power transmission on a line.

Let the traveling waves at the receiving end be

$$\mathbf{V}_+ e^{-\gamma\ell} + \mathbf{V}_- e^{\gamma\ell} = \mathbf{V}_{tr} e^{-\gamma\ell} \tag{3-2-15}$$

$$\frac{\mathbf{V}_+}{\mathbf{Z}_0} e^{-\gamma\ell} - \frac{\mathbf{V}_-}{\mathbf{Z}_0} e^{\gamma\ell} = \frac{\mathbf{V}_{tr}}{\mathbf{Z}_\ell} e^{-\gamma\ell} \tag{3-2-16}$$

Multiplication of Eq. (3-2-16) by $\mathbf{Z}_\ell$ and substitution of the result in Eq. (3-2-15) yield

$$\Gamma_\ell = \frac{\mathbf{V}_- e^{\gamma\ell}}{\mathbf{V}_+ e^{-\gamma\ell}} = \frac{\mathbf{Z}_\ell - \mathbf{Z}_0}{\mathbf{Z}_\ell + \mathbf{Z}_0} \tag{3-2-17}$$

which, in turn, on substitution back into Eq. (3-2-15), results in

$$\mathbf{T} = \frac{\mathbf{V}_{tr}}{\mathbf{V}_+} = \frac{2\mathbf{Z}_\ell}{\mathbf{Z}_\ell + \mathbf{Z}_0} \tag{3-2-18}$$

The power carried by the two waves in the side of the incident and reflected waves is

$$P_{inr} = P_{inc} - P_{ref} = \frac{(\mathbf{V}_+ e^{-\alpha\ell})^2}{2\mathbf{Z}_0} - \frac{(\mathbf{V}_- e^{\alpha\ell})^2}{2\mathbf{Z}_0} \tag{3-2-19}$$

The power carried to the load by the transmitted waves is

$$P_{tr} = \frac{(\mathbf{V}_{tr} e^{-\alpha\ell})^2}{2\mathbf{Z}_\ell} \tag{3-2-20}$$

By setting $P_{inr} = P_{tr}$ and using Eqs. (3-2-17) and (3-2-18), we have

$$\mathbf{T}^2 = \frac{\mathbf{Z}_\ell}{\mathbf{Z}_0}(1 - \Gamma_\ell^2) \tag{3-2-21}$$

This relation verifies the previous statement that the transmitted power is equal to the difference of the incident power and reflected power.

Example 3-2-1:   Reflection Coefficient and Transmission Coefficient

A certain transmission line has a characteristic impedance of $75 + j0.01$ Ω and is terminated in a load impedance of $70 + j50$ Ω. Compute **(a)** the reflection coefficient; **(b)** the transmission coefficient. Verify: **(c)** the relationship shown in Eq. (3-2-21); **(d)** the transmission coefficient equals the algebraic sum of 1 plus the reflection coefficient as shown in Eq. (2-3-18).

Solution

a. From Eq. (3-2-17) the reflection coefficient is

$$\Gamma = \frac{\mathbf{Z}_\ell - \mathbf{Z}_0}{\mathbf{Z}_\ell + \mathbf{Z}_0} = \frac{70 + j50 - (75 + j0.01)}{70 + j50 + (75 + j0.01)}$$

$$= \frac{50.24\underline{/95.71°}}{153.38\underline{/19.03°}} = 0.33\underline{/76.68°} = 0.08 + j0.32$$

b. From Eq. (3-2-18) the transmission coefficient is

$$\mathbf{T} = \frac{2\mathbf{Z}_\ell}{\mathbf{Z}_\ell + \mathbf{Z}_0} = \frac{2(70 + j50)}{70 + j50 + (75 + j0.01)}$$

$$= \frac{172.05\underline{/35.54°}}{153.38\underline{/19.03°}} = 1.12\underline{/16.51°} = 1.08 + j0.32$$

c.

$$\mathbf{T}^2 = (1.12\underline{/16.51°})^2 = 1.25\underline{/33.02°}$$

$$\frac{\mathbf{Z}_\ell}{\mathbf{Z}_0}(1 - \Gamma^2) = \frac{70 + j50}{75 + j0.01}[1 - (0.33\underline{/76.68°})^2]$$

$$= \frac{86\underline{/35.54°}}{75\underline{/0°}} \times 1.10\underline{/-2.6°} = 1.25\underline{/33°}$$

Thus Eq. (3-2-21) is verified.

d. From Eq. (2-3-18) we obtain

$$\mathbf{T} = 1.08 + j0.32 = 1 + 0.08 + j0.32 = 1 + \Gamma$$

3-3 STANDING WAVE AND STANDING-WAVE RATIO

3-3-1 Standing Wave

The general solutions of the transmission-line equation consist of two waves traveling in opposite directions with unequal amplitude as shown in Eqs. (3-1-23) and (3-1-24). Equation (3-1-23) can be written

$$\mathbf{V} = \mathbf{V}_+ e^{-\alpha z} e^{-j\beta z} + \mathbf{V}_- e^{\alpha z} e^{j\beta z}$$

$$= \mathbf{V}_+ e^{-\alpha z}[\cos(\beta z) - j\sin(\beta z)] + \mathbf{V}_- e^{\alpha z}[\cos(\beta z) + j\sin(\beta z)] \qquad (3\text{-}3\text{-}1)$$

$$= (\mathbf{V}_+ e^{-\alpha z} + \mathbf{V}_- e^{\alpha z})\cos(\beta z) - j(\mathbf{V}_+ e^{-\alpha z} - \mathbf{V}_- e^{\alpha z})\sin(\beta z)$$

With no loss in generality it can be assumed that $\mathbf{V}_+ e^{-\alpha z}$ and $\mathbf{V}_- e^{\alpha z}$ are real. Then the voltage-wave equation can be expressed as

$$\mathbf{V}_s = \mathbf{V}_0 e^{-j\phi} \tag{3-3-2}$$

This is called the *equation of the voltage standing wave*, where

$$\mathbf{V}_0 = [(\mathbf{V}_+ e^{-\alpha z} + \mathbf{V}_- e^{\alpha z})^2 \cos^2(\beta z) + (\mathbf{V}_+ e^{-\alpha z} - \mathbf{V}_- e^{-\alpha z})^2 \sin^2(\beta z)]^{1/2} \tag{3-3-3}$$

which is called the *standing-wave pattern* of the voltage wave or the amplitude of the standing wave, and

$$\phi = \arctan\left(\frac{\mathbf{V}_+ e^{-\alpha z} - \mathbf{V}_- e^{\alpha z}}{\mathbf{V}_+ e^{-\alpha z} + \mathbf{V}_- e^{\alpha z}} \tan(\beta z)\right) \tag{3-3-4}$$

which is called the *phase pattern of the standing wave*. The maximum and minimum values of Eq. (3-3-3) can be found as usual by differentiating the equation with respect to βz and equating the result to zero. By doing so and substituting the proper values of βz in the equation, we find that

1. The maximum amplitude is

$$\mathbf{V}_{\max} = \mathbf{V}_+ e^{-\alpha z} + \mathbf{V}_- e^{\alpha z} = \mathbf{V}_+ e^{-\alpha z}(1 + |\Gamma|) \tag{3-3-5}$$

   and this occurs at $\beta z = n\pi$, where $n = 0, \pm 1, \pm 2, \ldots$.
2. The minimum amplitude is

$$\mathbf{V}_{\min} = \mathbf{V}_+ e^{-\alpha z} - V_- e^{\alpha z} = \mathbf{V}_+ e^{-\alpha z}(1 - |\Gamma|) \tag{3-3-6}$$

   and this occurs at $\beta z = (2n - 1)\pi/2$, where $n = 0, \pm 1, \pm 2, \ldots$.
3. The distance between any two successive maxima or minima is one-half wavelength, since

$$\beta z = n\pi \qquad z = \frac{n\pi}{\beta} = \frac{n\pi}{2\pi/\lambda} = n\frac{\lambda}{2} \qquad (n = 0, \pm 1, \pm 2, \ldots)$$

Then

$$z_1 = \frac{\lambda}{2} \tag{3-3-7}$$

It is evident that there are no zeros in the minimum. Similarly,

$$\mathbf{I}_{\max} = \mathbf{I}_+ e^{-\alpha z} + \mathbf{I}_- e^{\alpha z} = \mathbf{I}_+ e^{-\alpha z}(1 + |\Gamma|) \tag{3-3-8}$$

$$\mathbf{I}_{\min} = \mathbf{I}_+ e^{-\alpha z} - \mathbf{I}_- e^{\alpha z} = \mathbf{I}_+ e^{-\alpha z}(1 - |\Gamma|) \tag{3-3-9}$$

The standing-wave patterns of two oppositely traveling waves with unequal amplitude in lossy or lossless line are shown in Figs. 3-3-1 and 3-3-2.

A further study of Eq. (3-3-3) reveals that

1. When $\mathbf{V}_+ \neq 0$ and $\mathbf{V}_- = 0$, the standing-wave pattern becomes

$$\mathbf{V}_0 = \mathbf{V}_+ e^{-\alpha z} \tag{3-3-10}$$
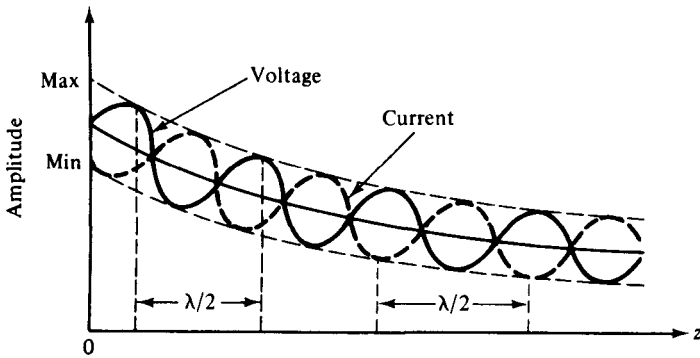
Figure 3-3-1   Standing-wave pattern in a lossy line.
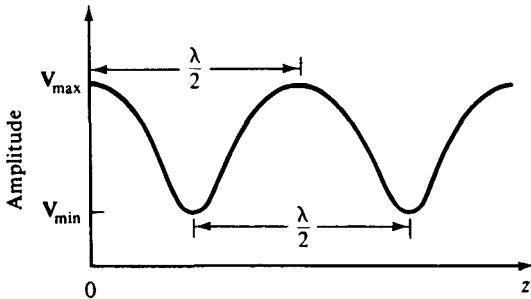
Figure 3-3-2   Voltage standing-wave pattern in a lossless line.

2. When $V_+ = 0$ and $V_- \neq 0$, the standing-wave pattern becomes

$$V_0 = V_- e^{\alpha z} \tag{3-3-11}$$

3. When the positive wave and the negative wave have equal amplitudes (that is, $|V_+ e^{-\alpha z}| = |V_- e^{\alpha z}|$) or the magnitude of the reflection coefficient is unity, the standing-wave pattern with a zero phase is given by

$$V_s = 2V_+ e^{-\alpha z} \cos{(\beta z)} \tag{3-3-12}$$

which is called a *pure standing wave*.

Similarly, the equation of a pure standing wave for the current is

$$I_s = -j2Y_0 V_+ e^{-\alpha z} \sin{(\beta z)} \tag{3-3-13}$$

Equations (3-3-12) and (3-3-13) show that the voltage and current standing waves are 90° out of phase along the line. The points of zero current are called the *current nodes*. The voltage nodes and current nodes are interlaced a quarter wavelength apart.

The voltage and current may be expressed as real functions of time and space:

$$v_s = (z, t) = \text{Re}[V_s(z)e^{j\omega t}] = 2V_+ e^{-\alpha z} \cos{(\beta z)} \cos{(\omega t)} \tag{3-3-14}$$

$$i_s = (z, t) = \text{Re}[I_s(z)e^{j\omega t}] = 2Y_0 V_+ e^{-\alpha z} \sin{(\beta z)} \sin{(\omega t)} \tag{3-3-15}$$

The amplitudes of Eqs. (3-3-14) and (3-3-15) vary sinusoidally with time; the

voltage is a maximum at the instant when the current is zero and vice versa. Figure 3-3-3 shows the pure-standing-wave patterns of the phasor of Eqs. (3-3-12) and (3-3-13) for an open-terminal line.
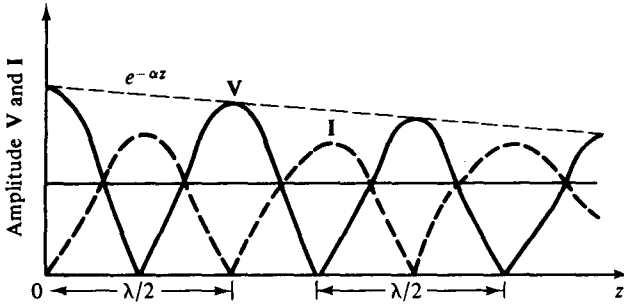
Figure 3-3-3 Pure standing waves of voltage and current.

3-3-2 Standing-Wave Ratio

Standing waves result from the simultaneous presence of waves traveling in opposite directions on a transmission line. The ratio of the maximum of the standing-wave pattern to the minimum is defined as the standing-wave ratio, designated by ρ. That is,

$$\text{Standing-wave ratio} \equiv \frac{\text{maximum voltage or current}}{\text{minimum voltage or current}}$$

$$\rho \equiv \frac{|\mathbf{V}_{max}|}{|\mathbf{V}_{min}|} = \frac{|\mathbf{I}_{max}|}{|\mathbf{I}_{min}|} \tag{3-3-16}$$

The standing-wave ratio results from the fact that the two traveling-wave components of Eq. (3-3-1) add in phase at some points and subtract at other points. The distance between two successive maxima or minima is $\lambda/2$. The standing-wave ratio of a pure traveling wave is unity and that of a pure standing wave is infinite. It should be noted that since the standing-wave ratios of voltage and current are identical, no distinctions are made between VSWR and ISWR.

When the standing-wave ratio is unity, there is no reflected wave and the line is called a *flat line*. The standing-wave ratio cannot be defined on a lossy line because the standing-wave pattern changes markedly from one position to another. On a low-loss line the ratio remains fairly constant, and it may be defined for some region. For a lossless line, the ratio stays the same throughout the line.

Since the reflected wave is defined as the product of an incident wave and its reflection coefficient, the standing-wave ratio ρ is related to the reflection coefficient Γ by

$$\rho = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{3-3-17}$$

and vice versa

$$|\Gamma| = \frac{\rho - 1}{\rho + 1} \qquad (3\text{-}3\text{-}18)$$

This relation is very useful for determining the reflection coefficient from the standing-wave ratio, which is usually found from the Smith chart. The curve in Fig. 3-3-4 shows the relationship between reflection coefficient $|\Gamma|$ and standing-wave ratio ρ.
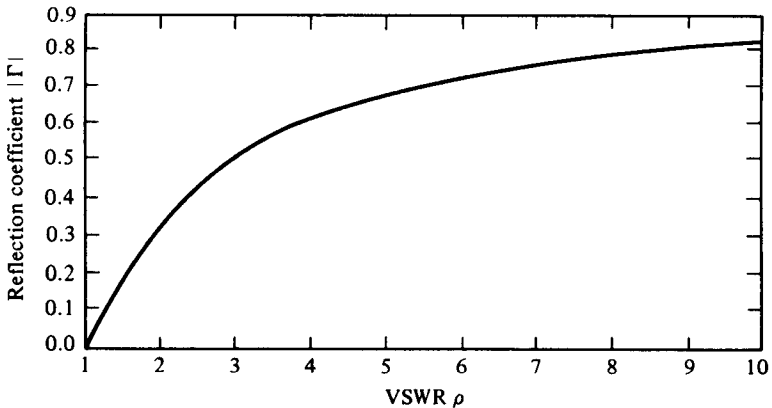
Figure 3-3-4 SWR versus reflection coefficient.

As a result of Eq. (3-3-17), since $|\Gamma| \leq 1$, the standing-wave ratio is a positive real number and never less than unity, $\rho \geq 1$. From Eq. (3-3-18) the magnitude of the reflection coefficient is never greater than unity.

Example 3-3-1:   Standing-Wave Ratio

A transmission line has a characteristic impedance of $50 + j0.01$ Ω and is terminated in a load impedance of $73 - j42.5$ Ω. Calculate: **(a)** the reflection coefficient; **(b)** the standing-wave ratio.

Solution

a. From Eq. (3-2-8) the reflection coefficient is

$$\Gamma = \frac{Z_\ell - Z_0}{Z_\ell + Z_0} = \frac{73 - j42.5 - (50 + j0.01)}{73 - j42.5 + (50 + j0.01)} = 0.377\underline{/-42.7^\circ}$$

b. From Eq. (3-3-17) the standing-wave ratio is

$$\rho = \frac{1 + |\Gamma|}{1 - |\Gamma|} = \frac{1 + 0.377}{1 - 0.377} = 2.21$$

3-4 LINE IMPEDANCE AND ADMITTANCE

3-4-1 Line Impedance

The line impedance of a transmission line is the complex ratio of the voltage phasor at any point to the current phasor at that point. It is defined as

$$Z \equiv \frac{V(z)}{I(z)} \tag{3-4-1}$$

Figure 3-4-1 shows a diagram for a transmission line.
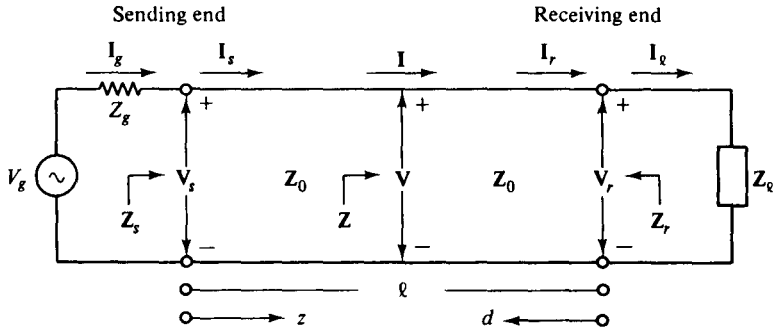
Figure 3-4-1   Diagram of a transmission line showing notations.

In general, the voltage or current along a line is the sum of the respective incident wave and reflected wave—that is,

$$V = V_{\text{inc}} + V_{\text{ref}} = V_+ e^{-\gamma z} + V_- e^{\gamma z} \tag{3-4-2}$$

$$I = I_{\text{inc}} + I_{\text{ref}} = Y_0(V_+ e^{-\gamma z} - V_- e^{\gamma z}) \tag{3-4-3}$$

At the sending end $z = 0$; then Eqs. (3-4-2) and (3-4-3) become

$$I_s Z_s = V_+ + V_- \tag{3-4-4}$$

$$I_s Z_0 = V_+ - V_- \tag{3-4-5}$$

By solving these two equations for V_+ and V_-, we obtain

$$V_+ = \frac{I_s}{2}(Z_s + Z_0) \tag{3-4-6}$$

$$V_- = \frac{I_s}{2}(Z_s - Z_0) \tag{3-4-7}$$

Substitution of V_+ and V_- in Eqs. (3-4-2) and (3-4-3) yields

$$V = \frac{I_s}{2}[(Z_s + Z_0)e^{-\gamma z} + (Z_s - Z_0)e^{\gamma z}] \tag{3-4-8}$$

$$I = \frac{I_s}{2Z_0}[(Z_s + Z_0)e^{-\gamma z} - (Z_s - Z_0)e^{\gamma z}] \tag{3-4-9}$$

Then the line impedance at any point z from the sending end in terms of $\mathbf{Z}_s$ and $\mathbf{Z}_0$ is expressed as

$$\mathbf{Z} = \mathbf{Z}_0 \frac{(\mathbf{Z}_s + \mathbf{Z}_0)e^{-\gamma z} + (\mathbf{Z}_s - \mathbf{Z}_0)e^{\gamma z}}{(\mathbf{Z}_s + \mathbf{Z}_0)e^{-\gamma z} - (\mathbf{Z}_s - \mathbf{Z}_0)e^{\gamma z}} \qquad (3\text{-}4\text{-}10)$$

At $z = \ell$ the line impedance at the receiving end in terms of $\mathbf{Z}_s$ and $\mathbf{Z}_0$ is given by

$$\mathbf{Z}_r = \mathbf{Z}_0 \frac{(\mathbf{Z}_s + \mathbf{Z}_0)e^{-\gamma \ell} + (\mathbf{Z}_s - \mathbf{Z}_0)e^{\gamma \ell}}{(\mathbf{Z}_s + \mathbf{Z}_0)e^{-\gamma \ell} - (\mathbf{Z}_s - \mathbf{Z}_0)e^{\gamma \ell}} \qquad (3\text{-}4\text{-}11)$$

Alternatively, the line impedance can be expressed in terms of $\mathbf{Z}_\ell$ and $\mathbf{Z}_0$. At $z = \ell$, $\mathbf{V}_r = \mathbf{I}_\ell \mathbf{Z}_\ell$; then

$$\mathbf{I}_\ell \mathbf{Z}_\ell = \mathbf{V}_+ e^{-\gamma \ell} + \mathbf{V}_- e^{\gamma \ell} \qquad (3\text{-}4\text{-}12)$$

$$\mathbf{I}_\ell \mathbf{Z}_0 = \mathbf{V}_+ e^{-\gamma \ell} - \mathbf{V}_- e^{\gamma \ell} \qquad (3\text{-}4\text{-}13)$$

Solving these two equations for $\mathbf{V}_+$ and $\mathbf{V}_-$, we have

$$\mathbf{V}_+ = \frac{\mathbf{I}_\ell}{2}(\mathbf{Z}_\ell + \mathbf{Z}_0)e^{\gamma \ell} \qquad (3\text{-}4\text{-}14)$$

$$\mathbf{V}_- = \frac{\mathbf{I}_\ell}{2}(\mathbf{Z}_\ell - \mathbf{Z}_0)e^{-\gamma \ell} \qquad (3\text{-}4\text{-}15)$$

Then substituting these results in Eqs. (3-4-2) and (3-4-3) and letting $z = \ell - d$, we obtain

$$\mathbf{V} = \frac{\mathbf{I}_\ell}{2}[(\mathbf{Z}_\ell + \mathbf{Z}_0)e^{\gamma d} + (\mathbf{Z}_\ell - \mathbf{Z}_0)e^{-\gamma d}] \qquad (3\text{-}4\text{-}16)$$

$$\mathbf{I} = \frac{\mathbf{I}_\ell}{2\mathbf{Z}_0}[(\mathbf{Z}_\ell + \mathbf{Z}_0)e^{\gamma d} - (\mathbf{Z}_\ell - \mathbf{Z}_0)e^{-\gamma d}] \qquad (3\text{-}4\text{-}17)$$

Next, the line impedance at any point from the receiving end in terms of $\mathbf{Z}_\ell$ and $\mathbf{Z}_0$ is expressed as

$$\mathbf{Z} = \mathbf{Z}_0 \frac{(\mathbf{Z}_\ell + \mathbf{Z}_0)e^{\gamma d} + (\mathbf{Z}_\ell - \mathbf{Z}_0)e^{-\gamma d}}{(\mathbf{Z}_\ell + \mathbf{Z}_0)e^{\gamma d} - (\mathbf{Z}_\ell - \mathbf{Z}_0)e^{-\gamma d}} \qquad (3\text{-}4\text{-}18)$$

The line impedance at the sending end can also be found from Eq. (3-4-18) by letting $d = \ell$:

$$\mathbf{Z}_s = \mathbf{Z}_0 \frac{(\mathbf{Z}_\ell + \mathbf{Z}_0)e^{\gamma \ell} + (\mathbf{Z}_\ell - \mathbf{Z}_0)e^{-\gamma \ell}}{(\mathbf{Z}_\ell + \mathbf{Z}_0)e^{\gamma \ell} - (\mathbf{Z}_\ell - \mathbf{Z}_0)e^{-\gamma \ell}} \qquad (3\text{-}4\text{-}19)$$

It is a tedious task to solve Eqs. (3-4-10), (3-4-11), (3-4-18), or (3-4-19) for the line impedance. These equations can be simplified by replacing the exponential factors with either hyperbolic functions or circular functions. The hyperbolic functions are obtained from

$$e^{\pm \gamma z} = \cosh(\gamma z) \pm \sinh(\gamma z) \qquad (3\text{-}4\text{-}20)$$

Substitution of the hyperbolic functions in Eq. (3-4-10) yields the line impedance at

any point from the sending end in terms of the hyperbolic functions:

$$\mathbf{Z} = \mathbf{Z}_0 \frac{\mathbf{Z}_s \cosh{(\gamma z)} - \mathbf{Z}_0 \sinh{(\gamma z)}}{\mathbf{Z}_0 \cosh{(\gamma z)} - \mathbf{Z}_s \sinh{(\gamma z)}} = \mathbf{Z}_0 \frac{\mathbf{Z}_s - \mathbf{Z}_0 \tanh{(\gamma z)}}{\mathbf{Z}_0 - \mathbf{Z}_s \tanh{(\gamma z)}} \quad (3\text{-}4\text{-}21)$$

Similarly, substitution of the hyperbolic functions in Eq. (3-4-18) yields the line impedance from the receiving end in terms of the hyperbolic function:

$$\mathbf{Z} = \mathbf{Z}_0 \frac{\mathbf{Z}_\ell \cosh{(\gamma d)} + \mathbf{Z}_0 \sinh{(\gamma d)}}{\mathbf{Z}_0 \cosh{(\gamma d)} + \mathbf{Z}_\ell \sinh{(\gamma d)}} = \mathbf{Z}_0 \frac{\mathbf{Z}_\ell + \mathbf{Z}_0 \tanh{(\gamma d)}}{\mathbf{Z}_0 + \mathbf{Z}_\ell \tanh{(\gamma d)}} \quad (3\text{-}4\text{-}22)$$

For a lossless line, $\gamma = j\beta$; and by using the following relationships between hyperbolic and circular functions

$$\sinh{(j\beta z)} = j \sin{(\beta z)} \quad (3\text{-}4\text{-}23)$$

$$\cosh{(j\beta z)} = \cos{(\beta z)} \quad (3\text{-}4\text{-}24)$$

the impedance of a lossless transmission line ($\mathbf{Z}_0 = R_0$) can be expressed in terms of the circular functions:

$$\mathbf{Z} = R_0 \frac{\mathbf{Z}_s \cos{(\beta z)} - jR_0 \sin{(\beta z)}}{R_0 \cos{(\beta z)} - j\mathbf{Z}_s \sin{(\beta z)}} = R_0 \frac{\mathbf{Z}_s - jR_0 \tan{(\beta z)}}{R_0 - j\mathbf{Z}_s \tan{(\beta z)}} \quad (3\text{-}4\text{-}25)$$

and

$$\mathbf{Z} = R_0 \frac{\mathbf{Z}_\ell \cos{(\beta d)} + jR_0 \sin{(\beta d)}}{R_0 \cos{(\beta d)} + j\mathbf{Z}_\ell \sin{(\beta d)}} = R_0 \frac{\mathbf{Z}_\ell + jR_0 \tan{(\beta d)}}{R_0 + j\mathbf{Z}_\ell \tan{(\beta d)}} \quad (3\text{-}4\text{-}26)$$

Impedance in terms of reflection coefficient or standing-wave ratio.
Rearrangement of Eq. (3-4-18) gives the line impedance—looking at it from the receiving end—as

$$\mathbf{Z} = \mathbf{Z}_0 \frac{1 + \mathbf{\Gamma}_\ell e^{-2\gamma d}}{1 - \mathbf{\Gamma}_\ell e^{-2\gamma d}} \quad (3\text{-}4\text{-}27)$$

in which the following substitution is made by

$$\mathbf{\Gamma}_\ell = \frac{\mathbf{Z}_\ell - \mathbf{Z}_0}{\mathbf{Z}_\ell + \mathbf{Z}_0} \quad (3\text{-}4\text{-}28)$$

From Eq. (3-2-12) the reflection coefficient at a distance d from the receiving end is given by

$$\mathbf{\Gamma} = \mathbf{\Gamma}_\ell e^{-2\gamma d} = |\mathbf{\Gamma}_\ell| e^{-2\alpha d} e^{j(\theta_\ell - 2\beta d)} \quad (3\text{-}4\text{-}29)$$

Then the simple equation for the line impedance at a distance d from the load is expressed by

$$\mathbf{Z} = \mathbf{Z}_0 \frac{1 + \mathbf{\Gamma}}{1 - \mathbf{\Gamma}} \quad (3\text{-}4\text{-}30)$$

The reflected coefficient is usually a complex quantity and can be written

$$\mathbf{\Gamma} = |\mathbf{\Gamma}| e^{j\phi} \quad (3\text{-}4\text{-}31)$$

where $|\Gamma| = |\Gamma_\ell| e^{-2\alpha d}$

$\phi = \theta_\ell - 2\beta d$

The impedance variation along a lossless line can be found as follows:

$$\mathbf{Z}(d) = \mathbf{Z}_0 \frac{1 + |\Gamma| e^{j\phi}}{1 - |\Gamma| e^{j\phi}} = R_0 \frac{1 + |\Gamma|(\cos\phi + j\sin\phi)}{1 - |\Gamma|(\cos\phi + j\sin\phi)}$$

$$= R(d) + jX(d) = |\mathbf{Z}(d)| e^{j\theta_d} \tag{3-4-32}$$

where $|\mathbf{Z}(d)| = R_0 \sqrt{\dfrac{1 + 2|\Gamma|\cos\phi + |\Gamma|^2}{1 - 2|\Gamma|\cos\phi + |\Gamma|^2}}$ $\tag{3-4-33}$

$$R(d) = R_0 \frac{1 - |\Gamma|^2}{1 - 2|\Gamma|\cos\phi + |\Gamma|^2} \tag{3-4-34}$$

$$X(d) = R_0 \frac{2|\Gamma|\sin\phi}{1 - 2|\Gamma|\cos\phi + |\Gamma|^2} \tag{3-4-35}$$

$$\theta_d = \arctan\left(\frac{X}{R}\right) = \arctan\left(\frac{2|\Gamma|\sin\phi}{1 - |\Gamma|^2}\right) \tag{3-4-36}$$

Since $\phi = \theta_\ell - 2\beta d$, then $\phi = \theta_\ell - 2\pi$ if $\beta d = \pi$. However, $\cos(\theta_\ell - 2\pi) = \cos\theta_\ell$ and $\sin(\theta_\ell - 2\pi) = \sin\theta_\ell$; then

$$\mathbf{Z}(d) = \mathbf{Z}\left(d + \frac{\pi}{\beta}\right) = \mathbf{Z}\left(d + \frac{\lambda}{2}\right) \tag{3-4-37}$$

It is concluded that the impedance along a lossless line will be repeated for every interval at a half-wavelength distance.

Furthermore, the magnitude of a reflection coefficient $|\Gamma|$ is related to the standing-wave ratio ρ by

$$|\Gamma| = \frac{\rho - 1}{\rho + 1} \quad \text{and} \quad \rho = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{3-4-38}$$

The line impedance at any location from the receiving end can be written

$$\mathbf{Z} = R_0 \frac{(\rho + 1) + (\rho - 1)e^{j\phi}}{(\rho + 1) - (\rho - 1)e^{j\phi}} \tag{3-4-39}$$

This is a very useful equation for determining the line impedance in terms of standing-wave ratio ρ, since ρ can easily be measured by a detector or a standing-wave meter.

Determination of characteristic impedance.   A common procedure for determining the characteristic impedance and propagation constant of a given transmission line is to take two measurements:

1. Measure the sending-end impedance with the receiving end short-circuited and record the result:

$$\mathbf{Z}_{sc} = \mathbf{Z}_0 \tanh(\gamma\ell) \tag{3-4-40}$$

2. Measure the sending-end impedance with the receiving end open-circuited and record the result:

$$\mathbf{Z}_{oc} = \mathbf{Z}_0 \coth (\gamma \ell) \tag{3-4-41}$$

Then the characteristic impedance of the measured transmission line is given by

$$\mathbf{Z}_0 = \sqrt{\mathbf{Z}_{sc} \mathbf{Z}_{oc}} \tag{3-4-42}$$

and the propagation constant of the line can be computed from

$$\gamma = \alpha + j\beta = \frac{1}{\ell} \operatorname{arctanh} \sqrt{\frac{\mathbf{Z}_{sc}}{\mathbf{Z}_{oc}}} \tag{3-4-43}$$

Normalized impedance.   The normalized impedance of a transmission line is defined as

$$\mathbf{z} \equiv \frac{\mathbf{Z}}{\mathbf{Z}_0} = \frac{1 + \mathbf{\Gamma}}{1 - \mathbf{\Gamma}} = r \pm jx \tag{3-4-44}$$

It should be noted that the lowercase letters are commonly designated for normalized quantities in describing the distributed transmission-line circuits.

An examination of Eqs. (3-4-39), (3-4-40), and (3-4-44) shows that the normalized impedance for a lossless line has the following significant features:

1. The maximum normalized impedance is

$$z_{max} = \frac{Z_{max}}{R_0} = \frac{|V_{max}|}{R_0|I_{min}|} = \frac{1 + |\mathbf{\Gamma}|}{1 - |\mathbf{\Gamma}|} = \rho \tag{3-4-45}$$

Here z_{max} is a positive real value and it is equal to the standing-wave ratio ρ at the location of any maximum voltage on the line.

2. The minimum normalized impedance is

$$z_{min} = \frac{Z_{min}}{R_0} = \frac{|V_{min}|}{R_0|I_{max}|} = \frac{1 - |\mathbf{\Gamma}|}{1 + |\mathbf{\Gamma}|} = \frac{1}{\rho} \tag{3-4-46}$$

Here z_{min} is a positive real number also but equals the reciprocal of the standing-wave ratio at the location of any minimum voltage on the line.

3. For every interval of a half-wavelength distance along the line, z_{max} or z_{min} is repeated:

$$z_{max}(z) = z_{max}\left(z \pm \frac{\lambda}{2}\right) \tag{3-4-47}$$

$$z_{min}(z) = z_{min}\left(z \pm \frac{\lambda}{2}\right) \tag{3-4-48}$$

4. Since V_{max} and V_{min} are separated by a quarter-wavelength, z_{max} is equal to the reciprocal of z_{min} for every $\lambda/4$ separation:

$$z_{max}\left(z \pm \frac{\lambda}{4}\right) = \frac{1}{z_{min}(z)} \tag{3-4-49}$$

3-4-2 Line Admittance

When a transmission line is branched, it is better to solve the line equations for the line voltage, current, and transmitted power in terms of admittance rather than impedance. The characteristic admittance and the generalized admittance are defined as

$$\mathbf{Y}_0 = \frac{1}{\mathbf{Z}_0} = G_0 \pm jB_0 \tag{3-4-50}$$

$$\mathbf{Y} = \frac{1}{\mathbf{Z}} = G \pm jB \tag{3-4-51}$$

Then the normalized admittance can be written

$$y \equiv \frac{\mathbf{Y}}{\mathbf{Y}_0} = \frac{\mathbf{Z}_0}{\mathbf{Z}} = \frac{1}{\mathbf{z}} = g \pm jb \tag{3-4-52}$$

Example 3-4-1:   Line Impedance

A lossless line has a characteristic impedance of 50 Ω and is terminated in a load resistance of 75 Ω. The line is energized by a generator which has an output impedance of 50 Ω and an open-circuit output voltage of 30 V (rms). The line is assumed to be 2.25 wavelengths long. Determine:

a. The input impedance
b. The magnitude of the instantaneous load voltage
c. The instantaneous power delivered to the load

Solution

a. From Eq. (3-4-26) the line that is 2.25 wavelengths long looks like a quarter-wave line. Then

$$\beta d = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{4} = \frac{\pi}{2}$$

From Eq. (3-4-26) the input impedance is

$$\mathbf{Z}_{in} = \frac{R_0^2}{R_\ell} = \frac{(50)^5}{75} = 33.33 \ \Omega$$

b. The reflection coefficient is

$$\Gamma_\ell = \frac{R_\ell - R_0}{R_\ell + R_0} = \frac{75 - 50}{75 + 50} = 0.20$$

Then the instantaneous voltage at the load is

$$\mathbf{V}_\ell = \mathbf{V}_+ e^{-j\beta\ell}(1 + \Gamma_\ell) = 30(1 + 0.20) = 36 \ \text{V}$$

c. The instantaneous power delivered to the load is

$$\mathbf{P}_\ell = \frac{(36)^2}{75} = 17.28 \ \text{W}$$

3-5 SMITH CHART

Many of the computations required to solve transmission-line problems involve the use of rather complicated equations. The solution of such problems is tedious and difficult because the accurate manipulation of numerous equations is necessary. To simplify their solution, we need a graphic method of arriving at a quick answer.

A number of impedance charts have been designed to facilitate the graphic solution of transmission-line problems. Basically all the charts are derived from the fundamental relationships expressed in the transmission equations. The most popular chart is that developed by Phillip H. Smith [1]. The purpose of this section is to present the graphic solutions of transmission-line problems by using the Smith chart.

The Smith chart consists of a plot of the normalized impedance or admittance with the angle and magnitude of a generalized complex reflection coefficient in a unity circle. The chart is applicable to the analysis of a lossless line as well as a lossy line. By simple rotation of the chart, the effect of the position on the line can be determined. To see how a Smith chart works, consider the equation of reflection coefficient at the load for a transmission line as shown in Eq. (3-2-8):

$$\Gamma_\ell = \frac{Z_\ell - Z_0}{Z_\ell + Z_0} = |\Gamma_\ell| e^{j\theta_\ell} = \Gamma_\gamma + j\Gamma_i \qquad (3\text{-}5\text{-}1)$$

Since $|\Gamma_\ell| \leq 1$, the value of Γ_ℓ must lie on or within the unity circle with a radius of 1. The reflection coefficient at any other location along a line as shown in Eq. (3-2-12) is

$$\Gamma_d = \Gamma_\ell e^{-2\alpha d} e^{-j2\beta d} = |\Gamma_\ell| e^{-2\alpha d} e^{j(\theta_\ell - 2\beta d)} \qquad (3\text{-}5\text{-}2)$$

which is also on or within the unity circle. Figure 3-5-1 shows circles for a constant reflection coefficient Γ and constant electrical-length radials βd.

From Eqs. (3-4-29) and (3-4-44) the normalized impedance along a line is given by

$$z = \frac{Z}{Z_0} = \frac{1 + \Gamma_\ell e^{-2\gamma d}}{1 - \Gamma_\ell e^{-2\gamma d}} \qquad (3\text{-}5\text{-}3)$$

With no loss in generality, it is assumed that $d = 0$; then

$$z = \frac{1 + \Gamma_\ell}{1 - \Gamma_\ell} = \frac{Z_\ell}{Z_0} = \frac{R_\ell + jX_\ell}{Z_0} = r + jx \qquad (3\text{-}5\text{-}4)$$

and

$$\Gamma_\ell = \frac{z - 1}{z + 1} = \Gamma_r + j\Gamma_i \qquad (3\text{-}5\text{-}5)$$

Substitution of Eq. (3-5-5) into Eq. (3-5-4) yields

$$r = \frac{1 - \Gamma_r^2 - \Gamma_i^2}{(1 - \Gamma_r)^2 + \Gamma_i^2} \qquad (3\text{-}5\text{-}6)$$
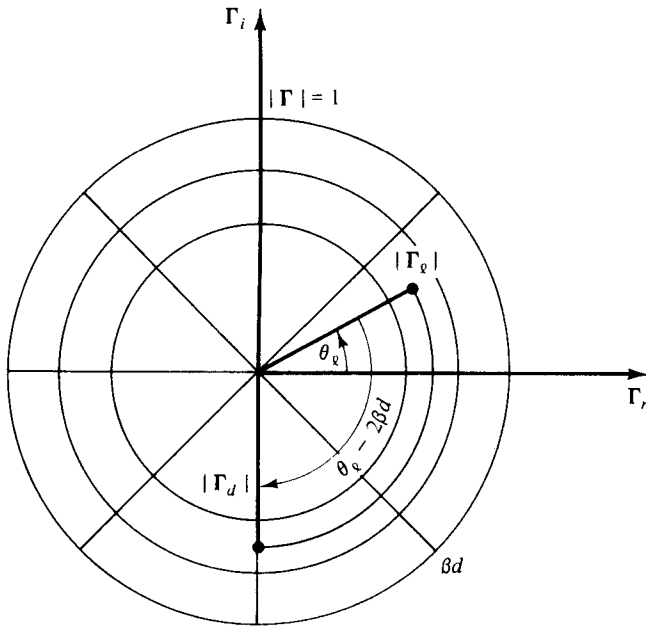
Figure 3-5-1 Constant Γ circles and electrical-length radials βd.

and

$$x = \frac{2\Gamma_i}{(1 - \Gamma_r)^2 + \Gamma_i^2} \qquad (3\text{-}5\text{-}7)$$

Equations (3-5-6) and (3-5-7) can be rearranged as

$$\left(\Gamma_r - \frac{r}{1+r}\right)^2 + \Gamma_i^2 = \left(\frac{1}{1+r}\right)^2 \qquad (3\text{-}5\text{-}8)$$

and

$$(\Gamma_r - 1)^2 + \left(\Gamma_i - \frac{1}{x}\right)^2 = \left(\frac{1}{x}\right)^2 \qquad (3\text{-}5\text{-}9)$$

Equation (3-5-8) represents a family of circles in which each circle has a constant resistance r. The radius of any circle is $1/(1 + r)$, and the center of any circle is $r/(1 + r)$ along the real axis in the unity circle, where r varies from zero to infinity. All constant resistance circles are plotted in Fig. 3-5-2 according to Eq. (3-5-8).

Equation (3-5-9) also describes a family of circles, but each of these circles specifies a constant reactance x. The radius of any circle is $(1/x)$, and the center of any circle is at

$$\Gamma_r = 1 \qquad \Gamma_i = \frac{1}{x} \qquad (\text{where } -\infty \le x \le \infty)$$

All constant reactance circles are plotted in Fig. 3-5-3 according to Eq. (3-5-9).
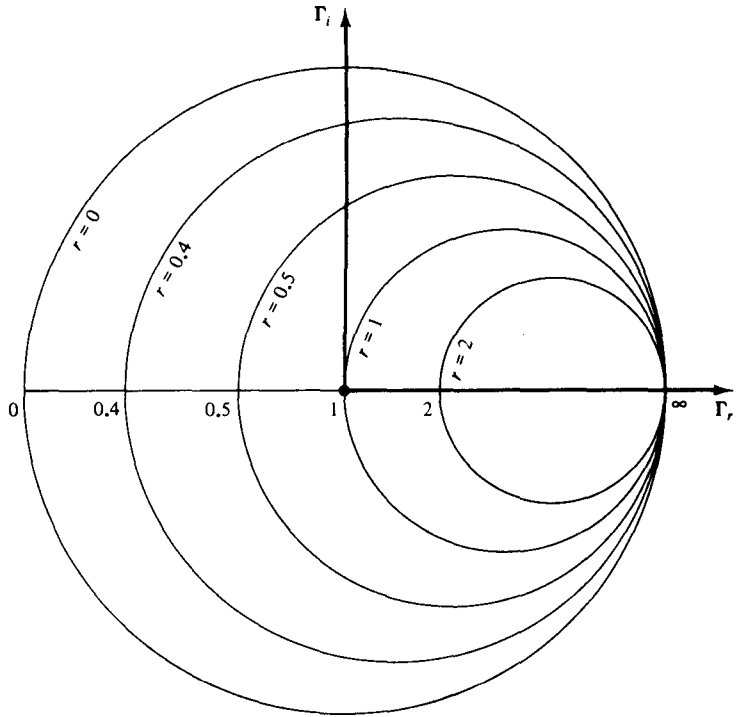
Figure 3-5-2   Constant resistance r circles.
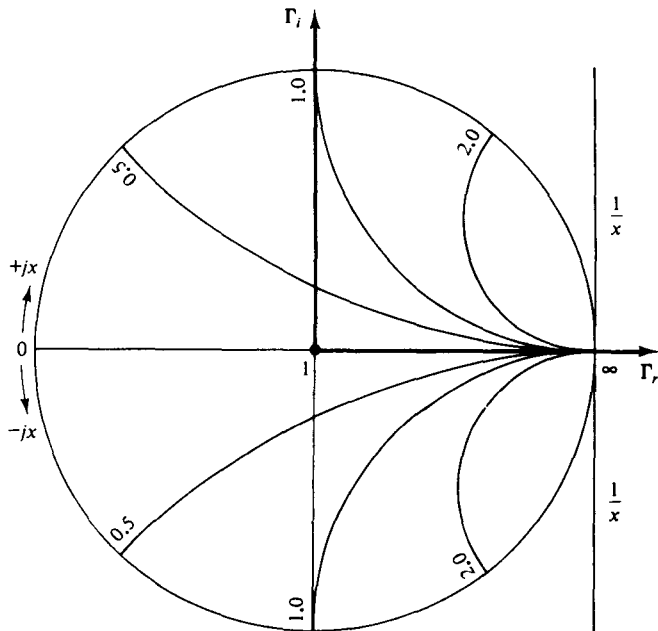
Figure 3-5-3   Constant reactance x circles.

There are relative distance scales in wavelength along the circumference of the Smith chart. Also, there is a phase scale specifying the angle of the reflection coefficient. When a normalized impedance **z** is located on the chart, the normalized impedance of any other location along the line can be found by use of Eq. (3-5-3):

$$\mathbf{z} = \frac{1 + \mathbf{\Gamma}_\ell e^{-2\gamma d}}{1 - \mathbf{\Gamma}_\ell e^{-2\gamma d}} \qquad (3\text{-}5\text{-}10)$$

where

$$\mathbf{\Gamma}_\ell e^{-2\gamma d} = |\mathbf{\Gamma}_\ell| e^{-2\alpha d} e^{j(\theta_\ell - 2\beta d)} \qquad (3\text{-}5\text{-}11)$$

The Smith chart may also be used for normalized admittance. This is evident since

$$\mathbf{Y}_0 = \frac{1}{\mathbf{Z}_0} = G_0 + jB_0 \quad \text{and} \quad \mathbf{Y} = \frac{1}{\mathbf{Z}} = G + jB \qquad (3\text{-}5\text{-}12)$$

Then the normalized admittance is

$$y = \frac{\mathbf{Y}}{\mathbf{Y}_0} = \frac{\mathbf{Z}_0}{\mathbf{Z}} = \frac{1}{\mathbf{z}} = g + jb \qquad (3\text{-}5\text{-}13)$$

Figure 3-5-4 shows a Smith chart which superimposes Figs. 3-5-2 and 3-5-3 into one chart. The characteristics of the Smith chart are summarized as follows:

1. The constant r and constant x loci form two families of orthogonal circles in the chart.
2. The constant r and constant x circles all pass through the point ($\Gamma_r = 1$, $\Gamma_i = 0$).
3. The upper half of the diagram represents $+jx$.
4. The lower half of the diagram represents $-jx$.
5. For admittance the constant r circles become constant g circles, and the constant x circles become constant susceptance b circles.
6. The distance around the Smith chart once is one-half wavelength ($\lambda/2$).
7. At a point of $z_{min} = 1/\rho$, there is a V_{min} on the line.
8. At a point of $z_{max} = \rho$, there is a V_{max} on the line.
9. The horizontal radius to the right of the chart center corresponds to V_{max}, I_{min}, z_{max}, and ρ (SWR).
10. The horizontal radius to the left of the chart center corresponds to V_{min}, I_{max}, z_{min}, and $1/\rho$.
11. Since the normalized admittance **y** is a reciprocal of the normalized impedance **z**, the corresponding quantities in the admittance chart are 180° out of phase with those in the impedance chart.
12. The normalized impedance or admittance is repeated for every half wavelength of distance.
13. The distances are given in wavelengths toward the generator and also toward the load.

IMPEDANCE OR ADMITTANCE COORDINATES
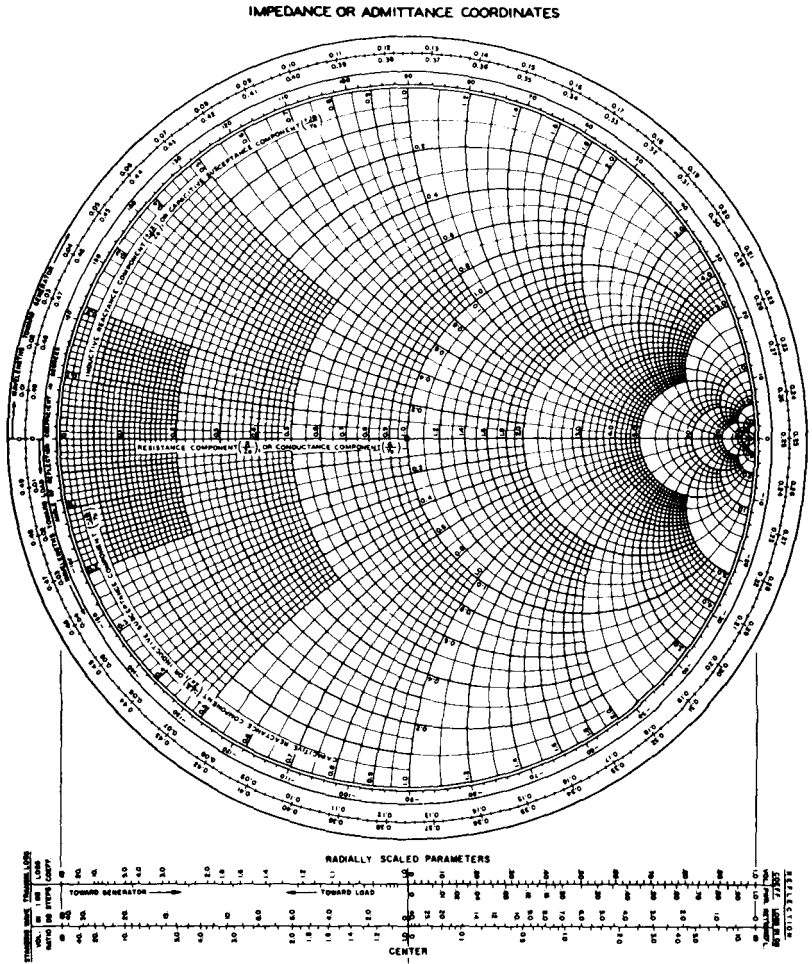
Figure 3-5-4   Smith chart.

The magnitude of the reflection coefficient is related to the standing-wave ratio by the following expression:

$$|\Gamma| = \frac{\rho - 1}{\rho + 1} \tag{3-5-14}$$

A Smith chart or slotted line can be used to measure a standing-wave pattern directly and then the magnitudes of the reflection coefficient, reflected power, transmitted power, and the load impedance can be calculated from it. The use of the Smith chart is illustrated in the following examples.

Example 3-5-1:   Location Determination of Voltage Maximum and Minimum from Load

Given the normalized load impedance $z_\ell = 1 + j1$ and the operating wavelength
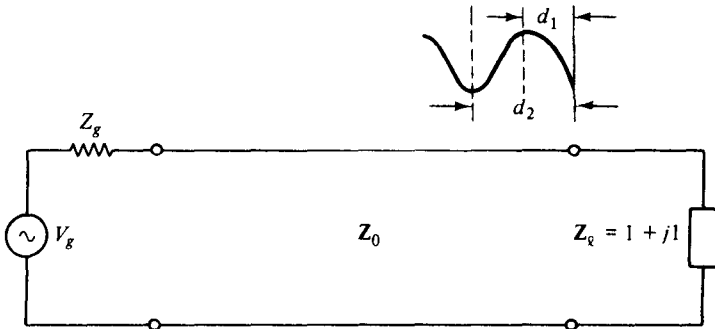
Figure 3-5-5   Diagram for Example 3-5-1.

$\lambda = 5$ cm, determine the first V_{max}, first V_{min} from the load, and the VSWR ρ as shown in Fig. 3-5-5.

Solution

1. Enter $z_\ell = 1 + j1$ on the chart as shown in Fig. 3-5-6.
2. Read 0.162λ on the distance scale by drawing a dashed-straight line from the center of the chart through the load point and intersecting the distance scale.
3. Move a distance from the point at 0.162λ toward the generator and first stop at the voltage maximum on the right-hand real axis at 0.25λ. Then

$$d_1(V_{max}) = (0.25 - 0.162)\lambda = (0.088)(5) = 0.44 \text{ cm}$$

4. Similarly, move a distance from the point of 0.162λ toward the generator and first stop at the voltage minimum on the left-hand real axis at 0.5λ. Then

$$d_2(V_{min}) = (0.5 - 0.162)\lambda = (0.338)(5) = 1.69 \text{ cm}$$
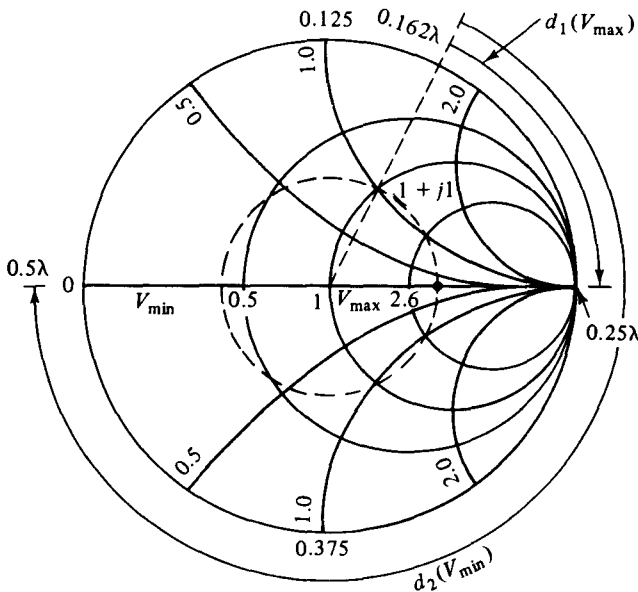
Figure 3-5-6   Graphic solution for Example 3-5-1.

5. Make a standing-wave circle with the center at $(1, 0)$ and pass the circle through the point of $1 + j1$. The location intersected by the circle at the right portion of the real axis indicates the SWR. This is $\rho = 2.6$.

Example 3-5-2: Impedance Determination with Short-Circuit Minima Shift

The location of a minimum instead of a maximum is usually specified because it can be determined more accurately. Suppose that the characteristic impedance of the line R_0 is 50 Ω, and the SWR $\rho = 2$ when the line is loaded. When the load is shorted, the minima shift 0.15λ toward the load. Determine the load impedance. Figure 3-5-7 shows the diagram for the example.
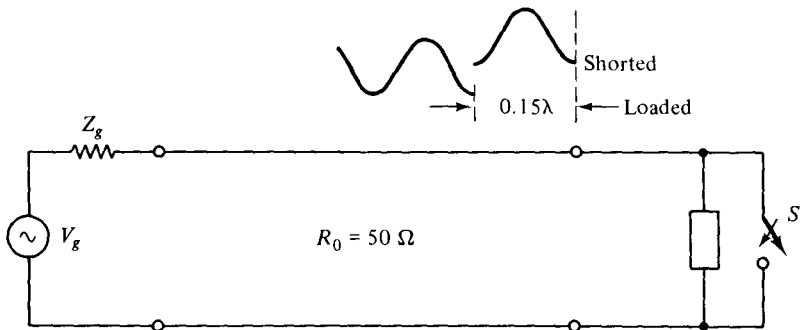
Figure 3-5-7 Diagram for Example 3-5-2.

Solution

1. When the line is shorted, the first voltage minimum occurs at the place of the load as shown in Fig. 3-5-8.
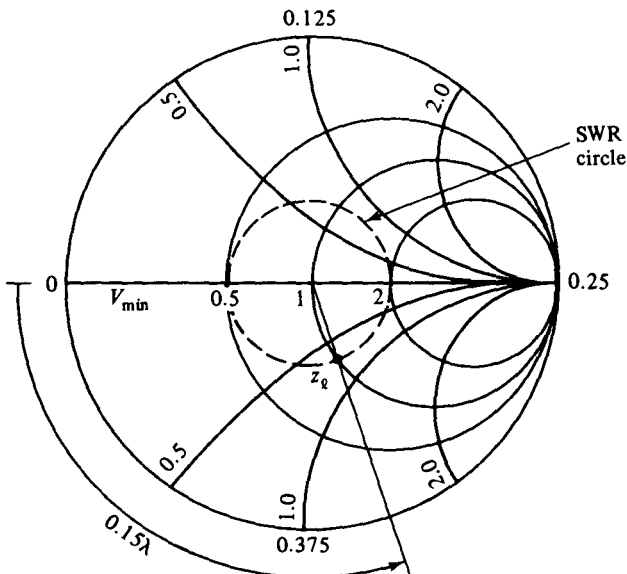
Figure 3-5-8 Graphic solution for Example 3-5-2.

2. When the line is loaded, the first voltage minimum shifts 0.15λ from the load. The distance between two successive minima is one-half wavelength.

3. Plot a SWR circle for $\rho = 2$.

4. Move a distance of 0.15λ from the minimum point along the distance scale toward the load and stop at 0.15λ.

5. Draw a line from this point to the center of the chart.

6. The intersection between the line and the SWR circle is

$$z_\ell = 1 - j0.65$$

7. The load impedance is

$$\mathbf{Z}_\ell = (1 - j0.65)(50) = 50 - j32.5 \ \Omega$$

3-6 IMPEDANCE MATCHING

Impedance matching is very desirable with radio frequency (RF) transmission lines. Standing waves lead to increased losses and frequently cause the transmitter to malfunction. A line terminated in its characteristic impedance has a standing-wave ratio of unity and transmits a given power without reflection. Also, transmission efficiency is optimum where there is no reflected power. A "flat" line is nonresonant; that is, its input impedance always remains at the same value $\mathbf{Z}_0$ when the frequency changes.

Matching a transmission line has a special meaning, one differing from that used in circuit theory to indicate equal impedance seen looking both directions from a given terminal pair for maximum power transfer. In circuit theory, maximum power transfer requires the load impedance to be equal to the complex conjugate of the generator. This condition is sometimes referred to as a *conjugate match*. In transmission-line problems *matching* means simply terminating the line in its characteristic impedance.

A common application of RF transmission lines is the one in which there is a feeder connection between a transmitter and an antenna. Usually the input impedance to the antenna itself is not equal to the characteristic impedance of the line. Furthermore, the output impedance of the transmitter may not be equal to the $\mathbf{Z}_0$ of the line. Matching devices are necessary to flatten the line. A complete matched transmission-line system is shown in Fig. 3-6-1.
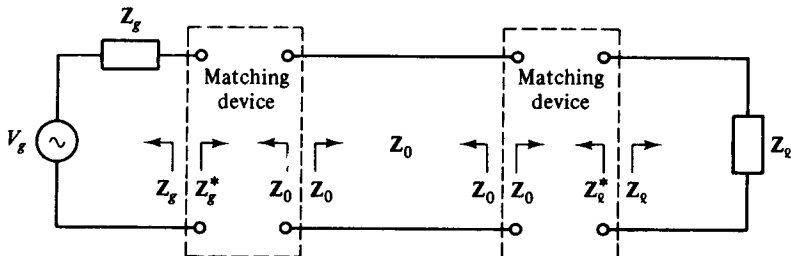
Figure 3-6-1 Matched transmission-line system.

For a low-loss or lossless transmission line at radio frequency, the characteristic impedance $\mathbf{Z}_0$ of the line is resistive. At every point the impedances looking in opposite directions are conjugate. If $\mathbf{Z}_0$ is real, it is its own conjugate. Matching can be tried first on the load side to flatten the line; then adjustment may be made on the transmitter side to provide maximum power transfer. At audio frequencies an iron-cored transformer is almost universally used as an impedance-matching device. Occasionally an iron-cored transformer is also used at radio frequencies. In a practical transmission-line system, the transmitter is ordinarily matched to the coaxial cable for maximum power transfer. Because of the variable loads, however, an impedance-matching technique is often required at the load side.

Since the matching problems involve parallel connections on the transmission line, it is necessary to work out the problems with admittances rather than impedances. The Smith chart itself can be used as a computer to convert the normalized impedance to admittance by a rotation of 180°, as described earlier.

3-6-1 Single-Stub Matching

Although single-lumped inductors or capacitors can match the transmission line, it is more common to use the susceptive properties of short-circuited sections of transmission lines. Short-circuited sections are preferable to open-circuited ones because a good short circuit is easier to obtain than a good open circuit.

For a lossless line with $\mathbf{Y}_g = \mathbf{Y}_0$, maximum power transfer requires $\mathbf{Y}_{11} = \mathbf{Y}_0$, where $\mathbf{Y}_{11}$ is the total admittance of the line and stub looking to the right at point 1-1 (see Fig. 3-6-2). The stub must be located at that point on the line where the real part of the admittance, looking toward the load, is $\mathbf{Y}_0$. In a normalized unit y_{11} must be in the form

$$y_{11} = y_d \pm y_s = 1 \qquad (3\text{-}6\text{-}1)$$
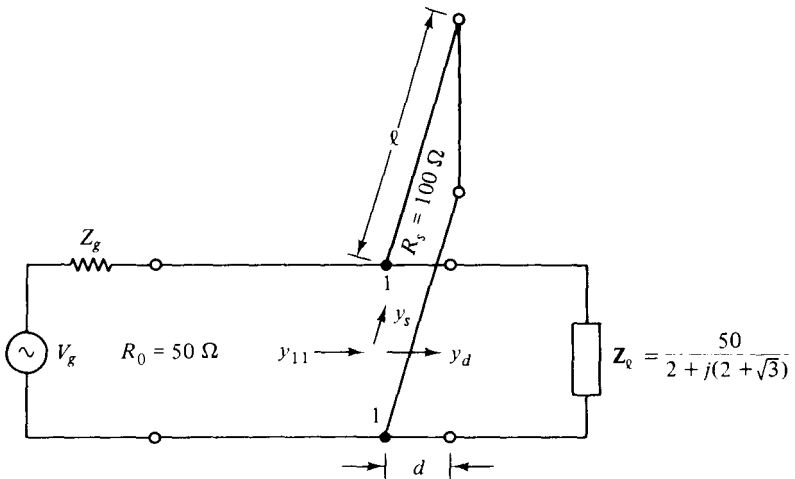
Figure 3-6-2   Single-stub matching for Example 3-6-1.

if the stub has the same characteristic impedance as that of the line. Otherwise

$$\mathbf{Y}_{11} = \mathbf{Y}_d \pm \mathbf{Y}_s = \mathbf{Y}_0 \tag{3-6-2}$$

The stub length is then adjusted so that its susceptance just cancels out the susceptance of the line at the junction.

Example 3-6-1: Single-Stub Matching

A lossless line of characteristic impedance $R_0 = 50 \ \Omega$ is to be matched to a load $\mathbf{Z}_\ell = 50/[2 + j(2 + \sqrt{3})] \ \Omega$ by means of a lossless short-circuited stub. The characteristic impedance of the stub is $100 \ \Omega$. Find the stub position (closest to the load) and length so that a match is obtained.

Solution

1. Compute the normalized load admittance and enter it on the Smith chart (see Fig. 3-6-3).

$$y_\ell = \frac{1}{z_\ell} = \frac{R_0}{\mathbf{Z}_\ell} = 2 + j(2 + \sqrt{3}) = 2 + j3.732$$

2. Draw a SWR circle through the point of y_ℓ so that the circle intersects the unity circle at the point y_d.

$$y_d = 1 - j2.6$$

Note that there are an infinite number of y_d. Take the one that allows the stub to be attached as closely as possible to the load.
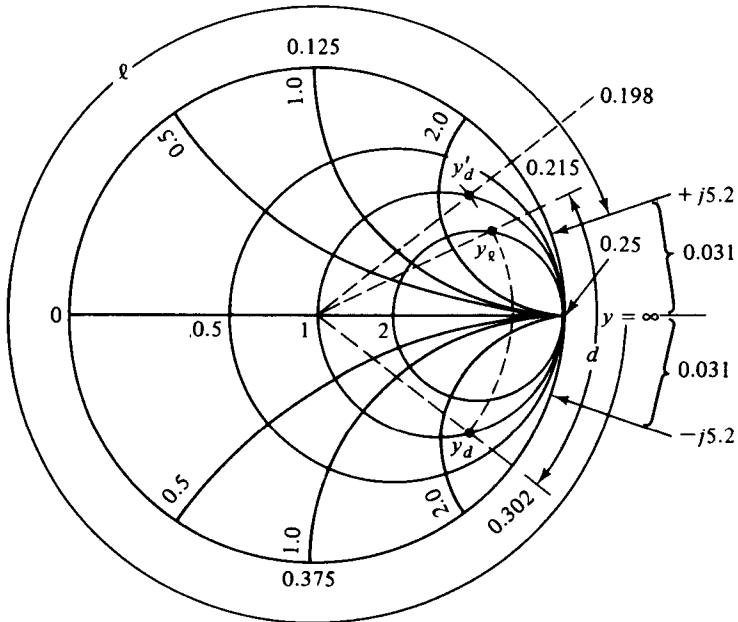
Figure 3-6-3 Graphic solution for Example 3-6-1.

3. Since the characteristic impedance of the stub is different from that of the line, the condition for impedance matching at the junction requires

$$\mathbf{Y}_{11} = \mathbf{Y}_d + \mathbf{Y}_s$$

where $\mathbf{Y}_s$ is the susceptance that the stub will contribute.

It is clear that the stub and the portion of the line from the load to the junction are in parallel, as seen by the main line extending to the generator. The admittances must be converted to normalized values for matching on the Smith chart. Then Eq. (3-6-2) becomes

$$y_{11}\mathbf{Y}_0 = y_d\mathbf{Y}_0 + y_s\mathbf{Y}_{0s}$$

$$y_s = (y_{11} - y_d)\left(\frac{\mathbf{Y}_0}{\mathbf{Y}_{0s}}\right) = [1 - (1 - j2.6)]\frac{100}{50} = +j5.20$$

4. The distance between the load and the stub position can be calculated from the distance scale as

$$d = (0.302 - 0.215)\lambda = 0.087\lambda$$

5. Since the stub contributes a susceptance of $+j5.20$, enter $+j5.20$ on the chart and determine the required distance ℓ from the short-circuited end ($z = 0$, $y = \infty$), which corresponds to the right side of the real axis on the chart, by transversing the chart toward the generator until the point of $+j5.20$ is reached. Then

$$\ell = (0.50 - 0.031)\lambda = 0.469\lambda$$

When a line is matched at the junction, there will be no standing wave in the line from the stub to the generator.

6. If an inductive stub is required,

$$y_d' = 1 + j2.6$$

the susceptance of the stub will be

$$y_s' = -j5.2$$

7. The position of the stub from the load is

$$d' = [0.50 - (0.215 - 0.198)]\lambda = 0.483\lambda$$

and the length of the short-circuited stub is

$$\ell' = 0.031\lambda$$

3-6-2 Double-Stub Matching

Since single-stub matching is sometimes impractical because the stub cannot be placed physically in the ideal location, double-stub matching is needed. Double-stub devices consist of two short-circuited stubs connected in parallel with a fixed length between them. The length of the fixed section is usually one-eighth, three-eighths, or five-eighths of a wavelength. The stub that is nearest the load is used to adjust the susceptance and is located at a fixed wavelength from the constant conductance

unity circle ($g = 1$) on an appropriate constant-standing-wave-ratio circle. Then the admittance of the line at the second stub as shown in Fig. 3-6-4 is

$$y_{22} = y_{d2} \pm y_{s2} = 1 \qquad (3\text{-}6\text{-}3)$$

$$\mathbf{Y}_{22} = \mathbf{Y}_{d2} \pm \mathbf{Y}_{s2} = \mathbf{Y}_0 \qquad (3\text{-}6\text{-}4)$$

In these two equations it is assumed that the stubs and the main line have the same characteristic admittance. If the positions and lengths of the stubs are chosen properly, there will be no standing wave on the line to the left of the second stub measured from the load.
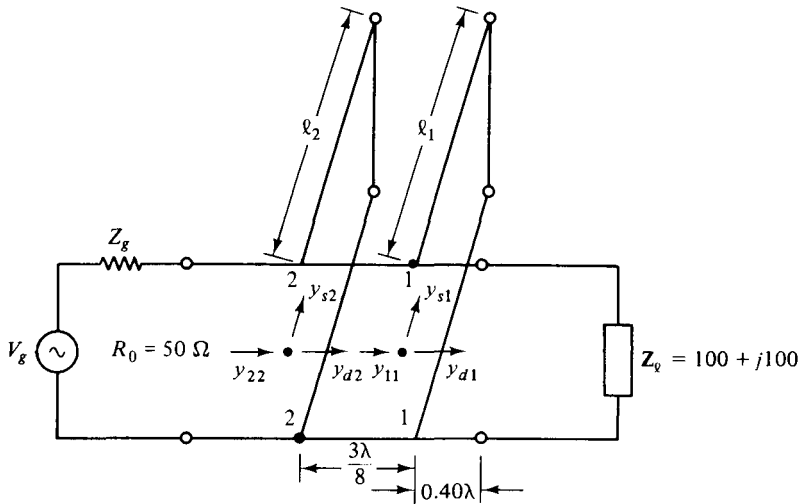
Figure 3-6-4   Double-stub matching for Example 3-6-2.

Example 3-6-2:   Double-Stub Matching

The terminating impedance $\mathbf{Z}_\ell$ is $100 + j100$ Ω, and the characteristic impedance $\mathbf{Z}_0$ of the line and stub is 50 Ω. The first stub is placed at 0.40λ away from the load. The spacing between the two stubs is $\frac{3}{8}\lambda$. Determine the length of the short-circuited stubs when the match is achieved. What terminations are forbidden for matching the line by the double-stub device?

Solution

1. Compute the normalized load impedance z_ℓ and enter it on the chart as shown in Fig. 3-6-5:

$$z_\ell = \frac{100 + j100}{50} = 2 + j2$$

2. Plot a SWR ρ circle and read the normalized load admittance 180° out of phase with z_ℓ on the SWR circle:

$$y_\ell = 0.25 - j0.25$$

3. Draw the spacing circle of $\frac{3}{8}\ell$ by rotating the constant-conductance unity circle
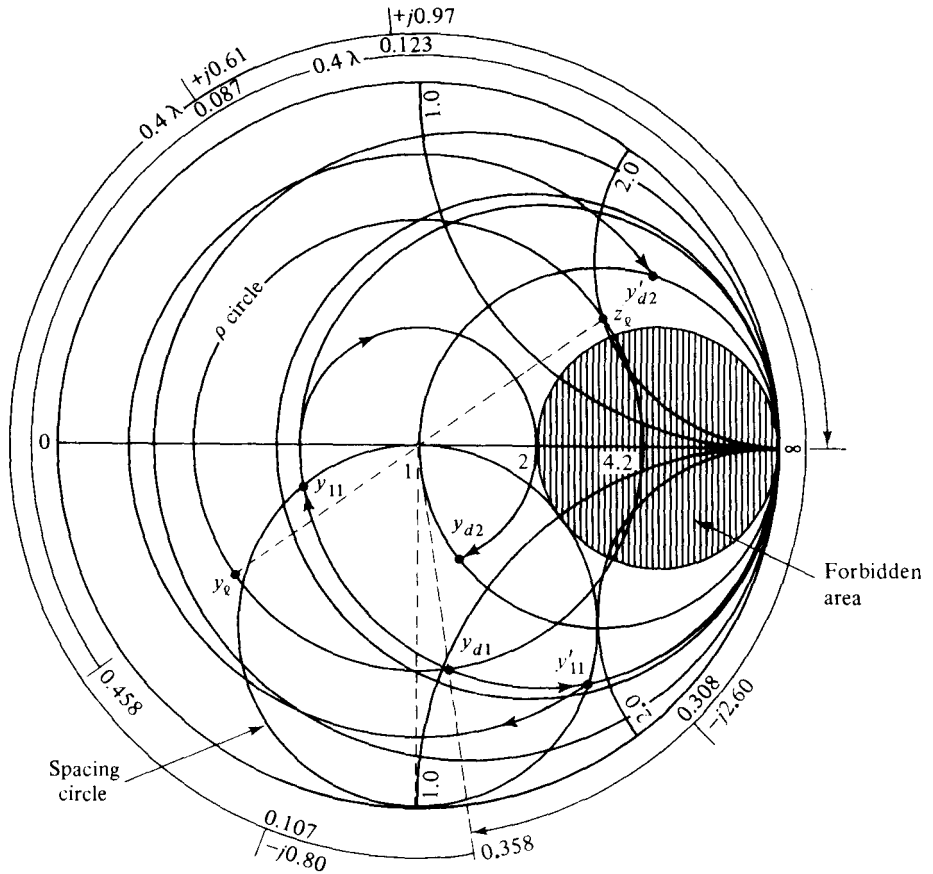
Figure 3-6-5 Graphic solution for Example 3-6-2.

($g = 1$) through a phase angle of $2\beta d = 2\beta \frac{3}{8}\lambda = \frac{3}{2}\pi$ toward the load. Now y_{11} must be on this spacing circle, since y_{d2} will be on the $g = 1$ circle (y_{11} and y_{d2} are $\frac{3}{8}\lambda$ apart).

4. Move y_ℓ for a distance of 0.40λ from 0.458 to 0.358 along the SWR ρ circle toward the generator and read y_{d1} on the chart:

$$y_{d1} = 0.55 - j1.08$$

5. There are two possible solutions for y_{11}. They can be found by carrying y_{d1} along the constant-conductance ($g = 0.55$) circle that intersects the spacing circle at two points:

$$y_{11} = 0.55 - j0.11$$

$$y'_{11} = 0.55 - j1.88$$

6. At the junction 1-1,

$$y_{11} = y_{d1} + y_{s1}$$

Then

$$y_{s1} = y_{11} - y_{d1} = (0.55 - j0.11) - (0.55 - j1.08) = +j0.97$$

Similarly,

$$y'_{s1} = -j.080$$

7. The lengths of stub 1 are found as

$$\ell_1 = (0.25 + 0.123)\lambda = 0.373\lambda$$
$$\ell'_1 = (0.25 - 0.107)\lambda = 0.143\lambda$$

8. The $\frac{3}{8}\lambda$ section of line transforms y_{11} to y_{d2} and y_{11} to y'_{d2} along their constant standing-wave circles, respectively. That is,

$$y_{d2} = 1 - j0.61$$
$$y'_{d2} = 1 + j2.60$$

9. Then stub 2 must contribute

$$y_{s2} = +j0.61$$
$$y'_{s2} = -j2.60$$

10. The lengths of stub 2 are found as

$$\ell_2 = (0.25 + 0.087)\lambda = 0.337\lambda$$
$$\ell'_2 = (0.308 - 0.25)\lambda = 0.058\lambda$$

11. It can be seen from Fig. 3-6-5 that a normalized admittance y_ℓ located inside the hatched area cannot be brought to lie on the locus of y_{11} or y'_{11} for a possible match by the parallel connection of any short-circuited stub because the spacing circle and $g = 2$ circle are mutually tangent. Thus the area of a $g = 2$ circle is called the *forbidden region* of the normalized load admittance for possible match.

Normally the solution of a double-stub-matching problem can be worked out backward from the load toward the generator, since the load is known and the distance of the first stub away from the load can be arbitrarily chosen. In quite a few practical matching problems, however, some stubs have a different $\mathbf{Z}_0$ from that of the line, the length of a stub may be fixed, and so on. So it is hard to describe a definite procedure for solving the double-matching problems.

The flexible coaxial lines are available in different types. Their diameters vary from 0.635 cm (0.25 in.) to about 2.54 cm (1 in.), depending on the power requirement. In some coaxial cables, the inner conductor is stranded or a solid wire, but the outer conductor is a single braid or double. The dielectric material used in these coaxial lines is polyethylene, which has low loss at radio frequencies. Particularly for the RG series, the dielectric is either solid or foam polyethylene. The loss per unit length for foam polyethylene is even appreciably less than the equivalent solid polyethylene.

3-7 MICROWAVE COAXIAL CONNECTORS

For high-frequency operation the average circumference of a coaxial cable must be limited to about one wavelength, in order to reduce multimodal propagation and eliminate erratic reflection coefficients, power losses, and signal distortion. The standardization of coaxial connectors during World War II was mandatory for microwave operation to maintain a low reflection coefficient or a low voltage standing-wave ratio (VSWR). Since that time many modifications and new designs for microwave connectors have been proposed and developed. Seven types of microwave coaxial connectors are described below (see Fig. 3-7-1; TNC is not shown).

1. **APC–3.5**   The APC–3.5 (Amphenol Precision Connector-3.5 mm) was originally developed by Hewlett-Packard, but is now manufactured by Amphenol. The connector provides repeatable connections and has a very low voltage standing-wave ratio (VSWR). Either the male or female end of this 50-ohm connector can mate with the opposite type of SMA connector. The APC–3.5 connector can work at frequencies up to 34 GHz.

2. **APC–7**   The APC–7 (Amphenol Precision Connector-7 mm) was also developed by Hewlett-Packard in the mid 1960s, but it was recently improved and is now manufactured by Amphenol. The connector provides a coupling mechanism without male or female distinction and is the most repeatable connecting device used for very accurate 50-ohm measurement applications. Its VSWR is extremely low, in the range of 1.02 to 18 GHz. Maury Microwave also has an MPC series available.

3. **BNC**   The BNC (Bayonet Navy Connector) was originally designed for military system applications during World War II. The connector operates very well at frequencies up to about 4 GHz; beyond that it tends to radiate electromagnetic energy. The BNC can accept flexible cables with diameters of up to 6.35 mm (0.25 in.) and characteristic impedance of 50 to 75 ohms. It is now the most commonly used connector for frequencies under 1 GHz.

4. **SMA**   The SMA (Sub-Miniature A) connector was originally designed by Bendix Scintilla Corporation, but it has been manufactured by Omni-Spectra Inc. (as the OSM connector) and many other electronic companies. The main application of SMA connectors is on components for microwave systems. The connector is seldom used above 24 GHz because of higher-order modes.

5. **SMC**   The SMC (Sub-Miniature C) is a 50-ohm connector that is smaller than the SMA. The connector is manufactured by Sealectro Corporation and can accept flexible cables with diameters of up to 3.17 mm (0.125 in.) for a frequency range of up to 7 GHz.

6. **TNC**   The TNC (Threaded Navy Connector) is merely a threaded BNC. The function of the thread is to stop radiation at higher frequencies, so that the connector can work at frequencies up to 12 GHz.

7. **Type N**   The Type N (Navy) connector was originally designed for military systems during World War II and is the most popular measurement connector

8473C
DETECTOR—NEG

8473B
DETECTOR—NEG
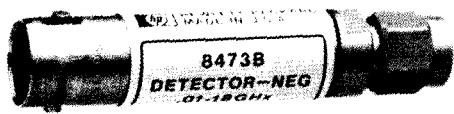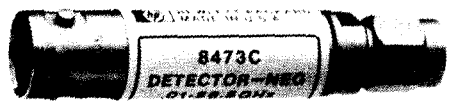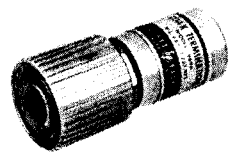
BNC female   APC-3.5 male

(a) APC-3.5

New coupling nut

Old coupling nut

(b) APC-7

8471A
RF DETECTOR

BNC male   BNC female

(c) BNC

8493A   10DB
ATTENUATOR
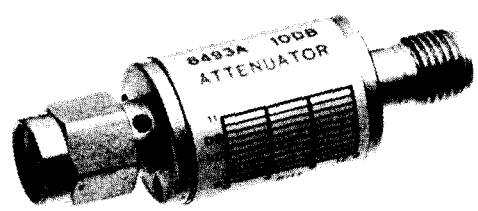
SMA male   SMA female

(d) SMA

APC-3.5
male

DETECTOR
.01–18GHz

DETECTOR
.01–26.5GHz
NEGATIVE

SMC male
(plug)
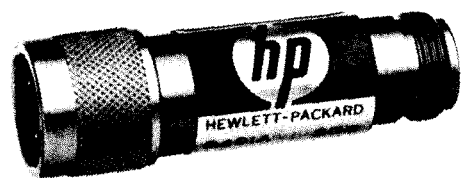
Type N male   Type N female

(e) SMC

(f) Type N

Figure 3-7-1   Microwave coaxial connectors.

for the frequency range of 1 to 18 GHz. It is 50- or 75-ohm connector and its VSWR is extremely low, less than 1.02.

REFERENCES

[1] SMITH, P. H.
    Transmission line calculator. *Electronics*, **12**, 29–31 (1939).
    An improved transmission line calculator. *Electronics*, **17**, 130–133 and 318–325 (1944).
    Smith Charts—Their Development and Use. A series published at intervals by the Kay Electric Co.; No. 1 is dated March 1962, and No. 9 is dated December 1966.

SUGGESTED READINGS

BROWN, R. G., et al., *Lines, Waves, and Antennas,* 2nd ed. John Wiley & Sons, New York, 1970.

DWORSKY, L. N., *Modern Transmission Line Theory and Applications.* John Wiley & Sons, New York, 1979.

LIAO, S. Y., *Engineering Applications of Electromagnetic Theory.* West Publishing Company, St. Paul, Minn., 1988.

RIZZI, PETER A., *Microwave Engineering: Passive Circuits.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.

SINNEMA, W., *Electronic Transmission Technology: Lines, Waves, and Applications.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1979.

PROBLEMS

3-1. A transmission line has a characteristic impedance of 300 Ω and is terminated in a load of $300 - j300$ Ω. The propagation constant of the line is $0.054 + j3.53$ per meter. Determine:
   a. The reflection coefficient at the load
   b. The transition coefficient at the load
   c. The reflection coefficient at a point 2 m away from the load

3-2. A lossless transmission line has a characteristic impedance of 50 Ω and is terminated in a load of 100 Ω. The magnitude of a voltage wave incident to the line is 20 V (rms). Determine:
   a. The VSWR on the line
   b. The maximum voltage V_{max} and minimum voltage V_{min} on the line
   c. The maximum current I_{max} and minimum current I_{min} on the line
   d. The power transmitted by the line

3-3. A lossless line has a characteristic impedance of 75 Ω and is terminated in a load of

300 Ω. The line is energized by a generator which has an open-circuit output voltage of 20 V (rms) and output impedance of 75 Ω. The line is assumed to be $2\frac{1}{4}$ wavelengths long.

a. Find the sending-end impedance.

b. Determine the magnitude of the receiving-end voltage.

c. Calculate the receiving-end power at the load.

3-4. A lossless transmission line has a characteristic impedance of 100 Ω and is terminated in a load of 75 Ω. The line is 0.75 wavelength long. Determine:

a. The sending-end impedance

b. The reactance which, if connected across the sending end of the line, will make the input impedance a pure resistance

3-5. A coaxial line with a solid polyethylene dielectric is to be used at a frequency of 3 GHz. Its characteristic impedance Z_0 is 50 Ω and its attenuation constant α is 0.0156 Np/m. The velocity factor which is defined as the ratio of phase velocity over the velocity of light in free space is 0.660. The line is 100 m long and is terminated in its characteristic impedance. A generator, which has an open-circuit voltage of 50 V (rms) and an internal impedance of 50 Ω, is connected to the sending end of the line. The frequency is tuned at 3 GHz. Compute:

a. The magnitude of the sending-end voltage and of the receiving-end voltage

b. The sending-end power and the receiving-end power

c. The wavelengths of the line

3-6. An open-wire transmission line has $R = 5$ Ω/m, $L = 5.2 \times 10^{-8}$H/m, $G = 6.2 \times 10^{-3}$mho/m, and $C = 2.13 \times 10^{-10}$ F/m. The signal frequency is 4 GHz. Calculate:

a. The characteristic impedance of the line in both rectangular form and polar form

b. The propagation constant of the wave along the line

c. The normalized impedance of a load $100 + j100$

d. The reflection coefficient at the load

e. The sending-end impedance if the line is assumed a quarter-wavelength long

3-7. A quarter-wave lossless line has a characteristic impedance of 50 Ω and is terminated in a load of 100 Ω. The line is energized by a generator of 20 V (rms) with an internal resistance of 50 Ω. Calculate:

a. The sending-end impedance

b. The magnitude of the receiving-end voltage

c. The power delivered to the load.

3-8. A lossless transmission line is terminated in an open circuit. The sending end is energized by a generator which has an open-circuit output voltage of V_g (rms) and an interval impedance equal to the characteristic impedance of the line. Show that the sending-end voltage is equal to the output voltage of the generator.

3-9. A lossless transmission line has a characteristic impedance of 300 Ω and is operated at a frequency of 10 GHz. The observed standing-wave ratio on the line is 5.0. It is proposed to use a short-circuited stub to match a pure resistor load to the line.

a. Determine the distance in centimeters from the load to the place where the stub should be located. (Two possible solutions.)

b. Find the length of the stub in centimeters. (Two possible solutions.)

3-10. A lossless line has a characteristic impedance of 50 Ω and is loaded by $60 - j60$ Ω. One stub is at the load, and the other is at $3\lambda/8$ distance away from the first.

a. Determine the lengths in wavelength of the two short-circuited stubs when a match is achieved.

    b. Locate and crosshatch the *forbidden region* of the normalized admittance for possible match.

3-11. A lossless transmission line has a characteristic impedance of 300 Ω and is terminated by an impedance Z_ℓ. The observed standing-wave ratio on the line is 6, and the distance of the first voltage minimum from the load is 0.166λ.

    a. Determine the load Z_ℓ.

    b. Find the lengths in λ of two shorted stubs, one at the load and one at $\lambda/4$ from the load, which are required to match the load to the line.

3-12. A single-stub tuner is to match a lossless line of 400 Ω to a load of $800 - j300$ Ω. The frequency is 3 GHz.

    a. Find the distance in meters from the load to the tuning stub.

    b. Determine the length in meters of the short-circuited stub.

3-13. A half-wave-dipole antenna has a driving-point impedance of $73 + j42.5$ Ω. A lossless transmission line connected to a TV set has a characteristic impedance of 300 Ω. The problem is to design a shorted stub with the same characteristic impedance to match the antenna to the line. The stub may be placed at a location closest to the antenna. The reception is assumed to be Channel 83 at a frequency of 0.88525 GHz.

    a. Determine the susceptance contributed by the stub.

    b. Calculate the length in centimeters of the stub.

    c. Find the distance in centimeters between the antenna and the point where the stub is placed. [*Note:* There are two sets of solutions.]

3-14. A lossless transmission line has a characteristic impedance of 100 Ω and is loaded by $100 + j100$ Ω. A single shorted stub with the same characteristic impedance is inserted at $\lambda/4$ from the load to match the line. The load current is measured to be 2 A. The length of the stub is $\lambda/8$.

    a. Determine the magnitude and the phase of the voltage across the stub location.

    b. Find the magnitude and the phase of the current flowing through the end of the stub.

3-15. A double-stub matching line is shown in Fig. P3-15. The characteristic resistances of the lossless line and stubs are 100 Ω, respectively. The spacing between the two stubs is $\lambda/8$. The load is $100 + j100$. One stub is located at the load. Determine:

    a. The reactances contributed by the stub

    b. The lengths of the two shorted double-stub tuners [*Note:* There are two sets of solutions.]
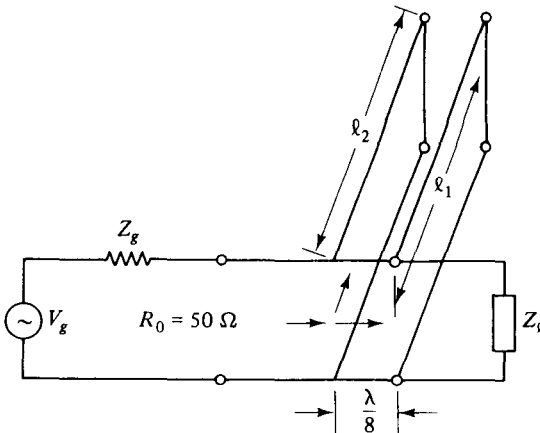
Figure P3-15

3-16. A lossless transmission line has a characteristic impedance Z_0 of 100 Ω and is loaded by an unknown impedance. Its voltage standing-wave ratio is 4 and the first voltage maximum is $\lambda/8$ from the load.

    a. Find the load impedance.

    b. To match the load to the line, a quarter section of a different line with a characteristic impedance $Z_{01} < Z_0$ is to be inserted somewhere between (in cascade with) the load and the original line. Determine the minimum distance between the load and matching section, and calculate the characteristic impedance Z_{01} in terms of Z_0.

3-17. A lossless transmission line has a characteristic impedance of 100 Ω and is loaded by an unknown impedance. The standing-wave ratio along the line is 2. The first two voltage minima are located at $z = -10$ and -35 cm from the load where $z = 0$. Determine the load impedance.

3-18. A matched transmission line is shown in Fig. P3-18.

    a. Find ℓ_1 and d which provide a proper match.

    b. With the line and load properly matched determine the VSWR on the section of line between the stubs.
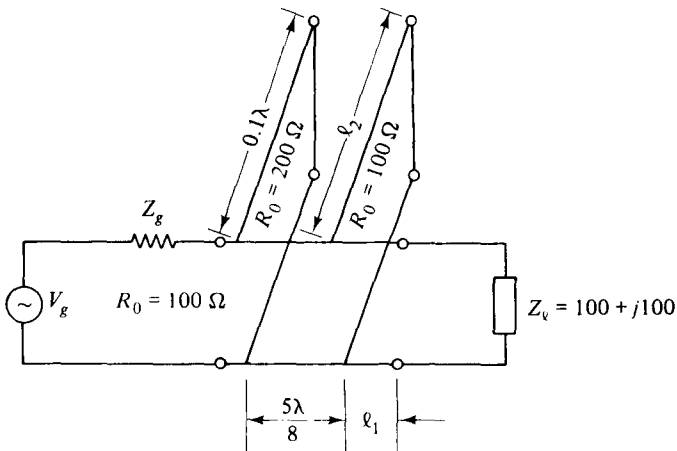
Figure P3-18

Chapter 4

Microwave Waveguides and Components

4-0 INTRODUCTION

In general, a waveguide consists of a hollow metallic tube of a rectangular or circular shape used to guide an electromagnetic wave. Waveguides are used principally at frequencies in the microwave range; inconveniently large guides would be required to transmit radio-frequency power at longer wavelengths. At frequency range X band from 8.00 to 12.0 GHz, for example, the U.S. standard rectangular waveguide WR-90 has an inner width of 2.286 cm (0.9 in.) and an inner height of 1.016 cm (0.4 in.); but its outside dimensions are 2.54 cm (1 in.) wide and 1.27 cm (0.5 in.) high [1].

In waveguides the electric and magnetic fields are confined to the space within the guides. Thus no power is lost through radiation, and even the dielectric loss is negligible, since the guides are normally air-filled. However, there is some power loss as heat in the walls of the guides, but the loss is very small.

It is possible to propagate several modes of electromagnetic waves within a waveguide. These modes correspond to solutions of Maxwell's equations for particular waveguides. A given waveguide has a definite cutoff frequency for each allowed mode. If the frequency of the impressed signal is above the cutoff frequency for a given mode, the electromagnetic energy can be transmitted through the guide for that particular mode without attenuation. Otherwise the electromagnetic energy with a frequency below the cutoff frequency for that particular mode will be attenuated to a negligible value in a relatively short distance. *The dominant mode in a particular guide is the mode having the lowest cutoff frequency.* It is advisable to choose the dimensions of a guide in such a way that, for a given input signal, only the energy of the dominant mode can be transmitted through the guide.

The process of solving the waveguide problems may involve three steps:

1. The desired wave equations are written in the form of either rectangular or cylindrical coordinate systems suitable to the problem at hand.
2. The boundary conditions are then applied to the wave equations set up in step 1.
3. The resultant equations usually are in the form of partial differential equations in either time or frequency domain. They can be solved by using the proper method.

4-1 RECTANGULAR WAVEGUIDES

A rectangular waveguide is a hollow metallic tube with a rectangular cross section. The conducting walls of the guide confine the electromagnetic fields and thereby guide the electromagnetic wave. A number of distinct field configurations or modes can exist in waveguides. When the waves travel longitudinally down the guide, the plane waves are reflected from wall to wall. This process results in a component of either electric or magnetic field in the direction of propagation of the resultant wave; therefore the wave is no longer a *transverse electromagnetic* (TEM) wave. Figure 4-1-1 shows that any uniform plane wave in a lossless guide may be resolved into TE and TM waves.
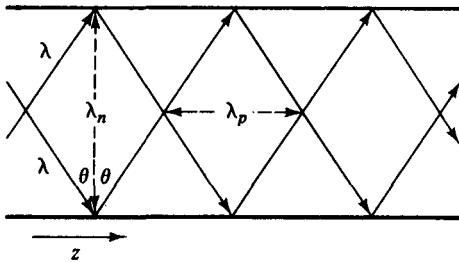
Figure 4-1-1   Plane wave reflected in a waveguide.

It is clear that when the wavelength λ is in the direction of propagation of the incident wave, there will be one component λ_n in the direction normal to the reflecting plane and another λ_p parallel to the plane. These components are

$$\lambda_n = \frac{\lambda}{\cos \theta} \tag{4-1-1}$$

$$\lambda_p = \frac{\lambda}{\sin \theta} \tag{4-1-2}$$

where θ = angle of incidence
       λ = wavelength of the impressed signal in unbounded medium

A plane wave in a waveguide resolves into two components: one standing wave in the direction normal to the reflecting walls of the guide and one traveling wave in the direction parallel to the reflecting walls. In lossless waveguides the modes may be classified as either *transverse electric* (TE) mode or *transverse magnetic* (TM) mode. In rectangular guides the modes are designated TE_{mn} or TM_{mn}. The integer m

denotes the number of half waves of electric or magnetic intensity in the x direction, and n is the number of half waves in the y direction if the propagation of the wave is assumed in the positive z direction.

4-1-1 Solutions of Wave Equations in Rectangular Coordinates

As stated previously, there are time-domain and frequency-domain solutions for each wave equation. However, for the simplicity of the solution to the wave equation in three dimensions plus a time-varying variable, only the sinusoidal steady-state or the frequency-domain solution will be given. A rectangular coordinate system is shown in Fig. 4-1-2.
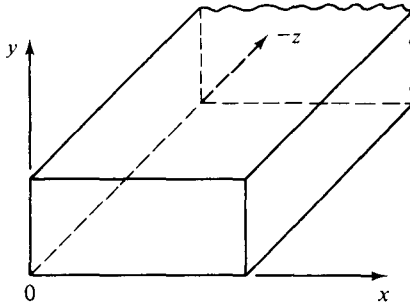
Figure 4-1-2   Rectangular coordinates.

The electric and magnetic wave equations in frequency domain in Eqs. (2-1-20) and (2-1-21) are given by

$$\nabla^2 \mathbf{E} = \gamma^2 \mathbf{E} \qquad (4\text{-}1\text{-}3)$$

$$\nabla^2 \mathbf{H} = \gamma^2 \mathbf{H} \qquad (4\text{-}1\text{-}4)$$

where $\gamma = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)} = \alpha + j\beta$. These are called the *vector wave equations*.

Rectangular coordinates are the usual right-hand system. The rectangular components of $\mathbf{E}$ or $\mathbf{H}$ satisfy the complex scalar wave equation or Helmholtz equation

$$\nabla^2 \psi = \gamma^2 \psi \qquad (4\text{-}1\text{-}5)$$

The Helmholtz equation in rectangular coordinates is

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = \gamma^2 \psi \qquad (4\text{-}1\text{-}6)$$

This is a linear and inhomogeneous partial differential equation in three dimensions. By the method of separation of variables, the solution is assumed in the form of

$$\psi = X(x)Y(y)Z(z) \qquad (4\text{-}1\text{-}7)$$

where $X(x)$ = a function of the x coordinate only
$\quad\ \ Y(y)$ = a function of the y coordinate only
$\quad\ \ Z(z)$ = a function of the z coordinate only

Substitution of Eq. (4-1-7) in Eq. (4-1-6) and division of the resultant by Eq. (4-1-7) yield

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} = \gamma^2 \tag{4-1-8}$$

Since the sum of the three terms on the left-hand side is a constant and each term is independently variable, it follows that each term must be equal to a constant.

Let the three terms be k_x^2, k_y^2, and k_z^2, respectively; then the separation equation is given by

$$-k_x^2 - k_y^2 - k_z^2 = \gamma^2 \tag{4-1-9}$$

The general solution of each differential equation in Eq. (4-1-8)

$$\frac{d^2X}{dx^2} = -k_x^2 X \tag{4-1-10}$$

$$\frac{d^2Y}{dy^2} = -k_y^2 Y \tag{4-1-11}$$

$$\frac{d^2Z}{dz^2} = -k_z^2 Z \tag{4-1-12}$$

will be in the form of

$$X = A \sin(k_x x) + B \cos(k_x x) \tag{4-1-13}$$

$$Y = C \sin(k_y y) + D \cos(k_y y) \tag{4-1-14}$$

$$Z = E \sin(k_z z) + F \cos(k_z z) \tag{4-1-15}$$

The total solution of the Helmholtz equation in rectangular coordinates is

$$\psi = [A \sin(k_x x) + B \cos(k_x x)][C \sin(k_y y) + D \cos(k_y y)]$$
$$\times [E \sin(k_z z) + F \cos(k_z z)] \tag{4-1-16}$$

The propagation of the wave in the guide is conventionally assumed in the positive z direction. It should be noted that the propagation constant γ_g in the guide differs from the intrinsic propagation constant γ of the dielectric. Let

$$\gamma_g^2 = \gamma^2 + k_x^2 + k_y^2 = \gamma^2 + k_c^2 \tag{4-1-17}$$

where $k_c = \sqrt{k_x^2 + k_y^2}$ is usually called the *cutoff wave number*. For a lossless dielectric, $\gamma^2 = -\omega^2 \mu\epsilon$. Then

$$\gamma_g = \pm\sqrt{\omega^2 \mu\epsilon - k_c^2} \tag{4-1-18}$$

There are three cases for the propagation constant γ_g in the waveguide.

Case I.   There will be no wave propagation (evanescence) in the guide if $\omega_c^2 \mu\epsilon = k_c^2$ and $\gamma_g = 0$. This is the critical condition for cutoff propagation. The cutoff frequency is expressed as

$$f_c = \frac{1}{2\pi\sqrt{\mu\epsilon}}\sqrt{k_x^2 + k_y^2} \qquad (4\text{-}1\text{-}19)$$

Case II.   The wave will be propagating in the guide if $\omega^2\mu\epsilon > k_c^2$ and

$$\gamma_g = \pm j\beta_g = \pm j\omega\sqrt{\mu\epsilon}\sqrt{1 - \left(\frac{f_c}{f}\right)^2} \qquad (4\text{-}1\text{-}20)$$

This means that the operating frequency must be above the cutoff frequency in order for a wave to propagate in the guide.

Case III.   The wave will be attenuated if $\omega^2\mu\epsilon < k_c^2$ and

$$\gamma_g = \pm\alpha_g = \pm\omega\sqrt{\mu\epsilon}\sqrt{\left(\frac{f_c}{f}\right)^2 - 1} \qquad (4\text{-}1\text{-}21)$$

This means that if the operating frequency is below the cutoff frequency, the wave will decay exponentially with respect to a factor of $-\alpha_g z$ and there will be no wave propagation because the propagation constant is a real quantity. Therefore the solution to the Helmholtz equation in rectangular coordinates is given by

$$\psi = [A \sin(k_x x) + B \cos(k_x x)][C \sin(k_y y) + D \cos(k_y y)]e^{-j\beta_g z} \qquad (4\text{-}1\text{-}22)$$

4-1-2 TE Modes in Rectangular Waveguides

It has been previously assumed that the waves are propagating in the positive z direction in the waveguide. Figure 4-1-3 shows the coordinates of a rectangular waveguide.

Figure 4-1-3   Coordinates of a rectangular guide.

The TE_{mn} modes in a rectangular guide are characterized by $E_z = 0$. In other words, the z component of the magnetic field, H_z, must exist in order to have energy transmission in the guide. Consequently, from a given Helmholtz equation,

$$\nabla^2 H_z = \gamma^2 H_z \qquad (4\text{-}1\text{-}23)$$

a solution in the form of

$$H_z = \left[A_m \sin\left(\frac{m\pi x}{a}\right) + B_m \cos\left(\frac{m\pi x}{a}\right) \right] \times \left[C_n \sin\left(\frac{n\pi y}{b}\right) + D_n \cos\left(\frac{n\pi y}{b}\right) \right] e^{-j\beta_g z}$$

$$(4\text{-}1\text{-}24)$$

will be determined in accordance with the given boundary conditions, where $k_x = m\pi/a$ and $k_y = n\pi/b$ are replaced. For a lossless dielectric, Maxwell's curl equations in frequency domain are

$$\nabla \times \mathbf{E} = -j\omega\mu\,\mathbf{H} \qquad (4\text{-}1\text{-}25)$$

$$\nabla \times \mathbf{H} = j\omega\epsilon\,\mathbf{E} \qquad (4\text{-}1\text{-}26)$$

In rectangular coordinates, their components are

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -j\omega\mu H_x \qquad (4\text{-}1\text{-}27)$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \qquad (4\text{-}1\text{-}28)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \qquad (4\text{-}1\text{-}29)$$

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = j\omega\epsilon E_x \qquad (4\text{-}1\text{-}30)$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = j\omega\epsilon E_y \qquad (4\text{-}1\text{-}31)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\epsilon E_z \qquad (4\text{-}1\text{-}32)$$

With the substitution $\partial/\partial z = -j\beta_g$ and $E_z = 0$, the foregoing equations are simplified to

$$\beta_g E_y = -\omega\mu H_x \qquad (4\text{-}1\text{-}33)$$

$$\beta_g E_x = \omega\mu H_y \qquad (4\text{-}1\text{-}34)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_z}{\partial y} = -j\omega\mu H_z \qquad (4\text{-}1\text{-}35)$$

$$\frac{\partial H_z}{\partial y} + j\beta_g H_y = j\omega\epsilon E_x \qquad (4\text{-}1\text{-}36)$$

$$-j\beta_g H_x - \frac{\partial H_z}{\partial x} = j\omega\epsilon E_y \qquad (4\text{-}1\text{-}37)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = 0 \qquad (4\text{-}1\text{-}38)$$

Solving these six equations for E_x, E_y, H_x, and H_y in terms of H_z will give the TE-

mode field equations in rectangular waveguides as

$$E_x = \frac{-j\omega\mu}{k_c^2} \frac{\partial H_z}{\partial y} \tag{4-1-39}$$

$$E_y = \frac{j\omega\mu}{k_c^2} \frac{\partial H_z}{\partial x} \tag{4-1-40}$$

$$E_z = 0 \tag{4-1-41}$$

$$H_x = \frac{-j\beta_g}{k_c^2} \frac{\partial H_z}{\partial x} \tag{4-1-42}$$

$$H_y = \frac{-j\beta_g}{k_c^2} \frac{\partial H_z}{\partial y} \tag{4-1-43}$$

$$H_z = \text{Eq. (4-1-24)} \tag{4-1-44}$$

where $k_c^2 = \omega^2\mu\epsilon - \beta_g^2$ has been replaced.

Differentiating Eq. (4-1-24) with respect to x and y and then substituting the results in Eqs. (4-1-39) through (4-1-43) yield a set of field equations. The boundary conditions are applied to the newly found field equations in such a manner that either the tangent **E** field or the normal **H** field vanishes at the surface of the conductor. Since $E_x = 0$, then $\partial H_z/\partial y = 0$ at $y = 0, b$. Hence $C_n = 0$. Since $E_y = 0$, then $\partial H_z/\partial x = 0$ at $x = 0, a$. Hence $A_m = 0$.

It is generally concluded that the normal derivative of H_z must vanish at the conducting surfaces—that is,

$$\frac{\partial H_z}{\partial n} = 0 \tag{4-1-45}$$

at the guide walls. Therefore the magnetic field in the positive z direction is given by

$$H_z = H_{0z} \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-46}$$

where H_{0z} is the amplitude constant.

Substitution of Eq. (4-1-46) in Eqs. (4-1-39) through (4-1-43) yields the TE_{mn} field equations in rectangular waveguides as

$$E_x = E_{0x} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-47}$$

$$E_y = E_{0y} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-48}$$

$$E_z = 0 \tag{4-1-49}$$

$$H_x = H_{0x} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-50}$$

$$H_y = H_{0y} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-51}$$

$$H_z = \text{Eq. (4-1-46)} \tag{4-1-52}$$

where $m = 0, 1, 2, \ldots$
$n = 0, 1, 2, \ldots$
$m = n = 0$ excepted

The cutoff wave number k_c, as defined by Eq. (4-1-17) for the TE_{mn} modes, is given by

$$k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} = \omega_c \sqrt{\mu\epsilon} \tag{4-1-53}$$

where a and b are in meters. The cutoff frequency, as defined in Eq. (4-1-19) for the TE_{mn} modes, is

$$f_c = \frac{1}{2\sqrt{\mu\epsilon}} \sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}} \tag{4-1-54}$$

The propagation constant (or the phase constant here) β_g, as defined in Eq. (4-1-18), is expressed by

$$\beta_g = \omega\sqrt{\mu\epsilon} \sqrt{1 - \left(\frac{f_c}{f}\right)^2} \tag{4-1-55}$$

The phase velocity in the positive z direction for the TE_{mn} modes is shown as

$$v_g = \frac{\omega}{\beta_g} = \frac{v_p}{\sqrt{1 - (f_c/f)^2}} \tag{4-1-56}$$

where $v_p = 1/\sqrt{\mu\epsilon}$ is the phase velocity in an unbounded dielectric.

The characteristic wave impedance of TE_{mn} modes in the guide can be derived from Eqs. (4-1-33) and (4-1-34):

$$Z_g = \frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{\omega\mu}{\beta_g} = \frac{\eta}{\sqrt{1 - (f_c/f)^2}} \tag{4-1-57}$$

where $\eta = \sqrt{\mu/\epsilon}$ is the intrinsic impedance in an unbounded dielectric. The wavelength λ_g in the guide for the TE_{mn} modes is given by

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (f_c/f)^2}} \tag{4-1-58}$$

where $\lambda = v_p/f$ is the wavelength in an unbounded dielectric.

Since the cutoff frequency shown in Eq. (4-1-54) is a function of the modes and guide dimensions, the physical size of the waveguide will determine the propagation of the modes. Table 4-1-1 tabulates the ratio of cutoff frequency of some modes with respect to that of the dominant mode in terms of the physical dimension.

Whenever two or more modes have the same cutoff frequency, they are said to be *degenerate modes*. In a rectangular guide the corresponding TE_{mn} and TM_{mn} modes are always degenerate. In a square guide the TE_{mn}, TE_{nm}, TM_{mn}, and TM_{nm} modes form a foursome of degeneracy. Rectangular guides ordinarily have dimensions of $a = 2b$ ratio. The mode with the lowest cutoff frequency in a particular

TABLE 4-1-1   MODES OF $(f_c)_{mn}/f_c$ FOR $a \geq b$

Modes f/f_{10} a/b	TE_{10}	TE_{01}	TE_{11} TM_{11}	TE_{20}	TE_{02}	TE_{21} TM_{21}	TE_{12} TM_{12}	TE_{22} TM_{22}	TE_{30}
1	1	1	1.414	2	2	2.236	2.236	2.828	3
1.5	1	1.5	1.803	2	3	2.500	3.162	3.606	3
2	1	2	2.236	2	4	2.828	4.123	4.472	3
3	1	3	3.162	2	6	3.606	6.083	6.325	3
∞	1	∞	∞	2	∞	∞	∞	∞	3

guide is called the *dominant mode*. The dominant mode in a rectangular guide with $a > b$ is the TE_{10} mode. Each mode has a specific mode pattern (or field pattern).

It is normal for all modes to exist simultaneously in a given waveguide. The situation is not very serious, however. Actually, only the dominant mode propagates, and the higher modes near the sources or discontinuities decay very fast.

Example 4-1-1: TE_{10} in Rectangular Waveguide

An air-filled rectangular waveguide of inside dimensions 7×3.5 cm operates in the dominant TE_{10} mode as shown in Fig. 4-1-4.
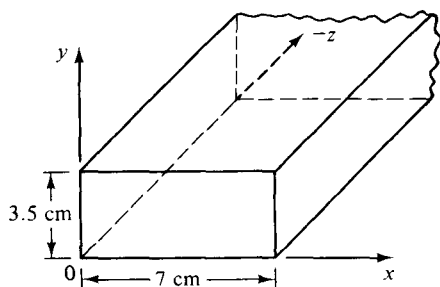
Figure 4-1-4   Rectangular waveguide for Example 4-1-1.

a. Find the cutoff frequency.
b. Determine the phase velocity of the wave in the guide at a frequency of 3.5 GHz.
c. Determine the guided wavelength at the same frequency.

Solution

a. $f_c = \dfrac{c}{2a} = \dfrac{3 \times 10^8}{2 \times 7 \times 10^{-2}} = 2.14$ GHz

b. $v_g = \dfrac{c}{\sqrt{1 - (f_c/f)^2}} = \dfrac{3 \times 10^8}{\sqrt{1 - (2.14/3.5)^2}} = 3.78 \times 10^8$ m/s

c. $\lambda_g = \dfrac{\lambda_0}{\sqrt{1 - (f_c/f)^2}} = \dfrac{3 \times 10^8/(3.5 \times 10^9)}{\sqrt{1 - (2.14/3.5)^2}} = 10.8$ cm
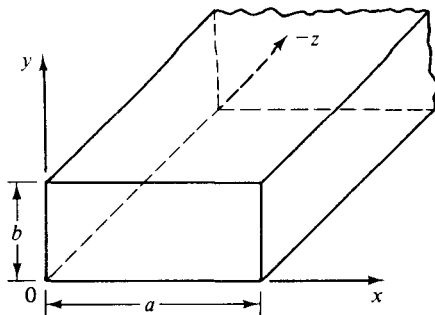
4-1-3 TM Modes in Rectangular Waveguides

The TM_{mn} modes in a rectangular guide are characterized by $H_z = 0$. In other words, the z component of an electric field E must exist in order to have energy transmission in the guide. Consequently, the Helmholtz equation for E in the rectangular coordinates is given by

$$\nabla^2 E_z = \gamma^2 E_z \qquad (4\text{-}1\text{-}59)$$

A solution of the Helmholtz equation is in the form of

$$E_z = \left[A_m \sin\left(\frac{m\pi x}{a}\right) + B_m \cos\left(\frac{m\pi x}{a}\right)\right]\left[C_n \sin\left(\frac{n\pi y}{b}\right) + D_n \cos\left(\frac{n\pi y}{b}\right)\right]e^{-j\beta_g z} \qquad (4\text{-}1\text{-}60)$$

which must be determined according to the given boundary conditions. The procedures for doing so are similar to those used in finding the TE-mode wave.

The boundary conditions on E_z require that the field vanishes at the waveguide walls, since the tangent component of the electric field E_z is zero on the conducting surface. This requirement is that for $E_z = 0$ at $x = 0$, a, then $B_m = 0$, and for $E_z = 0$ at $y = 0$, b, then $D_n = 0$. Thus the solution as shown in Eq. (4-1-60) reduces to

$$E_z = E_{0z} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right)e^{-j\beta_g z} \qquad (4\text{-}1\text{-}61)$$

where $m = 1, 2, 3, \ldots$
$n = 1, 2, 3, \ldots$

If either $m = 0$ or $n = 0$, the field intensities all vanish. So there is no TM_{01} or TM_{10} mode in a rectangular waveguide, which means that TE_{10} is the dominant mode in a rectangular waveguide for $a > b$. For $H_z = 0$, the field equations, after expanding $\nabla \times \mathbf{H} = j\omega\epsilon\,\mathbf{E}$, are given by

$$\frac{\partial E_z}{\partial y} + j\beta_g E_y = -j\omega\mu H_x \qquad (4\text{-}1\text{-}62)$$

$$j\beta_g E_x + \frac{\partial E_z}{\partial x} = j\omega\mu H_y \qquad (4\text{-}1\text{-}63)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0 \qquad (4\text{-}1\text{-}64)$$

$$\beta_g H_y = \omega\epsilon E_x \qquad (4\text{-}1\text{-}65)$$

$$-\beta_g H_x = \omega\epsilon E_y \qquad (4\text{-}1\text{-}66)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\epsilon E_z \qquad (4\text{-}1\text{-}67)$$

These equations can be solved simultaneously for E_x, E_y, H_x, and H_y in terms of E_z.

The resultant field equations for TM modes are

$$E_x = \frac{-j\beta_g}{k_c^2} \frac{\partial E_z}{\partial x} \tag{4-1-68}$$

$$E_y = \frac{-j\beta_g}{k_c^2} \frac{\partial E_z}{\partial y} \tag{4-1-69}$$

$$E_z = \text{Eq. (4-1-61)} \tag{4-1-70}$$

$$H_x = \frac{j\omega\epsilon}{k_c^2} \frac{\partial E_z}{\partial y} \tag{4-1-71}$$

$$H_y = \frac{-j\omega\epsilon}{k_c^2} \frac{\partial E_z}{\partial x} \tag{4-1-72}$$

$$H_z = 0 \tag{4-1-73}$$

where $\beta_g^2 - \omega^2\mu\epsilon = -k_c^2$ is replaced.

Differentiating Eq. (4-1-61) with respect to x or y and substituting the results in Eqs. (4-1-68) through (4-1-73) yield a new set of field equations. The TM$_{mn}$ mode field equations in rectangular waveguides are

$$E_x = E_{0x} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-74}$$

$$E_y = E_{0y} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-75}$$

$$E_z = \text{Eq. (4-1-61)} \tag{4-1-76}$$

$$H_x = H_{0x} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-77}$$

$$H_y = H_{0y} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-j\beta_g z} \tag{4-1-78}$$

$$H_z = 0 \tag{4-1-79}$$

Some of the TM-mode characteristic equations are identical to those of the TE modes, but some are different. For convenience, all are shown here:

$$f_c = \frac{1}{2\sqrt{\mu\epsilon}} \sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}} \tag{4-1-80}$$

$$\beta_g = \omega\sqrt{\mu\epsilon} \sqrt{1 - \left(\frac{f_c}{f}\right)^2} \tag{4-1-81}$$

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (f_c/f)^2}} \tag{4-1-82}$$

$$v_g = \frac{v_p}{\sqrt{1 - (f_c/f)^2}} \tag{4-1-83}$$

$$Z_g = \frac{\beta_g}{\omega\epsilon} = \eta\sqrt{1 - \left(\frac{f_c}{f}\right)^2} \qquad (4\text{-}1\text{-}84)$$

4-1-4 Power Transmission in Rectangular Waveguides

The power transmitted through a waveguide and the power loss in the guide walls can be calculated by means of the complex Poynting theorem described in Chapter 2. It is assumed that the guide is terminated in such a way that there is no reflection from the receiving end or that the guide is infinitely long compared with the wavelength. From the Poynting theorem in Section 2-2, the power transmitted through a guide is given by

$$P_{tr} = \oint \mathbf{p} \cdot d\mathbf{s} = \oint \frac{1}{2}(\mathbf{E} \times \mathbf{H}^*) \cdot d\mathbf{s} \qquad (4\text{-}1\text{-}85)$$

For a lossless dielectric, the time-average power flow through a rectangular guide is given by

$$P_{tr} = \frac{1}{2Z_g} \int_a |E|^2 \, da = \frac{Z_g}{2} \int_a |H|^2 \, da \qquad (4\text{-}1\text{-}86)$$

where $Z_g = \dfrac{E_x}{H_y} = -\dfrac{E_y}{H_x}$

$|E|^2 = |E_x|^2 + |E_y|^2$

$|H|^2 = |H_x|^2 + |H_y|^2$

For TE_{mn} modes, the average power transmitted through a rectangular waveguide is given by

$$P_{tr} = \frac{\sqrt{1 - (f_c/f)^2}}{2\eta} \int_0^b \int_0^a (|E_x|^2 + |E_y|^2) \, dx \, dy \qquad (4\text{-}1\text{-}87)$$

For TM_{mn} modes, the average power transmitted through a rectangular waveguide is given by

$$P_{tr} = \frac{1}{2\eta\sqrt{1 - (f_c/f)^2}} \int_0^b \int_0^a (|E_x|^2 + |E_y|^2) \, dx \, dy \qquad (4\text{-}1\text{-}88)$$

where $\eta = \sqrt{\mu/\epsilon}$ is the intrinsic impedance in an unbounded dielectric.

4-1-5 Power Losses in Rectangular Waveguides

There are two types of power losses in a rectangular waveguide:

1. Losses in the dielectric
2. Losses in the guide walls

First we shall consider power losses caused by dielectric attenuation. In a low-

loss dielectric (that is, $\sigma \ll \mu\epsilon$), the propagation constant for a plane wave traveling in an unbounded lossy dielectric is given in Eq. (2-5-20) by

$$\alpha = \frac{\sigma}{2}\sqrt{\frac{\mu}{\epsilon}} = \frac{\eta\sigma}{2} \qquad (4\text{-}1\text{-}89)$$

The attenuation caused by the low-loss dielectric in the rectangular waveguide for the TE_{mn} or TM_{mn} modes is given by

$$\alpha_g = \frac{\sigma\eta}{2\sqrt{1 - (f_c/f)^2}} \qquad \text{for TE mode} \qquad (4\text{-}1\text{-}90)$$

$$\alpha_g = \frac{\sigma\eta}{2}\sqrt{1 - (f_c/f)^2} \qquad \text{for TM mode} \qquad (4\text{-}1\text{-}90a)$$

As $f \gg f_c$, the attenuation constant in the guide approaches that for the unbounded dielectric given by Eq. (4-1-89). However, if the operating frequency is way below the cutoff frequency, $f \ll f_c$, the attenuation constant becomes very large and non-propagation occurs.

Now we shall consider power losses caused by the guide walls. When the electric and magnetic intensities propagate through a lossy waveguide, their magnitudes may be written

$$|E| = |E_{0z}|e^{-\alpha_g z} \qquad (4\text{-}1\text{-}91)$$

$$|H| = |H_{0z}|e^{-\alpha_g z} \qquad (4\text{-}1\text{-}92)$$

where E_{0z} and H_{0z} are the field intensities at $z = 0$. It is interesting to note that, for a low-loss guide, the time-average power flow decreases proportionally to $e^{-2\alpha_g z}$. Hence

$$P_{tr} = (P_{tr} + P_{loss})e^{-2\alpha_g z} \qquad (4\text{-}1\text{-}93)$$

For $P_{loss} \ll P_{tr}$ and $2\alpha_g z \ll 1$,

$$\frac{P_{loss}}{P_{tr}} + 1 = 1 + 2\alpha_g z \qquad (4\text{-}1\text{-}94)$$

Finally,

$$\alpha_g = \frac{P_L}{2P_{tr}} \qquad (4\text{-}1\text{-}95)$$

where P_L is the power loss per unit length. Consequently, the attenuation constant of the guide walls is equal to the ratio of the power loss per unit length to twice the power transmitted through the guide.

Since the electric and magnetic field intensities established at the surface of a low-loss guide wall decay exponentially with respect to the skin depth while the waves progress into the walls, it is better to define a surface resistance of the guide walls as

$$R_s \equiv \frac{\rho}{\delta} = \frac{1}{\sigma\delta} = \frac{\alpha_g}{\sigma} = \sqrt{\frac{\pi f\mu}{\sigma}} \qquad \Omega/\text{square} \qquad (4\text{-}1\text{-}96)$$

where ρ = resistivity of the conducting wall in ohms-meter
$\quad\ \sigma$ = conductivity in mhos per meter
$\quad\ \delta$ = skin depth or depth of penetration in meters

The power loss per unit length of guide is obtained by integrating the power density over the surface of the conductor corresponding to the unit length of the guide. This is

$$P_L = \frac{R_s}{2} \int_s |H_t|^2 \, ds \qquad \text{W/unit length} \qquad (4\text{-}1\text{-}97)$$

where H_t is the tangential component of magnetic intensity at the guide walls.
Substitution of Eqs. (4-1-86) and (4-1-97) in Eq. (4-1-95) yields

$$\alpha_g = \frac{R_s \int_s |H_t|^2 \, ds}{2Z_g \int_a |H|^2 \, da} \qquad (4\text{-}1\text{-}98)$$

where

$$|H|^2 = |H_z|^2 + |H_y|^2 \qquad (4\text{-}1\text{-}99)$$

$$|H_t|^2 = |H_{tx}|^2 + |H_{ty}|^2 \qquad (4\text{-}1\text{-}100)$$

Example 4-1-2:   TE₁₀ Mode in Rectangular Waveguide

An airfilled waveguide with a cross section 2 × 1 cm transports energy in the TE_{10} mode at the rate of 0.5 hp. The impressed frequency is 30 GHz. What is the peak value of the electric field occurring in the guide? (Refer to Fig. 4-1-5.)

Solution   The field components of the dominant mode TE_{10} can be obtained by substituting $m = 1$ and $n = 0$ in Eqs. (4-1-47) through (4-1-52). Then

$$E_x = 0 \qquad\qquad\qquad H_x = \frac{E_{0y}}{Z_g} \sin\left(\frac{\pi x}{a}\right) e^{-j\beta_g z}$$

$$E_y = E_{0y} \sin\left(\frac{\pi x}{a}\right) e^{-j\beta_g z} \qquad H_y = 0$$

$$E_z = 0 \qquad\qquad\qquad H_z = H_{0z} \cos\left(\frac{\pi x}{a}\right) e^{-j\beta_g z}$$

where $Z_g = \omega\mu_0/\beta_g$.
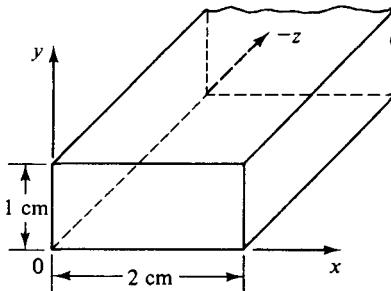
Figure 4-1-5 Rectangular waveguide for Example 4-1-2.

The phase constant β_g can be found from

$$\beta_g = \sqrt{\omega^2 \mu_0 \epsilon_0 - \frac{\pi^2}{a^2}} = \pi \sqrt{\frac{(2f)^2}{c^2} - \frac{1}{a^2}} = \pi \sqrt{\frac{4 \times 9 \times 10^{20}}{9 \times 10^{16}} - \frac{1}{4 \times 10^{-4}}}$$

$$= 193.5\pi = 608.81 \text{ rad/m}$$

The power delivered in the z direction by the guide is

$$P = \text{Re}\left[\frac{1}{2} \int_0^b \int_0^a (\mathbf{E} \times \mathbf{H}^*)\right] \cdot dx\, dy\, \mathbf{u}_z$$

$$= \frac{1}{2} \int_0^b \int_0^a \left[\left(E_{0y} \sin\left(\frac{\pi x}{a}\right) e^{-j\beta_g z} \mathbf{u}_y\right) \times \left(\frac{-\beta_g}{\omega\mu_0} E_{0y} \sin\left(\frac{\pi x}{a}\right) e^{+j\beta_g z}\mathbf{u}_x\right)\right] \cdot dx\, dy\, \mathbf{u}_z$$

$$= \frac{1}{2} E_{0y}^2 \frac{\beta_g}{\omega\mu_0} \int_0^b \int_0^a \left(\sin\left(\frac{\pi x}{a}\right)\right)^2 dx\, dy$$

$$= \frac{1}{4} E_{0y}^2 \frac{\beta_g}{\omega\mu_0} ab$$

$$373 = \frac{1}{4} E_{0y}^2 \frac{193.5\pi (10^{-2})(2 \times 10^{-2})}{2\pi (3 \times 10^{10})(4\pi \times 10^7)}$$

$$E_{0y} = 53.87 \text{ kV/m}$$

The peak value of the electric intensity is 53.87 kV/m.

4-1-6 Excitations of Modes in Rectangular Waveguides

In general, the field intensities of the desired mode in a waveguide can be established by means of a probe or loop-coupling device. The probe may be called a monopole antenna; the coupling loop, the loop antenna. A probe should be located so as to excite the electric field intensity of the mode, and a coupling loop in such a way as to generate the magnetic field intensity for the desired mode. If two or more probes or loops are to be used, care must be taken to ensure the proper phase relationship between the currents in the various antennas. This factor can be achieved by inserting additional lengths of transmission line in one or more of the antenna feeders. Impedance matching can be accomplished by varying the position and depth of the antenna in the guide or by using impedance-matching stubs on the coaxial line feeding the waveguide. A device that excites a given mode in the guide can also serve reciprocally as a receiver or collector of energy for that mode. The methods of excitation for various modes in rectangular waveguides are shown in Fig. 4-1-6.

In order to excite a TE_{10} mode in one direction of the guide, the two exciting antennas should be arranged in such a way that the field intensities cancel each other in one direction and reinforce in the other. Figure 4-1-7 shows an arrangement for launching a TE_{10} mode in one direction only. The two antennas are placed a quarter-wavelength apart and their phases are in time quadrature. Phasing is compensated by use of an additional quarter-wavelength section of line connected to the antenna
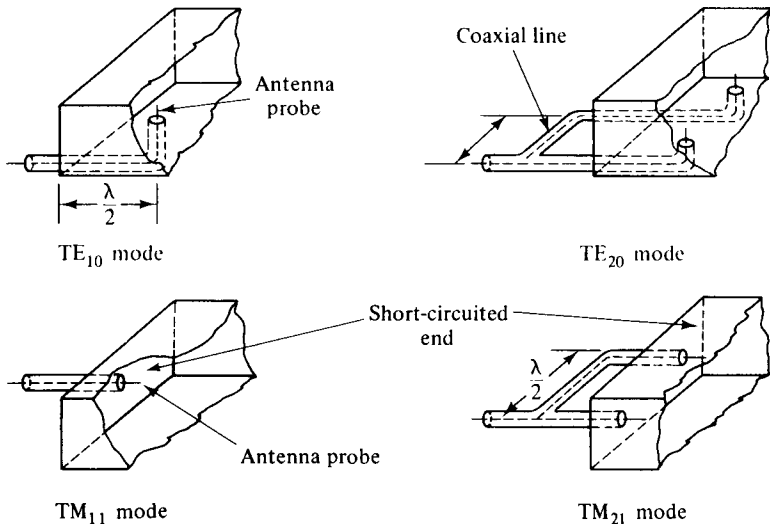
Figure 4-1-6   Methods of exciting various modes in rectangular waveguides.
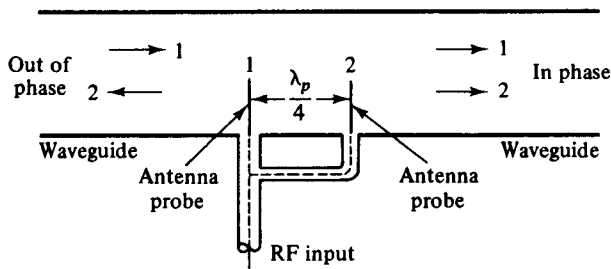
Figure 4-1-7   A method of launching a TE$_{10}$ mode in one direction only.

feeders. The field intensities radiated by the two antennas are in phase opposition to the left of the antennas and cancel each other, whereas in the region to the right of the antennas the field intensities are in time phase and reinforce each other. The resulting wave thus propagates to the right in the guide.

Some higher modes are generated by discontinuities of the waveguide such as obstacles, bends, and loads. However, the higher-order modes are, in general, more highly attenuated than the corresponding dominant mode. On the other hand, the dominant mode tends to remain as a dominant wave even when the guide is large enough to support the higher modes.

4-1-7 Characteristics of Standard Rectangular Waveguides

Rectangular waveguides are commonly used for power transmission at microwave frequencies. Their physical dimensions are regulated by the frequency of the signal being transmitted. For example, at X-band frequencies from 8 to 12 GHz, the outside dimensions of a rectangular waveguide, designated as EIA WR (90) by the Electronic Industry Association, are 2.54 cm (1.0 in.) wide and 1.27 cm (0.5 in.)

TABLE 4-1-7　CHARACTERISTICS OF STANDARD RECTANGULAR WAVEGUIDES

EIA[a] designation WR[b]( )	Physical dimensions				Cutoff frequency for air-filled waveguide in GHz	Recommended frequency range for TE$_{10}$ mode in GHZ
	Inside, in cm (in.)		Outside, in cm (in.)			
	Width	Height	Width	Height		
2300	58.420 (23.000)	29.210 (11.500)	59.055 (23.250)	29.845 (11.750)	0.257	0.32–0.49
2100	53.340 (21.000)	26.670 (10.500)	53.973 (21.250)	27.305 (10.750)	0.281	0.35–0.53
1800	45.720 (18.000)	22.860 (9.000)	46.350 (18.250)	23.495 (9.250)	0.328	0.41–0.62
1500	38.100 (15.000)	19.050 (7.500)	38.735 (15.250)	19.685 (7.750)	0.394	0.49–0.75
1150	29.210 (11.500)	14.605 (5.750)	29.845 (11.750)	15.240 (6.000)	0.514	0.64–0.98
975	24.765 (9.750)	12.383 (4.875)	25.400 (10.000)	13.018 (5.125)	0.606	0.76–1.15
770	19.550 (7.700)	9.779 (3.850)	20.244 (7.970)	10.414 (4.100)	0.767	0.96–1.46
650	16.510 (6.500)	8.255 (3.250)	16.916 (6.660)	8.661 (3.410)	0.909	1.14–1.73
510	12.954 (5.100)	6.477 (2.500)	13.360 (5.260)	6.883 (2.710)	1.158	1.45–2.20
430	10.922 (4.300)	5.461 (2.150)	11.328 (4.460)	5.867 (2.310)	1.373	1.72–2.61
340	8.636 (3.400)	4.318 (1.700)	9.042 (3.560)	4.724 (1.860)	1.737	2.17–3.30
284	7.214 (2.840)	3.404 (1.340)	7.620 (3.000)	3.810 (1.500)	2.079	2.60–3.95
229	5.817 (2.290)	2.908 (1.145)	6.142 (2.418)	3.233 (1.273)	2.579	3.22–4.90
187	4.755 (1.872)	2.215 (0.872)	5.080 (2.000)	2.540 (1.000)	3.155	3.94–5.99
159	4.039 (1.590)	2.019 (0.795)	4.364 (1.718)	2.344 (0.923)	3.714	4.64–7.05
137	3.485 (1.372)	1.580 (0.622)	3.810 (1.500)	1.905 (0.750)	4.304	5.38–8.17
112	2.850 (1.122)	1.262 (0.497)	3.175 (1.250)	1.588 (0.625)	5.263	6.57–9.99
90	2.286 (0.900)	1.016 (0.400)	2.540 (1.000)	1.270 (0.500)	6.562	8.20–12.50
75	1.905 (0.750)	0.953 (0.375)	2.159 (0.850)	1.207 (0.475)	7.874	9.84–15.00
62	1.580 (0.622)	0.790 (0.311)	1.783 (0.702)	0.993 (0.391)	9.494	11.90–18.00
51	1.295 (0.510)	0.648 (0.255)	1.499 (0.590)	0.851 (0.335)	11.583	14.50–22.00
42	1.067 (0.420)	0.432 (0.170)	1.270 (0.500)	0.635 (0.250)	14.058	17.60–26.70

[a] Electronic Industry Association

[b] Rectangular Waveguide

TABLE 4-1-7   CHARACTERISTICS OF STANDARD RECTANGULAR WAVEGUIDES (Cont.)

EIA[a] designation WR[b]( )	Physical dimensions				Cutoff frequency for air-filled waveguide in GHz	Recommended frequency range for TE_{10} mode in GHZ
	Inside, in cm (in.)		Outside, in cm (in.)			
	Width	Height	Width	Height		
34	0.864 (0.340)	0.432 (0.170)	1.067 (0.420)	0.635 (0.250)	17.361	21.70–33.00
28	0.711 (0.280)	0.356 (0.140)	0.914 (0.360)	0.559 (0.220)	21.097	26.40–40.00
22	0.569 (0.224)	0.284 (0.112)	0.772 (0.304)	0.488 (0.192)	26.362	32.90–50.10
19	0.478 (0.188)	0.239 (0.094)	0.681 (0.268)	0.442 (0.174)	31.381	39.20–59.60
15	0.376 (0.148)	0.188 (0.074)	0.579 (0.228)	0.391 (0.154)	39.894	49.80–75.80
12	0.310 (0.122)	0.155 (0.061)	0.513 (0.202)	0.358 (0.141)	48.387	60.50–91.90
10	0.254 (0.100)	0.127 (0.050)	0.457 (0.180)	0.330 (0.130)	59.055	73.80–112.00
8	0.203 (0.080)	0.102 (0.040)	0.406 (0.160)	0.305 (0.120)	73.892	92.20–140.00
7	0.165 (0.065)	0.084 (0.033)	0.343 (0.135)	0.262 (0.103)	90.909	114.00–173.00
5	0.130 (0.051)	0.066 (0.026)	0.257 (0.101)	0.193 (0.076)	115.385	145.00–220.00
4	0.109 (0.043)	0.056 (0.022)	0.211 (0.083)	0.157 (0.062)	137.615	172.00–261.00
3	0.086 (0.034)	0.043 (0.017)	0.163 (0.064)	0.119 (0.047)	174.419	217.00–333.00

high, and its inside dimensions are 2.286 cm (0.90 in.) wide and 1.016 cm (0.40 in.) high. Table 4-1-7 tabulates the characteristics of the standard rectangular waveguides.

4-2 CIRCULAR WAVEGUIDES

A circular waveguide is a tubular, circular conductor. A plane wave propagating through a circular waveguide results in a transverse electric (TE) or transerse magnetic (TM) mode. Several other types of waveguides, such as elliptical and reentrant guides, also propagate electromagnetic waves.

4-2-1 Solutions of Wave Equations in Cylindrical Coordinates

As described in Section 4-1 for rectangular waveguides, only a sinusoidal steady-state or frequency-domain solution will be attempted for circular waveguides. A cylindrical coordinate system is shown in Fig. 4-2-1.
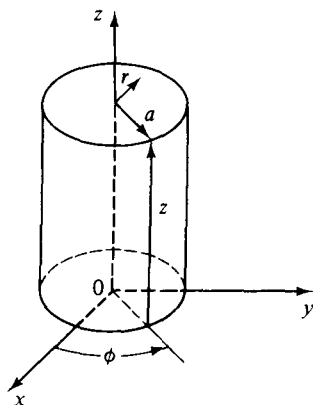
Figure 4-2-1 Cylindrical coordinates.

The scalar Helmholtz equation in cylindrical coordinates is given by

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \psi}{\partial \phi^2} + \frac{\partial^2 \psi}{\partial z^2} = \gamma^2 \psi \qquad (4\text{-}2\text{-}1)$$

Using the method of separation of variables, the solution is assumed in the form of

$$\Psi = R(r)\Phi(\phi)Z(z) \qquad (4\text{-}2\text{-}2)$$

where $R(r)$ = a function of the r coordinate only
$\Phi(\phi)$ = a function of the ϕ coordinate only
$Z(z)$ = a function of the z coordinate only

Substitution of Eq. (4-2-2) in (4-2-1) and division of the resultant by (4-2-2) yield

$$\frac{1}{rR}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \frac{1}{r^2\Phi}\frac{d^2\Phi}{d\phi^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} = \gamma^2 \qquad (4\text{-}2\text{-}3)$$

Since the sum of the three independent terms is a constant, each of the three terms must be a constant. The third term may be set equal to a constant γ_g^2:

$$\frac{d^2Z}{dz^2} = \gamma_g^2 Z \qquad (4\text{-}2\text{-}4)$$

The solutions of this equation are given by

$$Z = Ae^{-\gamma_g z} + Be^{\gamma_g z} \qquad (4\text{-}2\text{-}5)$$

where γ_g = propagation constant of the wave in the guide.
Inserting γ_g^2 for the third term in the left-hand side of Eq. (4-2-3) and multi-plying the resultant by r^2 yield

$$\frac{r}{R}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} - (\gamma^2 - \gamma_g^2)r^2 = 0 \qquad (4\text{-}2\text{-}6)$$

The second term is a function of ϕ only; hence equating the second term to a con-

stant $(-n^2)$ yields

$$\frac{d^2\Phi}{d\phi^2} = -n^2\Phi \tag{4-2-7}$$

The solution of this equation is also a harmonic function:

$$\Phi = A_n \sin(n\phi) + B_n \cos(n\phi) \tag{4-2-8}$$

Replacing the Φ term by $(-n^2)$ in Eq. (4-2-6) and multiplying through by R, we have

$$r\frac{d}{dr}\left(r\frac{dR}{dr}\right) + [(k_c r)^2 - n^2]R = 0 \tag{4-2-9}$$

This is Bessel's equation of order n in which

$$k_c^2 + \gamma^2 = \gamma_g^2 \tag{4-2-10}$$

This equation is called the *characteristic equation* of Bessel's equation. For a lossless guide, the characteristic equation reduces to

$$\beta_g = \pm\sqrt{\omega^2\mu\epsilon - k_c^2} \tag{4-2-11}$$

The solutions of Bessel's equation are

$$R = C_n J_n(k_c r) + D_n N_n(k_c r) \tag{4-2-12}$$

where $J_n(k_c r)$ is the nth-order Bessel function of the first kind, representing a standing wave of $\cos(k_c r)$ for $r < a$ as shown in Fig. 4-2-2. $N_n(k_c r)$ is the nth-order Bessel function of the second kind, representing a standing wave of $\sin(k_c r)$ for $r > a$ as shown in Fig. 4-2-3.

Therefore the total solution of the Helmholtz equation in cylindrical coordinates is given by

$$\Psi = [C_n J_n(k_c r) + D_n N_n(k_c r)][A_n \sin(n\phi) + B_n \cos(n\phi)]e^{\pm j\beta_g z} \tag{4-2-13}$$
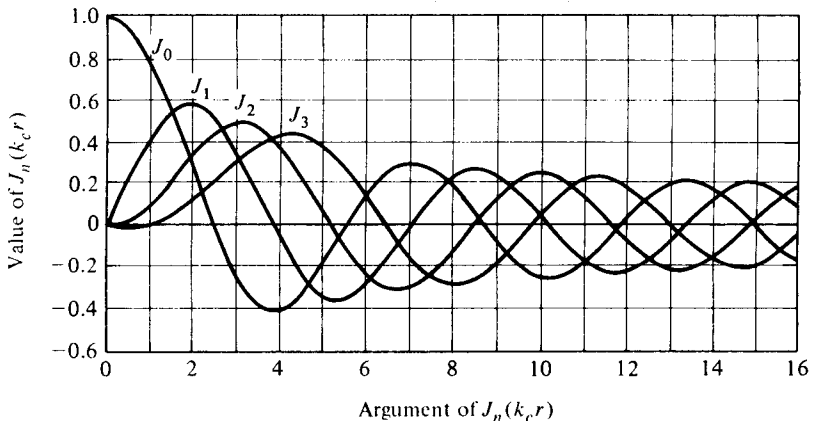
Figure 4-2-2   Bessel functions of the first kind.
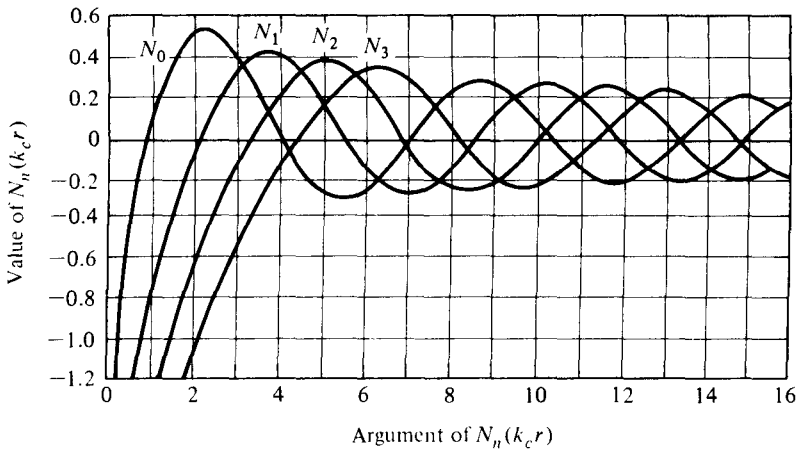
Figure 4-2-3 Bessel functions of the second kind.

At $r = 0$, however, $k_c r = 0$; then the function N_n approaches infinity, so $D_n = 0$. This means that at $r = 0$ on the z axis, the field must be finite. Also, by use of trigonometric manipulations, the two sinusoidal terms become

$$A_n \sin (n\phi) + B_n \cos (n\phi) = \sqrt{A_n^2 + B_n^2} \cos \left[n\phi + \tan^{-1}\left(\frac{A_n}{B_n}\right) \right]$$

$$= F_n \cos (n\phi) \qquad (4\text{-}2\text{-}14)$$

Finally, the solution of the Helmholtz equation is reduced to

$$\Psi = \Psi_0 J_n(k_c r) \cos (n\phi) e^{-j\beta_g z} \qquad (4\text{-}2\text{-}15)$$

4-2-2 TE Modes in Circular Waveguides

It is commonly assumed that the waves in a circular waveguide are propagating in the positive z direction. Figure 4-2-4 shows the coordinates of a circular guide.

The TE_{np} modes in the circular guide are characterized by $E_z = 0$. This means
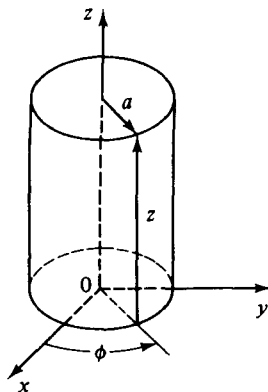
Figure 4-2-4 Coordinates of a circular waveguide.

that the z component of the magnetic field H_z must exist in the guide in order to have electromagnetic energy transmission. A Helmholtz equation for H_z in a circular guide is given by

$$\nabla^2 H_z = \gamma^2 H_z \qquad (4\text{-}2\text{-}16)$$

Its solution is given in Eq. (4-2-15) by

$$H_z = H_{0z} J_n(k_c r) \cos(n\phi) e^{-j\beta_g z} \qquad (4\text{-}2\text{-}17)$$

which is subject to the given boundary conditions.

For a lossless dielectric, Maxwell's curl equations in frequency domain are given by

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \qquad (4\text{-}2\text{-}18)$$

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E} \qquad (4\text{-}2\text{-}19)$$

In cylindrical coordinates, their components are expressed as

$$\frac{1}{r}\frac{\partial E_z}{\partial \phi} - \frac{\partial E_\phi}{\partial z} = -j\omega\mu H_r \qquad (4\text{-}2\text{-}20)$$

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = -j\omega\mu H_\phi \qquad (4\text{-}2\text{-}21)$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rE_\phi) - \frac{1}{r}\frac{\partial E_r}{\partial \phi} = -j\omega\mu H_z \qquad (4\text{-}2\text{-}22)$$

$$\frac{1}{r}\frac{\partial H_z}{\partial \phi} - \frac{\partial H_\phi}{\partial z} = j\omega\epsilon E_r \qquad (4\text{-}2\text{-}23)$$

$$-j\beta_g H_r - \frac{\partial H_z}{\partial r} = j\omega\epsilon E_\phi \qquad (4\text{-}2\text{-}24)$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\phi) - \frac{1}{r}\frac{\partial H_r}{\partial \phi} = j\omega\epsilon E_z \qquad (4\text{-}2\text{-}25)$$

When the differentiation $\partial/\partial z$ is replaced by $(-j\beta_g)$ and the z component of electric field E_z by zero, the TE-mode equations in terms of H_z in a circular waveguide are expressed as

$$E_r = -\frac{j\omega\mu}{k_c^2}\frac{1}{r}\frac{\partial H_z}{\partial \phi} \qquad (4\text{-}2\text{-}26)$$

$$E_\phi = \frac{j\omega\mu}{k_c^2}\frac{\partial H_z}{\partial r} \qquad (4\text{-}2\text{-}27)$$

$$E_z = 0 \qquad (4\text{-}2\text{-}28)$$

$$H_r = \frac{-j\beta_g}{k_c^2}\frac{\partial H_z}{\partial r} \qquad (4\text{-}2\text{-}29)$$

$$H_\phi = \frac{-j\beta_g}{k_c^2} \frac{1}{r} \frac{\partial H_z}{\partial \phi} \tag{4-2-30}$$

$$H_z = H_{0z} J_n(k_c r) \cos(n\phi) e^{-j\beta_g z} \tag{4-2-31}$$

where $k_c^2 = \omega^2 \mu\epsilon - \beta_g^2$ has been replaced.

The boundary conditions require that the ϕ component of the electric field E_ϕ, which is tangential to the inner surface of the circular waveguide at $r = a$, must vanish or that the r component of the magnetic field H_r, which is normal to the inner surface of $r = a$, must vanish. Consequently

$$E_\phi = 0 \text{ at } r = a \qquad \therefore \left. \frac{\partial H_z}{\partial r} \right|_{r=a} = 0$$

or

$$H_r = 0 \text{ at } r = a \qquad \therefore \left. \frac{\partial H_z}{\partial r} \right|_{r=a} = 0$$

This requirement is equivalent to that expressed in Eq. (4-2-17):

$$\left. \frac{\partial H_z}{\partial r} \right|_{r=a} = H_{0z} J_n'(k_c a) \cos(n\phi) e^{-j\beta_g z} = 0 \tag{4-2-32}$$

Hence

$$J_n'(k_c a) = 0 \tag{4-2-33}$$

where J_n' indicates the derivative of J_n.

Since the J_n are oscillatory functions, the $J_n'(k_c a)$ are also oscillatory functions. An infinite sequence of values of $(k_c a)$ satisfies Eq. (4-2-32). These points, the roots of Eq. (4-2-32), correspond to the maxima and minima of the curves $J_n'(k_c a)$, as shown in Fig. 4-2-2. Table 4-2-1 tabulates a few roots of $J_n'(k_c a)$ for some lower-order n.

TABLE 4-2-1 pth ZEROS OF $J_n'(K_c a)$ FOR TE$_{np}$ MODES

p	$n =$ 0	1	2	3	4	5
1	3.832	1.841	3.054	4.201	5.317	6.416
2	7.016	5.331	6.706	8.015	9.282	10.520
3	10.173	8.536	9.969	11.346	12.682	13.987
4	13.324	11.706	13.170			

The permissible values of k_c can be written

$$k_c = \frac{X_{np}'}{a} \tag{4-2-34}$$

Substitution of Eq. (4-2-17) in Eqs. (4-2-26) through (4-2-31) yields the complete field equations of the TE$_{np}$ modes in circular waveguides:

$$E_r = E_{0r} J_n\left(\frac{X_{np}' r}{a}\right) \sin(n\phi) e^{-j\beta_g z} \tag{4-2-35}$$

$$E_\phi = E_{0\phi} J_n' \left(\frac{X_{np}' r}{a} \right) \cos{(n\phi)} e^{-j\beta_g z} \qquad (4\text{-}2\text{-}36)$$

$$E_z = 0 \qquad (4\text{-}2\text{-}37)$$

$$H_r = -\frac{E_{0\phi}}{Z_g} J_n \left(\frac{X_{np}' r}{a} \right) \cos{(n\phi)} e^{-j\beta_g z} \qquad (4\text{-}2\text{-}38)$$

$$H_\phi = \frac{E_{0r}}{Z_g} J_n \left(\frac{X_{np}' r}{a} \right) \sin{(n\phi)} e^{-j\beta_g z} \qquad (4\text{-}2\text{-}39)$$

$$H_z = H_{0z} J_n \left(\frac{X_{np}' r}{a} \right) \cos{(n\phi)} e^{-j\beta_g z} \qquad (4\text{-}2\text{-}40)$$

where $Z_g = E_r/H_\phi = -E_\phi/H_r$ has been replaced for the wave impedance in the guide and where $n = 0, 1, 2, 3, \ldots$ and $p = 1, 2, 3, 4, \ldots$.

The first subscript n represents the number of full cycles of field variation in one revolution through 2π rad of ϕ. The second subscript p indicates the number of zeros of E_ϕ—that is, $J_n'(X_{np}' r/a)$ along the radial of a guide, but the zero on the axis is excluded if it exists.

The mode propagation constant is determined by Eqs. (4-2-26) through (4-2-31) and Eq. (4-2-34):

$$\beta_g = \sqrt{\omega^2 \mu\epsilon - \left(\frac{X_{np}'}{a} \right)^2} \qquad (4\text{-}2\text{-}41)$$

The cutoff wave number of a mode is that for which the mode propagation constant vanishes. Hence

$$k_c = \frac{X_{np}'}{a} = \omega_c \sqrt{\mu\epsilon} \qquad (4\text{-}2\text{-}42)$$

The cutoff frequency for TE modes in a circular guide is then given by

$$f_c = \frac{X_{np}'}{2\pi a \sqrt{\mu\epsilon}} \qquad (4\text{-}2\text{-}43)$$

and the phase velocity for TE modes is

$$v_g = \frac{\omega}{\beta_g} = \frac{v_p}{\sqrt{1 - (f_c/f)^2}} \qquad (4\text{-}2\text{-}44)$$

where $v_p = 1/\sqrt{\mu\epsilon} = c/\sqrt{\mu_r \epsilon_r}$ is the phase velocity in an unbounded dielectric.

The wavelength and wave impedance for TE modes in a circular guide are given, respectively, by

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (f_c/f)^2}} \qquad (4\text{-}2\text{-}45)$$

and

$$Z_g = \frac{\omega\mu}{\beta_g} = \frac{\eta}{\sqrt{1 - (f_c/f)^2}} \qquad (4\text{-}2\text{-}46)$$

where $\lambda = \dfrac{v_p}{f}$ = wavelength in an unbounded dielectric

$\eta = \sqrt{\dfrac{\mu}{\epsilon}}$ = intrinsic impedance in an unbounded dielectric

Example 4-2-1:   TE Mode in Circular Waveguide

A TE_{11} mode is propagating through a circular waveguide. The radius of the guide is 5 cm, and the guide contains an air dielectric (refer to Fig. 4-2-5).
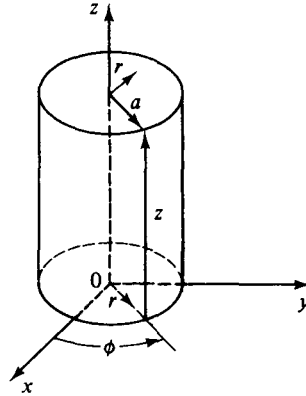
Figure 4-2-5   Diagram for Example 4-2-1.

   a. Determine the cutoff frequency.
   b. Determine the wavelength λ_g in the guide for an operating frequency of 3 GHz.
   c. Determine the wave impedance Z_g in the guide.

Solution

   a. From Table 4-2-1 for TE_{11} mode, $n = 1$, $p = 1$, and $X'_{11} = 1.841 = k_c a$. The cutoff wave number is

$$k_c = \frac{1.841}{a} = \frac{1.841}{5 \times 10^{-2}} = 36.82$$

The cutoff frequency is

$$f_c = \frac{k_c}{2\pi\sqrt{\mu_0\epsilon_0}} = \frac{(36.82)(3 \times 10^8)}{2\pi} = 1.758 \times 10^9 \text{ Hz}$$

   b. The phase constant in the guide is

$$\beta_g = \sqrt{\omega^2\mu_0\epsilon_0 - k_c^2}$$

$$= \sqrt{(2\pi \times 3 \times 10^9)^2(4\pi \times 10^{-7} \times 8.85 \times 10^{-12}) - (36.82)^2}$$

$$= 50.9 \text{ rads/m}$$

The wavelength in the guide is

$$\lambda_g = \frac{2\pi}{\beta_g} = \frac{6.28}{50.9} = 12.3 \text{ cm}$$

c. The wave impedance in the guide is

$$Z_g = \frac{\omega\mu_0}{\beta_g} = \frac{(2\pi \times 3 \times 10^9)(4\pi \times 10^{-7})}{50.9} = 465 \ \Omega$$

4-2-3 TM Modes in Circular Waveguides

The TM_{np} modes in a circular guide are characterized by $H_z = 0$. However, the z component of the electric field E_z must exist in order to have energy transmission in the guide. Consequently, the Helmholtz equation for E_z in a circular waveguide is given by

$$\nabla^2 E_z = \gamma^2 E_z \qquad (4\text{-}2\text{-}47)$$

Its solution is given in Eq. (4-2-15) by

$$E_z = E_{0z} J_n(k_c r) \cos (n\phi) e^{-j\beta_g z} \qquad (4\text{-}2\text{-}48)$$

which is subject to the given boundary conditions.

The boundary condition requires that the tangential component of electric field E_z at $r = a$ vanishes. Consequently,

$$J_n(k_c a) = 0 \qquad (4\text{-}2\text{-}49)$$

Since $J_n(k_c r)$ are oscillatory functions, as shown in Fig. 4-2-2, there are infinite numbers of roots of $J_n(k_c r)$. Table 4-2-2 tabulates a few of them for some lower-order n.

TABLE 4-2-2 pth ZEROS OF $J_n(K_c a)$ FOR TM_{np} MODES

p	$n =$	0	1	2	3	4	5
1		2.405	3.832	5.136	6.380	7.588	8.771
2		5.520	7.106	8.417	9.761	11.065	12.339
3		8.645	10.173	11.620	13.015	14.372	
4		11.792	13.324	14.796			

For $H_z = 0$ and $\partial/\partial z = -j\beta_g$, the field equations in the circular guide, after expanding $\nabla \times E = -j\omega\mu H$ and $\nabla \times H = j\omega\epsilon E$, are given by

$$E_r = \frac{-j\beta_g}{k_c^2} \frac{\partial E_z}{\partial r} \qquad (4\text{-}2\text{-}50)$$

$$E_\phi = \frac{-j\beta_g}{k_c^2} \frac{1}{r} \frac{\partial E_z}{\partial \phi} \qquad (4\text{-}2\text{-}51)$$

$$E_z = \text{Eq. } (4\text{-}2\text{-}48) \qquad (4\text{-}2\text{-}52)$$

$$H_r = \frac{j\omega\epsilon}{k_c^2} \frac{1}{r} \frac{\partial E_z}{\partial \phi} \qquad (4\text{-}2\text{-}53)$$

$$H_\phi = \frac{j\omega\epsilon}{k_c^2} \frac{\partial E_z}{\partial r} \tag{4-2-54}$$

$$H_z = 0 \tag{4-2-55}$$

where $k_c^2 = \omega^2\mu\epsilon - \beta_g^2$ has been replaced.

Differentiation of Eq. (4-2-48) with respect to z and substitution of the result in Eqs. (4-2-50) through (4-2-55) yield the field equations of TM_{np} modes in a circular waveguide:

$$E_r = E_{0r} J_n'\left(\frac{X_{np}r}{a}\right) \cos (n\phi)e^{-j\beta_g z} \tag{4-2-56}$$

$$E_\phi = E_{0\phi} J_n\left(\frac{X_{np}r}{a}\right) \sin (n\phi)e^{-j\beta_g z} \tag{4-2-57}$$

$$E_z = E_{0z} J_n\left(\frac{X_{np}r}{a}\right) \cos (n\phi)e^{-j\beta_g z} \tag{4-2-58}$$

$$H_r = \frac{E_{0\phi}}{Z_g} J_n\left(\frac{X_{np}r}{a}\right) \sin (n\phi)e^{-j\beta_g z} \tag{4-2-59}$$

$$H_\phi = \frac{E_{0r}}{Z_g} J_n'\left(\frac{X_{np}r}{a}\right) \cos (n\phi)e^{-j\beta_g z} \tag{4-2-60}$$

$$H_z = 0 \tag{4-2-61}$$

where $Z_g = E_r/H_\phi = -E_\phi/H_r = \beta_g/(\omega\epsilon)$ and $k_c = X_{np}/a$ have been replaced and where $n = 0, 1, 2, 3, \ldots$ and $p = 1, 2, 3, 4, \ldots$.

Some of the TM-mode characteristic equations in the circular guide are identical to those of the TE mode, but some are different. For convenience, all are shown here:

$$\beta_g = \sqrt{\omega^2\mu\epsilon - \left(\frac{X_{np}}{a}\right)^2} \tag{4-2-62}$$

$$k_c = \frac{X_{np}}{a} = \omega_c\sqrt{\mu\epsilon} \tag{4-2-63}$$

$$f_c = \frac{X_{np}}{2\pi a\sqrt{\mu\epsilon}} \tag{4-2-64}$$

$$v_g = \frac{\omega}{\beta_g} = \frac{v_p}{\sqrt{1 - (f_c/f)^2}} \tag{4-2-65}$$

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (f_c/f)^2}} \tag{4-2-66}$$

$$Z_g = \frac{\beta_g}{\omega\epsilon} = \eta\sqrt{1 - \left(\frac{f_c}{f}\right)^2} \tag{4-2-67}$$

It should be noted that the dominant mode, or the mode of lowest cutoff frequency

in a circular waveguide, is the mode of TE_{11} that has the smallest value of the product, $k_c a = 1.841$, as shown in Tables 4-2-1 and 4-2-2.

Example 4-2-2:   Wave Propagation in Circular Waveguide

An air-filled circular waveguide has a radius of 2 cm and is to carry energy at a frequency of 10 GHz. Find all the TE_{np} and TM_{np} modes for which energy transmission is possible.

Solution   Since the physical dimension of the guide and the frequency of the wave remain constant, the product of $(k_c a)$ is also constant. Thus

$$k_c a = (\omega_c \sqrt{\mu_0 \epsilon_0}) a = \frac{2\pi \times 10^{10}}{3 \times 10^8} (2 \times 10^{-2}) = 4.18$$

Any mode having a product of $(k_c a)$ less than or equal to 4.18 will propagate the wave with a frequency of 10 GHz. This is

$$k_c a \leqq 4.18$$

The possible modes are

$$TE_{11}(1.841) \qquad TM_{01}(2.405)$$

$$TE_{21}(3.054) \qquad TM_{11}(3.832)$$

$$TE_{01}(3.832)$$

4-2-4 TEM Modes in Circular Waveguides

The transverse electric and transverse magnetic (TEM) modes or transmission-line modes are characterized by

$$E_z = H_z = 0$$

This means that the electric and magnetic fields are completely transverse to the direction of wave propagation. This mode cannot exist in hollow waveguides, since it requires two conductors, such as the coaxial transmission line and two-open-wire line. Analysis of the TEM mode illustrates an excellent analogous relationship between the method of circuit theory and that of the field theory. Figure 4-2-6 shows a coaxial line.

Maxwell's curl equations in cylindrical coordinates

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \qquad (4\text{-}2\text{-}68)$$

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E} \qquad (4\text{-}2\text{-}69)$$

become

$$B_g E_r = \omega\mu H_\phi \qquad (4\text{-}2\text{-}70)$$

$$B_g E_\phi = \omega\mu H_r \qquad (4\text{-}2\text{-}71)$$
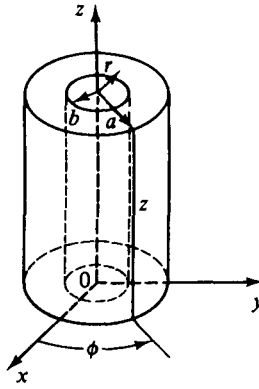
Figure 4-2-6   Coordinates of a coaxial line.

$$\frac{\partial}{\partial r}(rE_\phi) - \frac{\partial Er}{\partial \phi} = 0 \qquad (4\text{-}2\text{-}72)$$

$$\beta_g H_r = -\omega\epsilon E_\phi \qquad (4\text{-}2\text{-}73)$$

$$\beta_g H_\phi = \omega\epsilon E_r \qquad (4\text{-}2\text{-}74)$$

$$\frac{\partial}{\partial r}(rH_\phi) - \frac{\partial H_r}{\partial \phi} = 0 \qquad (4\text{-}2\text{-}75)$$

where $\partial/\partial r = -j\beta_g$ and $E_z = H_z = 0$ are replaced.

Substitution of Eq. (4-2-71) in (4-2-73) yields the propagation constant of the TEM mode in a coaxial line:

$$\beta_g = \omega\sqrt{\mu\epsilon} \qquad (4\text{-}2\text{-}76)$$

which is the phase constant of the wave in a lossless transmission line with a dielectric.

In comparing the preceding equation with the characteristic equation of the Helmholtz equation in cylindrical coordinates as given in Eq. (4-2-11) by

$$\beta_g = \sqrt{\omega^2\mu\epsilon - k_c^2} \qquad (4\text{-}2\text{-}77)$$

it is evident that

$$k_c = 0 \qquad (4\text{-}2\text{-}78)$$

This means that the cutoff frequency of the TEM mode in a coaxial line is zero, which is the same as in ordinary transmission lines.

The phase velocity of the TEM mode can be expressed from Eq. (4-2-76) as

$$v_p = \frac{\omega}{\beta_g} = \frac{1}{\sqrt{\mu\epsilon}} \qquad (4\text{-}2\text{-}79)$$

which is the velocity of light in an unbounded dielectric.

The wave impedance of the TEM mode is found from either Eqs. (4-2-70) and (4-2-73) or Eqs. (4-2-71) and (4-2-74) as

$$\eta\,(\text{TEM}) = \sqrt{\frac{\mu}{\epsilon}} \tag{4-2-80}$$

which is the wave impedance of a lossless transmission line in a dielectric.

Ampère's law states that the line integral of **H** about any closed path is exactly equal to the current enclosed by that path. This is

$$\oint \mathbf{H} \cdot d\ell = I = I_0 e^{-j\beta_g Z} = 2\pi r H_\phi \tag{4-2-81}$$

where I is the complex current that must be supported by the center conductor of a coaxial line. This clearly demonstrates that the TEM mode can only exist in the two-conductor system—not in the hollow waveguide because the center conductor does not exist.

In summary, the properties of TEM modes in a lossless medium are as follows:

1. Its cutoff frequency is zero.
2. Its transmission line is a two-conductor system.
3. Its wave impedance is the impedance in an unbounded dielectric.
4. Its propagation constant is the constant in an unbounded dielectric.
5. Its phase velocity is the velocity of light in an unbounded dielectric.

4-2-5 Power Transmission in Circular Waveguides or Coaxial Lines

In general, the power transmitted through circular waveguides and coaxial lines can be calculated by means of the complex Poynting theorem described in Section 2-2. For a lossless dielectric, the time-average power transmitted through a circular guide can be given by

$$P_{\text{tr}} = \frac{1}{2Z_g} \int_0^{2\pi} \int_0^a [|E_\phi|^2 + |E_\phi|^2] r\, dr\, d\phi \tag{4-2-82}$$

$$P_{\text{tr}} = \frac{Z_g}{2} \int_0^{2\pi} \int_0^a [|H_r|^2 + |H_\phi|^2] r\, dr\, d\phi \tag{4-2-83}$$

where $Z_g = \dfrac{E_r}{H_\phi} = -\dfrac{E_\phi}{H_r}$ = wave impedance in the guide

a = radius of the circular guide

Substitution of Z_g for a particular mode in Eq. (4-2-82) yields the power transmitted by that mode through the guide.

For TE_{np} modes, the average power transmitted through a circular guide is given by

$$P_{tr} = \frac{\sqrt{1 - (f_c/f)^2}}{2\eta} \int_0^{2\pi} \int_0^a [|E_r|^2 + |E_\phi|^2] r\, dr\, d\phi \qquad (4\text{-}2\text{-}84)$$

where $\eta = \sqrt{\mu/\epsilon}$ is the intrinsic impedance in an unbounded dielectric.

For TM_{np} modes, the average power transmitted through a circular guide is given by

$$P_{tr} = \frac{1}{2\eta \sqrt{1 - (f_c/f)^2}} \int_0^{2\pi} \int_0^a [|E_r|^2 + |E_\phi|^2] r\, dr\, d\phi \qquad (4\text{-}2\text{-}85)$$

For TEM modes in coaxial lines, the average power transmitted through a coaxial line or two-open-wire line is given by

$$P_{tr} = \frac{1}{2\eta} \int_0^{2\pi} \int_0^a [|E_r|^2 + |E_\phi|^2] r\, dr\, d\phi \qquad (4\text{-}2\text{-}86)$$

If the current carried by the center conductor of a coaxial line is assumed to be

$$I_z = I_0 e^{-j\beta_g Z} \qquad (4\text{-}2\text{-}87)$$

the magnetic intensity induced by the current around the center conductor is given by Ampère's law as

$$H_\phi = \frac{I_0}{2\pi r} e^{-j\beta_g Z} \qquad (4\text{-}2\text{-}88)$$

The potential rise from the outer conductor to the center conductor is given by

$$V_r = -\int_b^a E_r dr = -\int_b^a \eta H_\phi dr = \frac{I_0 \eta}{2\pi} \ln\left(\frac{b}{a}\right) e^{-j\beta_g Z} \qquad (4\text{-}2\text{-}89)$$

The characteristic impedance of a coaxial line is

$$Z_0 = \frac{V}{I} = \frac{\eta}{2\pi} \ln\left(\frac{b}{a}\right) \qquad (4\text{-}2\text{-}90)$$

where $\eta = \sqrt{\mu/\epsilon}$ is the intrinsic impedance in an unbounded dielectric.

The power transmitted by TEM modes in a coaxial line can be expressed from Eq. (4-2-86) as

$$P_{tr} = \frac{1}{2\eta} \int_0^{2\pi} \int_a^b |\eta H_\phi|^2 r\, dr\, d\phi = \frac{\eta I_0^2}{4\pi} \ln\left(\frac{b}{a}\right) \qquad (4\text{-}2\text{-}91)$$

Substitution of $|V_r|$ from Eq. (4-2-89) into Eq. (4-2-91) yields

$$P_{tr} = \tfrac{1}{2} V_0 I_0 \qquad (4\text{-}2\text{-}92)$$

This shows that the power transmission derived from the Poynting theory is the same as from the circuit theory for an ordinary transmission line.

4-2-6 Power Losses in Circular Waveguides or Coaxial Lines

The theory and equations derived in Section 4-1-5 for TE and TM modes in rectangular waveguides are applicable to TE and TM modes in circular guides. The power losses for the TEM mode in coaxial lines can be computed from transmission-line theory by means of

$$P_L = 2\alpha P_{tr} \tag{4-2-93}$$

where P_L = power loss per unit length
P_{tr} = transmitted power
α = attenuation constant

For a low-loss conductor, the attenuation constant is given by

$$\alpha = \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \tag{4-2-94}$$

4-2-7 Excitations of Modes in Circular Waveguides

As described earlier, TE modes have no z component of an electric field, and TM modes have no z component of magnetic intensity. If a device is inserted in a circular guide in such a way that it excites only a z component of electric intensity, the wave propagating through the guide will be the TM mode; on the other hand, if a device is placed in a circular guide in such a manner that only the z component of magnetic intensity exists, the traveling wave will be the TE mode. The methods of excitation for various modes in circular waveguides are shown in Fig. 4-2-7.
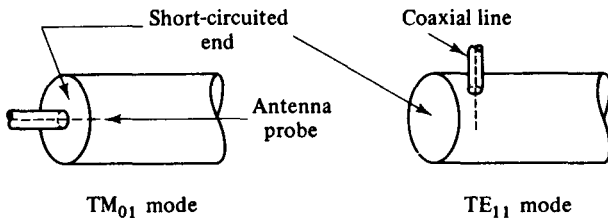
Figure 4-2-7   Methods of exciting various modes in circular waveguides.

A common way to excite TM modes in a circular guide is by a coaxial line as shown in Fig. 4-2-8. At the end of the coaxial line a large magnetic intensity exists in the ϕ direction of wave propagation. The magnetic field from the coaxial line will excite the TM modes in the guide. However, when the guide is connected to the source by a coaxial line, a discontinuity problem at the junction will increase the standing-wave ratio on the line and eventually decrease the power transmission. It is often necessary to place a turning device around the junction in order to suppress the reflection.
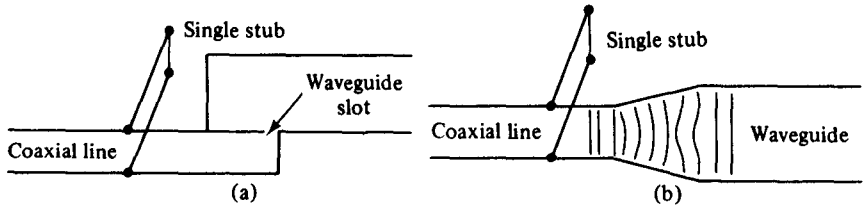
Figure 4-2-8   Methods of exciting TM modes in a circular waveguide. (a) Coaxial line with a slotted waveguide. (b) Coaxial line in series with a circular waveguide.

TABLE 4-2-8   CHARACTERISTICS OF STANDARD CIRCULAR WAVEGUIDES

EIA[a] designation WC[b]( )	Inside diameter in cm (in.)		Cutoff frequency for air-filled waveguide in GHz	Recommended frequency range for TE$_{11}$ mode in GHz
992	25.184	(9.915)	0.698	0.80–1.10
847	21.514	(8.470)	0.817	0.94–1.29
724	18.377	(7.235)	0.957	1.10–1.51
618	15.700	(6.181)	1.120	1.29–1.76
528	13.411	(5.280)	1.311	1.51–2.07
451	11.458	(4.511)	1.534	1.76–2.42
385	9.787	(3.853)	1.796	2.07–2.83
329	8.362	(3.292)	2.102	2.42–3.31
281	7.142	(2.812)	2.461	2.83–3.88
240	6.104	(2.403)	2.880	3.31–4.54
205	5.199	(2.047)	3.381	3.89–5.33
175	4.445	(1.750)	3.955	4.54–6.23
150	3.810	(1.500)	4.614	5.30–7.27
128	3.254	(1.281)	5.402	6.21–8.51
109	2.779	(1.094)	6.326	7.27–9.97
94	2.383	(0.938)	7.377	8.49–11.60
80	2.024	(0.797)	8.685	9.97–13.70
69	1.748	(0.688)	10.057	11.60–15.90
59	1.509	(0.594)	11.649	13.40–18.40
50	1.270	(0.500)	13.842	15.90–21.80
44	1.113	(0.438)	15.794	18.20–24.90
38	0.953	(0.375)	18.446	21.20–29.10
33	0.833	(0.328)	21.103	24.30–33.20
28	0.714	(0.281)	24.620	28.30–38.80
25	0.635	(0.250)	27.683	31.80–43.60
22	0.556	(0.219)	31.617	36.40–49.80
19	0.478	(0.188)	36.776	42.40–58.10
17	0.437	(0.172)	40.227	46.30–63.50
14	0.358	(0.141)	49.103	56.60–77.50
13	0.318	(0.125)	55.280	63.50–87.20
11	0.277	(0.109)	63.462	72.70–99.70
9	0.239	(0.094)	73.552	84.80–116.00

[a] Electronic Industry Association
[b] Circular Waveguide

4-2-8 Characteristics of Standard Circular Waveguides

The inner diameter of a circular waveguide is regulated by the frequency of the signal being transmitted. For example: at X-band frequencies from 8 to 12 GHz, the inner diameter of a circular waveguide designated as EIA WC(94) by the Electronic Industry Association is 2.383 cm (0.938 in.). Table 4-2-8 tabulates the characteristics of the standard circular waveguides.

4-3 MICROWAVE CAVITIES

In general, a cavity resonator is a metallic enclosure that confines the electromagnetic energy. The stored electric and magnetic energies inside the cavity determine its equivalent inductance and capacitance. The energy dissipated by the finite conductivity of the cavity walls determines its equivalent resistance. In practice, the rectangular-cavity resonator, circular-cavity resonator, and reentrant-cavity resonator are commonly used in many microwave applications.

Theoretically a given resonator has an infinite number of resonant modes, and each mode corresponds to a definite resonant frequency. When the frequency of an impressed signal is equal to a resonant frequency, a maximum amplitude of the standing wave occurs, and the peak energies stored in the electric and magnetic fields are equal. The mode having the lowest resonant frequency is known as the *dominant mode*.

4-3-1 Rectangular-Cavity Resonator

The electromagnetic field inside the cavity should satisfy Maxwell's equations, subject to the boundary conditions that the electric field tangential to and the magnetic field normal to the metal walls must vanish. The geometry of a rectangular cavity is illustrated in Fig. 4-3-1.

The wave equations in the rectangular resonator should satisfy the boundary condition of the zero tangential **E** at four of the walls. It is merely necessary to choose the harmonic functions in z to satisfy this condition at the remaining two end
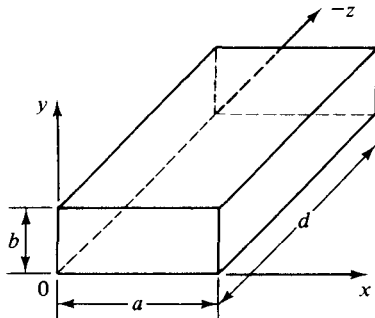
Figure 4-3-1   Coordinates of a rectangular cavity.

walls. These functions can be found if

$$H_z = H_{0z} \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \sin\left(\frac{p\pi z}{d}\right) \qquad (\text{TE}_{mnp}) \qquad (4\text{-}3\text{-}1)$$

where $m = 0, 1, 2, 3, \ldots$ represents the number of the half-wave periodicity in the x direction

$n = 0, 1, 2, 3, \ldots$ represents the number of the half-wave periodicity in the y direction

$p = 1, 2, 3, 4, \ldots$ represents the number of the half-wave periodicity in the z direction

and

$$E_z = E_{0z} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi x}{b}\right) \cos\left(\frac{p\pi z}{d}\right) \qquad (\text{TM}_{mnp}) \qquad (4\text{-}3\text{-}2)$$

where $m = 1, 2, 3, 4, \ldots$

$n = 1, 2, 3, 4, \ldots$

$p = 0, 1, 2, 3, \ldots$

The separation equation for both TE and TM modes is given by

$$k^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 + \left(\frac{p\pi}{d}\right)^2 \qquad (4\text{-}3\text{-}3)$$

For a lossless dielectric, $k^2 = \omega^2 \mu\epsilon$; therefore, the resonant frequency is expressed by

$$f_r = \frac{1}{2\sqrt{\mu\epsilon}} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2 + \left(\frac{p}{d}\right)^2} \qquad (\text{TE}_{mnp}, \text{TM}_{mnp}) \qquad (4\text{-}3\text{-}4)$$

For $a > b < d$, the dominant mode is the TE_{101} mode.

In general, a straight-wire probe inserted at the position of maximum electric intensity is used to excite a desired mode, and the loop coupling placed at the position of maximum magnetic intensity is utilized to launch a specific mode. Figure 4-3-2 shows the methods of excitation for the rectangular resonator. The maximum amplitude of the standing wave occurs when the frequency of the impressed signal is equal to the resonant frequency.
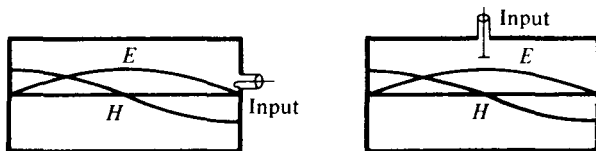
Figure 4-3-2   Methods of exciting wave modes in a resonator.

4-3-2 Circular-Cavity Resonator and Semicircular-Cavity Resonator

Circular-cavity resonator.   A circular-cavity resonator is a circular waveguide with two ends closed by a metal wall (see Fig. 4-3-3). The wave function in the circular resonator should satisfy Maxwell's equations, subject to the same
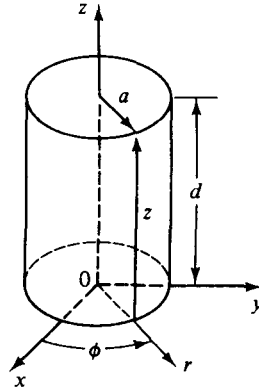
Figure 4-3-3 Coordinates of a circular resonator.

boundary conditions described for a rectangular-cavity resonator. It is merely necessary to choose the harmonic functions in z to satisfy the boundary conditions at the remaining two end walls. These can be achieved if

$$H_z = H_{0z} J_n\left(\frac{X'_{np} r}{a}\right) \cos(n\phi) \sin\left(\frac{q\pi z}{d}\right) \qquad (\text{TE}_{npq}) \qquad (4\text{-}3\text{-}5)$$

where $n = 0, 1, 2, 3, \ldots$ is the number of the periodicity in the ϕ direction
$\qquad p = 1, 2, 3, 4, \ldots$ is the number of zeros of the field in the radial direction
$\qquad q = 1, 2, 3, 4, \ldots$ is the number of half-waves in the axial direction
$\qquad J_n$ = Bessell's function of the first kind
$\qquad H_{0z}$ = amplitude of the magnetic field

and

$$E_z = E_{0z} J_n\left(\frac{X_{np} r}{a}\right) \cos(n\phi) \cos\left(\frac{q\pi z}{d}\right) \qquad (\text{TM}_{npq}) \qquad (4\text{-}3\text{-}6)$$

where $n = 0, 1, 2, 3, \ldots$
$\qquad p = 1, 2, 3, 4, \ldots$
$\qquad q = 0, 1, 2, 3, \ldots$
$\qquad E_{0z}$ = amplitude of the electric field

The separation equations for TE and TM modes are given by

$$k^2 = \left(\frac{X'_{np}}{a}\right)^2 + \left(\frac{q\pi}{d}\right)^2 \qquad (\text{TE mode}) \qquad (4\text{-}3\text{-}7)$$

$$k^2 = \left(\frac{X_{np}}{a}\right)^2 + \left(\frac{q\pi}{d}\right)^2 \qquad (\text{TM mode}) \qquad (4\text{-}3\text{-}8)$$

Substitution of $k^2 = \omega^2 \mu\epsilon$ in Eqs. (4-3-7) and (4-3-8) yields the resonant frequencies for TE and TM modes, respectively, as

$$f_r = \frac{1}{2\pi\sqrt{\mu\epsilon}} \sqrt{\left(\frac{X'_{np}}{a}\right)^2 + \left(\frac{q\pi}{d}\right)^2} \qquad (\text{TE}) \qquad (4\text{-}3\text{-}9)$$

$$f_r = \frac{1}{2\pi\sqrt{\mu\epsilon}}\sqrt{\left(\frac{X_{np}}{a}\right)^2 + \left(\frac{q\pi}{d}\right)^2} \qquad \text{(TM)} \qquad (4\text{-}3\text{-}10)$$

It is interesting to note that the TM_{110} mode is dominant where $2a > d$, and that the TE_{111} mode is dominant when $d \geq 2a$.

Semicircular-cavity resonator. A semicircular-cavity resonator is shown in Fig. 4-3-4. The wave function of the TE_{npq} mode in the semicircular resonator can be written

$$H_z = H_{0z}J_n\left(\frac{X'_{np}r}{a}\right)\cos(n\phi)\sin\left(\frac{q\pi z}{d}\right) \qquad \text{(TE mode)} \qquad (4\text{-}3\text{-}11)$$

where $n = 0, 1, 2, 3, \ldots$
$\quad p = 1, 2, 3, 4, \ldots$
$\quad q = 1, 2, 3, 4, \ldots$
$\quad a = $ radius of the semicircular-cavity resonator
$\quad d = $ length of the resonator
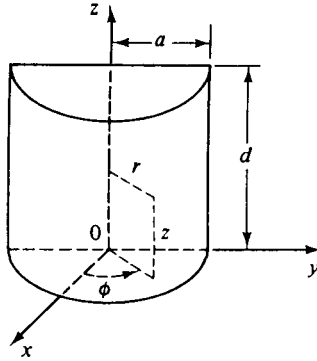
Figure 4-3-4 Semicircular resonator.

The wave function of the TM_{npq} mode in the semicircular-cavity resonator can be written

$$E_z = E_{02}J_n\left(\frac{X_{np}r}{a}\right)\sin(n\phi)\cos\left(\frac{q\pi}{d}z\right) \qquad \text{(TM mode)} \qquad (4\text{-}3\text{-}12)$$

where $n = 1, 2, 3, 4, \ldots$
$\quad p = 1, 2, 3, 4, \ldots$
$\quad q = 0, 1, 2, 3, \ldots$

With the separation equations given in Eqs. (4-3-7) and (4-3-8), the equations of resonant frequency for TE and TM modes in a semicircular-cavity resonator are the same as in the circular-cavity resonator. They are repeated as follows:

$$f_r = \frac{1}{2\pi a\sqrt{\mu\epsilon}}\sqrt{(X'_{np})^2 + \left(\frac{q\pi a}{d}\right)^2} \qquad (TE_{npq} \text{ mode}) \qquad (4\text{-}3\text{-}13)$$

$$f_r = \frac{1}{2\pi a \sqrt{\mu\epsilon}} \sqrt{(X_{np})^2 + \left(\frac{q\pi a}{d}\right)^2} \quad (\text{TM}_{npq} \text{ mode}) \qquad (4\text{-}3\text{-}14)$$

However, the values of the subscripts n, p, and q differ from those for the circular-cavity resonator. Also, it must be emphasized that the TE_{111} mode is dominant if $d > a$ and that the TM_{110} mode is dominant if $d < a$.

4-3-3 Q Factor of a Cavity Resonator

The quality factor Q is a measure of the frequency selectivity of a resonant or antiresonant circuit, and it is defined as

$$Q \equiv 2\pi \frac{\text{maximum energy stored}}{\text{energy dissipated per cycle}} = \frac{\omega W}{P} \qquad (4\text{-}3\text{-}15)$$

where W is the maximum stored energy and P is the average power loss.

At resonant frequency, the electric and magnetic energies are equal and in time quadrature. When the electric energy is maximum, the magnetic energy is zero and vice versa. The total energy stored in the resonator is obtained by integrating the energy density over the volume of the resonator:

$$W_e = \int_v \frac{\epsilon}{2} |E|^2 \, dv = W_m = \int_v \frac{\mu}{2} |H|^2 \, dv = W \qquad (4\text{-}3\text{-}16)$$

where $|E|$ and $|H|$ are the peak values of the field intensities.

The average power loss in the resonator can be evaluated by integrating the power density as given in Eq. (2-5-12) over the inner surface of the resonator. Hence

$$P = \frac{R_s}{2} \int_s |H_t|^2 \, da \qquad (4\text{-}3\text{-}17)$$

where H_t is the peak value of the tangential magnetic intensity and R_s is the surface resistance of the resonator.

Substitution of Eqs. (4-3-16) and (4-3-17) in Eq. (4-3-15) yields

$$Q = \frac{\omega\mu \int_v |H|^2 \, dv}{R_s \int_s |H_t|^2 \, da} \qquad (4\text{-}3\text{-}18)$$

Since the peak value of the magnetic intensity is related to its tangential and normal components by

$$|H|^2 = |H_t|^2 + |H_n|^2$$

where H_n is the peak value of the normal magnetic intensity, the value of $|H_t|^2$ at the resonator walls is approximately twice the value of $|H|^2$ averaged over the volume. So the Q of a cavity resonator as shown in Eq. (4-3-18) can be expressed approximately by

$$Q = \frac{\omega\mu \, (\text{volume})}{2R_s (\text{surface areas})} \qquad (4\text{-}3\text{-}19)$$

An unloaded resonator can be represented by either a series or a parallel resonant circuit. The resonant frequency and the unloaded Q_0 of a cavity resonator are

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{4-3-20}$$

$$Q_0 = \frac{\omega_0 L}{R} \tag{4-3-21}$$

If the cavity is coupled by means of an ideal $N{:}1$ transformer and a series inductance L_s to a generator having internal impedance Z_g, then the coupling circuit and its equivalent are as shown in Fig. 4-3-5.
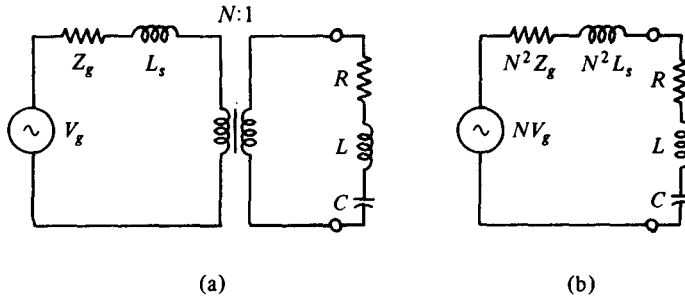
Figure 4-3-5 Cavity coupled to a generator. (a) Coupling circuit. (b) Equivalent circuit.

The loaded Q_ℓ of the system is given by

$$Q_\ell = \frac{\omega_0 L}{R + N^2 Z_g} \qquad \text{for } |N^2 L_s| \ll |R + N^2 Z_g| \tag{4-3-22}$$

The coupling coefficient of the system is defined as

$$K \equiv \frac{N^2 Z_g}{R} \tag{4-3-23}$$

and the loaded Q_ℓ would become

$$Q_\ell = \frac{\omega_0 L}{R(1 + K)} = \frac{Q_0}{1 + K} \tag{4-3-24}$$

Rearrangement of Eq. (4-3-24) yields

$$\frac{1}{Q_\ell} = \frac{1}{Q_0} + \frac{1}{Q_{\text{ext}}} \tag{4-3-25}$$

where $Q_{\text{ext}} = Q_0/K = \omega_0 L/(KR)$ is the external Q.

There are three types of coupling coefficients:

1. Critical coupling: If the resonator is matched to the generator, then

$$K = 1 \tag{4-3-26}$$

The loaded Q_ℓ is given by

$$Q_\ell = \tfrac{1}{2}Q_{\text{ext}} = \tfrac{1}{2}Q_0 \qquad (4\text{-}3\text{-}27)$$

2. **Overcoupling:** If $K > 1$, the cavity terminals are at a voltage maximum in the input line at resonance. The normalized impedance at the voltage maximum is the standing-wave ratio ρ. That is

$$K = \rho \qquad (4\text{-}3\text{-}28)$$

The loaded Q_ℓ is given by

$$Q_\ell = \frac{Q_0}{1 + \rho} \qquad (4\text{-}3\text{-}29)$$

3. **Undercoupling:** If $K < 1$, the cavity terminals are at a voltage minimum and the input terminal impedance is equal to the reciprocal of the standing-wave ratio. That is,

$$K = \frac{1}{\rho} \qquad (4\text{-}3\text{-}30)$$

The loaded Q_ℓ is given by

$$Q_\ell = \frac{\rho}{\rho + 1}Q_0 \qquad (4\text{-}3\text{-}31)$$

The relationship of the coupling coefficient K and the standing-wave ratio is shown in Fig. 4-3-6.
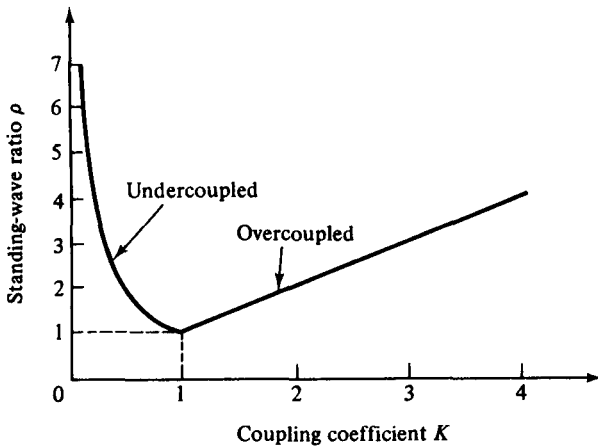
Figure 4-3-6 Coupling coefficient versus standing-wave ratio.

4-4 MICROWAVE HYBRID CIRCUITS

A microwave circuit ordinarily consists of several microwave devices connected in some way to achieve the desired transmission of a microwave signal. The interconnection of two or more microwave devices may be regarded as a microwave junction. Commonly used microwave junctions include such waveguide tees as the *E*-

plane tee, H-plane tee, magic tee, hybrid ring (rat-race circuit), directional coupler, and the circulator. This section describes these microwave hybrids, which are shown in Fig. 4-4-1.

A two-port network is shown in Fig. 4-4-2. From network theory a two-port device can be described by a number of parameter sets, such as the H, Y, Z, and $ABCD$ parameters.

H parameters:
$$\begin{bmatrix} V_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ V_2 \end{bmatrix}$$

$$V_1 = h_{11}I_1 + h_{12}V_2 \qquad (4\text{-}4\text{-}1)$$
$$I_2 = h_{21}I_1 + h_{22}V_2 \qquad (4\text{-}4\text{-}2)$$

Y parameters:
$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

$$I_1 = y_{11}V_1 + y_{12}V_2 \qquad (4\text{-}4\text{-}3)$$
$$I_2 = y_{21}V_1 + y_{22}V_1 \qquad (4\text{-}4\text{-}4)$$

(a) E-plane tee

(b) H-plane tee

(c) Magic tee

(d) Hybrid ring

(e) Directional coupler
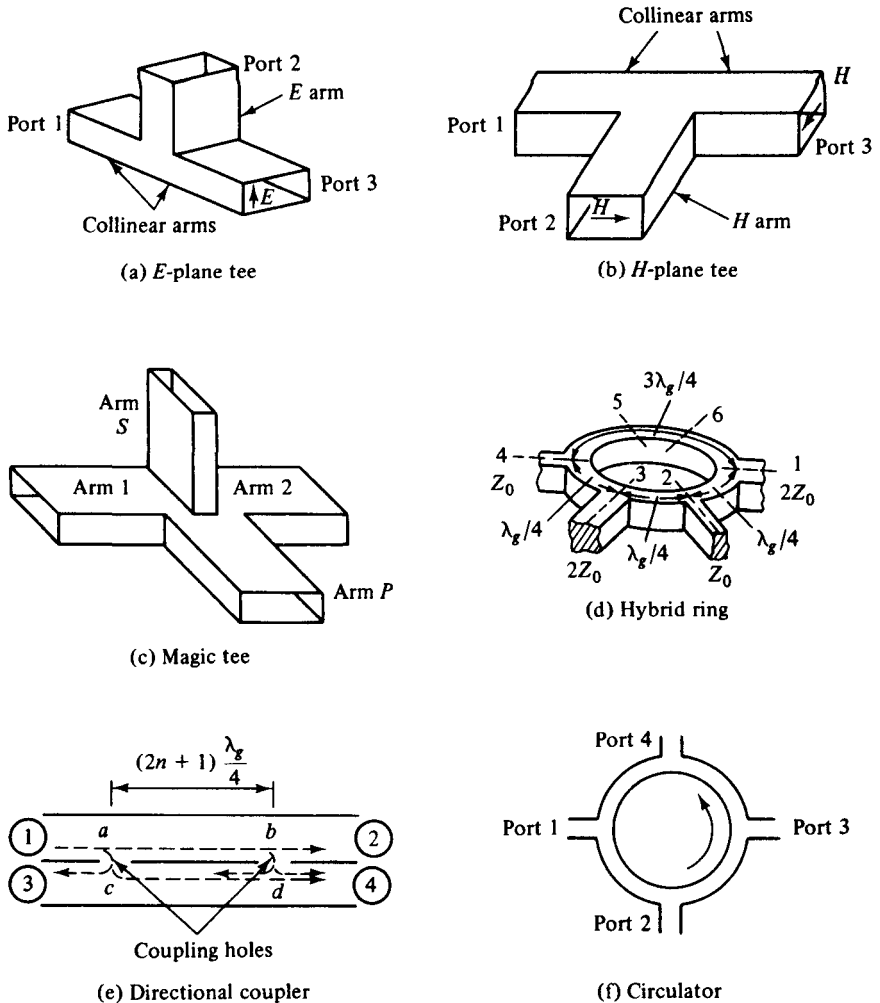
(f) Circulator

Figure 4-4-1   Microwave hybrids. (a) E-plane tee. (b) H-plane tee. (c) Magic tee. (d) Hybrid ring. (e) Directional coupler. (f) Circulator.
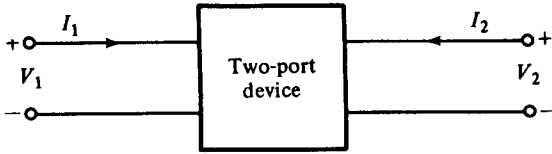
Figure 4-4-2   Two-port network.

$$Z \text{ parameters:} \quad \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \qquad \begin{aligned} V_1 &= z_{11}I_1 + z_{12}I_2 & (4\text{-}4\text{-}5) \\ V_2 &= z_{21}I_1 + z_{22}I_2 & (4\text{-}4\text{-}6) \end{aligned}$$

$$ABCD \text{ parameters:} \quad \begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix} \qquad \begin{aligned} V_1 &= AV_2 - BI_2 & (4\text{-}4\text{-}7) \\ I_1 &= CV_2 - DI_2 & (4\text{-}4\text{-}8) \end{aligned}$$

All these network parameters relate total voltages and total currents at each of the two ports. For instance,

$$h_{11} = \frac{V_1}{I_1}\bigg|_{V_2=0} \qquad \text{(short circuit)} \qquad (4\text{-}4\text{-}9)$$

$$h_{12} = \frac{V_1}{V_2}\bigg|_{I_1=0} \qquad \text{(open circuit)} \qquad (4\text{-}4\text{-}10)$$

If the frequencies are in the microwave range, however, the H, Y, and Z parameters cannot be measured for the following reasons:

1. Equipment is not readily available to measure total voltage and total current at the ports of the network.
2. Short and open circuits are difficult to achieve over a broad band of frequencies.
3. Active devices, such as power transistors and tunnel diodes, frequently will not have stability for a short or open circuit.

Consequently, some new method of characterization is needed to overcome these problems. The logical variables to use at the microwave frequencies are traveling waves rather than total voltages and total currents. These are the S parameters, which are expressed as

$$b_1 = S_{11}a_1 + S_{12}a_2 \qquad (4\text{-}4\text{-}11\text{a})$$

$$b_2 = S_{21}a_1 + S_{22}a_2 \qquad (4\text{-}4\text{-}11\text{b})$$

Figure 4-4-3 shows the S parameters of a two-port network.
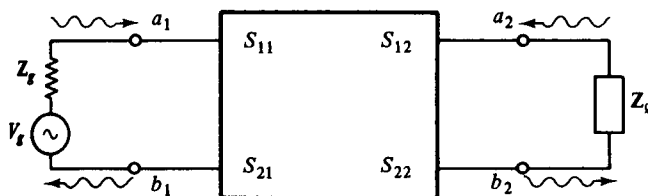
Figure 4-4-3   Two-port network.

4-4-1 Waveguide Tees

As noted, waveguide tees may consist of the E-plane tee, H-plane tee, magic tee, hybrid rings, corners, bends, and twists. All such waveguide components are discussed in this section.

Tee junctions.   In microwave circuits a waveguide or coaxial-line junction with three independent ports is commonly referred to as a *tee junction*. From the S-parameter theory of a microwave junction it is evident that a tee junction should be characterized by a matrix of third order containing nine elements, six of which should be independent. The characteristics of a three-port junction can be explained by three theorems of the tee junction. These theorems are derived from the equivalent-circuit representation of the tee junction. Their statements follow

1. A short circuit may always be placed in one of the arms of a three-port junction in such a way that no power can be transferred through the other two arms.
2. If the junction is symmetric about one of its arms, a short circuit can always be placed in that arm so that no reflections occur in power transmission between the other two arms. (That is, the arms present matched impedances.)
3. It is impossible for a general three-port junction of arbitrary symmetry to present matched impedances at all three arms.

The E-plane tee and H-plane tee are described below.

E-plane tee (series tee).   An E-plane tee is a waveguide tee in which the axis of its side arm is parallel to the E field of the main guide (see Fig. 4-4-4). If the collinear arms are symmetric about the side arm, there are two different transmission characteristics (see Fig. 4-4-5). It can be seen from Fig. 4-4-4 that if the E-plane tee is perfectly matched with the aid of screw tuners or inductive or capacitive windows at the junction, the diagonal components of the scattering matrix, S_{11}, S_{22}, and S_{33}, are zero because there will be no reflection. When the waves are fed into the side arm (port 3), the waves appearing at port 1 and port 2 of the collinear arm will be in opposite phase and in the same magnitude. Therefore

$$S_{13} = -S_{23} \tag{4-4-12}$$

It should be noted that Eq. (4-4-12) does not mean that S_{13} is always positive and S_{23}
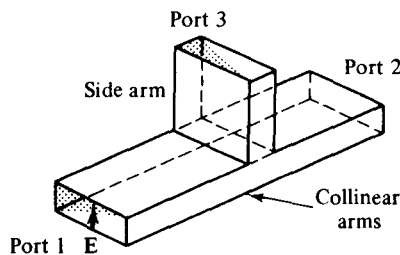
Figure 4-4-4 E-plane tee
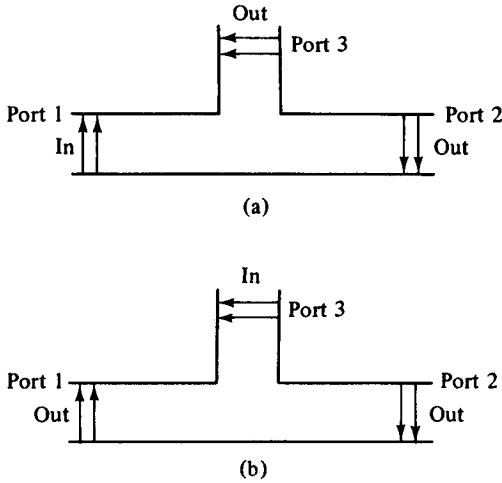
Figure 4-4-5 Two-way transmission of E-plane tee. (a) Input through main arm. (b) Input from side arm.

is always negative. The negative sign merely means that S_{13} and S_{23} have opposite signs. For a matched junction, the **S** matrix is given by

$$\mathbf{S} = \begin{bmatrix} 0 & S_{12} & S_{13} \\ S_{21} & 0 & S_{23} \\ S_{31} & S_{32} & 0 \end{bmatrix} \tag{4-4-13}$$

From the symmetry property of **S** matrix, the symmetric terms in Eq. (4-4-13) are equal and they are

$$S_{12} = S_{21} \qquad S_{13} = S_{31} \qquad S_{23} = S_{32} \tag{4-4-14}$$

From the zero property of **S** matrix, the sum of the products of each term of any column (or row) multiplied by the complex conjugate of the corresponding terms of any other column (or row) is zero and it is

$$S_{11} S_{12}^* + S_{21} S_{22}^* + S_{31} S_{32}^* = 0 \tag{4-4-15}$$

Hence

$$S_{13} S_{23}^* = 0 \tag{4-4-16}$$

This means that either S_{13} or S_{23}^*, or both, should be zero. However, from the unity property of **S** matrix, the sum of the products of each term of any one row (or column) multiplied by its complex conjugate is unity; that is,

$$S_{21} S_{21}^* + S_{31} S_{31}^* = 1 \tag{4-4-17}$$

$$S_{12} S_{12}^* + S_{32} S_{32}^* = 1 \tag{4-4-18}$$

$$S_{13} S_{13}^* + S_{23} S_{23}^* = 1 \tag{4-4-19}$$

Substitution of Eq. (4-4-14) in (4-4-17) results in

$$|S_{12}|^2 = 1 - |S_{13}|^2 = 1 - |S_{23}|^2 \tag{4-4-20}$$

Equations (4-4-19) and (4-4-20) are contradictory, for if $S_{13} = 0$, then S_{23} is also

zero and thus Eq. (4-4-19) is false. In a similar fashion, if $S_{23} = 0$, then S_{13} becomes zero and therefore Eq. (4-4-20) is not true. This inconsistency proves the statement that the tee junction cannot be matched to the three arms. In other words, the diagonal elements of the **S** matrix of a tee junction are not all zeros.

In general, when an E-plane tee is constructed of an empty waveguide, it is poorly matched at the tee junction. Hence $S_{ij} \neq 0$ if $i = j$. However, since the collinear arm is usually symmetric about the side arm, $|S_{13}| = |S_{23}|$ and $S_{11} = S_{22}$. Then the **S** matrix can be simplified to

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{12} & S_{11} & -S_{13} \\ S_{13} & -S_{13} & S_{33} \end{bmatrix} \tag{4-4-21}$$

H-plane tee (shunt tee). An H-plane tee is a waveguide tee in which the axis of its side arm is "shunting" the E field or parallel to the H field of the main guide as shown in Fig. 4-4-6.
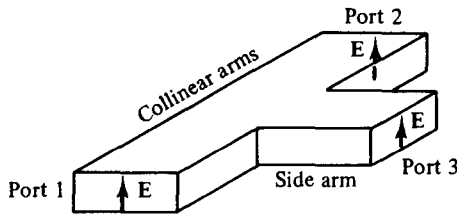
Figure 4-4-6 H-plane tee.

It can be seen that if two input waves are fed into port 1 and port 2 of the collinear arm, the output wave at port 3 will be in phase and additive. On the other hand, if the input is fed into port 3, the wave will split equally into port 1 and port 2 in phase and in the same magnitude. Therefore the **S** matrix of the H-plane tee is similar to Eqs. (4-4-13) and (4-4-21) except that

$$S_{13} = S_{23} \tag{4-4-22}$$

4-4-2 Magic Tees (Hybrid Tees)

A magic tee is a combination of the E-plane tee and H-plane tee (refer to Fig. 4-4-7). The magic tee has several characteristics:
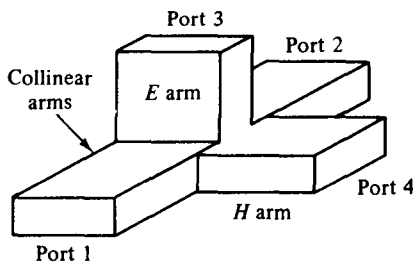
Figure 4-4-7 Magic tee.

1. If two waves of equal magnitude and the same phase are fed into port 1 and port 2, the output will be zero at port 3 and additive at port 4.
2. If a wave is fed into port 4 (the H arm), it will be divided equally between port 1 and port 2 of the collinear arms and will not appear at port 3 (the E arm).
3. If a wave is fed into port 3 (the E arm), it will produce an output of equal magnitude and opposite phase at port 1 and port 2. The output at port 4 is zero. That is, $S_{43} = S_{34} = 0$.
4. If a wave is fed into one of the collinear arms at port 1 or port 2, it will not appear in the other collinear arm at port 2 or port 1 because the E arm causes a phase delay while the H arm causes a phase advance. That is, $S_{12} = S_{21} = 0$.

Therefore the **S** matrix of a magic tee can be expressed as

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & S_{13} & S_{14} \\ 0 & 0 & S_{23} & S_{24} \\ S_{31} & S_{32} & 0 & 0 \\ S_{41} & S_{42} & 0 & 0 \end{bmatrix} \tag{4-4-23}$$

The magic tee is commonly used for mixing, duplexing, and impedance measurements. Suppose, for example, there are two identical radar transmitters in equipment stock. A particular application requires twice more input power to an antenna than either transmitter can deliver. A magic tee may be used to couple the two transmitters to the antenna in such a way that the transmitters do not load each other. The two transmitters should be connected to ports 3 and 4, respectively, as shown in Fig. 4-4-8. Transmitter 1, connected to port 3, causes a wave to emanate from port 1 and another to emanate from port 2; these waves are equal in magnitude but opposite in phase. Similarly, transmitter 2, connected to port 4, gives rise to a wave at port 1 and another at port 2, both equal in magnitude and in phase. At port 1 the two opposite waves cancel each other. At port 2 the two in-phase waves add together; so double output power at port 2 is obtained for the antenna as shown in Fig. 4-4-8.
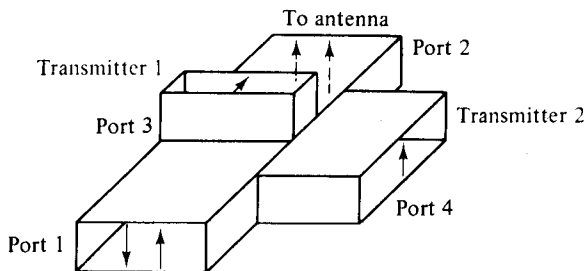
Figure 4-4-8 Magic tee-coupled transmitters to antenna.

4-4-3 Hybrid Rings (Rat-Race Circuits)

A hybrid ring consists of an annular line of proper electrical length to sustain standing waves, to which four arms are connected at proper intervals by means of series or parallel junctions. Figure 4-4-9 shows a hybrid ring with series junctions.
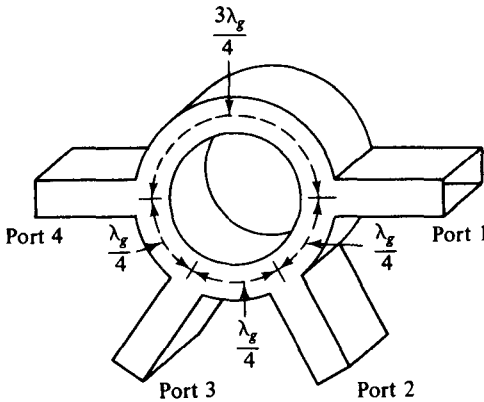
Figure 4-4-9 Hybrid ring.

The hybrid ring has characteristics similar to those of the hybrid tee. When a wave is fed into port 1, it will not appear at port 3 because the difference of phase shifts for the waves traveling in the clockwise and counterclockwise directions is 180°. Thus the waves are canceled at port 3. For the same reason, the waves fed into port 2 will not emerge at port 4 and so on.

The **S** matrix for an ideal hybrid ring can be expressed as

$$\mathbf{S} = \begin{bmatrix} 0 & S_{12} & 0 & S_{14} \\ S_{21} & 0 & S_{23} & 0 \\ 0 & S_{32} & 0 & S_{34} \\ S_{41} & 0 & S_{43} & 0 \end{bmatrix} \tag{4-4-24}$$

It should be noted that the phase cancellation ocurs only at a designated frequency for an ideal hybrid ring. In actual hybrid rings there are small leakage couplings, and therefore the zero elements in the matrix of Eq. (4-4-24) are not quite equal to zero.

4-4-4 Waveguide Corners, Bends, and Twists

The waveguide corner, bend, and twist are shown in Fig. 4-4-10. These waveguide components are normally used to change the direction of the guide through an arbitrary angle.

In order to minimize reflections from the discontinuities, it is desirable to have the mean length L between continuities equal to an odd number of quarter-wavelengths. That is,

$$L = (2n + 1)\frac{\lambda_g}{4} \tag{4-4-25}$$

where $n = 0, 1, 2, 3, \ldots$, and λ_g is the wavelength in the waveguide. If the mean length L is an odd number of quarter wavelengths, the reflected waves from both ends of the waveguide section are completely canceled. For the waveguide bend, the minimum radius of curvature for a small reflection is given by Southworth [2] as

$$R = 1.5b \qquad \text{for an } E \text{ bend} \tag{4-4-26}$$

$$R = 1.5a \qquad \text{for an } H \text{ bend} \qquad (4\text{-}4\text{-}27)$$

where a and b are the dimensions of the waveguide bend as illustrated in Fig. 4-4-10.
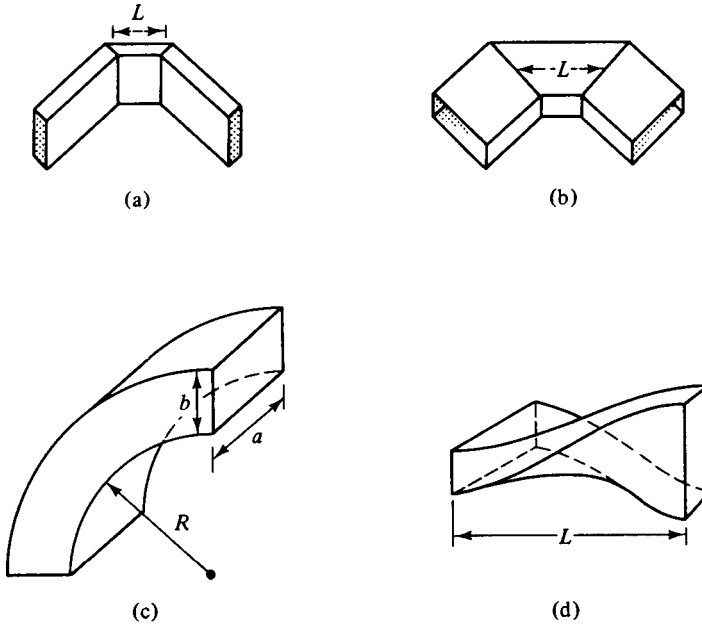
Figure 4-4-10   Waveguide corner, bend, and twist. (a) E-plane corner. (b) H-plane corner. (c) Bend. (d) Continuous twist.

4-5 DIRECTIONAL COUPLERS

A *directional coupler* is a four-port waveguide junction as shown in Fig. 4-5-1. It consists of a primary waveguide 1–2 and a secondary waveguide 3–4. When all ports are terminated in their characteristic impedances, there is free transmission of power, without reflection, between port 1 and port 2, and there is no transmission of power between port 1 and port 3 or between port 2 and port 4 because no coupling exists between these two pairs of ports. The degree of coupling between port 1 and port 4 and between port 2 and port 3 depends on the structure of the coupler.

The characteristics of a directional coupler can be expressed in terms of its coupling factor and its directivity. Assuming that the wave is propagating from port 1 to port 2 in the primary line, the coupling factor and the directivity are defined,
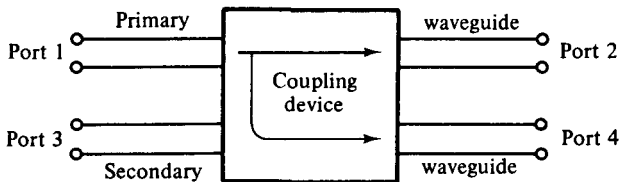
Figure 4-5-1   Directional coupler.

respectively, by

$$\text{Coupling factor (dB)} = 10 \log_{10} \frac{P_1}{P_4} \qquad (4\text{-}5\text{-}1)$$

$$\text{Directivity (dB)} = 10 \log_{10} \frac{P_4}{P_3} \qquad (4\text{-}5\text{-}2)$$

where P_1 = power input to port 1
    P_3 = power output from port 3
    P_4 = power output from port 4

It should be noted that port 2, port 3, and port 4 are terminated in their characteristic impedances. *The coupling factor is a measure of the ratio of power levels* in the primary and secondary lines. Hence if the coupling factor is known, a fraction of power measured at port 4 may be used to determine the power input at port 1. This significance is desirable for microwave power measurements because no disturbance, which may be caused by the power measurements, occurs in the primary line. The directivity is a measure of how well the forward traveling wave in the primary waveguide couples only to a specific port of the secondary waveguide. An ideal directional coupler should have infinite directivity. In other words, the power at port 3 must be zero because port 2 and port 4 are perfectly matched. Actually, well-designed directional couplers have a directivity of only 30 to 35 dB.

Several types of directional couplers exist, such as a two-hole directional couler, four-hole directional coupler, reverse-coupling directional coupler (Schwinger coupler), and Bethe-hole directional coupler (refer to Fig. 4-5-2). Only the very commonly used two-hole directional coupler is described here.
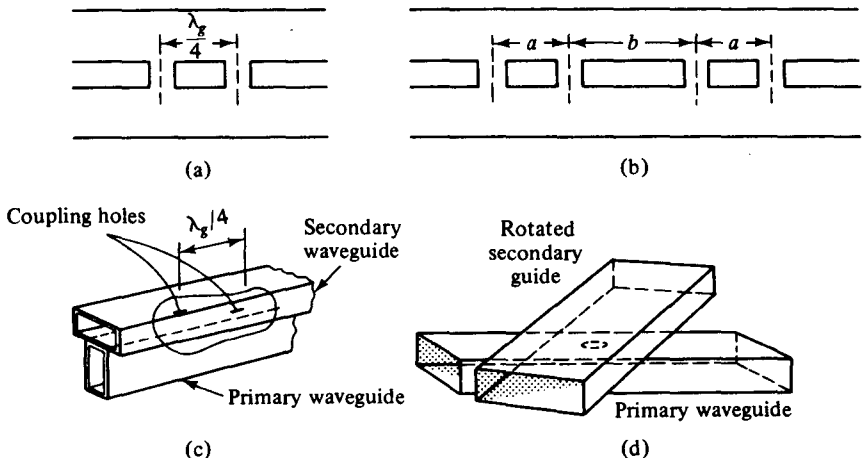
Figure 4-5-2 Different directional couplers. (a) Two-hole directional coupler. (b) Four-hole directional coupler. (c) Schwinger coupler. (d) Bethe-hole directional coupler.

4-5-1 Two-Hole Directional Couplers

A two-hole directional coupler with traveling waves propagating in it is illustrated in Fig. 4-5-3. The spacing between the centers of two holes must be

$$L = (2n + 1)\frac{\lambda_g}{4} \qquad (4\text{-}5\text{-}3)$$

where n is any positive integer.
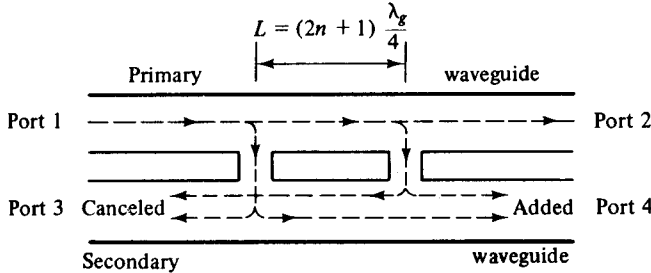
Figure 4-5-3   Two-hole directional coupler.

A fraction of the wave energy entered into port 1 passes through the holes and is radiated into the secondary guide as the holes act as slot antennas. The forward waves in the secondary guide are in the same phase, regardless of the hole space, and are added at port 4. The backward waves in the secondary guide (waves are progressing from right to left) are out of phase by $(2L/\lambda_g)2\pi$ rad and are canceled at port 3.

4-5-2 S Matrix of a Directional Coupler

In a directional coupler all four ports are completely matched. Thus the diagonal elements of the **S** matrix are zeros and

$$S_{11} = S_{22} = S_{33} = S_{44} = 0 \qquad (4\text{-}5\text{-}4)$$

As noted, there is no coupling between port 1 and port 3 and between port 2 and port 4. Thus

$$S_{13} = S_{31} = S_{24} = S_{42} = 0 \qquad (4\text{-}5\text{-}5)$$

Consequently, the **S** matrix of a directional coupler becomes

$$\mathbf{S} = \begin{bmatrix} 0 & S_{12} & 0 & S_{14} \\ S_{21} & 0 & S_{23} & 0 \\ 0 & S_{32} & 0 & S_{34} \\ S_{41} & 0 & S_{43} & 0 \end{bmatrix} \qquad (4\text{-}5\text{-}6)$$

Equation (4-5-6) can be further reduced by means of the zero property of the **S** matrix, so we have

$$S_{12}S_{14}^* + S_{32}S_{34}^* = 0 \qquad (4\text{-}5\text{-}7)$$

$$S_{21} S_{23}^* + S_{41} S_{43}^* = 0 \tag{4-5-8}$$

Also from the unity property of the **S** matrix, we can write

$$S_{12} S_{12}^* + S_{14} S_{14}^* = 1 \tag{4-5-9}$$

Equations (4-5-7) and (4-5-8) can also be written

$$|S_{12}||S_{14}| = |S_{32}||S_{34}| \tag{4-5-10}$$

$$|S_{21}||S_{23}| = |S_{41}||S_{43}| \tag{4-5-11}$$

Since $S_{12} = S_{21}$, $S_{14} = S_{41}$, $S_{23} = S_{32}$, and $S_{34} = S_{43}$, then

$$|S_{12}| = |S_{34}| \tag{4-5-12}$$

$$|S_{14}| = |S_{23}| \tag{4-5-13}$$

Let

$$S_{12} = S_{34} = p \tag{4-5-14}$$

where p is positive and real. Then from Eq. (4-5-8)

$$p(S_{23}^* + S_{41}) = 0 \tag{4-5-15}$$

Let

$$S_{23} = S_{41} = jq \tag{4-5-16}$$

where q is positive and real. Then from Eq. (4-5-9)

$$p^2 + q^2 = 1 \tag{4-5-17}$$

The **S** matrix of a directional coupler is reduced to

$$\mathbf{S} = \begin{bmatrix} 0 & p & 0 & jq \\ p & 0 & jq & 0 \\ 0 & jq & 0 & p \\ jq & 0 & p & 0 \end{bmatrix} \tag{4-5-18}$$

Example 4-5-1:   Directional Coupler

A symmetric directional coupler with infinite directivity and a forward attenuation of 20 dB is used to monitor the power delivered to a load Z_ℓ (see Fig. 4-5-4). Bolometer 1 introduces a VSWR of 2.0 on arm 4; bolometer 2 is matched to arm 3. If bolometer 1
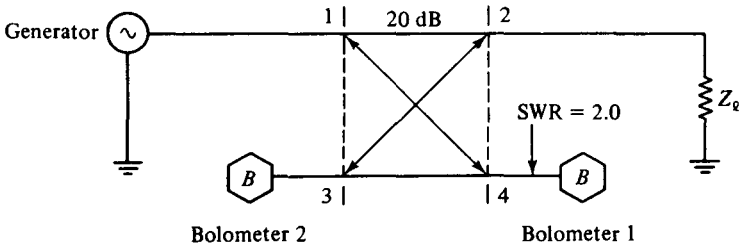
Figure 4-5-4   Power measurements by directional coupler.

reads 8 mW and bolometer 2 reads 2 mW, find: **(a)** the amount of power dissipated in the load Z_ℓ; **(b)** the VSWR on arm 2.

Solution   The wave propagation in the directional coupler is shown in Fig. 4-5-5.
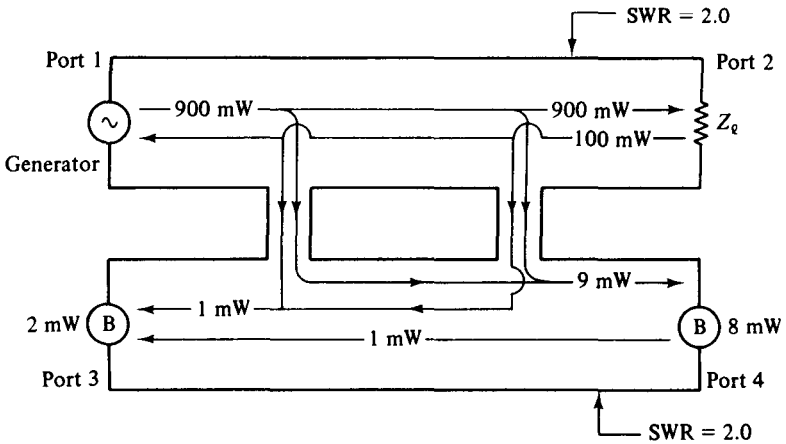
Figure 4-5-5   Wave propagation in the directional coupler.

a. Power dissipation at Z_ℓ.

1. The reflection coefficient at port 4 is

$$|\Gamma| = \frac{\rho - 1}{\rho + 1} = \frac{2 - 1}{2 + 1} = \frac{1}{3}$$

2. Since the incident power and reflected power are related by

$$P^- = P^+ |\Gamma|^2$$

where P^+ = incident power and P^- = reflected power, then

$$|\Gamma| = \frac{1}{3} = \sqrt{\frac{P^-}{P^+}} = \sqrt{\frac{P^-}{8 + P^-}}$$

The incident power to port 4 is $P_4^+ = 9$ mW, and the reflected power from port 4 is $P_4^- = 1$ mW.

3. Since port 3 is matched and the bolometer at port 3 reads 2 mW, then 1 mW must be radiated through the holes.

4. Since 20 dB is equivalent to a power ratio of 100:1, the power intput at port 1 is given by

$$P_1 = 100 P_4^+ = 900 \text{ mW}$$

and the power reflected from the load is

$$P_2^- = 100 \times (1 \text{ mW}) = 100 \text{ mW}$$

5. The power dissipated in the load is

$$P_\ell = P_2^+ - P_2^- = 900 - 100 = 800 \text{ mW}$$

b. The reflection coefficient is calculated as

$$|\Gamma| = \sqrt{\frac{P^-}{P^+}} = \sqrt{\frac{100}{900}} = \frac{1}{3}$$

Then the VSWR on arm 2 is

$$\rho = \frac{1 + |\Gamma|}{1 - |\Gamma|} = \frac{1 + \frac{1}{3}}{1 - \frac{1}{3}} = 2.0$$

4-5-3 Hybrid Couplers

Hybrid couplers are interdigitated microstrip couplers consisting of four parallel strip lines with alternate lines tied together. A single ground plane, a single dielectric, and a single layer of metallization are used. This type of coupler, called a Lange hybrid coupler [3], has four ports, as shown in Fig. 4-5-6.
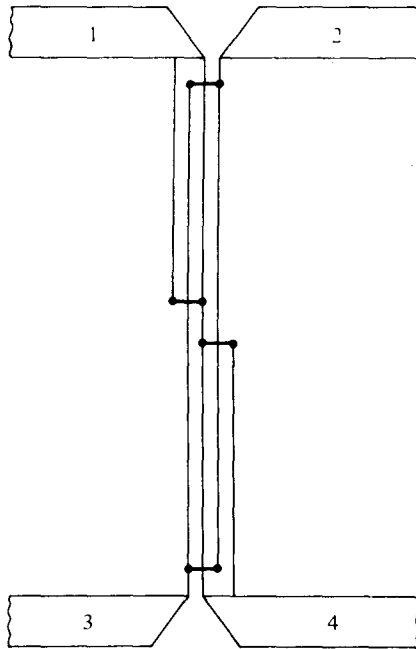
Figure 4-5-6   Lange hybrid coupler.

A signal wave incident in port 1 couples equal power into ports 2 and 4, but none into port 3. There are two basic types of Lange couplers: 180° hybrids and 90° (quadrature) hybrids. The latter are also called 3-dB directional couplers.

Hybrid couplers are frequently used as components in microwave systems or subsystems such as attenuators, balanced amplifiers, balanced mixers, modulators, discriminators, and phase shifters. The hybrid has a directivity of over 27 dB, a return loss of over 25 dB, an insertion loss of less than 0.13 dB, and an imbalance of less than 0.25 dB over a 40% bandwidth.

In modern microwave circuit design, Lange hybrid couplers are commonly

used in balanced amplifier circuitry for high-power and broad-bandwidth applications, as shown in Fig. 4-5-7.

Single-stage or cascaded double-stage GaAs MESFET chips are connected in parallel to two 3-dB and 90-degree Lange hybrid couplers. Their basic relationship can be expressed by the following three equations:

$$\mathbf{S}_{11} = \tfrac{1}{2}(\mathbf{S}_{11a} - \mathbf{S}_{11b}) \qquad (4\text{-}5\text{-}19)$$

$$\mathbf{S}_{22} = \tfrac{1}{2}(\mathbf{S}_{22a} - \mathbf{S}_{22b}) \qquad (4\text{-}5\text{-}20)$$

and

$$\text{Gain} = |\mathbf{S}_{21}|^2 = \tfrac{1}{4}|\mathbf{S}_{21a} + \mathbf{S}_{21b}|^2 \qquad (4\text{-}5\text{-}21)$$

where a and b indicate the two GaAs MESFET chips, and 1 and 2 refer to the input and output ports, respectively. The VSWRs of the balanced amplifier can be expressed as

$$\text{VSWR} = \frac{1 + |\mathbf{S}_{11}|}{1 - |\mathbf{S}_{11}|} \qquad \text{for the input port} \qquad (4\text{-}5\text{-}22)$$

and

$$\text{VSWR} = \frac{1 + |\mathbf{S}_{22}|}{1 - |\mathbf{S}_{22}|} \qquad \text{for the output port} \qquad (4\text{-}5\text{-}23)$$
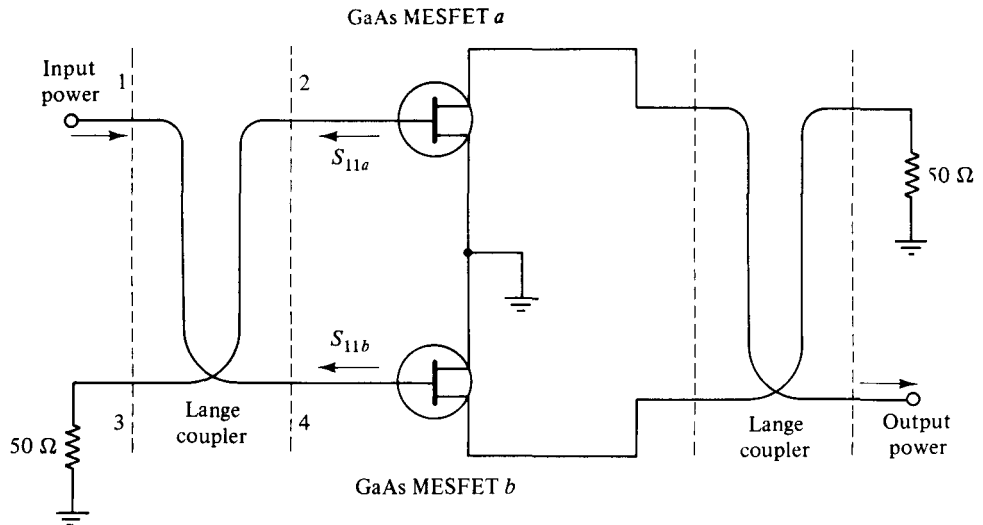
Figure 4-5-7   Balanced amplifier with Lange couplers.

Theoretically, if the two GaAs MESFET chips (or four chips in a double-stage amplifier circuit) are identical, the amplifier is balanced and its VSWR will be unity. Practically, however, characteristics of the two GaAs MESFET chips are not actually measured and they may not be the same. When their characteristics are different, the amplifier will not be balanced and manual tuning will be needed to balance it. Therefore, for mass production it is necessary to characterize the GaAs MESFET

chips in advance before placing them in the microwave integrated circuit in order to minimize the tuning work, reduce the production cost, and increase the hybrid reproducibility.

Example 4-5-2: Operation of a Balanced Amplifier

A GaAs MESFET balanced amplifier with two Lange couplers has the following parameters:

S parameters:	$S_{11a} = S_{11b}$
	$S_{22a} = S_{22b}$
Input signal power:	$P_{in} = 200$ mW
Power gain of each GaAs chip:	Gain $= 10$ dB

Determine: **(a)** the input and output VSWRs; **(b)** the output power in watts; **(c)** the linear output power gain in dB.

Solution

a. From Eqs. (4-5-22) and (4-5-23), the input and output VSWRs are unity.
b. The output power is

$$P_{out} = 200 \times 10 \times 2 = 4000 \text{ mW} = 4 \text{ W}$$

c. Because two GaAs chips are in parallel, the linear output power gain is

$$\text{Gain} = 10 \log (2) = 3 \text{ dB}$$

4-6 CIRCULATORS AND ISOLATORS

Both microwave circulators and microwave isolators are nonreciprocal transmission devices that use the property of Faraday rotation in the ferrite material. In order to understand the operating principles of circulators and isolators, let us describe the behavior of ferrites in the nonreciprocal phase shifter.

A *nonreciprocal phase shifter* consists of a thin slab of ferrite placed in a rectangular waveguide at a point where the dc magnetic field of the incident wave mode is circularly polarized. Ferrite is a family of $MeO \cdot Fe_2O_3$, where Me is a divalent iron metal. When a piece of ferrite is affected by a dc magnetic field, the ferrite exhibits Faraday rotation. It does so because the ferrite is nonlinear material and its permeability is an asymmetric tensor [4], as expressed by

$$\mathbf{B} = \hat{\mu}\mathbf{H} \tag{4-6-1}$$

where

$$\hat{\mu} = \mu_0(1 + \hat{\chi}_m) \tag{4-6-2}$$

$$\hat{\chi}_m = \begin{bmatrix} \chi_m & j\kappa & 0 \\ j\kappa & \chi_m & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4-6-3}$$

which is the tensor magnetic susceptibility. Here χ is the diagonal susceptibility and κ is the off-diagonal susceptibility.

When a dc magnetic field is applied to a ferrite, the unpaired electrons in the ferrite material tend to line up with the dc field because of their magnetic dipole moment. However, the nonreciprocal precession of unpaired electrons in the ferrite causes their relative permeabilities (μ_r^+, μ_r^-) to be unequal and the wave in the ferrite is then circularly polarized. The propagation constant for a linearly polarized wave inside the ferrite can be expressed as [4]

$$\gamma^\pm = j\omega \sqrt{\epsilon\mu_0(\mu + \kappa)} \qquad (4\text{-}6\text{-}4)$$

where

$$\mu = 1 + \hat{\chi}_m \qquad (4\text{-}6\text{-}5)$$

$$\mu_r^+ = \mu + \kappa \qquad (4\text{-}6\text{-}6)$$

$$\mu_r^- = \mu - \kappa \qquad (4\text{-}6\text{-}7)$$

The relative permeability μ_r changes with the applied dc magnetic field as given by

$$\mu_r^\pm = 1 + \frac{\gamma_e M_e}{|\gamma_e| H_{dc} \mp \omega} \qquad (4\text{-}6\text{-}8)$$

where γ_e = gyromagnetic ratio of an electron
M_e = saturation magnetization
ω = angular frequency of a microwave field
H_{dc} = dc magnetic field
μ_r^+ = relative permeability in the clockwise direction (right or positive circular polarization)
μ_r^- = relative permeability in the counterclockwise direction (left or negative circular polarization)

It can be seen from Eq. (4-6-8) that if $\omega = |\gamma_e| H_{dc}$, then μ_r^+ is infinite. This phenomenon is called the *gyromagnetic resonance* of the ferrite. A graph of μ_r is plotted as a function of H_{dc} for longitudinal propagation in Fig. 4-6-1.

If μ_r^+ is much larger than μ_r^- ($\mu_r^+ \gg \mu_r^-$), the wave in the ferrite is rotated in
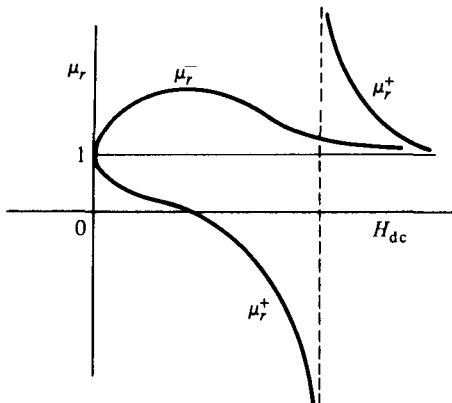
Figure 4-6-1 Curves of μ_r versus H_{dc} for axial propagation.

the clockwise direction. Consequently, the propagation phase constant β^+ for the forward direction differs from the propagation phase constant β^- for the backward direction. By choosing the length of the ferrite slab and the dc magnetic field so that

$$\omega = (\beta^+ - \beta^-)\ell = \frac{\pi}{2} \qquad (4\text{-}6\text{-}9)$$

a differential phase shift of 90° for the two directions of propagation can be obtained.

4-6-1 Microwave Circulators

A *microwave circulator* is a multiport waveguide junction in which the wave can flow only from the nth port to the $(n + 1)$th port in one direction (see Fig. 4-6-2). Although there is no restriction on the number of ports, the four-port microwave circulator is the most common. One type of four-port microwave circulator is a combination of two 3-dB side-hole directional couplers and a rectangular waveguide with two nonreciprocal phase shifters as shown in Fig. 4-6-3.
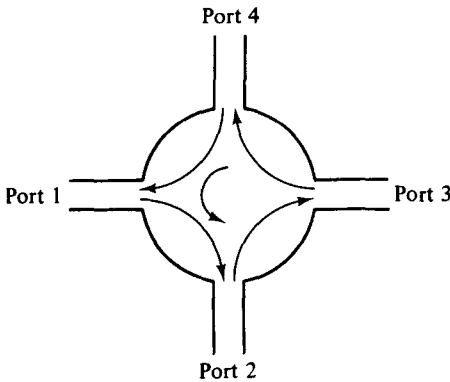
Figure 4-6-2 The symbol of a circulator.
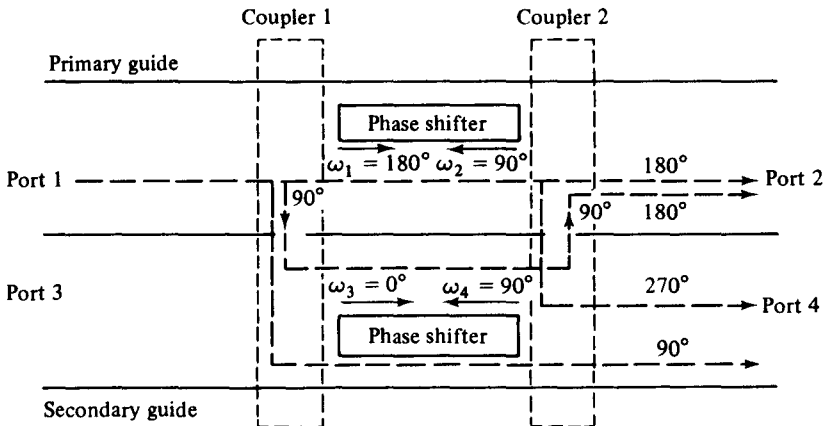
Figure 4-6-3 Schematic diagram of four-port circulator.

The operating principle of a typical microwave circulator can be analyzed with the aid of Fig. 4-6-3. Each of the two 3-dB couplers in the circulator introduces a phase shift of 90°, and each of the two phase shifters produces a certain amount of phase change in a certain direction as indicated. When a wave is incident to port 1, the wave is split into two components by coupler 1. The wave in the primary guide arrives at port 2 with a relative phase change of 180°. The second wave propagates through the two couplers and the secondary guide and arrives at port 2 with a relative phase shift of 180°. Since the two waves reaching port 2 are in phase, the power transmission is obtained from port 1 to port 2. However, the wave propagates through the primary guide, phase shifter, and coupler 2 and arrives at port 4 with a phase change of 270°. The wave travels through coupler 1 and the secondary guide, and it arrives at port 4 with a phase shift of 90°. Since the two waves reaching port 4 are out of phase by 180°, the power transmission from port 1 to port 4 is zero. In general, the differential propagation constants in the two directions of propagation in a waveguide containing ferrite phase shifters should be

$$\omega_1 - \omega_3 = (2m + 1)\pi \qquad \text{rad/s} \qquad (4\text{-}6\text{-}10)$$

$$\omega_2 - \omega_4 = 2n\pi \qquad \text{rad/s} \qquad (4\text{-}6\text{-}11)$$

where m and n are any integers, including zeros. A similar analysis shows that a wave incident to port 2 emerges at port 3 and so on. As a result, the sequence of power flow is designated as $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$.

Many types of microwave circulators are in use today. However, their principles of operation remain the same. Figure 4-6-4 shows a four-port circulator constructed of two magic tees and a phase shifter. The phase shifter produces a phase shift of 180°. The explanation of how this circulator works is left as an exercise for the reader.
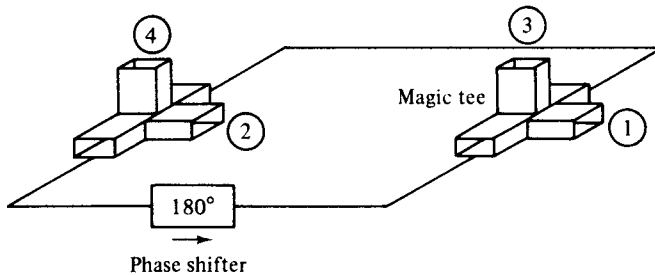
Figure 4-6-4   A four-port circulator.

A perfectly matched, lossless, and nonreciprocal four-port circulator has an **S** matrix of the form

$$\mathbf{S} = \begin{bmatrix} 0 & S_{12} & S_{13} & S_{14} \\ S_{21} & 0 & S_{23} & S_{24} \\ S_{31} & S_{32} & 0 & S_{34} \\ S_{41} & S_{42} & S_{43} & 0 \end{bmatrix} \qquad (4\text{-}6\text{-}12)$$

Using the properties of S parameters as described previously, the **S** matrix in Eq.

(4-6-12) can be simplified to

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{4-6-13}$$

4-6-2 Microwave Isolators

An *isolator* is a nonreciprocal transmission device that is used to isolate one component from reflections of other components in the transmission line. An ideal isolator completely absorbs the power for propagation in one direction and provides lossless transmission in the opposite direction. Thus the isolator is usually called *uniline*. Isolators are generally used to improve the frequency stability of microwave generators, such as klystrons and magnetrons, in which the reflection from the load affects the generating frequency. In such cases, the isolator placed between the generator and load prevents the reflected power from the unmatched load from returning to the generator. As a result, the isolator maintains the frequency stability of the generator.

Isolators can be constructed in many ways. They can be made by terminating ports 3 and 4 of a four-port circulator with matched loads. On the other hand, isolators can be made by inserting a ferrite rod along the axis of a rectangular waveguide as shown in Fig. 4-6-5. The isolator here is a Faraday-rotation isolator. Its operating principle can be explained as follows [5]. The input resistive card is in the y-z plane, and the output resistive card is displaced 45° with respect to the input card. The dc magnetic field, which is applied longitudinally to the ferrite rod, rotates the wave plane of polarization by 45°. The degrees of rotation depend on the length and diameter of the rod and on the applied dc magnetic field. An input TE_{10} dominant mode is incident to the left end of the isolator. Since the TE_{10} mode wave is perpendicular to the input resistive card, the wave passes through the ferrite rod without attenuation. The wave in the ferrite rod section is rotated clockwise by 45° and is normal to the output resistive card. As a result of rotation, the wave arrives at the output
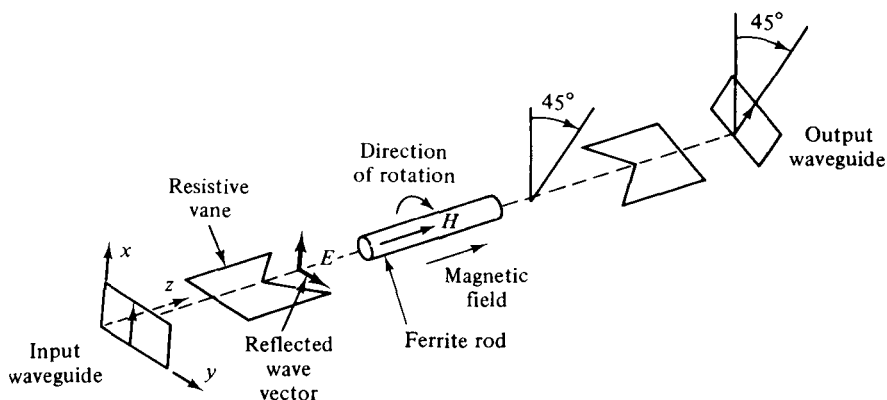
Figure 4-6-5   Faraday-rotation isolator.

end without attenuation at all. On the contrary, a reflected wave from the output end is similarly rotated clockwise 45° by the ferrite rod. However, since the reflected wave is parallel to the input resistive card, the wave is thereby absorbed by the input card. The typical performance of these isolators is about 1-dB insertion loss in forward transmission and about 20- to 30-dB isolation in reverse attenuation.

REFERENCES

[1] SAAD, T., and R. C. HANSEN, *Microwave Engineer's Handbook*, Vol. 1. Artech House, Dedham, Mass., 1971.

[2] SOUTHWORTH, G. E., *Principles and Applications of Waveguide Transmission*. Chapters 8 and 9. D. Van Nostrand Company, Princeton, N.J., 1950.

[3] LANGE, J. Interdigited stripline quadrature hybrid. *IEEE Trans. on Microwave Theory and Techniques*, **MTT-17**, No. 12, 1150–1151, December 1969.

[4] SOOHOO, R. F., *Theory and Applications of Ferrites*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1960.

[5] BOWNESS, C., Microwave ferrites and their applications. *Microwave J.*, **1,** 13–21, July-August 1958.

SUGGESTED READINGS

LIAO, S. Y., *Engineering Applications of Electromagnetic Theory*. West Publishing Company, St. Paul, Minn., 1988.

RIZZI, P. A., *Microwave Engineering: Passive Circuits*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.

PROBLEMS

Rectangular waveguides

4-1. An air-filled rectangular waveguide has dimensions of $a = 6$ cm and $b = 4$ cm. The signal frequency is 3 GHz. Compute the following for the TE_{10}, TE_{01}, TE_{11}, and TM_{11} modes:
  a. Cutoff frequency
  b. Wavelength in the waveguide
  c. Phase constant and phase velocity in the waveguide
  d. Group velocity and wave impedance in the waveguide

4-2. Show that the TM_{01} and TM_{10} modes in a rectangular waveguide do not exist.

4-3. The dominant mode TE_{10} is propagated in a rectangular waveguide of dimensions $a = 6$ cm and $b = 4$ cm. The distance between a maximum and a minimum is 4.47 cm. Determine the signal frequency of the dominant mode.

4-4. A TE_{11} mode of 10 GHz is propagated in an air-filled rectangular waveguide. The mag-

netic field in the z direction is given by

$$\mathbf{H}_z = \mathbf{H}_0 \cos\left(\frac{\pi x}{\sqrt{6}}\right) \cos\left(\frac{\pi y}{\sqrt{6}}\right) \quad \text{A/m}$$

The phase constant is $\beta = 1.0475$ rad/cm, the quantities x and y are expressed in centimeters, and $a = b = \sqrt{6}$ are also in centimeters. Determine the cutoff frequency f_c, phase velocity v_g, guided wavelength λ_g, and the magnetic field intensity in the y direction.

4-5. A rectangular waveguide is designed to propagate the dominant mode TE_{10} at a frequency of 5 GHz. The cutoff frequency is 0.8 of the signal frequency. The ratio of the guide height to width is 2. The time-average power flowing through the guide is 1 kW. Determine the magnitudes of electric and magnetic intensities in the guide and indicate where these occur in the guide.

4-6. An air-filled rectangular waveguide has dimensions of $a = 6$ cm and $b = 4$ cm. The guide transports energy in the dominant mode TE_{10} at a rate of 1 horsepower (746 J). If the frequency is 20 GHz, what is the peak value of electric field occurring in the guide?

4-7. An impedance of $(0.5 - j0.4)Z_0$ is connected to a rectangular waveguide. A capacitive window with a susceptance $jB = j0.4Y_0$ is located at a distance of 0.2λ from the load.
   a. Determine the VSWR on the line in the absence of the window.
   b. Find the VSWR on the line in the presence of the window.

4-8. An air-filled rectangular waveguide with dimensions of 3 cm × 1 cm operates in the TE_{10} mode at 10 GHz. The waveguide is perfectly matched and the maximum E field existing everywhere in the guide is 10^3 V/m. Determine the voltage, current, and wave impedance in the waveguide.

4-9. The dominant mode TE_{10} is propagated in a rectangular waveguide of dimensions $a = 2.25$ cm and $b = 1$ cm. Assume an air dielectric with a breakdown gradient of 30 kV/cm and a frequency of 10 GHz. There are no standing waves in the guide. Determine the maximum average power that can be carried by the guide.

4-10. A rectangular waveguide is terminated in an unknown impedance at $z = 25$ cm. A dominant mode TE_{10} is propagated in the guide, and its VSWR is measured as 2.8 at a frequency of 8 GHz. The adjacent voltage minima are located at $z = 9.46$ cm and $z = 12.73$ cm.
   a. Determine the value of the load impedance in terms of Z_0.
   b. Find the position closest to the load where an inductive window is placed in order to obtain a VSWR of unity.
   c. Determine the value of the window admittance.

4-11. A rectangular waveguide is filled by dielectric material of $\epsilon_r = 9$ and has inside dimensions of 7 × 3.5 cm. It operates in the dominant TE_{10} mode.
   a. Determine the cutoff frequency.
   b. Find the phase velocity in the guide at a frequency of 2 GHz.
   c. Find the guided wavelength λ_g at the same frequency.

4-12. The electric field intensity of the dominant TE_{10} mode in a lossless rectangular waveguide is

$$E_y = E_0 \sin\left(\frac{\pi x}{a}\right)e^{-j\beta_g z} \quad \text{for } f > f_c$$

   a. Find the magnetic field intensity $\mathbf{H}$.
   b. Compute the cutoff frequency and the time-average transmitted power.

Circular waveguides

4-13. An air-filled circular waveguide is to be operated at a frequency of 6 GHz and is to have dimensions such that $f_c = 0.8f$ for the dominant mode. Determine:
   a. The diameter of the guide
   b. The wavelength λ_g and the phase velocity v_g in the guide

4-14. An air-filled circular waveguide of 2 cm inside radius is operated in the TE_{01} mode.
   a. Compute the cutoff frequency.
   b. If the guide is to be filled with a dielectric material of $\epsilon_r = 2.25$, to what value must its radius be changed in order to maintain the cutoff frequency at its original value?

4-15. An air-filled circular waveguide has a radius of 1.5 cm and is to carry energy at a frequency of 10 GHz. Find all TE and TM modes for which transmission is possible.

4-16. A TE_{11} wave is propagating through a circular waveguide. The diameter of the guide is 10 cm, and the guide is air-filled.
   a. Find the cutoff frequency.
   b. Find the wavelength λ_g in the guide for a frequency of 3 GHz.
   c. Determine the wave impedance in the guide.

4-17. An air-filled circular waveguide has a diameter of 4 cm and is to carry energy at a frequency of 10 GHz. Determine all TE_{np} modes for which transmission is possible.

4-18. A circular waveguide has a cutoff frequency of 9 GHz in dominant mode.
   a. Find the inside diameter of the guide if it is air-filled.
   b. Determine the inside diameter of the guide if the guide is dielectric-filled. The relative dielectric constant is $\epsilon_r = 4$.

Microwave cavities

4-19. A coaxial resonator is constructed of a section of coaxial line and is open-circuited at both ends. The resonator is 5 cm long and filled with dielectric of $\epsilon_r = 9$. The inner conductor has a radius of 1 cm and the outer conductor has a radius of 2.5 cm.
   a. Find the resonant frequency of the resonator.
   b. Determine the resonant frequency of the same resonator with one end open and one end shorted.

4-20. An air-filled circular waveguide has a radius of 3 cm and is used as a resonator for TE_{01} mode at 10 GHz by placing two perfectly conducting plates at its two ends. Determine the minimum distance between the two end plates.

4-21. A four-port circulator is constructed of two magic tees and one phase shifter as shown in Fig. 4-6-4. The phase shifter produces a phase shift of 180°. Explain how this circulator works.

4-22. A coaxial resonator is constructed of a section of coaxial line 6 cm long and is short-circuited at both ends. The circular cavity has an inner radius of 1.5 cm and an outer radius of 3.5 cm. The line is dielectric-filled with $\epsilon_r = 2.25$.
   a. Determine the resonant frequency of the cavity for TEM_{001}.
   b. Calculate the quality Q of the cavity.

4-23. A rectangular-cavity resonator has dimensions of $a = 5$ cm, $b = 2$ cm, and $d = 15$ cm. Compute:
   a. The resonant frequency of the dominant mode for an air-filled cavity
   b. The resonant frequency of the dominant mode for a dielectric-filled cavity of $\epsilon_r = 2.56$

4-24. An undercoupled resonant cavity is connected to a lossless transmission line as shown in Fig. P4-24. The directional coupler is assumed to be ideal and matched on all arms. The unloaded Q of the cavity is 1000 and the VSWR at resonance is 2.5.
  a. Calculate the loaded Q_ℓ of the cavity.
  b. Find the reading of bolometer 2 if bolometer 1 reads 4 mW.
  c. Compute the power dissipated in the cavity.

P4-24
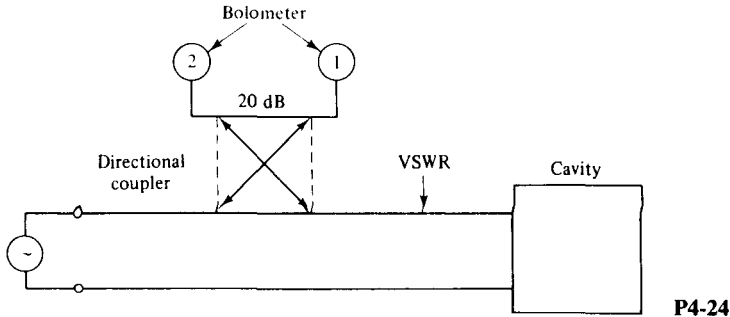
Hybrid circuits

4-25. A microwave transmission system consists of a generator, an overcoupled cavity, two ideal but not identical dual directional couplers with matched bolometers, and a load Z_ℓ. The lossless transmission line has a characteristic impedance Z_0. The readings of the four bolometers (1, 2, 3, and 4) are 2 mW, 4 mW, 0 and 1 mW, respectively. The system is shown in Fig. P4-25.
  a. Find the load impedance Z_ℓ in terms of Z_0.
  b. Calculate the power dissipated by Z_ℓ.
  c. Compute the power dissipated in the cavity.
  d. Determine the VSWR on the input transmission line.
  e. Find the ratio of Q_ℓ/Q_0 for the cavity.

P4-25
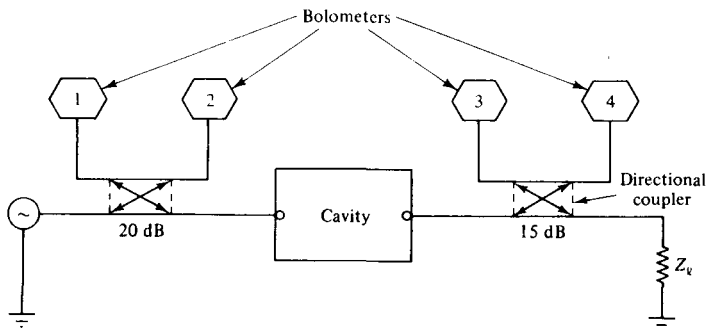
4-26. A symmetric directional coupler has an infinite directivity and a forward attenuation of 20 dB. The coupler is used to monitor the power delivered to a load Z_ℓ as shown in Fig. P4-26. Bolometer 1 introduces a VSWR of 2.0 on arm 1; bolometer 2 is matched to arm 2. If bolometer 1 reads 9 mW and bolometer 2 reads 3 mW:
  a. Find the amount of power dissipated in the load Z_ℓ.
  b. Determine the VSWR on arm 3.

Bolometer 2 ... VSWR = 2.0 ... Bolometer 1 ... 20 dB ... Z_ϱ

P4-26
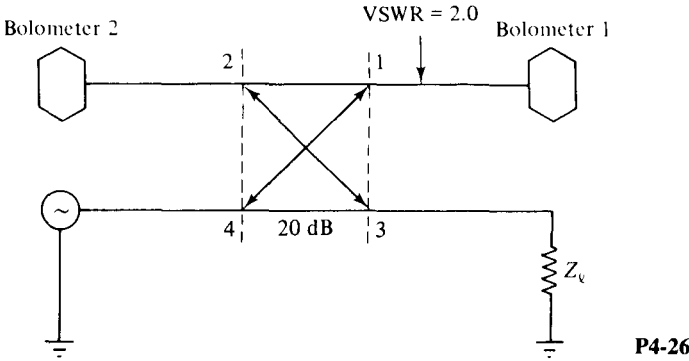
4-27. A semicircular-cavity resonator has a length of 5 cm and a radius of 2.5 cm.
   a. Calculate the resonant frequency for the dominant mode if the cavity is air-filled.
   b. Repeat part (a) if the cavity is loaded by a dielectric with a relative constant of 9.

4-28. The impedance matrix of a certain lumped-element network is given by

$$[z_{ij}] = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

Determine the scattering matrix by using S-parameter theory and indicate the values of the components:

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

4-29. A hybrid waveguide is constructed of two identical rectangular waveguides across each other at the center and works as a four-port device. Write a general scattering matrix and then simplify it as much as possible by inspection of geometric symmetry and by use of the known phases of the electric waves.

4-30. A helical slow-wave structure has a pitch P of 2 mm and a diameter of 4 cm. Calculate the wave velocity in the axial direction of the helix.

4-31. Two 3-dB quadrature Lange couplers are used in a GaAs MESFET balanced amplifier circuit with the following parameters:

MESFET a: Reflection coefficients $S_{11a} = 0.7488\underline{/-158.3°}$

$S_{22a} = 0.8521\underline{/-155.7°}$

Forward transmission $S_{21a} = 1.3500\underline{/-8.5°}$
coefficient

MESFET b: Reflection coefficients $S_{11b} = 0.6210\underline{/-175.9°}$

$S_{22b} = 0.7727\underline{/-151.4°}$

Forward transmission $S_{21b} = 1.2200\underline{/-19.1°}$
coefficient

Compute: **a.** The input and output VSWRs
   b. The power gain in dB for the balanced amplifier
   c. The power loss in dB if one MESFET fails
   d. The linear output power gain in dB

Microwave Transistors and Tunnel Diodes

5-0 INTRODUCTION

Microwave solid-state devices are becoming increasingly important at microwave frequencies. These devices can be broken down into four groups. In the first group are the microwave bipolar junction transistor (BJT), the heterojunction bipolar transistor (HBT), and the tunnel diodes. This group is discussed in this chapter. The second group includes microwave field-effect transistors (FETs) such as the junction field-effect transistors (JFETs), metal-semiconductor field-effect transistors (MESFETs), high electron mobility transistors (HEMTs), metal-oxide-semiconductor field-effect transistors (MOSFETs), the metal-oxide-semiconductor transistors and memory devices, and the charge-coupled devices (CCDs). This group is described in Chapter 6. The third group, which is characterized by the bulk effect of the semiconductor, is called the transferred electron device (TED). These devices include the Gunn diode, limited space-charge-accumulation diode (LSA diode), indium phosphide diode (InP diode), and cadmium telluride diode (CdTe diode). This group is analyzed in Chapter 7. The devices of the fourth group, which are operated by the avalanche effect of the semiconductor, are referred to as avalanche diodes: the impact ionization avalanche transit-time diodes (IMPATT diodes), the trapped plasma avalanche triggered transit-time diodes (TRAPATT diodes), and the barrier injected transit-time diodes (BARITT diodes). The avalanche diodes are studied in Chapter 8. All those microwave solid-state devices are tabulated in Table 5-0-1.

In studying microwave solid-state devices, the electrical behavior of solids is the first item to be investigated. In this section it will be seen that the transport of charge through a semiconductor depends not only on the properties of the electron but also on the arrangement of atoms in the solids. Semiconductors are a group of substances having electrical conductivities that are intermediate between metals and insulators. Since the conductivity of the semiconductors can be varied over wide

TABLE 5-0-1   MICROWAVE SOLID-STATE DEVICES

	Microwave transistor	Microwave BJT
		HBT
		Tunnel diode
	Field-effect transistors	JFET
		MESFET, HEMT
		MOSFET
		NMOS, PMOS, CMOS
		Memories
		CCD
Microwave solid-state devices	Transferred electron devices	Gunn diode
		LSA diodes
		InP diodes
		CdTe diodes
	Avalanche transit-time devices	Read diode
		IMPATT diode
		TRAPATT diode
		BARITT diode
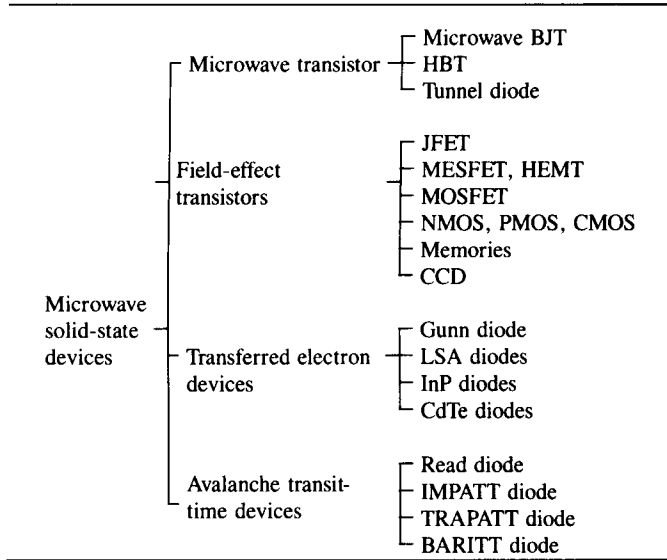
ranges by changes in their temperature, optical excitation, and impurity content, they are the natural choices for electronic devices. The properties of important semiconductors are tabulated in Table 5-0-2.

The energy bands of a semiconductor play a major role in their electrical behavior. For any semiconductor, there is a forbidden energy region in which no allowable states can exist. The energy band above the forbidden region is called the *conduction band,* and the bottom of the conduction band is designated by E_c. The energy band below the forbidden region is called the *valence band,* and the top of

TABLE 5-0-2   PROPERTIES OF IMPORTANT SEMICONDUCTORS

Semiconductor	Bandgap energy (eV)		Mobility at 300°K (cm²/V · s)		Relative dielectric constant
	0°K	300°K	Holes	Electrons	
C	5.51	5.47	1600	1800	5.5
Ge	0.89	0.803	1900	3900	16
Si	1.16	1.12	450	1600	11.8
AlSb	1.75	1.63	420	200	11
GaSb	0.80	0.67	1400	4000	15
GaAs	1.52	1.43	400	8500	13.1
GaP	2.40	2.26	75	110	10
InSb	0.26	1.80	750	78,000	17
InAs	0.46	0.33	460	33,000	14.5
InP	1.34	1.29	150	4600	14
CdS	2.56	2.42	50	300	10
CdSe	1.85	1.70		800	10
ZnO		3.20		200	9
ZnS	3.70	3.60		165	8

the valence band is designated by E_v. The separation between the energy of the lowest conduction band and that of the highest valence band is called the *bandgap energy* E_g, which is the most important parameter in semiconductors.

Electron energy is conventionally defined as positive when measured upward whereas the hole energy is positive when measured downward. A simplified band diagram is shown in Fig. 5-0-1.
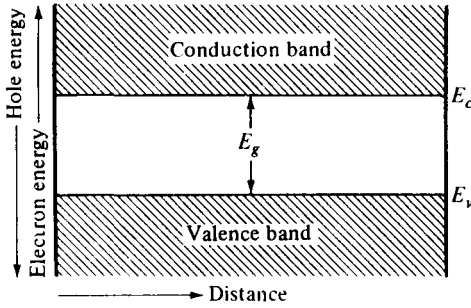
Figure 5-0-1   Energy-band diagram.

In the 1970s, it seemed that microwave transistors would be useful for generating power up to about 5 GHz. Since their inception, avalanche diodes have produced in excess of 4 W continuous wave (CW) at 5 GHz. Gunn diodes had been considered only for local oscillators or low-power transmitter applications, but recent results indicate that a single Gunn diode can generate an output power of 1 W at X band. At higher microwave frequencies, and even well into the millimeter range, limited space-charge-accumulation diodes (LSAS) can provide the highest peak power of any solid-state device, up to 250W in C band, 100 W in X band, and 50 W in Ku band. Since the pulsed Gunn and TRAPATT diodes are essentially transit-time devices, their operating frequency is approximately determined by the thickness of the active layer in the diode. An operating frequency of 10 GHz requires an active layer thickness on the order of 10 μm (microns). Thus only a limited voltage can be applied to such a thin layer because of breakdown limitations. Consequently, the
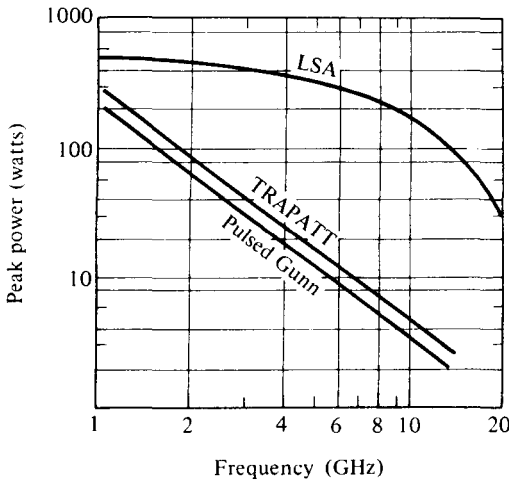
Figure 5-0-2   Peak-power levels achieved by microwave diode.

peak power capability of both the pulsed Gunn diodes and the TRAPATT diodes is greatly limited at higher frequencies. On the other hand, the peak power capability of an LSA diode is approximately proportional to the square of the thickness of the active layer because its operating frequency is independent of the thickness of the active layer. Thus the LSA diode is capable of producing higher peak power than either the pulsed Gunn diodes or the TRAPATT diodes. Figure 5-0-2 shows peak power versus frequency for these three devices.

Solid-state microwave power sources are widely used in radar, communications, navigational and industrial electronics, and medical and biological equipment. Representative applications for microwave solid-state devices are listed in Table 5-0-3.

TABLE 5-0-3 APPLICATIONS OF MICROWAVE SOLID-STATE DEVICES

Devices	Applications	Advantages
Transistor	L-band transmitters for telemetry systems and phased array radar systems L- and S-band transmitters for communications systems	Low cost, low power supply, reliable, high CW power output, light weight
TED	C-, X-, and Ku-band ECM amplifiers for wideband systems X- and Ku-band transmitters for radar systems, such as traffic control	Low power supply (12 V), low cost, light weight, reliable, low noise, high gain
IMPATT	Transmitters for millimeter-wave communications systems	Low power supply, low cost, reliable, high CW power output, light weight
TRAPATT	S-band pulsed transmitters for phased array radar systems	High peak and average power, reliable, low power supply, low cost
BARITT	Local oscillators in communications and radar receivers	Low cost, low power supply, reliable, low noise

5-1 MICROWAVE BIPOLAR TRANSISTORS

The invention of the transistor (contraction for transfer resistor) by William Shockley and his coworkers at Bell Laboratory in 1948 had a revolutionary impact on electronic technology in general and on solid-state devices in particular. Since then transistors and related semiconductor devices have replaced vacuum tubes for lower-power sources. Microwave power transistor technology has advanced significantly during the past three decades. The microwave transistor is a nonlinear device, and its principle of operation is similar to that of the low-frequency device, but requirements for dimensions, process control, heat sinking, and packaging are much more severe.

For microwave applications, the silicon (Si) bipolar transistors dominate for frequency range from UHF to about S band (about 3 GHz). As the technology improves, the upper frequency limit for these devices is continuously being extended, and at the present time the devices are capable of producing useful power up to 22 GHz. The majority of bipolar transistors of current interest are fabricated from silicon, although GaAs devices offer prospects for improvements in operating fre-

quency, in high temperatures, and in radiation hardness. The Si bipolar transistor is inexpensive, durable, integrative, and offers gain much higher than available with competing field-effect devices. It has moderate noise figure in RF amplifiers and $1/f$ noise characteristics that are about $10-20$ dB superior to GaAs MESFETs. For these reasons, the Si bipolar transistors dominate in amplifier applications for the lower microwave frequencies and are often the devices of choice for local oscillators.

5-1-1 Physical Structures

All microwave transistors are now planar in form and almost all are of the silicon *n-p-n* type. The geometry can be characterized as follows: (a) interdigitated, (b) overlay, and (c) matrix (also called mesh or emitter grid) as shown in Fig. 5-1-1. The interdigitated type is for a small signal and power, but the overlay type and matrix type are for small power only. The figure of merit for the three surface geometries shown in Fig. 5-1-1 is listed in Table 5-1-1 [1].
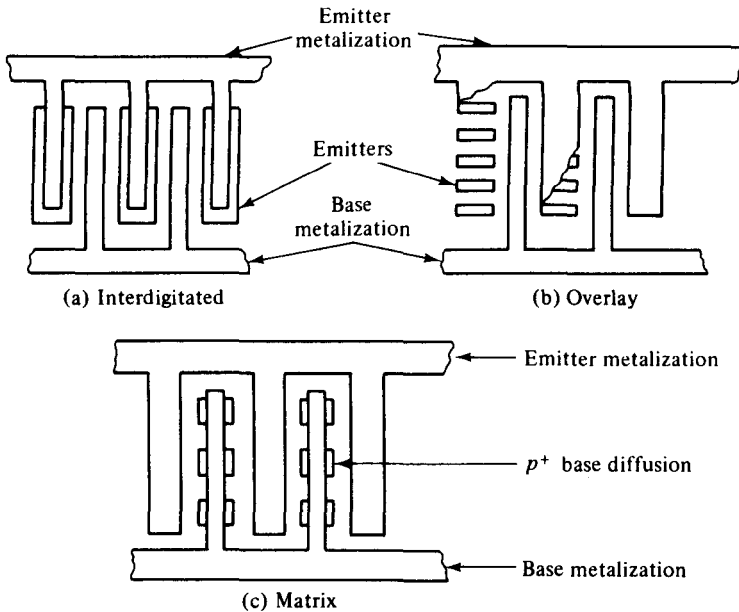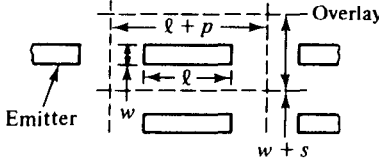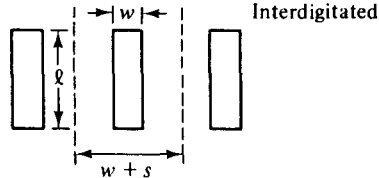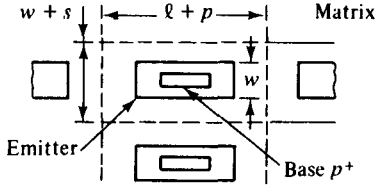
Figure 5-1-1 Surface geometries of microwave power transistor. (*From H. Sobol and F. Sterzer [1]: reprinted by permission of IEEE, Inc.*)

For high-frequency applications, the *n-p-n* structure is preferred because the electron mobility ($\mu_n = 1500$ cm^2/V · s) is much higher than the hole mobility ($\mu_p = 450$ cm^2/V · s). Figure 5-1-2 shows an example of the densities for an *n-p-n* transistor. The density unit is in cm^2/V · s.

Although there are many ways of fabricating a transistor, diffusion and ion implantation are generally used. For example, the structure would typically start with a lightly doped *n*-type epitaxial layer as the collector. The base region would be formed by counter-doping the base region *p*-type by diffusion. The emitter would

TABLE 5-1-1   FIGURE OF MERIT (M) OF VARIOUS SURFACE GEOMETRIES

Surface geometry and unit cell	$M = \dfrac{EP}{BA}$

$$\frac{2(\ell + w)}{(w + s)(\ell + p)}$$

$$\frac{2(\ell + w)}{\ell(w + s)}$$

$$\frac{2(\ell + w)}{(w + s)(\ell + p)}$$

Source: From H. Sobol and F. Sterzer [1]; reprinted by permission of IEEE, Inc.

be formed by a shallow heavily doped n-type diffusion or by ion implantation. The emitter and base contacts are generally located on the semiconductor surface in an interdigital, planar arrangement. The interdigital geometry always provides for $n + 1$ base fingers, where n is the number of emitter fingers. The number of fingers varies with the application, with more fingers required as the output power capabil-
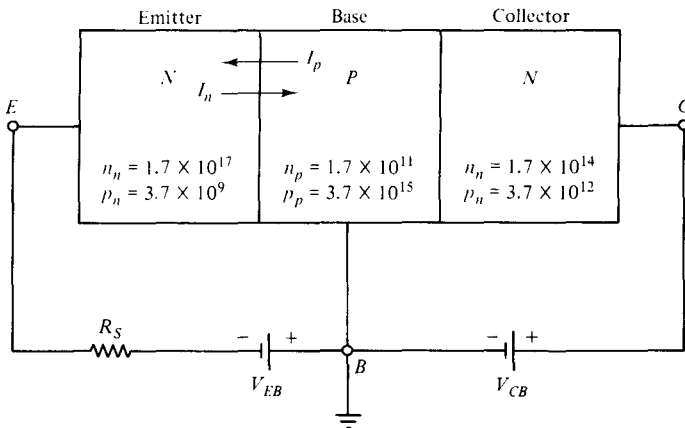
Figure 5-1-2   Carrier densities of an n-p-n transistor.

ity of the transistor increases. Additional fingers, however, increase the device parasitics and degrade the noise and upper frequency capability of the devices.

Figure 5-1-3 shows two schematic diagrams for a bipolar junction transistor (BJT): (a) the cross section of a discrete *n-p-n* planar BJT and (b) the cross section of a chip-type *n-p-n* integrated BJT.

The *p-n-p* bipolar junction transistor is a complementary structure of the *n-p-n* BJT by interchanging *p* for *n* and *n* for *p*. The *p-n-p* BJT is basically fabricated by first forming an *n*-type layer in the *p*-type substrate; then a *p⁺*-type region is devel-
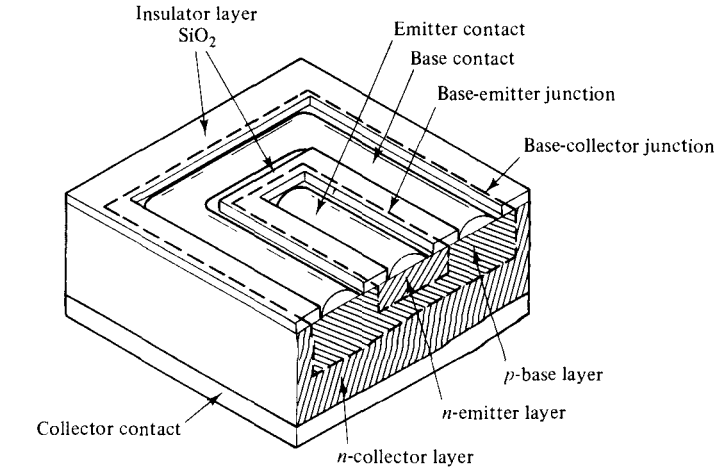
(a) Discrete *n-p-n* planar BJT
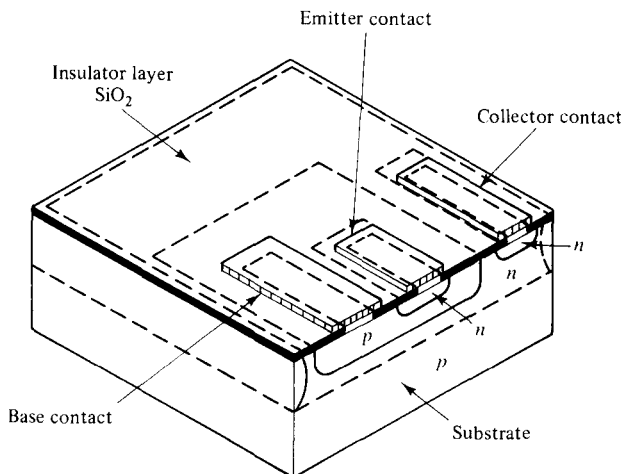
(b) Integrated chip-type *n-p-n* BJT

Figure 5-1-3   Schematic diagrams of bipolar junction transistors. (*After D. Navon* *[2].*)

oped in the n layer. Finally, metallic contacts are introduced to the p^+ region and p layer through the windows opened in the oxide layer and to the p region at the bottom.

5-1-2 Bipolar Transistor Configurations

In general, there are two types of bipolar transistors: *p-n-p* and *n-p-n*. In practical applications, a transistor can be connected as three different configurations: common base (CB), common emitter (CE), and common collector (CC), depending on the polarities of the bias voltages connected to its terminals.

Common-base configuration. The *common-base* (CB) *configuration* refers to the one where the emitter (input circuit) and collector (output circuit) terminals are common to the base as shown in Fig. 5-1-4.
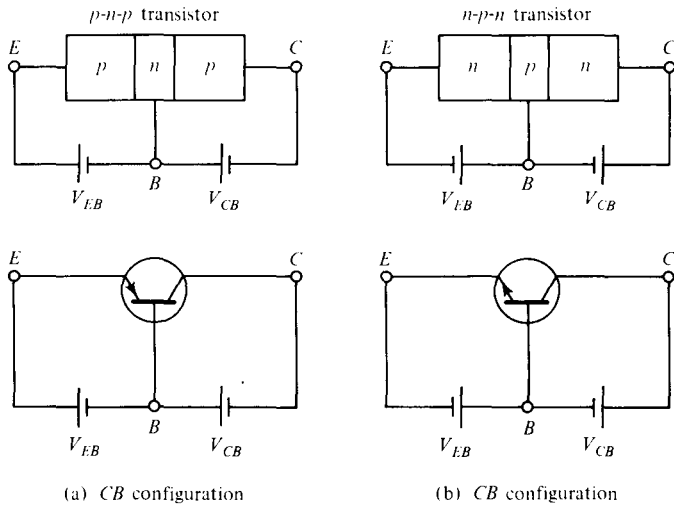
Figure 5-1-4   Common-base configurations for *p-n-p* and *n-p-n* transistors.

The CB configuration is also called the *grounded-base configuration*. For a *p-n-p* transistor, the largest current components are caused by holes. Holes flow from the emitter to the collector and down toward ground out of the base terminal. In an *n-p-n* transistor all current and voltage polarities are negative to those in a *p-n-p* transistor. The CB configuration of a transistor is usually used in amplifier applications. Its input voltage V_{EB} and output current I_C can be expressed in terms of the output voltage V_{CB} and input current I_E as

$$V_{EB} = \text{some function } (V_{CB}, I_E) \qquad (5\text{-}1\text{-}1)$$

$$I_C = \text{some function } (V_{CB}, I_E) \qquad (5\text{-}1\text{-}2)$$

Common-emitter configuration. Most transistors have their emitter, rather than their base, as the terminal to both input and output networks. Such a

configuration is known as *common-emitter* (CE) or *grounded-emitter configuration* as shown in Fig. 5-1-5.

In the CE configuration, the input current I_B and the output voltage V_{CE} are independent variables, whereas the input voltage V_{BE} and output current I_C can be written as

$$V_{BE} = \text{some function } (V_{CE}, I_B) \tag{5-1-3}$$

$$I_C = \text{some function } (V_{CE}, I_B) \tag{5-1-4}$$

The CE configuration is commonly used as a switch or pulse transistor amplifier. This is because the transistor is open at the cutoff mode and is closed at the saturation mode.
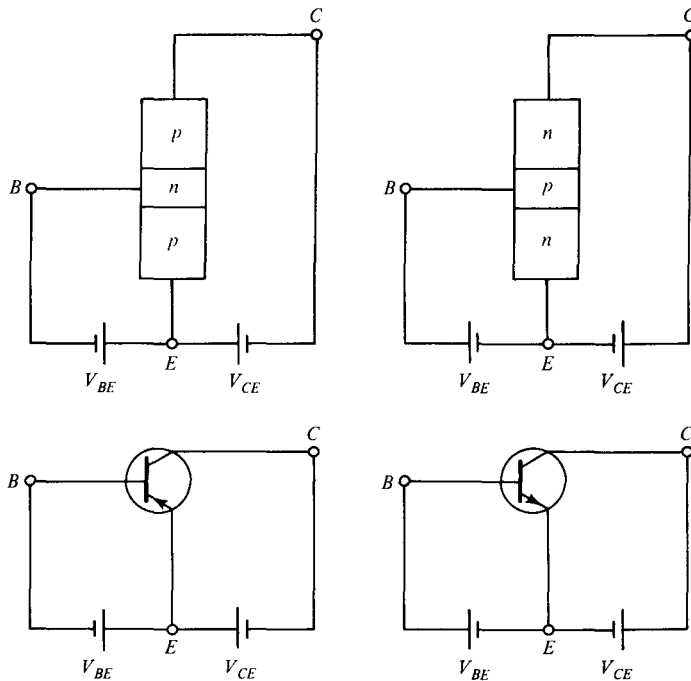
Figure 5-1-5   Common-emitter configurations for *p-n-p* and *n-p-n* transistors.

Common-collector configuration.   Another transistor-circuit configuration is called the *common-collector* or *grounded-collector* configuration as shown in Fig. 5-1-6.

In a common-collector (CC) configuration, the output voltage of the load is taken from the emitter terminal instead of the collector as in the common-base and common-emitter configurations. When the transistor is cut off, no current will flow in the emitter terminal at the load. When the transistor is operating in a saturation mode, the load current reaches toward its maximum. Therefore the CC configuration transistor can also be used as a switch or pulse amplifier. The significant differ-
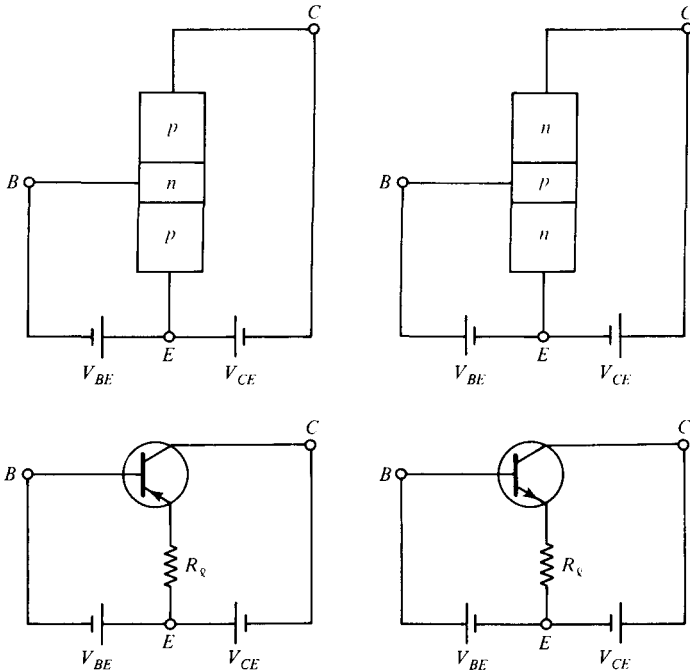
Figure 5-1-6   Common-collector configuration for *p-n-p* and *n-p-n* transistors.

ence, however, between common-emitter and common-collector configurations is that the common-collector amplifier has no voltage gain.

Hybrid-pi equivalent model.   The *hybrid-pi equivalent model* is commonly used in the normal active mode of the common-emitter configuration for small-signal operation. Fig. 5-1-7 shows the hybrid-pi equivalent model of the common-emitter configuration.

For a small-signal operation, the nonlinear or ac parameters of a hybrid-pi equivalent model can be expressed as

$$v_{ce} = h_{ie}i_b + h_{re}v_{ce} \tag{5-1-5}$$

$$i_c = h_{fe}i_b + h_{oe}v_{ce} \tag{5-1-6}$$

$$\text{where } h_{ie} = \left.\frac{\partial v_{be}}{\partial i_b}\right|_{v_{ce}=\text{constant}} \tag{5-1-7}$$

$$h_{re} = \left.\frac{\partial v_{be}}{\partial i_b}\right|_{i_b} \tag{5-1-8}$$

$$h_{fe} = \left.\frac{\partial i_c}{\partial i_b}\right|_{v_{ce}} \tag{5-1-9}$$

$$h_{oe} = \left.\frac{\partial i_c}{\partial v_{ce}}\right|_{i_b} \tag{5-1-10}$$
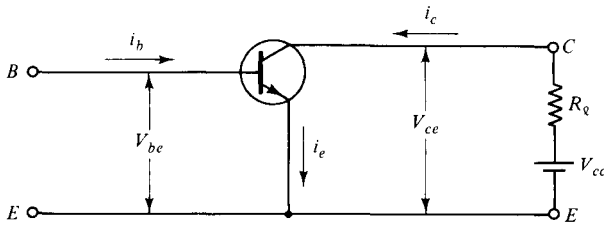
(a) Common-emitter n-p-n transistor
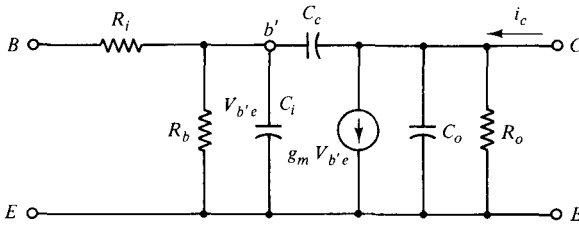
(b) Hybrid-pi model
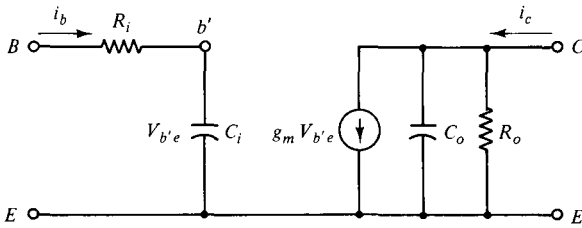
(c) Simplified hybrid-pi model at low frequency

Figure 5-1-7 Hybrid-pi equivalent model.

When the dimensions of a bipolar junction transistor become very small, their Z, Y, or H parameters cannot be measured because the input and output terminals cannot be openly and shortly realized. Therefore, the S parameters are commonly measured. In transistor design, it is necessary to convert the S parameters into Y parameters for the network component computations.

An incremental change of the emitter voltage $\Delta \mathcal{V}_{b'e}$ at the input terminal will induce an incremental change of the collector current Δi_c at the output terminal. Then the mutual conductance (or transconductance) of a small-signal transistor is defined by

$$g_m = \left. \frac{\partial i_c}{\partial v_{b'e}} \right|_{v_{ce}} \qquad (5\text{-}1\text{-}11)$$

From the diode junction theory the thermal equilibrium density at the junction is equal to the minority density times the forward-bias voltage factor. That is,

$$n_p(0) = n_{po} e^{V_f/V_T} \qquad (5\text{-}1\text{-}12)$$

and

$$|i_c| = \frac{qAD_nn_p(0)}{L_n} \tag{5-1-13}$$

Substitution of Eq. (5-1-12) into Eq. (5-1-13) and differentiation of the resultant yield

$$g_m = \frac{|i_c|}{V_T} \tag{5-1-14}$$

where $V_T = 26$ mV at 300° K is the voltage equivalent of temperature

As the width of the base region is very narrow, $C_b \gg C_c$, the diffusion capacitance in the base charge-storage is given by

$$C_{b'e} = \frac{dQ_b}{dV_{be}} \tag{5-1-15}$$

and the total charges stored in the base is

$$Q_b = \frac{qn_p(0)W_bA}{2} \tag{5-1-16}$$

Then the diffusion capacitance is expressed by

$$C_{b'e} = \frac{qAW_bn_p(0)}{2V_T} = g_m \frac{W_b^2}{2D_n} \tag{5-1-17}$$

The voltage across the diffusion capacitance can be written as

$$V_{be}' = \frac{i_b}{j\omega C} \tag{5-1-18}$$

The small-signal input conductance of the emitter junction looking at the input of the base is defined as

$$g_b = \frac{1}{R_b} = \frac{I_B}{V_T} = \frac{I_C}{h_{FE}V_T} = \frac{g_m}{h_{FE}} \tag{5-1-19}$$

where h_{FE} is the linear or dc common-emitter current gain factor and
   g_m is the mutual conductance.

Example 5-1-1:   Equivalent Elements of a Hybrid-Pi Common-Emitter Circuit

A Si *n-p-n* bipolar transistor has the following parameters:

Collector current:	$I_c = 6$ mA
Common-emitter current gain factor:	$h_{FE} = 120$
Operational temperature:	$T = 300°$ K
Cross-sectional area:	$W_b = 10^{-8}$ cm^2

Compute: (a) the mutual conductance g_m; (b) the input conductance g_b and resistance R_i; (c) the electron diffusion coefficient D_n; and (d) the diffusion capacitance C'_{be}.

Solution

   a. The mutual conductance is

$$g_m = \frac{I_c}{V_T} = \frac{6 \times 10^{-3}}{26 \times 10^{-3}} = 0.23 \qquad \text{mho}$$

   b. The input conductance and resistance are

$$g_b = \frac{g_m}{h_{FE}} = \frac{0.23}{120} = 1.92 \times 10^{-3} \qquad \text{mho}$$

$$R_i = 521 \qquad \text{ohms}$$

   c. The electron diffusion coefficient is

$$D_n = \mu_n \frac{KT}{q} = \mu_n V_T = 1600 \times 26 \times 10^{-3} = 41.6 \text{ cm}^2/\text{s}$$

   d. The diffusion capacitance is

$$C'_{be} = g_m \frac{W_b^2}{2D_n} = 0.23 \times \frac{10^{-8}}{2 \times 41.6} = 2.76 \text{ pF}$$

5-1-3 Principles of Operation

The bipolar junction transistor (BJT) is an active three-terminal device which is commonly used as an amplifier or switch. Its principles of operation are discussed in this section.

   Modes of operation.   A bipolar transistor can operate in four different modes depending on the voltage polarities across the two junctions: normal (active) mode, saturation mode, cutoff mode, and inverse (or inverted) mode as shown in Fig. 5-1-8.

1. **Normal Mode.** If the emitter junction of an n-p-n transistor is forward-biased and the collector is reverse-biased, the transistor is operated in the normal mode as shown in Fig. 5-1-8(a). The term *forward bias* means that the positive polarity of the bias voltage is connected to the p side and the negative polarity to the n side for a p-n junction; the opposite obtains for reverse bias. Most transistor amplifiers are operated in normal mode, and its common-base current gain *alpha* is known as the *normal alpha* α_N.
2. **Saturation Mode.** When both transistor junctions are forward-biased, the transistor is in its saturation mode with very low resistance, and acts like a short circuit, as shown in Fig. 5-1-8(b).
3. **Cutoff Mode.** If both transistor junctions are reverse-biased the transistor is operated in its cutoff mode. As the current is cut off, the transistor acts like an
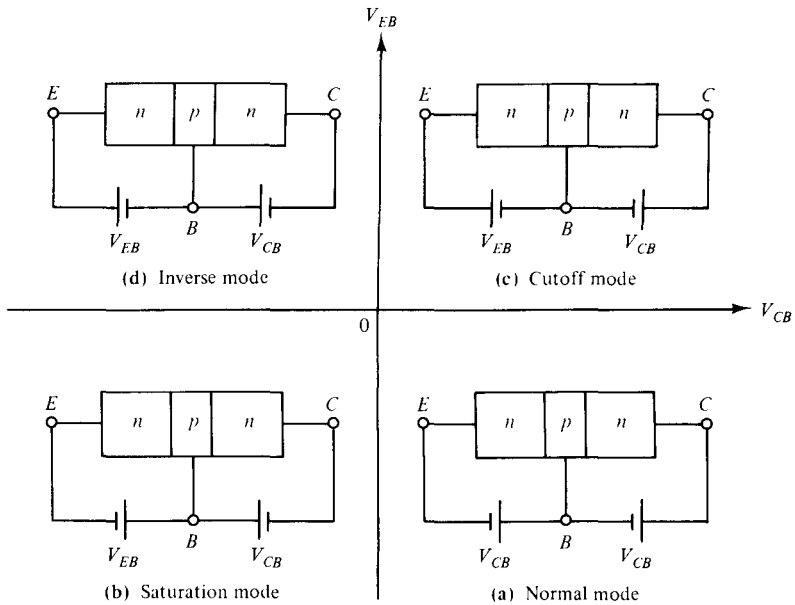
Figure 5-1-8   Operational modes of an *n-p-n* transistor.

open circuit. Both the cutoff and saturation modes of a transistor are used as switching devices for the OFF and ON states. Fig. 5-1-8(c) shows the cutoff-mode bias-voltage connection.

4. **Inverse Mode.** When the emitter is reverse-biased and the collector is forward-biased, the transistor is operated in the inverse (or inverted) mode, and its current gain is designated as the *inverse alpha* α_I. If the transistor is symmetric, the *normal alpha* α_N is nearly equal to the inverse alpha α_I. The two current gains, however, are not actually equal because of their unequal dopings. The inverse mode is shown in Fig. 5-1-8(d). In practice, the inverse mode is not commonly used except as a multiemitter transistor in TTL (transistor-transistor logic) logic gate.

Current flow in normal mode.   When a transistor is properly biased, the holes and electrons in the transistor will follow the field direction in motion. Figure 5-1-9 shows the current flow of an *n-p-n* transistor.

The current flow in an ideal *n-p-n* bipolar junction transistor is analyzed under the following assumptions:

1. The resistivities of the semiconductor regions are low
2. The injected current densities are low
3. The space-charge layer widening effects can be ignored
4. The current and voltage of *n-p* junction diodes follow the basic equation of

$$I = I_o(e^{V/V_T} - 1) \qquad (5\text{-}1\text{-}20)$$
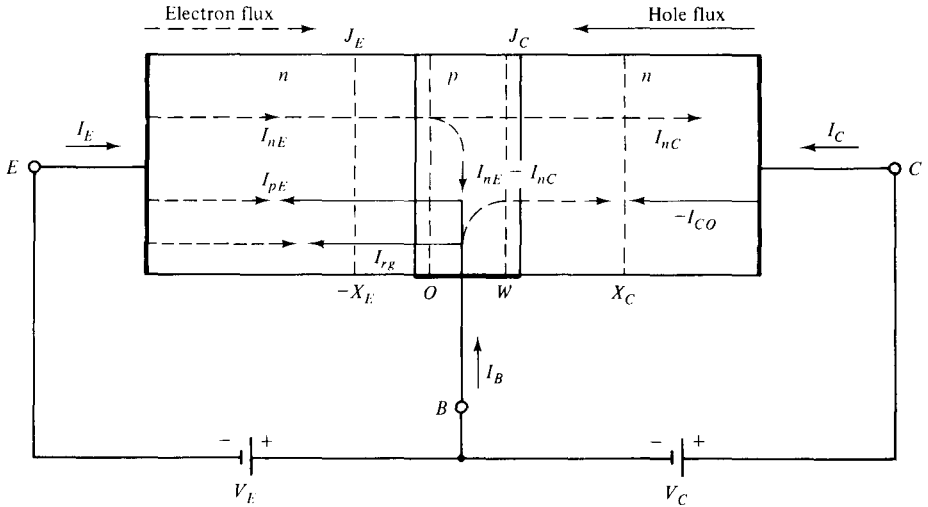
Figure 5-1-9   Current flow in an *n-p-n* transistor.

For a common-base *n-p-n* transistor, the emitter junction is forward-biased and the collector junction is reverse-biased as shown in Fig. 5-1-9. Consequently, the emitter current I_E consists of electron current I_{nE} crossing from the emitter into the base and the hole current I_{pE} crossing from the base into the emitter. Since the doping of the emitter is much larger than the doping of the base, the hole current is negligible. However, not all the electrons crossing the emitter junction J_E reach the collector junction J_C because some of them combine with the holes in the *p*-type base. If I_{nC} is the electron current at the collector junction J_C, there must be a recombination current $I_{nE} - I_{nC}$ leaving the base. When the emitter is open-circuited, then $I_E = 0$ and $I_{nC} = 0$. As a result, the collector current I_C is equal to the reverse saturation current I_{Co} because the junction between base and collector is reverse-biased. From Fig. 5-1-9, we have

$$I_B = I_{pE} - (I_{nE} - I_{nC}) - I_{CO} + I_{rg} \qquad (5\text{-}1\text{-}21)$$

$$I_E = -I_{pE} + I_{nE} - I_{rg} \qquad (5\text{-}1\text{-}22)$$

and

$$I_C = -I_{CO} - I_{nC} \qquad (5\text{-}1\text{-}23)$$

For an *n-p-n* transistor, I_{CO} consists of holes moving across the collector junction J_C from right to left (collector to base) and electrons crossing J_C in the opposite direction. Since the reference direction for I_{CO} in Fig. 5-1-9 is assumed from left to right, then, for an *n-p-n* transistor, I_{CO} is positive for forward-biased J_C junction and I_{CO} is negative for reverse-biased J_C junction. The saturation current I_{CO} at the J_C junction of an *n-p-n* transistor is given by

$$I_{CO} = \frac{AqDn_{po}}{W} + \frac{AqD_p p_{no}}{L_E} = Aqn_i^2\left(\frac{D_n}{WN_a} + \frac{D_p}{L_E N_d}\right) \qquad (5\text{-}1\text{-}24)$$

Also from Fig. 5-1-9, the sum of the three terminal currents should be zero and it is

$$I_E + I_C + I_B = 0 \qquad (5\text{-}1\text{-}25)$$

Equation (5-1-25) can be verified by adding together Eqs. (5-1-21) through (5-1-23).

Current flow in common-base *n-p-n* transistor.

In a common-base configuration of an *n-p-n* transistor as shown in Fig. 5-1-9, the emitter *n-p* junction is forward-biased and the collector *p-n* junction is reverse biased. Their total current flow can be found from the basic diffusion equation. The steady-state diffusion equation for an *n-p-n* transistor at low-level injection is given by

$$I_n = AqD_n \frac{dn_p}{dx} \qquad (5\text{-}1\text{-}25a)$$

and

$$D_n = \frac{d^2 n_p}{dx^2} - \frac{n_p - n_{po}}{\tau_n} = 0 \qquad (5\text{-}1\text{-}26)$$

where D_n = electron diffusion constant

n_p = minority electron carrier density in the p-type base layer

n_{po} = equilibrium minority electron carrier density in the p-type base layer

x = distance measured from the base region as shown in Fig. 5-1-10

τ_n = electron lifetime
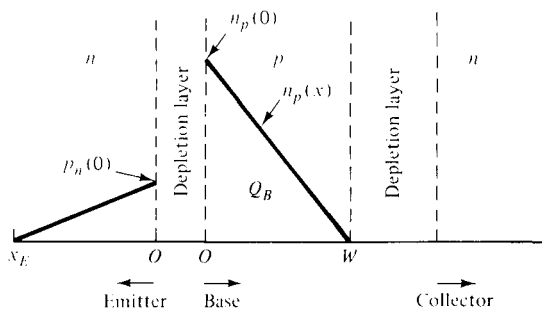
Figure 5-1-10 Minority-carrier densities under normal active bias with negligible recombination.

The general solution of Eq. (5-1-26) can be written as

$$n_p - n_{po} = C_1 e^{-x/L_n} + C_2 e^{x/L_n} \qquad (5\text{-}1\text{-}27)$$

where C_1 and C_2 are constants to be determined by the boundary conditions

$L_n = \sqrt{\tau_n D_n}$ is the electron diffusion length

$n_{po} = n_i^2/N_a$ is the mass-action law

The boundary conditions at the edge of the emitter depletion layer in the base side with a forward-biased emitter junction is

$$n_p(0) = n_{po} e^{V_E/V_T} \qquad (5\text{-}1\text{-}28)$$

where V_E = forward-biased voltage across the emitter junction
$V_T = 26 \times 10^{-3}$ V at 300° k is the voltage equivalent of temperature

The boundary condition at the depletion-layer edge of the reverse-biased collector junction is usually assumed to be

$$n_p(W) = 0 \qquad (5\text{-}1\text{-}29)$$

The two boundary conditions are shown in Fig. 5-1-10.
   The general solution of Eq. (5-1-26) can be written as

$$n_p(x) = n_{po}(e^{V_E/V_T} - 1)\left[\frac{\sinh\left[(W - x)/L_n\right]}{\sinh(W/L_n)} + n_{po}\left(1 - \frac{\sinh(x/L_n)}{\sinh(W/L_n)}\right)\right] \quad (5\text{-}1\text{-}30)$$

   In almost all transistors, the base width is made very narrow ($W \ll L_n$) so that the minority-carrier recombination in the base is negligible. As a result, the boundary conditions specify the two end points of the base carrier concentration with a straight line as shown in Fig. 5-1-11. (*Note:* $\sinh y \doteq y$, $\cosh y \doteq 1$, $\coth y \doteq 1/y$, and $\operatorname{sech} y \doteq 1 - y^2/2$ for $y \ll 1$.)
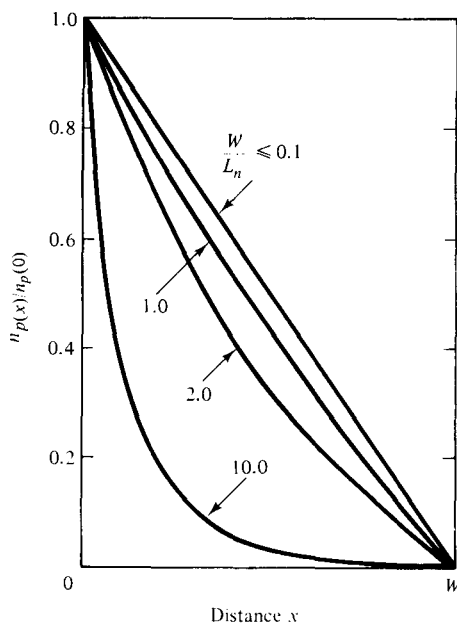
Figure 5-1-11 Minority carrier distributions in base region as a function of W/l_n. (*After S. Sze [3].*)

   For $W \ll L_n$, the minority carrier distribution of Eq. (5-1-30) can be simplified to

$$n_p(x) = n_p(0)\left(1 - \frac{x}{W}\right) = n_{po}e^{V_E/V_T}\left(1 - \frac{x}{W}\right) \qquad (5\text{-}1\text{-}31)$$

Then the boundary conditions for holes in the emitter and collector depletion regions can be expressed, respectively, as

$$p_E = p_{EO}e^{V_E/V_T} \qquad \text{at } x = -x_E \qquad (5\text{-}1\text{-}32)$$

and

$$p_C = p_{CO}e^{-V_C/V_T} \qquad \text{at } x = x_C \qquad (5\text{-}1\text{-}33)$$

where p_{EO} = equilibrium minority hole density in the emitter region

$\qquad p_{CO}$ = equilibrium minority hole density in the collector region

Substituting Eqs. (5-1-32) and (5-1-33) into Eq. (5-1-27) yields the minority distributions in the emitter and collector regions as

$$p_E(x) = p_{EO} + p_{EO}(e^{V_E/V_T} - 1) \exp\left[(x + x_E)/L_n\right] \qquad \text{for } x \leq -x_E \qquad (5\text{-}1\text{-}34)$$

and

$$p_C(x) = p_{CO} + p_{CO} \exp\left[-(x - x_C)/L_n\right] \qquad \text{for } x \geq x_C \qquad (5\text{-}1\text{-}35)$$

The total excess minority-carrier charge in the base region is given by

$$Q_B = Aq \int_O^W \left[n_p(x) - n_{po}(x)\right] dx \simeq \frac{AqWn_p(0)}{2} \qquad (5\text{-}1\text{-}36)$$

where A = cross section

The base recombination current for a carrier lifetime τ_n is

$$In_B = \frac{AQ_B}{\tau_n} = \frac{AqWn_{po}}{2\tau_n} e^{V_E/V_T} \qquad (5\text{-}1\text{-}37)$$

According to the carrier injection processes, the minority electron density at the reverse-biased collector junction J_C is usually assumed to be zero. That is,

$$n_{pC} = 0 \qquad \text{at } J_C \text{ for } x = W \qquad (5\text{-}1\text{-}38)$$

This assumption is reasonable because the electric field of the collector junction sweeps carriers into the collector so that the collector is almost a perfect sink.

The electron current I_{nE} which is injected from the emitter into the base at $x = 0$ for $L_n \gg W$ is proportional to the gradient of the minority carrier density and is expressed as

$$I_{nE} = AqD_n \frac{dn_p}{dx}\bigg|_{x=0} = \frac{-AqD_n n_{po}}{L_n} \coth(W/L_n)\left[(e^{V_E/V_T} - 1) + \frac{1}{\cosh(W/L_n)}\right]$$

$$= \frac{-AqD_n n_i^2}{N_a W}(e^{V_E/V_T} - 1) - \frac{AqD_n n_i^2}{N_a W} \doteq -\frac{AqD_n n_i^2}{N_a W} e^{V_E/V_T} \qquad (5\text{-}1\text{-}39)$$

The electron current which reaches the collector at $x = W$ is

$$I_{nC} = Aq \frac{dn_p}{dx}\bigg|_{x=W} = \frac{-AqD_n n_{po}}{L_n} \frac{1}{\sinh(W/L_n)}\left[(e^{V_E/V_T} - 1) + \cosh(W/L_n)\right]$$

$$= \frac{-AqD_n n_i^2}{N_a W}(e^{V_E/V_T} - 1) - \frac{AqD_n n_i^2}{N_a W} = -\frac{AqD_n n_i^2}{N_a W} e^{V_E/V_T} \qquad (5\text{-}1\text{-}40)$$

The collector current I_{nC} can also be expressed as

$$I_{nC} = I_{nE} - I_{nB} = I_{nE}\left(1 - \frac{I_{nB}}{I_{nE}}\right)$$

$$= I_{nE}\left(1 - \frac{W^2}{2\tau_n D_n}\right) = I_n^E\left(1 - \frac{W^2}{2L_E^2}\right) \qquad (5\text{-}1\text{-}41)$$

where L_E = diffusion length of the emitter and exp $(V_E/V_T) \gg 1$ is assumed

Similarly, the hole currents are

$$I_{pE} = \frac{AqD_E p_{EO}}{LE}(e^{V_E/V_T} - 1) \qquad \text{for } x = -x_E \qquad (5\text{-}1\text{-}42)$$

and

$$I_{pC} = \frac{AqD_C p_{CO}}{L_C} \qquad \text{for } x = x_C \qquad (5\text{-}1\text{-}43)$$

where D_E and D_C are the hole diffusion constants in the emitter and collector,
respectively
L_C = diffusion length of the collector

The current flow in an n-p-n transistor as described so far is an ideal model, and its recombination-generation current is not counted. If the recombination current is considered, the current flow is the sum of the drift, the diffusion, and the recombination-generation currents. That is,

$$I = I_{dr} \text{ (drift)} + I_{df} \text{ (diffusion)} + I_{rg} \text{ (recombination-generation)} \qquad (5\text{-}1\text{-}44)$$

The recombination-generation current can be computed from the following equation

$$I_{rg} = \frac{Aqn_i x_d}{\tau O}e^{V_E/(2V_T)} \qquad (5\text{-}1\text{-}45)$$

where x_d = depletion-layer width
τ_o = effective minority-carrier lifetime in the depletion layer

Figure 5-1-12 shows current-voltage (I-V) characteristics of an ideal n-p-n bipolar transistor for a common-base configuration.

There are three regions for the I-V characteristics of an n-p-n bipolar transistor:

1. **Active Region:** In this region the emitter junction is forward-biased and the collector junction is reverse-biased. The collector current I_C is essentially independent of collector voltage and depends only on the emitter current I_E. When the emitter current is zero, the collector current is equal to the reverse saturation current I_{Co}.

2. **Saturation Region:** In this region, as shown on the left side of Fig. 5-1-12, both emitter and collector junctions are forward-biased. The electron current flows from the n side across the collector junction to the p-type base. As a result, the collector current increases sharply.
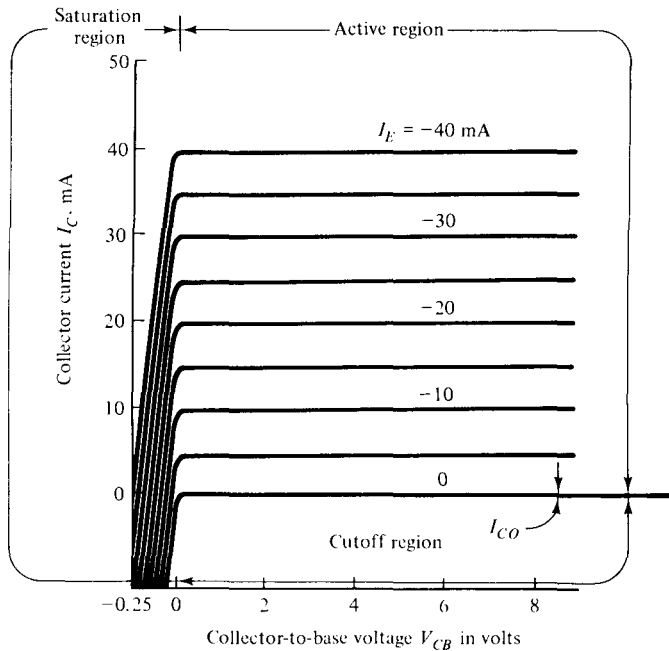
Figure 5-1-12   Current-Voltage (I-V) characteristics of an *n-p-n* transistor.

3. Cutoff Region: In this region the emitter and collector junctions are both reverse-biased. Consequently, the emitter current is cut off to zero, as shown in the lower right side of Fig. 5-1-12.

Example 5-1-2:   I-V Characteristics of an *n-p-n* Transistor

A silicon *n-p-n* transistor at 300° K has the following parameters:

Base width:	$w = 10^{-5}$ cm
Diffusion length in emitter:	$L_E = 10^{-4}$ cm
Diffusion length in collector:	$L_C = 5 \times 10^{-4}$ cm
Base resistivity:	$\rho_B = 0.15$ Ω-cm
Emitter resistivity:	$\rho_E = 0.006$ Ω-cm
Collector resistivity:	$\rho_C = 16$ Ω-cm
Emitter junction voltage:	$V_E = 0.5$ V
Collector junction voltage:	$V_C = 0.6$ V
Cross-section area:	$A = 2 \times 10^{-2}$ cm^2

Find:

    a. The impurity densities in the emitter, base, and collector regions
    b. The mobilities in the emitter, base, and collector regions
    c. The diffusion lengths in the emitter, base, and collector regions

   d. The equilibrium densities in the emitter, base, and collector regions
   e. The terminal currents

Solution

   a. The impurity densities are read from Fig. A-1 in Appendix A as
   $N_{dE} = 1 \times 10^{19}$ cm^{-3} in the n-type emitter region
   $N_{aB} = 1.5 \times 10^{17}$ cm^{-3} in the p-type base region
   $N_{dC} = 3 \times 10^{14}$ cm^{-3} in the n-type collector region

   b. The mobilities are read from Fig. A-2 in Appendix A as
   $\mu_{pE} = 80$ $cm^2/V \cdot s$ in the emitter
   $\mu_{nE} = 105$ $cm^2/V \cdot s$ in the emitter
   $\mu_{pB} = 400$ $cm^2/V \cdot s$ in the base
   $\mu_{nC} = 1600$ $cm^2/V \cdot s$ in the collector

   c. The diffusion constants are computed to be
   $D_{pE} = \mu_{pE}V_T = 80 \times 26 \times 10^{-3} = 2.08$ cm^2/s
   $D_{nE} = \mu_{nE}V_T = 105 \times 26 \times 10^{-3} = 2.73$ cm^2/s
   $D_{pB} = \mu_{pB}V_T = 400 \times 26 \times 10^{-3} = 10.4$ cm^2/s
   $D_{nC} = \mu_{nC}V_T = 1600 \times 26 \times 10^{-3} = 41.6$ cm^2/s

   d. The equilibrium densities are
   $n_{PB} = n_i^2/N_{aB} = (1.5 \times 10^{10})^2/(1.5 \times 10^{17}) = 1.5 \times 10^3$ cm^{-3}
   $p_{EO} = n_i^2/N_{dE} = (1.5 \times 10^{10})^2/(1 \times 10^{19}) = 2.5 \times 10^4$ cm^{-3}
   $p_{CO} = n_i^2/N_{dC} = (1.5 \times 10^{10})^2/(3 \times 10^{14}) = 7.5 \times 10^5$ cm^{-3}

   e. The terminal currents are computed as follows:
   From Eq. (5-1-39), the electron current in the emitter is

$$I_{nE} = -\frac{AqD_n n_i^2}{N_a W} e^{V_E/V_T} = \frac{A_q D_p n_i^2}{L_E N_d}(e^{V_E/ \div V_T} - 1)$$

$$= -\frac{2 \times 10^{-2} \times 1.6 \times 10^{-19} \times 2.73 \times (1.5 \times 10^{10})^2}{1.5 \times 10^{17} \times 10^{-5}}$$

$$\times \exp[0.5/(26 \times 10^{-3})]$$

$$= -13.104 \times 10^{-13} \times 2.248 \times 10^8$$

$$= -0.2946 \text{ mA}$$

   From Eq. (5-1-42), the hole current in the emitter is

$$I_{pE} = \frac{AqD_E p_{EO}}{L_E}(e^{V_E/V_T} - 1) = \frac{AqD_p n_i^2}{L_E N_d}(e^{V_E/V_T)} - 1)$$

$$= \frac{2 \times 10^{-2} \times 1.6 \times 10^{-19} \times 2.08 \times (1.5 \times 10^{10})^2}{10^{-4} \times 1 \times 10^{19}} \times (2.248 \times 10^8 - 1)$$

$$= 14.976 \times 10^{-16} \times 2.248 \times 10^8$$

$$= 0.337 \ \mu A$$

   From Eq. (5-1-24), the reverse saturation current in the collector is

$$I_{CO} = \frac{-AqD_n n_i^2}{N_a W} - \frac{AqD_p p_{no}}{L_E}$$

$$= -13.104 \times 10^{-13} - 14.976 \times 10^{-16}$$

$$= -1.312 \ pA$$

From Eq. (5-1-40), the electron current which reaches the collector is

$$I_{nC} = -\frac{AqD_n n_i^2}{N_a W} e^{V_E/V_T} = -13.104 \times 10^{-13} \times 2.248 \times 10^8$$

$$= -0.2946 \ mA$$

The emitter current is

$$I_E = -I_{pE} + I_{nE} = -33.67 \times 10^{-8} - 0.295 \times 10^{-3}$$

$$= -0.295 \ mA$$

The collector current is

$$I_C = -I_{CO} - I_{nC} = 1.312 \times 10^{-12} - (-0.295 \times 10^{-3})$$

$$= 0.295 \ mA$$

The current in the base terminal is

$$I_B = I_{pE} - (I_{nE} - I_{nC}) + I_{CO}$$

$$= 33.67 \times 10^{-8} - (-29.46 \times 10^{-5} + 29.46 \times 10^{-5}) + 1.312 \times 10^{-12}$$

$$= 0.337 \ \mu A$$

Note: The recombination-generation currents in the space-charge regions are not counted.

5-1-4 Amplification Phenomena

Bipolar transistors are usually used for signal amplification. The amplification phenomena can be described from the common-base and common-emitter transistors.

Common-base *n-p-n* transistor.   The ratio of the output current to the input current for a small signal in a bipolar junction transistor is known as the *current gain alpha* α, or h_{fb}. The current gain of a common-base *p-n-p* transistor is defined by the current components crossing the emitter and collector junctions as

$$\alpha = -\frac{I_C + I_{CO}}{I_E} = h_{fb} \tag{5-1-46}$$

where I_{CO} = collector-junction reverse saturation current with zero emitter current

Since I_C and I_E have opposite signs, the alpha α, as defined, as always positive. Typical numerical values of α are between 0.9 and 0.995.

The emitter efficiency (or injection efficiency) is defined as

$$\gamma = \frac{\text{current of injected carriers at } J_E}{\text{total emitter current}}$$

$$= \frac{I_{nE}}{I_{nE} + I_{pE}} = \frac{1}{1 + I_{pE}/I_{nE}} = \left(1 + \frac{D_p N_a W}{D_n N_d L_E}\right)^{-1} \tag{5-1-47}$$

where I_{nE} = injected electron diffusion current at emitter junction J_E
  I_{pE} = injected hole diffusion current at emitter junction J_E

The transport factor β^* is defined as

$$\beta^* = \frac{\text{injected carrier current reaching } J_C}{\text{injected carrier current at } J_E} \tag{5-1-48}$$

$$= \frac{I_{nC}}{I_{nE}} = 1 - \frac{W^2}{2\tau_n D_n} = 1 - \frac{W^2}{2L_n^2}$$

where I_{nC} = injected electron diffusion current at J_C
  I_{nE} = injected electron diffusion current at J_E
  $L_n = \sqrt{\tau_n D_n}$ is the electron diffusion length

At the collector we have

$$I_C = -I_{CO} - I_{nC} \tag{5-1-49}$$

Then the current gain can be expressed as

$$\alpha = \frac{I_{nC}}{I_E} = \beta^* \gamma \tag{5-1-50}$$

In the normal active mode, the collector current is given by

$$I_C = -\alpha I_E - I_{CO} \tag{5-1-51}$$

The current I_{CO} is the current crossing the p-n junction, and it is expressed in Eq. (5-1-24). Then the complete expression of I_C for any V_C and I_E is

$$I_C = -\alpha I_E + I_{CO}(1 - e^{V_C/V_T}) \tag{5-1-52}$$

where $I_O = -I_{CO}$ is replaced.

Example 5-1-3:   Silicon Bipolar Transistor

  A silicon n-p-n bipolar transistor operates in common-base mode at 300° K and has the following parameters:

Silicon intrinsic density:	$n_i = 1.5 \times 10^{10}$	cm^{-3}
Acceptor density in base region:	$N_a = 5 \times 10^{16}$	cm^{-3}
Donor density in emitter region:	$N_d = 5 \times 10^{18}$	cm^{-3}
Hole lifetime:	$\tau_p = 1$	μs

Electron lifetime:	$\tau_n = 1$	μs
Cross section:	$A = 10^{-4}$	cm^2
Base width:	$W = 10^{-3}$	cm
Emitter length:	$L_E = 10^{-2}$	cm

Determine:

a. The mobilities μ_n and μ_p
b. The diffusion coefficients D_n and D_p
c. The emitter efficiency factor γ
d. The transport factor β
e. The current gain α

Solution

a. The mobilities are read from Fig. A-2 in Appendix A as
$\mu_n = 200$ cm^2/V·s   for $N_{dE} = 5 \times 10^{18}$ cm^{-3}
$\mu_p = 500$ cm^2/V·s   for $N_a = 5 \times 10^{16}$ cm^{-3}
b. The diffusion coefficients are
$D_n = \mu_n V_T = 200 \times 26 \times 10^{-3} = 5.20$ cm^2/s
$D_p = 500 \times 26 \times 10^{-3} = 13.0$ cm^2/s
c. The emitter efficiency factor is

$$\gamma = \left(1 + \frac{13.0 \times 5 \times 10^{16} \times 10^{-3}}{5.20 \times 5 \times 10^{18} \times 10^{-2}}\right)^{-1} = 0.997$$

d. The transport factor is

$$\beta^* = 1 - \frac{(10^{-3})^2}{2 \times 10^{-6} \times 5.20} = 0.904$$

e. The current gain is

$$\alpha = \beta^* \gamma = 0.904 \times 0.997 = 0.90$$

Common-emitter *n-p-n* transistor.   In the active region of a common-emitter *n-p-n* transistor, the emitter junction is forward-biased and the collector junction is reverse-biased. The base current is

$$I_B = -(I_C + I_E) \tag{5-1-53}$$

Combining this equation with Eq. (5-1-51), we have

$$I_C = \frac{-I_{CO}}{1 - \alpha} + \frac{\alpha I_B}{1 - \alpha} \tag{5-1-54}$$

In the cutoff region, if $I_B = 0$, then $I_E = -I_C$ and the collector current is given by

$$I_C = -I_E = \frac{-I_{CO}}{1 - \alpha} = I_{CEO} \tag{5-1-55}$$

The actual collector current with the collector junction reverse-biased and the base junction open-circuited is designated by the symbol I_{CEO}. We define the common-emitter current gain β or h_{fe} as

$$\beta = h_{fe} = \frac{I_C + I_{CO}}{I_B} \tag{5-1-56}$$

Since $I_E + I_B + I_C + I_{CO} = 0$, we have

$$\beta = \frac{\alpha}{1 - \alpha} \tag{5-1-57}$$

5-1-5 Power-Frequency Limitations

The question then arises as to whether microwave power transistors have any limitations on their frequency and output power. The answer is yes. Several authors have discussed this subject. Early [4] first introduced the power-frequency limitations inherent in (1) the limiting velocity of carriers in semiconductors and (2) the maximum fields attainable in semiconductors without the onset of avalanche multiplication. These basic ideas were later developed and discussed in detail by Johnson [5], who made three assumptions:

1. There is a maximum possible velocity of carriers in a semiconductor. This is the "saturated drift velocity v_s," and is on the order of 6×10^6 cm/s for electrons and holes in silicon and germanium.
2. There is a maximum electric field E_m that can be sustained in a semiconductor without having dielectric breakdown. This field is about 10^5 V/cm in germanium and 2×10^5 V/cm in silicon.
3. The maximum current that a microwave power transistor can carry is limited by the base width.

With these three postulates Johnson derived four basic equations for the power-frequency limitations on microwave power transistors.

First equation: Voltage-Frequency Limitation:

$$V_m f_T = \frac{E_m v_s}{2\pi} = \begin{cases} 2 \times 10^{11} \text{ V/s for silicon} \\ 1 \times 10^{11} \text{ V/s for germanium} \end{cases} \tag{5-1-58}$$

where $f_T = \dfrac{1}{2\pi\tau}$ is the charge-carrier transit-time cutoff frequency

$\quad\tau = \dfrac{L}{v}$ is the average time for a charge carrier moving at an average velocity v to traverse the emitter-collector distance L

$\quad V_m = E_m L_{min}$ is the maximum allowable applied voltage

$\quad v_s =$ maximum possible saturated drift velocity

$\quad E_m =$ maximum electric field

With the carriers moving at a velocity v_s of 6×10^6 cm/s, the transit time can

be reduced even further by decreasing the distance L. The lower limit on L can be reached when the electric field becomes equal to the dielectric breakdown field. However, the present state of the art of microwave transistor fabrication limits the emitter-collector length L to about 25 μm for overlay and matrix devices and to nearly 250 μm for interdigitated devices. Consequently, there is an upper limit on cutoff frequency. In practice, the attainable cutoff frequency is considerably less than the maximum possible frequency indicated by Eq. (5-1-58) because the saturated velocity v_s and the electric field intensity will not be uniform.

Second equation: Current-Frequency Limitation:

$$(I_m X_c) f_T = \frac{E_m v_s}{2\pi} \tag{5-1-59}$$

where I_m = maximum current of the device

$X_c = \dfrac{1}{\omega_T C_0} = \dfrac{1}{2\pi f_T C_0}$ is the reactive impedance

C_0 = collector-base capacitance

It should be noted that the relationship $2\pi f_T \tau_0 \approx 2\pi f_T \tau = 1$ is used in deriving Eq. (5-1-59) from Eq. (5-1-58). In practice, no maximum current exists because the area of the device cannot be increased without bound. If the impedance level is zero, the maximum current through a velocity-saturated sample might be infinite. However, the limited impedance will limit the maximum current for a maximum attainable power.

Third equation: Power-Frequency Limitation:

$$(P_m X_c)^{1/2} f_T = \frac{E_m v_s}{2\pi} \tag{5-1-60}$$

This equation was obtained by multiplying Eq. (5-1-58) by Eq. (5-1-59) and replacing $V_m I_m$ by P_m. It is significant that, for a given device impedance, the power capacity of a device must be decreased as the device cutoff frequency is increased. For a given product of $E_m v_s$ (that is, a given material), the maximum power that can be delivered to the carriers traversing the transistor is infinite if the cross section of the transistor can be made as large as possible. In other words, the value of the reactance X_c must approach zero. Thus Eq. (5-1-60) allows the results to be predicted. Figure 5-1-13 shows a graph of Eq. (5-1-60) and the experimental results reported from manufacturers [6].

Fourth equation: Power Gain-Frequency Limitation:

$$(G_m V_{th} V_m)^{1/2} f = \frac{E_m v_s}{2\pi} \tag{5-1-61}$$

where G_m = maximum available power gain

$V_{th} = KT/e$ is the thermal voltage

K = Boltzmann's constant, 1.38×10^{-23} J/°K
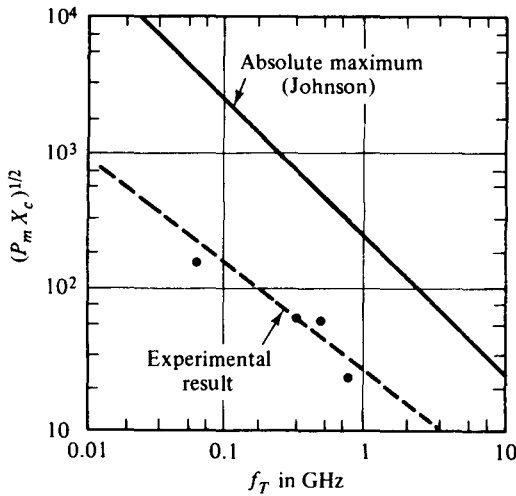
Figure 5-1-13 $(P_m X_c)^{1/2}$ versus f_T.
(*After B. C. De Loach [6]: reprinted by permission of Academic Press.*)

$$T = \text{absolute temperature in degrees Kelvin}$$
$$e = \text{electron charge } (1.60 \times 10^{-19} \text{ C})$$

The maximum available power gain of a transistor was derived by Johnson [5] as

$$G_m = \left(\frac{f_T}{f}\right)^2 \frac{\mathbf{Z}_{\text{out}}}{\mathbf{Z}_{\text{in}}} \qquad (5\text{-}1\text{-}62)$$

where $\mathbf{Z}_{\text{out}}$ and $\mathbf{Z}_{\text{in}}$ are the output and input impedances, respectively. If the electrode series resistances are assumed to be zero, the ratio of the output impedance to the input impedance can be written

$$\frac{\mathbf{Z}_{\text{out}}}{\mathbf{Z}_{\text{in}}} = \frac{C_{\text{in}}}{C_{\text{out}}} \qquad (5\text{-}1\text{-}63)$$

where C_{in} is the input capacitance and C_{out} is the output (base-collector) capacitance. When the maximum total carrier charges Q_m move to the collector in a carrier-base transit time τ_b and with a thermal voltage V_{th}, the input capacitance C_{in} and the emitter diffusion capacitance C_d are related by

$$C_{\text{in}} = C_d \approx \frac{Q_m}{V_{\text{th}}} = \frac{I_m \tau_b}{V_{\text{th}}} \qquad (5\text{-}1\text{-}64)$$

The output capacitance is given by

$$C_{\text{out}} = \frac{I_m \tau_0}{V_m} \qquad (5\text{-}1\text{-}65)$$

Substitution of Eqs. (5-1-58), (5-1-64), and (5-1-65) in Eq. (5-1-62) yields Eq. (5-1-61). The actual performance of a microwave transistor will fall far short of that predicted by Eq. (5-1-61). At present the high-frequency limit of a 28-V silicon *n-p-n* transistor operating at the 1-W level is approximately 10 Ghz. Typical power gains of microwave transistors lie in the 6- to 10-dB range.

Example 5-1-4:   Power-Frequency Limitation

A certain silicon microwave transistor has the following parameters:

Reactance:	$X_c = 1\ \Omega$
Transit-time cutoff frequency:	$fr = 4$ GHz
Maximum electric field:	$E_m = 1.6 \times 10^5$ V/cm
Saturation drift velocity:	$v_s = 4 \times 10^5$ cm/s

Determine the maximum allowable power that the transistor can carry.

Solution.   From Eq. (5-1-60) the maximum allowable power is

$$P_m = \frac{1}{X_c f_T^2}\left(\frac{E_m v_s}{2\pi}\right)^2 = \frac{1}{1 \times (4 \times 10^9)^2}\left(\frac{1.6 \times 10^5 \times 4 \times 10^5}{2\pi}\right)^2 = 6.48\ \text{W}$$

5-2 HETEROJUNCTION BIPOLAR TRANSISTORS (HBTs)

Bipolar transistors can be constructed as homojunction or heterojunction types of transistors. When the transistor junction is jointed by two similar materials such as silicon to silicon or germanium to germanium, it is a *homojunction transistor*. The transistor junction formed by two different materials, such as Ge to GaAs, is called a *heterojunction transistor*. So far only the ordinary homojunction transistors have been discussed. In this section we study the heterojunction transistors.

5-2-1 Physical Structures

When the lattice constants of two semiconductor materials are matched, they can be formed together as a heterojunction transistor. This lattice condition is very important because the lattice mismatch could introduce a large number of interface states and degrade the heterojunction operation. Currently, Ge and GaAs are the two materials commonly used for heterojunction structures because their lattice constants ($a = 5.646$ Å for Ge and $a = 5.653$ Å for GaAs) are matched to within 1%. Since each material may be either p type or n type, there are four possible heterojunction combinations:

1. p-Ge to p-GaAs junction
2. p-Ge to n-GaAs junction
3. n-Ge to p-GaAs junction
4. n-Ge to n-GaAs junction

Fig. 5-2-1 shows the model diagram of a heterojunction transistor formed by n-Ge, p-GaAs, and n-GaAs materials.
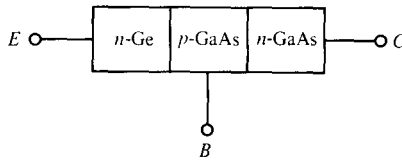
Figure 5-2-1   Model diagram of a heterojunction transistor.

5-2-2 Operational Mechanism

When an n-Ge and a p-GaAs are isolated, their Fermi energy levels are not aligned, as shown in Fig. 5-2-2 [7].

In Fig. 5-2-2, the vacuum level is used as reference, the work function is denoted by ϕ, n-Ge is designated as 1, and p-GaAs is referred to as 2. The different energies of the conduction-band edge and the valence-band edge are given by

$$\Delta E_c = \chi_1 - \chi_2 \tag{5-2-1}$$

and

$$\Delta E_v = E_{g2} - E_{g1} - \Delta E_c \tag{5-2-2}$$

where χ = electron affinity in eV
E_g = bandgap energy in eV
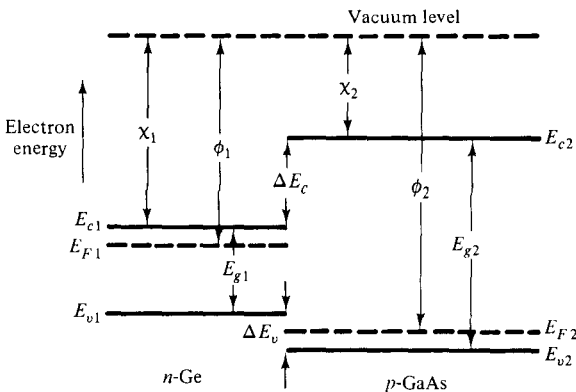
Figure 5-2-2   Energy-band diagram for isolated n-Ge and p-GaAs.

Example 5-2-1:   Heterojunction Bipolar Transistor (HBT)

A Ge-GaAs heterojunction transistor has the following parameters:

Lattice constant: Ge	$a_1 = 5.646$ Å	
GaAs	$a_2 = 5.653$ Å	
Electron affinity: Ge	$\chi_1 = 4.0$	eV
GaAs	$\chi_2 = 4.07$	eV
Energy gap: Ge	$E_{g1} = 0.80$	eV
GaAs	$E_{g2} = 1.43$	eV

Determine:

  a. The lattice match in percent
  b. The conduction-band differential between Ge and GaAs
  c. The valence-band differential between Ge and GaAs

Solution

  a. The lattice match is within 1%
  b. The conduction-band differential is

$$\Delta E_c = \chi_1 - \chi_2 = 4.0 - 4.07 = -0.07 \quad \text{eV}$$

  c. The valance-band differential is

$$E_v = E_{g2} - E_{g1} - \Delta E_c = 1.43 - 0.8 - (-0.07)$$

$$= 0.70 \text{ eV}$$

When the two semiconductor materials are jointed together, their Fermi energy levels are aligned and their energy bands are depleted at the junction, as shown in Fig. 5-2-3 [7].
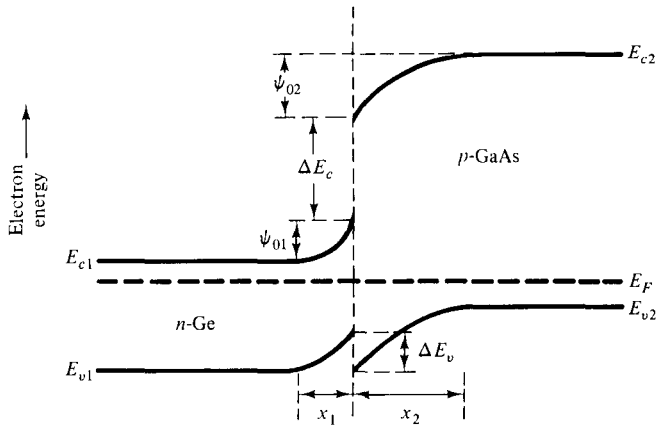
Figure 5-2-3   Energy-band diagram for an n-Ge–p-GaAs junction.

When the two materials are jointed together, the electrons in the n-Ge are injected into the p-GaAs, and the holes in the p-GaAs are transferred to the n-Ge until their Fermi energy levels are aligned. As a result, the energy bands at the junction are depleted or bent. The bending energy then creates a built-in voltage in both sides of the junction. The total built-in voltage is expressed by

$$\psi_o = \psi_{o1} + \psi_{o2} \tag{5-2-3}$$

where ψ_{o1} = barrier potential or portion of built-in voltage in n-Ge
    ψ_{o2} = barrier potential or portion of built-in voltage in p-GaAs

At the junction, the electric flux D is continuous. That is,

$$D = \epsilon_o \epsilon_{r1} \mathscr{E}_1 = \epsilon_o \epsilon_{r2} \mathscr{E}_2 \qquad (5\text{-}2\text{-}4)$$

$$= \epsilon_1 \mathscr{E}_1 = \epsilon_2 \mathscr{E}_2$$

where ϵ_o = permittivity of free space
$\quad\;\epsilon$ = permittivity
$\quad\;\epsilon_r$ = relative permittivity or dielectric constant
$\quad\;\mathscr{E}$ = electric field

The space charges on both sides of the junction are equal and it is given by

$$x_1 N_d = x_2 N_a \qquad (5\text{-}2\text{-}5)$$

where x_1 = depletion width in n-Ge
$\quad\; x_2$ = depletion width in p-GaAs
$\quad\; N_d$ = donor density
$\quad\; N_a$ = acceptor density

The electric fields in both sides can be written as

$$\mathscr{E}_1 = \frac{\psi_{o1} - V_1}{x_1} \qquad (5\text{-}2\text{-}6)$$

and

$$\mathscr{E}_2 = \frac{\psi_{o2} - V_2}{x_2} \qquad (5\text{-}2\text{-}7)$$

where V_1 = portion of bias-voltage in n-Ge
$\quad\; V_2$ = portion of bias-voltage in p-GaAs

Substitution of Eq. (5-2-4) into Eqs. (5-2-6) and (5-2-7) results in

$$(\psi_{o1} - V_1) \epsilon_{r1} N_{d1} = (\psi_{o2} - V_2) \epsilon_{r2} N_{a2} \qquad (5\text{-}3\text{-}8)$$

For the heterojunction shown in Fig. 5-2-3, the electron current from n-Ge to p-GaAs is very small because the potential barrier of $(\psi_{o1} + \psi_{o2} + \Delta E_c/q)$ across the junction for electron injection is very high. In contrast, the hole current from the p-GaAs side to the n-Ge side is dominant because of the low potential barrier ψ_{o2} for hole injection. Therefore, the junction current can be approximated as shown in Eq. (5-1-42) to be

$$I = \frac{A q D_p p_{no}}{L_p} (e^{V/V_T} - 1) \qquad (5\text{-}2\text{-}9)$$

where A = cross section
$\quad\; q$ = electron charge
$\quad\; D_p$ = hole diffusion constant
$\quad\; P_{no}$ = minority or equilibrium hole density in n-Ge
$\quad\; L_p$ = hole diffusion length
$\quad\; V$ = bias voltage
$\quad\; V_T$ = voltage equivalent of temperature

5-2-3 Electronic Applications

Heterojunctions have been studied since 1951 and have many applications in photon devices as described in Chapter 10. The heterojunction bipolar transistor is a potential candidate for high-speed switching devices such as GaAs MESFETs. The analysis described previously can be applied to the structures of Ge-GaAs and GaAs-AlGaAs. In other heterojunction transistors, such as the Ge-Si structure, the lattice mismatch ($a = 5.646$ Å for Ge and $a = 5.431$ Å for Si) causes a high interface state density and recombination- and tunneling-current components must be counted.

Example 5-2-2: n-Ge–p-GaAs–n-GaAs HBT

An n-Ge–p-GaAs–n-GaAs heterojunction transistor at 300° K has the following parameters:

Donor density in n-Ge region:	$N_d = 5 \times 10^{18}$ cm^{-3}
Acceptor density in p-GaAs region:	$N_a = 6 \times 10^{16}$ cm^{-3}
Hole lifetime:	$\tau_p = 6 \times 10^{-6}$ sec
Bias voltage at emitter junction:	$V_E = 1$ V
Cross section:	$A = 2 \times 10^{-2}$ cm^2

Compute:

a. The built-in voltage in the p-GaAs side
b. The hole mobility
c. The hole diffusion constant
d. The minority hole density in the n-Ge region
e. The minority electron density in the p-GaAs region
f. The hole diffusion length
g. The emitter-junction current

Solution

a. The built-in voltage in the p-GaAs side is

$$\psi_{02} = -26 \times 10^{-3} \ln\left(\frac{6 \times 10^{16}}{1.8 \times 10^6}\right) = -0.63 \text{ V}$$

b. The hole mobility is read from Fig. A-2 in Appendix A as

$$\mu_p = 400 \text{ cm}^2/\text{V} \cdot \text{s}$$

c. The hole diffusion constant is

$$D_p = 400 \times 26 \times 10^{-3} = 10.40 \text{ cm}^2/\text{s}$$

d. The minority hole density in n-Ge is

$$P_{no} = \frac{(1.5 \times 10^{10})^2}{5 \times 10^{18}} = 45 \text{ cm}^{-3}$$

e. The minority electron density in the p-GaAs region is

$$n_{po} = \frac{(1.8 \times 10^6)^2}{6 \times 10^{16}} = 5.4 \times 10^{-3} \quad cm^{-3}$$

f. The hole diffusion length is

$$L_p = \sqrt{6 \times 10^{-6} \times 10.40} = 7.90 \times 10^{-3} \quad cm$$

g. The emitter-junction current is computed from Eq. (5-2-9) as

$$I = \frac{2 \times 10^{-2} \times 1.6 \times 10^{-19} \times 10.4 \times 45}{7.9 \times 10^{-3}} [e^{1/(26 \times 10^{-3})} - 1]$$

$$= 0.958 \text{ A}$$

5-3 MICROWAVE TUNNEL DIODES

After the publication of Esaki's classic paper on tunnel diodes in 1958, the potential of tunnel diodes for microwave applications was quickly established. Prior to 1958 the anomalous characteristics of some p-n junctions were observed by many scientists, but the irregularities were rejected immediately because they did not follow the "classic" diode equation. Esaki, however, described this anomalous phenomenon by applying a quantum tunneling theory. The tunneling phenomenon is a majority carrier effect. The tunneling time of carriers through the potential energy barrier is not governed by the classic transit-time concept—that the transit time is equal to the barrier width divided by the carrier velocity—but rather by the quantum transition probability per unit time. Tunnel diodes are useful in many circuit applications in microwave amplification, microwave oscillation, and binary memory because of their low cost, light weight, high speed, low-power operation, low noise, and high peak-current to valley-current ratio.

5-3-1 Principles of Operation

The tunnel diode is a negative-resistance semiconductor p-n junction diode. The negative resistance is created by the tunnel effect of electrons in the p-n junction. The doping of both the p and n regions of the tunnel diode is very high—impurity concentrations of 10^{19} to 10^{20} atoms/cm^3 are used—and the depletion-layer barrier at the junction is very thin, on the order of 100 Å or 10^{-6} cm. Classically, it is possible for those particles to pass over the barrier if and only if they have an energy equal to or greater than the height of the potential barrier. Quantum mechanically, however, if the barrier is less than 3 Å there is an appreciable probability that particles will tunnel through the potential barrier even though they do not have enough kinetic energy to pass over the same barrier. In addition to the barrier thinness, there must also be filled energy states on the side from which particles will tunnel and allowed empty states on the other side into which particles penetrate through at the same energy level. In order to understand the tunnel effects fully, let us analyze the energy-band pictures of a heavily doped p-n diode. Figure 5-3-1 shows energy-band diagrams of a tunnel diode.
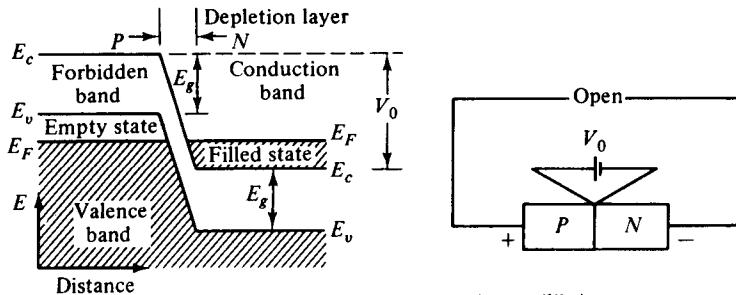
(a) Tunnel diode under zero-bias equilibrium
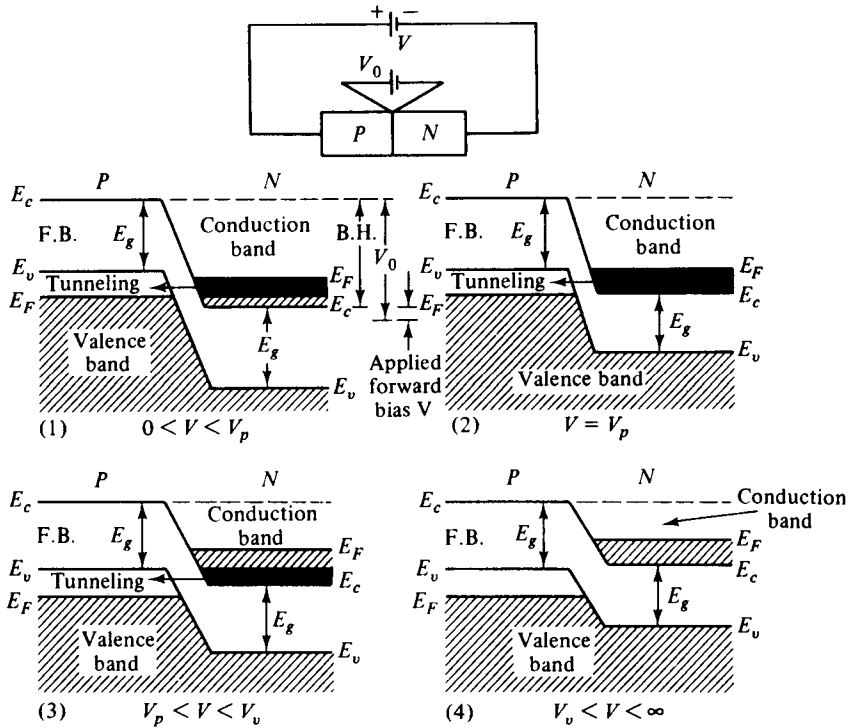
(b) Tunnel diode with applied forward bias

E_F is the Fermi level representing the energy state with 50%
    probability of being filled if no forbidden band exists
V_0 is the potential barrier of the junction
E_g is the energy required to break a covalent bond, which is
    0.72 eV for germanium and 1.10 eV for silicon
E_c is the lowest energy in the conduction band
E_v is the maximum energy in the valence band
V is the applied forward bias
F.B. stands for the forbidden band
B.H. represents the barrier height

Figure 5-3-1   Energy-band diagrams of tunnel diode.

Under open-circuit conditions or at zero-bias equilibrium, the upper levels of electron energy of both the p type and n type are lined up at the same Fermi level as shown in Fig. 5-3-1(a). Since there are no filled states on one side of the junction that are at the same energy level as empty allowed states on the other side, there is no flow of charge in either direction across the junction and the current is zero, as shown at point (a) of the volt-ampere characteristic curve of a tunnel diode in Fig. 5-3-2.

(a) Tunnelling current (solid curve) and injection current (dashed curve)

(b) I–V characteristic of tunnel diode
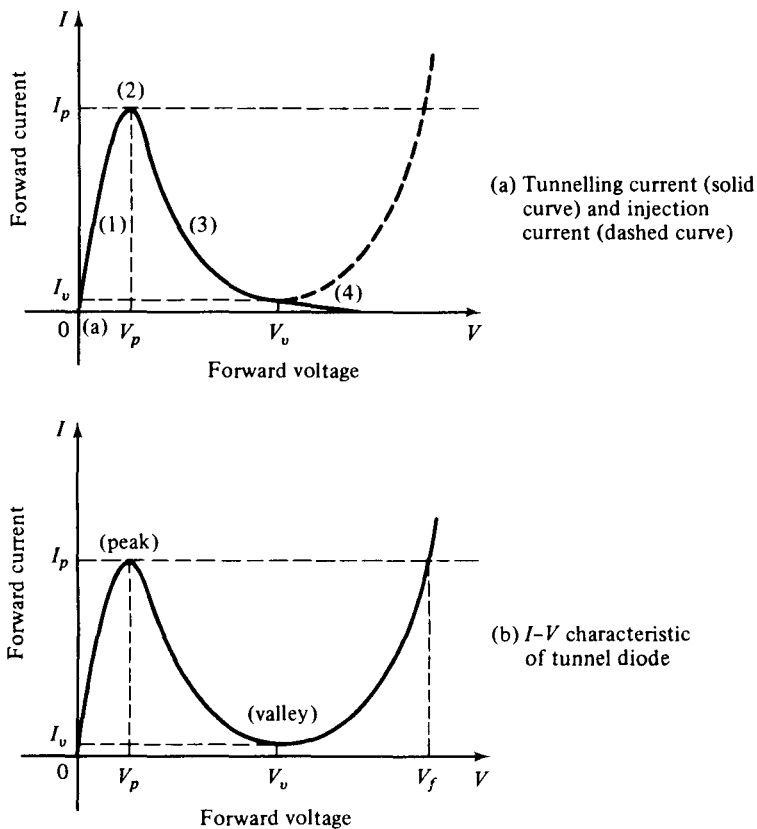
Figure 5-3-2   Ampere-voltage characteristics of tunnel diode.

In ordinary diodes the Fermi level exists in the forbidden band. Since the tunnel diode is heavily doped, the Fermi level exists in the valence band in p-type and in the conduction band in n-type semiconductors. When the tunnel diode is forward-biased by a voltage between zero and the value that would produce peak tunneling current $I_p(0 < V < V_p)$, the energy diagram is shown in part (1) of Fig. 5-3-1(b). Accordingly, the potential barrier is decreased by the magnitude of the applied forward-bias voltage. A difference in Fermi levels in both sides is created. Since there are filled states in the conduction band of the n type at the same energy level as allowed empty states in the valence band of the p type, the electrons tunnel through the barrier from the n type to the p type, giving rise to a forward tunneling current

from the p type to the n type as shown in sector (1) of Fig. 5-3-2(a). As the forward bias is increased to V_p, the picture of the energy band is as shown in part (2) of Fig. 5-3-1(b). A maximum number of electrons can tunnel through the barrier from the filled states in the n type to the empty states in the p type, giving rise to the peak current I_p in Fig. 5-3-2(a). If the bias voltage is further increased, the condition shown in part (3) of Fig. 5-3-1(b) is reached. The tunneling current decreases as shown in sector (3) of Fig. 5-3-2(a). Finally, at a very large bias voltage, the band structure of part (4) of Fig. 5-3-1(b) is obtained. Since there are now no allowed empty states in the p type at the same energy level as filled states in the n type, no electrons can tunnel through the barrier and the tunneling current drops to zero as shown at point (4) of Fig. 5-3-2(a).

When the forward-bias voltage V is increased above the valley voltage V_v, the ordinary injection current I at the p-n junction starts to flow. This injection current is increased exponentially with the forward voltage as indicated by the dashed curve of Fig. 5-3-2(a). The total current, given by the sum of the tunneling current and the injection current, results in the volt-ampere characteristic of the tunnel diode as shown in Fig. 5-3-2(b). It can be seen from the figure that the total current reaches a minimum value I_v (or valley current) somewhere in the region where the tunnel-diode characteristic meets the ordinary p-n diode characteristic. The ratio of peak current to valley current (I_p/I_v) can theoretically reach 50 to 100. In practice, however, this ratio is about 15.

5-3-2 Microwave Characteristics

The tunnel diode is useful in microwave oscillators and amplifiers because the diode exhibits a negative-resistance characteristic in the region between peak current I_p and valley current I_v. The I-V characteristic of a tunnel diode with the load line is shown in Fig. 5-3-3.

Here the *abc* load line intersects the characteristic curve in three points. Points a and c are stable points, and point b is unstable. If the voltage and current vary about b, the final values of I and V would be given by point a or c, but not by b. Since the tunnel diode has two stable states for this load line, the circuit is called *bistable,* and it can be utilized as a binary device in switching circuits. However, mi-
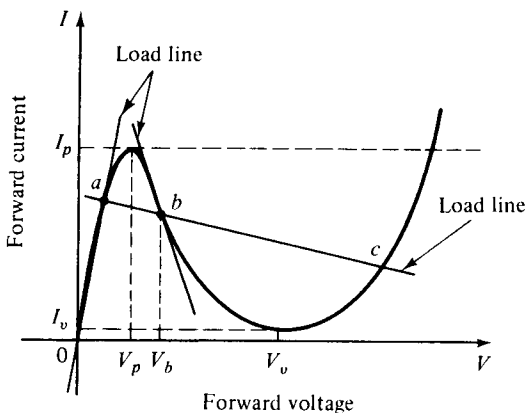
Figure 5-3-3 *I-V* characteristic of tunnel diode with load line.

crowave oscillation or amplification generated by the tunnel diode is our major concern in this section. The second load line intersects the *I-V* curve at point *b* only. This point is stable and shows a dynamic negative conductance that enables the tunnel diode to function as a microwave amplifier or oscillator. The circuit with a load line crossing point *b* in the negative-resistance region is called *astable*. Another load line crossing point *a* in the positive-resistance region indicates a *monostable* circuit. The negative conductance in Fig. 5-3-3 is given by

$$-g = \frac{\partial i}{\partial v}\bigg|_{v_b} = \frac{1}{-R_n} \tag{5-3-1}$$

where R_n is the magnitude of negative resistance.

For a small variation of the forward voltage about V_b, the negative resistance is constant and the diode circuit behavior is stable. A small-signal equivalent circuit for the tunnel diode operated in the negative-resistance region is shown in Fig. 5-3-4. Here R_s and L_s denote the inductance and resistance of the packaging circuit of a tunnel diode. The junction capacitance C of the diode is usually measured at the valley point; R_n is the negative resistance of the diode. Typical values of these parameters for a tunnel diode having a peak current I_p of 10 mA are

$$-R_n = -30 \ \Omega \qquad R_s = 1 \ \Omega \qquad L_s = 5 \ \text{nH} \qquad C = 20 \ \text{pF}$$
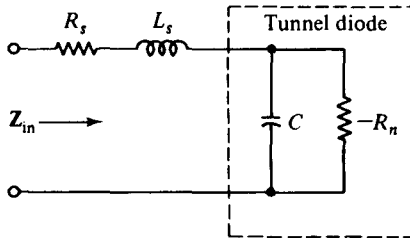
Figure 5-3-4   Equivalent circuit of tunnel diode.

The input impedance $\mathbf{Z}_{\text{in}}$ of the equivalent circuit shown in Fig. 5-3-4 is given by

$$\mathbf{Z}_{\text{in}} = R_s + j\omega L_s + \frac{R_n[\,j/(\omega C)]}{-R_n - j/(\omega C)}$$

$$\mathbf{Z}_{\text{in}} = R_s - \frac{R_n}{1 + (\omega R_n C)^2} + j\left[\omega L_s - \frac{\omega R_n^2 C}{1 + (\omega R_n C)^2}\right] \tag{5-3-2}$$

For the resistive cutoff frequency, the real part of the input impedance $\mathbf{Z}_{\text{in}}$ must be zero. Consequently, from Eq. (5-3-2) the resistive cutoff frequency is given by

$$f_c = \frac{1}{2\pi R_n C} \sqrt{\frac{R_n}{R_s} - 1} \tag{5-3-3}$$

For the self-resonance frequency, the imaginary part of the input impedance must be zero. Thus,

$$f_r = \frac{1}{2\pi R_n C} \sqrt{\frac{R_n^2 C}{L_s} - 1} \tag{5-3-4}$$

The tunnel diode can be connected either in parallel or in series with a resistive load as an amplifier; its equivalent circuits are shown in Fig. 5-3-5.

(a) Parallel loading   (b) Series loading
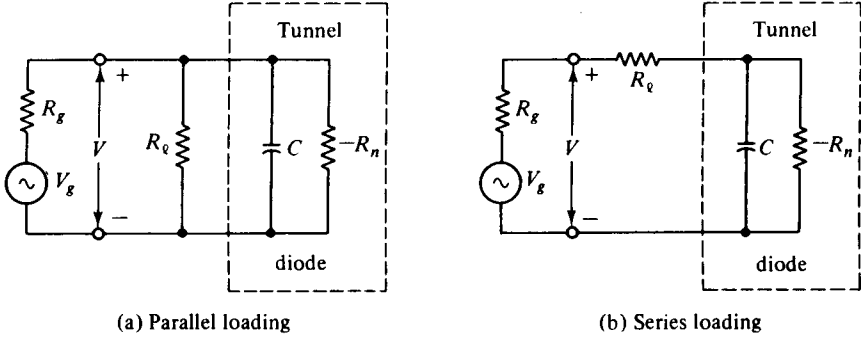
Figure 5-3-5 Equivalent circuits of tunnel diodes.

Parallel loading.   It can be seen from Fig. 5-3-5(a) that the output power in the load resistance is given by

$$P_{\text{out}} = \frac{V^2}{R_\ell} \tag{5-3-5}$$

One part of this output power is generated by the small input power through the tunnel diode amplifier with a gain of A, and this part can be written

$$P_{\text{in}} = \frac{V^2}{AR_\ell} \tag{5-3-6}$$

Another part of the output power is generated by the negative resistance, and it is expressed as

$$P_n = \frac{V^2}{R_n} \tag{5-3-7}$$

Therefore

$$\frac{V^2}{AR_\ell} + \frac{V^2}{R_n} = \frac{V^2}{R_\ell} \tag{5-3-8}$$

and the gain equation of a tunnel diode amplifier is given by

$$A = \frac{R_n}{R_n - R_\ell} \tag{5-3-9}$$

When the negative resistance R_n of the tunnel diode approaches the load resistance R_ℓ, the gain A approaches infinity and the system goes into oscillation.

Series loading.   In the series circuit shown in Fig. 5-3-5(b) the power gain A is given by

$$A = \frac{R_\ell}{R_\ell - R_n} = \frac{1}{1 - R_n/R_\ell} \tag{5-3-10}$$

The device remains stable in the negative-resistance region without switching if $R_\ell < R_n$.

A tunnel diode can be connected to a microwave circulator to make a negative-resistance amplifier as shown in Fig. 5-3-6. A microwave circulator is a multiport junction in which the power may flow only from port 1 to port 2, port 2 to port 3, and so on in the direction shown. Although the number of ports is not restricted, microwave circulators with four ports are most commonly used. If the circulator is perfect and has a positive real characteristic impedance R_0, an amplifier with infinite gain can be built by selecting a negative-resistance tunnel diode whose input impedance has a real part equal to $-R_0$ and an imaginary part equal to zero. The reflection coefficient from Fig. 5-3-6 is infinite. In general, the reflection coefficient is given by

$$\Gamma = \frac{-R_n - R_0}{-R_n + R_0} \tag{5-3-11}$$
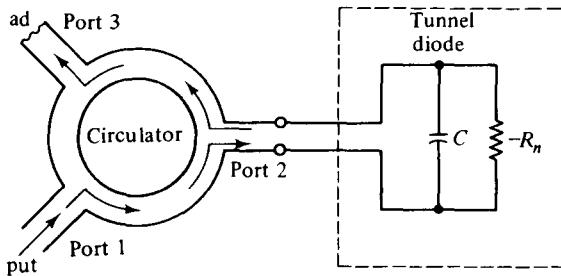
Figure 5-3-6   Tunnel diode connected to circulator.

REFERENCES

[1] SOBOL, H., and F. STERZER, Solid-state microwave power sources. *IEEE Spectrum*, **4**, 32, April 1972.

[2] NAVON, D. H., *Semiconductor Microdevices and Materials*. Holt, Rinehart and Winston, New York, 1986.

[3] SZE, S. M., *Semiconductor Devices: Physics and Technology*, Fig. 6 in Chapter 4. John Wiley & Sons, New York, 1985.

[4] EARLY, J. M., Maximum rapidly switchable power density in junction triodes. *IRE Trans. on Electron Devices*, **ED-6**, 322–335 (1959).

[5] JOHNSON, E. O., Physical limitations on frequency and power parameters of transistors. *RCA Rev.*, **26**, No. 6, 163–177, June 1965.

[6] DELOACH, B. C., JR., Recent advances in solid state microwave generators. *Advances in Microwaves*, Vol. 2. Academic Press, New York, 1967.

[7] ANDERSON, R. L., Experiments on Ge-GaAs heterojunction. *Solid State Electronics*, 5, p. 341. Pergamon Press, London, 1962.

SUGGESTED READINGS

IEEE Proceedings. **70,** No. 1, January 1982. Special issue on very fast solid-state technology.

LIAO, S. Y., *Semiconductor Electronic Devices*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1990.

NAVON, D. H., *Semiconductor Microdevices and Materials*. Holt, Rinehart and Winston, New York, 1986.

SZE, S. M., *Semiconductor Devices: Physics and Technology*. John Wiley & Sons, New York, 1985.

PROBLEMS

Microwave Bipolar Transistors

5-1. A GaAs n-p-n transistor at 300° K has the following parameters:

Base width	$W = 2 \times 10^{-5}$ cm
Diffusion length in emitter	$L_E = 2 \times 10^{-4}$ cm
Diffusion length in collector	$L_C = 4 \times 10^{-4}$ cm
Base resistivity	$\rho_B = 20$ Ω-cm
Emitter resistivity	$\rho_E = 1$ Ω-cm
Collector resistivity	$\rho_C = 10$ Ω-cm
Emitter junction voltage	$V_E = 0.4$ V
Collector junction voltage	$V_C = 0.5$ V
Cross-section area	$A = 0.01$ cm^2

Find: **a.** the impurity densities in the emitter, base and collector
     b. the mobilities in the emitter, base, and collector
     c. the diffusion lengths in the emitter, base, and collector
     d. the equilibrium densities in the emitter, base, and collector
     e. the coupling coefficients
     f. the emitter and collector currents

5-2. Derive Eq. (5-1-30) from Eq. (5-1-27).

5-3. Derive Eq. (5-1-31) from Eq. (5-1-30).

5-4. Verify Eq. (5-1-34).

5-5. Derive Eqs. (5-1-39), (5-1-40), and (5-1-41).

5-6. A certain silicon bipolar transistor has a maximum electric field intensity E_m of 3×10^5 volts/cm and its carrier has a saturated drift velocity v_s of 4×10^6 cm/sec. The emitter-collector length L is 4 microns.
    a. Calculate the maximum allowable applied voltage.
    b. Compute the transit time for a charge to transverse the emitter-collector length L.
    c. Determine the maximum possible transit frequency.

5-7. A bipolar transistor has voltage-frequency and current-frequency limitations as shown in Eqs. (5-1-58) and (5-1-59). Derive the power-frequency limitation as shown in Eq. (5-1-60).

5-8. A bipolar transistor has a voltage-current-power limitation. Derive the power gain-frequency relationship as shown in Eq. (5-1-61).

Heterojunction Bipolar Transistors

5-9. An n-Ge–p-GaAs–n-GaAs heterojunction transistor at 300° K has the following parameters:

Donor density in the n-Ge region:	$N_d = 2 \times 10^{19}$ cm^{-3}
Acceptor density in p-GaAs region:	$N_a = 3 \times 10^{17}$ cm^{-3}
Hole lifetime:	$\tau_p = 4 \times 10^{-6}$ sec
Bias voltage at emitter junction:	$V_E = 1.50$ V
Cross section:	$A = 4 \times 10^{-2}$ cm^2

Compute: **a.** The built-in voltage in the p-GaAs side
   b. The hole mobility (read the value from Fig. 2-4-2)
   c. The hole diffusion constant
   d. The minority hole density in the n-Ge region
   e. The minority electron density in the p-GaAs region
   f. The hole diffusion length
   g. The emitter-junction current

5-10. An n-Ge–p-GaAs–n-GaAs heterojunction transistor at 300° K has the following parameters:

Donor density in the n-Ge region:	$N_d = 5 \times 10^{18}$ cm^{-3}
Acceptor density in the p-GaAs region:	$N_a = 6 \times 10^{16}$ em^{-3}
Hole lifetime:	$\tau_p = 4 \times 10^{-6}$ sec
Bias voltage across n-p junction:	$V_E = 0.80$ V
Cross section:	$A = 2 \times 10^{-2}$ cm^2

Calculate: **a.** The built-in voltage in the p-GaAs side
   b. The hole mobility (read from Fig. A-2 in Appendix A)
   c. The hole diffusion constant
   d. The minority hole density in the n-Ge region
   e. The minority electron density in the p-GaAs region
   f. The hole diffusion length
   g. The emitter-junction current

Describe the electronic applications of an HBT.

Tunnel Diodes

5-11. A negative-resistance device is connected through a 1-kΩ resistor in series and a 0.01-μF capacitor in parallel to a combination of a supply voltage V of 10 V and a sig-

(a)

(b)
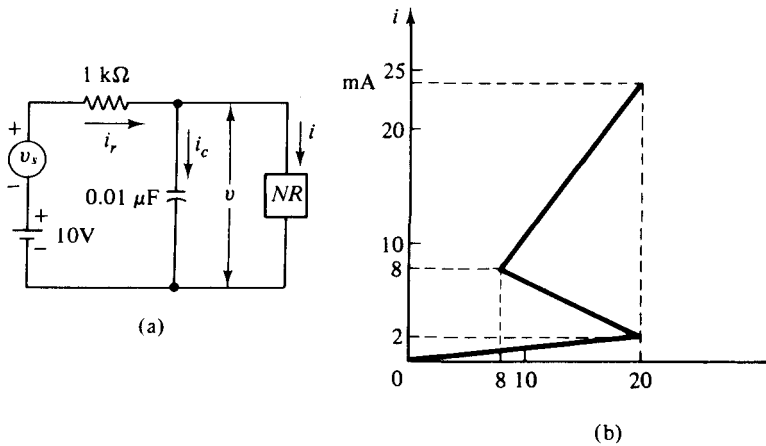
Figure P5-11

nal source V_s as shown in Fig. P5-11(a). Its I-V characteristic curve is shown in Fig. P5-11(b).

a. Find the negative resistance and forward resistances of the device.

b. Draw the load line on the I-V curve.

c. Determine the quiescent operating point of the circuit by the values of voltage and current.

d. Determine the new operating point by the values of voltage and current when a signal voltage of 14 V is applied to the circuit.

e. Draw the new load line on the I-V curve.

f. Find the time constant of the circuit.

g. Compute v, i, i_c, and i_r as a function of time after the triggering signal is applied and before the transition takes place.

h. Find the transition time T in microseconds.

i. Calculate v, i, i_c, and i_r immediately after transition.

5-12. A microwave tunnel diode has a negative resistance R_n and the resonant circuit has a circuit resistance R_ℓ. Derive an equation for the gain of a microwave tunnel-diode amplifier.

5-13. A certain microwave tunnel diode has a negative resistance of $69 + j9.7\ \Omega$. Determine the resonant-circuit impedance so that the microwave tunnel-diode amplifier will produce a power gain of 15 dB.

Chapter 6

Microwave Field-Effect Transistors

6-0 INTRODUCTION

After Shockley and his coworkers invented the transistor in 1948, he proposed in 1952 a new type of field-effect transistor (FET) in which the conductivity of a layer of a semiconductor is modulated by a transverse electric field [1]. In a conventional transistor both the majority and the minority carriers are involved; hence this type of transistor is customarily referred to as a *bipolar* transistor. In a field-effect transistor the current flow is carried by majority carriers only; this type is referred to as a *unipolar* transistor. In addition, the field-effect transistors are controlled by a voltage at the third terminal rather than by a current as with bipolar transistors. Our purpose here is to describe the physical structures, principles of operation, microwave characteristics, and power-frequency limitations of unipolar field-effect transistors. Since the microwave field-effect transistor has the capability of amplifying small signals up to the frequency range of X band with low-noise figures, it has lately replaced the parametric amplifier in airborne radar systems because the latter is complicated in fabrication and expensive in production.

The unipolar field-effect transistor has several advantages over the bipolar junction transistor:

1. It may have voltage gain in addition to current gain.
2. Its efficiency is higher than that of a bipolar transistor.
3. Its noise figure is low.
4. Its operating frequency is up to X band.
5. Its input resistance is very high, up to several megohms.

Among the unipolar field-effect transistors are the junction field-effect transistors, metal-semiconductor field-effect transistors, high electron-mobility transistors, and the metal-oxide-semiconductor field-effect transistors. All these devices are analyzed in this chapter.

6-1 JUNCTION FIELD-EFFECT TRANSISTORS (JFETs)

Unipolar field-effect transistors may be in the form of either a *p-n* junction gate or a Schottky-barrier gate. The former is called a *junction field-effect transistor* (JFET), and the latter is referred to as a MESFET or *metal-semiconductor field-effect transistor*. The JFET was originally proposed by Shockley [1]. Figure 6-1-1 shows the schematic diagram and circuit symbol for an *n*-channel JFET.

6-1-1 Physical Structure

The *n*-type material is sandwiched between two highly doped layers of *p*-type material that is designated P^+. This type of device is called an *n-channel JFET*. If the middle part is a *p*-type semiconductor, the device is called a *p-channel JFET*. The two *p*-type regions in the *n*-channel JFET shown in Fig. 6-1-1 are referred to as the *gates*. Each end of the *n* channel is joined by a metallic contact. In accordance with the directions of the biasing voltages shown in Fig. 6-1-1, the left-hand contact which supplies the source of the flowing electrons is referred to as the *source*, whereas the right-hand contact which drains the electrons out of the material is called the *drain*. The circuit symbol for an *n*-channel JFET is also shown in Fig. 6-1-1. The direction of the drain current I_d is flowing from the drain to the device.
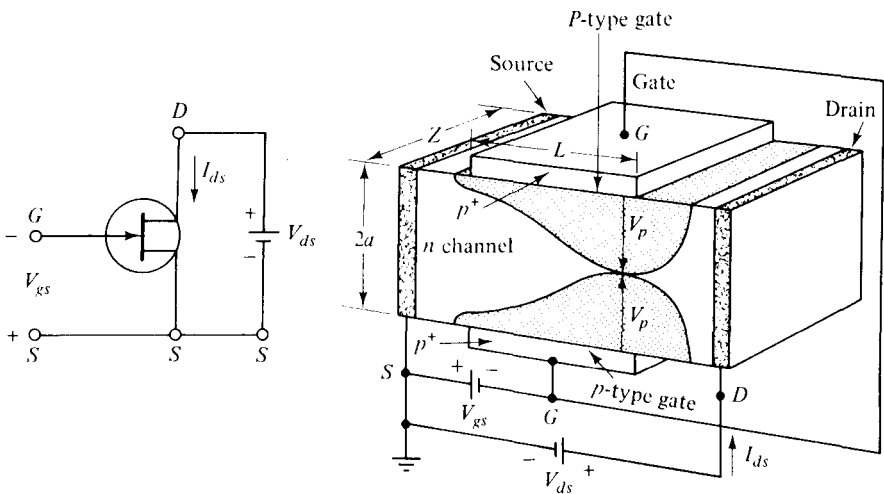
Figure 6-1-1   Schematic diagram and circuit symbol for an *n*-channel JFET.

For a p-channel JFET, the polarities of the two biasing voltages V_g and V_d are inter-changed, the head of the arrow points away from the device, and the drain current I_d flows away from the device. Since electrons have higher mobility than holes, the n-channel JFET provides higher conductivity and higher speed and is preferred in most applications to the p-channel JFET.

6-1-2 Principles of Operation

Under normal operating conditions, when the gate voltage V_g is zero, the drain current I_d is also zero. The channel between the gate junctions is entirely open. When a small drain voltage V_d is applied between the drain and source, the n-type semiconductor bar acts as a simple resistor, and the current I_d increases linearly with V_d. If a reverse gate voltage V_g is applied across the p-n gate junctions, the majority of free electrons are depleted from the channel, and the space-charge regions are extended into the channel. As the drain voltage V_d is further increased, the space-charge regions expand and join together, so that all free electron carriers are completely depleted in the joined region. This condition is called *pinch off*. When the channel is pinched off, the drain current I_d remains almost constant, but the drain voltage V_d is continuously increased.

Pinch-off voltage Vp.   The pinch-off voltage is the gate reverse voltage that removes all the free charge from the channel. The Poisson's equation for the voltage in the n channel, in terms of the volume charge density is given by

$$\frac{d^2 V}{dy^2} = -\frac{\rho}{\epsilon_s} = -\frac{qN_d}{\epsilon_s} = -\frac{qN_d}{\epsilon_r \epsilon_o} \qquad (6\text{-}1\text{-}1)$$

where ρ = volume charge density in coulombs per cubic meter
q = charge in coulombs
N_d = electron concentration in electrons per cubic meter
ϵ_s = permittivity of the material in farads per meter
$\epsilon_s = \epsilon_r \epsilon_o$, ϵ_r is the relative dielectric constant
$\epsilon_o = 8.854 \times 10^{-12}$ F/m is the permittivity of free space

Integration of Eq. (6-1-1) once and application of the boundary condition of the electric field $E = -\dfrac{dV}{dy} = 0$ at $y = a$ yield

$$\frac{dV}{dy} = -\frac{qN_d}{\epsilon_s}(y - a) \qquad \text{volts per meter} \qquad (6\text{-}1\text{-}2)$$

Integration of Eq. (6-1-2) once and application of the boundary condition $V = 0$ at $y = 0$ result in

$$V = -\frac{qN_d}{2\epsilon_s}(y^2 - 2ay) \qquad \text{volts} \qquad (6\text{-}1\text{-}3)$$

Then the pinch-off voltage V_p at $y = a$ is expressed as

$$V_p = \frac{qN_d a^2}{2\epsilon_s} \qquad \text{volts} \qquad (6\text{-}1\text{-}4)$$

where a is the height of the channel in meters.

Equation (6-1-4) indicates that the pinch-off voltage is a function of the doping concentration N_d and the channel height a. Doping may be increased to the limit set by the gate breakdown voltage, and the pinch-off voltage may be made large enough so that drift-saturation effects just become dominant.

From Fig. 6-1-1, the pinch-off voltage under saturation condition can be expressed as

$$V_p = V_d + |V_g| + \psi_o \qquad (6\text{-}1\text{-}5)$$

where $|V_g|$ = absolute value of the gate voltage
ψ_o = built-in or barrier voltage at the junction

The saturation drain voltage is then given by

$$V_{d\,\text{sat}} = \frac{qN_d a^2}{2\epsilon_s} - |V_g| - \psi_o \qquad (6\text{-}1\text{-}6)$$

The JFET has a conducting channel between the source and the drain electrodes when the gate bias voltage is zero. This is the ON state, and the transistor is called a *normally ON JFET*. In order to reach the OFF state, a gate voltage must be applied to deplete all carriers in the channel. As a result, this device is referred to as the *depletion-mode JFET* or *D-JFET*.

Example 6-1-1:   Pinch-Off Voltage of a Silicon JFET

A certain Si JFET has the following parameters:

Channel height: $a = 0.1 \, \mu\text{m}$
Electron concentration: $N_d = 8 \times 10^{17} \, \text{cm}^{-3}$
Relative dielectric constant: $\epsilon_r = 11.80$

Calculate the pinch-off voltage.

Solution   From Eq. (6-1-6), the pinch-off voltage is

$$V_p = \frac{qN_d a^2}{2\epsilon_s} = \frac{1.6 \times 10^{-19} \times 8 \times 10^{23} \times (0.1 \times 10^{-6})^2}{2 \times 8.854 \times 10^{-12} \times 11.8}$$

$$= 6.66 \text{ volts}$$

6-1-3 Current-Voltage (I-V) Characteristics

The drain current of an n-channel JFET is dependent on the drain and gate voltages and the channel resistance. The n-channel resistance can be expressed as

$$R = \frac{\rho L}{A} = \frac{L}{\sigma A} = \frac{L}{q\mu_n N_d A} = \frac{L}{2q\mu_n N_d Z(a - W)} \qquad (6\text{-}1\text{-}7)$$

where μ_n = electron mobility
$\quad\quad Z$ = distance in z direction
$\quad\quad L$ = length in x direction
$\quad\quad W$ = depletion-layer width
$\quad\quad a$ = width between two p-n junctions

The drain voltage across an elemental section dx of the channel is given by

$$dV(x) = I_d\,dR = \frac{I_d\,dx}{2q\mu_n N_d Z[a - W(x)]} \tag{6-1-8}$$

From Eq. (6-1-6), the depletion-layer width can be written as

$$W(x) = \left[\frac{2\epsilon_s[V(x) + |V_g| + \psi_o]}{qN_d}\right]^{1/2} \tag{6-1-9}$$

Then

$$W\,dw = \frac{\epsilon_s}{qN_d}\,dV \tag{6-1-10}$$

The drain current I_d is expressed by

$$I_d = 2q\mu_n N_d Z[a - W(x)]\frac{dV}{dx} \tag{6-1-11}$$

Substituting Eq. (6-1-10) into Eq. (6-1-11) and integrating from $x = 0$ to $x = L$ give the drain current as

$$\begin{aligned} I_d &= \frac{2q^2\mu_n N_d^2 Z}{L\epsilon_s}\int_{w_1}^{w_2} (a - W)W\,dw \\ &= \frac{\mu_n q^2 N_d^2 Z}{L\epsilon_s}\left[a(W_2^2 - W_1^2) - \frac{2}{3}(W_2^3 - W_1^3)\right] \end{aligned} \tag{6-1-12}$$

Boundary conditions.   The channel-to-gate voltage can be written from Eq. (6-1-4) as

$$V_{cg} = \frac{qN_d W^2}{2\epsilon_s} = \frac{qN_d a^2 W^2}{2\epsilon_s a^2} = V_p\frac{W^2}{a^2} = V(x) + |V_g| + \psi_o \tag{6-1-13}$$

The depletion-layer width is

$$W = a\left(\frac{V(x) + |V_g| + \psi_o}{V_p}\right)^{1/2} \tag{6-1-14}$$

Then

$$W_1^2 = a^2\left(\frac{|V_g| + \psi_o}{V_p}\right) \qquad \text{at } x = 0 \text{ and } V(x) = 0 \tag{6-1-15}$$

and

$$W_2^2 = a^2\left(\frac{V_d + |V_g| + \psi_o}{V_p}\right) \qquad \text{at } x = L \text{ and } V(x) = V_d \tag{6-1-16}$$

Substituting Eqs. (6-1-15) and (6-1-16) into Eq. (6-1-12) yields the drain current

$$I_d = I_p \left[\frac{V_d}{V_p} - \frac{2}{3}\left(\frac{V_d + |V_g| + \psi_o}{V_p} \right)^{3/2} + \frac{2}{3}\left(\frac{|V_g| + \psi_o}{V_p} \right)^{3/2} \right] \qquad (6\text{-}1\text{-}17)$$

where $I_p = \dfrac{\mu_n q^2 N_d^2 Z a^3}{L \epsilon_s}$ is the pinch-off current.

Figure 6-1-2 shows the normalized ideal current-voltage characteristics for a pinch-off voltage of 3.2 volts [1].
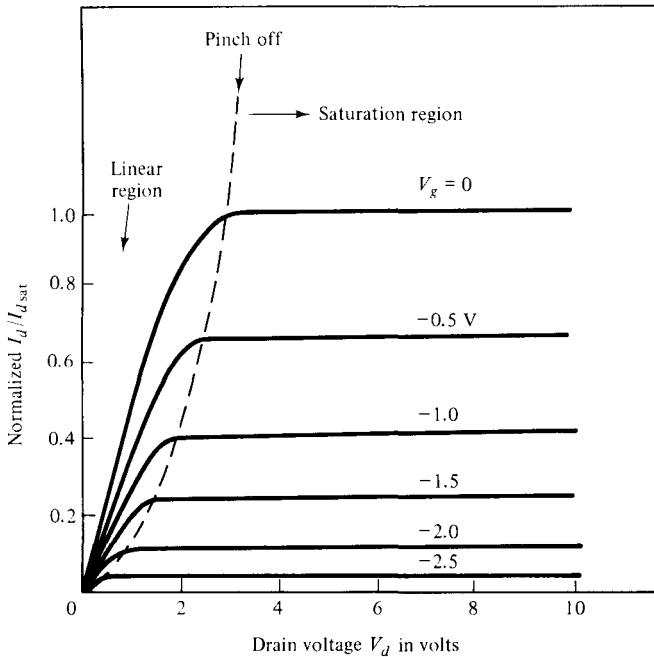
Figure 6-1-2   Normalized ideal I-V characteristics of a typical JFET.

Linear region.   In the linear region at $V_d \ll (|V_g| + \psi_o)$, the drain current can be expressed from Eq. (6-1-12) as

$$I_d = \frac{I_p V_d}{V_p} \left[1 - \left(\frac{|V_g| + \psi_o}{V_p} \right)^{1/2} \right] \qquad (6\text{-}1\text{-}18)$$

The channel conductance (or the drain conductance) is given by

$$g_d = \frac{\partial I_d}{\partial V_d} \bigg|_{V_g = \text{constant}} = \frac{I_p}{V_p} \left[1 - \left(\frac{|V_g| + \psi_o}{V_p} \right)^{1/2} \right] \qquad (6\text{-}1\text{-}19)$$

The mutual conductance (or transconductance) is expressed as

$$g_m = \frac{\partial I_d}{\partial V_g} \bigg|_{V_d = \text{constant}} = \frac{I_p V_d}{2 V_p^2} \left[1 - \left(\frac{V_p}{|V_g| + \psi_o} \right)^{1/2} \right] \qquad (6\text{-}1\text{-}20)$$

Saturation region.   At the pinch-off voltage, the drain current becomes saturated. By setting $V_p = V_d + |V_g| + \psi_o$, the saturation drain current is given by

$$I_{d\,\text{sat}} = I_p\left[\frac{1}{3} - \left(\frac{|V_g| + \psi_o}{V_p}\right) + \frac{2}{3}\left(\frac{|V_g| + \psi_o}{V_p}\right)^{3/2}\right] \qquad (6\text{-}1\text{-}21)$$

and the corresponding saturation drain voltage as

$$V_{d\,\text{sat}} = V_p + |V_g| + \psi_o \qquad (6\text{-}1\text{-}22)$$

The mutual conductance in the saturation region is then expressed as

$$g_m = \frac{\partial I_d}{\partial V_g}\bigg|_{V_d = \text{constant}} = \frac{I_p}{V_p}\left[1 - \left(\frac{|V_g| + \psi_o}{V_p}\right)^{1/2}\right] \qquad (6\text{-}1\text{-}23)$$

where $\dfrac{I_p}{V_p} = \dfrac{2\mu_n q N_d a Z}{L}$

It should be noted that the g_m in the saturation region is identical to the one in the linear region.

For a small signal, the drain or output resistance is defined as

$$r_d = \frac{\partial V_d}{\partial I_d}\bigg|_{V_g = \text{constant}} \qquad (6\text{-}1\text{-}24)$$

Then the amplification factor for a JFET may be defined as

$$\mu = \frac{\partial V_d}{\partial V_g}\bigg|_{I_d = \text{constant}} = r_d g_m \qquad (6\text{-}1\text{-}25)$$

The cutoff frequency is given by

$$f_c = \frac{g_m}{2\pi C_g} = \frac{I_p/V_p}{2\pi Z L \epsilon_s/(2a)} = \frac{2\mu_n q N_d a^2}{\pi \epsilon_s L^2} \qquad (6\text{-}1\text{-}26)$$

where $C_g = \dfrac{L Z \epsilon_s}{2a}$ is the capacitance between gate and source

Pinch-off region.   When an electric field appears along the x axis between the drain and the source, the drain end of the gate is more reverse-biased than the source end. Hence the boundaries of the depletion region are not parallel to the center of the channel, but converged as shown in Fig. 6-1-2. As the drain voltage V_d and drain current I_d are further increased, the channel is finally pinched off.

Breakdown voltage.   As the drain voltage V_d increases for a constant gate voltage V_g, the bias-voltage causes avalanche breakdown across the gate junction, and the drain current I_d increases sharply. The breakdown voltage is shown in Fig. 6-1-3 by the relationship

$$V_b = V_d + |V_g|$$
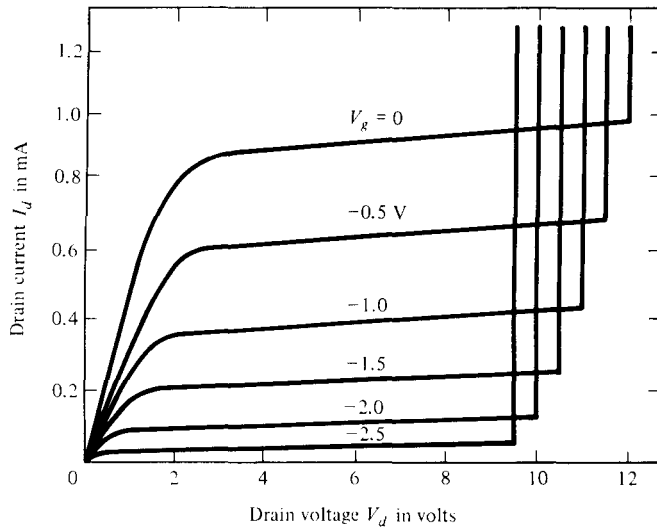
Figure 6-1-3   Breakdown voltage for a typical JFET.

Example 6-1-2:   Current of a JFET

A silicon JFET at 300° K has the following parameters:

Electron density:	$N_d = 1 \times 10^{17}$ cm^{-3}
Hole density:	$N_a = 1 \times 10^{19}$ cm^{-3}
Relative dielectric constant:	$\epsilon_r = 11.8$
Channel height:	$a = 0.2 \times 10^{-4}$ cm
Channel length:	$L = 8 \times 10^{-4}$ cm
Channel width:	$Z = 50 \times 10^{-4}$ cm
Electron mobility:	$\mu_n = 800$ cm^2/V·s
Drain voltage:	$V_d = 10$V
Gate voltage:	$V_g = -1.5$ V

Compute:

a. The pinch-off voltage in volts
b. The pinch-off current in mA
c. The built-in voltage in V
d. The drain current in mA
e. The saturation drain current at $V_g = 0$
f. The cutoff frequency

Solution

a. The pinch-off voltage is

$$V_p = \frac{1.6 \times 10^{-19} \times 1 \times 10^{17} \times (0.2 \times 10^{-4})^2}{2 \times 8.854 \times 10^{-14} \times 11.8} = 3.06 \text{ V}$$

b. The pinch-off current is

$$I_p = \frac{800 \times (1.6 \times 10^{-19})^2 (1 \times 10^{17})^2 (50 \times 10^{-4})(0.2 \times 10^{-4})^3}{8 \times 10^{-4} \times 8.854 \times 10^{-14} \times 11.8}$$

$$= 9.80 \text{ mA}$$

c. The built-in voltage is

$$\psi_o = 26 \times 10^{-3} \ln\left[\frac{1 \times 10^{17} \times 1 \times 10^{19}}{(1.5 \times 10^{10})^2}\right] = 0.936 \text{ V}$$

d. The drain current is computed from Eq. (6-1-17) as

$$I_d = 9.8 \times 10^{-3}\left[\frac{10}{3.06} - \frac{2}{3}\left(\frac{10 + 1.5 + 0.936}{3.06}\right)^{3/2} + \frac{2}{3}\left(\frac{1.5 + 0.936}{3.06}\right)^{3/2}\right]$$

$$= 6.80 \text{ mA}$$

e. The saturation drain current is computed from Eq. (6-1-21) to be

$$I_{d\,\text{sat}} = 6.8 \times 10^{-3}\left[\frac{1}{3} - \left(\frac{1.5 + 0.936}{3.06}\right) + \frac{2}{3}\left(\frac{1.5 + 0.936}{3.06}\right)^{3/2}\right]$$

$$= 0.088 \text{ mA}$$

f. The cutoff frequency is computed from Eq. (6-1-26) as

$$f_c = \frac{2 \times 800 \times 1.6 \times 10^{-19} \times 1 \times 10^{17} \times (0.2 \times 10^{-4})^2}{3.1416 \times 8.854 \times 10^{-14} \times 11.8 \times (8 \times 10^{-4})^2}$$

$$= 4.8 \text{ GHz}$$

The operating frequency must be less than 4.8 GHz.

6-2 METAL-SEMICONDUCTOR FIELD-EFFECT TRANSISTORS (MESFETs)

In 1938, Schottky suggested that the potential barrier could arise from stable space charges in the semiconductor alone without the presence of a chemical layer [2]. The model derived from his theory is known as the *Schottky barrier*.

If the field-effect transistor is constructed with a metal-semiconductor Schottky-barrier diode, the device is called a *metal-semiconductor field-effect transistor* (MESFET). The material may be either silicon or gallium arsenide (GaAs), and the channel type may be either n channel or p channel.

Since GaAs MEFSETs have the capability of amplifying small signals up to the frequence range of X band with low-noise figure, they have lately replaced the parametric amplifiers in airborne radar systems because the latters are complicated to fabricate and expensive to produce.

The GaAs MESFET has higher electron mobility, higher electric field, and higher electron saturation drift velocity than silicon devices, so its output power is also greater. Another special feature is its lower noise figure, accounted for by its

higher electron mobility. Therefore the GaAs MESFETs are very commonly used in microwave integrated circuits for high-power, low-noise, and broadband amplifier applications.

6-2-1 Physical Structures

The MESFET was developed by many scientists and engineers, such as Mead [3] and Hooper [4], and it is sometimes also called the *Schottky-barrier field-effect transistor*.

The unipolar transistor such as a GaAs MESFET can be developed by using either the epitaxial process or the ion implantation method and those methods are discussed in Chapter 12. Figure 6-2-1 shows schematically a simple MESFET in GaAs.
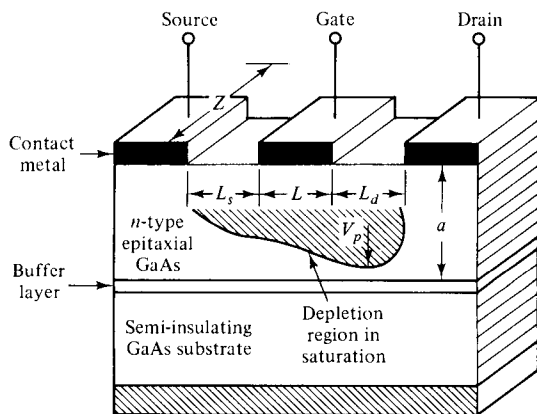
Figure 6-2-1   Schematic diagram of a GaAs MESFET.

In GaAs MESFETs the substrate is doped with chromium (Cr), which has an energy level near the center of the GaAs bandgap. As Cr is the dominant impurity, the Fermi level is pinned near the center of the bandgap. Thus, a very high-resistivity substrate (near 10^8 ohm-cm) generally results, and it is commonly called the *semi-insulator GaAs substrate*. On this nonconducting substrate a thin layer of lightly doped n-type GaAs is grown epitaxially to form the channel region of the field-effect transistor. In many cases a high resistivity GaAs epitaxial layer, called the *buffer layer*, is grown between the n-type GaAs layer and the substrate. The photolithographic process may be used to define the patterns in the metal layers such as Au–Ge for source and drain ohmic contacts and in the Al layer for the Schottky barrier-gate contact. The reason for using GaAs instead of Si is that GaAs has higher electron mobility and can operate at higher temperature and higher power.

The GaAs MESFET can also be grown by using ion implantation. A thin n-type layer can be formed at the surface of the substrate by implanting Si or a donor impurity Se from column VI of the element periodic table. However, the ion implantation process requires an anneal to remove the radiation damage. In either the fully implanted device or the epitaxial device, the source and drain contacts may be improved by further n^+ implantation in these regions.

After the fabrication processes are completed, the individual transistor is separated from the wafer, and this discrete transistor is called a *chip device*. The chip device is then alloyed to a header to provide a contact to the collector region, and Au or Al wires are bonded to the metallized regions to serve as leads to the emitter and base. This bonded chip device is then named a *package transistor*. Figure 6-2-2 shows both the package and chip GaAs MESFET devices.
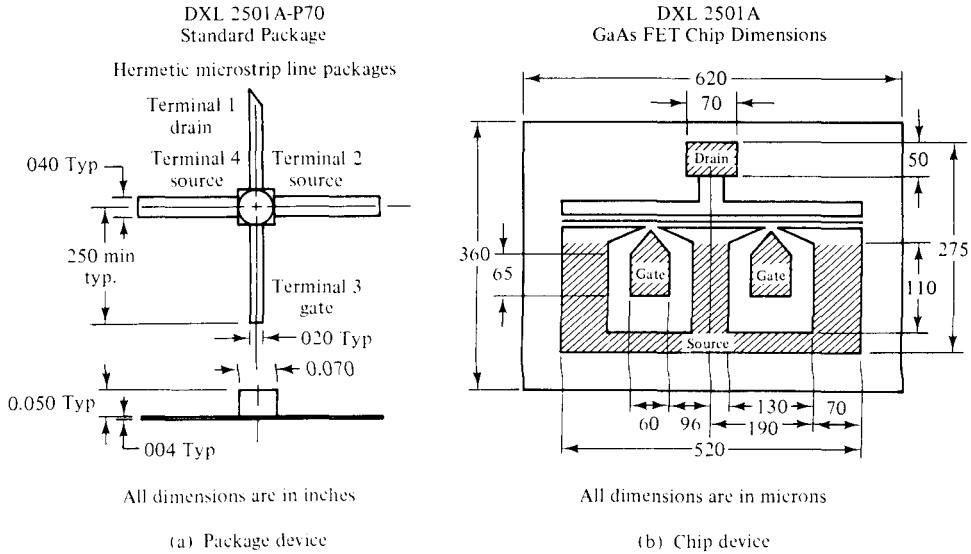
Figure 6-2-2   Dimensions of package and chip GaAs MESFETs.

The MESFET device is of an interdigitated structure, fabricated by using an *n*-type GaAs epitaxial film about 0.15- to 0.35-μm thick on a semi-insulating substrate about 100 μ. The *n*-channel layer is doped with either sulphur or tin in a doping concentration N between $8 \times 10^{16}/cm^3$ and $2 \times 10^{17}/cm^3$. The electron mobility in the layer is in the range of 3000 to 4500 $cm^2/V{\cdot}s$. The Schottky barrier-gate is evaporated aluminum. The source and drain contacts are Au–Ge, Au–Te, or Au–Te–Ge alloys. A contact metallization pattern of gold is used to bring the source, drain, and gate contacts out to bonding pads over the semi-insulating substrate. A buffer layer of about 3 μm with a doping concentration of 10^{15} to 10^{16} cm^{-3} is often fabricated between the active *n*-type epitaxial layer and the semi-insulating substrate.

6-2-2 Principles of Operation

In Fig. 6-2-3 and in Fig. 6-1-1 for JFET, a voltage is applied in the direction to reverse-bias the *p-n* junction between the source and the gate, while the source and the drain electrodes are forward-biased. Under this bias condition, the majority carriers (electrons) flow in the *n*-type epitaxial layer from the source electrode, through the channel beneath the gate, to the drain electrode. The current in the channel
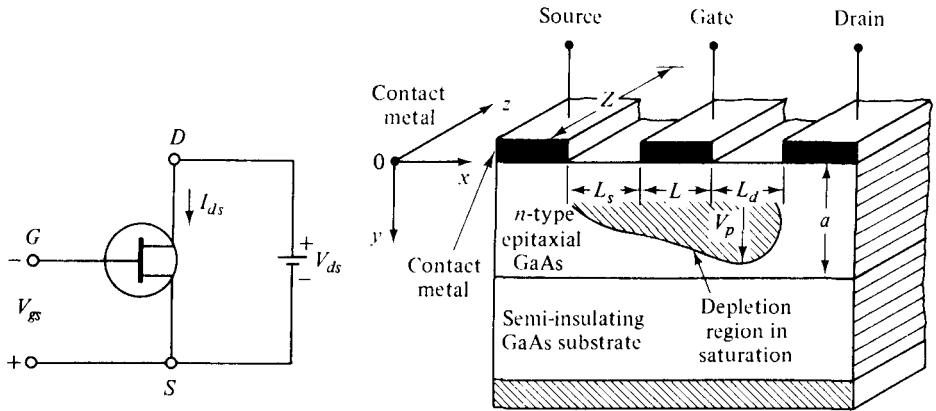
Figure 6-2-3   Schematic diagram and circuit symbol of a GaAs MESFET.

causes a voltage drop along its length so that the Schottky barrier-gate electrode becomes progressively more reverse-biased toward the drain electrode. As a result, a charge-depletion region is set up in the channel and gradually pinches off the channel against the semi-insulating substrate toward the drain end. As the reverse bias between the source and the gate increases, so does the height of the charge-depletion region. The decrease of the channel height in the nonpinched-off region will increase the channel resistance. Consequently, the drain current I_d will be modulated by the gate voltage V_g. This phenomenon is analogous to the characteristics of the collector current I_c versus the collector voltage V_c with the base current I_b as a parameter in a bipolar transistor. In other words, a family of curves of the drain current I_d versus the voltage V_{ds} between the source and drain with the gate voltage V_g as a parameter will be generated in an unipolar GaAs MESFET, as shown in Figure 6-2-4.

The transconductance of a field-effect transistor (FET) is expressed as

$$g_m = \frac{dI_d}{dV_g}\bigg|_{V_d=\text{constant}} \qquad \text{mhos} \qquad (6\text{-}2\text{-}1)$$

For a fixed drain-to-source voltage V_{ds}, the drain current I_d is a function of the reverse-biasing gate voltage V_g. Because the drain current I_d is controlled by the field effect of the gate voltage V_g, this device is referred to as the field-effect transistor. When the drain current I_d is continuously increasing, the ohmic voltage drop between the source and the channel reverse-biases the p-n junction further. As a result, the channel is eventually pinched off. When the channel is pinched off, the drain current I_d will remain almost constant even though the drain-to-source voltage V_d is continuously increased.

Pinch-off voltage V_p.   The pinch-off voltage is the gate reverse voltage that removes all the free charge from the channel. Poisson's equation for the voltage in the n channel, in terms of the volume charge density is given by

$$\frac{d^2 V}{dy^2} = -\frac{\rho}{\epsilon_s} = -\frac{qN_d}{\epsilon_s} = -\frac{qN_d}{\epsilon_r \epsilon_o} \qquad (6\text{-}2\text{-}2)$$
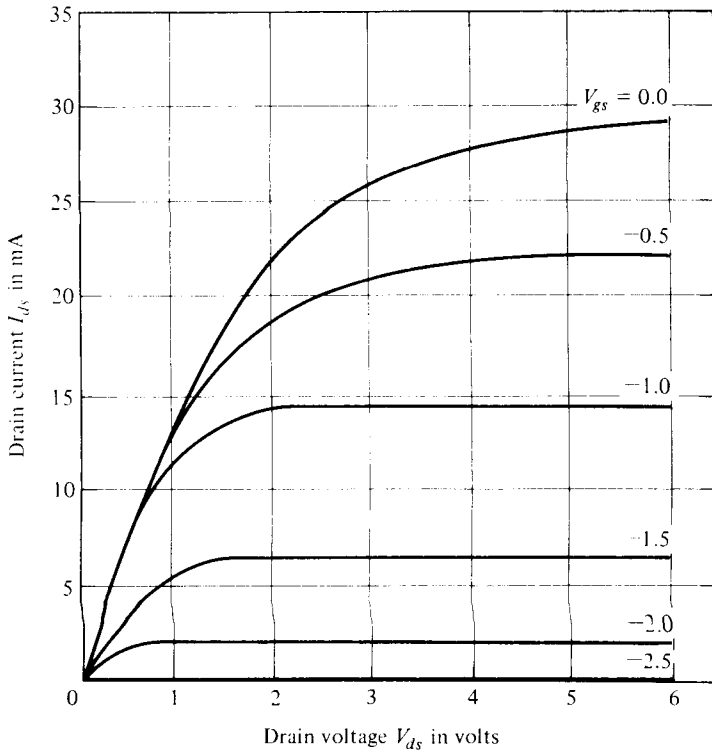
Figure 6-2-4   Current-voltage characteristics of a typical n-channel GaAs MESFET.

where ρ = volume charge density in coulombs per cubic meter
  q = charge in coulombs
  N_d = electron concentration in electrons per cubic meter
  ϵ_s = permittivity of the material in farads per meter
  $\epsilon_s = \epsilon_r \epsilon_o$, ϵ_r is the relative dielectric constant
  $\epsilon_o = 8.854 \times 10^{-12}$ F/m is the permittivity of free space

Integration of Eq. (6-2-2) twice and application of the boundary condition yield the pinch-off voltage at $y = a$ as

$$V_p = \frac{qN_d a^2}{2\epsilon_s} \qquad \text{volts} \qquad (6\text{-}2\text{-}3)$$

where a is the height of the channel in meters.

Equation (6-2-3) indicates that the pinch-off voltage is a function of the doping concentration N_d and the channel height a. Doping may be increased to the limit set by the gate breakdown voltage, and the pinch-off voltage may be made large enough so that drift saturation effects just become dominant.

Example 6-2-1:   **Pinch-Off Voltage of a MESFET**

A certain GaAs MESFET has the following parameters:

Channel height:	$a = 0.1 \ \mu\text{m}$
Electron concentration:	$N_d = 8 \times 10^{17} \ \text{cm}^{-3}$
Relative dielectric constant:	$\epsilon_r = 13.10$

Calculate the pinch-off voltage.

Solution   From Eq. (6-2-3) the pinch-off voltage is

$$V_p = \frac{qN_da^2}{2\epsilon_s} = \frac{1.6 \times 10^{-19} \times 8 \times 10^{23} \times (0.1 \times 10^{-6})^2}{2 \times 8.854 \times 10^{-12} \times 13.10}$$

$$= 6.00 \text{ volts}$$

6-2-3 Small-Signal Equivalent Circuit

For microwave frequencies, the metal-semiconductor field-effect transistor (MESFET) has a very short channel length, and its velocity saturation occurs in the channel before reaching the pinched path. The electronic characteristics of a MESFET depend not only on the intrinsic parameters, such as g_m, G_d, R_i, C_{gs}, and C_{gd}, but also on the extrinsic parameters R_g, R_s, C_{ds}, R_p and C_p. Figure 6-2-5 shows the cross section of a MESFET and its equivalent circuit.

The microwave properties of metal-semiconductor field-effect transistors were investigated and analyzed by many scientists and engineers. The noise behavior of these transistors at microwave frequencies has been investigated and measured by van der Ziel and others [5].

1. Intrinsic elements:

$$g_m = \text{transconductance of the MESFET}$$

$$G_d = \text{drain conductance}$$

$$R_i = \text{input resistance}$$

$$C_{gs} = \text{gate-source capacitance}$$

$$C_{gd} = \text{gate-drain (or feedback) capacitance}$$

2. Extrinsic elements:

$$R_g = \text{gate metallization resistance}$$

$$R_s = \text{source-gate resistance}$$

$$C_{ds} = \text{drain-source capacitance}$$

$$R_p = \text{gate bonding-pad parasitic resistance}$$
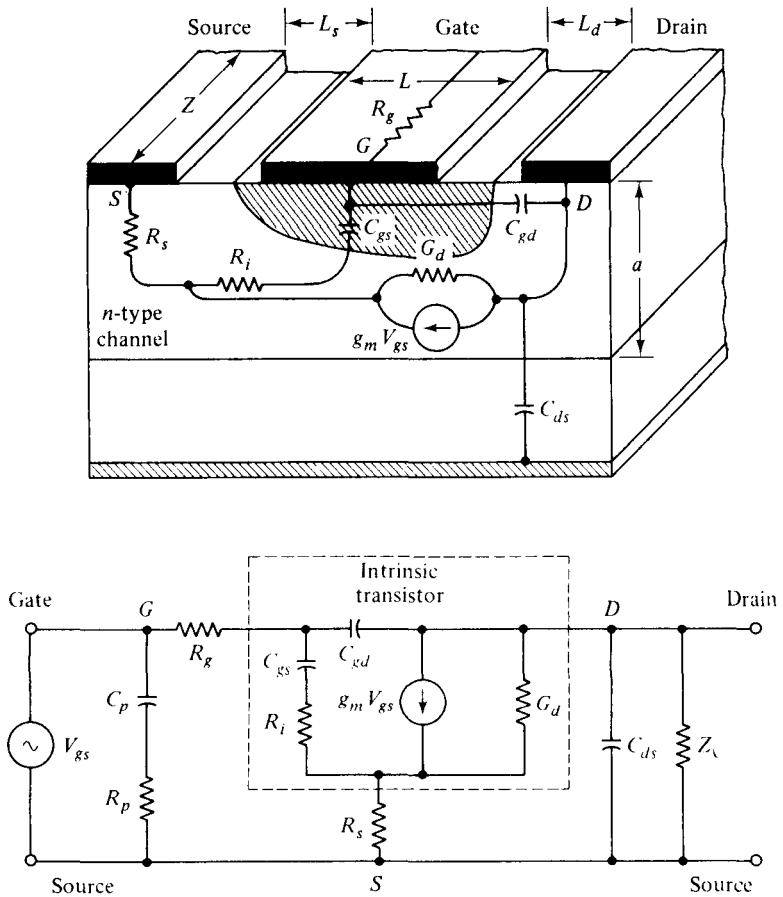
Figure 6-2-5   Cross section and equivalent circuit of a MESFET.

C_p = gate bonding-pad parasitic capacitance

Z_ℓ = load impedance

The values of these intrinsic and extrinsic elements depend on the channel type, material, structure, and dimensions of the Schottky barrier-gate FET. The large values of the extrinsic resistances will seriously decrease the power gain and efficiency and increase the noise figure of the MESFET. It is advantageous to increase the channel doping N as high as possible in order to decrease the relative influence of the feedback capacitance C_{gd} and to increase the transconductance g_m and the dc open-circuit voltage gain. However, an increase in concentration N decreases the breakdown voltage of the gate. A doping of 10^{18} per cubic centimeter might be an upper limit.

6-2-4 Drain Current I_d

The drain current I_d of a Schottky barrier-gate FET is expressed [6] as

$$I_d = I_p \frac{3(\mu^2 - \rho^2) - 2(\mu^3 - \rho^3)}{1 + \eta(\mu^2 - \rho^2)} \qquad \text{amperes} \qquad (6\text{-}2\text{-}4)$$

where $I_p = \dfrac{qN_d\mu aZV_p}{3L}$ is the saturation current for the Shockley case at $V_g = 0$

μ = low-field mobility in square meters per volt-second

$q = 1.6 \times 10^{-19}$ coulomb is the electron charge

N_d = doping concentration in electrons per cubic meter

a = channel height

Z = channel depth or width

L = gate length

V_p = pinch-off voltage as defined in Eq. (6-2-3)

$u = \left(\dfrac{V_d + |V_g|}{V_p}\right)^{1/2}$ is the normalized sum of the drain and gate voltages with respect to the pinch-off voltage

$\rho = \left(\dfrac{|V_g|}{V_p}\right)^{1/2}$ is the normalized gate voltage with respect to the pinch-off voltage

$\eta = \dfrac{\mu|V_p|}{v_s L} = \dfrac{v}{v_s}$ is the normalized drift velocity with respect to the saturation drift velocity

v_s = saturation drift velocity

$v = \dfrac{\mu E_x}{1 + \dfrac{\mu E_x}{v_s}}$ is the drift velocity in the channel

E_x = absolute value of the electric field in the channel

Figure 6-2-4 shows a plot of the drain current I_d versus the drain voltage V_{ds} with the gate voltage V_{gs} as a parameter for a typical n-channel GaAs MESFET. The drain current has a maximum at $u = u_m$ given by

$$u_m^3 + 3u_m\left(\frac{1}{z} - \rho^2\right) + 2\rho^3 - \frac{3}{z} = 0 \qquad (6\text{-}2\text{-}5)$$

where $z = \mu|V_p|/(v_x L)$ is the ratio between the small-field velocity extrapolated linearly to the field V_p/L and the saturation velocity

Substitution of Eq. (6-2-5) into Eq. (6-2-4) yields the saturation drain current as

$$I_{d(\text{sat})} = 3I_{dss}(1 - u_m)/z \qquad (6\text{-}2\text{-}6)$$

The transconductance in the saturation region is given by

$$g_m = \frac{\partial I_{d\,\text{sat}}}{\partial|V_g|}\bigg|_{V_d = \text{constant}} = \frac{g_{\max}(u_m - \rho)}{1 + z(u_m^2 - \rho^2)} \qquad (6\text{-}2\text{-}7)$$

In practice, the drain current and mutual conductance of a GaAs MESFET can be expressed by

$$I_{ds} = I_{dss}\left(1 + \frac{|V_g|}{V_p}\right)^2 \qquad (6\text{-}2\text{-}8)$$

and

$$g_m = \frac{2I_{dss}}{|V_p|}\left(1 + \frac{|V_g|}{V_p}\right) \qquad (6\text{-}2\text{-}9)$$

The velocity-field curves of a GaAs MESFET are very complicated. Figure 6-2-6 shows these curves in the saturation region [7]. The narrowest channel cross section is located under the drain end of the gate. The peak electric field appears near the drain. The drift velocity rises to a peak at x_1, close to the center of the channel, and falls to the low saturated value under the gate edge. To preserve current continuity, heavy electron accumulation has to form in this region because the channel cross section is narrowing. In addition, the electrons are moving progressively slower with increasing x. Exactly the opposite occurs between x_2 and x_3. The channel widens and the electrons move faster, causing a strong depletion layer. The charges in the accumulation and depletion layers are nearly equal, and most of the drain voltage drops in this stationary dipole-layer.

For a Si MESFET, the drain current is expressed as [7]

$$I_d = Zqn(x)v(x)d(x) \qquad (6\text{-}2\text{-}10)$$

where Z = channel depth or width
$\quad n(x) = N_d$ is the density of conduction electrons
$\quad v(x)$ = drift velocity
$\quad d(x)$ = conductive layer thickness
$\quad\quad x$ = coordinate in the direction of the electron drift

The current-voltage characteristics of a Si MESFET are shown in Fig. 6-2-7 [7]. In Fig. 6-2-7(a), there is no metal gate electrode. At the surface of the conducting layer, the source and drain contacts are made. When a positive voltage V_{ds} is applied to the drain, electrons will flow from source to drain. In Fig. 6-2-7(b), a metal gate is added and shorted to the source. When a small drain voltage is applied, a depletion layer is created. The current I_{ds} flowing from drain to source is indicated by Eq. 6-2-7. When the drain voltage V_{ds} is increased, the depletion layer becomes wider. The resulting decrease in conductive cross section d must be compensated by an increase of electron velocity v to maintain a constant current through the channel. As the drain voltage is increased further, the electrons reach their saturation velocity v_s
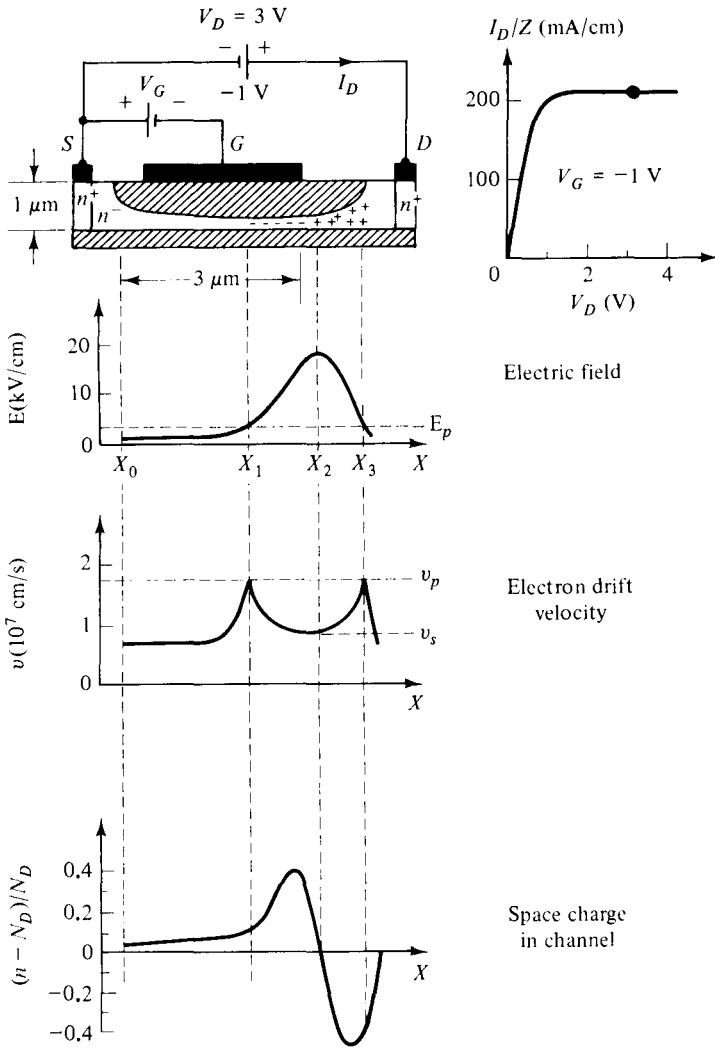
Figure 6-2-6   Channel cross section, electric field, drift velocity, and space charge distribution in the channel of a GaAs MESFET in the saturation region.

as shown in Fig. 6-2-7(c). When the drain voltage is increased beyond $V_{d\,sat}$, the depletion layer widens toward the drain as indicated in Fig. 6-2-7(d). When a negative voltage is applied to the gate as shown in Fig. 6-2-7(e), the gate-to-channel junction is reverse-biased, and the depletion layer becomes wider. As the gate voltage V_{gs} is more negative, the channel is almost pinched off as shown in Fig. 6-2-7(f), and the drain current I_{ds} is nearly cut off.
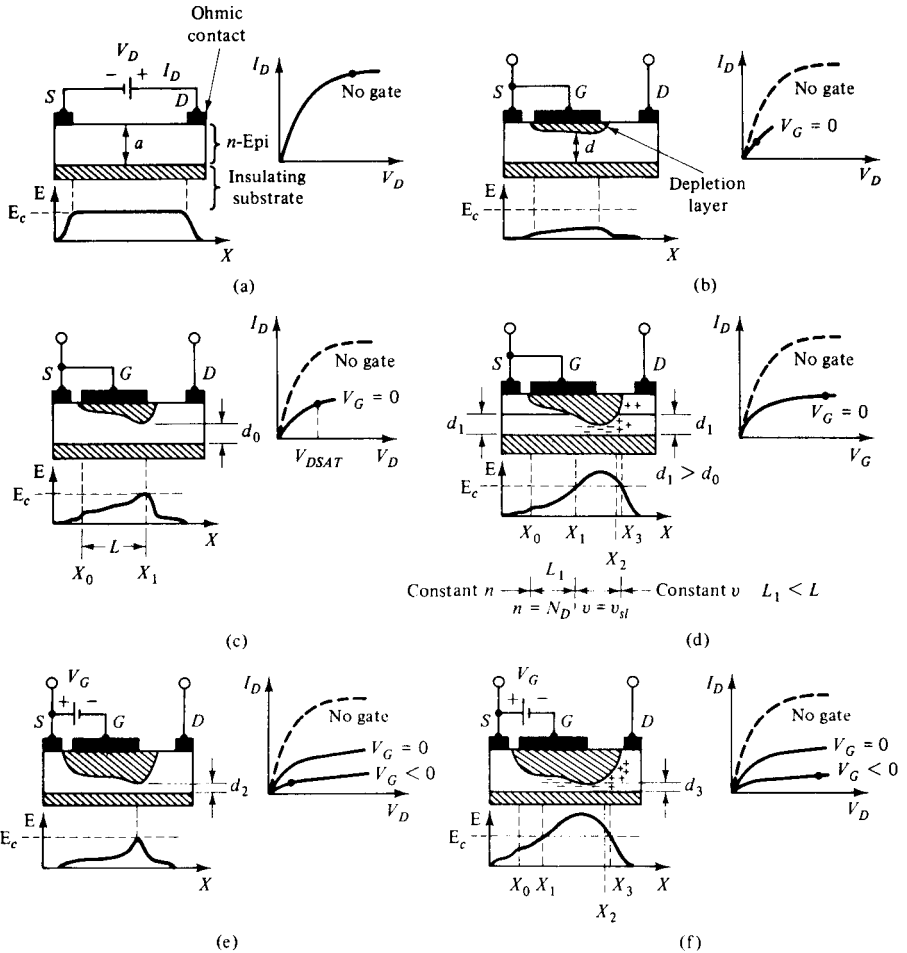
Figure 6-2-7 Electric-field diagrams and current-voltage characteristics of a Si MESFET.

Example 6-2-2: Current-Voltage (I-V) Characteristics of a GaAs MESFET

A typical n-channel GaAs MESFET has the following parameters:

Electron concentration:	$N_d = 8 \times 10^{17}$ cm^{-3}
Channel height:	$a = 0.1$ μm
Relative dielectric constant:	$\epsilon_r = 13.1$
Channel length:	$L = 14$ μm
Channel width:	$Z = 36$ μm
Electron mobility:	$\mu = 0.08$ m^2/V$\cdot$s
	$= 800$ cm^2/V$\cdot$s

Drain voltage: $V_d = 5$ volts
Gate voltage: $V_g = -2$ volts
Saturation drift velocity: $v_s = 2 \times 10^5$ m/s

a. Calculate the pinch-off voltage.
b. Compute the velocity ratio.
c. Determine the saturation current at $V_g = 0$.
d. Find the drain current I_d.

Solution

a. From Eq. (6-2-3), the pinch-off voltage is

$$V_p = 1.6 \times 10^{-19} \times 8 \times 10^{23} \times 10^{-14}/(2 \times 8.854 \times 10^{-12} \times 13.1)$$

$$= 5.52 \text{ volts}$$

b. From Eq. (6-2-4), the velocity ratio is

$$\eta = 0.08 \times 5.52/(2 \times 10^5 \times 14 \times 10^{-6})$$

$$= 0.158$$

c. From Eq. (6-2-4), the saturation current at $V_g = 0$ is

$$I_p = \frac{qN_d\mu aZV_p}{2L}$$

$$= \frac{1.6 \times 10^{-19} \times 8 \times 10^{23} \times 0.08 \times 10^{-7} \times 36 \times 10^{-6} \times 5.52}{3 \times 14 \times 10^{-6}}$$

$$= 4.845 \text{ mA}$$

d. From Eq. (6-2-4), the u and ρ factors are

$$u = \left(\frac{5+2}{5.52}\right)^{1/2} = 1.126 \qquad u^2 = 1.268 \qquad u^3 = 1.428$$

$$\rho = \left(\frac{2}{5.52}\right)^{1/2} = 0.60 \qquad \rho^2 = 0.362 \qquad \rho^3 = 0.217$$

then the drain current is

$$I_d = 4.845 \times 10^{-3} \times \frac{3(1.268 - 0.362) - 2(1.428 - 0.217)}{1 - 0.158(1.268 - 0.362)}$$

$$= 4.845 \times 10^{-3} \times 0.259$$

$$= 1.26 \text{ mA}$$

6-2-5 Cutoff Frequency f_{co} and Maximum Oscillation Frequency f_{max}

The cutoff frequency of a Schottky barrier-gate FET in a circuit depends on the way in which the transistor is being made. In a wideband lumped circuit the cutoff frequency is expressed [7] as

$$f_{co} = \frac{g_m}{2\pi C_{gs}} = \frac{v_s}{4\pi L} \qquad \text{Hz} \qquad (6\text{-}2\text{-}11)$$

where g_m = transconductance

C_{gs} = gate-source capacitance $= \left.\dfrac{dQ}{dV_{gs}}\right|_{V_{gd}=\text{constant}}$

L = gate length

v_s = saturation drift velocity

It is interesting to note that the cutoff frequency of the lumped circuit analysis as shown in Eq. (6-2-11) is different from the charge-carrier transit-time cutoff frequency as shown in Eq. (5-1-58) by a factor of one-half.

The maximum frequency of oscillation depends on the device transconductance and the drain resistance in a distributed circuit. It is expressed [6] as

$$f_{max} = \frac{f_{co}}{2}(g_m R_d)^{1/2} = \frac{f_{co}}{2}\left[\frac{\mu E_p(u_m - \rho)}{v_s(1 - u_m)}\right]^{1/2} \qquad \text{Hz} \qquad (6\text{-}2\text{-}12)$$

where R_d = drain resistance

g_m = device transconductance

E_p = electric field at the pinch-off region in the channel

u_m = saturation normalization of u

v_s, μ, and ρ are defined previously

For $\rho = 0$ (i.e. $V_{gs} = 0$) and $\eta \gg 1$, so that $f_{co} = \dfrac{v_s}{4\pi L}$ and $u_m = \left(\dfrac{3}{\eta}\right)^{1/3} \ll 1$. We have

$$f_{mas} = \gamma \frac{v_s}{L}\left(\frac{3}{\eta}\right)^{1/6} \qquad \text{Hz} \qquad (6\text{-}2\text{-}13)$$

where $\gamma = 0.14$ for $\mu E_p/v_s = 13$ and $\gamma = 0.18$ for $\mu E_p/v_s = 20$ in the case of GaAs.

It has been found experimentally [6] that the maximum frequency of oscillation for a gallium arsenide FET with the gate length less than 10 μm is

$$f_{max} = \frac{33 \times 10^3}{L} \qquad \text{Hz} \qquad (6\text{-}2\text{-}14)$$

where L = gate length in meters. The best value of L is 0.5 μm.

The maximum frequency of oscillation f_{max} is similar to the cutoff frequency f_{co} determined by the transit time. From Eq. (5-1-58) the charge-carrier transit-time cutoff frequency is

$$f_{co} = \frac{1}{2\pi\tau} = \frac{v_s}{2\pi L} \qquad \text{Hz} \qquad (6\text{-}2\text{-}15)$$

where $\tau = \dfrac{L}{v_s}$ is the transit time in seconds

L = gate length in meters

v_s = saturation drift velocity in meters per second

It is evident that the gallium arsenide FET has a better figure of merit than the silicon FET for an X-band amplifier because the saturation drift velocity $\mathcal{V}_s$ is 2×10^7 cm/sec for GaAs at an electric field of 3 kV/cm and 8×10^6 cm/sec for silicon at 15 kV/cm. In comparing Eq. (6-2-15) with Eq. (6-2-11), the difference is a factor of one-half.

The highest frequency of oscillation for maximum power gain with the input and output networks matched is given [8] as

$$f_{max} = \frac{f_{co}}{2}\left(\frac{R_d}{R_s + R_g + R_i}\right)^{1/2} \qquad \text{Hz} \qquad (6\text{-}2\text{-}16)$$

where R_d = drain resistance

R_s = source resistance

R_g = gate metallization resistance

R_i = input resistance

Example 6-2-3:   Cutoff Frequency of a MESFET

A certain GaAs MESFET has the following parameters:

$$R_g = 3\ \Omega \qquad R_i = 2.5\ \Omega \qquad g_m = 50\ \text{m}\mho$$

$$R_d = 450\ \Omega \qquad R_s = 2.5\ \Omega \qquad C_{gs} = 0.60\ \text{pF}$$

a. Determine the cutoff frequency.

b. Find the maximum operating frequency.

Solution

a. From Eq. (6-2-11), the cutoff frequency is

$$f_{co} = \frac{g_m}{2\pi C_{gs}} = \frac{0.05}{2\pi \times 0.6 \times 10^{-12}}$$

$$= 13.26\ \text{GHz}$$

b. From Eq. (6-2-16), the maximum frequency is

$$f_{max} = \frac{f_{co}}{2}\left(\frac{R_d}{R_s + R_g + R_i}\right)^{1/2}$$

$$= \frac{13.26 \times 10^9}{2}\left(\frac{450}{2.5 + 3 + 2.5}\right)^{1/2}$$

$$= 49.73 \text{ GHz}$$

6-3 HIGH ELECTRON-MOBILITY TRANSISTORS (HEMTs)

The evolution of high-speed GaAs integrated circuits (ICs) is the result of continuous technological progress utilizing the superior electronic properties of gallium arsenide compared with those of silicon. Electron mobility in the MESFET channel with typical donor concentrations of about 10^{17} cm^{-3} ranges from 4000 to 5000 cm^2/V·s at room temperature. The mobility in the channel at 77° K is not too much higher than at room temperature because of ionized impurity scattering. In undoped GaAs, however, electron mobility of 2 to 3 $\times$ 10^5 cm^2/V·s has been achieved at 77° K. The mobility of GaAs with feasibly high electron concentrations for facilitating the fabrication of devices was found to increase through modulation-doping technique demonstrated in GaAs–AlGaAs superlattices [8]. A high electron-mobility transistor (HEMT), based on a modulation-doped GaAs–AlGaAs single heterojunction structure, was developed [9]. HEMTs have exhibited lower noise figure and higher gain at microwave frequencies up to 70 GHz, and it is possible to construct HEMT amplifiers at even higher frequencies of operation. The major improvements over MESFETs include shorter gate lengths, reduced gate and source contact resistances, and optimized doping profiles.

6-3-1 Physical Structure

The basic structure of a HEMT is a selectively doped GaAs–AlGaAs heterojunction structure as shown in Fig. 6-3-1.

An undoped GaAs layer and an Si-doped n-type AlGaAs layer are successively grown on a semi-insulating GaAs substrate. A two-dimensional electron gas (2-DEG) is created between the undoped and n-type layers. A buffer layer is sandwiched between the undoped GaAs layer and the semi-insulator substrate.

The HEMT can be fabricated by using the integrated-circuit techniques. Fig. 6-3-2 indicates the sequence for the self-aligned gate procedures in the fabrication of large-scale integration HEMTs, including the E-(enhacement-mode) and D-(depletion-mode) HEMTs.

The processing steps include the following:

1. Ohmic contact formation: The active region is isolated by a shallow mesa step (180 nm), which is almost achieved in a single process, and can be made nearly planar. The source and drain for E- and D-HEMTs are metallized with
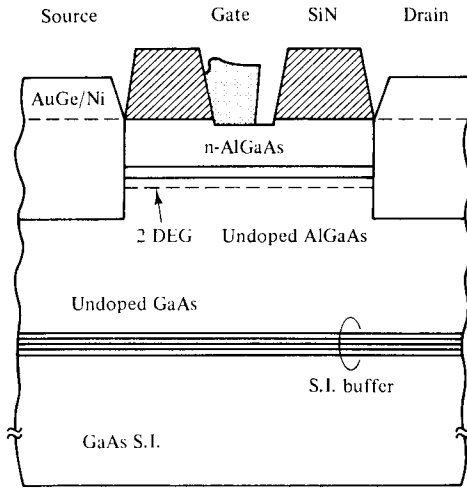
Figure 6-3-1 A cross section of a HEMT (*From K. Togashi et al. [10]; reprinted by permission of* Microwave Journal.)
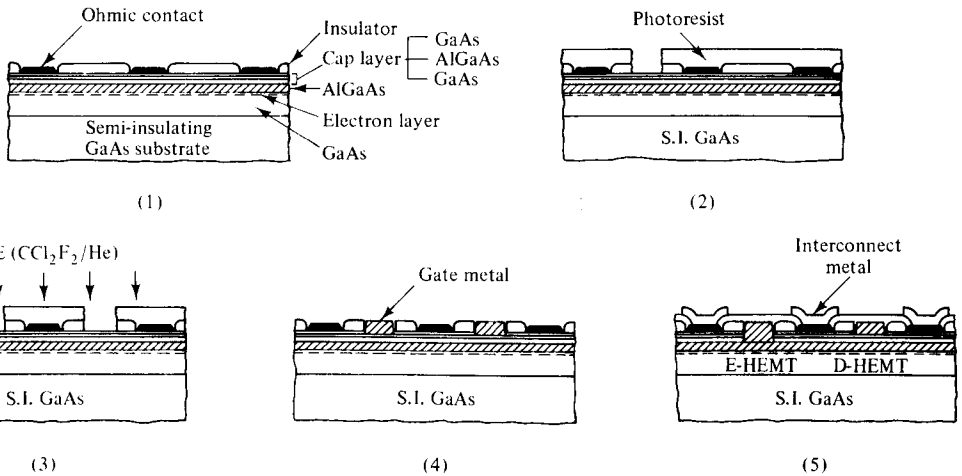
Figure 6-3-2 Processing steps for HEMT direct-coupled FET logic (DCFL) circuits. (*After M. Abe and others [9]; reprinted by permission of IEEE, Inc.*)

an AuGe eutectic alloy and an Au overlay alloy to produce ohmic contacts with the electron layer.

2. Opening gate windows: The fine gate patterns are formed for E-HEMTs as the top GaAs layer and thin $Al_{0.3}Ga_{0.7}As$ stopper are etched off by nonselective chemical etching.

3. Selective dry etching: With the same photoresist process for formation of gate patterns in D-HEMTs, selective dry etching is performed to remove the top GaAs layer for D-HEMTs and also to remove the GaAs layer under the thin $Al_{0.3}Ga_{0.7}As$ stopper for E-HEMTs.

4. Gate metallization: Schottky contacts for the E- and D-HEMT gate are pro-vided by depositing Al. The Schottky gate contacts and the GaAs top layer for ohmic contacts are then self-aligned to achieve high-speed performance.

5. Interconnect metallization: Finally, electrical connections from the intercon-necting metal, composed of Ti, Pt, and Au, to the device terminals are pro-vided through contact holes etched in a crossover insulator film.

Figure 6-3-3 shows the cross-sectional view of a typical self-aligned structure of E- and D-HEMTs forming an inverter for the direct-coupled FET logic (DCFL) ciruit configuration [9].
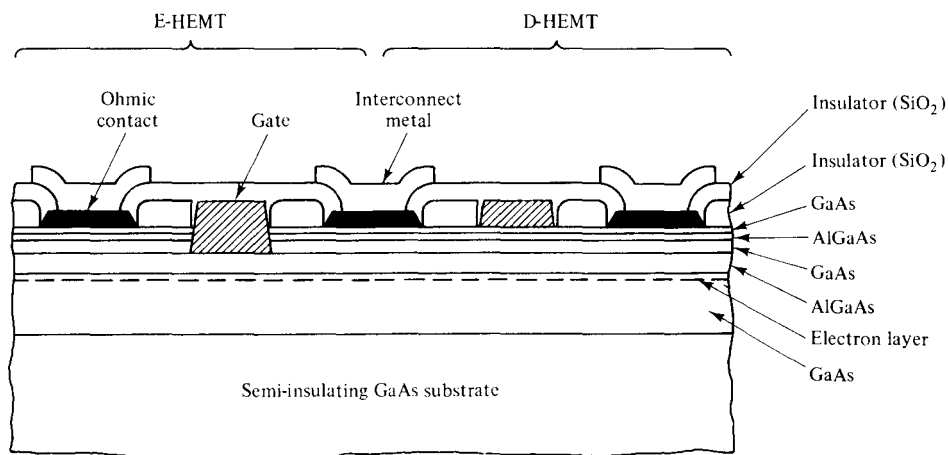
Figure 6-3-3   Cross-sectional view of a DCFL HEMT. (*After M. Abe et al. [9]; reprinted by permission of IEEE, Inc.*)

6-3-2 Operational Mechanism

Since GaAs has higher electron affinity, free electrons in the AlGaAs layer are trans-ferred to the undoped GaAs layer where they form a two-dimensional high-mobility electron gas within 100 Å of the interface. The n-type AlGaAs layer of the HEMT is depleted completely through two depletion mechanisms (11):

1. Trapping of free electrons by surface states causes the surface depletion.
2. Transfer of electrons into the undoped GaAs layer brings about the interface depletion.

The Fermi energy level of the gate metal is matched to the pinning point, which is 1.2 eV below the conduction band. With the reduced AlGaAs layer thick-ness, the electrons supplied by donors in the AlGaAs layer are insufficient to pin the surface Fermi level, and the space-charge region is extended into the undoped GaAs layer. As a result, band bending is moving upward and the two-dimensional electron gas (2-DEG) does not appear. When a positive voltage higher than the threshold

voltage is applied to the gate, electrons accumulate at the interface and form a two-dimensional electron gas.

The electron concentration can control D-(depletion-mode) and E-(enhancement-mode) HEMT operations. As temperature decreases, electron mobility, which is about 8000 $cm^2/V \cdot s$ at $300°$ K, increases dramatically to 2×10^5 $cm^2/V \cdot s$ at $77°$ because of reduced phonon scattering. When the temperatures decrease further, the electron mobility of 1.5×10^6 $cm^2/V \cdot s$ at $50°K$ and 2.5×10^6 $cm^2/V \cdot s$ at $4.5°K$ have been demonstrated.

6-3-3 Performance Characteristics

High electron mobility transistor (HEMT) amplifiers for 40 to 70 GHz have been constructed. The 60-GHz amplifier exhibited a gain of 4.5 to 6.5 dB across the frequency band 56 to 62 GHz, and had an associated noise figure of 6 dB measured at 57.5 GHz. The 72-GHz amplifier achieved a gain of 4 to 5 dB with a bandwidth of 2.5 GHz [11].

Current–voltage (I–V) characteristics.   The drain current can be evaluated from the following basic equation

$$I_{ds} = qn(z)Wv(z) \qquad (6\text{-}3\text{-}1)$$

where q = electron charge
  $n(z)$ = concentration of the two-dimensional electron gas
  W = gate width
  $v(z)$ = electron velocity

Figure 6-3-4 shows the I-V characteristics of a HEMT amplifier with $W = 150 \, \mu m$ and L_g (gate length) $= 0.6 \, \mu m$ at 30 GHz.

The major advantages of a HEMT are higher frequency, lower noise, and higher speed. Table 6-3-1 compares HEMT data with other semiconductor electronic devices.

Example 6-3-1:   Current of a HEMT

A HEMT has the following parameters:

Gate width:	$W = 150 \, \mu m$
Electron velocity:	$v(z) = 2 \times 10^5$ m/s
Two-dimensional electron-gas density:	$n(z) = 5.21 \times 10^{15}$ m^{-2}

Determine the drain current of the HEMT.

Solution   From Eq. (6-3-1), the drain current is

$$I_{ds} = qn(z)Wv(z)$$

$$= 1.6 \times 10^{-19} \times 5.21 \times 10^{15} \times 150 \times 10^{-6} \times 2 \times 10^5$$

$$= 25 \text{ mA}$$
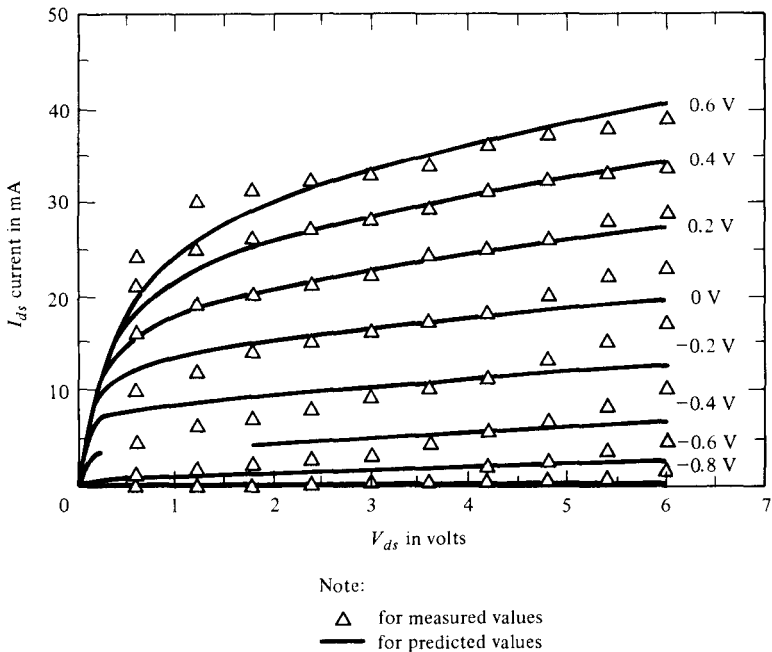
Figure 6-3-4   *I-V* characteristics of a HEMT.

TABLE 6-3-1   HEMT PERFORMANCE COMPARED WITH OTHER DEVICES

Device	Frequency (GHz)	Noise	Power	Speed
HEMT	Up to 70	Very Good	Very Good	Excellent
GaAs MESFET	40	Good	Good	Good
GaAs–AlGaAs HBT*	20	Good	Good	Excellent
Si MOSFET	10	Poor	Very good	Very poor
Si bipolar transistor	1	Poor	Poor	Good

*HBT = heterojunction bipolar transistor

Equivalent circuit.   In order to predict or calculate the values for a small- or large-signal HEMT amplifier, the following high-frequency equivalent circuit, shown in Fig. 6-3-5, may be useful.

It is difficult to compare the optimized performance of most major semiconductor devices. The switching delay time of GaAs MESFETs is two or three times longer than that of HEMTs. The GaAs–AlGaAs heterojunction bipolar transistor (HBT) can achieve the same high-speed performance as the HEMT. The ultimate speed capability, limited by cutoff frequency f_T, is more than 100 GHz, and the HBT also has the merit of flexible fan-out loading capability. The silicon MOSFET and the bipolar transistor have excellent performance in threshold voltage uniformity and controllability with no material problems, and they are easy to fabricate despite complex processing steps. HEMTs are very promising devices for high-speed, very large-scale integration (VLSI) with low-power dissipation, but they require new
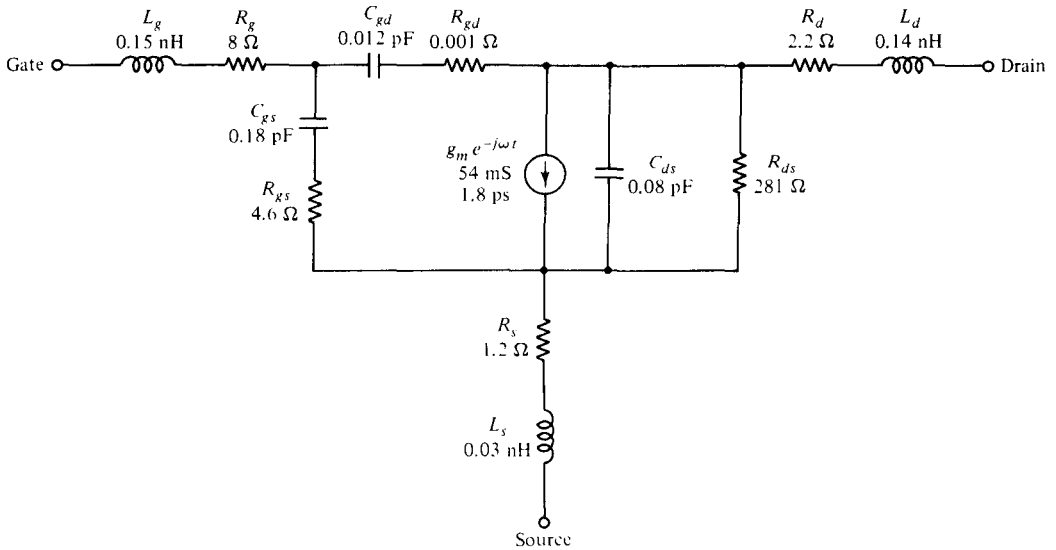
Figure 6-3-5   Equivalent circuit of a HEMT.

technological breakthroughs to achieve the large-scale integration (LSI) quality of GaAs–AlGaAs material. Such technologies include molecular-beam epitaxy (MBE), organic metal-vapor phase epitaxy (OMPVE), and self-alignment device fabrications. The excellent controllability of MBE growth can regulate the ratio of standard deviation of threshold voltage to the logic voltage swing.

The vertical threshold sensitivity is defined by the differential threshold voltage to the thickness of the $Al_x Ga_{1-x}$ As layer and is expressed as

$$\frac{dV_{th}}{d\ell} = -[2qN_d(\psi_{ms} - \Delta E_c - V_{th})/\epsilon]^{1/2} \qquad (6\text{-}3\text{-}2)$$

where V_{th} = threshold voltage in volts

  ℓ = AlGaAs layer thickness in meters

  q = electron charge

  N_d = donor concentration

  ψ_{ms} = metal-semicondcutor Schottky barrier potential between Al and GaAs

  E_c = conduction bandedge difference between GaAs and AlGaAs

  ϵ = $\epsilon_o \epsilon_r$ is the permittivity of AlGaAs

  ϵ_r = dielectric constant of AlGaAs

Example 6-3-2:   Sensitivity of HEMT

  A HEMT has the following parameters:

Threshold voltage:	$V_{th} = 0.13$ V
Donor concentration:	$N_d = 2 \times 10^{24}$ m^{-3}
Metal-semiconductor Schottky barrier potential:	$\psi_{ms} = 0.8$ V

GaAs bandgap: $E_{gg} = 1.43$ V
AlGaAs bandgap: $E_{ga} = 1.80$ V
AlGaAs dielectric constant $\epsilon_r = 4.43$

Compute:

a. The conduction band-edge difference between GaAs and AlGaAs
b. The sensitivity of the HEMT

Solution

a. The conduction band-edge difference is

$$\Delta E_c = E_{ga} - E_{gg} = 1.80 - 1.43 = 0.37 \text{ V}$$

b. The sensitivity of the HEMT is

$$\frac{dV_{th}}{d\ell} = -[2qN_d(\psi_{ms} - \Delta E_c - V_{th})/\epsilon]^{1/2}$$

$$= -[2 \times 1.6 \times 10^{-19} \times 2 \times 10^{24}(0.80 - 0.37$$
$$- 0.13)/(8.854 \times 10^{-12} \times 4.43)]^{1/2}$$

$$= -\left[\frac{0.49 \times 10^4}{10^{-12}}\right]^{1/2} = -70 \text{ mV/nm}$$

$$\left|\frac{dV_{th}}{d}\right| = 70 \text{ mV/nm}$$

6-3-4 Electronic Applications

The switching speed of a HEMT is about three times as fast as that of a GaAs MES-FET. The largest-scale logic integrated circuit (IC) with HEMT technology has achieved the highest speed ever reported among 8 × 8 bit multipliers. The switching delay time of a HEMT is below 10 picosecond with a power dissipation reported at about 100 μW. Therefore, HEMTs are promising devices for very large-scale integration, especially in very high-speed supercomputers, star wars, and space communications.

Information processing in the 1990s will require ultrahigh-speed computers, having high-speed large-scale integration circuits with logic delays in the sub-hundred-picosecond range. A 4K × 1 bit static random-access memory (SRAM) device consists of a memory cell of 55 μm × 39 μm. Its normal read-write operation was confirmed both at 300° K and 77° K. The minimum address access time obtained was 2.0 nanosecond with a chip dissipation power of 1.6 W and a supply voltage of 1.54 V. The HEMT SRAMs have demonstrated better performance than the SiMOS, BJT, and GaAs MESFET SRAMs.

6-4 METAL-OXIDE-SEMICONDUCTOR FIELD-EFFECT TRANSISTORS (MOSFETs)

The metal-oxide-semiconductor field-effect transistor (MOSFET) is a four-terminal device, and its current flow is controlled by an applied vertical electric field. The four terminals are denoted as the *source, gate, drain,* and *substrate.* When the gate bias voltage is zero, the two back-to-back *p-n* junctions between the source and drain prevent current flow in either direction. When a positive voltage is applied to the gate with respect to the source (source and substrate are common), negative charges are induced in the channel to provide current flow. Since the current is controlled by the electric field, this type of device is called the *junction field-effect transistor* (JFET). The MOSFETs have superseded the bipolar junction transistors in many electronic applications because their structures are simple and their costs are low. In addition, the *n*-channel MOSFET (NMOS), *p*-channel MOSFET (PMOS), complementary MOSFET (CMOS), logic-gate memories, and the charge-coupled devices (CCDs) have emerged as important semiconductor electronic devices. All these devices are discussed in this section.

The metal-insulator-semiconductor field-effect transistor (MISFET) may be formed by a metal such as aluminum (Al) and a semiconductor, such as Ge, Si, or GaAs, with an insulator such as SiO_2, Si_3N_4, or Al_2O_3 sandwiched between. If the structure is in the form of Al-SiO_2-Si, it is called a MOSFET. This device is very useful in very large-scale integrated microwave circuits. In the near future, a microwave-device chip of 1-μm dimension containing 1 million or more devices will be commercially available. The basic component of a MOSFET is the MIS diode which was described previously.

6-4-1 Physical Structures

The metal-oxide-semiconductor field-effect transistor (MOSFET) is a four-terminal device. There are both *n*-channel and *p*-channel MOSFETs. The *n*-channel MOSFET consists of a slightly doped *p*-type semiconductor substrate into which two highly doped n^+ sections are diffused, as shown in Fig. 6-4-1. These n^+ sections, which act as the source and the drain, are separated by about 0.5 μm. A thin layer of insulating silicon dioxide (SiO_2) is grown over the surface of the structure. The metal contact on the insulator is called the *gate.*

Similarly, the *p*-channel MOSFET is made of a slightly doped *n*-type semiconductor with two highly doped p^+-type regions for the source and drain. The heavily doped polysilicon or a combination of silicide and polysilicon can also be used as the gate electrode. In practice, a MOSFET is commonly surrounded by a thick oxide to isolate it from the adjacent devices in a microwave integrated circuit. The basic device parameters of a MOSFET are as follows: L is the channel length, which is the distance between the two n^+-p junctions just beneath the insulator (say, 0.5 μm), Z is the channel depth (say, 5μm), d is the insulator thickness (say, 0.1 μm), and r_j is the junction thickness of the n^+ section (say, 0.2 μm).
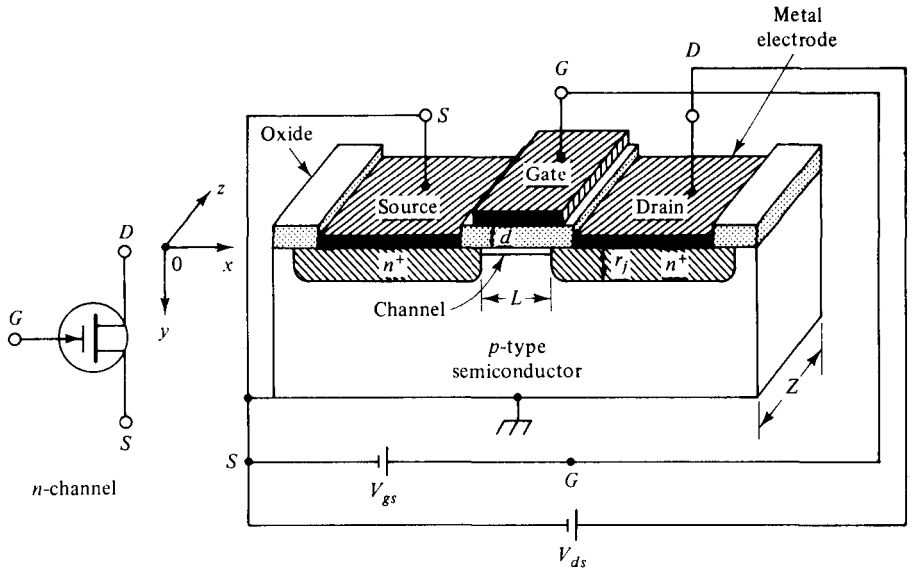
Figure 6-4-1   Schematic diagram of an n-channel MOSFET.

6-4-2 Electronic Mechanism

When no voltage is applied to the gate of an n-channel MOSFET, the connection between the source electrode and the drain electrode corresponds to a link of two p-n junctions connected back-to-back. The only current that can flow from the source to the drain is the reverse leakage current. When a positive voltage is applied to the gate relative to the source (the semiconductor substrate is grounded or connected to the source), positive charges are deposited on the gate metal. As a result, negative charges are induced in the p-type semiconductor at the insulator-semiconductor interface. A depletion layer with a thin surface region containing mobile electrons is formed. These induced electrons form the n-channel of the MOSFET and allow the current to flow from the drain electrode to the source electrode. For a given value of the gate voltage V_g, the drain current I_d will be saturated for some drain voltages V_d.

A minimum gate voltage is required to induce the channel, and it is called the *threshold voltage* V_{th}. For an n-channel MOSFET, the positive gate voltage V_g must be larger than the threshold voltage V_{th} before a conducting n-channel (mobile electrons) is induced. Similarly, for a p-channel MOSFET, the gate voltage V_g must be more negative than the threshold voltage V_{th} before the p-channel (mobile holes) is formed.

6-4-3 Modes of Operation

There are basically four modes of operation for n-channel and p-channel MOSFETs.

   1. n-Channel Enhancement Mode (normally OFF). When the gate voltage is
   zero, the channel conductance is very low and it is not conducting. A positive

voltage must be applied to the gate to form an n channel for conduction. The drain current is enhanced by the positive voltage. This type is called the enhancement-mode (normally OFF) n-channel MOSFET.

2. **n-Channel Depletion Mode (normally ON).** If an n channel exists at equilibrium (that is, at zero bias), a negative gate voltage must be applied to deplete the carriers in the channel. In effect, the channel conductance is reduced, and the device is turned OFF. This type is called the depletion-mode (normally ON) n-channel MOSFET.

3. **p-Channel Enhancement Mode (normally OFF).** A negative voltage must be applied to the gate to induce a p channel for conduction. This type is called the enhancement-mode (normally OFF) p-channel MOSFET.

4. **p-Channel Depletion Mode (normally ON).** A positive voltage must be applied to the gate to deplete the carriers in the channel for nonconduction. This type is called the depletion-mode (normally ON) p-channel MOSFET.

Figure 6-4-2 shows the four modes of the MOSFETs, and Fig. 6-4-3 illustrates their electric symbols, output I-V characteristics, and transfer characteristics.
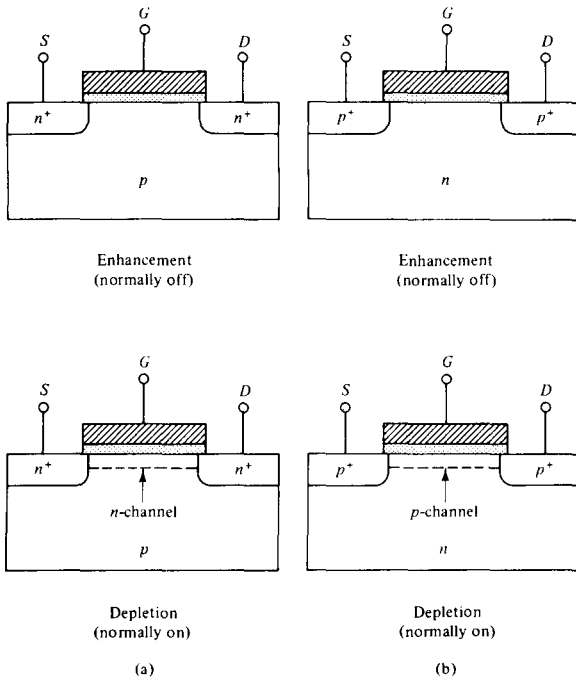
Figure 6-4-2   Four modes of operation for MOSFETs: (a) n channel and (b) p channel.

6-4-4 Drain Current and Transconductance

Drain current.   The drain current I_d of the MOSFET is dependent of the drain voltage V_d. It first increases linearly with the drain voltage in the linear region and then gradually levels off to a saturated value in the saturation region. Figure 6-4-4 shows the current-voltage characteristic curves of an n-channel MOSFET [12].

Type	Electrical symbol	Output characteristic	Transfer characteristic
n-channel enhancement (normally off)	G, S, D	I_D, V_D, V_T, +	I_D, $0\ V_T$, V_G
n-channel depletion (normally on)	G, V_G, I_D, S, D, BS	I_D, V_D, $V_G = 0$	I_D, V_G
p-channel enhancement (normally off)	G, S, D	V_D, 0, V_T, I_D	V_G, $V_T\ 0$, I_D
p-channel depletion (normally on)	G, S, D	V_D, 0, $V_G = 0$, I_D	V_G, 0, I_D
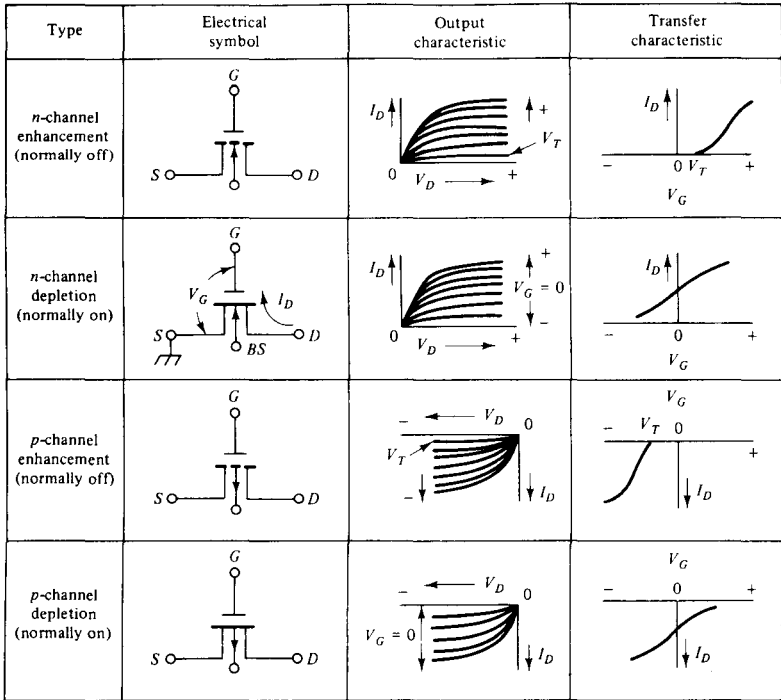
Figure 6-4-3 Electric symbols and output and transfer characteristics of the four modes of MOSFETs.

The drain current I_d is given by [12] as

$$I_d = \frac{Z}{L}\mu_n C_i \left\{ \left(V_g - 2\psi_b - \frac{V_d}{2} \right) V_d - \frac{2}{3C_i}(2\epsilon_s q N_a)^{1/2}\left[(V_d + 2\psi_b)^{1/2} - (2\psi_b)^{1/2} \right] \right\}$$

$$(6\text{-}4\text{-}1)$$

where μ_n = electron carrier mobility

$\qquad C_i = \dfrac{\epsilon_i}{d}$ is the insulator capacitance per unit area

$\qquad \epsilon_i$ = insulator permittivity

$\qquad V_g$ = gate voltage

$\qquad \psi_b = (E_i - E_F)/q$ is the potential difference between the Fermi level E_F and the intrinsic Fermi level E_i

$\qquad V_d$ = drain voltage

$\qquad \epsilon_s$ = semiconductor permittivity

$\qquad q$ = carrier charge

$\qquad N_a$ = acceptor concentration

In the linear region, the drain voltage is small and Eq. (6-4-1) becomes

$$I_d \simeq \frac{Z}{L}\mu_n C_i \left[(V_g - V_{th})V_d - \left(\frac{1}{2} + \frac{\sqrt{\epsilon_s q N_a/\psi_b}}{4C_i} \right) V_d^2 \right]$$

$$(6\text{-}4\text{-}2)$$
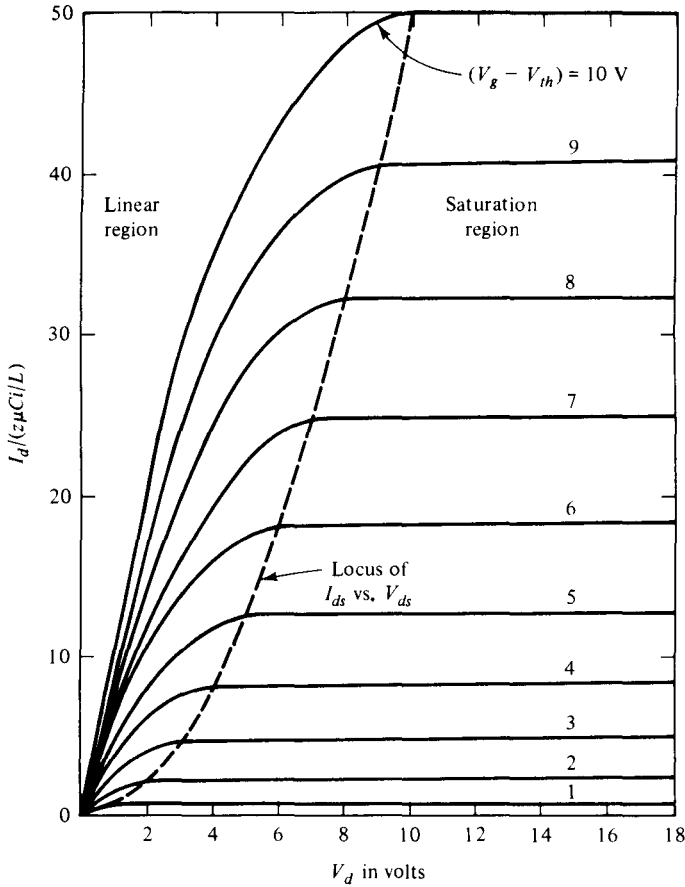
Figure 6-4-4   Current-Voltage (I-V) curves of an n-channel MOSFET.

or

$$I_d \simeq \frac{Z}{L} \mu_n C_i (V_g - V_{th}) V_d \qquad \text{for } V_d \ll (V_g - V_{th}) \qquad (6\text{-}4\text{-}3)$$

where $V_{th} = 2\psi_b + \frac{2}{C_i}(\epsilon_s q N_a \psi_b)^{1/2}$

In the saturation region, the drain current is given by

$$I_{d \text{ sat}} \simeq \frac{mZ}{L} \mu_n C_i (V_g - V_{th})^2 \qquad (6\text{-}4\text{-}4)$$

where $m = 0.5$ is the low doping factor, and $V_{d \text{ sat}} \simeq V_g - V_{th}$ may be assumed

Transconductance.   The transconductance g_m, which is also called the mutual conductance, in the linear region can be found from Eq. (6-4-3) as

$$g_m = \frac{\partial I_d}{\partial V_g}\Bigg|_{V_d = \text{constant}} = \frac{Z}{L}\mu_n C_i V_d \tag{6-4-5}$$

In the saturation region, the transconductance becomes

$$g_{m\,\text{sat}} = \frac{2mZ}{L}\mu_n C_i (V_g - V_{th}) \tag{6-4-6}$$

The channel conductance g_d is given by

$$g_d = \frac{\partial I_d}{\partial V_d}\Bigg|_{V_g = \text{constant}} = \frac{Z}{L}\mu_n C_i (V_g - V_{th}) \tag{6-4-7}$$

All equations derived here so far are based on the idealized n-channel MOSFET. For an idealized p-channel MOSFET, all voltages V_g, V_d, and V_{th} are negative, and the drain current I_d flows from the source to the drain. For a real n-channel MOSFET (say, Al-Oxide-Si structure), the saturation drain current is

$$I_{d\,\text{sat}} = ZC_i (V_g - V_{th})v_s \tag{6-4-8}$$

and the transconductance becomes

$$g_m = ZC_i v_s \tag{6-4-9}$$

where $v_s = L/\tau$ is the carrier drift velocity

The threshold voltage V_{th} is given by

$$V_{th} = \frac{\phi_{ms}}{q} - \frac{Q_f}{C_i} + 2\psi_b + \frac{2}{C_i}(\epsilon_s q N_a \psi_b)^{1/2} \tag{6-4-10}$$

where $\phi_{ms} = \phi_m - \phi_s$ is the work function difference (in eV) between the metal work function ϕ_m and the semiconductor work function ϕ_s

Q_f = fixed oxide charges

Example 6-4-1:   Threshold Voltage of an Ideal MOSFET

A certain p-channel MOSFET has the following parameters:

Doping concentration:	$N_a = 3 \times 10^{17}$ cm^{-3}
Relative dielectric constant:	$\epsilon_r = 11.8$
Relative dielectric constant of SiO$_2$:	$\epsilon_{ir} = 4$
Insulator depth:	$d = 0.01$ μm
Operating temperature:	$T = 300°$K

a. Calculate the surface potential ψ_s (inv) for strong inversion.
b. Compute the insulator capacitance.
c. Determine the threshold voltage.

Solution

a. From Eq. (6-4-4), the surface potential for strong inversion is

$$\psi_{s\,inv} = 2 \times 26 \times 10^{-3} \ell n\left(\frac{3 \times 10^{17}}{1.5 \times 10^{10}}\right)$$

$$= 0.874 \text{ volt}$$

b. From Eq. (6-4-1), the insulator capacitance is

$$C_i = \frac{\epsilon_i}{d} = \frac{4 \times 8.854 \times 10^{-12}}{0.01 \times 10^{-6}} = 3.54 \text{ mF/m}^2$$

c. From Eq. (6-4-3), the threshold voltage is

$$V_{th} = 0.874 + \frac{2}{3.54 \times 10^{-3}} \times (8.854 \times 10^{-12} \times 11.8$$

$$\times 1.6 \times 10^{-19} \times 3 \times 10^{23} \times 0.437)^{1/2}$$

$$= 0.874 + 0.56 \times 10^3 \times 14.80 \times 10^{-4}$$

$$= 1.70 \text{ volts}$$

6-4-5 Maximum Operating Frequency

The maximum operating frequency of a MOSFET is determined by its circuit parameters. A common-source equivalent circuit of a MOSFET is shown in Fig. 6-4-5.

where g_{in} = input conductance due to the leakage current. Since the leakage current is very small, say 10^{-10} A/cm^2, g_{in} is negligible

C_{in} = ZLC_i is the input capacitance

C_{fb} = feedback capacitance

g_{out} = g_d is output conductance

C_{out} = $C_i C_s/(C_i + C_s)$ is the sum of the two p-n junction capacitances in series with the semiconductor capacitance per unit area
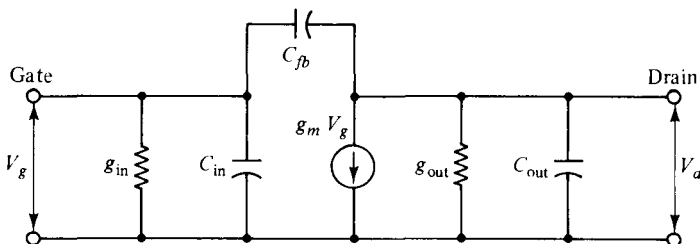
Figure 6-4-5   Equivalent circuit of a common-source MOSFET.

The maximum operating frequency of a MOSFET in the linear region can be expressed as

$$f_m = \frac{\omega_m}{2\pi} = \frac{g_m}{2\pi C_{\text{in}}} = \frac{\mu_n V_d}{2\pi L^2} \tag{6-4-11}$$

In the saturation region, $V_g \gg V_{th}$, the transconductance is reduced to

$$g_{m\,\text{sat}} = \frac{Z}{L}\mu_n C_i V_g = Z C_i v_s \tag{6-4-12}$$

6-4-6 Electronic Applications

The MOSFETs are often used as power amplifiers because they offer two advantages over MESFETs and JFETs.

1. In the active region of an enhancement-mode MOSFET, the input capacitance and the transconductance are almost independent of gate voltage, and the output capacitance is independent of the drain voltage. This leads to very linear (Class A) power amplification.
2. The active gate-voltage range can be larger because n-channel depletion-type MOSFETs can be operated from the depletion-mode region $(-V_g)$ to the enhancement-mode region $(+V_g)$.

Example 6-4-2: Characteristics of a MOSFET

A certain n-channel MOSFET has the following parameters:

Channel length:	$L = 4\ \mu\text{m}$
Channel depth:	$Z = 12\ \mu\text{m}$
Insulator thickness:	$d = 0.05\ \mu\text{m}$
Gate voltage:	$V_g = 5$ V
Doping factor:	$m = 1$
Threshold voltage:	$V_{th} = 0.10$ V
Electron mobility:	$\mu_n = 1350 \times 10^{-4}\ \text{m}^2/\text{V} \cdot \text{s}$
Electron velocity:	$\mathcal{V}_s = 1.70 \times 10^7$ cm/s
Relative dielectric constant of SiO_2:	$\epsilon_{ir} = 3.9$

a. Compute the insulator capacitance in F/m^2.
b. Calculate the saturation drain current in mA.
c. Determine the transconductance in the saturation region in millimhos.
d. Estimate the maximum operating frequency in the saturation region in GHz.

Solution

a. The capacitance of the insulator SiO_2 is

$$C_i = \frac{\epsilon_i}{d} = \frac{3.9 \times 8.854 \times 1C^{-12}}{0.05 \times 10^{-6}} = 6.91 \times 10^{-4} \, \text{F/m}^2$$

b. From Eq. (6-4-8), the saturation drain current is

$$I_{d\,\text{sat}} = ZC_i(V_g - V_{th})v_s$$

$$= 12 \times 10^{-6} \times 6.91 \times 10^{-4} \times (5 - 0.1) \times 1.7 \times 10^5$$

$$= 6.91 \, \text{mA}$$

c. From Eq. (6-4-9), the transconductance in the saturation region is

$$g_{m\,\text{sat}} = ZC_iv_s$$

$$= 12 \times 10^{-6} \times 6.91 \times 10^{-4} \times 1.7 \times 10^5$$

$$= 1.41 \, \text{m}\mho$$

d. From Eq. (6-4-13), the maximum operating frequency is

$$f_m = \frac{v_s}{2\pi L} = \frac{1.7 \times 10^5}{2\pi \times 4 \times 10^{-6}} = 6.76 \, \text{GHz}$$

6-5 MOS TRANSISTORS AND MEMORY DEVICES

As discussed in Section 6-2, the source, channel, and the drain of the MOS transistor are surrounded by a depletion region, so there is no need to isolate individual components. This elimination of isolation regions in MOS transistors allows a much greater packing density on a semiconductor chip than is possible with bipolar junction transistors. The MOSFET can be subdivided into two groups:

1. The n-channel MOSFET is commonly referred to as an *NMOS*.
2. The complementary MOSFET is usually called a *CMOS*. The CMOS provides n-channel and p-channel MOSFETs on the same chip.

The fabrication technology for an NMOS is much simpler than for the bipolar transistor. The CMOS circuit has lower power dissipation than both the bipolar transistor and the NMOS circuits. So both NMOS and CMOS devices are very useful in high density integrated circuits.

6-5-1 NMOS Devices

The metal-oxide-semiconductor field-effect transistor (MOSFET) is the dominant device used in very large-scale integration (VLSI) circuits. In the 1960s, the *p-channel MOS (PMOS)* was originally used in integrated circuits. *n-channel MOS (NMOS)*

devices, however, have dominated the IC market since the 1970s because their electron mobility is higher than that of holes.

NMOS structure. Figure 6-5-1 shows a three-dimensional view of an NMOS logic circuit.
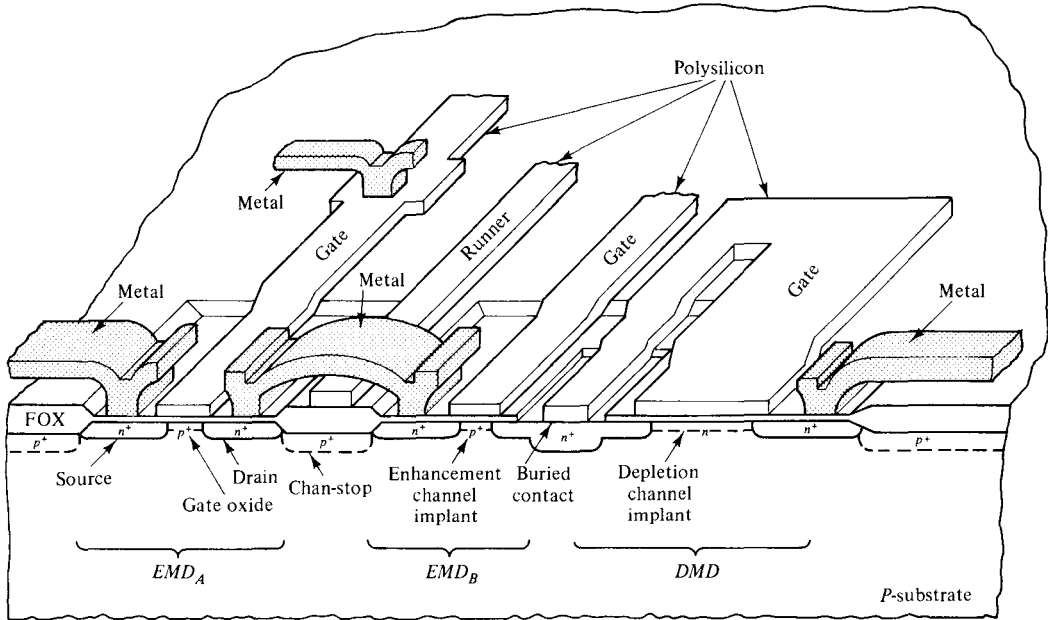
Figure 6-5-1 Three-dimensional view of an NMOS logic circuit. (*After Parrillo [13]; reprinted by permission of the Bell Laboratory.*)

From Fig. 6-5-1, it can be seen that two enhancement-mode (normally OFF) devices (EMD_A and EMD_B) are in series with a depletion-mode (normally ON) device (DMD). A field oxide (FOX) surrounds the transistors, and the gate and source of the DMD are connected at the buried contact. An intermediate dielectric layer separates the overlying metal layer from the underlying layers. In the JFET the high input resistance is obtained from the reverse-biased *p-n* junction. In the MOSFET the extremely high input resistance ($\sim 10^{14}$ ohms) is made possible by the insulator.

There are two basic structures for MOSFETs: the depletion type and the induced type as shown in Figs. 6-5-2 and 6-5-3, respectively. The major difference between them is that with the terminals of the device open-circuited, the depletion type has a conducting channel that links the drain to the source; the induced type has a channel of opposite type to that of the drain and the source linking the two regions. When the device has *p*-type source and drain regions, it is called the *p*-channel MOSFET or PMOS.

(a) Cross-sectional view

(b) I-V curves
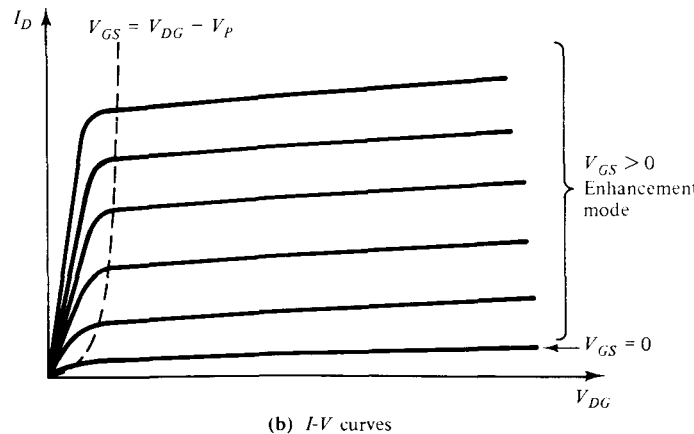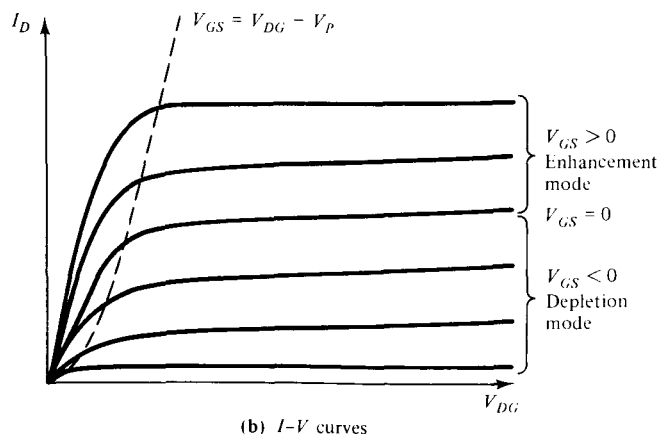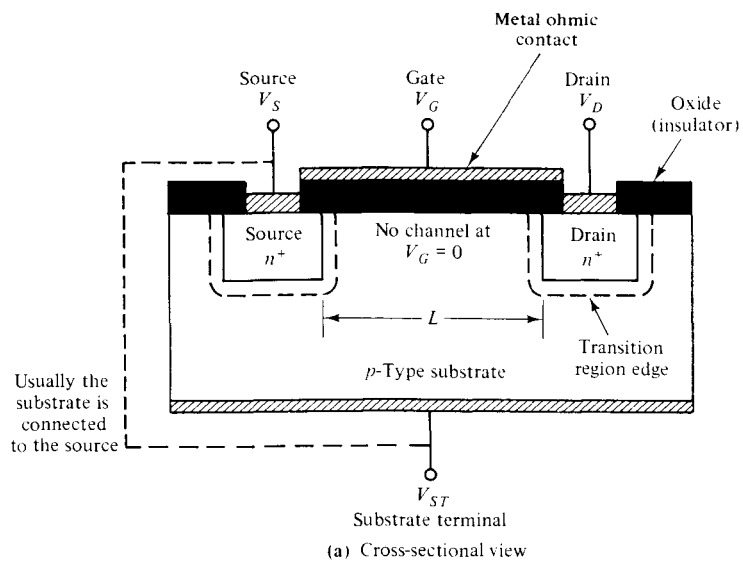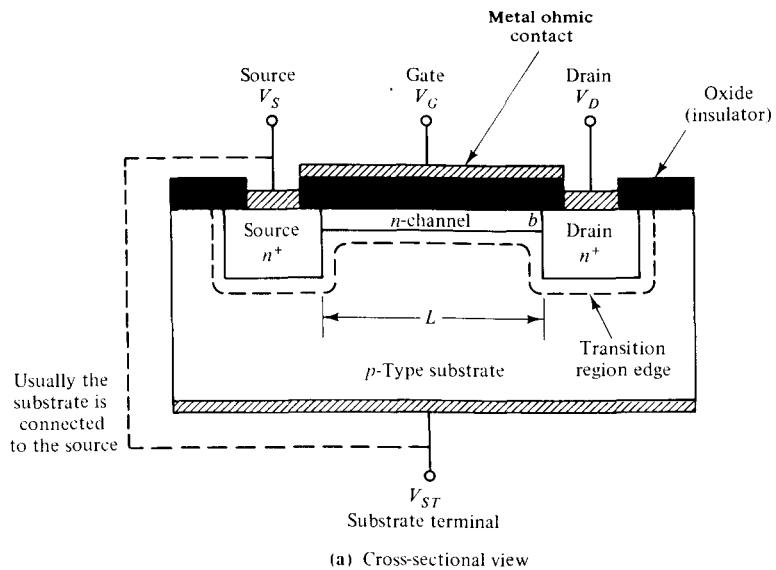
Figure 6-5-2 *n*-channel depletion-type MOSFET.

(a) Cross-sectional view

(b) I-V curves

Figure 6-5-3 *n*-channel induced-type MOSFET.

The structures of enhancement and depletion modes refer to the relative increase or decrease of the majority carrier density in the channel connecting the source to the drain. If a given gate bias tends to increase the majority carrier density in the channel, the device is said to be operated in the enhancement mode (normally OFF). Hence, if the gate in an NMOS is biased by a positive voltage with respect to the substrate, which tends to drain more electrons into the n-type channel, it is said to be operated in the enhancement mode. If a negative gate potential with respect to the substrate is applied in order to diminish the electron density in the n channel, then it is said to be operated in the depletion mode. For a p-channel MOSFET (PMOS), when the negative gate potential with respect to the substrate is biased, the device is operated in the enhancement mode; and when the positive gate voltage is applied, the device is operated in the depletion mode. When a MOS is operated in the enhancement mode, the drain current is higher because the majority carrier density is higher. When, however, the MOS is operated in the depletion mode, the drain current is lower because the majority carrier density is lower.

NMOS operation.   The operation of an NMOS can be described as a logic gate as shown in Fig. 6-5-4.

Two enhancement-mode devices (EMD) are in series with a depletion-mode device (DMD), and the three transistors are connected between the positive power supply V_{DD} and ground V_{SS} reference. The DMD is normally on at $V_{GS} = 0$ and acts as a current source for the two EMDs. Gates A and B of the two EMDs serve as inputs to the logic circuit, and the DMD's gate-source connection is the output electrode of the logic circuit. The output voltage of the two-input NAND circuit is low only when both EMDs are turned on at their logic-high level.

NMOS devices can be also used as NOR-gate circuits, as shown in Fig. 6-5-5.

The two-input enhancement-mode NMOS devices are called *drivers,* and the depletion-mode MOSFET, *the load.* When there is a logic O-state input signal (low voltage at V_{i1}) at the driver device, the device has a very small channel current and the output is at 1 state (the output voltage is close to V_{DD}). But when the input signal is at 1 state, the device conducts a large current with a small voltage drop across the driver, and the output is at O state. The input and output are inverted and a NOR logic is achieved.

6-5-2 CMOS Devices

The *complementary MOS* (*CMOS*) is made of both NMOS and PMOS devices, and its power consumption is quite low. In some CMOS designs, the NMOS circuit is incorporated in domino-CMOS to take advantage of the NMOS's high speed and the CMOS's low power.

CMOS structure.   There are three structures of CMOS devices: n tub, p tub, and twin tub. When a tub is formed in a p-type substrate, the device is called an n *tub,* just as when in an n-type substrate it is called a p *tub.* If an n-tub and a p-tub are combined on the same substrate, the device is referred to as a *twin tub.* A tub is

(a)

(b)

(c)
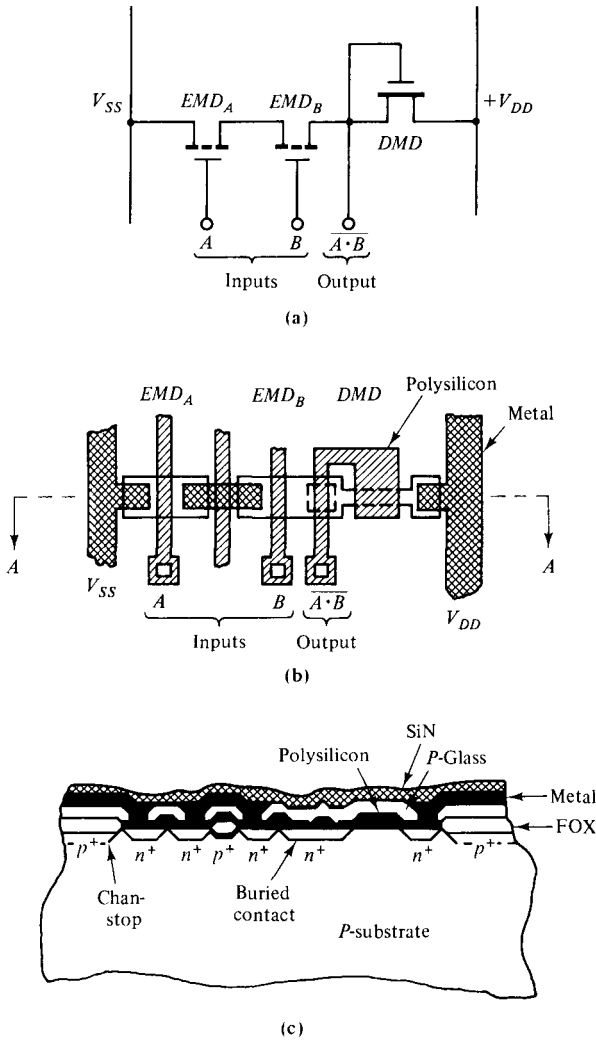
Figure 6-5-4   Two-input NAND logic gate. (*After Parrillo [13]; reprinted by permission of the Bell Laboratory.*)
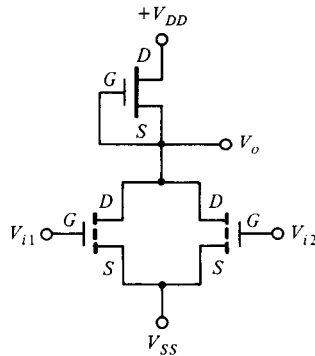
Figure 6-5-5   NOR logic gate.

(a) *n* tub

(b) *p* tub
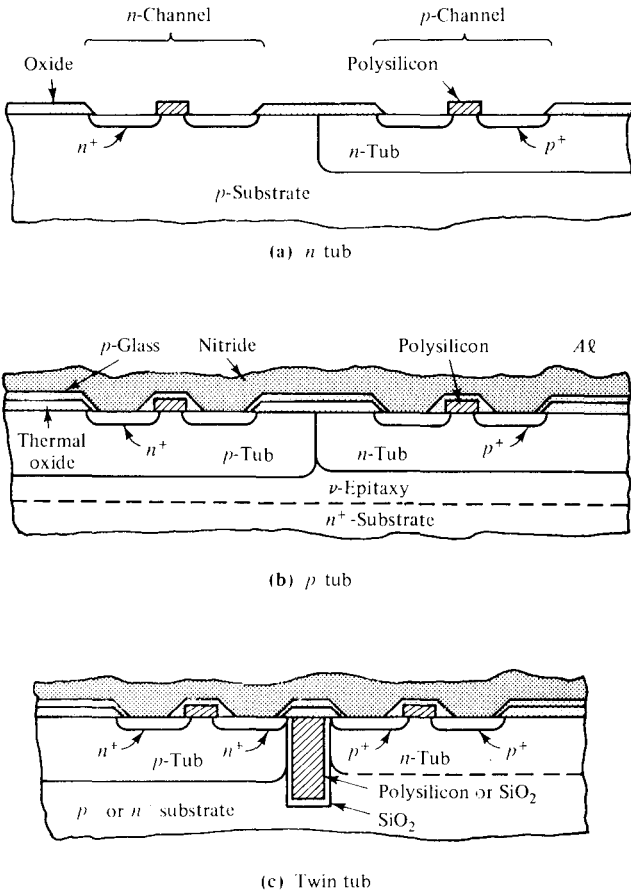
(c) Twin tub

Figure 6-5-6 CMOS structures. (*After L.C. Parrillo [13]; reprinted by permission of the Bell Laboratory.*)

also called a well, and it can be produced by extra diffusion steps. Figure 6-5-6 shows the structures of an *n* tub, a *p* tub, and a twin tub.

CMOS operation.   The operation of a CMOS inverter is shown in Fig. 6-5-7. The *p*-channel transistor is formed in the *n*-type structure, and the *n*-channel transistor is grown in the *p* region, which in turn is formed in the *n*-type substrate. The *p* region acts as the *n*-channel transistor's substrate (back gate) and it is commonly referred to as a tub or well. The gates of the *n*- and *p*-channel transistors are connected and serve as the input to the inverter. The common drains of each device are the output of the inverter. The threshold voltage of the *n*- and *p*-channel transistors are V_{tn} and V_{tp}, respectively, ($V_{tp} < 0$). Figure 6-5-7(c) shows the dependence of the output voltage V_o on the input voltage V_I of the CMOS inverter. For $V_I = 0$, the *n*-channel transistor is OFF ($V_I \ll V_{tn}$), while the *p*-channel transistor is heavily turned ON (that is, the gate-to-source voltage of the *p* channel is $-V_{DD}$, which is much more negative than V_{tp}). Therefore, the output voltage is $V_o = V_{DD}$. As the input voltage V_I is increased above zero, the *n*-channel transistor eventually is turned

(a) Circuit diagram

(b) Device cross-section
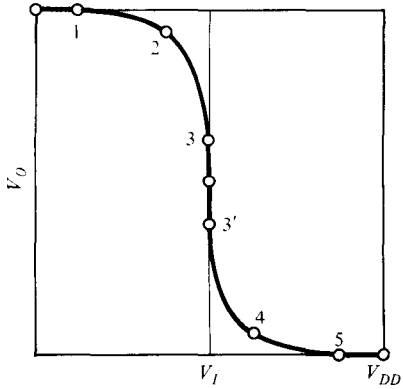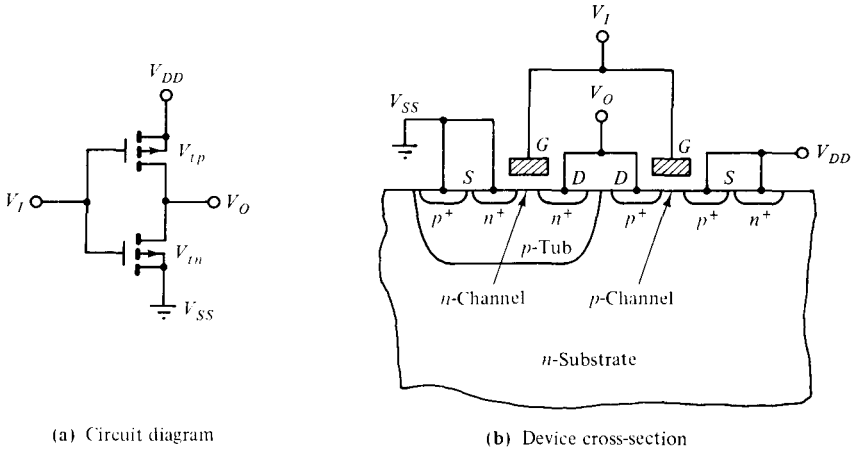
(c) V_o vs V_I
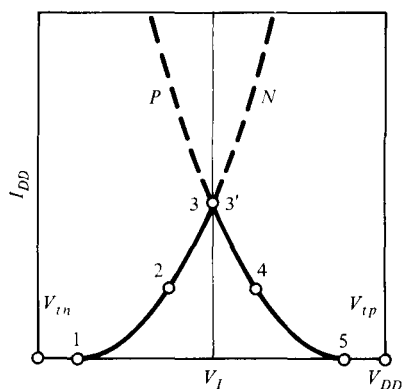
(d) Current curve

Figure 6-5-7 CMOS operation. (*After Parrillo [13]; reprinted by permission of the Bell Laboratory.*)

ON, while the p-channel transistor finally turns OFF. When $V_I > (V_{DD} - |V_{tp}|)$, the output voltage is $V_o = V_{SS}$.

The key feature of a CMOS gate is that in either logic state ($V_o = V_{DD}$ or V_{SS}) one of the two transistors is OFF and the current conducted between V_{DD} and V_{SS} is negligible. Figure 6-5-7(d) shows the current I_{DD} as a function of V_I. A significant current is conducted through the CMOS circuit only when both transistors are ON at the same switching time. The low power consumption of CMOS is one of its most important contributions. The performance and simplification of circuit design are other attractive features of the CMOS device. CMOS provide the circuit designer with flexibility in designing circuits that are either static CMOS (a p-channel transistor for every n-channel transistor) or dynamic CMOS (unequal number of n- and p-channel transistors).

6-5-3 Memory Devices

Memories are devices that can store digital data or information in terms of bits (binary digits). Many memory chips were designed and developed by using NMOS devices. The major semiconductor memory categories are ROM, PROM, EPROM, EEPROM, SRAM, BSRAM, and DRAM. The unit cells of the memory types are shown in circuit schematics in Fig. 6-5-8 [14].
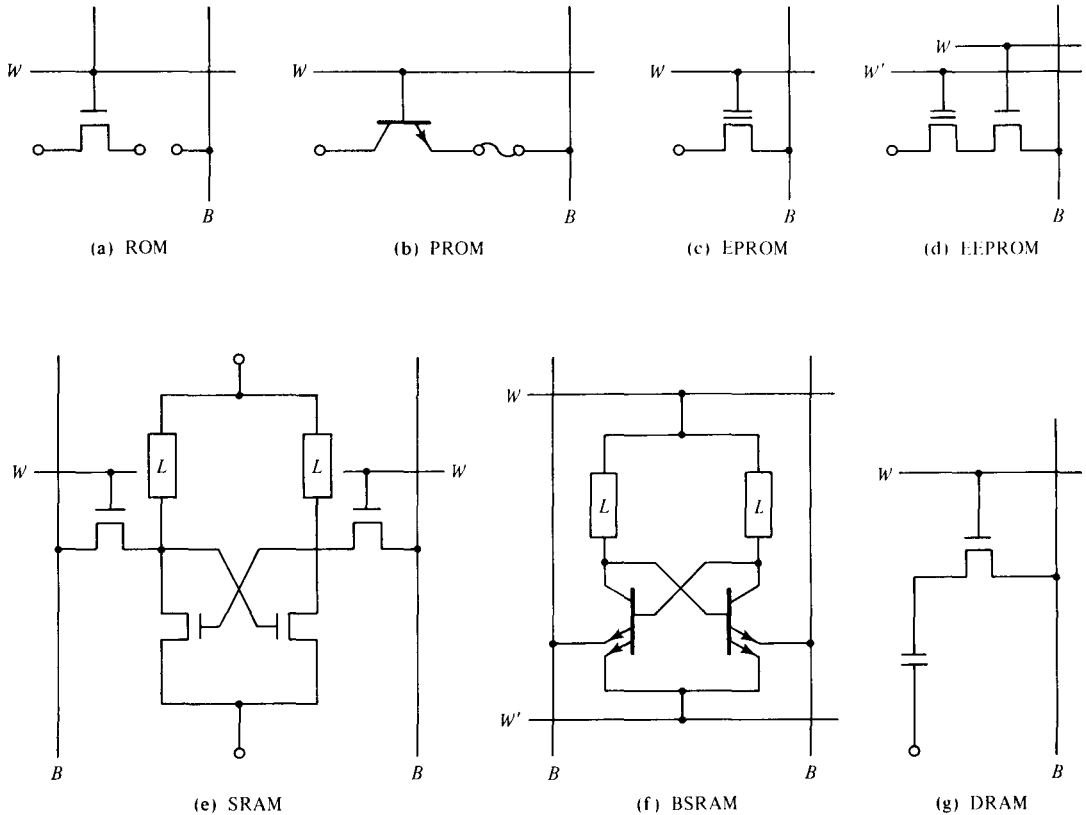
Figure 6-5-8   Circuit diagrams of memory types. (*After S. Asai [14]; reprinted by permission of IEEE, Inc.*)

ROM.   The *read-only memory* (ROM) is also called *mask ROM*. During fabrication stage, the information is inscribed (and cannot be altered) in the form of presence or absence of a link between the word (access) line and the bit (sense) line. This causes the presence or absence of a readout signal on the bit line when the word line is activated. The essential part of the ROM is the way the link is provided, since the link determines the cell size (and thus the cost per bit) and the turnaround time (TAT). Fast, high-density ROMs are in great demand for personal computers.

PROM.   The *programmable read-only memory* (PROM) is one of the ROMs that are field-programmable but lack "erase" capability. PROM uses cells with a fuse that can be blown open electrically, or a *p-n* diode that can be short-circuited by an avalanching pulse. A bipolar PROM has 64-Kbit *p-n* diode cells, and MOS PROMs have higher density but slower speed.

EPROM.   EPROM stands for the *erasable programmable read-only memory*. In the EPROM cell, the presence or absence of charge in the floating gate of a double-poly-gate MOSFET determines the logic state. Programming is done by injecting energetic carriers generated by drain *p-n* junction avalanche breakdown, thereby increasing the threshold voltage V_{th} of the memory transistor. The EPROM is, therefore, also called the FAMOS (*floating-gate avalanche injection MOS*) device. When ultraviolet light is shone on the device, the charge in the floating gate is released, thereby erasing the memory. EPROMs are therefore packaged with a glass window to permit the erasing operation to occur. The original *p*-channel FAMOS cell, which has two transistors in series—one for storage and the other for access, has been replaced by *n*-channel, single-transistor cells. The original programming voltage of 30 V has been reduced to around 12 V. The most advanced 1-Mbit EPROMs today have a cell size of $19-29$ μm^2 and an access time of $80-140$ ns.

EEPROM.   EEPROM designates the *electrically erasable programmable read-only memory,* and it is the most sophisticated in its principle of operation. For a ROM to be electrically erasable, it must have the means to inject and extract charge carriers into and from a floating gate. The first proposal for an EEPROM used the electron-trapping states at the nitride-oxide interface in a metal-nitride-oxide-silicon (MNOS) structure with a very thin (about 2 nm) oxide. The recent dominant technology of EEPROM uses a floating gate separated from the silicon by an oxide (about 150Å thick). Programming and erasing in either type of EEPROM is achieved by forcing the channel current to flow between the gate and substrate with the control biased and negative, respectively.

Both EPROM and EEPROM store charge on a conductive region in the middle of a MOS-gate oxide, and are, therefore, critically dependent on MOS structure, especially high-field carrier transport in both silicon and oxide. Because carriers in the floating island stay there even after the power supply is turned off, EPROM and EEPROM are also nonvolatile memories.

SRAM.   The *random-access memory* (RAM) is one of the largest memories. In a RAM, memory cells are organized in a matrix form, and they can be accessed in random order to read (retrieve) or write (store) data. A *static random-access memory* (SRAM) can retain stored data indefinitely, and it can be implemented as a flip-flop circuit to store one bit of information. Since SRAMs use flip-flop ciruits in the memory cells, they are the most basic of all semiconductor memories. Active devices for access and drive are either MOS transistors or bipolar transistors.

BSRAM. The *bipolar SRAM* (BSRAM) is the fastest of all the semiconductor memory types, but MOS SRAM is the fastest among MOS memories. Figure 6-5-9 shows the equivalent-circuit diagrams for bipolar SRAMs.

(a) Cross-coupled *p-n-p* load cell   (b) Equivalent circuit for bipolar SRAM
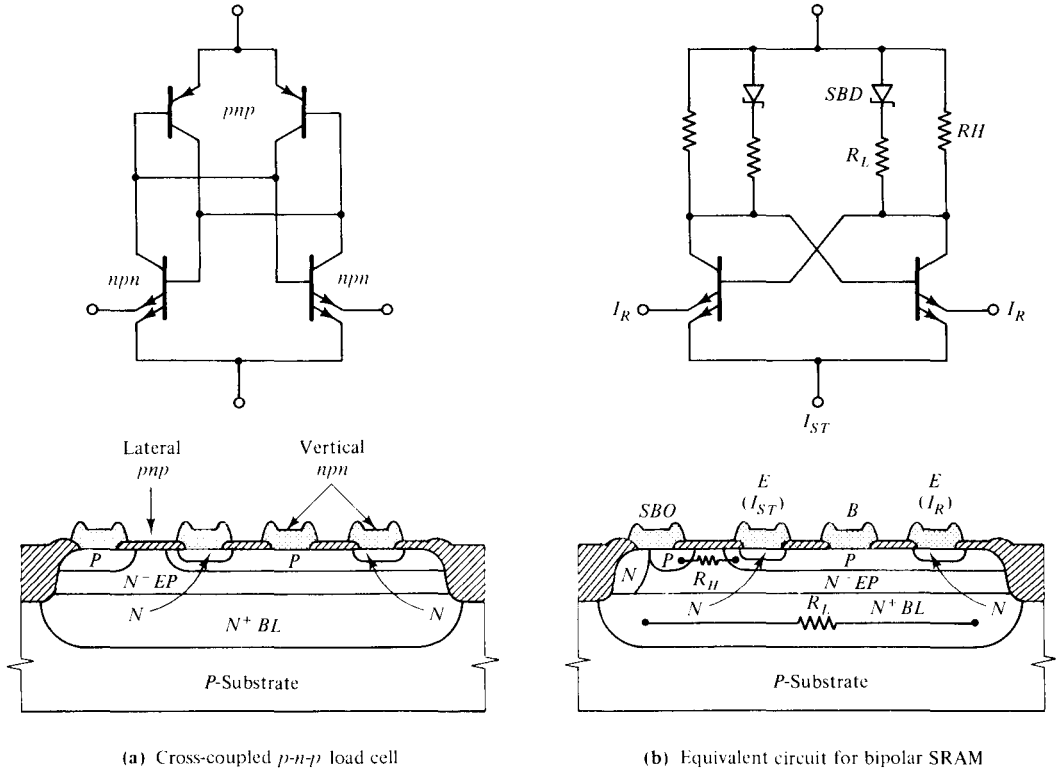
Figure 6-5-9 Circuit diagram for BSRAM.

DRAM. The first *dynamic random-access memory* (DRAM) used one-half of the static memory cell, or three transistors—one for the driver, another for the load, and the third for the access. This device has evolved to today's one-transistor, one-capacitor DRAM cell, in which the single transistor is used as the access to the capacitive reservoir. Since the stored charge is gradually lost because of the space-charge generation-recombination process, cells have to be read and refleshed at predetermined intervals on the order of milliseconds (hence, the name *dynamic*). The DRAM readout signal is small, on the order of 150–200 mV, and is subject to various kinds of noise sources.

6-6 CHARGE-COUPLED DEVICES (CCDs)

The *charge-coupled device* (CCD) is a metal-oxide-semiconductor (MOS) diode structure that was proposed in 1969 by Boyle and Smith [15, 16]. The CCD can move the charges in the MOS diode along a predetermined path under the control of

clock pulses, and so it is also called the charge-transfer device (CTD). CCDs have many microwave applications, such as in infrared detection and imaging and digital signal processing. There are three basic types of CCDs: surface-channel CCD (SCCD), buried-channel CCD (BCCD), and junction CCD (JCCD). In the SCCD or BCCD the charge is stored and transferred at the semiconductor surface or in the semiconductor interior, respectively; whereas in the JCCD the store and transfer of the charge packet occur at the *p-n* junction.

The motion of the charge packets in a charge-coupled device is transversely controlled by the applied gate voltages. This phenomenon is similar to the carrier motion in a microwave field-effect transistor like MESFET or MOSFET. In effect, the CCD can be referred to as the field-effect CCD.

6-6-1 Operational Mechanism

A charge-coupled device (CCD) is an array of many MOS or MIS diodes. In operation, the information (or signal) is stored in the form of electrical charge packets in the potential wells created in a MOS diode. Under the control of externally applied voltage (i.e., gate voltage), the potential wells and the charge packets can be shifted from one well to an adjacent one rapidly through the entire CCD structure.

Three separate mechanisms allow the charge packets to move from one well to another: self-induced drift, thermal diffusion, and fringing field drift. Thermal diffusion results in an exponential decay of the remaining charge under the transferring electrode. The fringing field is the electric field in the direction of charge flow and it can help speed the charge-transfer process considerably. Self-induced drift (or a charge-repulsion effect) is only important at relatively large signal-charge densities. It is the dominant mechanism in the transfer of the first 99% of the charge signal.

Energy band of MIS diode.   A single MIS structure on an *n*-type semiconductor (or *p*-type semiconductor) is the basic element of the CCD. Figure 6-6-1 shows the energy band diagrams for a MIS structure [15].
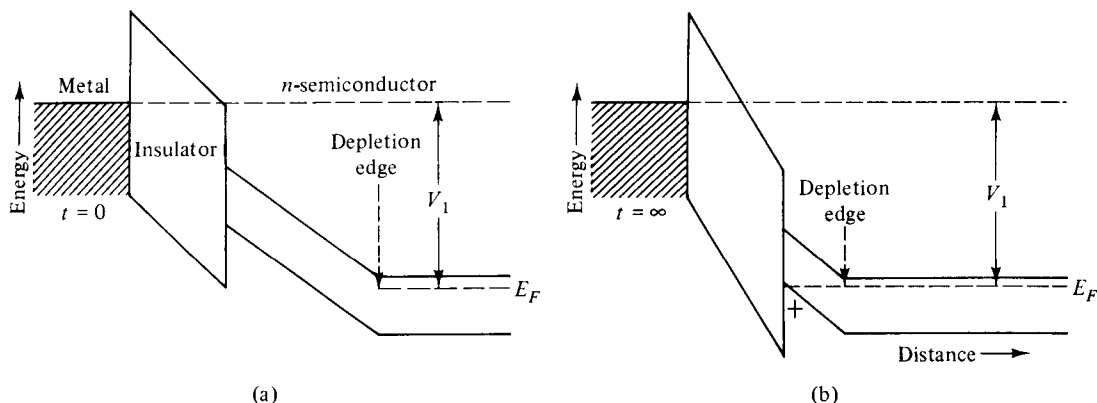
Figure 6-6-1   Energy-band diagrams of MIS structure. (*From W. S. Boyle and G. E. Smith [15]; reprinted with permission from The Bell System, AT&T.*)

The voltage applied to the metal electrode is negative with respect to the semiconductor and large enough to cause depletion. When the voltage is first applied at $t = 0$, there are no holes at the insulator-semiconductor interface [see Fig. 6-6-1(a)]. As holes are introduced into the depletion region, they will accumulate at the interface and cause the surface potential to be more positive [see Fig. 6-6-1(b)].

Three-phase structure.   The CCD can be constructed in the form of a typical three-phase structure as shown in Fig. 6-6-2 [16].

(a)

(b)
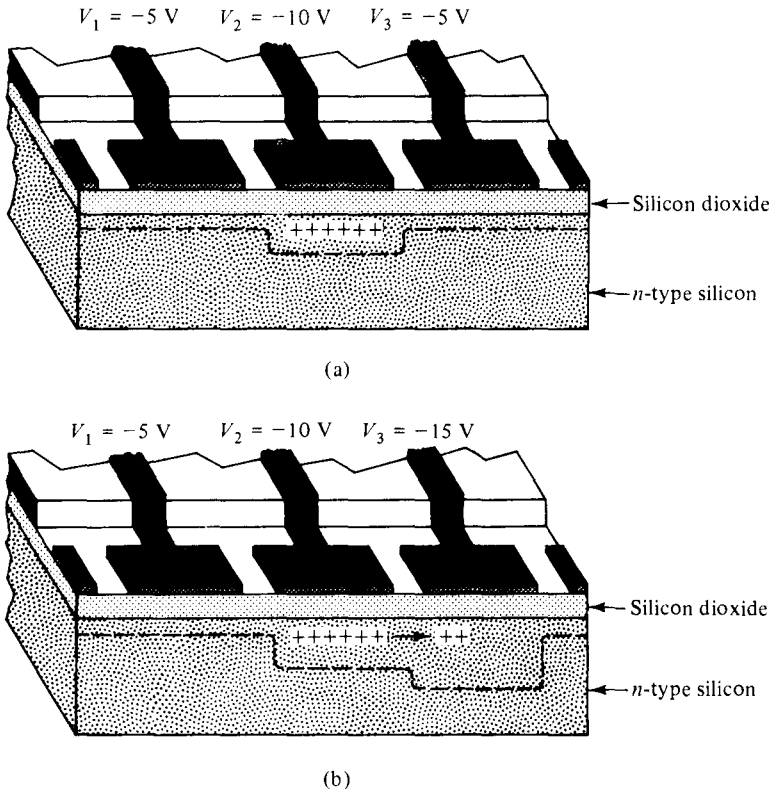
Figure 6-6-2 Cutaway of CCD. (*From W. S. Boyle and G. E. Smith [16]; reprinted by permission of the IEEE, Inc.*)

The CCD consists of a closely spaced array of MIS diodes on an n-type semiconductor substrate with a large negative gate voltage applied. Its basic function is to store and transfer the charge packets from one potential well to an adjacent one. As shown in Fig. 6-6-2(a), $V_1 = V_3$ and V_2 is more negative. In effect, a potential well with stored holes is created at gate electrode 2. The stored charge is temporary because a thermal effect will diffuse the holes out of the wells. Therefore the switching time of the voltage clock must be fast enough to move all charges out of the occupied well to the next empty one. When the voltage V_3 is pulsed to be more negative than the other two voltages V_1 and V_2, the charge begins to transfer to the potential well at gate electrode 3 as shown in Fig. 6-6-2(b).

Store and transfer of charge packets.   A linear array of MIS diodes on an n-type semiconductor is shown in Fig. 6-6-3 [15].
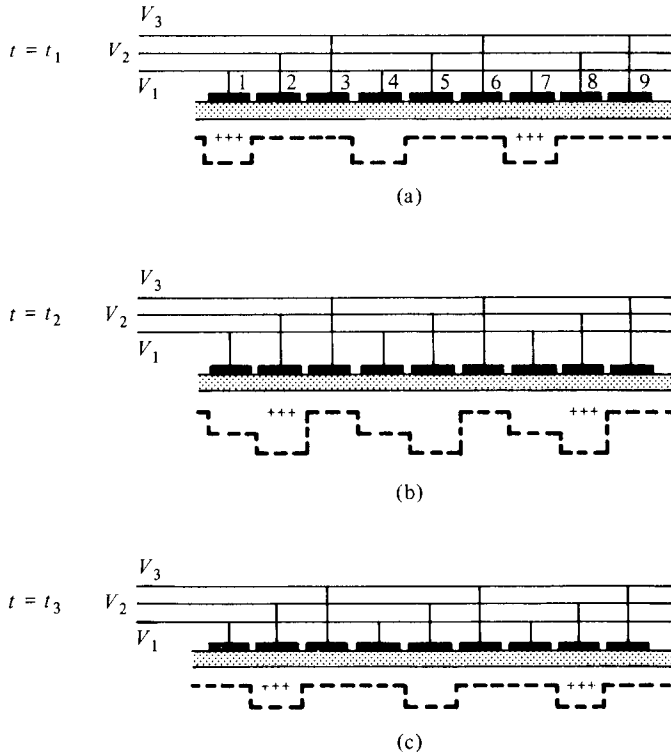
Figure 6-6-3   Store and transfer of charge packets for a three-phase CCD. (*From W. S. Boyle and G. E. Smith [15]; reprinted by permission of The Bell System, AT&T.*)

For a three-phase CCD, every third gate electrode is connected to a common line (see Fig. 6-6-3). At $t = t_1$, a more negative voltage V_1 is applied to gate electrodes 1, 4, 7, and so on, and less negative voltages V_2 and V_3 ($V_2 = V_3$) are applied to the other gate electrodes. It is assumed that the semiconductor substrate is grounded and that the magnitude of V_1 is larger than the threshold voltage V_{th} for the production of inversion under steady-state conditions. As a result, positive charges are stored in the potential wells under electrodes 1, 4, 7, . . . as shown in Fig. 6-6-3(a).

At $t = t_2$, when voltage V_2 at gate electrodes 2, 5, 8, . . . is pulsed to be more negative than V_1 and V_3, the charge packets will be transferred from the potential wells at gate electrodes 1, 4, 7, . . . to the potential minimum under gate electrodes 2, 5, 8, . . . as shown in Fig. 6-6-3(b).

At $t = t_3$, when voltages $V_1 = V_3$ and voltage V_2 remains more negative, the charge packets have been transferred one spatial position and the sequence is ready to continue as shown in Fig. 6-6-3(c).

6-6-2 Surface-Channel Charge-Coupled Devices (SCCDs)

A surface-channel charge-coupled device (SCCD) is basically an MOS-diode structure as shown in Fig. 6-6-4(a) [28]. Its energy band diagrams are illustrated in parts (b) and (c).
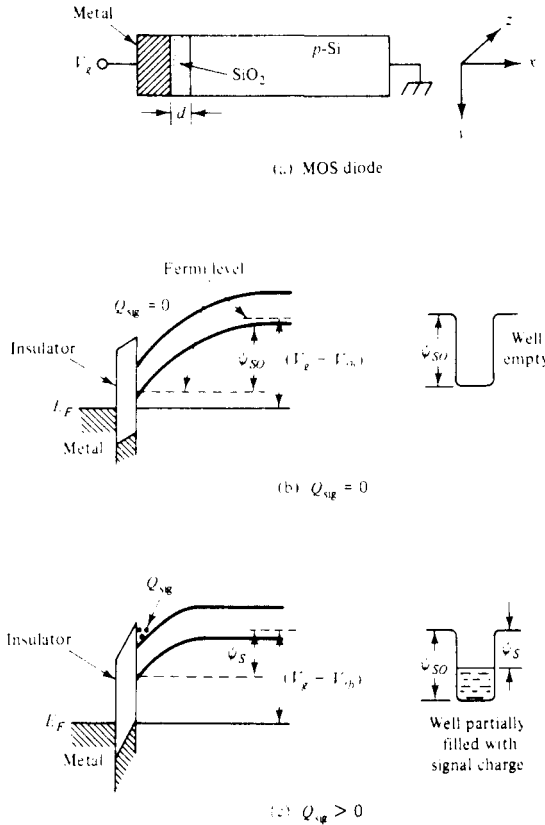
Figure 6-6-4   Physical structure and energy-band diagrams of a BCCD. (*From D. F. Barle [28]; reprinted with permission of the IEEE, Inc.*)

There are two cases for the creation of a charge packet under deep depletion.

1. Zero Signal-Charge ($Q_{sig} = 0$). When the signal charge is zero, an empty well is formed by the potential minimum at the semiconductor surface as shown in Fig. 6-6-4(b). Gate voltage V_g and surface potential ψ_s are related by

$$V_g - V_{fb} = V_i + \psi_s = \frac{qN_aW}{C_i} + \psi_s \tag{6-6-1}$$

where V_{fb} = flatband voltage

V_i = voltage across the insulator

$W = \left(\dfrac{2\epsilon_s \psi_s}{qN_a}\right)^{1/2}$ in the depletion width

Substitution of the depletion width into Eq. (6-6-1) yields

$$V_g - V_{fb} = \frac{1}{C_i}(2\epsilon_s qN_a\psi_s)^{1/2} + \psi_s \qquad (6\text{-}6\text{-}2)$$

2. **Stored Signal-Charge** ($Q_{sig} > 0$). When a signal-charge packet is stored at the semiconductor surface, the surface potential decreases and the potential well is partially filled as shown in Fig. 6-6-4(c). The surface potential equation of Eq. (6-6-2) becomes

$$V_g - V_{fb} = \frac{Q_{sig}}{C_i} + \frac{1}{C_i}(2\epsilon_s qN_a\psi_s)^{1/2} + \psi_s \qquad (6\text{-}6\text{-}3)$$

The maximum charge density (electron or hole) that can be stored on an MOS capacitor is approximately equal to

$$N_{max} \simeq \frac{C_i V_g}{q} \qquad \text{for } V_g \gg 1 \qquad (6\text{-}6\text{-}4)$$

6-6-3 Dynamic Characteristics

The dynamic characteristics can be described in terms of charge-transfer efficiency η, frequency response, and power dissipation.

Charge-transfer efficiency η.   The charge-transfer efficiency is defined as the fraction of charge transferred from one well to the next in a CCD. The fraction left behind is the transfer loss and it is denoted by ϵ. Therefore the charge-transfer efficiency is given by

$$\eta = 1 - \epsilon \qquad (6\text{-}6\text{-}5)$$

If a single charge pulse with an initial amplitude P_0 transfers down a CCD register, after n transfers, the amplitude P_n becomes

$$P_n = P_0\eta^n = P_0(1 - n\epsilon) \qquad \text{for } \epsilon \ll 1 \qquad (6\text{-}6\text{-}6)$$

where n equals the number of transfers or phases.

If many transfers are required, the transfer loss ϵ must be very small. For example, if a transfer efficiency of 99.99% is required for a three-phase, 330-stage shift register, the transfer loss must be less than 0.01%. The maximum achievable transfer efficiency depends on two factors: how fast the free charge can be transferred between adjacent gates and how much of the charge gets trapped at each gate location by stationary states.

Frequency response.   There are, in fact, upper and lower frequency limitations for CCDs. The potential well will not remain indefinitely, and thermally generated electrons (or holes) eventually fill the well completely. Also, the time stored by the charge must be much shorter than the thermal relaxation time of the CCD's capacitor. So the maximum frequency is limited by the channel length L.

Power dissipation.   The power dissipation per bit is given by

$$P = nfVQ_{max} \tag{6-6-7}$$

Example 6-6-1:   Power Dissipation of a Three-Phase CCD

A three-phase CCD is operating under the following conditions:

Applied voltage:	$V = 10$ volts
Number of phases:	$n = 3$
Maximum stored charges:	$Q_{max} = 0.04$ pC
Clock frequency:	$f = 10$ MHz

Determine the power dissipation per bit.

Solution   From Eq. (6-6-7) the power dissipation per bit is

$$P = nfVQ_{max} = 3 \times 10^7 \times 10 \times 0.04 \times 10^{-12}$$

$$= 12 \ \mu W$$

Charge-coupled devices (CCDs) have many microwave applications in electronic components and systems, such as in infrared systems and signal processing.

Infrared detection and imaging.   Because varying amounts of charge corresponding to information can be introduced into the potential wells at one end of the CCD structure to emerge after some delay at the other end, the CCD is capable of detecting and imaging the infrared light from a target [17]. Ten years ago, when the CCD was first developed, only an array of 12×12 photodetectors could be used to detect the infrared images of a target. Because of the availability of greatly improved CCDs, today an array of photodetectors, such as Indium Antimonide (InSb) 128×128, are used to form charge packets that are proportional to the light intensity of a target; and these packets are shifted to a detector point for detection, readout, multiplexing, and time delay and integration (TDI). In scanned IR systems TDI is one of the most important functions performed by CCDs. In such a system the scene is mechanically scanned across an array of detector elements. By using CCD columns to shift the detector output signals (in the form of charge packets) along the focal plane with the same speed as the mechanical scan moves the scene across the array, the signal-to-noise ratio can be improved by the square root of the number of detector elements in the TDI column.

Signal processing.   The CCD can perform several analog and digital signal processing functions, such as delay, multiplexing, demultiplexing, transversal

filtering, recursive filtering, integration, analog memory, digital memory, and digital logic. Thus CCDs are being used widely in special applications for the very large scale integration (VLSI) circuits.

Example 6-6-2:   Design of an N-Type Three-Phase Surface-Channel CCD

The n-type three-phase surface-channel CCD has the following specifications:

Electron density:	$N_{max} = 2 \times 10^{12}$ cm^{-2}
Insulator relative dielectric constant:	$\epsilon_{ir} = 3.9$
Insulator thickness:	$d = 0.15$ μm
Insulator cross section:	$A = 0.5 \times 10^{-4}$ cm^2
Power dissipation allowable per bit:	$P = 0.67$ mW

a. Compute the insulator capacitance in farads per square centimeter.
b. Determine the maximum stored charges per well in coulombs.
c. Find the required applied gate voltage in volts.
d. Choose the clock frequency in megahertz.

Solution

a. The insulator capacitance is

$$C_i = \frac{\epsilon_i}{d} = \frac{3.9 \times 8.854 \times 10^{-12}}{0.15 \times 10^{-6}} = 23 \text{ nF/cm}^2$$

b. The maximum stored charges per well is

$$Q_{max} = N_{max}qA = 2 \times 10^{12} \times 1.6 \times 10^{-19} \times 0.5 \times 10^{-4}$$
$$= 16 \text{ pC}$$

c. From Eq. (6-6-4) the required applied gate voltage is

$$V_g = \frac{N_{max}q}{C_i} = \frac{2 \times 10^{12} \times 1.6 \times 10^{-19}}{23 \times 10^{-9}} = 14 \text{ V}$$

d. From Eq. (6-6-7) the clock frequency is

$$f = \frac{P}{nVQ_{max}} = \frac{0.67 \times 10^{-3}}{3 \times 14 \times 16 \times 10^{-12}} = 1 \text{ MHz}$$

REFERENCES

[1] SHOCKLEY, W., A unipolar "field-effect" transistor. *Proc. IRE,* **40,** No. 11, 1365–1376, November 1952.

[2] SCHOTTKY, W., Naturwiss, 26, 843, (1938).

[3] MEAD, C. A., Schottky barrier gate field-effect transistor. *Proc. IEEE,* **54,** No. 2, 307–308, February 1966.

[4] Hooper, W. W., et al, An epitaxial GaAs field-effect transistor. *Proc. IEEE,* **55,** No. 7, 1237–1238, July 1967.

[5] VanDerZiel, A. et al., Gate noise in field-effect transistors at moderately high frequencies. *Proc. IEEE,* **51,** No. 3, 461–467, March 1963.

[6] Zuleeg, R., and K. Lehovec, High frequency and temperature characteristics of GaAs junction field-effect transistors in hot electron range. *Proc. Symp. GaAs,* Institute of Physics Conf. Series No. 9, 240–245, 1970.

[7] Lehovec, K., and R. Zuleeg, Voltage-current characteristics of GaAs J-FET's in the hot electron range. *Solid-State Electronics,* Vol. 13, 1415–1426. Pergamon Press, London, 1970.

[8] Dingle, R., et al., Electron mobilities in modulation-doped semiconductor heterojunction superlattices. *Appl. Phys. Letters,* Vol. 33, 665–667 (1978).

[9] Abe, M., et al., Recent advance in ultra-light speed HEMT technology. *IEEE J. Quantum Electronics.* **QE-22,** No. 9, 1870–1879, September 1986.

[10] Togashi, K., et al., Reliability of low-noise microwave HEMTs made by MOCVD. *Microwave J.,* 123–132, April 1987.

[11] Pavlidis, D., and M. Weiss, The influence of device physical parameters on HEMT large-signal characteristics. *IEEE Trans. on Microwave Theory and Techniques,* **MTT-36,** No. 2, 239–249, February 1988.

[12] Sze, S. M., *Physics of Semiconductor Devices,* 2nd ed., p. 440, John Wiley & Sons, New York, 1981.

[13] Parrillo, L. C., *VLSI Process Integration in VLSI Technology.* Chapter 11, ed. S. M. Sze. McGraw-Hill Book Company, New York, 1983.

[14] Asai, S., Semiconductor memory trend. *Proc. IEEE,* **74,** No. 12, 1623–1635, December 1986.

[15] Boyle, W. S., and G. E. Smith, Charge couple semiconductor devices. *Bell Syst. Tech. J.,* **49,** 587–593 (1977).

[16] Boyle, W. S., and G. E. Smith, Charge-coupled devices—a new approach to MIS device structures. *IEEE Spectrum,* **8,** No. 7, 18027, July 1971.

[17] Steckl, A. J., et al., Application of charge-coupled devices to infrared detection and imaging. *Proc. IEEE,* **63,** No. 1, 67–74, January 1975.

SUGGESTED READINGS

Hess, K., *Advanced Theory of Semiconductor Devices.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.

IEEE Transactions on Electron Devices. Special issues on microwave solid-state devices.
ED-27, No. 2, February 1980.
ED-27, No. 6, June 1980.
ED-28, No. 2, February 1981.
ED-28, No. 8, August 1981.

IEEE Transactions on Microwave Theory and Techniques. Special issues on microwave solid-state devices.
MTT-21, No. 11, November 1973.
MTT-24, No. 11, November 1976.

MTT-27, No. 5, May 1979.

MTT-28, No. 12, December 1980.

MTT-30, No. 4, April 1982.

MTT-30, No. 10, October 1982.

LIAO, S. Y., *Microwave Solid-State Devices*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1985.

LIAO, S. Y., *Semiconductor Electronic Devices*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1990.

NAVON, D. H., *Semiconductor Microdevices and Materials*. Holt, Rinehart and Winston, New York 1986.

PULFREY, DAVID L., and N. GARRY TARR., *Introduction to Microelectronic Devices*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1989.

PROBLEMS

JFETs

6-1. A silicon JFET has a gate length of 5 μm. Calculate the maximum frequency of oscillation for the device.

6-2. An n-channel silicon JFET has the following parameters:

Electron density:	$N_d = 2 \times 10^{17}$ cm^{-3}
Channel height:	$a = 0.4 \times 10^{-4}$ cm
Relative dielectric constant:	$\epsilon_r = 11.8$

Compute the pinch-off voltage.

6-3. An n-channel silicon JFET at 300° K has the following parameters:

Electron density:	$N_d = 5 \times 10^{17}$ cm^{-3}
Hole density:	$N_a = 8 \times 10^{18}$ cm^{-3}
Dielectric constant:	$\epsilon_r = 11.8$
Channel height:	$a = 0.3 \times 10^{-4}$ cm
Channel length:	$L = 6 \times 10^{-4}$ cm
Channel width:	$Z = 45 \times 10^{-4}$ cm
Electron mobility:	$\mu_n = 520$ cm^2/V·s
Drain voltage:	$V_d = 8$ V
Gate voltage:	$V_g = -1$ V

Compute:
a. The pinch-off voltage in volts
b. The pinch-off current in mA
c. The built-in voltage in V
d. The drain current in mA
e. The saturation drain current at $V_g = -1$ V in mA
f. The cut-off frequency in GHz

6-4. An n-channel silicon JFET at $300°$ K has the following parameters:

Electron density:	$N_d = 3 \times 10^{17}$ cm^{-3}
Hole density:	$N_a = 2 \times 10^{19}$ cm^{-3}
Relative dielectric constant:	$\epsilon_r = 11.8$
Channel height:	$a = 0.5 \times 10^{-4}$ cm
Channel length:	$L = 10 \times 10^{-4}$ cm
Channel width:	$Z = 20 \times 10^{-4}$ cm
Drain voltage:	$V_d = 7$ V
Gate voltage:	$V_g = -1.5$ V

Determine:
a. The electron mobility μ_n
b. The pinch-off voltage in V
c. The pinch-off current in mA
d. The drain current in mA
e. The saturation current in mA at $V_g = 0$
f. The cutoff frequency in GHz

6-5. Verify Eq. (6-1-17) from Eq. (6-1-12).

6-6. Write a FORTRAN program to compute the drain current of an n-channel silicon JFET. The device parameters are as follows:

Electron concentration:	$N_d = 2 \times 10^{23}$ m^{-3}
Channel height:	$a = 0.2 \times 10^{-6}$ m
Relative dielectric constant:	$\epsilon_r = 11.8$
Channel length:	$L = 10 \times 10^{-6}$ m
Channel width:	$Z = 60 \times 10^{-6}$ m
Drain voltage:	$V_d = 0$ to 6 V
Gate voltage:	$V_g = 0$ to -3 V
Saturation drift velocity:	$v_s = 10^5$ m/s

The program specifications are

1. The drain voltage V_d varies from 0 to 6 volts with an increment of 1 volt per step
2. The gate voltage V_g varies from 0 to -3 volts with a decrease of -0.5 volt per step
3. The electron mobility μ varies from 0.9 to 0.3 m^2/V·s with a decrease of 0.1 per step
4. Use F10.5 format for numerical outputs and Hollerith format for character outputs
5. Print the outputs in three columns for drain voltage V_d (volts), gate voltage V_g (volts), and drain current I_d (mA)

MESFETs

6-7. A MESFET has a gate width of 5 μm. Calculate the maximum frequency of oscillation for the device.

6-8. A GaAs has a thickness of 0.40 μm and a doping concentration N_d of 5×10^{17} cm^{-3}. The relative dielectric constant ϵ_r of GaAs is 13.10. Calculate the pinch-off voltage in volts.

6-9. An n-channel GaAs MESFET at 300° K has the following parameters:

Donor density:	$N_d = 2 \times 10^{17}$ cm^{-3}
Relative dielectric constant:	$\epsilon_r = 13.10$
Channel height:	$a = 0.3 \times 10^{-4}$ cm
Channel length:	$L = 10 \times 10^{-4}$ cm
Channel width:	$Z = 50 \times 10^{-4}$ cm
Drain voltage:	$V_d = 7$ V
Gate voltage:	$V_g = -1.5$ V
Saturation drift velocity:	$v_s = 10^{-7}$ cm/s

Compute:
a. The electron mobility (read from Fig. A-2 in Appendix A)
b. The pinch-off voltage in V
c. The saturation current at $V_g = 0$
d. The drain current I_d in mA

6-10. An n-channel GaAs MESFET at 300° K has the following parameters:

Donor density:	$N_d = 3 \times 10^{17}$ cm^{-3}
Relative dielectric constant:	$\epsilon_r = 13.10$
Channel height:	$a = 0.5 \times 10^{-4}$ cm
Channel length:	$L = 8 \times 10^{-4}$ cm
Channel width:	$Z = 60 \times 10^{-4}$ cm
Drain voltage:	$V_d = 6$ V
Gate voltage:	$V_g = -2.0$ V
Saturation drift velocity:	$v_s = 1 \times 10^7$ cm/s

Calculate:
a. The electron mobility (read from Fig. A-2 in Appendix A)
b. The pinch-off voltage
c. The saturation current at $V_g = 0$
d. The drain current I_d in mA

6-11. A certain n-channel GaAs MESFET has the following parameters:

Electron concentration:	$N = 2.38 \times 10^{23}$ m^{-3}
Channel height:	$a = 0.2$ μm
Relative dielectric constant:	$\epsilon_r = 13.10$
Channel length:	$L = 10$ μm
Channel width:	$Z = 60$ μm
Electron mobility:	$\mu = 0.3$ m^2/V $\cdot$ s
Drain voltage:	$V_{ds} = 6$ volts
Gate voltage:	$V_g = -3$ volts
Saturation drift velocity:	$v_s = 10^5$ m/s

a. Calculate the pinch-off voltage.
b. Compute the velocity ratio.
c. Determine the saturation drain current at $V_g = 0$.
d. Find the drain current I_d.

HEMTs

6-12. Describe the structure of a HEMT and its fabrication processes.

6-13. Describe the operating principles of a HEMT.

6-14. A HEMT has the following parameters:

Gate width:	$W = 140$ m
Electron velocity:	$v(z) = 3 \times 10^5$ m/s
2-D electron-gas density:	$n(z) = 6 \times 10^{15}$ m^{-2}

Compute the drain current of the HEMT.

6-15. A HEMT has the following parameters:

Gate width:	$W = 100$ m
Electron velocity:	$v(z) = 4 \times 10^5$ m/s
2-D electron-gas density:	$n(z) = 4 \times 10^{15}$ m^{-2}

Calculate the drain current of the HEMT.

6-16. A HEMT has the following parameters:

Threshold voltage:	$V_{th} = 0.15$ V
Donor concentration:	$N_d = 3 \times 10^{24}$ m^{-3}
Metal-semiconductor	$\psi_{ms} = 0.8$ V
Schottky barrier potential:	
AlGaAs dielectric constant:	$\epsilon_r = 4.43$

Compute:
a. The conduction–band-edge difference between GaAs and AlGaAs
b. The sensitivity of the HEMT

6-17. A HEMT has the following parameters:

Threshold voltage:	$V_{th} = 0.12$ V
Donor concentration:	$N_d = 5 \times 10^{24}$ m^{-3}
AlGaAs dielectric constant:	$\epsilon_r = 4.43$
Metal-semiconductor	
Schottky barrier potential:	$\psi_{ms} = 0.8$

Calculate:
a. The conduction–band-edge difference between GaAs and AlGaAs
b. The sensitivity of the HEMT

6-18. Describe the applications of a HEMT.

MOSFETs

6-19. A basic MOSFET is formed of Al metal, SiO_2 insulator, and Si semiconductor. The insulator capacitance is 4 pF and the channel length L is 12 μm.
  a. Determine the drain current $I_{d\,sat}$ in the saturation region with $V_{th} = 2V$ and $V_{gs} = 4V$ for the enhancement mode.
  b. Compute the transconductance g_m for the same mode.
  c. Calculate the drain current $I_{d\,sat}$ in the saturation region with $V_{th} = -1.5V$ and $V_{gs} = 0V$ for the depletion mode.
  d. Find the transconductance g_m for the depletion mode.

6-20. An Al-Oxide-Si MOSFET has an insulator capacitance of 3 pF and a channel length L of 10 μm. The gate voltage V_{gs} is 10 V and the threshold voltage is 1.5 V.
  a. Determine the carrier drift velocity in real case.
  b. Calculate the drain current.
  c. Compute the carrier transit time.
  d. Find the maximum operating frequency in GHz.

6-21. The insulator SiO_2 in a Si MOSFET has a relative dielectric constant ϵ_{ir} of 11.8 and a depth d of 0.08 μm. The channel length L is 15 μm and the channel depth Z is 150 μm. Calculate the insulator capacitance C_{in}.

6-22. Compare the advantages and disadvantages of a GaAs MOSFET with those of a Si MOSFET.

6-23. A certain p-channel MOSFET has the following parameters:

Doping concentration:	$N_a = 2 \times 10^{17} \text{ cm}^{-3}$
Relative dielectric constant:	$\epsilon_r = 11.8$
Insulator relative dielectric constant:	$\epsilon_{ir} = 4$
Insulator thickness:	$d = 0.01 \ \mu$m
Operating temperature:	$T = 320° \text{ K}$

  a. Calculate the surface potential $\psi_s(\text{inv})$ for strong inversion.
  b. Compute the insulator capacitance.
  c. Determine the threshold voltage.

6-24. A certain n-channel MOSFET has the following parameters:

Channel length:	$L = 5 \ \mu$m
Channel depth:	$Z = 10 \ \mu$m
Insulator thickness:	$d = 0.02 \ \mu$m
Gate voltage:	$V_g = 8$ V
Threshold voltage:	$V_{th} = 1.5$ V
Electron velocity:	$v_s = 2 \times 10^7$ cm/s
Insulator relative dielectric constant:	$\epsilon_r = 4$

  a. Compute the insulator capacitance in mF/m^2.
  b. Calculate the saturation drain current in mA.

c. Determine the saturation transconductance in millimhos.

d. Estimate the maximum saturation operating frequency in GHz.

NMOS, CMOS, and Memory Devices

6-25. Describe the structures of NMOS, PMOS, and CMOS devices.

6-26. Explain the operational principles of the NMOS, PMOS, and CMOS devices.

6-27. Discuss the logic-gate operations of NMOS, PMOS, and CMOS.

6-28. Describe the major memory devices.

CCDs

6-29. A charge-coupled device has 484 (or 22 × 22 array) elements each with a transfer inefficiency of 10^{-4} and is clocked at a frequency of 100 KHz.

a. Determine the delay time between input and output.

b. Find the percentage of input charge appearing at the output terminal.

6-30. A surface-channel CCD is operated by a gate voltage V_g of 10 volts. The insulator has a relative dielectric constant ϵ_r of 6 and a depth d of 0.1 μm. Determine the stored charge density on this MOS capacitor.

6-31. A p-type surface-channel CCD (Al-Oxide-Si) has the following parameters:

$$N_a = 10^{14} \text{ cm}^{-3} \qquad Q_{in} = 10^{-8} \text{ C/cm}^2$$
$$\phi_{ms} = -0.85 \text{ eV} \qquad W = 1 \text{ } \mu\text{m}$$
$$\epsilon_{ir} = 4 \qquad d = 0.1 \text{ } \mu\text{m}$$

a. Determine the flat-band potential.

b. Calculate the voltage across the insulator.

6-32. A three-phase CCD has $Q_{max} = 0.06$ pC operating at a clock frequency of 20 MHz with 10 volts applied. Determine the power dissipation per bit.

6-33. A three-phase JCCD has a length L of 8 μm, an electric field E_x of 480 V/cm, and an electron mobility μ_n of 1000 cm²/V · s. Calculate the signal carrier transit time.

6-34. An n-type three-phase SCCD has the following parameters:

Insulator relative dielectric constant:	$\epsilon_{ir} = 4$
Insulator thickness:	$d = 0.3$ μm
Insulator cross section:	$A = 0.6 \times 10^{-4}$ cm²
Electron density:	$N_{max} = 4 \times 10^{12}$ cm^{-2}
Power dissipation allowable per bit:	$P = 0.8$ mW

a. Compute the insulator capacitance in F/cm².

b. Calculate the maximum stored charges per well in coulombs.

c. Find the required applied gate voltage in volts.

d. Choose the clock frequency in MHz.

Chapter 7

Transferred Electron Devices (TEDs)

7-0 INTRODUCTION

The application of two-terminal semiconductor devices at microwave frequencies has been increased usage during the past decades. The CW, average, and peak power outputs of these devices at higher microwave frequencies are much larger than those obtainable with the best power transistor. The common characteristic of all active two-terminal solid-state devices is their negative resistance. The real part of their impedance is negative over a range of frequencies. In a positive resistance the current through the resistance and the voltage across it are in phase. The voltage drop across a positive resistance is positive and a power of $(I^2 R)$ is dissipated in the resistance. In a negative resistance, however, the current and voltage are out of phase by 180°. The voltage drop across a negative resistance is negative, and a power of $(-I^2 R)$ is generated by the power supply associated with the negative resistance. In other words, positive resistances absorb power (passive devices), whereas negative resistances generate power (active devices). In this chapter the transferred electron devices (TEDs) are analyzed.

  The differences between microwave transistors and transferred electron devices (TEDs) are fundamental. Transistors operate with either junctions or gates, but TEDs are bulk devices having no junctions or gates. The majority of transistors are fabricated from elemental semiconductors, such as silicon or germanium, whereas TEDs are fabricated from compound semiconductors, such as gallium arsenide (GaAs), indium phosphide (InP), or cadmium telluride (CdTe). Transistors operate with "warm" electrons whose energy is not much greater than the thermal energy (0.026 eV at room temperature) of electrons in the semiconductor, whereas TEDs

operate with "hot" electrons whose energy is very much greater than the thermal energy. Because of these fundamental differences, the theory and technology of transistors cannot be applied to TEDs.

7-1 GUNN-EFFECT DIODES—GaAs DIODE

Gunn-effect diodes are named after J. B. Gunn, who in 1963 discovered a periodic fluctuations of current passing through the n-type gallium arsenide (GaAs) specimen when the applied voltage exceeded a certain critical value. Two years later, in 1965, B. C. DeLoach, R. C. Johnston, and B. G. Cohen discovered the impact ionization avalanche transit-time (IMPATT) mechanism in silicon, which employs the avalanching and transit-time properties of the diode to generate microwave frequencies. In later years the limited space-charge-accumulation diode (LSA diode) and the indium phosphide diode (InP diode) were also successfully developed. These are bulk devices in the sense that microwave amplification and oscillation are derived from the bulk negative-resistance property of uniform semiconductors rather than from the junction negative-resistance property between two different semiconductors, as in the tunnel diode.

7-1-1 Background

After inventing the transistor, Shockley suggested in 1954 that two-terminal negative-resistance devices using semiconductors may have advantages over transistors at high frequencies [1]. In 1961 Ridley and Watkins described a new method for obtaining negative differential mobility in semiconductors [2]. The principle involved is to heat carriers in a light-mass, high-mobility subband with an electric field so that the carriers can transfer to a heavy-mass, low-mobility, higher-energy subband when they have a high enough temperature. Ridley and Watkins also mentioned that Ge–Si alloys and some III–V compounds may have suitable subband structures in the conduction bands. Their theory for achieving negative differential mobility in bulk semiconductors by transferring electrons from high-mobility energy bands to low-mobility energy bands was taken a step further by Hilsum in 1962 [3]. Hilsum carefully calculated the transferred electron effect in several III–V compounds and was the first to use the terms transferred electron amplifiers (TEAs) and oscillators (TEOs). He predicted accurately that a TEA bar of semi-insulating GaAs would be operated at 373°K at a field of 3200 V/cm. Hilsum's attempts to verify his theory experimentally failed because the GaAs diode available to him at that time was not of sufficiently high quality.

It was not until 1963 that J. B. Gunn of IBM discovered the so-called Gunn effect from thin disks of n-type GaAs and n-type InP specimens while studying the noise properties of semiconductors [4]. Gunn did not connect—and even immediately rejected—his discoveries with the theories of Ridley, Watkins, and Hilsum. In 1963 Ridley predicted [5] that the field domain is continually moving down through the crystal, disappearing at the anode and then reappearing at a favored nucleating

center, and starting the whole cycle once more. Finally, Kroemer stated [6] that the origin of the negative differential mobility is Ridley–Watkins–Hilsum's mechanism of electron transfer into the satellite valleys that occur in the conduction bands of both the n-type GaAs and the n-type InP and that the properties of the Gunn effect are the current oscillations caused by the periodic nucleation and disappearance of traveling space-charge instability domains. Thus the correlation of theoretical predictions and experimental discoveries completed the theory of transferred electron devices (TEDs).

7-1-2 Gunn Effect

A schematic diagram of a uniform n-type GaAs diode with ohmic contacts at the end surfaces is shown in Fig. 7-1-1. J. B. Gunn observed the Gunn effect in the n-type GaAs bulk diode in 1963, an effect best explained by Gunn himself, who published several papers about his observations [7 to 9]. He stated in his first paper [7] that

> Above some critical voltage, corresponding to an electric field of 2000–4000 volts/cm, the current in every specimen became a fluctuating function of time. In the GaAs specimens, this fluctuation took the form of a periodic oscillation superimposed upon the pulse current. . . . The frequency of oscillation was determined mainly by the specimen, and not by the external circuit. . . . The period of oscillation was usually inversely proportional to the specimen length and closely equal to the transit time of electrons between the electrodes, calculated from their estimated velocity of slightly over 10^7 cm/s. . . . The peak pulse microwave power delivered by the GaAs specimens to a matched load was measured. Value as high as 0.5 W at 1 Gc/s, and 0.15 W at 3 Gc/s, were found, corresponding to 1–2% of the pulse input power.*
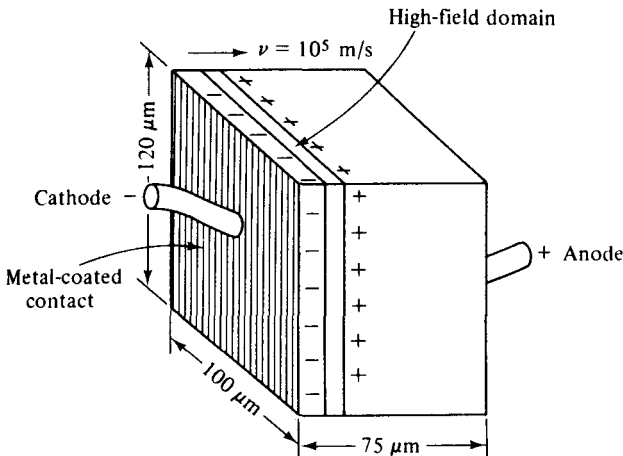
Figure 7-1-1 Schematic diagram for n-type GaAs diode.

*After J. B. Gunn [7]; reproduced by permission of IBM, Inc.

From Gunn's observation the carrier drift velocity is linearly increased from zero to a maximum when the electric field is varied from zero to a threshold value. When the electric field is beyond the threshold value of 3000 V/cm for the n-type GaAs, the drift velocity is decreased and the diode exhibits negative resistance. This situation is shown in Fig. 7-1-2.
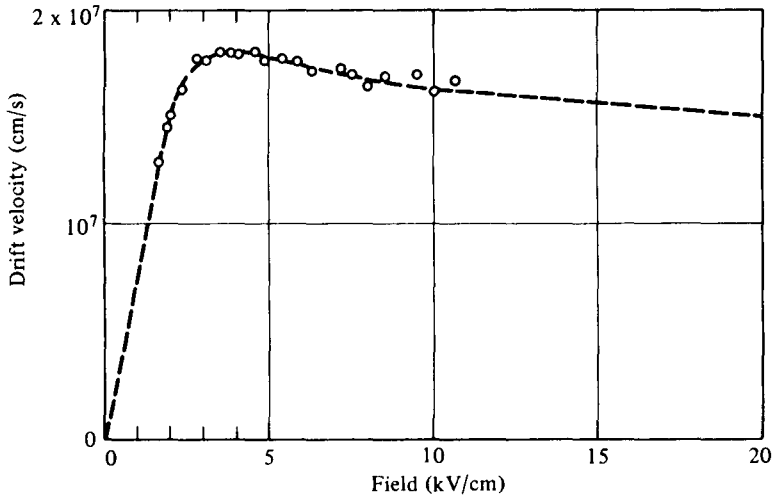
Figure 7-1-2 Drift velocity of electrons in n-type GaAs versus electric field. (*After J. B. Gunn [8]; reprinted by permission of IBM, Inc.*)

The current fluctuations are shown in Fig. 7-1-3. The current waveform was produced by applying a voltage pulse of 16-V amplitude and 10-ns duration to a specimen of n-type GaAs 2.5×10^{-3} cm in length. The oscillation frequency was 4.5 GHz. The lower trace had 2 ns/cm in the horizontal axis and 0.23 A/cm in the vertical axis. The upper trace was the expanded view of the lower trace. Gunn found that the period of these oscillations was equal to the transit time of the electrons through the specimen calculated from the threshold current.
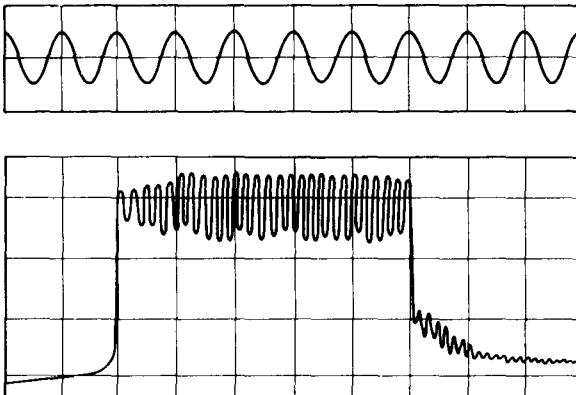
Figure 7-1-3 Current waveform of n-type GaAs reported by Gunn. (*After J. B. Gunn [8]; reprinted by permission of IBM, Inc.*)

Gunn also discovered that the threshold electric field E_{th} varied with the length and type of material. He developed an elaborate capacitive probe for plotting the electric field distribution within a specimen of n-type GaAs of length $L = 210\ \mu m$ and cross-sectional area $3.5 \times 10^{-3}\ cm^2$ with a low-field resistance of $16\ \Omega$. Current instabilities occurred at specimen voltages above 59 V, which means that the threshold field is

$$E_{th} = \frac{V}{L} = \frac{59}{210 \times 10^{-6} \times 10^2} = 2810 \text{ volts/cm} \qquad (7\text{-}1\text{-}1)$$

7-2 RIDLEY–WATKINS–HILSUM (RWH) THEORY

Many explanations have been offered for the Gunn effect. In 1964 Kroemer [6] suggested that Gunn's observations were in complete agreement with the Ridley–Watkins–Hilsum (RWH) theory.

7-2-1 Differential Negative Resistance

The fundamental concept of the Ridley–Watkins–Hilsum (RWH) theory is the differential negative resistance developed in a bulk solid-state III-V compound when either a voltage (or electric field) or a current is applied to the terminals of the sample. There are two modes of negative-resistance devices: voltage-controlled and current-controlled modes as shown in Fig. 7-2-1(a) and (b), respectively [5].

(a) Voltage-controlled mode   (b) Current-controlled mode
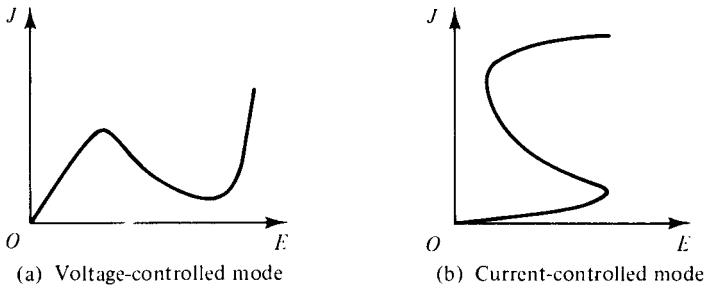
Figure 7-2-1   Diagram of negative resistance. (*From B. K. Ridley [5]; reprinted by permission of the Institute of Physics.*)

In the voltage-controlled mode the current density can be multivalued, whereas in the current-controlled mode the voltage can be multivalued. The major effect of the appearance of a differential negative-resistance region in the current-density-field curve is to render the sample electrically unstable. As a result, the initially homogeneous sample becomes electrically heterogeneous in an attempt to reach stability. In the voltage-controlled negative-resistance mode high-field domains are formed, separating two low-field regions. The interfaces separating low- and high-field domains lie along equipotentials; thus they are in planes perpendicular to the current direction as shown in Fig. 7-2-2(a). In the current-controlled negative-resistance mode splitting the sample results in high-current filaments running along the field direction as shown in Fig. 7-2-2(b).
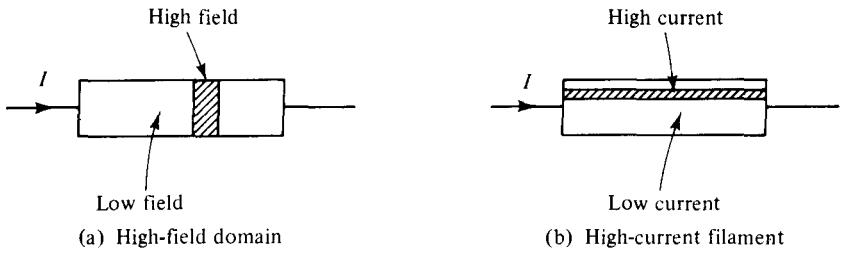
Figure 7-2-2   Diagrams of high field domain and high current filament. (*From B. K. Ridley [5]; reprinted by permission of the Institute of Physics.*)

Expressed mathematically, the negative resistance of the sample at a particular region is

$$\frac{dI}{dV} = \frac{dJ}{dE} = \text{negative resistance} \qquad (7\text{-}2\text{-}1)$$

If an electric field E_0 (or voltage V_0) is applied to the sample, for example, the current density J_0 is generated. As the applied field (or voltage) is increased to E_2 (or V_2), the current density is decreased to J_2. When the field (or voltage) is decreased to E_1 (or V_1), the current density is increased to J_1. These phenomena of the voltage-controlled negative resistance are shown in Fig. 7-2-3(a). Similarly, for the current-controlled mode, the negative-resistance profile is as shown in Fig. 7-2-3(b).
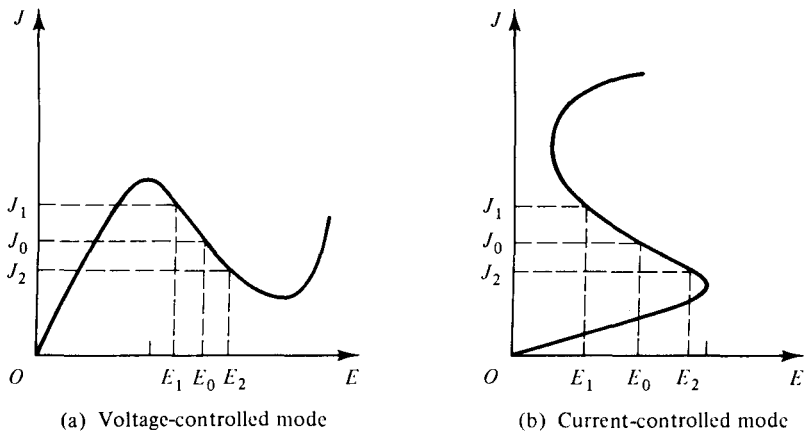
Figure 7-2-3   Multiple values of current density for negative resistance. (*From B. K. Ridley [5]; reprinted by permission of the Institute of Physics.*)

7-2-2 Two-Valley Model Theory

A few years before the Gunn effect was discovered, Kroemer proposed a negative-mass microwave amplifier in 1958 [10] and 1959 [11]. According to the energy band theory of the n-type GaAs, a high-mobility lower valley is separated by an energy of 0.36 eV from a low-mobility upper valley as shown in Fig. 7-2-4. Table 7-2-1 lists

Upper valley

$m_{eu} = 1.2$
$\mu_u = 180 \text{ cm}^2/V \cdot s$

Lower valley

$m_{e\ell} = 0.068$
$\mu_\ell = 8000 \text{ cm}^2/V \cdot s$

$\Delta E = 0.36 \text{ eV}$

Conduction band

Forbidden band

$E_g = 1.43 \text{ eV}$

K

$< 100 >$

Valence band
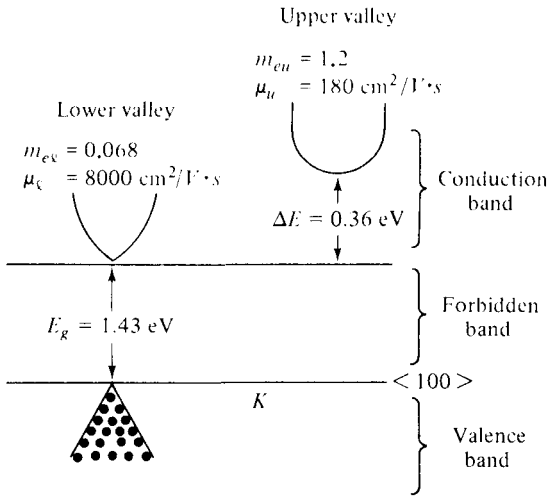
Figure 7-2-4 Two-valley model of electron energy versus wave number for n-type GaAs.

TABLE 7-2-1 DATA FOR TWO VALLEYS IN GaAs

Valley	Effective Mass M_e	Mobility μ	Separation ΔE
Lower	$M_{e\ell} = 0.068$	$\mu_\ell = 8000 \text{ cm}^2/V\text{-sec}$	$\Delta E = 0.36 \text{ eV}$
Upper	$M_{eu} = 1.2$	$\mu_u = 180 \text{ cm}^2/V\text{-sec}$	$\Delta E = 0.36 \text{ eV}$

the data for the two valleys in the n-type GaAs and Table 7-2-2 shows the data for two-valley semiconductors.

Electron densities in the lower and upper valleys remain the same under an equilibrium condition. When the applied electric field is lower than the electric field of the lower valley ($E < E_\ell$), no electrons will transfer to the upper valley as shown in Fig. 7-2-5(a). When the applied electric field is higher than that of the lower valley and lower than that of the upper valley ($E_\ell < E < E_u$), electrons will begin to transfer to the upper valley as shown in Fig. 7-2-5(b). And when the applied electric

TABLE 7-2-2 DATA FOR TWO-VALLEY SEMICONDUCTORS

Semiconductor	Gap energy (at 300°K) $E_g(\text{eV})$	Separation energy between two valleys $\Delta E(\text{eV})$	Threshold field $E_{th}(\text{KV/cm})$	Peak velocity $v_p(10^7 \text{ cm/s})$
Ge	0.80	0.18	2.3	1.4
GaAs	1.43	0.36	3.2	2.2
InP	1.33	0.60	10.5	2.5
		0.80		
CdTe	1.44	0.51	13.0	1.5
InAs	0.33	1.28	1.60	3.6
InSb	0.16	0.41	0.6	5.0

Note: InP is a three-valley semiconductor: 0.60 eV is the separation energy between the middle and lower valleys, 0.8 eV that between the upper and lower valleys.
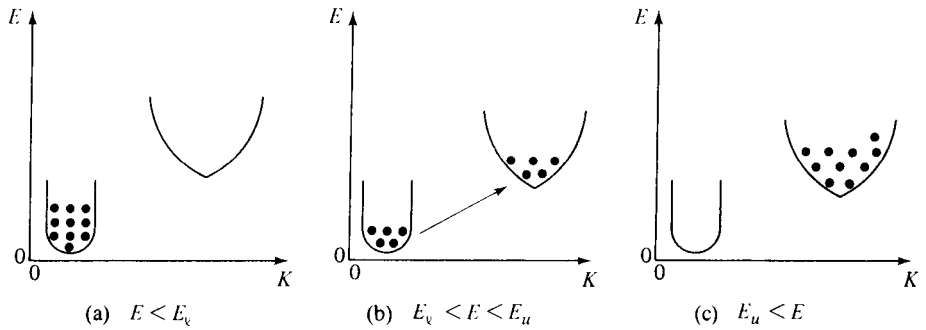
Figure 7-2-5   Transfer of electron densities.

field is higher than that of the upper valley ($E_u < E$), all electrons will transfer to the upper valley as shown in Fig. 7-2-5(c).

If electron densities in the lower and upper valleys are n_ℓ and n_u, the conductivity of the n-type GaAs is

$$\sigma = e(\mu_\ell n_\ell + \mu_u n_u) \qquad (7\text{-}2\text{-}2)$$

where e = the electron charge
μ = the electron mobility
$n = n_\ell + n_u$ is the electron density

Example 7-2-1:   Conductivity of an n-Type GaAs Gunn Diode

Electron density: $n = 10^{18}$ cm^{-3}
Electron density at lower valley: $n_\ell = 10^{10}$ cm^{-3}
Electron density at upper valley: $n_u = 10^8$ cm^{-3}
Temperature: $T = 300°$K

Determine the conductivity of the diode.

Solution   From Eq. (7-2-2) the conductivity is

$$\sigma = e(\mu_\ell n_\ell + \mu_u n_u)$$
$$= 1.6 \times 10^{-19}(8000 \times 10^{-4} \times 10^{16} + 180 \times 10^{-4} \times 10^{14})$$
$$\simeq 1.6 \times 10^{-19} \times 8000 \times 10^{-4} \times 10^{16} \qquad \text{for } n_\ell \gg n_u$$
$$= 1.28 \text{ mmhos}$$

When a sufficiently high field E is applied to the specimen, electrons are accelerated and their effective temperature rises above the lattice temperature. Furthermore, the lattice temperature also increases. Thus electron density n and mobility μ are both functions of electric field E. Differentiation of Eq. (7-2-2) with respect to E yields

$$\frac{d\sigma}{dE} = e\left(\mu_\ell \frac{dn_\ell}{dE} + \mu_u \frac{dn_u}{dE}\right) + e\left(n_\ell \frac{d\mu_\ell}{dE} + n_u \frac{d\mu_u}{dE}\right) \qquad (7\text{-}2\text{-}3)$$

If the total electron density is given by $n = n_\ell + n_u$ and it is assumed that μ_ℓ and μ_u are proportional to E^p, where p is a constant, then

$$\frac{d}{dE}(n_\ell + n_u) = \frac{dn}{dE} = 0 \qquad (7\text{-}2\text{-}4)$$

$$\frac{dn_\ell}{dE} = -\frac{dn_u}{dE} \qquad (7\text{-}2\text{-}5)$$

and
$$\frac{d\mu}{dE} \propto \frac{dE^p}{dE} = pE^{p-1} = p\frac{E^p}{E} \propto p\frac{\mu}{E} = \mu\frac{p}{E} \qquad (7\text{-}2\text{-}6)$$

Substitution of Eqs. (7-2-4) to (7-2-6) into Eq. (7-2-3) results in

$$\frac{d\sigma}{dE} = e(\mu_\ell - \mu_u)\frac{dn_\ell}{dE} + e(n_\ell\mu_\ell + n_u\mu_u)\frac{p}{E} \qquad (7\text{-}2\text{-}7)$$

Then differentiation of Ohm's law $J = \sigma E$ with respect to E yields

$$\frac{dJ}{dE} = \sigma + \frac{d\sigma}{dE}E \qquad (7\text{-}2\text{-}8)$$

Equation (7-2-8) can be rewritten

$$\frac{1}{\sigma}\frac{dJ}{dE} = 1 + \frac{d\sigma/dE}{\sigma/E} \qquad (7\text{-}2\text{-}9)$$

Clearly, for negative resistance, the current density J must decrease with increasing field E or the ratio of dJ/dE must be negative. Such would be the case only if the right-hand term of Eq. (7-2-9) is less than zero. In other words, the condition for negative resistance is

$$-\frac{d\sigma/dE}{\sigma/E} > 1 \qquad (7\text{-}2\text{-}10)$$

Substitution of Eqs. (7-2-2) and (7-2-7) with $f = n_u/n_\ell$ results in [2]

$$\left[\left(\frac{\mu_\ell - \mu_u}{\mu_\ell + \mu_u f}\right)\left(-\frac{E}{n_\ell}\frac{dn_\ell}{dE}\right) - p\right] > 1 \qquad (7\text{-}2\text{-}11)$$

Note that the field exponent p is a function of the scattering mechanism and should be negative and large. This factor makes impurity scattering quite undesirable because when it is dominant, the mobility rises with an increasing field and thus p is positive. When lattice scattering is dominant, however, p is negative and must depend on the lattice and carrier temperature. The first bracket in Eq. (7-2-11) must be positive in order to satisfy inequality. This means that $\mu_\ell > \mu_u$. Electrons must begin in a low-mass valley and transfer to a high-mass valley when they are heated by the electric field. The maximum value of this term is unity—that is, $\mu_\ell \gg \mu_u$. The factor dn_ℓ/dE in the second bracket must be negative. This quantity represents the rate of the carrier density with field at which electrons transfer to the upper valley; this rate depends on differences between electron densities, electron temperature, and gap energies in the two valleys.

On the basis of the Ridley–Watkins–Hilsum theory as described earlier, the band structure of a semiconductor must satisfy three criteria in order to exhibit negative resistance [12].

1. The separation energy between the bottom of the lower valley and the bottom of the upper valley must be several times larger than the thermal energy (about 0.026 eV) at room temperature. This means that $\Delta E > kT$ or $\Delta E > 0.026$ eV.
2. The separation energy between the valleys must be smaller than the gap energy between the conduction and valence bands. This means that $\Delta E < E_g$. Otherwise the semiconductor will break down and become highly conductive before the electrons begin to transfer to the upper valleys because hole-electron pair formation is created.
3. Electrons in the lower valley must have high mobility, small effective mass, and a low density of state, whereas those in the upper valley must have low mobility, large effective mass, and a high density of state. In other words, electron velocities (dE/dk) must be much larger in the lower valleys than in the upper valleys.

The two most useful semiconductors—silicon and germanium—do not meet all these criteria. Some compound semiconductors, such as gallium arsenide (GaAs), indium phosphide (InP), and cadmium telluride (CdTe) do satisfy these criteria. Others such as indium arsenide (InAs), gallium phosphide (GaP), and indium antimonide (InSb) do not. Figure 7-2-6 shows a possible current versus field characteristic of a two-valley semiconductor.
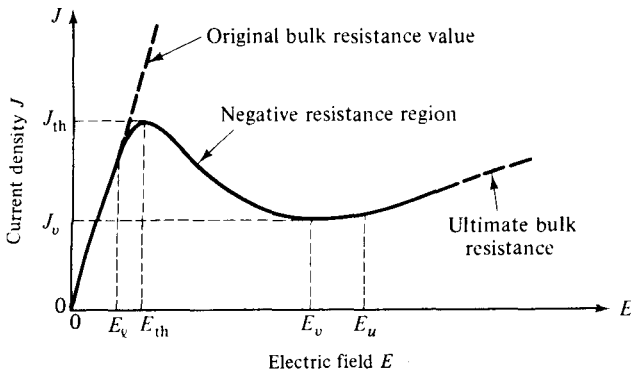
Figure 7-2-6 Current versus field characteristic of a two-valley semiconductor.

A mathematical analysis of differential negative resistance requires a detailed analysis of high-field carrier transports [13–14]. From electric field theory the magnitude of the current density in a semiconductor is given by

$$J = qnv \tag{7-2-12}$$

where q = electric charge
n = electron density, and
v = average electron velocity.
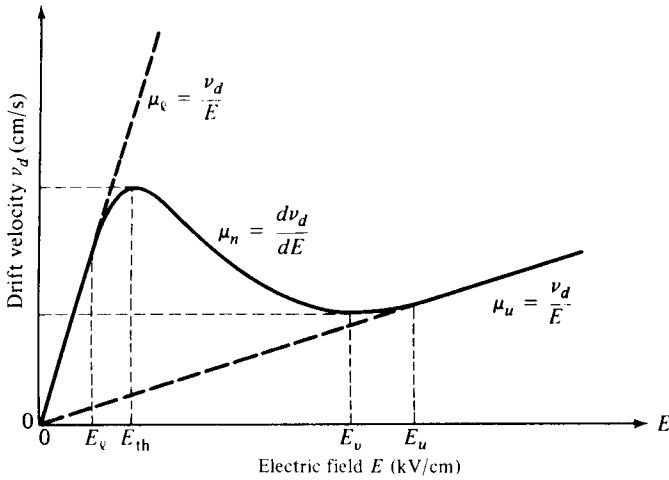
Figure 7-2-7   Electron drift velocity versus electric field.

Differentiation of Eq. (7-2-12) with respect to electric field E yields

$$\frac{dJ}{dE} = qn\frac{dv}{dE} \qquad (7\text{-}2\text{-}13)$$

The condition for negative differential conductance may then be written

$$\frac{dv_d}{dE} = \mu_n < 0 \qquad (7\text{-}2\text{-}14)$$

where μ_n denotes the negative mobility, which is shown in Fig. 7-2-7.

The direct measurement of the dependence of drift velocity on the electric field and direct evidence for the existence of the negative differential mobility were made by Ruch and Kino [15]. Experimental results, along with the theoretical results of analysis by Butcher and Fawcett [13], are shown in Fig. 7-2-8.
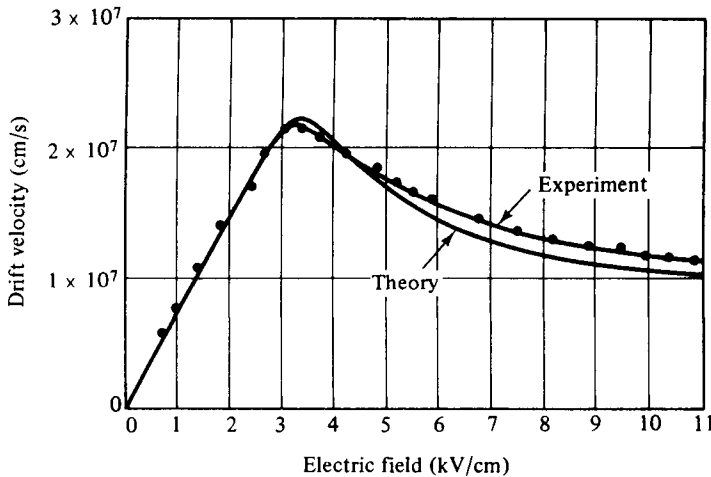
Figure 7-2-8   Theoretical and experimental velocity-field characteristics of a GaAs diode.

Example 7-2-2: Characteristics of a GaAs Gunn Diode

A typical n-type GaAs Gunn diode has the following parameters:

Threshold field	$E_{th} = 2800$ V/cm
Applied field	$E = 3200$ V/cm
Device length	$L = 10 \ \mu$m
Doping concentration	$n_0 = 2 \times 10^{14}$ cm^{-3}
Operating frequency	$f = 10$ GHz

a. Compute the electron drift velocity.
b. Calculate the current density.
c. Estimate the negative electron mobility.

Solution

a. The electron drift velocity is

$$v_d = 10 \times 10^9 \times 10 \times 10^{-6} = 10^5 \text{ m/sec} = 10^7 \text{ cm/sec}$$

b. From Eq. (7-2-12) the current density is

$$J = qnv = 1.6 \times 10^{-19} \times 2 \times 10^{20} \times 10 \times 10^9 \times 10^{-5}$$

$$= 3.2 \times 10^6 \text{ A/m}^2$$

$$= 320 \text{ A/cm}^2$$

c. The negative electron mobility is

$$\mu_n = -\frac{v_d}{E} = -\frac{10^7}{3200} = -3100 \text{ cm}^2/\text{V} \cdot \text{sec}$$

7-2-3 High-Field Domain

In the last section we described how differential resistance can occur when an electric field of a certain range is applied to a multivalley semiconductor compound, such as the n-type GaAs. In this section we demonstrate how a decrease in drift velocity with increasing electric field can lead to the formation of a high-field domain for microwave generation and amplification.

In the n-type GaAs diode the majority carriers are electrons. When a small voltage is applied to the diode, the electric field and conduction current density are uniform throughout the diode. At low voltage the GaAs is ohmic, since the drift velocity of the electrons is proportional to the electric field. This situation was shown in Fig. 7-1-2. The conduction current density in the diode is given by

$$\mathbf{J} = \sigma \mathbf{E}_x = \frac{\sigma V}{L} \mathbf{U}_x = \rho v_x \mathbf{U}_x \qquad (7\text{-}2\text{-}15)$$

where $\mathbf{J}$ = conduction current density
$\quad \sigma$ = conductivity
$\quad \mathbf{E}_x$ = electric field in the x direction

L = length of the diode
V = applied voltage
ρ = charge density
v = drift velocity
$\mathbf{U}$ = unit vector

The current is carried by free electrons that are drifting through a background of fixed positive charge. The positive charge, which is due to impurity atoms that have donated an electron (donors), is sometimes reduced by impurity atoms that have accepted an electron (acceptors). As long as the fixed charge is positive, the semiconductor is n type, since the principal carriers are the negative charges. The density of donors less the density of acceptors is termed *doping*. When the space charge is zero, the carrier density is equal to the doping.

When the applied voltage is above the threshold value, which was measured at about 3000 V/cm times the thickness of the GaAs diode, a high-field domain is formed near the cathode that reduces the electric field in the rest of the material and causes the current to drop to about two-thirds of its maximum value. This situation occurs because the applied voltage is given by

$$V = -\int_0^L E_x \, dx \qquad (7\text{-}2\text{-}16)$$

For a constant voltage V an increase in the electric field within the specimen must be accompanied by a decrease in the electric field in the rest of the diode. The high-field domain then drifts with the carrier stream across the electrodes and disappears at the anode contact. When the electric field increases, the electron drift velocity decreases and the GaAs exhibits negative resistance.

Specifically, it is assumed that at point A on the J-E plot as shown in Fig. 7-2-9(b) there exists an excess (or accumulation) of negative charge that could be caused by a random noise fluctuation or possibly by a permanent nonuniformity in doping in the n-type GaAs diode. An electric field is then created by the accumulated charges as shown in Fig. 7-2-9(d). The field to the left of point A is lower than that to the right. If the diode is biased at point E_A on the J-E curve, this situation implies that the carriers (or current) flowing into point A are greater than those flowing out of point A, thereby increasing the excess negative space charge at A. Furthermore, when the electric field to the left of point A is lower than it was before, the field to the right is then greater than the original one, resulting in an even greater space-charge accumulation. This process continues until the low and high fields both reach values outside the differential negative-resistance region and settle at points 1 and 2 in Fig. 7-2-9(a) where the currents in the two field regions are equal. As a result of this process, a traveling space-charge accumulation is formed. This process, of course, depends on the condition that the number of electrons inside the crystal is large enough to allow the necessary amount of space charge to be built up during the transit time of the space-charge layer.

The pure accumulation layer discussed above is the simplest form of space-charge instability. When positive and negative charges are separated by a small distance, then a dipole domain is formed as shown in Fig. 7-2-10. The electric field in-
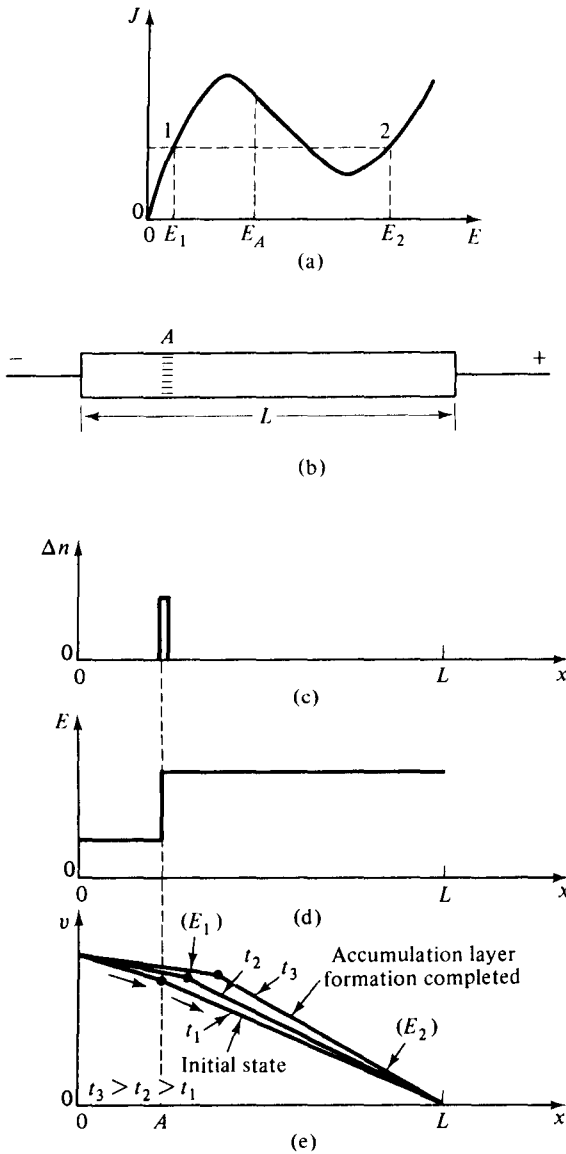
Figure 7-2-9 Formation of an electron accumulation layer in GaAs. (*After H. Kroemer [16]; reprinted by permission of IEEE, Inc.*)

side the dipole domain would be greater than the fields on either side of the dipole in Fig. 7-2-10(c). Because of the negative differential resistance, the current in the low-field side would be greater than that in the high-field side. The two field values will tend toward equilibrium conditions outside the differential negative-resistance region, where the low and high currents are the same as described in the previous section. Then the dipole field reaches a stable condition and moves through the specimen toward the anode. When the high-field domain disappears at the anode, a new dipole field starts forming at the cathode and the process is repeated.

(a)

(b)

(c)

(d)
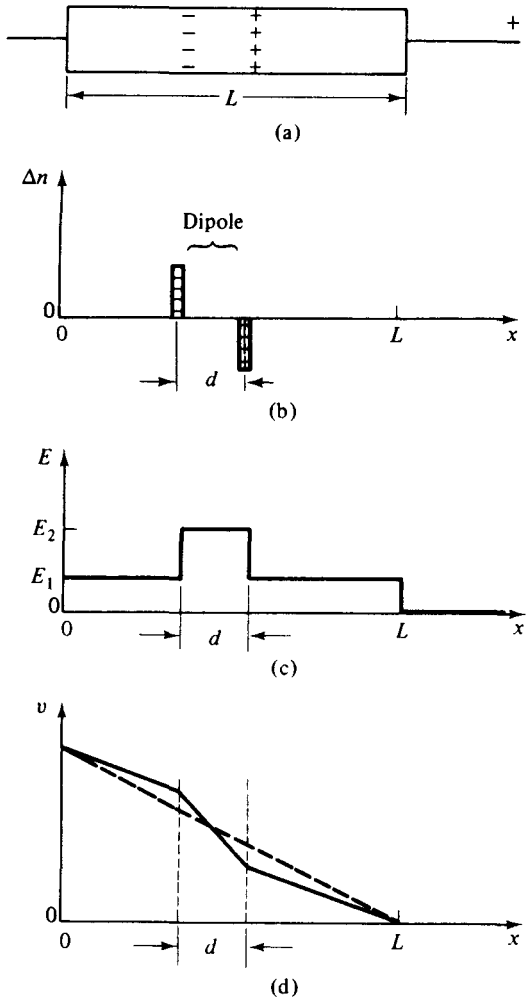
Figure 7-2-10   Formation of an electron dipole layer in GaAs. (*After H. Kroemer [16]; reprinted by permission of IEEE, Inc.*)

In general, the high-field domain has the following properties [12]:

1. A domain will start to form whenever the electric field in a region of the sample increases above the threshold electric field and will drift with the carrier stream through the device. When the electric field increases, the electron drift velocity decreases and the GaAs diode exhibits negative resistance.
2. If additional voltage is applied to a device containing a domain, the domain will increase in size and absorb more voltage than was added and the current will decrease.
3. A domain will not disappear before reaching the anode unless the voltage is dropped appreciably below threshold (for a diode with uniform doping and area).

4. The formation of a new domain can be prevented by decreasing the voltage slightly below threshold (in a nonresonant circuit).

5. A domain will modulate the current through a device as the domain passes through regions of different doping and cross-sectional area, or the domain may disappear. The effective doping may be varied in regions along the drift path by additional contacts.

6. The domain's length is generally inversely proportional to the doping; thus devices with the same product of doping multiplied by length will behave similarly in terms of frequency multiplied by length, voltage/length, and efficiency.

7. As a domain passes a point in the device, the domain can be detected by a capacitive contact, since the voltage changes suddenly as the domain passes. The presence of a domain anywhere in a device can be detected by a decreased current or by a change in differential impedance.

It should be noted that properties 3 and 6 are valid only when the length of the domain is much longer than the thermal diffusion length for carriers, which for GaAs is about 1 μm for a doping of 10^{16} per cubic centimeter and about 10 μm for a doping of 10^{14} per cubic centimeter.

7-3 MODES OF OPERATION

Since Gunn first announced his observation of microwave oscillation in the n-type GaAs and n-type InP diodes in 1963, various modes of operation have been developed, depending on the material parameters and operating conditions. As noted, the formation of a strong space-charge instability depends on the conditions that enough charge is available in the crystal and that the specimen is long enough so that the necessary amount of space charge can be built up within the transit time of the electrons. This requirement sets up a criterion for the various modes of operation of bulk negative-differential-resistance devices. Copeland proposed four basic modes of operation of uniformly doped bulk diodes with low-resistance contacts [17] as shown in Fig. 7-3-1.

1. Gunn oscillation mode: This mode is defined in the region where the product of frequency multiplied by length is about 10^7 cm/s and the product of doping multiplied by length is greater than 10^{12}/cm^2. In this region the device is unstable because of the cyclic formation of either the accumulation layer or the high-field domain. In a circuit with relatively low impedance the device operates in the high-field domain mode and the frequency of oscillation is near the intrinsic frequency. When the device is operated in a relatively high-Q cavity and coupled properly to the load, the domain is quenched or delayed (or both) before nucleating. In this case, the oscillation frequency is almost entirely determined by the resonant frequency of the cavity and has a value of several times the intrinsic frequency.
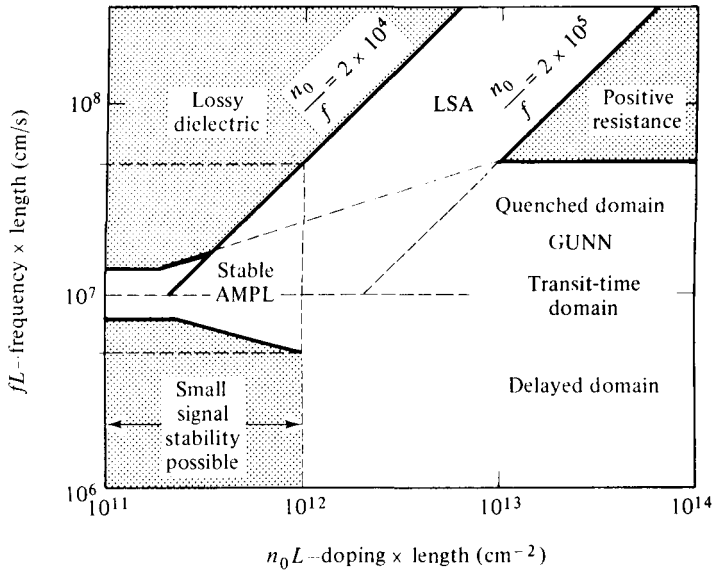
Figure 7-3-1 Modes of operation for Gunn diodes. (*After J. A. Copeland [17];* *reprinted by permission of IEEE, Inc.*)

2. Stable amplification mode: This mode is defined in the region where the product of frequency times length is about 10^7 cm/s and the product of doping times length is between 10^{11} and 10^{12}/cm^2.

3. LSA oscillation mode: This mode is defined in the region where the product of frequency times length is above 10^7 cm/s and the quotient of doping divided by frequency is between 2×10^4 and 2×10^5.

4. Bias-circuit oscillation mode: This mode occurs only when there is either Gunn or LSA oscillation, and it is usually at the region where the product of frequency times length is too small to appear in the figure. When a bulk diode is biased to threshold, the average current suddenly drops as Gunn oscillation begins. The drop in current at the threshold can lead to oscillations in the bias circuit that are typically 1 kHz to 100 MHz [18].

The first three modes are discussed in detail in this section. Before doing so, however, let us consider the criterion for classifying the modes of operation.

7-3-1 Criterion for Classifying the Modes of Operation

The Gunn-effect diodes are basically made from an n-type GaAs, with the concentrations of free electrons ranging from 10^{14} to 10^{17} per cubic centimeter at room temperature. Its typical dimensions are 150×150 μm in cross section and 30 μm long. During the early stages of space-charge accumulation, the time rate of growth

of the space-charge layers is given by

$$Q(X, t) = Q(X - vt, 0) \exp\left(\frac{t}{\tau_d}\right) \tag{7-3-1}$$

where $\tau_d = \dfrac{\epsilon}{\sigma} = \dfrac{\epsilon}{en_0|\mu_n|}$ is the magnitude of the negative dielectric relaxation time

ϵ = semiconductor dielectric permittivity
n_0 = doping concentration
μ_n = negative mobility
e = electron charge
σ = conductivity

Figure 7-3-2 clarifies Eq. (7-3-1).
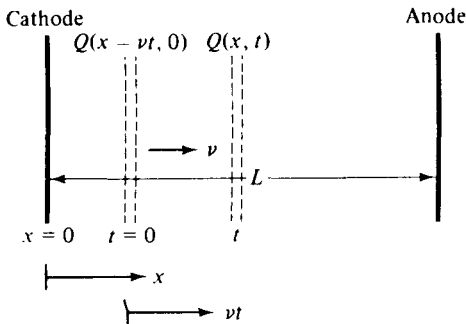
Figure 7-3-2 Space-charge accumulation with a velocity of v.

If Eq. (7-3-1) remains valid throughout the entire transit time of the space-charge layer, the factor of maximum growth is given by

$$\text{Growth factor} = \frac{Q(L, L/v)}{Q(0, 0)} = \exp\left(\frac{L}{v\tau_d}\right) = \exp\left(\frac{Ln_0e|\mu_n|}{\epsilon v}\right) \tag{7-3-2}$$

In Eq. (7-3-2) the layer is assumed to start at the cathode at $t = 0$, $X = 0$, and arrive at the anode at $t = L/v$ and $X = L$. For a large space-charge growth, this factor must be larger than unity. This means that

$$n_0 L > \frac{\epsilon v}{e|\mu_n|} \tag{7-3-3}$$

This is the criterion for classifying the modes of operation for the Gunn-effect diodes. For n-type GaAs, the value of $\epsilon v/(e|\mu_n|)$ is about $10^{12}/\text{cm}^2$, where $|\mu_n|$ is assumed to be 150 cm²/V · s.

Example 7-3-1:   Criterion of Mode Operation

An n-type GaAs Gunn diode has the following parameters:

Electron drift velocity:	$v_d = 2.5 \times 10^5$ m/s
Negative electron mobility:	$|\mu_n| = 0.015$ m²/V · s
Relative dielectric constant:	$\epsilon_r = 13.1$

Determine the criterion for classifying the modes of operation.

Solution   From Eq. (7-3-3) the criterion is

$$\frac{\epsilon v_d}{e|\mu_n|} = \frac{8.854 \times 10^{-12} \times 13.1 \times 2.5 \times 10^5}{1.6 \times 10^{-19} \times 0.015}$$

$$= 1.19 \times 10^{16}/\text{m}^2$$

$$= 1.19 \times 10^{12}/\text{cm}^2$$

This means that the product of the doping concentration and the device length must be

$$n_0 L > 1.19 \times 10^{12}/\text{cm}^2$$

7-3-2 Gunn Oscillation Modes ($10^{12}/\text{cm}^2 \leqslant (n_0 L) < 10^{14}/\text{cm}^2$)

Most Gunn-effect diodes have the product of doping and length ($n_0 L$) greater than $10^{12}/\text{cm}^2$. However, the mode that Gunn himself observed had a product $n_0 L$ that is much less. When the product of $n_0 L$ is greater than $10^{12}/\text{cm}^2$ in GaAs, the space-charge perturbations in the specimen increase exponentially in space and time in accordance with Eq. (7-3-1). Thus a high-field domain is formed and moves from the cathode to the anode as described earlier. The frequency of oscillation is given by the relation [19]

$$f = \frac{v_{\text{dom}}}{L_{\text{eff}}} \tag{7-3-4}$$

where v_{dom} is the domain velocity and L_{eff} is the effective length that the domain travels from the time it is formed until the time that a new domain begins to form.

Gunn described the behavior of Gunn oscillators under several circuit configurations [20]. When the circuit is mainly resistive or the voltage across the diode is constant, the period of oscillation is the time required for the domain to drift from the cathode to the anode. This mode is not actually typical of microwave applications. Negative conductivity devices are usually operated in resonant circuits, such as high-Q resonant microwave cavities. When the diode is in a resonant circuit, the frequency can be tuned to a range of about an octave without loss of efficiency [21].

As described previously, the normal Gunn domain mode (or Gunn oscillation mode) is operated with the electric field greater than the threshold field ($E > E_{\text{th}}$). The high-field domain drifts along the specimen until it reaches the anode or until the low-field value drops below the sustaining field E_s required to maintain v_s as shown in Fig. 7-3-3. The sustaining drift velocity for GaAs is $v_s = 10^7$ cm/s. Since the electron drift velocity v varies with the electric field, there are three possible domain modes for the Gunn oscillation mode.

Transit-time domain mode ($fL \approx 10^7$ cm/s). When the electron drift velocity v_d is equal to the sustaining velocity v_s, the high-field domain is stable. In other words, the electron drift velocity is given by

$$v_d = v_s = fL \approx 10^7 \text{ cm/s} \tag{7-3-5}$$

Then the oscillation period is equal to the transit time—that is, $\tau_0 = \tau_t$. This situation is shown in Fig. 7-3-4(a). The efficiency is below 10% because the current is collected only when the domain arrives at the anode.
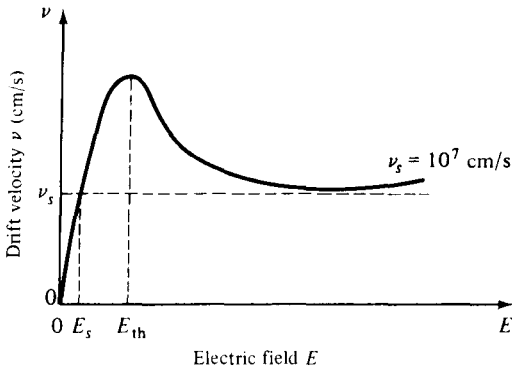
Figure 7-3-3 Electron drift velocity versus electric field.
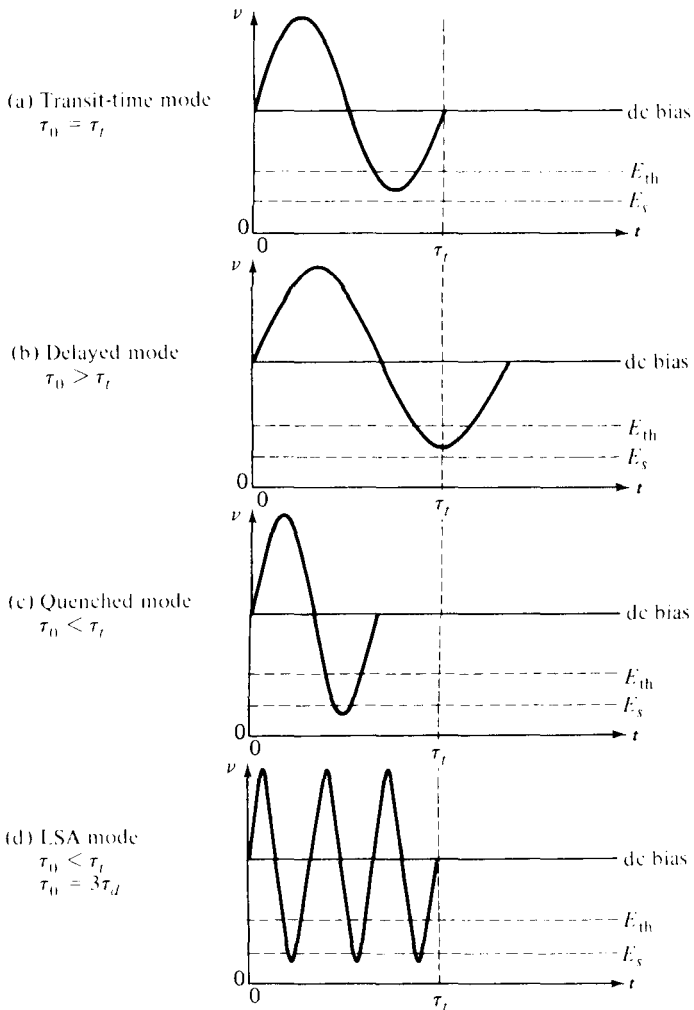
Figure 7-3-4 Gunn domain modes.

Delayed domain mode (10^6 cm/s $< fL < 10^7$ cm/s). When the transit time is chosen so that the domain is collected while $E < E_{th}$ as shown in Fig. 7-3-4(b), a new domain cannot form until the field rises above threshold again. In this case, the oscillation period is greater than the transit time—that is, $\tau_0 > \tau_t$. This delayed mode is also called *inhibited mode*. The efficiency of this mode is about 20%.

Quenched domain mode ($fL > 2 \times 10^7$ cm/s). If the bias field drops below the sustaining field E_s during the negative half-cycle as shown in Fig. 7-3-4(c), the domain collapses before it reaches the anode. When the bias field swings back above threshold, a new domain is nucleated and the process repeats. Therefore the oscillations occur at the frequency of the resonant circuit rather than at the transit-time frequency. It has been found that the resonant frequency of the circuit is several times the transit-time frequency, since one dipole does not have enough time to readjust and absorb the voltage of the other dipoles [22, 23]. Theoretically, the efficiency of quenched domain oscillators can reach 13% [22].

7-3-3 Limited-Space-Charge Accumulation (LSA) Mode ($fL > 2 \times 10^7$ cm/s)

When the frequency is very high, the domains do not have sufficient time to form while the field is above threshold. As a result, most of the domains are maintained in the negative conductance state during a large fraction of the voltage cycle. Any accumulation of electrons near the cathode has time to collapse while the signal is below threshold. Thus the LSA mode is the simplest mode of operation, and it consists of a uniformly doped semiconductor without any internal space charges. In this instance, the internal electric field would be uniform and proportional to the applied voltage. The current in the device is then proportional to the drift velocity at this field level. The efficiency of the LSA mode can reach 20%.

The oscillation period τ_0 should be no more than several times larger than the magnitude of the dielectric relaxation time in the negative conductance region τ_d. The oscillation indicated in Fig. 7-3-4(d) is $\tau_0 = 3\tau_d$. It is appropriate here to define the LSA boundaries. As described earlier, the sustaining drift velocity is 10^7 cm/s as shown in Eq. (7-3-5) and Fig. 7-3-3. For the *n*-type GaAs, the product of doping and length ($n_0 L$) is about $10^{12}/\text{cm}^2$. At the low-frequency limit, the drift velocity is taken to be

$$v_\ell = fL = 5 \times 10^6 \text{ cm/s} \qquad (7\text{-}3\text{-}6)$$

The ratio of $n_0 L$ to fL yields

$$\frac{n_0}{f} = 2 \times 10^5 \qquad (7\text{-}3\text{-}7)$$

At the upper-frequency limit it is assumed that the drift velocity is

$$v_u = fL = 5 \times 10^7 \text{ cm/s} \qquad (7\text{-}3\text{-}8)$$

and the ratio of $n_0 L$ to fL is

$$\frac{n_0}{f} = 2 \times 10^4 \qquad (7\text{-}3\text{-}9)$$

Both the upper and lower boundaries of the LSA mode are indicated in Fig. 7-3-1. The LSA mode is discussed further in Section 7-4.

7-3-4 Stable Amplification Mode ($n_0 L < 10^{12}/cm^2$)

When the $n_0 L$ product of the device is less than about $10^{12}/cm^2$, the device exhibits amplification at the transit-time frequency rather than spontaneous oscillation. This situation occurs because the negative conductance is utilized without domain formation. There are too few carriers for domain formation within the transit time. Therefore amplification of signals near the transit-time frequency can be accomplished. This mode was first observed by Thim and Barber [23]. Furthermore, Uenohara showed that there are types of amplification depending on the fL product of the device [24] as shown in Fig. 7-3-5.
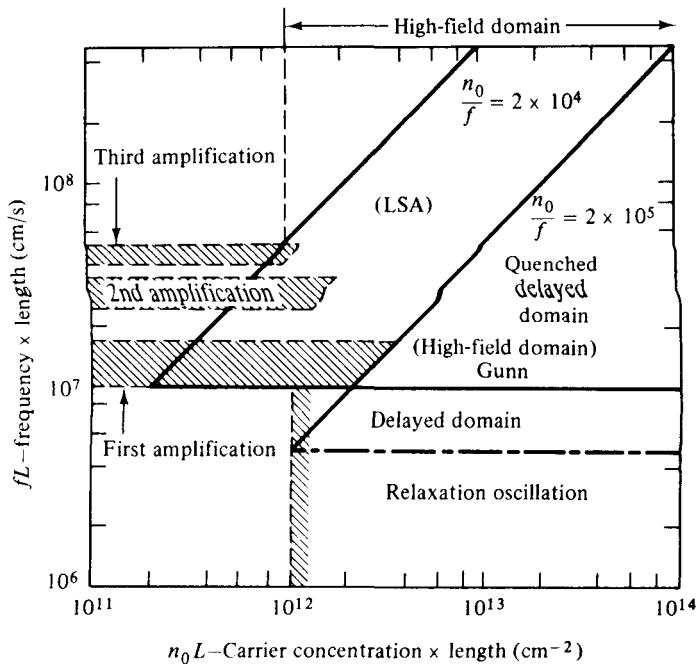
Figure 7-3-5　Mode chart. (*After M. Uenohara [24]; reprinted by permission of McGraw-Hill Book Company.*)

The various modes of operation of Gunn diodes can be classified on the basis of the times in which various processes occur. These times are defined as follows:

τ_t = domain transit time

τ_d = dielectric relaxation time at low field

τ_g = domain growth time

τ_0 = natural period of oscillation of a high-Q external electric circuit

The modes described previously are summarized in Table 7-3-1.

TABLE 7-3-1 MODES OF OPERATION OF GUNN OSCILLATORS

Mode	Time relationship	Doping level	Nature of circuit
Stable amplifier	$\tau_0 \geq \tau_t$	$n_0 L < 10^{12}$	Nonresonant
Gun domain	$\tau_g \leq \tau_t$	$n_0 L > 10^{12}$	Nonresonant;
	$\tau_0 = \tau_t$		constant voltage
Quenched domain	$\tau_g \leq \tau_t$	$n_0 L > 10^{12}$	Resonant;
	$\tau_0 = \tau_t$		finite impedance
Delayed domain	$\tau_g \leq \tau_t$	$n_0 L > 10^{12}$	Resonant;
	$\tau_0 > \tau_t$		finite impedance
LSA	$\tau_0 < \tau_g$	$2 \times 10^4 < \left(\dfrac{n_0}{f}\right) < 2 \times 10^5$	Multiple resonances;
	$\tau_0 > \tau_d$		high impedance;
			high dc bias

7-4 LSA DIODES

The abbreviation LSA stands for the limited space-charge accumulation mode of the Gunn diode. As described previously, if the product $n_0 L$ is larger than 10^{12}/cm^2 and if the ratio of doping n_0 to frequency f is within 2×10^5 to 2×10^4 s/cm^3, the high-field domains and the space-charge layers do not have sufficient time to build up. The magnitude of the RF voltage must be large enough to drive the diode below threshold during each cycle in order to dissipate space charge. Also, the portion of each cycle during which the RF voltage is above threshold must be short enough to prevent the domain formation and the space-charge accumulation. Only the primary accumulation layer forms near the cathode; the rest of the sample remains fairly homogeneous. Thus with limited space-charge formation the remainder of the sample appears as a series negative resistance that increases the frequency of the oscillations in the resonant circuit. Copeland discovered the LSA mode of the Gunn diode in 1966 [25]. In the LSA mode the diode is placed in a resonator tuned to an oscillation frequency of

$$f_0 = \frac{1}{\tau_0} \qquad (7\text{-}4\text{-}1)$$

The device is biased to several times the threshold voltage (see Fig. 7-4-1).

As the RF voltage swings beyond the threshold, the space charge starts building up at the cathode. Since the oscillation period τ_0 of the RF signal is less than the domain-growth time constant τ_g, the total voltage swings below the threshold before the domain can form. Furthermore, since τ_0 is much greater than the dielectric relaxation time τ_d, the accumulated space charge is drained in a very small fraction of the RF cycle. Therefore the device spends most of the RF cycle in the negative-resistance region, and the space charge is not allowed to build up. The frequency of oscillation in the LSA mode is independent of the transit time of the carriers and is determined solely by the circuit external to the device. Also, the power-impedance product does not fall off as $1/f_0^2$; thus the output power in the LSA mode can be greater than that in the other modes.

The LSA mode does have limitations. It is very sensitive to load conditions, temperatures, and doping fluctuations [26]. In addition, the RF circuit must allow the field to build up quickly in order to prevent domain formation. The power output
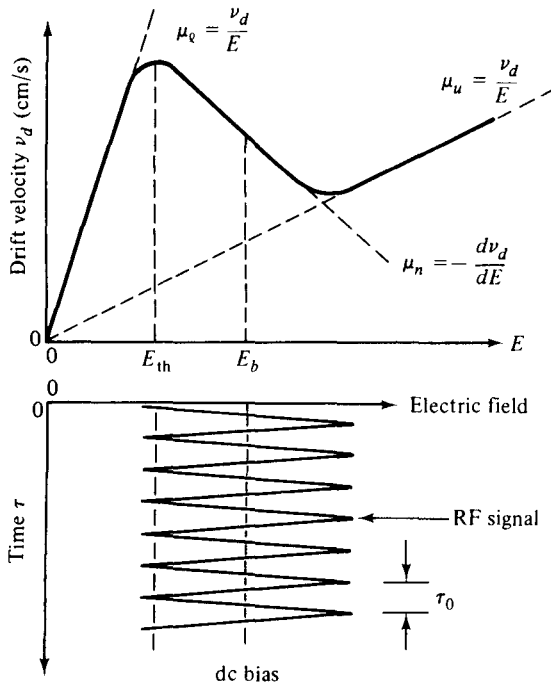
Figure 7-4-1 LSA mode operation.

of an LSA oscillator can be simply written

$$P = \eta V_0 I_0 = \eta (M E_{th} L)(n_0 e v_0 A) \qquad (7\text{-}4\text{-}2)$$

where η = dc-to-RF conversion efficiency (primarily a function of material and circuit considerations)

V_0 = operating voltage

I_0 = operating current

M = multiple of the operating voltage above negative-resistance threshold voltage

E_{th} = threshold field (about 3400 V/cm)

L = device length (about 10 to 200 μm)

n_o = donor concentration (about 10^{15} e/cm³)

e = electron charge (1.6×10^{-19} C)

v_0 = average carrier drift velocity (about 10^7 cm/s)

A = device area (about 3×10^{-4} to 20×10^{-4} cm²)

For an LSA oscillator, n_0 is primarily determined by the desired operating frequency f_0 so that, for a properly designed circuit, peak power output is directly proportional to the volume (LA) of the device length L multiplied by the area A of the active layer. Active volume cannot be increased indefinitely. In the theoretical limit, this is due to electrical wavelength and skin-depth considerations. In the practical limit, however, available bias, thermal dissipation capability, or technological problems associated with material uniformity limit device length. The power capabilities of LSA oscillators vary from 6 kW (pulse) at 1.75 GHz to 400 W (pulse) at 51 GHz.

Example 7-4-1:   Output Power of an LSA Oscillator

An LSA oscillator has the following parameters:

Conversion efficiency:	$\eta = 0.06$
Multiplication factor:	$M = 3.5$
Threshold field:	$E_{th} = 320$ kV/m
Device length:	$L = 12\mu$ m
Donor concentration:	$n_0 = 10^{21}$ m^3
Average carrier velocity:	$v_0 = 1.5 \times 10^5$ m/s
Area:	$A = 3 \times 10^{-8}$ m^2

Determine the output power in milliwatts.

Solution   From Eq. (7-4-2) the output power is

$$P = \eta\,(ME_{th}L)(n_0ev_0A) = 0.06 \times (3.5 \times 320 \times 10^3 \times 12 \times 10^{-6})$$
$$\times\ (10^{21} \times 1.6 \times 10^{-19} \times 1.5 \times 10^5 \times 3 \times 10^{-8})$$
$$= 0.06 \times 13.44 \times 7.2 \times 10^{-1} = 581 \text{ mW}$$

7-5 InP DIODES

When Gunn first announced his Gunn effect in 1963, the diodes he investigated were of gallium arsenide (GaAs) and indium phosphide (InP). The GaAs diode was described earlier in the chapter. In this section the n-type InP diode is discussed. Both the GaAs diode and the InP diode operate basically the same way in a circuit with dc voltage applied at the electrodes. In the ordinary Gunn effect in the n-type GaAs, the two-valley model theory is the foundation for explaining the electrical behavior of the Gunn effect. However, Hilsum proposed that indium phosphide and some alloys of indium gallium antimonide should work as three-level devices [27]. Figure 7-5-1 shows the three-valley model for indium phosphide.

It can be seen that InP, besides having an upper-valley energy level and a lower-valley energy level similar to the model shown in Fig. 7-2-4 for n-type GaAs, also has a third middle-valley energy level. In GaAs the electron transfer process from the lower valley to the upper valley is comparatively slow. At a particular voltage above threshold current flow consists of a larger contribution of electrons from the lower valley rather than from the upper valley. Because of this larger contribution from the lower-energy level, a relatively low peak-to-valley current ratio results, which is shown in Fig. 7-5-2.

As shown in Fig. 7-5-3, the InP diode has a larger peak-to-valley current ratio because an electron transfer proceeds rapidly as the field increases. This situation occurs because the coupling between the lower valley and upper valley in InP is weaker than in GaAs. The middle-valley energy level provides the additional energy loss mechanism required to avoid breakdown caused by the high energies acquired by the lower-valley electrons from the weak coupling. It can be seen from Fig. 7-5-1
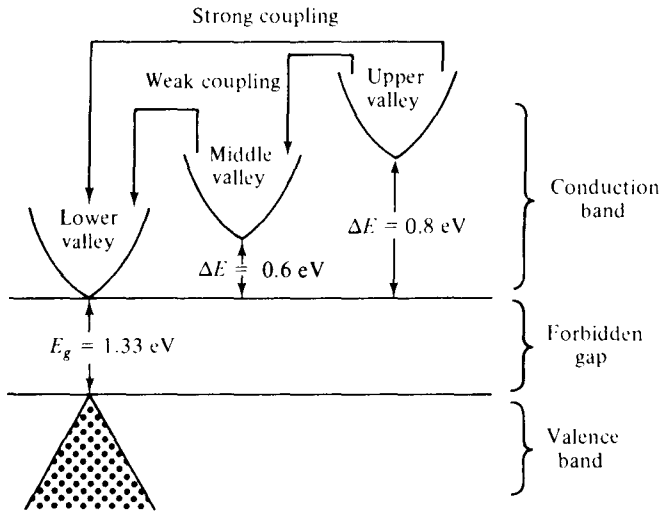
Figure 7-5-1   Three-valley-model energy level for InP diode.
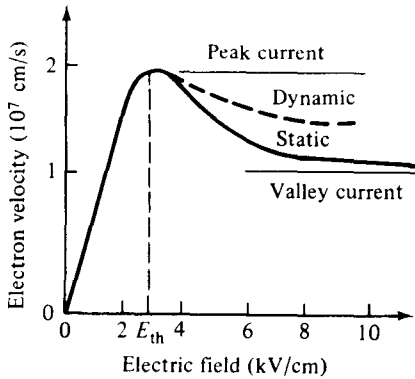
Figure 7-5-2   Peak-to-valley current ratio for n-type GaAs.
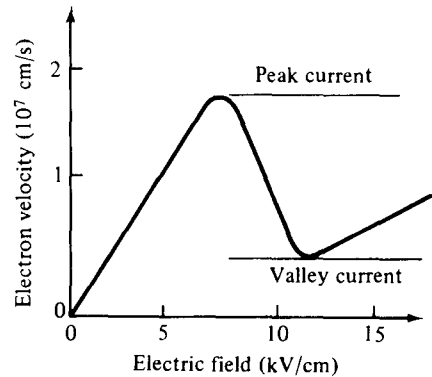
Figure 7-5-3   Peak-to-valley current ratio for InP.

that the lower valley is weakly coupled to the middle valley but strongly coupled to the upper valley to prevent breakdown. This situation ensures that under normal operating conditions electrons concentrate in the middle valley. Because InP has a greater energy separation between the lower valley and the nearest energy levels, the thermal excitation of electrons has less effect, and the degradation of its peak-to-valley current ratio is about four times less than in GaAs [28].

The mode of operation of InP is unlike the domain oscillating mode in which a high-field domain is formed that propagates with a velocity of about 10^7 cm/s. As a result, the output current waveform of an InP diode is transit-time dependent. This mode reduces the peak-to-valley current ratio so that efficiency is reduced. For this
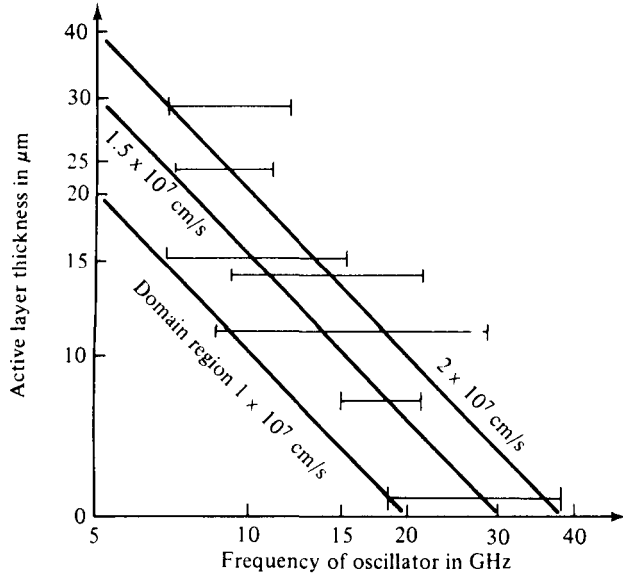
Figure 7-5-4 Active-layer thickness versus frequency for InP diode. (*From B. C. Taylor and D. J. Colliver [29]; reprinted by permission of IEEE, Inc.*)

reason, an operating mode is usually sought where charge domains are not formed. The three-valley model of InP inhibits the formation of domains because the electron diffusion coefficient is increased by the stronger coupling [28]. From experiments performed by Taylor and Colliver [29] it was determined that epitaxial InP oscillators operate through a transit-time phenomenon and do not oscillate in a bulk mode of the LSA type. From their findings it was determined that it is not appropriate to attempt to describe the space-charge oscillations in InP in terms of modes known to exist in GaAs devices. Taylor and Colliver also determined that the frequencies obtained from a device were dependent on the active-layer thickness. The InP oscillator could be tuned over a large frequency range, bounded only by the thickness, by adjusting the cavity size. Figure 7-5-4 shows the frequency ranges for the different active-layer thicknesses and lines of constant electron velocity [28, 29]. It can be seen from the figure that only a few InP devices operate in the domain formation area. In each case on the graph, the maximum efficiency occurs at about midband [29]. Table 7-5-1 summarizes the highest power and efficiencies for InP diodes.

TABLE 7-5-1 BEST PERFORMANCE OF InP DIODE

Frequency (GHz)	Thickness (μm)	Power (W)	Operation	Efficiency (%)	Reference
5.5	—	3.05	pulsed	14.7	[28]
8.5	28	0.95	pulsed	7.0	[29]
10.75	—	1.33	pulsed	12.0	[28]
13.8	11	0.50	pulsed	6.3	[29]
15	—	1.13	pulsed	15	[28]
18	11	1.05	pulsed	4.2	[29]
18	—	0.20	CW	10.2	[28]
25	5.4	0.65	pulsed	2.6	[29]
26	—	0.15	CW	6	[28]
29.4	5.4	0.23	pulsed	2	[29]
33	5.4	0.10	pulsed	0.4	[29]
37	—	0.01	CW	1	[28]

7-6 CdTe DIODES

The Gunn effect, first observed by Gunn as a time variation in the current through samples of n-type GaAs when the voltage across the sample exceeded a critical value, has since been observed in n-type InP, n-type CdTe, alloys of n-GaAs and n-GaP, and in InAs. In n-type cadmium telluride (CdTe), the Gunn effect was first seen by Foyt and McWhorter [30], who observed a time variation of the current through samples 250 to 300 μm long with a carrier concentration of 5×10^{14}/cm^3 and a room temperature mobility of 1000 cm^2/V $\cdot$ s. Ludwig, Halsted, and Aven [31] confirmed the existence of current oscillations in n-CdTe, and Ludwig has further reported studies of the Gunn effect in CdTe over a wider range of sample doping levels and lengths [32]. It has been confirmed that the same mechanism—the field-induced transfer of electrons to a higher conduction band minimum (Gunn effect)—applies in CdTe just as it does in GaAs. From the two-valley model theory in CdTe, as in GaAs, the $\langle 000 \rangle$ minimum is the lowest in energy. The effective mass $m_{\text{eff}} =$ 0.11 m (electron mass) and the intrinsic mobility $\mu \approx 1100$ cm^2/V $\cdot$ s at room temperature. Hilsum has estimated that $\langle 111 \rangle$ minima are the next lowest in energy, being 0.51 eV higher than $\langle 000 \rangle$ minimum [33]. In comparing the Gunn effect in CdTe to that in GaAs, a major difference is the substantially higher threshold field, about 13 kV/cm for CdTe compared with about 3 kV/cm for GaAs [34]. Qualitatively, the higher threshold can be thought of as associated with the relatively strong coupling of the electrons to longitudinal optical phonons, which limits the mobility—and hence the rate of energy acquisition from the applied field—and also provides an efficient mechanism for transferring energy to the lattice, thereby minimizing the kinetic energy in the electron distribution.

    The ratio of peak-to-valley current is another parameter of interest. In CdTe, as in GaAs, the spike amplitude can be as large as 50% of the maximum total current. A similar maximum efficiency for CdTe and GaAs can be expected. Since the domain velocities in CdTe and GaAs are approximately equal, samples of the same length will operate at about the same frequency in the transit-time mode. The high threshold field of CdTe combined with its poor thermal conductivity creates a heating problem. If sufficiently short pulses are used so that the heat can be dissipated, however, the high operating field of the sample can be an advantage.

7-7 MICROWAVE GENERATION AND AMPLIFICATION

7-7-1 Microwave Generation

As described earlier in this section, if the applied field is less than threshold the specimen is stable. If, however, the field is greater than threshold, the sample is unstable and divides up into two domains of different conductivity and different electric field but the same drift velocity. Figure 7-7-1 shows the stable and unstable regions.

    At the initial formation of the accumulation layer, the field behind the layer decreases and the field in front of it increases. This process continues as the layer travels from the cathode toward the anode. As the layer approaches the anode, the field
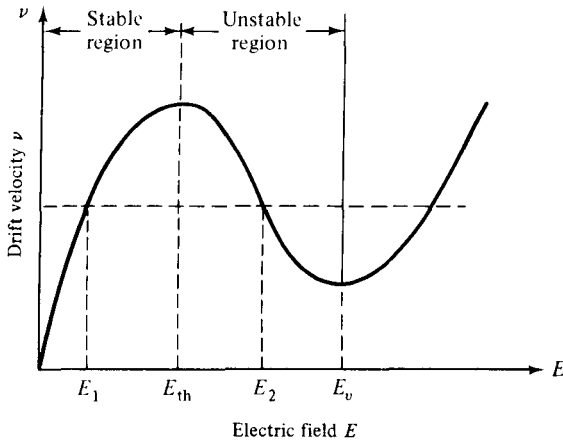
Figure 7-7-1 Electric field versus drift velocity.

behind it begins to increase again; and after the layer is collected by the anode, the field in the whole sample is higher than threshold. When the high-field domain disappears at the anode, a new dipole field starts forming again at the cathode and the process repeats itself. Since current density is proportional to the drift velocity of the electrons, a pulsed current output is obtained. The oscillation frequency of the pulsed current is given by

$$f = \frac{v_d}{L_{\text{eff}}} \tag{7-7-1}$$

where v_d is the velocity of the domain or approximately the drift velocity of the electrons and L_{eff} is the effective length that the domain travels. Experiments have shown that the n-type GaAs diodes have yielded 200-W pulses at 3.05 GHz and 780-mW CW power at 8.7 GHz. Efficiencies of 29% have been obtained in pulsed operation at 3.05 GHz and 5.2% in CW operation at 24.8 GHz. Predictions have been made that 250-kW pulses from a single block of n-type GaAs are theoretically possible up to 100 GHz.

The source generation of solid-state microwave devices has many advantages over the vacuum tube devices they are beginning to replace. However, at present they also have serious drawbacks that could prevent more widespread application. The most important disadvantages are:

1. Low efficiency at frequencies above 10 GHz
2. Small tuning range
3. Large dependence of frequency on temperature
4. High noise

These problems are common to both avalanche diodes and transferred electron devices [35].

Figure 7-7-2 shows the latest state-of-the-art performance for GaAs and InP Gunn diodes [36]. The numbers adjacent to the data points indicate efficiency in percent. Gunn diode oscillators have better noise performance than IMPATTs. They are
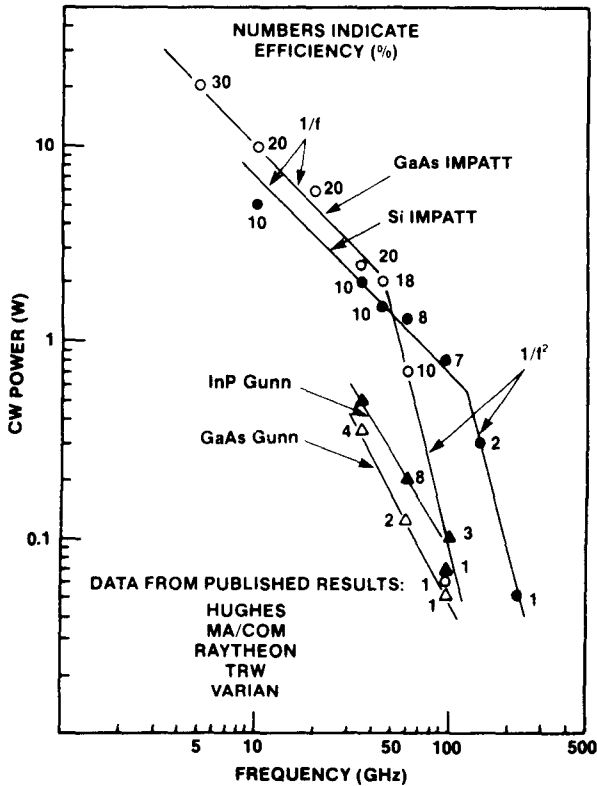
Figure 7-7-2   State-of-the-art perfor-
mance for GaAs and InP Gunn diodes.
(*From H. Hieslmair et al. [36]; reprinted
by permission of* Microwave Journal,
Inc.)

used as local oscillators for receivers and as primary sources where CW powers of
up to 100 mW are required. InP Gunn diodes have higher power and efficiency than
GaAs Gunn diodes.

7-7-2 Microwave Amplification

When an RF signal is applied to a Gunn oscillator, amplification of the signal oc-
curs, provided that the signal frequency is low enough to allow the space charge in
the domain to readjust itself. There is a critical value of fL above which the device
will not amplify. Below this frequency limit the sample presents an impedance with
a negative real part that can be utilized for amplification. If $n_0 L$ becomes less than
$10^{12}/cm^2$, domain formation is inhibited and the device exhibits a nonuniform field
distribution that is stable with respect to time and space. Such a diode can amplify
signals in the vicinity of the transit-time frequency and its harmonics without oscilla-
tion. If this device is used in a circuit with enough positive feedback, it will oscil-
late. Hakki has shown that the oscillation diode can amplify at nearby frequencies or
can be used simultaneously as an amplifier and local oscillator [37]. However, the
output power of a stable amplifier is quite low because of the limitation imposed by
the value of $n_0 L$.

In contrast to the stable amplifier, the Gunn-effect diode must oscillate at the transit-time frequency while it is amplifying at some other frequency. The value of $n_0 L$ must be larger than $10^{12}/\text{cm}^2$ in order to establish traveling domain oscillations; hence substantially larger output power can be obtained. Because of the presence of high-field domains, this amplifier is called a *traveling domain amplifier* (TDA).

Although a large number of possible amplifier circuits exist, the essential feature of each is to provide both a broadband circuit at the signal frequency and a short circuit at the Gunn oscillation frequency. In order to maintain stability with respect to the signal frequency, the Gunn diode must see a source admittance whose real part is larger than the negative conductance of the diode. The simplest circuit satisfying this condition is shown in Fig. 7-7-3. An average gain of 3 dB was exhibited between 5.5 and 6.5 GHz.

Gunn diodes have been used in conjunction with circulator-coupled networks in the design of high-level wideband transferred electron amplifiers that have a voltage gain-bandwidth product in excess of 10 dB for frequencies from 4 to about 16 GHz. Linear gains of 6 to 12 dB per stage and saturated-output-power levels in excess of 0.5 W have been realized [39]. Table 7-7-1 lists the performance of several amplifiers that have been designed since 1970.
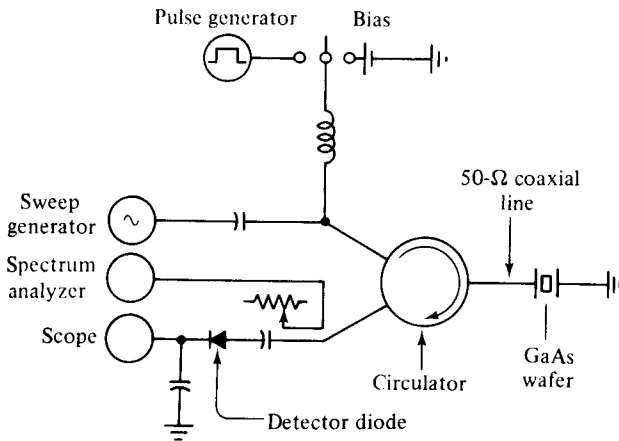
Figure 7-7-3 Gunn-diode amplifier circuit. (*After H. W. Thim [38]; reprinted by permission of IEEE, Inc.*)

TABLE 7-7-1 CW GUNN-DIODE AMPLIFIER PERFORMANCE

Frequency band	3-dB bandwidth (GHz)	Small signal gain (dB)	Power gain (dB)	Efficiency (%)
C	4.5– 8.0	8	3	3
X	7.5–10.75	12	1.65	2.3
	8.0–12.0	6	1.8	2.5
Ku	12.0–16.0	6	1.5	2.5
	13.0–15.0	8	0.36	2

Source: After B.S. Perlman et al. [39]; reprinted by permission of IEEE, Inc.

REFERENCES

[1] SHOCKLEY, W., Negative resistance arising from transit time in semiconductor diodes. *Bell System Tech. J.*, **33**, 799–826, July 1954.

[2] RIDLEY, B. K., and T. B. WATKINS, The possibility of negative resistance effects in semiconductors. *Proc. Phys. Soc.*, **78**, 293–304, August 1961.

[3] HILSUM, C., Transferred electron amplifiers and oscillators. *Proc. IEEE*, **50**, 185–189, February 1962.

[4] GUNN, J. B., Microwave oscillations of current in III-V semiconductors. *Solid-state Communications*, **1**, 89–91, September 1963.

[5] RIDLEY, B. K., Specific negative resistance in solids. *Proc. Phys. Soc. (London)*, **82**, 954–966, December 1963.

[6] KROEMER, H., Theory of the Gunn effect. *Proc. IEEE*, **52**, 1736 (1964).

[7] GUNN, J. B., Microwave oscillations of current in III-V semiconductors. *Solid-state Communications*, **1**, 88–91 (1963).

[8] GUNN, J. B., Instabilities of current in III-V semiconductors. *IBM J. Res. Develop.*, **8**, 141–159, April 1964.

[9] GUNN, J. B., Instabilities of current and of potential distribution in GaAs and InP. *7th Int. Conf. on Physics of Semiconductor "Plasma Effects in Solids,"* 199–207, Tokyo, 1964.

[10] KROEMER, H., Proposed negative-mass microwave amplifier. *Physical Rev.*, **109**, No. 5. 1856, March 1, 1958.

[11] KROEMER, H., The physical principles of a negative-mass amplifier. *Proc. IRE*, **47**, 397–406, March 1959.

[12] COPELAND, J. A., Bulk negative-resistance semiconductor devices. *IEEE Spectrum*, No. 5, 40, May 1967.

[13] BUTCHER, P. N., and W. FAWCETT, Calculation of the velocity-field characteristics of gallium arsenide. *Appl. Phys. Letters*, **21**, 498 (1966).

[14] CONWELL, E. M. and M. O. VASSELL, High-field distribution function in GaAs. *IEEE Trans. on Electron Devices*, **ED-13**, 22 (1966).

[15] RUCH, J. G., and G. S. KINO, Measurement of the velocity-field characteristics of gallium arsenide. *Appl. Phys. Letters*, **10**, 50 (1967).

[16] KROEMER, H., Negative conductance in semiconductors. *IEEE Spectrum*, **5**, No. 1, 47, January 1968.

[17] COPELAND, J. A., Characterization of bulk negative-resistance diode behavior. *IEEE Trans. on Electron Devices*, **ED-14**, No. 9, 436–441, September 1967.

[18] ELLIOTT, B. J., J. B. GUNN, and J. C. McGRODDY, Bulk negative differential conductivity and traveling domains in n-type germanium. *Appl. Phys. Letters*, **11**, 253 (1967).

[19] COPELAND, J. A., Stable space-charge layers in two-valley semiconductors. *J. Appl. Phys.*, **37**, No. 9, 3602, August 1966.

[20] GUNN, J. B., Effect of domain and circuit properties on oscillations in GaAs. *IBM J. Res. Develop.*, 310–320, July 1966.

[21] HOBSON, G. S., Some properties of Gunn-effect oscillations in a biconical cavity. *IEEE Trans. on Electron Devices*, **ED-14**, No. 9, 526–531, September 1967.

[22] THIM, H. W., Computer study of bulk GaAs devices with random one-dimensional doping fluctuations. *J. Appl. Phys.*, **39**, 3897 (1968).

[23] THIM, H. W., and M. R. BARBER, Observation of multiple high-field domains in *n*-GaAs. *Proc. IEEE*, **56,** 110 (1968).

[24] UENOHARA, M., Bulk gallium arsenide devices. Chapter 16 in H. A. Watson (Ed.), *Microwave Semiconductor Devices and Their Circuit Application*. McGraw-Hill Book Company, New York, 1969.

[25] COPELAND, J. A., CW operation of LSA oscillator diodes—44 to 88 GHz. *Bell System Tech. J.*, **46,** 284–287, January 1967.

[26] WILSON, W. E., Pulsed LSA and TRAPATT sources for microwave systems. *Microwave J.*, **14,** No. 8, August 1971, 87–90.

[27] G. B. L., Three-level oscillator in indium phosphide. *Physics Today*, **23,** 19–20, December 1970.

[28] COLLIVER, D., and B. PREW, Indium phosphide: Is it practical for solid state microwave sources? *Electronics*, 110–113, April 10, 1972.

[29] TAYLOR, B. C., and D. J. COLLIVER, Indium phosphide microwave oscillators. *IEEE Trans. on Electron Devices*, **ED-18,** No. 10, 835–840, October 1971.

[30] FOYT, A. G., and A. L. McWHORTER, The Gunn effect in polar semiconductors. *IEEE Trans. on Electron Devices*, **ED-13,** 79–87, January 1966.

[31] LUDWIG, G. W., R. E. HALSTED, and M. AVEN, Current saturation and instability in CdTe and ZnSe. *IEEE Trans. on Electron Devices*, **ED-13,** 671, August–September 1966.

[32] LUDWIG, G. W., Gunn effect in CdTe. *IEEE Trans. on Electron Devices*, **ED-14,** No. 9, 547–551, September 1967.

[33] BUTCHER, P. N., and W. FAWCETT, *Proc. Phys. Soc. (London)*, **86,** 1205 (1965).

[34] OLIVER, M. R., and A. G. FOYT, The Gunn effect in *n*-CdTe. *IEEE Trans. on Electron Devices*, **ED-14,** No. 9, 617–618, September 1967.

[35] HILSUM, C., New developments in transfered electron effects. *Proc. 3rd Conf. on High Frequency Generation and Amplification: Devices and Applications*. August 17–19, 1971, Cornell University.

[36] HIESLMAIR, H., ET AL., State of the art of solid-state and tube transmitters. *Microwave J.*, **26,** No. 10, 46–48, October 1983.

[37] HAKKI, B. W., GaAs post-threshold microwave amplifier, mixer, and oscillator. *Proc. IEEE (Letters)*, **54,** 299–300, February 1966.

[38] THIM, H. W., Linear microwave amplification with Gunn oscillators. *IEEE Trans. on Electron Devices*, **ED-14,** No. 9, 520–526, September 1967.

SUGGESTED READINGS

EASTMAN, L. F., *Gallium Arsenide Microwave Bulk and Transit-Time Devices*. Artech House, Dedham, Mass., 1973.

MILNES, A. G., *Semiconductor Devices and Integrated Electronics*. Van Nostrand Reinhold Company, New York, 1980.

SOOHOO, R. F., *Microwave Electronics*. Addison-Wesley Publishing Company, Reading, Mass., 1971.

SZE, S. M., *Physics of Semiconductor Devices*, 2nd ed. John Wiley & Sons, New York, 1981.

PROBLEMS

Transferred Electron Devices (TEDs)

7-1. **a.** Spell out the following abbreviated terms: LSA, InP, and CdTe.

   b. Describe in detail the principles of the following terms: Gunn effect, high-field domain theory, two-valley theory, and three-valley theory.

   c. Discuss the differences between transferred electron devices and avalanche transit-time devices.

7-2. Describe the operating principles of tunnel diodes, Gunn diodes, and LSA diodes.

7-3. Derive Eq. (7-2-11).

7-4. Describe the modes of operation for Gunn diodes.

7-5. Describe the Ridley–Watkins–Hilsum theory.

7-6. The LSA oscillation mode is defined between 2×10^4 and 2×10^5 ratios of doping over frequency as shown in Fig. 7-3-1. Derive Eqs. (7-3-7) and (7-3-9).

7-7. For a transit-time domain mode, the domain velocity is equal to the carrier drift velocity and is about 10^7 cm/s. Determine the drift length of the diode at a frequency of 8 GHz.

Chapter 8

Avalanche Transit-Time Devices

8-0 INTRODUCTION

Avalanche transit-time diode oscillators rely on the effect of voltage breakdown across a reversé-biased *p-n* junction to produce a supply of holes and electrons. Ever since the development of modern semiconductor device theory scientists have speculated on whether it is possible to make a two-terminal negative-resistance device. The tunnel diode was the first such device to be realized in practice. Its operation depends on the properties of a forward-biased *p-n* junction in which both the *p* and *n* regions are heavily doped. The other two devices are the transferred electron devices and the avalanche transit-time devices. In this chapter the latter type is discussed.

The transferred electron devices or the Gunn oscillators operate simply by the application of a dc voltage to a bulk semiconductor. There are no *p-n* junctions in this device. Its frequency is a function of the load and of the natural frequency of the circuit. The avalanche diode oscillator uses carrier impact ionization and drift in the high-field region of a semiconductor junction to produce a negative resistance at microwave frequencies. The device was originally proposed in a theoretical paper by Read [1] in which he analyzed the negative-resistance properties of an idealized n^+-p-i-p^+ diode. Two distinct modes of avalanche oscillator have been observed. One is the IMPATT mode, which stands for *impact ionization avalanche transit-time* operation. In this mode the typical dc-to-RF conversion efficiency is 5 to 10%, and frequencies are as high as 100 GHz with silicon diodes. The other mode is the TRAP-ATT mode, which represents *trapped plasma avalanche triggered transit* operation. Its typical conversion efficiency is from 20 to 60%.

Another type of active microwave device is the BARITT (*barrier injected transit-time*) diode [2]. It has long drift regions similar to those of IMPATT diodes. The carriers traversing the drift regions of BARITT diodes, however, are generated

by minority carrier injection from forward-biased junctions rather than being extracted from the plasma of an avalanche region. Several different structures have been operated as BARITT diodes, such as *p-n-p, p-n-v-p, p-n-*metal, and metal-*n-*metal. BARITT diodes have low noise figures of 15 dB, but their bandwidth is relatively narrow with low output power.

8-1 READ DIODE

8-1-1 Physical Description

The basic operating principle of IMPATT diodes can be most easily understood by reference to the first proposed avalanche diode, the Read diode [1]. The theory of this device was presented by Read in 1958, but the first experimental Read diode was reported by Lee et al. in 1965 [3]. A mode of the original Read diode with a doping profile and a dc electric field distribution that exists when a large reverse bias is applied across the diode is shown in Fig. 8-1-1.
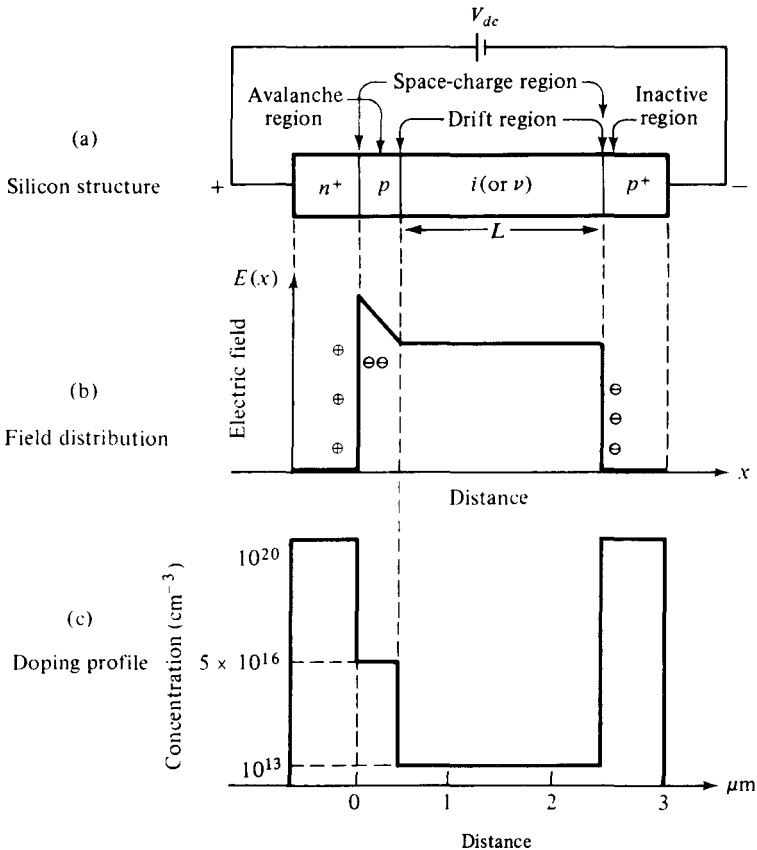
Figure 8-1-1   Read diode.

The Read diode is an n^+-p-i-p^+ structure, where the superscript plus sign denotes very high doping and the i or v refers to intrinsic material. The device consists essentially of two regions. One is the thin p region at which avalanche multiplication occurs. This region is also called the high-field region or the avalanche region. The other is the i or v region through which the generated holes must drift in moving to the p^+ contact. This region is also called the intrinsic region or the drift region. The p region is very thin. The space between the n^+-p junction and the i-p^+ junction is called the space-charge region. Similar devices can be built in the p^+-n-i-n^+ structure, in which electrons generated from avalanche multiplication drift through the i region.

The Read diode oscillator consists of an n^+-p-i-p^+ diode biased in reverse and mounted in a microwave cavity. The impedance of the cavity is mainly inductive and is matched to the mainly capacitive impedance of the diode to form a resonant circuit. The device can produce a negative ac resistance that, in turn, delivers power from the dc bias to the oscillation.

8-1-2 Avalanche Multiplication

When the reverse-biased voltage is well above the punchthrough or breakdown voltage, the space-charge region always extends from the n^+-p junction through the p and i regions to the i-p^+ junction. The fixed charges in the various regions are shown in Fig. 8-1-1(b). A positive charge gives a rising field in moving from left to right. The maximum field, which occurs at the n^+-p junction, is about several hundred kilovolts per centimeter. Carriers (holes) moving in the high field near the n^+-p junction acquire energy to knock valence electrons into the conduction band, thus producing hole-electron pairs. The rate of pair production, or avalanche multiplication, is a sensitive nonlinear function of the field. By proper doping, the field can be given a relatively sharp peak so that avalanche multiplication is confined to a very narrow region at the n^+-p junction. The electrons move into the n^+ region and the holes drift through the space-charge region to the p^+ region with a constant velocity v_d of about 10^7 cm/s for silicon. The field throughout the space-charge region is above about 5 kV/cm. The transit time of a hole across the drift i-region L is given by

$$\tau = \frac{L}{v_d} \tag{8-1-1}$$

and the avalanche multiplication factor is

$$M = \frac{1}{1 - (V/V_b)^n} \tag{8-1-1a}$$

where V = applied voltage
    V_b = avalanche breakdown voltage
    n = 3–6 for silicon is a numerical factor depending on the doping of p^+-n or n^+-p junction

The breakdown voltage for a silicon p^+-n junction can be expressed as

$$|V_b| = \frac{\rho_n \mu_n \epsilon_s |E_{\max}|_b^2}{2} \qquad (8\text{-}1\text{-}1b)$$

where ρ_n = resistivity
μ_n = electron mobility
ϵ_s = semiconductor permittivity
$E_{\max}$ = maximum breakdown of the electric field

Figure 8-1-2 shows the avalanche breakdown voltage as a function of impurity at a p^+-n junction for several semiconductors.
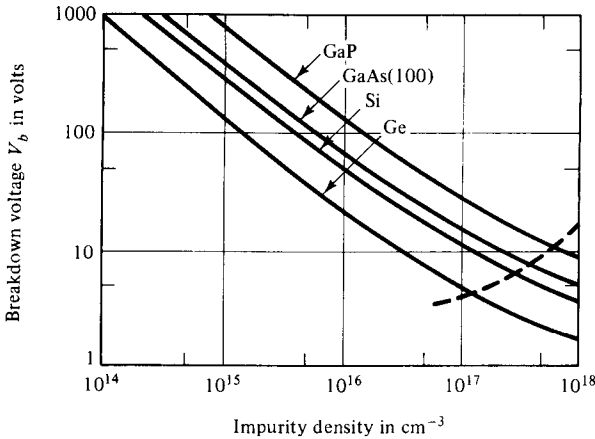
Figure 8-1-2   Breakdown voltage versus impurity doping.

8-1-3 Carrier Current I₀(t) and External Current Iₑ(t)

As described previously, the Read diode is mounted in a microwave resonant circuit. An ac voltage can be maintained at a given frequency in the circuit, and the total field across the diode is the sum of the dc and ac fields. This total field causes breakdown at the n^+-p junction during the positive half of the ac voltage cycle if the field is above the breakdown voltage, and the carrier current (or the hole current in this case) $I_0(t)$ generated at the n^+-p junction by the avalanche multiplication grows exponentially with time while the field is above the critical value. During the negative half cycle, when the field is below the breakdown voltage, the carrier current $I_0(t)$ decays exponentially to a small steady-state value. The carrier current $I_0(t)$ is the current at the junction only and is in the form of a pulse of very short duration as shown in Fig. 8-1-3(d). Therefore the carrier current $I_0(t)$ reaches its maximum in the middle of the ac voltage cycle, or one-quarter of a cycle later than the voltage. Under the influence of the electric field the generated holes are injected into the space-charge region toward the negative terminal. As the injected holes traverse the drift space, they induce a current $I_e(t)$ in the external circuit as shown in Fig. 8-1-3(d).
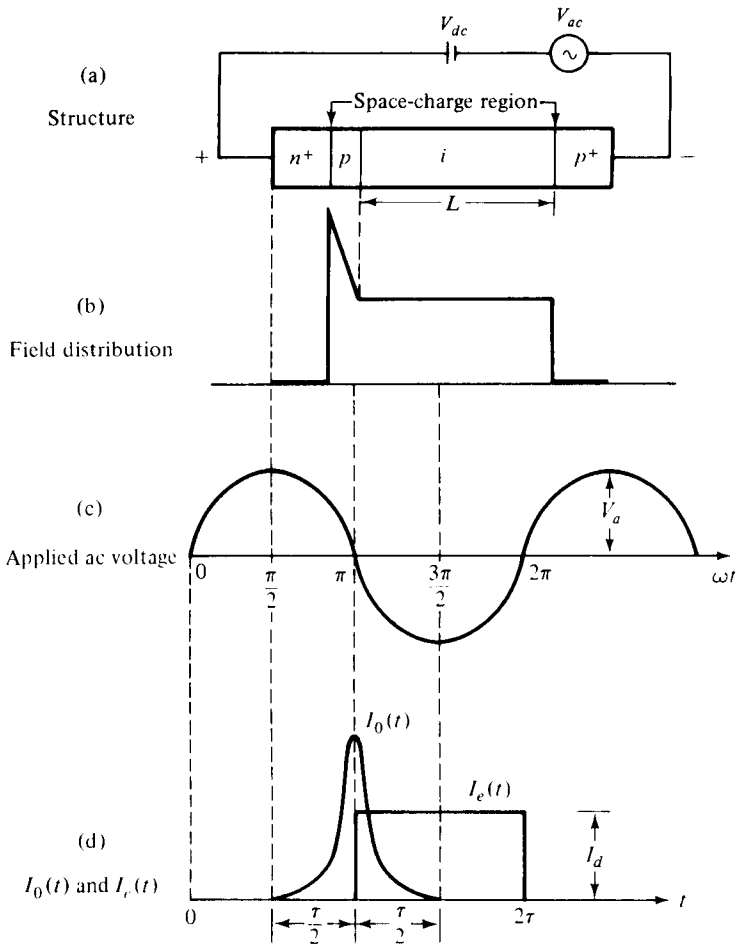
Figure 8-1-3 Field, voltage, and currents in Read diode. (*After Read [1];* *reprinted by permission of the Bell System, AT&T Co.*)

When the holes generated at the n^+-p junction drift through the space-charge region, they cause a reduction of the field in accordance with Poisson's equation:

$$\frac{dE}{dx} = -\frac{\rho}{\epsilon_s} \qquad (8\text{-}1\text{-}2)$$

where ρ is the volume charge density and ϵ_s is the semiconductor permittivity.

Since the drift velocity of the holes in the space-charge region is constant, the induced current $I_e(t)$ in the external circuit is simply equal to

$$I_e(t) = \frac{Q}{\tau} = \frac{v_d Q}{L} \qquad (8\text{-}1\text{-}3)$$

where Q = total charge of the moving holes
  v_d = hole drift velocity
  L = length of the drift i region

It can be seen that the induced current $I_e(t)$ in the external circuit is equal to the average current in the space-charge region. When the pulse of hole current $I_0(t)$ is suddenly generated at the n^+-p junction, a constant current $I_e(t)$ starts flowing in the external circuit and continues to flow during the time τ in which the holes are moving across the space-charge region. Thus, on the average, the external current $I_e(t)$ because of the moving holes is delayed by $\tau/2$ or 90° relative to the pulsed carrier current $I_0(t)$ generated at the n^+-p junction. Since the carrier $I_0(t)$ is delayed by one-quarter of a cycle or 90° relative to the ac voltage, the external current $I_e(t)$ is then delayed by 180° relative to the voltage as shown in Fig. 8-1-3(d). Therefore the cavity should be tuned to give a resonant frequency as

$$2\pi f = \frac{\pi}{\tau}$$

Then

$$f = \frac{1}{2\tau} = \frac{v_d}{2L} \tag{8-1-4}$$

Since the applied ac voltage and the external current $I_e(t)$ are out of phase by 180°, negative conductance occurs and the Read diode can be used for microwave oscillation and amplification. For example, taking $v_d = 10^7$ cm/s for silicon, the optimum operating frequency for a Read diode with an i-region length of 2.5 μm is 20 GHz.

8-1-4 Output Power and Quality Factor Q

The external current $I_e(t)$ approaches a square wave, being very small during the positive half cycle of the ac voltage and almost constant during the negative half cycle. Since the direct current I_d supplied by the dc bias is the average external current or conductive current, it follows that the amplitude of variation of $I_e(t)$ is approximately equal to I_d. If V_a is the amplitude of the ac voltage, the ac power delivered is found to be

$$P = 0.707 V_a I_d \qquad \text{W/unit area} \tag{8-1-5}$$

The quality factor Q of a circuit is defined as

$$Q = \omega \frac{\text{maximum stored energy}}{\text{average dissipated power}} \tag{8-1-6}$$

Since the Read diode supplies ac energy, it has a negative Q in contrast to the positive Q of the cavity. At the stable operating point, the negative Q of the diode is equal to the positive Q of the cavity circuit. If the amplitude of the ac voltage increases, the stored energy, or energy of oscillation, increases faster than the energy

delivered per cycle. This is the condition required in order for stable oscillation to be possible.

8-2 IMPATT DIODES

8-2-1 Physical Structures

A theoretical Read diode made of an n^+-p-i-p^+ or p^+-n-i-n^+ structure has been analyzed. Its basic physical mechanism is the interaction of the impact ionization avalanche and the transit time of charge carriers. Hence the Read-type diodes are called IMPATT diodes. These diodes exhibit a differential negative resistance by two effects:

1. The impact ionization avalanche effect, which causes the carrier current $I_0(t)$ and the ac voltage to be out of phase by 90°
2. The transit-time effect, which further delays the external current $I_e(t)$ relative to the ac voltage by 90°

The first IMPATT operation as reported by Johnston et al. [4] in 1965, however, was obtained from a simple p-n junction. The first real Read-type IMPATT diode was reported by Lee et al. [3], as described previously. From the small-signal theory developed by Gilden [5] it has been confirmed that a negative resistance of the IMPATT diode can be obtained from a junction diode with any doping profile. Many IMPATT diodes consist of a high doping avalanching region followed by a drift region where the field is low enough that the carriers can traverse through it without avalanching. The Read diode is the basic type in the IMPATT diode family. The others are the one-sided abrupt p-n junction, the linearly graded p-n junction (or double-drift region), and the p-i-n diode, all of which are shown in Fig. 8-2-1. The principle of operation of these devices, however, is essentially similar to the mechanism described for the Read diode.

8-2-2 Negative Resistance

Small-signal analysis of a Read diode results in the following expression for the real part of the diode terminal impedance [5]:

$$R = R_s + \frac{2L^2}{v_d \epsilon_s A} \frac{1}{1 - \omega^2/\omega_r^2} \frac{1 - \cos \theta}{\theta} \qquad (8\text{-}2\text{-}1)$$

where R_s = passive resistance of the inactive region
v_d = carrier drift velocity
L = length of the drift space-charge region
A = diode cross section
e_s = semiconductor dielectric permittivity

(a) Abrupt *p–n* junction

| p^+ | n | n^+ |
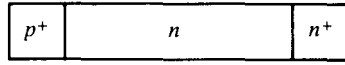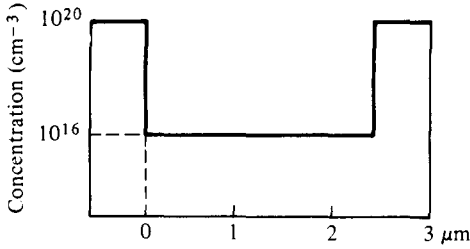
Doping profile

(b) Linearly graded
    p–n junction

| p^+ | p | n | n^+ |

Doping profile

(c) *p–i–n* diode

| p^+ | i | n^+ |
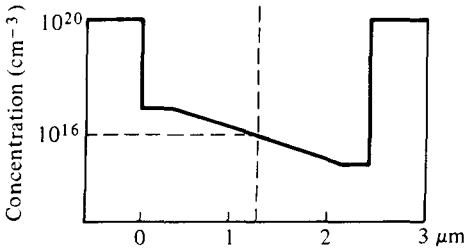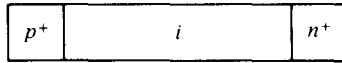
Doping profile
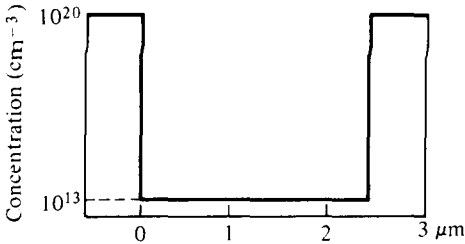
Figure 8-2-1   Three typical silicon
IMPATT diodes. (*After R. L. Johnston et
al. [4]; reprinted by permission of the
Bell System, AT&T Co.*)

Moreover, θ is the transit angle, given by

$$\theta = \omega\tau = \omega\frac{L}{v_d} \tag{8-2-2}$$

and ω_r is the avalanche resonant frequency, defined by

$$\omega_r \equiv \left(\frac{2\alpha' v_d I_0}{\epsilon_s A}\right)^{1/2} \tag{8-2-3}$$

In Eq. (8-2-3) the quantity α' is the derivative of the ionization coefficient with respect to the electric field. This coefficient, the number of ionizations per centimeter produced by a single carrier, is a sharply increasing function of the electric field. The variation of the negative resistance with the transit angle when $\omega > \omega_r$ is plotted in Fig. 8-2-2. The peak value of the negative resistance occurs near $\theta = \pi$. For transit angles larger than π and approaching $3\pi/2$, the negative resistance of the diode decreases rapidly. For practical purposes, the Read-type IMPATT diodes work well only in a frequency range around the π transit angle. That is,

$$f = \frac{1}{2\tau} = \frac{v_d}{2L} \tag{8-2-4}$$
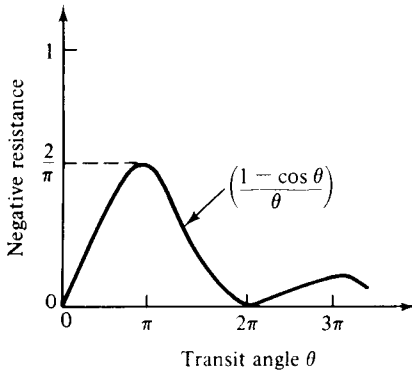
Figure 8-2-2 Negative resistance versus transit angle.

8-2-3 Power Output and Efficiency

At a given frequency the maximum output power of a single diode is limited by semiconductor materials and the attainable impedance levels in microwave circuitry. For a uniform avalanche, the maximum voltage that can be applied across the diode is given by

$$V_m = E_m L \tag{8-2-5}$$

where L is the depletion length and E_m is the maximum electric field. This maximum applied voltage is limited by the breakdown voltage. Furthermore, the maximum current that can be carried by the diode is also limited by the avalanche breakdown process, for the current in the space-charge region causes an increase in the electric field. The maximum current is given by

$$I_m = J_m A = \sigma E_m A = \frac{\epsilon_s}{\tau} E_m A = \frac{v_d \epsilon_s E_m A}{L} \tag{8-2-6}$$

Therefore the upper limit of the power input is given by

$$P_m = I_m V_m = E_m^2 \epsilon_s v_d A \tag{8-2-7}$$

The capacitance across the space-charge region is defined as

$$C = \frac{\epsilon_s A}{L} \tag{8-2-8}$$

Substitution of Eq. (8-2-8) in Eq. (8-2-7) and application of $2\pi f\tau = 1$ yield

$$P_m f^2 = \frac{E_m^2 v_d^2}{4\pi^2 X_c} \tag{8-2-9}$$

It is interesting to note that this equation is identical to Eq. (5-1-60) of the power-frequency limitation for the microwave power transistor. The maximum power that can be given to the mobile carriers decreases as $1/f^2$. For silicon, this electronic limit is dominant at frequencies as high as 100 GHz. The efficiency of the IMPATT diodes is given by

$$\eta = \frac{P_{ac}}{P_{dc}} = \left(\frac{V_a}{V_d}\right)\left(\frac{I_a}{I_d}\right) \tag{8-2-10}$$
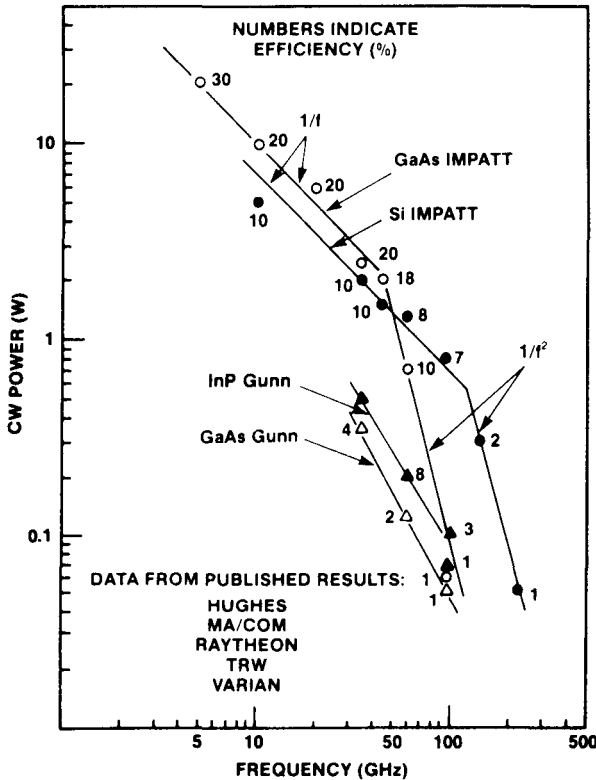
Figure 8-2-3 State-of-the-art performance for GaAs and Si IMPATTs. (*From H. Hieslmair, et al. [6]; reprinted by permission of* Microwave Journal.)

For an ideal Read-type IMPATT diode, the ratio of the ac voltage to the applied voltage is about 0.5 and the ratio of the ac current to the dc current is about $2/\pi$, so that the efficiency would be about $1/\pi$ or more than 30%. For practical IMPATT diodes, however, the efficiency is usually less than 30% because of the space-charge effect, the reverse-saturation-current effect, the high-frequency-skin effect, and the ionization-saturation effect.

IMPATT diodes are at present the most powerful CW solid-state microwave power sources. The diodes have been fabricated from germanium, silicon, and gallium arsenide and can probably be constructed from other semiconductors as well. IMPATT diodes provide potentially reliable, compact, inexpensive, and moderately efficient microwave power sources.

Figure 8-2-3 shows the latest state-of-the-art performance for GaAs and Si IMPATTs [6]. The numbers adjacent to the data points indicate efficiency in percent. Power output data for both the GaAs and Si IMPATTs closely follow the $1/f$ and $1/f^2$ slopes. The transition from the $1/f$ to the $1/f^2$ slope for GaAs falls between 50 and 60 GHz, and that for Si IMPATTs between 100 and 120 GHz. GaAs IMPATTs show higher power and efficiency in the 40- to 60-GHz region whereas Si IMPATTs are produced with higher reliability and yield in the same frequency region. On the contrary, the GaAs IMPATTs have higher powers and efficiencies below 40 GHz than do Si IMPATTs. Above 60 GHz the Si IMPATTs seem to outperform the GaAs devices.

Example 8-2-1:   CW Output Power of an IMPATT Diode

An IMPATT diode has the following parameters:

Carrier drift velocity:	$v_d = 2 \times 10^7$ cm/s
Drift-region length:	$L = 6~\mu$m
Maximum operating voltage:	$V_{0max} = 100$ V
Maximum operating current:	$I_{0max} = 200$ mA
Efficiency:	$\eta = 15\%$
Breakdown voltage:	$V_{bd} = 90$ V

Compute: (a) the maximum CW output power in watts; (b) the resonant frequency in gigahertz.

Solution

a. From Eq. (8-2-10) the CW output power is

$$P = \eta P_{dc} = 0.15 \times 100 \times 0.2 = 3\text{W}$$

b. From Eq. (8-2-4) the resonant frequency is

$$f = \frac{v_d}{2L} = \frac{2 \times 10^5}{2 \times 6 \times 10^{-6}} = 16.67\text{ GHz}$$

8-3 TRAPATT DIODES

8-3-1 Physical Structures

The abbreviation TRAPATT stands for *trapped plasma avalanche triggered transit* mode, a mode first reported by Prager et al. [7]. It is a high-efficiency microwave generator capable of operating from several hundred megahertz to several gigahertz. The basic operation of the oscillator is a semiconductor *p-n* junction diode reverse-biased to current densities well in excess of those encountered in normal avalanche operation. High-peak-power diodes are typically silicon n^+-p-p^+ (or p^+-n-n^+) structures with the *n*-type depletion region width varying from 2.5 to 12.5 μm. The doping of the depletion region is generally such that the diodes are well "punched through" at breakdown; that is, the dc electric field in the depletion region just prior to breakdown is well above the saturated drift-velocity level. The device's p^+ region is kept as thin as possible at 2.5 to 7.5 μm. The TRAPATT diode's diameter ranges from as small as 50 μm for CW operation to 750 μm at lower frequency for high-peak-power devices.

8-3-2 Principles of Operation

Approximate analytic solutions for the TRAPATT mode in p^+-n-n^+ diodes have been developed by Clorfeine et al. [8] and DeLoach [9] among others. These analyses have shown that a high-field avalanche zone propagates through the diode and fills the depletion layer with a dense plasma of electrons and holes that become trapped in the low-field region behind the zone. A typical voltage waveform for the TRAPATT mode of an avalanche p^+-n-n^+ diode operating with an assumed square-wave current drive is shown in Fig. 8-3-1. At point A the electric field is uniform throughout the sample and its magnitude is large but less than the value required for avalanche breakdown. The current density is expressed by

$$J = \epsilon_s \frac{dE}{dt} \tag{8-3-1}$$

where ϵ_s is the semiconductor dielectric permittivity of the diode.

At the instant of time at point A, the diode current is turned on. Since the only charge carriers present are those caused by the thermal generation, the diode initially charges up like a linear capacitor, driving the magnitude of the electric field above the breakdown voltage. When a sufficient number of carriers is generated, the particle current exceeds the external current and the electric field is depressed throughout the depletion region, causing the voltage to decrease. This portion of the cycle is shown by the curve from point B to point C. During this time interval the electric field is sufficiently large for the avalanche to continue, and a dense plasma of electrons and holes is created. As some of the electrons and holes drift out of the ends of the depletion layer, the field is further depressed and "traps" the remaining plasma. The voltage decreases to point D. A long time is required to remove the plasma because the total plasma charge is large compared to the charge per unit time
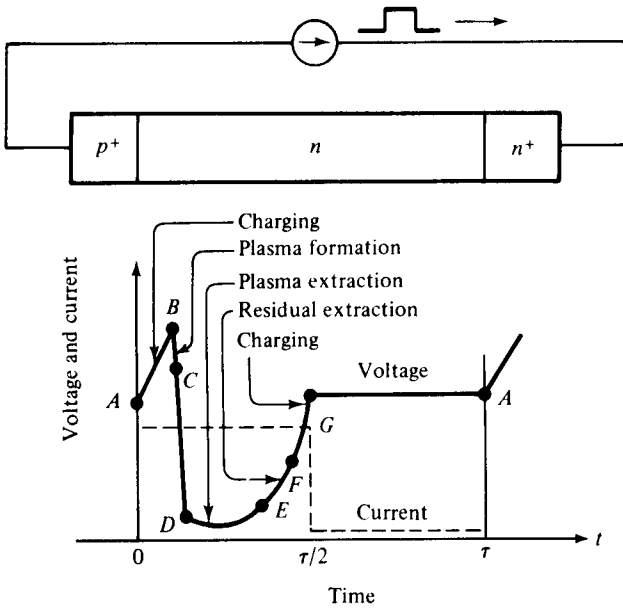
Figure 8-3-1   Voltage and current waveforms for TRAPATT diode. (*After A. S. Clorfeine et al. [8]; reprinted by permission of RCA Laboratory.*)

in the external current. At point E the plasma is removed, but a residual charge of electrons remains in one end of the depletion layer and a residual charge of holes in the other end. As the residual charge is removed, the voltage increases from point E to point F. At point F all the charge that was generated internally has been removed. This charge must be greater than or equal to that supplied by the external current; otherwise the voltage will exceed that at point A. From point F to point G the diode charges up again like a fixed capacitor. At point G the diode current goes to zero for half a period and the voltage remains constant at V_A until the current comes back on and the cycle repeats. The electric field can be expressed as

$$E(x, t) = E_m - \frac{qN_A}{\epsilon_s}x + \frac{Jt}{\epsilon_s} \qquad (8\text{-}3\text{-}2)$$

where N_A is the doping concentration of the n region and x is the distance.

Thus the value of t at which the electric field reaches E_m at a given distance x into the depletion region is obtained by setting $E(x, t) = E_m$, yielding

$$t = \frac{qN_A}{J}x \qquad (8\text{-}3\text{-}3)$$

Differentiation of Eq. (8-3-3) with respect to time t results in

$$v_z \equiv \frac{dx}{dt} = \frac{J}{qN_A} \qquad (8\text{-}3\text{-}4)$$

where v_z is the avalanche-zone velocity.

Example 8-3-1:   Avalanche-Zone Velocity of a TRAPATT Diode

A TRAPATT diode has the following parameters:

Doping concentration: $N_A = 2 \times 10^{15}$ cm^{-3}
Current density: $J = 20$ kA/cm2

Calculate the avalanche-zone velocity.

Solution   From Eq. (8-3-4) the avalanche-zone velocity is

$$v_z = \frac{J}{qN_A} = \frac{20 \times 10^3}{1.6 \times 10^{-19} \times 2 \times 10^{15}} = 6.25 \times 10^7 \text{ cm/s}$$

This means that the avalanche-zone velocity is much larger than the scattering-limited velocity.

Thus the avalanche zone (or avalanche shock front) will quickly sweep across most of the diode, leaving the diode filled by a highly conductng plasma of holes and electrons whose space charge depresses the voltage to low values. Because of the dependence of the drift velocity on the field, the electrons and holes will drift at velocities determined by the low-field mobilities, and the transit time of the carriers can become much longer than

$$\tau_s = \frac{L}{v_s} \tag{8-3-5}$$

where v_s is the saturated carrier drift velocity.

Thus the TRAPATT mode can operate at comparatively low frequencies, since the discharge time of the plasma—that is, the rate Q/I of its charge to its current—can be considerably greater than the nominal transit time τ_s of the diode at high field. Therefore the TRAPATT mode is still a transit-time mode in the real sense that the time delay of carriers in transit (that is, the time between injection and collection) is utilized to obtain a current phase shift favorable for oscillation.

8-3-3 Power Output and Efficiency

RF power is delivered by the diode to an external load when the diode is placed in a proper circuit with a load. The main function of this circuit is to match the diode effective negative resistance to the load at the output frequency while reactively terminating (trapping) frequencies above the oscillation frequency in order to ensure TRAPATT operation. To date, the highest pulse power of 1.2 kW has been obtained at 1.1 GHz (five diodes in series) [10], and the highest efficiency of 75% has been achieved at 0.6 GHz (11). Table 8-3-1 shows the current state of TRAPATT diodes [12].

The TRAPATT operation is a rather complicated means of oscillation, however, and requires good control of both device and circuit properties. In addition,

TABLE 8-3-1 TRAPATT OSCILLATOR CAPABILITIES

Frequency (GHz)	Peak power (W)	Average power (W)	Operating voltage (V)	Efficiency (%)
0.5	600	3	150	40
1.0	200	1	110	30
1.0	400	2	110	35
2.0	100	1	80	25
2.0	200	2	80	30
4.0	100	1	80	20
8.0	50	1	60	15

Source: After W. E. Wilson [12]; reprinted by permission of Horizon House.

the TRAPATT mode generally exhibits a considerably higher noise figure than the IMPATT mode, and the upper operating frequency appears to be practically limited to below the millimeter wave region.

8-4 BARITT DIODES

8-4-1 Physical Description

BARITT diodes, meaning *barrier injected transit-time* diodes, are the latest addition to the family of active microwave diodes. They have long drift regions similar to those of IMPATT diodes. The carriers traversing the drift regions of BARITT diodes, however, are generated by minority carrier injection from forward-biased junctions instead of being extracted from the plasma of an avalanche region.

Several different structures have been operated as BARITT diodes, including p-n-p, p-n-v-p, p-n-metal, and metal-n-metal. For a p-n-v-p BARITT diode, the forward-biased p-n junction emits holes into the ν region. These holes drift with saturation velocity through the ν region and are collected at the p contact. The diode exhibits a negative resistance for transit angles between π and 2π. The optimum transit angle is approximately 1.6π.

Such diodes are much less noisy than IMPATT diodes. Noise figures are as low as 15 dB at C-band frequencies with silicon BARITT amplifiers. The major disadvantages of BARITT diodes are relatively narrow bandwidth and power outputs limited to a few milliwatts.

8-4-2 Principles of Operation

A crystal n-type silicon wafer with 11 Ω-cm resistivity and 4×10^{14} per cubic centimeter doping is made of a 10-μm thin slice. Then the n-type silicon wafer is sandwiched between two PtSi Schottky barrier contacts of about 0.1 μm thickness. A schematic diagram of a metal-n-metal structure is shown in Fig. 8-4-1(a).

The energy-band diagram at thermal equilibrium is shown in Fig. 8-4-1(b), where ϕ_{n1} and ϕ_{n2} are the barrier heights for the metal-semiconductor contacts, re-
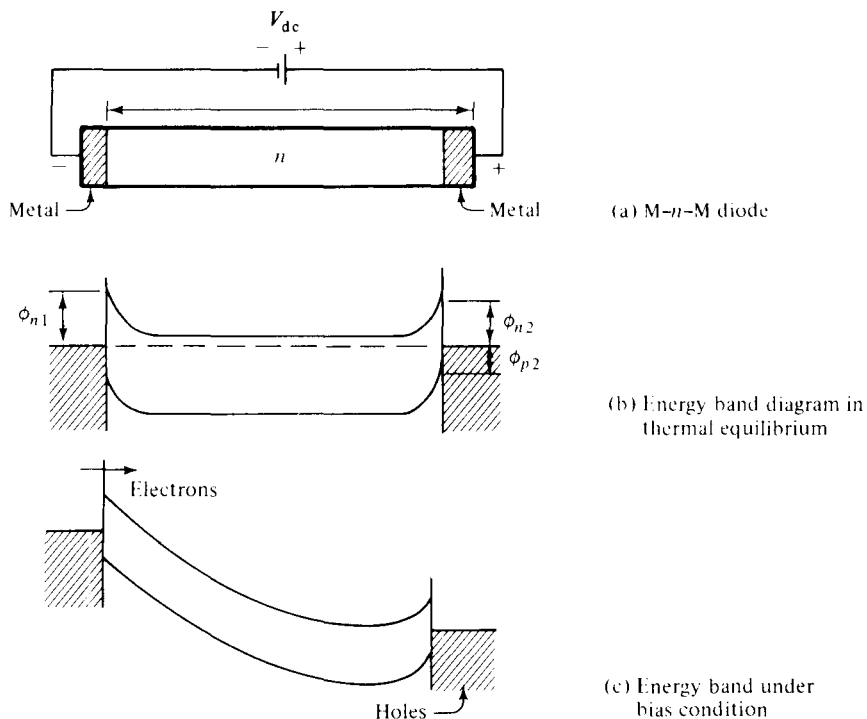
Figure 8-4-1 M-*n*-M diode. (*After D. J. Coleman and S. M. Sze [13]; reprinted by permission of the Bell System, AT&T Co.*)

spectively. For the PtSi–Si–PtSi structure mentioned previously, $\phi_{n1} = \phi_{n2} = 0.85$ eV. The hole barrier height ϕ_{p2} for the forward-biased contact is about 0.15 eV. Figure 8-4-1(c) shows the energy-band diagram when a voltage is applied. The mechanisms responsible for the microwave oscillations are derived from:

1. The rapid increase of the carrier injection process caused by the decreasing potential barrier of the forward-biased metal-semiconductor contact
2. An apparent $3\pi/2$ transit angle of the injected carrier that traverses the semiconductor depletion region

The rapid increase in terminal current with applied voltage (above 30 V) as shown in Fig. 8-4-2 is caused by thermionic hole injection into the semiconductor as the depletion layer of the reverse-biased contact reaches through the entire device thickness. The critical voltage is approximately given by

$$V_c = \frac{qNL^2}{2\epsilon_s} \tag{8-4-1}$$

where N = doping concentration
L = semiconductor thickness
ϵ_s = semiconductor dielectric permittivity
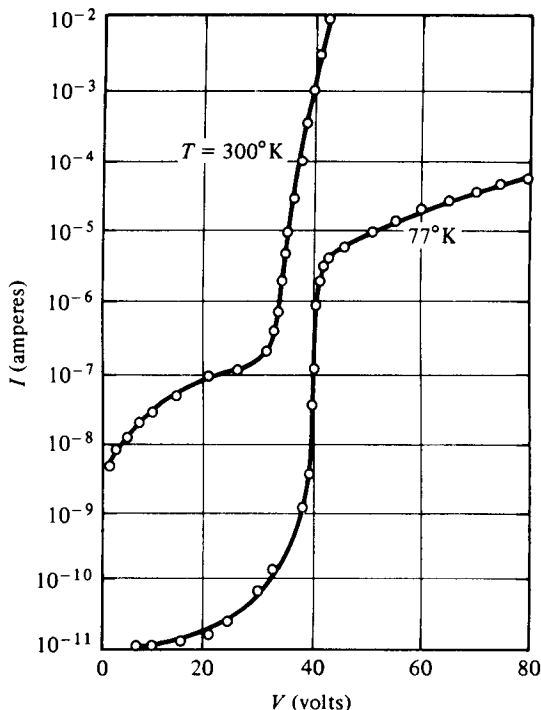
Figure 8-4-2 Current versus voltage of a BARITT diode (PtSi–Si–PtSi). (*After D. J. Coleman and S. M. Sze [13]; reprinted by permission of the Bell System, AT&T Co.*)

The current-voltage characteristics of the silicon MSM structure (PtSi–Si–PtSi) were measured at 77° K and 300° K. The device parameters are $L = 10$ μm, $N = 4 \times 10^{14}$ cm^3, $\phi_{n1} = \phi_{n2} = 0.85$ eV, and area $= 5 \times 10^{-4}$ cm^2.

The current increase is not due to avalanche multiplication, as is apparent from the magnitude of the critical voltage and its negative temperature coefficient. At 77° K the rapid increase is stopped at a current of about 10^{-5} A. This saturated current is expected in accordance with the thermionic emission theory of hole injection from the forward-biased contact with a hole barrier height (ϕ_{p2}) of about 0.15 eV.

8-4-3 Microwave Performance

Continuous-wave (CW) microwave performance of the M-n-M-type BARITT diode was obtained over the entire C band of 4 to 8 GHz. The maximum power observed was 50 mW at 4.9 GHz. The maximum efficiency was about 1.8%. The FM single-sideband noise measure at 1 MHz was found to be 22.8 dB at a 7-mA bias current. This noise measure is substantially lower than that of a silicon IMPATT diode and is comparable to that of a GaAs transfer-electron oscillator. Figure 8-4-3 shows some of the measured microwave power versus current with frequency of operation indicated on each curve for three typical devices tested.

The voltage enclosed in parentheses for each curve indicates the average bias voltage at the diode while the diode is in oscillation. The gain-bandwidth product of a 6-GHz BARITT diode was measured to be 19-dB gain at 5-mA bias current at 200 MHz. The small-signal noise measure was about 15 dB.
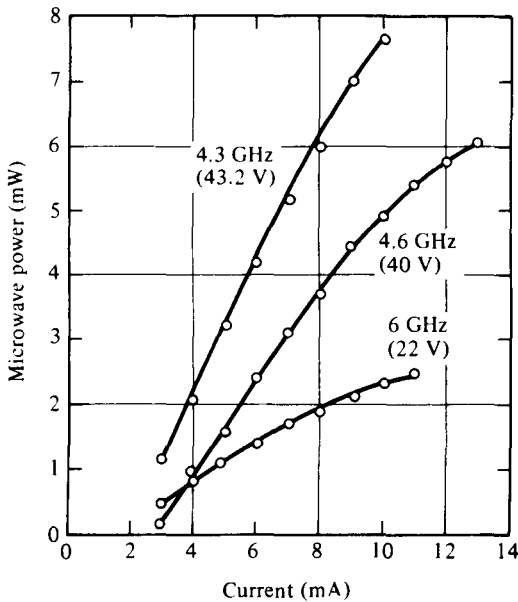
Figure 8-4-3 Power output versus current for three silicon M-n-M devices. (*After D. J. Coleman and S. M. Sze [13]; reprinted by permission of the Bell System, AT&T Co.*)

Example 8-4-1:   Breakdown Voltage of a BARITT Diode

An M-Si-M BARITT diode has the following parameters:

Relative dielectric constant of silicon: $\epsilon_r = 11.8$
Donor concentration: $N = 2.8 \times 10^{21}$ m^{-3}
Silicon length: $L = 6~\mu$m

Determine: **a.** the breakdown voltage; **b.** the breakdown electric field.

Solution

a. From Eq. (8-4-1) the breakdown voltage is double its critical voltage as

$$V_{bd} = \frac{qNL^2}{\epsilon_s} = \frac{1.6 \times 10^{-19} \times 2.8 \times 10^{21} \times (6 \times 10^{-6})^2}{8.854 \times 10^{-12} \times 11.8} = 154.36 \text{ V}$$

b. The breakdown electric field is

$$E_{bd} = \frac{V_{bd}}{L} = \frac{154.36}{6 \times 10^{-6}} = 2.573 \times 10^7 \text{ V/m} = 2.57 \times 10^5 \text{ V/cm}$$

8-5 PARAMETRIC DEVICES

8-5-1 Physical Description

A parametric device is one that uses a nonlinear reactance (capacitance or inductance) or a time-varying reactance. The word *parametric* is derived from the term *parametric excitation,* since the capacitance or inductance, which is a reactive

parameter, can be used to produce capacitive or inductive excitation. Parametric excitation can be subdivided into parametric amplification and oscillation. Many of the essential properties of nonlinear energy-storage systems were described by Faraday [14] as early as 1831 and by Lord Rayleigh [15] in 1883. The first analysis of the nonlinear capacitance was given by van der Ziel [16] in 1948. In his paper van der Ziel first suggested that such a device might be useful as a low-noise amplifier, since it was essentially a reactive device in which no thermal noise is generated. In 1949 Landon [17] analyzed and presented experimental results of such circuits used as amplifiers, converters, and oscillators. In the age of solid-state electronics, microwave electronics engineers dreamed of a solid-state microwave device to replace the noisy electron beam amplifier. In 1957 Suhl [18] proposed a microwave solid-state amplifier that used ferrite. The first realization of a microwave parametric amplifier following Suhl's proposal was made by Weiss [19] in 1957. After the work done by Suhl and Weiss, the parametric amplifier was at last discovered.

At present the solid-state varactor diode is the most widely used parametric amplifier. Unlike microwave tubes, transistors, and lasers, the parametric diode is of a reactive nature and thus generates a very small amount of Johnson noise (thermal noise). One of the distinguishing features of a parametric amplifier is that it utilizes an ac rather than a dc power supply as microwave tubes do. In this respect, the parametric amplifier is analogous to the quantum amplifier laser or maser in which an ac power supply is used.

8-5-2 Nonlinear Reactance and Manley–Rowe Power Relations

A *reactance* is defined as a circuit element that stores and releases electromagnetic energy as opposed to a *resistance*, which dissipates energy. If the stored energy is predominantly in the electric field, the reactance is said to be capacitive; if the stored energy is predominantly in the magnetic field, the reactance is said to be inductive. In microwave engineering it is more convenient to speak in terms of voltages and currents rather than electric and magnetic fields. A capacitive reactance may then be a circuit element for which capacitance is the ratio of charge on the capacitor over voltage across the capacitor. Then

$$C = \frac{Q}{V} \tag{8-5-1}$$

If the ratio is not linear, the capacitive reactance is said to be nonlinear. In this case, it is convenient to define a nonlinear capacitance as the partial derivative of charge with respect to voltage. That is,

$$C(v) = \frac{\partial Q}{\partial v} \tag{8-5-2}$$

The analogous definition of a nonlinear inductance is

$$L(i) = \frac{\partial \Phi}{\partial i} \tag{8-5-3}$$

In the operation of parametric devices, the mixing effects occur when voltages at two or more different frequencies are impressed on a nonlinear reactance.

Small-signal method. It is assumed that the signal voltage v_s is much smaller than the pumping voltage v_p, and the total voltage across the nonlinear capacitance $C(t)$ is given by

$$v = v_s + v_p = V_s \cos(\omega_s t) + V_p \cos(\omega_p t) \tag{8-5-4}$$

where $V_s \ll V_p$. The charge on the capacitor can be expanded in a Taylor series about the point $v_s = 0$, and the first two terms are

$$Q(v) = Q(v_1 + v_p) = Q(v_p) + \frac{dQ(v_p)}{dv}\bigg|_{v_s=0} v_s \tag{8-5-5}$$

For convenience, it is assumed that

$$C(v_p) = \frac{dQ(v_p)}{dv} = C(t) \tag{8-5-6}$$

where $C(v_p)$ is periodic with a fundamental frequency of ω_p. If the capacitance $C(v_p)$ is expanded in a Fourier series, the result is

$$C(v_p) = \sum_{n=0}^{\infty} C_n \cos(n\omega_p t) \tag{8-5-7}$$

Since v_p is a function of time, the capacitance $C(v_p)$ is also a function of time. Then

$$C(t) = \sum_{n=0}^{\infty} C_n \cos(n\omega_p t) \tag{8-5-8}$$

The coefficients C_n are the magnitude of each harmonic of the time-varying capacitance. In general, the coefficients C_n are not linear functions of the ac pumping voltage v_p. Since the junction capacitance $C(t)$ of a parametric diode is a nonlinear capacitance, the principle of superposition does not hold for arbitrary ac signal amplitudes.

The current through the capacitance $C(t)$ is the derivative of Eq. (8-5-5) with respect to time and it is

$$i = \frac{dQ}{dt} = \frac{dQ(v_p)}{dt} + \frac{d}{dt}[C(t)v_s] \tag{8-5-9}$$

It is evident that the nonlinear capacitance behaves like a time-varying linear capacitance for signals with amplitudes that are much smaller than the amplitude of the pumping voltage. The first term of Eq. (8-5-9) yields a current at the pump frequency f_p and is not related to the signal frequency f_s.

Large-signal method. If the signal voltage is not small compared with the pumping voltage, the Taylor series can be expanded about a dc bias voltage V_0 in a junction diode. In a junction diode the capacitance C is proportional to $(\phi_0 - V)^{-1/2} = V_0^{-1/2}$, where ϕ_0 is the junction barrier potential and V is a negative

voltage supply. Since

$$[V_0 + V_p \cos{(\omega_p t)}]^{-1/2} \approx V_0^{-1/2}\left(1 - \frac{V_p}{3V_0}\cos{(\omega_p t)}\right) \qquad \text{for } V_p \ll V_0$$

the capacitance $C(t)$ can be expressed as

$$C(t) = C_0[1 + 2\gamma \cos{(\omega_p t)}] \qquad\qquad (8\text{-}5\text{-}10)$$

The parameter γ is proportional to the pumping voltage v_p and indicates the coupling effect between the voltages at the signal frequency f_s and the output frequency f_0.

Manley–Rowe power relations.   Manley and Rowe [20] derived a set of general energy relations regarding power flowing into and out of an ideal nonlinear reactance. These relations are useful in predicting whether power gain is possible in a parametric amplifier. Figure 8-5-1 shows an equivalent circuit for Manley–Rowe derivation.
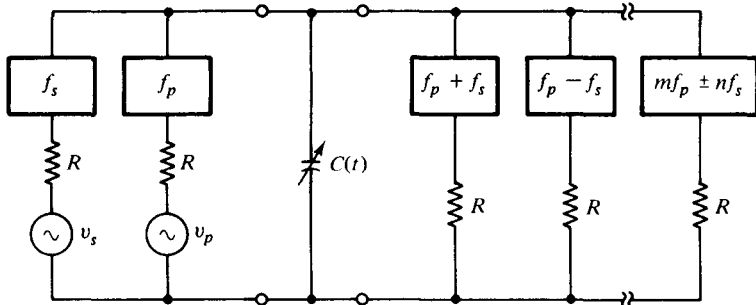
Figure 8-5-1   Equivalent circuit for Manley–Rowe derivation.

In Fig. 8-5-1, one signal generator and one pump generator at their respective frequencies f_s and f_p, together with associated series resistances and bandpass filters, are applied to a nonlinear capacitance $C(t)$. These resonating circuits of filters are designed to reject power at all frequencies other than their respective signal frequencies. In the presence of two applied frequencies f_s and f_p, an infinite number of resonant frequencies of $mf_p \pm nf_p$ are generated, where m and n are any integers from zero to infinity.

Each of the resonating circuits is assumed to be ideal. The power loss by the nonlinear susceptances is negligible. That is, the power entering the nonlinear capacitor at the pump frequency is equal to the power leaving the capacitor at the other frequencies through the nonlinear interaction. Manley and Rowe established the power relations between the input power at the frequencies f_s and f_p and the output power at the other frequencies $mf_p \pm nf_s$.

From Eq. (8-5-4) the voltage across the nonlinear capacitor $C(t)$ can be expressed in exponential form as

$$v = v_p + v_s = \frac{V_p}{2}(e^{j\omega_p t} + e^{-j\omega_p t}) + \frac{V_s}{2}(e^{j\omega_s t} + e^{-j\omega_s t}) \qquad (8\text{-}5\text{-}11)$$

The general expression of the charge Q deposited on the capacitor is given by

$$Q = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} Q_{m,n} e^{j(m\omega_p t + n\omega_s t)} \tag{8-5-12}$$

In order for the charge Q to be real, it is necessary that

$$Q_{m,n} = Q^*_{-m,-n} \tag{8-5-13}$$

The total voltage v can be expressed as a function of the charge Q. A similar Taylor series expansion of $v(Q)$ shows that

$$v = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} V_{m,n} e^{j(m\omega_p t + n\omega_s t)} \tag{8-5-14}$$

In order for the voltage v to be real, it is required that

$$V_{m,n} = V^*_{-m,-n} \tag{8-5-15}$$

The current flowing through $C(t)$ is the total derivative of Eq. (8-5-12) with respect to time. This is

$$i = \frac{dQ}{dt} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} j(m\omega_p + n\omega_s) Q_{m,n} e^{j(m\omega_p t + n\omega_s t)}$$

$$= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} I_{m,n} e^{j(m\omega_p t + n\omega_s t)} \tag{8-5-16}$$

where $I_{m,n} = j(m\omega_p + n\omega_s)Q_{m,n}$ and $I_{m,n} = I^*_{-m,-n}$. Since the capacitance $C(t)$ is assumed to be a pure reactance, the average power at the frequencies $mf_p + nf_s$ is

$$P_{m,n} = (V_{m,n} I^*_{m,n} + V^*_{m,n} I_{m,n})$$

$$= (V^*_{-m,-n} I_{-m,-n} + V_{-m,-n} I^*_{-m,-n}) = P_{-m,-n} \tag{8-5-17}$$

Then conservation of power can be written

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} P_{m,n} = 0 \tag{8-5-18}$$

Multiplication of Eq. (8-5-18) by a factor of $(m\omega_p + n\omega_s)/(m\omega_p + n\omega_s)$ and rearrangement of the resultant into two parts yield

$$\omega_p \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{m\omega_p + n\omega_s} + \omega_s \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{nP_{m,n}}{m\omega_p + n\omega_s} = 0 \tag{8-5-19}$$

Since

$$I_{m,n}/(m\omega_p + n\omega_s) = jQ_{m,n}$$

then $P_{m,n}/(m\omega_p + n\omega_s)$ becomes $-jV_{m,n}Q^*_{m,n} - jV_{-m,-n}Q^*_{-m,-n}$ and is independent of ω_p or ω_s. For any choice of the frequencies f_p and f_s, the resonating circuit external to that of the nonlinear capacitance $C(t)$ can be so adjusted that the currents may keep all the voltage amplitudes $V_{m,n}$ unchanged. The charges $Q_{m,n}$ are then also un-

changed, since they are functions of the voltages $V_{m,n}$. Consequently, the frequencies f_p and f_s can be arbitrarily adjusted in order to require

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{m\omega_p + n\omega_s} = 0 \tag{8-5-20}$$

$$\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{nP_{m,n}}{m\omega_p + n\omega_s} = 0 \tag{8-5-21}$$

Equation (8-5-20) can be expressed as two terms:

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{m\omega_p + n\omega_s} + \sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{-mP_{m,n}}{-m\omega_p - n\omega_s} = 0 \tag{8-5-22}$$

Since $P_{m,n} = P_{-m,-n}$, then

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{mP_{m,n}}{mf_p + nf_s} = 0 \tag{8-5-23}$$

Similarly,

$$\sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{nP_{m,n}}{mf_p + nf_s} = 0 \tag{8-5-24}$$

where ω_p and ω_s have been replaced by f_p and f_s, respectively.

Equations (8-5-23) and (8-5-24) are the standard forms for the Manley–Rowe power relations. The term $P_{m,n}$ indicates the real power flowing into or leaving the nonlinear capacitor at a frequency of $mf_p + nf_s$. The frequency f_p represents the fundamental frequency of the pumping voltage oscillator and the frequency f_s signifies the fundamental frequency of the signal voltage generator. The sign convention for the power term $P_{m,n}$ will follow that power flowing into the nonlinear capacitance or the power coming from the two voltage generators is positive, whereas the power leaving from the nonlinear capacitance or the power flowing into the load resistance is negative.

Consider, for instance, the case where the power output flow is allowed at a frequency of $f_p + f_s$ as shown in Fig. 8-5-1. All other harmonics are open-circuited. Then currents at the three frequencies f_p, f_s, and $f_p + f_s$ are the only ones existing. Under these restrictions m and n vary from -1 through zero to $+1$, respectively. Then Eqs. (8-5-23) and (8-5-24) reduce to

$$\frac{P_{1,0}}{f_p} + \frac{P_{1,1}}{f_p + f_s} = 0 \tag{8-5-25}$$

$$\frac{P_{0,1}}{f_s} + \frac{P_{1,1}}{f_p + f_s} = 0 \tag{8-5-26}$$

where $P_{1,0}$ and $P_{0,1}$ are the power supplied by the two voltage generators at the frequencies f_p and f_s, respectively, and they are considered positive. The power $P_{1,1}$ flowing from the reactance into the resistive load at a frequency of $f_p + f_s$ is considered negative.

The power gain, which is defined as a ratio of the power delivered by the capacitor at a frequency of $f_p + f_s$ to that absorbed by the capacitor at a frequency of f_s as shown in Eq. (8-5-26) is given by

$$\text{Gain} = \frac{f_p + f_s}{f_s} = \frac{f_0}{f_s} \qquad \text{(for modulator)} \qquad (8\text{-}5\text{-}27)$$

where $f_p + f_s = f_o$ and $(f_p + f_s) > f_p > f_s$. The maximum power gain is simply the ratio of the output frequency to the input frequency. This type of parametric device is called the *sum-frequency parametric amplifier* or *up-converter*.

If the signal frequency is the sum of the pump frequency and the output frequency, Eq. (8-5-26) predicts that the parametric device will have a gain of

$$\text{Gain} = \frac{f_s}{f_p + f_s} \qquad \text{(for demodulator)} \qquad (8\text{-}5\text{-}28)$$

where $f_s = f_p + f_o$ and $f_o = f_s - f_p$. This type of parametric device is called the *parametric down-converter* and its power gain is actually a loss.

If the signal frequency is at f_s, the pump frequency at f_p, and the output frequency at f_o, where $f_p = f_s + f_o$, the power $P_{1,1}$ supplied at f_p is positive. Both $P_{1,0}$ and $P_{0,1}$ are negative. In other words, the capacitor delivers power to the signal generator at f_s instead of absorbing it. The power gain may be infinite, which is an unstable condition, and the circuit may be oscillating both at f_s and f_o. This is another type of parametric device, often called a *negative-resistance parametric amplifier*.

8-5-3 Parametric Amplifiers

In a superheterodyne receiver a radio frequency signal may be mixed with a signal from the local oscillator in a nonlinear circuit (the mixer) to generate the sum and difference frequencies. In a parametric amplifier the local oscillator is replaced by a pumping generator such as a reflex klystron and the nonlinear element by a time-varying capacitor such as a varactor diode (or inductor) as shown in Fig. 8-5-2.

In Fig. 8-5-2, the signal frequency f_s and the pump frequency f_p are mixed in the nonlinear capacitor C. Accordingly, a voltage of the fundamental frequencies f_s and f_p as well as the sum and the difference frequencies $mf_p \pm nf_s$ appears across C. If a resistive load is connected across the terminals of the idler circuit, an output voltage can be generated across the load at the output frequency f_o. The output circuit, which does not require external excitation, is called the *idler circuit*. The output (or idler) frequency f_o in the idler circuit is expressed as the sum and the difference frequencies of the signal frequency f_s and the pump frequency f_p. That is,

$$f_o = mf_p \pm nf_s \qquad (8\text{-}5\text{-}29)$$

where m and n are positive integers from zero to infinity.

If $f_o > f_s$, the device is called a parametric *up-converter*. Conversely, if $f_0 < f_s$, the device is known as a parametric *down-converter*.
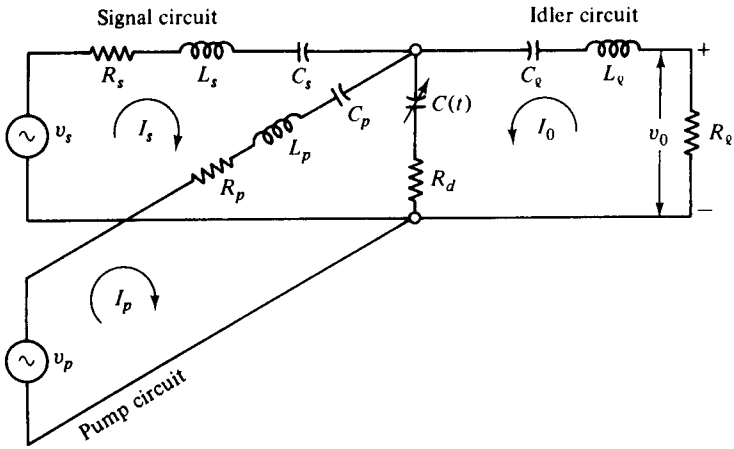
Figure 8-5-2 Equivalent circuit for a parametric amplifier.

Parametric up-converter. A parametric up-converter has the following properties:

1. The ouptut frequency is equal to the sum of the signal frequency and the pump frequency.
2. There is no power flow in the parametric device at frequencies other than the signal, pump, and output frequencies.

Power Gain.   When these two conditions are satisfied, the maximum power gain of a parametric up-converter [21] is expressed as

$$\text{Gain} = \frac{f_0}{f_s} \frac{x}{(1 + \sqrt{1 + x})^2} \tag{8-5-30}$$

where $f_0 = f_p + f_s$

$$x = \frac{f_s}{f_0}(\gamma Q)^2$$

$$Q = \frac{1}{2\pi f_s C R_d}$$

Moreover, R_d is the series resistance of a p-n junction diode and γQ is the figure of merit for the nonlinear capacitor. The quantity of $x/(1 + \sqrt{1 + x})^2$ may be regarded as a gain-degradation factor. As R_d approaches zero, the figure of merit γQ goes to infinity and the gain-degradation factor becomes equal to unity. As a result, the power gain of a parametric up-converter for a lossless diode is equal to f_0/f_s, which is predicted by the Manley–Rowe relations as shown in Eq. (8-5-27). In a typical microwave diode γQ could be equal to 10. If $f_0/f_s = 15$, the maximum gain given by Eq. (8-5-30) is 7.3 dB.

Noise Figure. One advantage of the parametric amplifier over the transistor amplifier is its low-noise figure because a pure reactance does not contribute thermal noise to the circuit. The noise figure F for a parametric up-converter [21] is given by

$$F = 1 + \frac{2T_d}{T_0}\left[\frac{1}{\gamma Q} + \frac{1}{(\gamma Q)^2}\right] \tag{8-5-31}$$

where T_d = diode temperature in degrees Kelvin

$T_0 = 300°K$ is the ambient temperature in degrees Kelvin

γQ = figure of merit for the nonlinear capacitor

In a typical microwave diode γQ could be equal to 10. If $f_o/f_s = 10$ and $T_d = 300°K$, the minimum noise figure is 0.90 dB, as calculated by using Eq. (8-5-31).

Bandwidth. The bandwidth of a parametric up-converter is related to the gain-degradation factor of the merit figure and the ratio of the signal frequency to the output frequency. The bandwidth equation [21] is given by

$$BW = 2\gamma\sqrt{\frac{f_o}{f_s}} \tag{8-5-32}$$

If $f_o/f_s = 10$ and $\gamma = 0.2$, the bandwidth (BW) is equal to 1.264.

Example 8-5-1:   Up-Converter Parametric Amplifier

An up-converter parametric amplifier has the following parameters:

Ratio of output frequency over signal frequency: $f_o/f_s = 25$
Figure of merit: $\gamma Q = 10$
Factor of merit figure: $\gamma = 0.4$
Diode temperature: $T_d = 350°K$

Calculate: **(a)** the power gain in decibels; **(b)** the noise figure in decibels; **(c)** the bandwidth.

Solution

a. From Eq. (8-5-30) the up-converter power gain is

$$\text{Power gain} = \frac{f_o}{f_s}\frac{x}{(1 + \sqrt{1 + x})^2} = 25 \times \frac{100/25}{(1 + \sqrt{1 + 100/25})^2}$$
$$= 9.55 = 9.80 \text{ dB}$$

b. From Eq. (8-5-31) the noise figure is

$$F = 1 + \frac{2T_d}{T_o}\left[\frac{1}{\gamma Q} + \frac{1}{(\gamma Q)^2}\right] = 1 + \frac{2 \times 350}{300}\left[\frac{1}{10} + \frac{1}{100}\right] = 1.26 = 1 \text{ dB}$$

c. From Eq. (8-5-32) the bandwidth is

$$BW = 2\gamma\sqrt{\frac{f_o}{f_s}} = 2 \times 0.4 \times (25)^{1/2} = 4$$

Parametric down-converter.   If a mode of down conversion for a parametric amplifier is desirable, the signal frequency f_s must be equal to the sum of the pump frequency f_p and the output frequency f_o. This means that the input power must feed into the idler circuit and the output power must move out from the signal circuit as shown in Fig. 8-5-2. The down-conversion gain (actually a loss) is given by [21]

$$\text{Gain} = \frac{f_s}{f_o} \frac{x}{(1 + \sqrt{1 + x})^2} \tag{8-5-33}$$

Negative-resistance parametric amplifier.   If a significant portion of power flows only at the signal frequency f_s, the pump frequency f_p, and the idler frequency f_i, a regenerative condition with the possibility of oscillation at both the signal frequency and the idler frequency will occur. The idler frequency is defined as the difference between the pump frequency and the signal frequency, $f_i = f_p - f_s$. When the mode operates below the oscillation threshold, the device behaves as a bilateral negative-resistance parametric amplifier.

Power Gain.   The output power is taken from the resistance R_i at a frequency f_i, and the conversion gain from f_s to f_i [21] is given by

$$\text{Gain} = \frac{4f_i}{f_s} \cdot \frac{R_g R_i}{R_{Ts} R_{Ti}} \cdot \frac{a}{(1 - a)^2} \tag{8-5-34}$$

where f_s = signal frequency
f_p = pump frequency
$f_i = f_p - f_s$ is the idler frequency
R_g = output resistance of the signal generator
R_i = output resistance of the idler generator
R_{Ts} = total series resistance at f_s
R_{Ti} = total series resistance at f_i
$a = R/R_{Ts}$
$R = \gamma^2/(\omega_s \omega_i C^2 R_{Ti})$ is the equivalent negative resistance

Noise Figure.   The optimum noise figure of a negative-resistance parametric amplifier [21] is expressed as

$$F = 1 + 2\frac{T_d}{T_0}\left[\frac{1}{\gamma Q} + \frac{1}{(\gamma Q)^2}\right] \tag{8-5-35}$$

where γQ = figure of merit for the nonlinear capacitor
T_0 = 300°K is the ambient temperature in degrees Kelvin
T_d = diode temperature in degrees Kelvin

It is interesting to note that the noise figure given by Eq. (8-5-35) is identical to that for the parametric up-converter in Eq. (8-5-31).

Bandwidth. The maximum gain bandwidth of a negative-resistance parametric amplifier [21] is given by

$$BW = \frac{\gamma}{2} \sqrt{\frac{f_i}{f_s \text{ gain}}} \tag{8-5-36}$$

If gain = 20 dB, $f_i = 4f_s$, and $\gamma = 0.30$, the maximum possible bandwidth for single-tuned circuits is about 0.03.

Degenerate parametric amplifier. The degenerate parametric amplifer or oscillator is defined as a negative-resistance amplifier with the signal frequency equal to the idler frequency. Since the idler frequency f_i is the difference between the pump frequency f_p and the signal frequency f_s, the signal frequency is just one-half the pump frequency.

Power Gain and Bandwidth. The power gain and bandwidth characteristics of a degenerate parametric amplifier are exactly the same as for the parametric up-converter. With $f_s = f_i$ and $f_p = 2f_s$, the power transferred from pump to signal frequency is equal to the power transferred from pump to idler frequency. At high gain the total power at the signal frequency is almost equal to the total power at the idler frequency. Hence the total power in the passband will have 3 dB more gain.

Noise Figure. The noise figures for a single-sideband and a double-sideband degenerate parametric amplifier [21] are given by, respectively,

$$F_{ssb} = 2 + \frac{2\bar{T}_d R_d}{T_0 R_g} \tag{8-5-37}$$

$$F_{dsb} = 1 + \frac{\bar{T}_d R_d}{T_0 R_g} \tag{8-5-38}$$

where $\bar{T}_d$ = average diode temperature in degrees Kelvin
T_0 = 300°K is the ambient noise temperature in degrees Kelvin
R_d = diode series resistance in ohms
R_g = external output resistance of the signal generator in ohms

It can be seen that the noise figure for double-sideband operation is 3 dB less than that for single-sideband operation.

8-5-4 Applications

The choice of which type of parametric amplifier to use depends on the microwave system requirements. The up-converter is a unilateral stable device with a wide bandwidth and low gain. The negative-resistance amplifier is inherently a bilateral and unstable device with narrow bandwidth and high gain. The degenerate parametric amplifier does not require a separate signal and idler circuit coupled by the diode and is the least complex type of parametric amplifier.

In general, the up-converter has the following advantages over the negative-resistance parametric amplifier:

1. A positive input impedance
2. Unconditionally stable and unilateral
3. Power gain independent of changes in its source impedance
4. No circulator required
5. A typical bandwidth on the order of 5%

At higher frequencies where the up-converter is no longer practical, the negative-resistance parametric amplifier operated with a circulator becomes the proper choice. When a low noise figure is required by a system, the degenerate parametric amplifier may be the logical choice, since its double-sideband noise figure is less than the optimum noise figure of the up-converter or the nondegenerate negative-resistance parametric amplifier. Furthermore, the degenerate amplifier is a much simpler device to build and uses a relatively low pump frequency. In radar systems the negative-resistance parametric amplifier may be the better choice, since the frequency required by the system may be higher than the X band. However, since the parametric amplifier is complicated in fabrication and expensive in production, there is a tendency in microwave engineering to replace it with the GaAs metal-semiconductor field-effect transistor (MESFET) amplifier in airborne radar systems.

REFERENCES

[1] READ, W. T., A proposed high-frequency negative-resistance diode. *Bell System Tech. J.,* **37,** 401–446 (1958)

[2] COLEMAN, D. J., JR., and S. M. SZE, A low-noise metal-semiconductor-metal (MSM) microwave oscillator. *Bell System Tech. J.,* **50,** 1695–1699, May–June 1971.

[3] LEE, C. A., ET AL., The Read diode, an avalanche, transit-time, negative-resistance oscillator. *Appl. Phys. Letters,* **6,** 89 (1965).

[4] JOHNSTON, R. L., B. C. DeLOACH, and G. B. COHEN, A silicon diode microwave oscillator. *Bell System Tech. J.,* **44,** 369–372, February 1965.

[5] GILDEN, M., and M. E. HINES, Electronic tuning effects in the Read microwave avalanche diode. *IEEE Trans. on Electron Devices,* **ED-13,** 5–12, January 1966.

[6] HIESLMAIR, H., ET AL., State of the art of solid-state and tube transmitters. *Microwave J.,* **26,** No. 10, 46–48, October 1983.

[7] PRAGER, H. J., ET AL., High-power, high-efficiency silicon avalanche diodes at ultra high frequencies. *Proc. IEEE (Letters),* **55,** 586–587, April 1967.

[8] CLORFEINE, A. S., ET AL., A theory for the high-efficiency mode of oscillation in avalanche diodes. *RCA Rev.,* **30,** 397–421, September 1969.

[9] DeLOACH, B. C., JR., and D. L. SCHARFETTER, Device physics of TRAPATT oscillators. *IEEE Trans. on Electron Devices,* **ED-17,** 9–21, January 1970.

[10] LIU, S. G., and J. J. RISKA, Fabrication and performance of kilowatt L-band avalanche diodes. *RCA Rev.*, **31**, 3, March 1970.

[11] KOSTICHACK, D. F., UHF avalanche diode oscillator providing 400 watts peak power and 75 percent efficiency. *Proc. IEEE (Letters)*, **58**, 1282–1283, August 1970.

[12] WILSON, W. E., Pulsed LSA and TRAPATT sources for microwave systems. *Microwave J.*, **14**, No. 8, 33–41, August 1971.

[13] COLEMAN, D. J., JR., and S. M. SZE, A low-noise metal-semiconductor-metal (MSM) microwave oscillator. *Bell System Tech. J.*, 1675–1695, May–June 1971.

[14] FARADAY, M., On a peculiar class of acoustical figures; and certain forms assumed by a group of particles upon vibrating elastic surface. *Phil. Trans. Roy. Soc. (London)*, **121**, 299–318, May 1831.

[15] LORD RAYLEIGH, and J. W. STRUTT, On the crispations of fluid resting upon a vibrating support. *Phil. Mag.*, **16**, 50–53, July 1883.

[16] VAN DER ZIEL, A., On the mixing properties of nonlinear capacitances. *J. Appl. Phys.*, **19**, 999–1006, November 1948.

[17] LANDON, V. D., The use of ferrite cored coils as converters, amplifiers, and oscillators. *RCA Rev.*, **10**, 387–396, September 1949.

[18] SUHL, H., Proposal for a ferromagnetic amplifier in the microwave range. *Phys. Rev.*, **106**, 384–385, April 15, 1957.

[19] WEISS, M. T., A solid-state microwave amplifier and oscillator using ferrites. *Phys Rev.*, **107**, 317, July 1957.

[20] MANLEY, J. M., and H. E. ROWE, Some general properties of nonlinear elements—Pt. I, General energy relations. *Proc. IRE*, **44**, 904–913, July 1956.

[21] BLACKWELL, L. A., and K. L. KOTZEBUE, *Semiconductor-Diode Parametric Amplifiers*, 41, 42, 45, 53, 57, 62, 70. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1961.

SUGGESTED READINGS

CHANG, K. K. N., *Parametric and Tunnel Diodes*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.

DELOACH, B. C., and D. L. SCHARFETTER, Device physics of TRAPATT oscillators. *IEEE Trans. on Electron Devices*, **ED-17**, No. 1, 9–21, January 1970.

EASTMAN, L. F., *Gallium Arsenide Microwave Bulk and Transit-Time Devices*. Artech House, Dedham, Mass., 1973.

HADDAD, G. I., ed., *Avalanche Transit-Time Devices*. Artech House, Dedham, Mass., 1973.

HADDAD, G. I., ET AL., Basic principles and properties of avalanche transit-time devices. *IEEE Trans. on Microwave Theory and Techniques*, **MTT-18,** No. 11, 752–772, November 1970.

MILNES, A. G., *Semiconductor Devices and Integrated Electronics*. Van Nostrand Reinhold Company, New York, 1980.

PARKER, D., TRAPATT oscillations in a *p-i-n* avalanche diode. *IEEE Trans. on Electron Devices*, **ED-18,** No. 5, 281–293, May 1971.

SZE, S. M., Microwave avalanche diodes. *IEEE Proc.*, **59,** No.8, 1140–1171, August 1971.

PROBLEMS

Avalanche Transit-Time Devices

8-1. Spell out the following abbreviated terms: IMPATT, TRAPATT, and BARITT.

8-2. Derive Eq. (8-2-9).

8-3. Describe the operating principles of the Read diode, IMPATT diode, TRAPATT diode, and BARITT diode.

8-4. An IMPATT diode has a drift length of 2 μm. Determine: (**a**) the drift time of the carriers and (**b**) the operating frequency of the IMPATT diode.

8-5. A Ku-band IMPATT diode has a pulsed-operating voltage of 100 V and a pulsed-operating current of 0.9 A. The efficiency is about 10%. Calculate:
a. The output power
b. The duty cycle if the pulse width is 0.01 ns and the frequency is 16 GHz

8-6. An M-Si-M BARITT diode has the following parameters:

Relative dielectric constant of Si: $\epsilon_r = 11.8$
Donor concentration: $N = 3 \times 10^{21} \text{ m}^{-3}$
Si length: $L = 6.2 \ \mu\text{m}$

Calculate: (**a**) the breakdown voltage and (**b**) the breakdown electric field.

Parametric Devices

8-7. **a.** Describe the advantages and disadvantages of the parametric devices.
b. Describe the applications of the parametric amplifiers.

8-8. The figure of merit for a diode nonlinear capacitor in an up-converter parametric amplifier is 8, and the ratio of the output (or idler) frequency f_o over the signal frequency f_s is 8. The diode temperature is 300°K.
a. Calculate the maximum power gain in decibels.
b. Compute the noise figure F in decibels.
c. Determine the bandwidth (BW) for $\gamma = 0.2$.

8-9. A negative-resistance parametric amplifier has the following parameters:

$$f_s = 2 \text{ GHz} \qquad R_i = 1 \text{ k}\Omega \qquad \gamma = 0.35$$

$$f_p = 12 \text{ GHz} \qquad R_g = 1 \text{ K}\Omega \qquad \gamma Q = 10$$

$$f_i = 10 \text{ GHz} \qquad R_{Ts} = 1 \text{ K}\Omega \qquad T_d = 300°\text{K}$$

$$f_i = 5 f_s \qquad R_{Ti} = 1 \text{ K}\Omega \qquad C = 0.01 \text{ pF}$$

a. Calculate the power gain in decibels.
b. Compute the noise figure F in decibels.
c. Determine the bandwidth (BW).

Chapter 9

Microwave Linear-Beam Tubes (O Type)

9-0 INTRODUCTION

We turn now to a quantitative and qualitative analysis of several conventional vacuum tubes and microwave tubes in common use. The conventional vacuum tubes, such as triodes, tetrodes, and pentodes, are still used as signal sources of low output power at low microwave frequencies. The most important microwave tubes at present are the linear-beam tubes (O type) tabulated in Table 9-0-1. The paramount O-type tube is the two-cavity klystron, and it is followed by the reflex klystron. The helix traveling-wave tube (TWT), the coupled-cavity TWT, the forward-wave amplifier (FWA), and the backward-wave amplifier and oscillator (BWA and BWO) are also O-type tubes, but they have nonresonant periodic structures for electron interactions. The Twystron is a hybrid amplifier that uses combinations of klystron and TWT components. The switching tubes such as krytron, thyratron, and planar triode are very useful in laser modulation. Although it is impossible to discuss all such tubes in detail, the common operating principles of many will be described.

   The advent of linear-beam tubes began with the Heil oscillators [1] in 1935 and the Varian brothers' klystron amplifier [1] in 1939. The work was advanced by the space-charge-wave propagation theory of Hahn and Ramo [1] in 1939 and continued with the invention of the helix-type traveling-wave tube (TWT) by R. Kompfner in 1944 [2]. From the early 1950s on, the low power output of linear-beam tubes made it possible to achieve high power levels, first rivaling and finally surpassing magnetrons, the early sources of microwave high power. Subsequently, several additional devices were developed, two of which have shown lasting importance. They are the extended interaction klystron [3] and the Twystron hybrid amplifier [4].

TABLE 9-0-1 LINEAR BEAM TUBES (*O* TYPE)
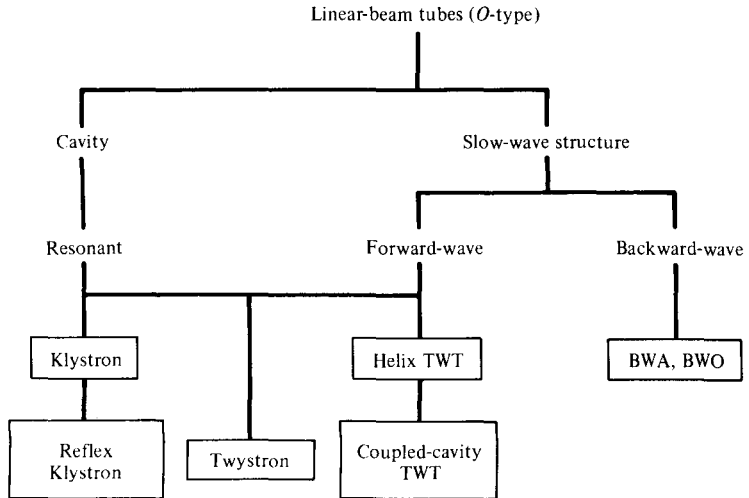
In a linear-beam tube a magnetic field whose axis coincides with that of the electron beam is used to hold the beam together as it travels the length of the tube. O-type tubes derive their name from the French *TPO* (*tubes à propagation des ondes*) or from the word *original* (meaning the original type of tube). In these tubes electrons receive potential energy from the dc beam voltage before they arrive in the microwave interaction region, and this energy is converted into their kinetic energy. In the microwave interaction region the electrons are either accelerated or decelerated by the microwave field and then bunched as they drift down the tube. The bunched electrons, in turn, induce current in the output structure. The electrons then give up their kinetic energy to the microwave fields and are collected by the collector.

O-type traveling-wave tubes are suitable for amplification. At present, klystron and TWT amplifiers can deliver a peak power output up to 30 MW (megawatts) with a beam voltage on the order of 100 kV at the frequency of 10 GHz. The average power outputs are up to 700 kW. The gain of these tubes is on the order of 30 to 70 dB, and the efficiency is from 15 to 60%. The bandwidth is from 1 to 8% for klystrons and 10 to 15% for TWTs.

Since the early 1960s, predictions have continued that microwave tubes will be displaced by microwave solid-state devices. This displacement has occurred only at the low-power and receiving circuit level of equipment. Microwave power tubes continue, however, as the only choice for high-power transmitter outputs and are expected to maintain this dominant role throughout the next generation and beyond. Figure 9-0-1 and 9-0-2 show the CW power and peak power state-of-the-art performances for various tube types [5].

The numbers by the data points represent efficiency in percent; the letter in parentheses stands for the developer of the tube. Impressive gain has been made in bandwidth of a single TWT device by Varian. More than 50 dB of gain is available
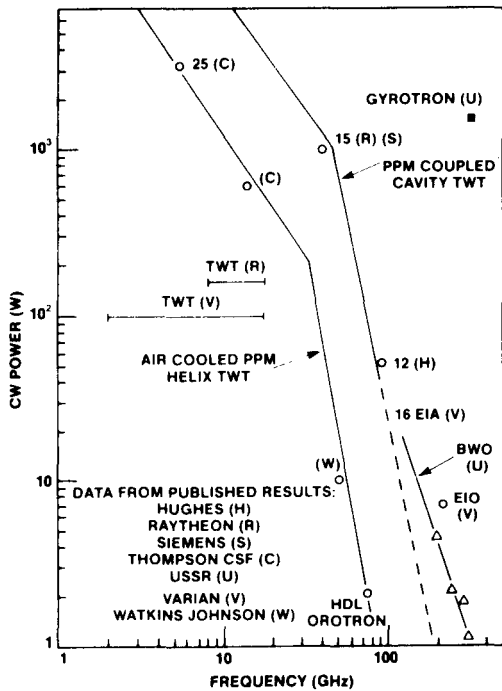
Figure 9-0-1 CW power state-of-the-art performance for various microwave tubes. (*After C. Hieslmair [5], reprinted by permission of* Microwave Journal.)
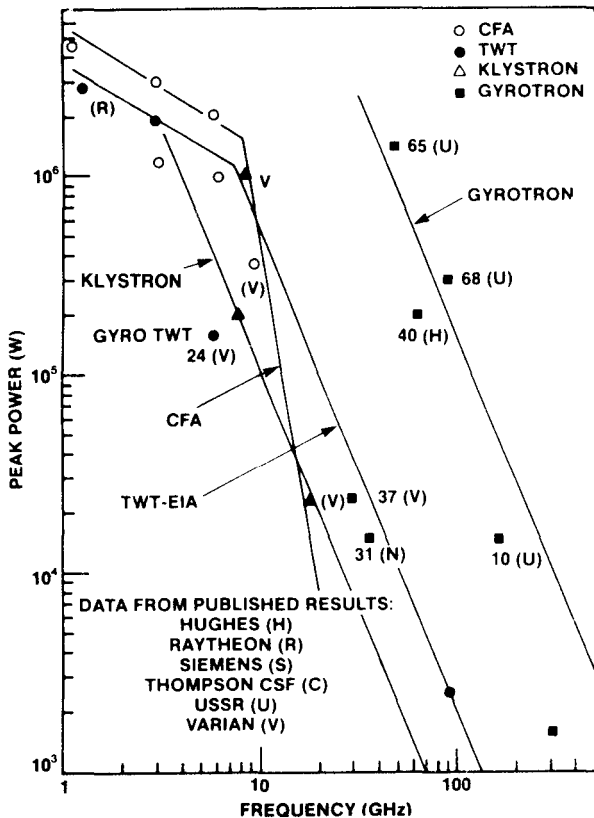
Figure 9-0-2 Peak power state-of-the-art performance for various microwave tubes. (*After C. Hieslmair [5], reprinted by permission of* Microwave Journal.)

in a 93- to 95-GHz TWT at the 50-watt average level by Hughes, and 2-kW peak power at 30-dB gain is provided by Varian. The most impressive power achievements at very good efficiencies continue to occur in the gyrotron area. The Naval Research Laboratory (NRL) reports results of work on a 30% bandwidth gyrotron TWT. Tube reliability continues to improve as a result of low-temperature cathode technology, better materials, and quality control in manufacture.

9-1 CONVENTIONAL VACUUM TRIODES, TETRODES, AND PENTODES

Conventional vacuum triodes, tetrodes, and pentodes are less useful signal sources at frequencies above 1 GHz because of lead-inductance and interelectrode-capacitance effects, transit-angle effects, and gain-bandwidth product limitations. These three effects are analyzed in detail in the following sections.

9-1-1 Lead-Inductance and Interelectrode-Capacitance Effects

At frequencies above 1 GHz conventional vacuum tubes are impaired by parasitic-circuit reactances because the circuit capacitances between tube electrodes and the circuit inductance of the lead wire are too large for a microwave resonant circuit. Furthermore, as the frequency is increased up to the microwave range, the real part of the input admittance may be large enough to cause a serious overload of the input circuit and thereby reduce the operating efficiency of the tube. In order to gain a better understanding of these effects, the triode ciruit shown in Fig. 9-1-1 should be studied carefully.
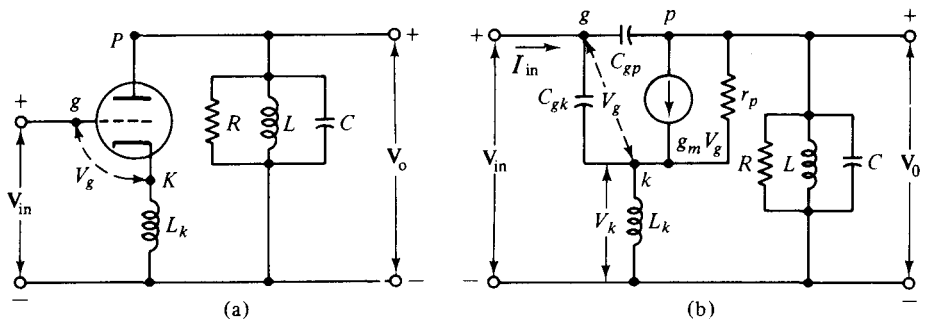
Figure 9-1-1   Triode circuit (a) and its equivalent (b).

Figure 9-1-1(b) shows the equivalent circuit of a triode circuit under the assumption that the interelectrode capacitances and cathode inductance are the only parasitic elements. Since $C_{gp} \ll C_{gk}$ and $\omega L_k \ll 1/(\omega C_{gk})$, the input voltage $\mathbf{V}_{in}$ can be written as

$$\mathbf{V}_{in} = V_g + V_k = V_g + j\omega L_k g_m V_g \qquad (9\text{-}1\text{-}1)$$

and the input current as

$$\mathbf{I}_{in} = j\omega C_{gk} V_g \qquad (9\text{-}1\text{-}2)$$

Substitution of Eq. (9-1-2) in Eq. (9-1-1) yields

$$\mathbf{V}_{in} = \frac{\mathbf{I}_{in}(1 + j\omega L_k g_m)}{j\omega C_{gk}} \qquad (9\text{-}1\text{-}3)$$

The input admittance of the tube is approximately

$$\mathbf{Y}_{in} = \frac{\mathbf{I}_{in}}{\mathbf{V}_{in}} = \frac{j\omega C_{gk}}{1 + j\omega L_k g_m} = \omega^2 L_k C_{gk} g_m + j\omega C_{gk} \qquad (9\text{-}1\text{-}4)$$

in which $\omega L_k g_m \ll 1$ has been replaced. The inequality is almost always true, since the cathode lead is usually short and is quite large in diameter, and the transconductance g_m is generally much less than one millimho.

The input impedance at very high frequencies is given by

$$\mathbf{Z}_{in} = \frac{1}{\omega^2 L_k C_{gk} g_m} - j\frac{1}{\omega^3 L_k^2 C_{gk} g_m^2} \qquad (9\text{-}1\text{-}5)$$

The real part of the impedance is inversely proportional to the square of the frequency, and the imaginary part is inversely proportional to the third order of the frequency. When the frequencies are above 1 GHz, the real part of the impedance becomes small enough to nearly short the signal source. Consequently, the output power is decreased rapidly. Similarly, the input admittance of a pentode circuit is expressed by

$$\mathbf{Y}_{in} = \omega^2 L_k C_{gk} g_m + j\omega (C_{gk} + C_{gs}) \qquad (9\text{-}1\text{-}6)$$

where C_{gs} is the capacitance between the grid and screen, and its input impedance is given by

$$\mathbf{Z}_{in} = \frac{1}{\omega^2 L_k C_{gk} g_m} - j\frac{C_{gk} + C_{gs}}{\omega^3 L_k^2 C_{gk}^2 g_m^2} \qquad (9\text{-}1\text{-}7)$$

There are several ways to minimize the inductance and capacitance effects, such as a reduction in lead length and electrode area. This minimization, however, also limits the power-handling capacity.

9-1-2 Transit-Angle Effects

Another limitation in the application of conventional tubes at microwave frequencies is the electron transit angle between electrodes. The electron transit angle is defined as

$$\theta_g \equiv \omega\tau_g = \frac{\omega d}{v_0} \qquad (9\text{-}1\text{-}8)$$

where $\tau_g = d/v_0$ is the transit time across the gap

d = separation between cathode and grid

$v_0 = 0.593 \times 10^6 \sqrt{V_0}$ is the velocity of the electron

V_0 = dc voltage

When frequencies are below microwave range, the transit angle is negligible. At microwave frequencies, however, the transit time (or angle) is large compared to the period of the microwave signal, and the potential between the cathode and the grid may alternate from 10 to 100 times during the electron transit. The grid potential during the negative half cycle thus removes energy that was given to the electron during the positive half cycle. Consequently, the electrons may oscillate back and forth in the cathode-grid space or return to the cathode. The overall result of transit-angle effects is to reduce the operating efficiency of the vacuum tube. The degenerate effect becomes more serious when frequencies are well above 1 GHz. Once electrons pass the grid, they are quickly accelerated to the anode by the high plate voltage.

When the frequency is below 1 GHz, the output delay is negligible in comparison with the phase of the grid voltage. This means that the transadmittance is a large real quantity, which is the usual transconductance g_m. At microwave frequencies the transit angle is not negligible, and the transadmittance becomes a complex number with a relatively small magnitude. This situation indicates that the output is decreased.

From the preceding analysis it is clear that the transit-angle effect can be minimized by first accelerating the electron beam with a very high dc voltage and then velocity-modulating it. This is indeed the principal operation of such microwave tubes as klystrons and magnetrons.

9-1-3 Gain-Bandwidth Product Limitation

In ordinary vacuum tubes the maximum gain is generally achieved by resonating the output circuit as shown in Fig. 9-1-2. In the equivalent circuit (see Fig. 9-1-2) it is assumed that $r_p \gg \omega L_k$. The load voltage is given by

$$\mathbf{V}_\ell = \frac{g_m V_g}{G + j[\omega C - 1/(\omega L)]} \tag{9-1-9}$$

where $G = 1/r_p + 1/R$
r_p = plate resistance
R = load resistance
L, C = tuning elements

The resonant frequency is expressed by

$$f_r = \frac{1}{2\pi\sqrt{LC}} \tag{9-1-10}$$
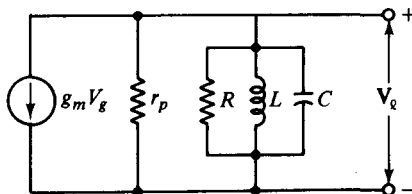
Figure 9-1-2   Output-tuned circuit of a pentode.

and the maximum voltage gain A_{max} at resonance by

$$A_{max} = \frac{g_m}{G} \qquad (9\text{-}1\text{-}11)$$

Since the bandwidth is measured at the half-power point, the denominator of Eq. (9-1-9) must be related by

$$G = \omega C - \frac{1}{\omega L} \qquad (9\text{-}1\text{-}12)$$

The roots of this quadratic equation are given by

$$\omega_1 = \frac{G}{2C} - \sqrt{\left(\frac{G}{2C}\right)^2 + \frac{1}{LC}} \qquad (9\text{-}1\text{-}13)$$

$$\omega_2 = \frac{G}{2C} + \sqrt{\left(\frac{G}{2C}\right)^2 + \frac{1}{LC}} \qquad (9\text{-}1\text{-}14)$$

Then the bandwidth can be expressed by

$$\text{BW} = \omega_2 - \omega_1 = \frac{G}{C} \qquad \text{for } \left(\frac{G}{2C}\right)^2 \gg \frac{1}{LC} \qquad (9\text{-}1\text{-}15)$$

Hence the gain-bandwidth product of the circuit of Fig. 9-1-2 is

$$A_m(\text{BW}) = \frac{g_m}{C} \qquad (9\text{-}1\text{-}16)$$

It is important to note that the gain-bandwidth product is independent of frequency. For a given tube, a higher gain can be achieved only at the expense of a narrower bandwidth. This restriction is applicable to a resonant circuit only. In microwave devices either reentrant cavities or slow-wave structures are used to obtain a possible overall high gain over a broad bandwidth.

9-2 KLYSTRONS

The two-cavity klystron is a widely used microwave amplifier operated by the principles of velocity and current modulation. All electrons injected from the cathode arrive at the first cavity with uniform velocity. Those electrons passing the first cavity gap at zeros of the gap voltage (or signal voltage) pass through with unchanged velocity; those passing through the positive half cycles of the gap voltage undergo an increase in velocity; those passing through the negative swings of the gap voltage undergo a decrease in velocity. As a result of these actions, the electrons gradually bunch together as they travel down the drift space. The variation in electron velocity in the drift space is known as *velocity modulation*. The density of the electrons in the second cavity gap varies cyclically with time. The electron beam contains an ac component and is said to be current-modulated. The maximum bunching should occur approximately midway between the second cavity grids during its retarding

phase; thus the kinetic energy is transferred from the electrons to the field of the second cavity. The electrons then emerge from the second cavity with reduced velocity and finally terminate at the collector. The charateristics of a two-cavity klystron amplifier are as follows:

1. Efficiency: about 40%.
2. Power output: average power (CW power) is up to 500 kW and pulsed power is up to 30 MW at 10 GHz.
3. Power gain: about 30 dB.

Figure 9-2-1 shows the present state of the art for U.S. high-power klystrons. Figure 9-2-2 shows the schematic diagram of a two-cavity klystron amplifier. The cavity close to the cathode is known as the *buncher cavity* or input cavity, which velocity-modulates the electron beam. The other cavity is called the *catcher cavity* or output cavity; it catches energy from the bunched electron beam. The beam then passes through the catcher cavity and is terminated at the collector. The quantitative analysis of a two-cavity klystron can be described in four parts under the following assumptions:

1. The electron beam is assumed to have a uniform density in the cross section of the beam.
2. Space-charge effects are negligible.
3. The magnitude of the microwave signal input is assumed to be much smaller than the dc accelerating voltage.

9-2-1 Reentrant Cavities

At a frequency well below the microwave range, the cavity resonator can be represented by a lumped-constant resonant circuit. When the operating frequency is increased to several tens of megahertz, both the inductance and the capacitance must be reduced to a minimum in order to maintain resonance at the operating frequency. Ultimately the inductance is reduced to a minimum by short wire. Therefore the reentrant cavities are designed for use in klystrons and microwave triodes. A reentrant cavity is one in which the metallic boundaries extend into the interior of the cavity. Several types of reentrant cavities are shown in Fig. 9-2-3. One of the commonly used reentrant cavities is the coaxial cavity shown in Fig. 9-2-4.

It is clear from Fig. 9-2-4 that not only has the inductance been considerably decreased but the resistance losses are markedly reduced as well, and the shelf-shielding enclosure prevents radiation losses. It is difficult to calculate the resonant frequency of the coaxial cavity. An approximation can be made, however, using transmission-line theory. The characteristic impedance of the coaxial line is given by

$$Z_0 = \frac{1}{2\pi} \sqrt{\frac{\mu}{\epsilon}} \, \ell n \frac{b}{a} \qquad \text{ohms} \qquad (9\text{-}2\text{-}1)$$
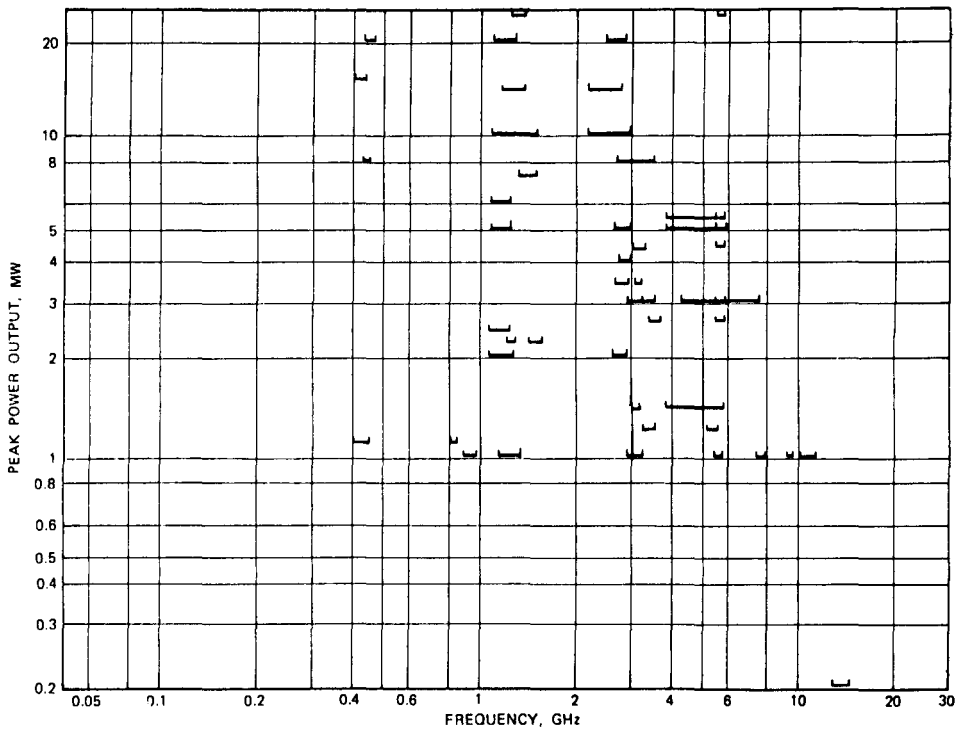
Figure 9-2-1 State of the art for U.S. high-power klystrons.
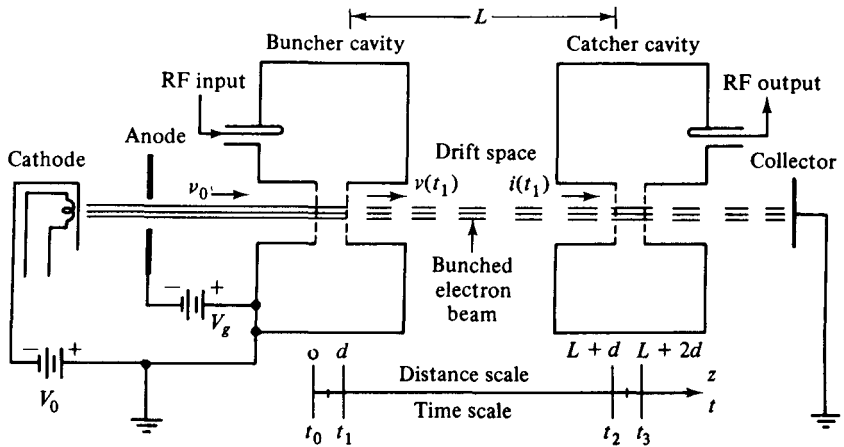
Figure 9-2-2 Two-cavity klystron amplifier.
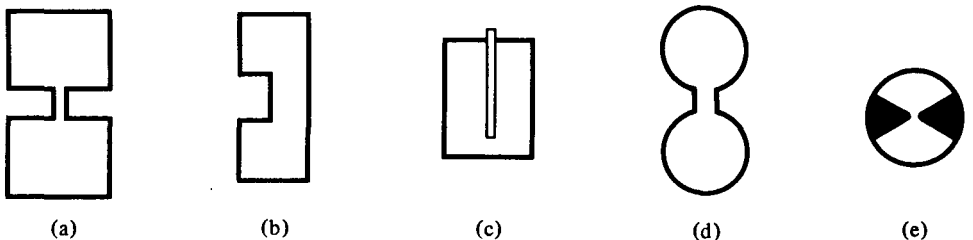
Figure 9-2-3 Reentrant cavities. (a) Coaxial cavity. (b) Radial cavity. (c) Tunable cavity. (d) Toroidal cavity. (c) Butterfly cavity.

(a) Coaxial cavity   (b) Cross-section of
               coaxial cavity
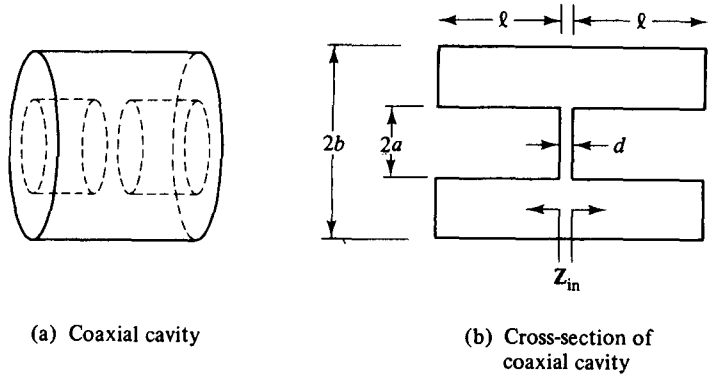
Figure 9-2-4 Coaxial cavity and its equivalent.

The coaxial cavity is similar to a coaxial line shorted at two ends and joined at the center by a capacitor. The input impedance to each shorted coaxial line is given by

$$Z_{in} = jZ_0 \tan(\beta \ell) \tag{9-2-2}$$

where ℓ is the length of the coaxial line.

Substitution of Eq. (9-2-1) in (9-2-2) results in

$$Z_{in} = j\frac{1}{2\pi} \sqrt{\frac{\mu}{\epsilon}} \ell n \frac{b}{a} \tan(\beta \ell) \tag{9-2-3}$$

The inductance of the cavity is given by

$$L = \frac{2X_{in}}{\omega} = \frac{1}{\pi\omega} \sqrt{\frac{\mu}{\epsilon}} \ell n \frac{b}{a} \tan(\beta \ell) \tag{9-2-4}$$

and the capacitance of the gap by

$$C_g = \frac{\epsilon \pi a^2}{d} \tag{9-2-5}$$

At resonance the inductive reactance of the two shorted coaxial lines in series is equal in magnitude to the capacitive reactance of the gap. That is, $\omega L = 1/(\omega C_g)$. Thus

$$\tan(\beta \ell) = \frac{dv}{\omega a^2 \ell n(b/a)} \tag{9-2-6}$$

where $v = 1/\sqrt{\mu\epsilon}$ is the phase velocity in any medium.

The solution to this equation gives the resonant frequency of a coaxial cavity. Since Eq. (9-2-6) contains the tangent function, it has an infinite number of solutions with larger values of frequency. Therefore this type of reentrant cavity can support an infinite number of resonant frequencies or modes of oscillation. It can be shown that a shorted coaxial-line cavity stores more magnetic energy than electric energy. The balance of the electric stored energy appears in the gap, for at resonance the magnetic and electric stored energies are equal.

The radial reentrant cavity shown in Fig. 9-2-5 is another commonly used reentrant resonator. The inductance and capacitance of a radial reentrant cavity is expressed by

$$L = \frac{\mu \ell}{2\pi} \ell n \frac{b}{a} \qquad (9\text{-}2\text{-}7)$$

$$C = \epsilon_0 \left[\frac{\pi a^2}{d} - 4a \, \ell n \frac{0.765}{\sqrt{\ell^2 + (b-a)^2}} \right] \qquad (9\text{-}2\text{-}8)$$

The resonant frequency [1] is given by

$$f_r = \frac{c}{2\pi \sqrt{\epsilon_r}} \left\{ a\ell \left[\frac{a}{2d} - \frac{2}{\ell} \ell n \frac{0.765}{\sqrt{\ell^2 + (b-a)^2}} \right] \ell n \frac{b}{a} \right\}^{-1/2} \qquad (9\text{-}2\text{-}9)$$

where $c = 3 \times 10^8$ m/s is the velocity of light in a vacuum.
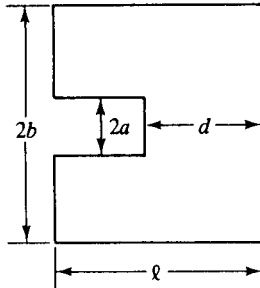
Figure 9-2-5 Radial reentrant cavity.

9-2-2 Velocity-Modulation Process

When electrons are first accelerated by the high dc voltage V_0 before entering the buncher grids, their velocity is uniform:

$$v_0 = \sqrt{\frac{2eV_0}{m}} = 0.593 \times 10^6 \sqrt{V_0} \qquad \text{m/s} \qquad (9\text{-}2\text{-}10)$$

In Eq. (9-2-10) it is assumed that electrons leave the cathode with zero velocity. When a microwave signal is applied to the input terminal, the gap voltage between the buncher grids appears as

$$V_s = V_1 \sin (\omega t) \qquad (9\text{-}2\text{-}11)$$

where V_1 is the amplitude of the signal and $V_1 \ll V_0$ is assumed.

In order to find the modulated velocity in the buncher cavity in terms of either the entering time t_0 or the exiting time t_1 and the gap transit angle θ_g as shown in Fig. 9-2-2 it is necessary to determine the average microwave voltage in the buncher gap as indicated in Fig. 9-2-6.

Since $V_1 \ll V_0$, the average transit time through the buncher gap distance d is

$$\tau \approx \frac{d}{v_0} = t_1 - t_0 \qquad (9\text{-}2\text{-}12)$$
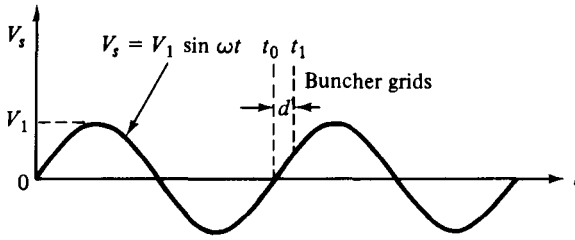
Figure 9-2-6   Signal voltage in the buncher gap.

The average gap transit angle can be expressed as

$$\theta_g = \omega\tau = \omega(t_1 - t_0) = \frac{\omega d}{v_0} \tag{9-2-13}$$

The average microwave voltage in the buncher gap can be found in the following way:

$$\langle V_s \rangle = \frac{1}{\tau}\int_{t_0}^{t_1} V_1 \sin(\omega t)dt = -\frac{V_1}{\omega\tau}[\cos(\omega t_1) - \cos(\omega t_0)]$$

$$= \frac{V_1}{\omega\tau}\left[\cos(\omega t_0) - \cos\left(\omega_0 + \frac{\omega d}{v_0}\right)\right] \tag{9-2-14}$$

Let

$$\omega t_0 + \frac{\omega d}{2v_0} = \omega t_0 + \frac{\theta_g}{2} = A$$

and

$$\frac{\omega d}{2v_0} = \frac{\theta_g}{2} = B$$

Then using the trigonometric identity that $\cos(A - B) - \cos(A + B) = 2\sin A \sin B$, Eq. (9-2-14) becomes

$$\langle V_s \rangle = V_1\frac{\sin[\omega d/2v_0]}{\omega d/(2v_0)}\sin\left(\omega t_0 + \frac{\omega d}{2v_0}\right) = V_1\frac{\sin(\theta_g/2)}{\theta_g/2}\sin\left(\omega t_0 + \frac{\theta_g}{2}\right) \tag{9-2-15}$$

It is defined as

$$\beta_i \equiv \frac{\sin[\omega d/(2v_0)]}{\omega d/(2v_0)} = \frac{\sin(\theta_g/2)}{\theta_g/2} \tag{9-2-16}$$

Note that β_i is known as the *beam-coupling coefficient* of the input cavity gap (see Fig. 9-2-7).

It can be seen that increasing the gap transit angle θ_g decreases the coupling between the electron beam and the buncher cavity; that is, the velocity modulation
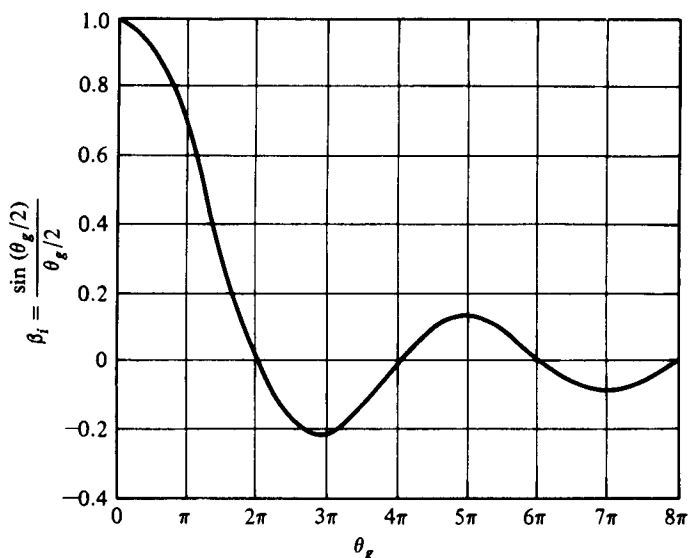
Figure 9-2-7   Beam-coupling coefficient versus gap transit angle.

of the beam for a given microwave signal is decreased. Immediately after velocity modulation, the exit velocity from the buncher gap is given by

$$v(t_1) = \sqrt{\frac{2e}{m}\left[V_0 + \beta_i V_1 \sin\left(\omega t_0 + \frac{\theta_g}{2}\right)\right]}$$

$$= \sqrt{\frac{2e}{m} V_0 \left[1 + \frac{\beta_i V_1}{V_0} \sin\left(\omega t_0 + \frac{\theta_g}{2}\right)\right]} \qquad (9\text{-}2\text{-}17)$$

where the factor $\beta_i V_1 / V_0$ is called the *depth of velocity modulation.*
    Using binomial expansion under the assumption of

$$\beta_i V_1 \ll V_0 \qquad (9\text{-}2\text{-}18)$$

Eq. (9-2-17) becomes

$$v(t_1) = v_0 \left[1 + \frac{\beta_i V_1}{2V_0} \sin\left(\omega t_0 + \frac{\theta_g}{2}\right)\right] \qquad (9\text{-}2\text{-}19)$$

Equation (9-2-19) is the equation of velocity modulation. Alternatively, the equation of velocity modulation can be given by

$$v(t_1) = v_0 \left[1 + \frac{\beta_i V_1}{2V_0} \sin\left(\omega t_1 - \frac{\theta_g}{2}\right)\right] \qquad (9\text{-}2\text{-}20)$$

9-2-3 Bunching Process

Once the electrons leave the buncher cavity, they drift with a velocity given by Eq. (9-2-19) or (9-2-20) along in the field-free space between the two cavities. The effect of velocity modulation produces bunching of the electron beam—or current modulation. The electrons that pass the buncher at $V_s = 0$ travel through with unchanged velocity v_0 and become the bunching center. Those electrons that pass the buncher cavity during the positive half cycles of the microwave input voltage V_s travel faster than the electrons that passed the gap when $V_s = 0$. Those electrons that pass the buncher cavity during the negative half cycles of the voltage V_s travel slower than the electrons that passed the gap when $V_s = 0$. At a distance of ΔL along the beam from the buncher cavity, the beam electrons have drifted into dense clusters. Figure 9-2-8 shows the trajectories of minimum, zero, and maximum electron acceleration.
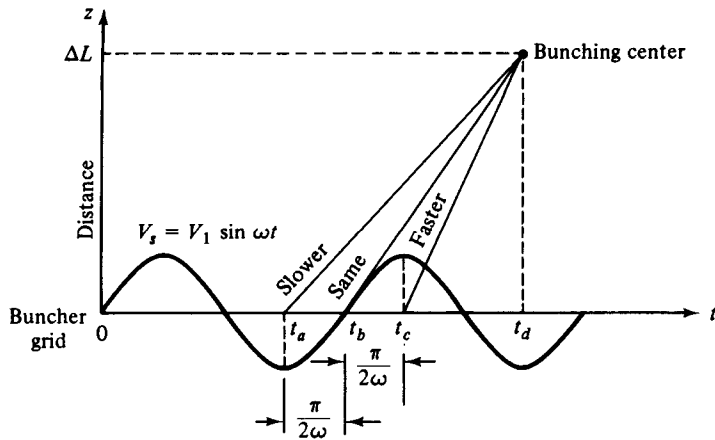
Figure 9-2-8   Bunching distance.

The distance from the buncher grid to the location of dense electron bunching for the electron at t_b is

$$\Delta L = v_0(t_d - t_b) \tag{9-2-21}$$

Similarly, the distances for the electrons at t_a and t_c are

$$\Delta L = v_{\min}(t_d - t_a) = v_{\min}\left(t_d - t_b + \frac{\pi}{2\omega}\right) \tag{9-2-22}$$

$$\Delta L = v_{\max}(t_d - t_c) = v_{\max}\left(t_d - t_b - \frac{\pi}{2\omega}\right) \tag{9-2-23}$$

From Eq. (9-2-19) or (9-2-20) the minimum and maximum velocities are

$$v_{\min} = v_0\left(1 - \frac{\beta_i V_1}{2V_0}\right) \tag{9-2-24}$$

$$v_{max} = v_0\left(1 + \frac{\beta_i V_1}{2V_0}\right) \qquad (9\text{-}2\text{-}25)$$

Substitution of Eqs. (9-2-24) and (9-2-25) in Eqs. (9-2-22) and (9-2-23), respectively, yields the distance

$$\Delta L = v_0(t_d - t_b) + \left[v_0\frac{\pi}{2\omega} - v_0\frac{\beta_i V_1}{2V_0}(t_d - t_b) - v_0\frac{\beta_i V_1}{2V_0}\frac{\pi}{2\omega}\right] \qquad (9\text{-}2\text{-}26)$$

and

$$\Delta L = v_0(t_d - t_b) + \left[-v_0\frac{\pi}{2\omega} + v_0\frac{\beta_i V_1}{2V_0}(t_d - t_b) + v_0\frac{\beta_i V_1}{2V_0}\frac{\pi}{2\omega}\right] \qquad (9\text{-}2\text{-}27)$$

The necessary condition for those electrons at t_a, t_b, and t_c to meet at the same distance ΔL is

$$v_0\frac{\pi}{2\omega} - v_0\frac{\beta_i V_1}{2V_0}(t_d - t_b) - v_0\frac{\beta_i V_1}{2V_0}\frac{\pi}{2\omega} = 0 \qquad (9\text{-}2\text{-}28)$$

and

$$-v_0\frac{\pi}{2\omega} + v_0\frac{\beta_i V_1}{2V_0}(t_d - t_b) + v_0\frac{\beta_i V_1}{2V_0}\frac{\pi}{2\omega} = 0 \qquad (9\text{-}2\text{-}29)$$

Consequently,

$$t_d - t_b \approx \frac{\pi V_0}{\omega\beta_i V_1} \qquad (9\text{-}2\text{-}30)$$

and

$$\Delta L = v_0\frac{\pi V_0}{\omega\beta_i V_1} \qquad (9\text{-}2\text{-}31)$$

It should be noted that the mutual repulsion of the space charge is neglected, but the qualitative results are similar to the preceding representation when the effects of repulsion are included. Furthermore, the distance given by Eq. (9-2-31) is not the one for a maximum degree of bunching. Figure 9-2-9 shows the distance-time plot or Applegate diagram.

What should the spacing be between the buncher and catcher cavities in order to achieve a maximum degree of bunching? Since the drift region is field free, the transit time for an electron to travel a distance of L as shown in Fig. 9-2-2 is given by

$$T = t_2 - t_1 = \frac{L}{v(t_1)} = T_0\left[1 - \frac{\beta_i V_1}{2V_0}\sin\left(\omega t_1 - \frac{\theta_g}{2}\right)\right] \qquad (9\text{-}2\text{-}32)$$

where the binomial expansion of $(1 + x)^{-1}$ for $|x| \ll 1$ has been replaced and $T_0 = L/v_0$ is the dc transit time. In terms of radians the preceding expression can be written

$$\omega T = \omega t_2 - \omega t_1 = \theta_0 - X\sin\left(\omega t_1 - \frac{\theta_g}{2}\right) \qquad (9\text{-}2\text{-}33)$$
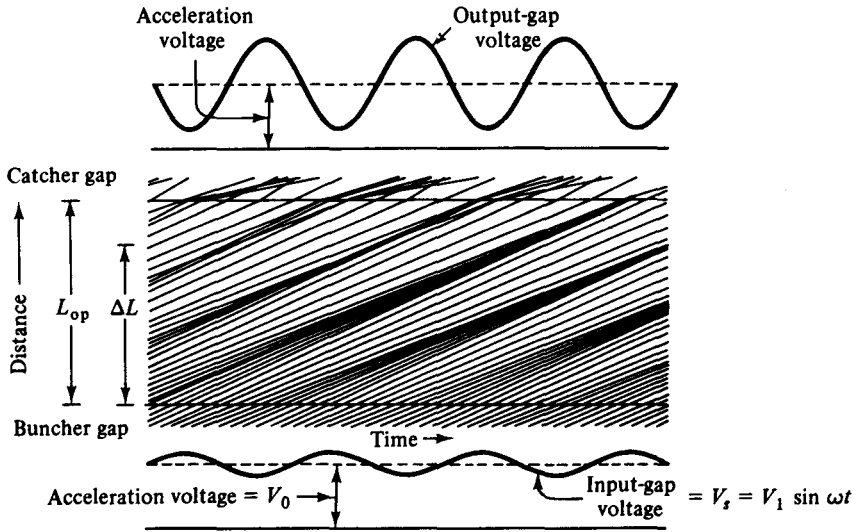
Figure 9-2-9   Applegate diagram.

where

$$\theta_0 = \frac{\omega L}{v_0} = 2\pi N \tag{9-2-34}$$

is the dc transit angle between cavities, N is the number of electron transit cycles in the drift space, and

$$X \equiv \frac{\beta_i V_1}{2V_0} \theta_0 \tag{9-2-35}$$

is defined as the *bunching parameter* of a klystron.

At the buncher gap a charge dQ_0 passing through at a time interval dt_0 is given by

$$dQ_0 = I_0 dt_0 \tag{9-2-36}$$

where I_0 is the dc current. From the principle of conservation of charges this same amount of charge dQ_0 also passes the catcher at a later time interval dt_2. Hence

$$I_0 |dt_0| = i_2 |dt_2| \tag{9-2-37}$$

where the absolute value signs are necessary because a negative value of the time ratio would indicate a negative current. Current i_2 is the current at the catcher gap. Rewriting Eq. (9-2-32) in terms of Eq. (9-2-19) yields

$$t_2 = t_0 + \tau + T_0 \left[1 - \frac{\beta_i V_1}{2V_0} \sin\left(\omega t_0 + \frac{\theta_g}{2}\right) \right] \tag{9-2-38}$$

Alternatively,

$$\omega t_2 - \left(\theta_0 + \frac{\theta_g}{2}\right) = \left(\omega t_0 + \frac{\theta_g}{2}\right) - X \sin\left(\omega t_0 + \frac{\theta_g}{2}\right) \tag{9-2-39}$$

where $(\omega t_0 + \theta_g/2)$ is the buncher cavity departure angle and $\omega t_2 - (\theta_0 + \theta_g/2)$ is the catcher cavity arrival angle. Figure 9-2-10 shows the curves for the catcher cav-
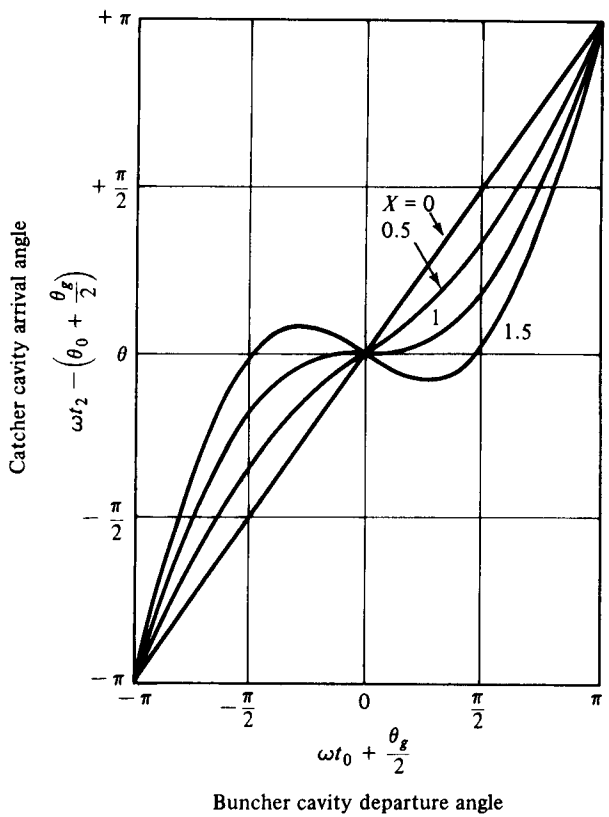
Figure 9-2-10 Catcher arrival angle versus buncher departure angle.

ity arrival angle as a function of the buncher cavity departure angle in terms of the bunching parameter X.

Differentiation of Eq. (9-2-38) with respect to t_0 results in

$$dt_2 = dt_0\left[1 - X \cos\left(\omega t_0 + \frac{\theta_g}{2}\right)\right] \qquad (9\text{-}2\text{-}40)$$

The current arriving at the catcher cavity is then given as

$$i_2(t_0) = \frac{I_0}{1 - X \cos(\omega t_0 + \theta_g/2)} \qquad (9\text{-}2\text{-}41)$$

In terms of t_2 the current is

$$i_2(t_2) = \frac{I_0}{1 - X \cos(\omega t_2 - \theta_0 - \theta_g/2)} \qquad (9\text{-}2\text{-}42)$$

In Eq. (9-2-42) the relationsip of $t_2 = t_0 + \tau + T_0$ is used—namely, $\omega t_2 = \omega t_0 + \omega \tau + \omega T_0 = \omega t_0 + \theta_g + \theta_0$. Figure 9-2-11 shows curves of the beam current $i_2(t_2)$ as a function of the catcher arrival angle in terms of the bunching parameter X.
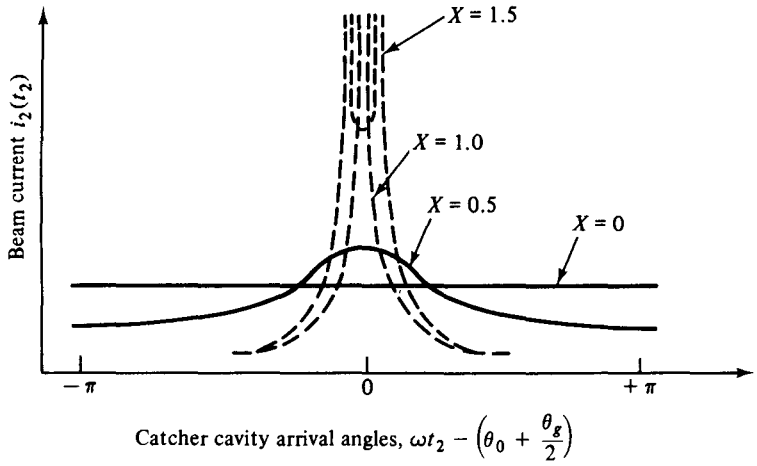
Figure 9-2-11 Beam current i_2 versus catcher cavity arrival angle.

The beam current at the catcher cavity is a periodic waveform of period $2\pi/\omega$ about dc current. Therefore the current i_2 can be expanded in a Fourier series and so

$$i_2 = a_0 + \sum_{n=1}^{\infty} [a_n \cos(n\omega t_2) + b_n \sin(n\omega t_2)] \qquad (9\text{-}2\text{-}43)$$

where n is an integer, excluding zero. The series coefficients a_0, a_n, and b_n in Eq. (9-2-43) are given by the integrals

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} i_2 \, d(\omega t_2)$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(n\omega t_2) \, d(\omega t_2) \qquad (9\text{-}2\text{-}44)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} i_2 \sin(n\omega t_2) \, d(\omega t_2)$$

Substitution of Eqs. (9-2-37) and (9-2-39) in Eq. (9-2-44) yields

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_0 \, d(\omega t_0) = I_0 \qquad (9\text{-}2\text{-}45)$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} I_0 \cos\left[(n\omega t_0 + n\theta_g + n\theta_0) + nX \sin\left(\omega t_0 + \frac{\theta_g}{2}\right) \right] d(\omega t_0) \qquad (9\text{-}2\text{-}46)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} I_0 \sin\left[(n\omega t_0 + n\theta_g + n\theta_0) + nX \sin\left(\omega t_0 + \frac{\theta_g}{2}\right) \right] d(\omega t_0) \qquad (9\text{-}2\text{-}47)$$

By using the trigonometric functions

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$$

and

$$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$$

the two integrals shown in Eqs. (9-2-40) and (9-2-41) involve cosines and sines of a sine function. Each term of the integrand contains an infinite number of terms of Bessel functions. These are

$$\cos\left[nX\sin\left(\omega t_0 + \frac{\theta_g}{2}\right)\right] = 2J_0(nX) + 2\left[J_2(nX)\cos 2\left(\omega t_0 + \frac{\theta_g}{2}\right)\right]$$

$$+ 2\left[J_4(nX)\cos 4\left(\omega t_0 + \frac{\theta_g}{2}\right)\right]$$

$$+ \cdots \qquad (9\text{-}2\text{-}48)$$

and

$$\sin\left[nX\sin\left(\omega t_0 + \frac{\theta_g}{2}\right)\right] = 2\left[J_1(nX)\sin\left(\omega t_0 + \frac{\theta_g}{2}\right)\right]$$

$$+ 2\left[J_3(nX)\sin 3\left(\omega t_0 + \frac{\theta_g}{2}\right)\right]$$

$$+ \cdots \qquad (9\text{-}2\text{-}49)$$

If these series are substituted into the integrands of Eqs. (9-2-46) and (9-2-47), respectively, the integrals are readily evaluated term by term and the series coefficients are

$$a_n = 2I_0 J_n(nX)\cos(n\theta_g + n\theta_0) \qquad (9\text{-}2\text{-}50)$$

$$b_n = 2I_0 J_n(nX)\sin(n\theta_g + n\theta_0) \qquad (9\text{-}2\text{-}51)$$

where $J_n(nX)$ is the nth-order Bessel function of the first kind (see Fig. 9-2-12).
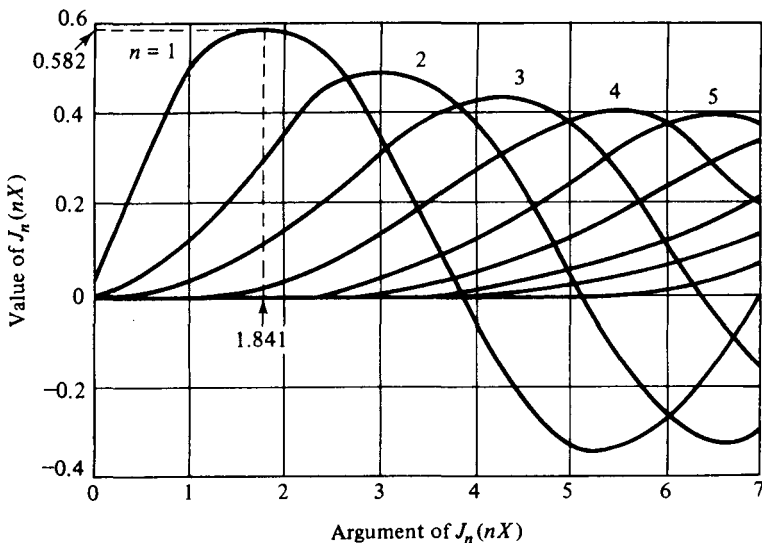
Figure 9-2-12   Bessel functions $J_n(nX)$.

Substitution of Eqs. (9-2-45), (9-2-50), and (9-2-51) in Eq. (9-2-43) yields the beam current i_2 as

$$i_2 = I_0 + \sum_{n=1}^{\infty} 2I_0 J_n(nX) \cos\left[n\omega(t_2 - \tau - T_0)\right] \qquad (9\text{-}2\text{-}52)$$

The fundamental component of the beam current at the catcher cavity has a magnitude

$$I_f = 2I_0 J_1(X) \qquad (9\text{-}2\text{-}53)$$

This fundamental component I_f has its maximum amplitude at

$$X = 1.841 \qquad (9\text{-}2\text{-}54)$$

The optimum distance L at which the maximum fundamental component of current occurs is computed from Eqs. (9-2-34), (9-2-35), and (9-2-54) as

$$L_{\text{optimum}} = \frac{3.682 v_0 V_0}{\omega \beta_i V_1} \qquad (9\text{-}2\text{-}55)$$

It is interesting to note that the distance given by Eq. (9-2-31) is approximately 15% less than the result of Eq. (9-2-55). The discrepancy is due in part to the approximations made in deriving Eq. (9-2-31) and to the fact that the maximum fundamental component of current will not coincide with the maximun electron density along the beam because the harmonic components exist in the beam.

9-2-4 Output Power and Beam Loading

The maximum bunching should occur approximately midway between the catcher grids. The phase of the catcher gap voltage must be maintained in such a way that the bunched electrons, as they pass through the grids, encounter a retarding phase. When the bunched electron beam passes through the retarding phase, its kinetic energy is transferred to the field of the catcher cavity. When the electrons emerge from the catcher grids, they have reduced velocity and are finally collected by the collector.

The induced current in the catcher cavity.   Since the current induced by the electron beam in the walls of the catcher cavity is directly proportional to the amplitude of the microwave input voltage V_1, the fundamental component of the induced microwave current in the catcher is given by

$$i_{2\text{ind}} = \beta_0 i_2 = \beta_0 2I_0 J_1(X) \cos[\omega(t_2 - \tau - T_0)] \qquad (9\text{-}2\text{-}56)$$

where β_0 is the beam coupling coefficient of the catcher gap. If the buncher and catcher cavities are identical, then $\beta_i = \beta_0$. The fundamental component of current induced in the catcher cavity then has a magnitude

$$I_{2\text{ind}} = \beta_0 I_2 = \beta_0 2I_0 J_1(X) \qquad (9\text{-}2\text{-}57)$$

Figure 9-2-13 shows an output equivalent circuit in which R_{sho} represents the wall resistance of catcher cavity, R_B the beam loading resistance, R_L the external load resistance, and R_{sh} the effective shunt resistance.
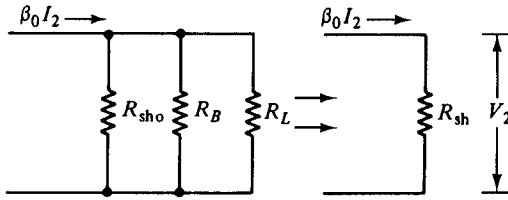
Figure 9-2-13 Output equivalent circuit.

The output power delivered to the catcher cavity and the load is given as

$$P_{\text{out}} = \frac{(\beta_0 I_2)^2}{2} R_{\text{sh}} = \frac{\beta_0 I_2 V_2}{2} \tag{9-2-58}$$

where R_{sh} is the total equivalent shunt resistance of the catcher circuit, including the load, and V_2 is the fundamental component of the catcher gap voltage.

Efficiency of klystron.   The electronic efficiency of the klystron amplifier is defined as the ratio of the output power to the input power:

$$\text{Efficiency} \equiv \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{\beta_0 I_2 V_2}{2 I_0 V_0} \tag{9-2-59}$$

in which the power losses to the beam loading and cavity walls are included.

If the coupling is perfect, $\beta_0 = 1$, the maximum beam current approaches $I_{2\text{max}} = 2I_0(0.582)$, and the voltage V_2 is equal to V_0. Then the maximum electronic efficiency is about 58%. In practice, the electronic efficiency of a klystron amplifier is in the range of 15 to 30%. Since the efficiency is a function of the catcher gap transit angle θ_g, Fig. 9-2-14 shows the maximum efficiency of a klystron as a function of catcher transit angle.
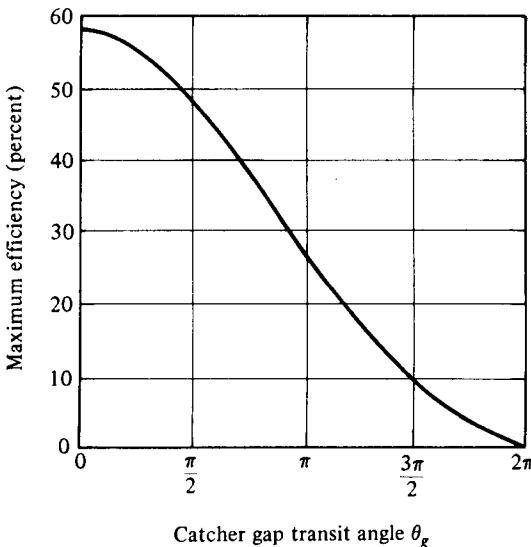
Figure 9-2-14 Maximum efficiency of klystron versus transit angle.

Mutual conductance of a klystron amplifier. The equivalent mutual conductance of the klystron amplifier can be defined as the ratio of the induced output current to input voltage. That is

$$|G_m| \equiv \frac{i_{2\,ind}}{V_1} = \frac{2\beta_0 I_0 J_1(X)}{V_1} \tag{9-2-60}$$

From Eq. (9-2-35) the input voltage V_1 can be expressed in terms of the bunching parameter X as

$$V_1 = \frac{2V_0}{\beta_0 \theta_0} X \tag{9-2-61}$$

In Eq. (9-2-61) it is assumed that $\beta_0 = \beta_i$. Substitution of Eq. (9-2-61) in Eq. (9-2-60) yields the normalized mutual conductance as

$$\frac{|G_m|}{G_0} = \beta_0^2 \theta_0 \frac{J_1(X)}{X} \tag{9-2-62}$$

where $G_0 = I_0/V_0$ is the dc beam conductance. The mutual conductance is not a constant but decreases as the bunching parameter X increases. Figure 9-2-15 shows the curves of normalized transconductance as a function of X.

Bunching parameter $\quad X = \dfrac{\beta_0 \theta_0 V_1}{2V_0}$
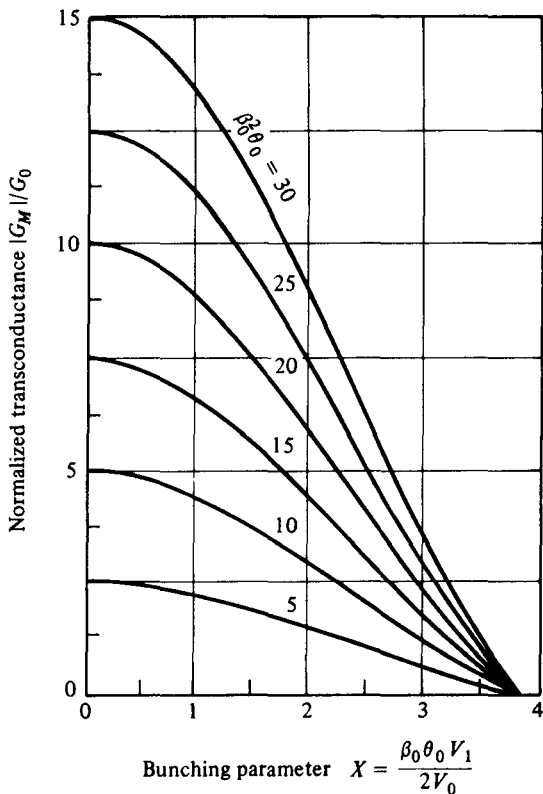
Figure 9-2-15 Normalized transconductance versus bunching parameter.

It can be seen from the curves that, for a small signal, the normalized transconductance is maximum. That is,

$$\frac{|G_m|}{G_0} = \frac{\beta_0^2 U_0}{2} \tag{9-2-63}$$

For a maximum output at $X = 1.841$, the normalized mutual conductance is

$$\frac{|G_m|}{G_0} = 0.316\beta_0^2 \theta_0 \tag{9-2-64}$$

The voltage gain of a klystron amplifier is defined as

$$A_v \equiv \frac{|V_2|}{|V_1|} = \frac{\beta_0 I_2 R_{\text{sh}}}{V_1} = \frac{\beta_0^2 \theta_0}{R_0} \frac{J_1(X)}{X} R_{\text{sh}} \tag{9-2-65}$$

where $R_0 = V_0/I_0$ is the dc beam resistance. Substitution of Eqs. (9-2-57) and (9-2-61) in Eq. (9-2-65) results in

$$A_v = G_m R_{\text{sh}} \tag{9-2-66}$$

Power required to bunch the electron beam.   As described earlier, the bunching action takes place in the buncher cavity. When the buncher gap transit angle is small, the average energy of electrons leaving the buncher cavity over a cycle is nearly equal to the energy with which they enter. When the buncher gap transit angle is large, however, the electrons that leave the buncher gap have greater average energy than when they enter. The difference between the average exit energy and the entrance energy must be supplied by the buncher cavity to bunch the electron beam. It is difficult to calculate the power required to produce the bunching action. Feenberg did some extensive work on beam loading [6]. The ratio of the power required to produce bunching action over the dc power required to perform the electron beam is given by Feenberg as

$$\frac{P_B}{P_0} = \frac{V_1^2}{2V_0^2}\left[\frac{1}{2}\beta_i^2 - \frac{1}{2}\beta_i \cos\left(\frac{\theta_g}{2}\right)\right] = \frac{V_1^2}{2V_0^2} F(\theta_g) \tag{9-2-67}$$

where

$$F(\theta_g) = \frac{1}{2}\left[\beta_i^2 - \beta_i \cos\left(\frac{\theta_g}{2}\right)\right]$$

The dc power is

$$P_0 = V_0^2 G_0 \tag{9-2-68}$$

where $G_0 = I_0/V_0$ is the equivalent electron beam conductance. The power given by the buncher cavity to produce beam bunching is

$$P_B = \frac{V_1^2}{2} G_B \tag{9-2-69}$$

where G_B is the equivalent bunching conductance. Substitution of Eqs. (9-2-68) and (9-2-69) in Eq. (9-2-67) yields the normalized electronic conductance as

$$\frac{G_B}{G_0} = \frac{R_0}{R_B} = F(\theta_g) \tag{9-2-70}$$

Figure 9-2-16 shows the normalized electronic conductance as a function of the buncher gap transit angle. It can be seen that there is a critical buncher gap transit angle for a minimum equivalent bunching resistance. When the transit angle θ_g is 3.5 rad, the equivalent bunching resistance is about five times the electron beam resistance. The power delivered by the electron beam to the catcher cavity can be expressed as

$$\frac{V_2^2}{2R_{sh}} = \frac{V_2^2}{2R_{sho}} + \frac{V_2^2}{2R_B} + \frac{V_2^2}{2R_L} \tag{9-2-71}$$

As a result, the effective impedance of the catcher cavity is

$$\frac{1}{R_{sh}} = \frac{1}{R_{sho}} + \frac{1}{R_B} + \frac{1}{R_L} \tag{9-2-72}$$

Finally, the loaded quality factor of the catcher cavity circuit at the resonant frequency can be written

$$\frac{1}{Q_L} = \frac{1}{Q_0} + \frac{1}{Q_B} + \frac{1}{Q_{ext}} \tag{9-2-73}$$

where Q_L = loaded quality factor of the whole catcher circuit
Q_0 = quality factor of the catcher walls
Q_B = quality factor of the beam loading
Q_{ext} = quality factor of the external load

Buncher gap transit angle θ_g, radians
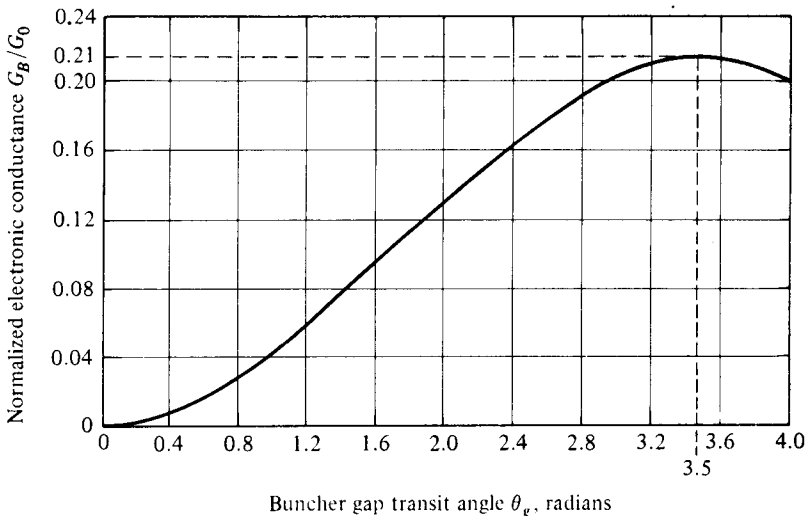
Figure 9-2-16 Normalized electronic conductance versus buncher gap transit angle.

Example 9-2-1:   Klystron Amplifier

A two-cavity klystron amplifier has the following parameters:

$$V_0 = 1000 \text{ V} \qquad R_0 = 40 \text{ k}\Omega$$

$$I_0 = 25 \text{ mA} \qquad f = 3 \text{ GHz}$$

Gap spacing in either cavity:	$d = 1$ mm
Spacing between the two cavities:	$L = 4$ cm
Effective shunt impedance, excluding beam loading:	$R_{sh} = 30$ kΩ

a. Find the input gap voltage to give maximum voltage V_2.
b. Find the voltage gain, neglecting the beam loading in the output cavity.
c. Find the efficiency of the amplifier, neglecting beam loading.
d. Calculate the beam loading conductance and show that neglecting it was justified in the preceding calculations.

Solution

a. For maximum V_2, $J_1(X)$ must be maximum. This means $J_1(X) = 0.582$ at $X = 1.841$. The electron velocity just leaving the cathode is

$$v_0 = (0.593 \times 10^6)\sqrt{V_0} = (0.593 \times 10^6)\sqrt{10^3} = 1.88 \times 10^7 \text{ m/s}$$

The gap transit angle is

$$\theta_g = \omega \frac{d}{v_0} = 2\pi (3 \times 10^9)\frac{10^{-3}}{1.88 \times 10^7} = 1 \text{ rad}$$

The beam-coupling coefficient is

$$\beta_i = \beta_0 = \frac{\sin (\theta_g/2)}{\theta_g/2} = \frac{\sin (1/2)}{1/2} = 0.952$$

The dc transit angle between the cavities is

$$\theta_0 = \omega T_0 = \omega \frac{L}{v_0} = 2\pi (3 \times 10^9)\frac{4 \times 10^{-2}}{1.88 \times 10^7} = 40 \text{ rad}$$

The maximum input voltage V_1 is then given by

$$V_{1 \text{ max}} = \frac{2V_0 X}{\beta_i \theta_0} = \frac{2(10^3)(1.841)}{(0.952)(40)} = 96.5 \text{ V}$$

b. The voltage gain is found as

$$A_v = \frac{\beta_0^2 \theta_0}{R_0} \frac{J_1(X)}{X} R_{sh} = \frac{(0.952)^2(40)(0.582)(30 \times 10^3)}{4 \times 10^4 \times 1.841} = 8.595$$

c. The efficiency can be found as follows:

$$I_2 = 2I_0 J_1(X) = 2 \times 25 \times 10^{-3} \times 0.582 = 29.1 \times 10^{-3} \text{ A}$$

$$V_2 = \beta_0 I_2 R_{sh} = (0.952)(29.1 \times 10^{-3})(30 \times 10^3) = 831 \text{ V}$$

$$\text{Efficiency} = \frac{\beta_0 I_2 V_2}{2I_0 V_0} = \frac{(0.952)(29.1 \times 10^{-3})(831)}{2(25 \times 10^{-3})(10^3)} = 46.2\%$$

d. Calculate the beam loading conductance (refer to Fig. 9-2-13). The beam loading conductance G_B is

$$G_B = \frac{G_0}{2}\left(\beta_0^2 - \beta_0 \cos \frac{\theta_g}{2}\right) = \frac{25 \times 10^{-6}}{2}[(0.952)^2 - (0.952) \cos (28.6°)]$$

$$= 8.8 \times 10^{-7} \text{ mho}$$

Then the beam loading resistance R_B is

$$R_B = \frac{1}{G_B} = 1.14 \times 10^6 \; \Omega$$

In comparison with R_L and R_{sho} or the effective shunt resistance R_{sh}, the beam loading resistance is like an open circuit and thus can be neglected in the preceding calculations.

9-2-5 State of the Art

Extended Interaction.   The most common form of extended interaction has been attained recently by coupling two or more adjacent klystron cavities. Figure 9-2-17 shows schematically a five-section extended-interaction cavity as compared to a single-gap klystron cavity.

(a)

(b)
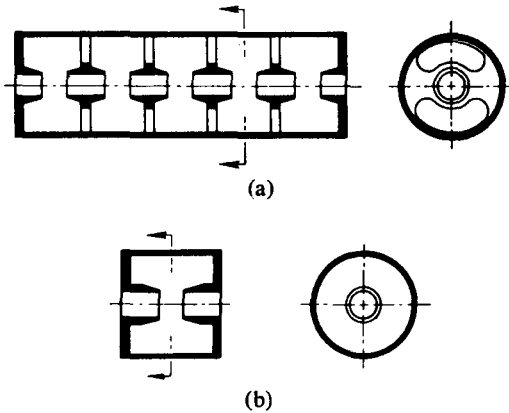
Figure 9-2-17   Comparison of a five-gap extended interaction cavity with a single-gap klystron cavity. (a) Five-gap coupled cavity resonator. (b)Single-gap klystron cavity. (*After A. Staprans et al.* [7]; reprinted by permission of IEEE.)

High efficiency and large power.   In the 1960s much effort was devoted to improving the efficiency of klystrons. For instance, a 50-kW experimental tube has demonstrated 75% efficiency in the industrial heating band [8]. The VA-884D klystron is a five-cavity amplifier whose operating characteristics are listed in Table 9-2-1.

One of the better-known high-peak-power klystrons is the tube developed specifically for use in the 2-mile Stanford Linear Accelerator [9] at Palo Alto, California. A cutaway view of the tube is shown in Fig. 9-2-18. The operating characteristics of this tube are listed in Table 9-2-2.

The Varian CW superpower klystron amplifier VKC-8269A as shown in Fig. 9-2-19 has an output power of 500 kW (CW) at frequency of 2.114 GHz. Its power

TABLE 9-2-1   VA-884D OPERATING CHARACTERISTICS

Frequency	5.9–6.45 GHz
Power output	14 kW
Gain	52 dB
Efficiency	36%
Electronic bandwidth (1 dB)	75 MHz
Beam voltage	16.5 kV
Beam current	2.4 A
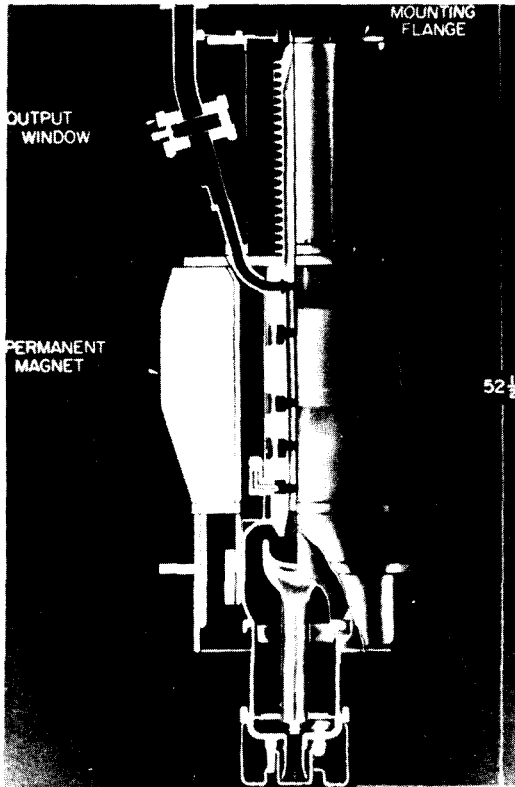
Figure 9-2-18   Cutaway of 24-MW S-band permanent-magnet-focused klystron. (*Courtesy of Stanford Linear Accelerator Center*.)

TABLE 9-2-2   OPERATING CHARACTERISTICS OF THE STANFORD LINEAR ACCELERATOR CENTER HIGH-POWER KLYSTRON [9]

Frequency	2.856	GHz
RF pulse width	2.5	μs
Pulse repetition rate	60 to 360	pps
Peak power output	24	MW
Beam voltage	250	kV
Beam current	250	A
Gain	50 to 55	dB
Efficiency	about 36	%
Weight of permanent (focusing) magnet	363	kg
Electronic bandwidth (1 dB)	20	MHz
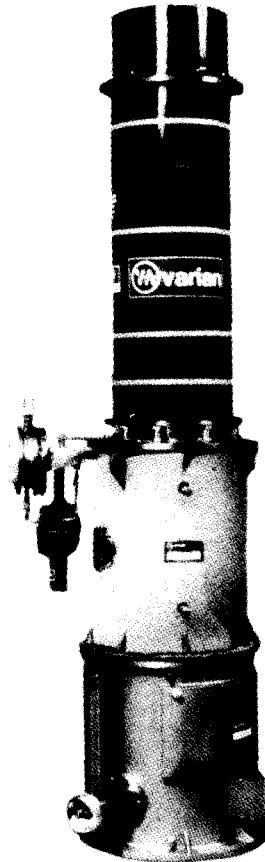
Figure 9-2-19 Photograph of Varian VKC-8269A CW super-power klystron amplifier. (*Courtesy of Varian Associates, Inc.*)

gain is 56 dB and efficiency is 50 percent. The beam voltage is 62 V (dc) and the beam current is 16.50 A (dc).

Long-life improvement. A new long-life klystron amplifier tube with a design life in excess of ten years, three times the current design life, has been developed by the Electron Dynamics Division of the Hughes Aircraft Company. Key to the long-life klystron was the development of a method of reducing the operating temperature of the tube's cathode. The cathode is made of porous tungsten impregnated with barium, calcium, and aluminum oxides. The new cathode is coated with a layer of osmium ruthenium alloy that lowers its work function, which is the temperature necessary for electrons to be emitted. This temperature reduction cuts evaporation of barium 10-fold and extends the life of the cathode.

9-3 MULTICAVITY KLYSTRON AMPLIFIERS

The typical power gain of a two-cavity klystron amplifier is about 30 dB. In order to achieve higher overall gain, one way is to connect several two-cavity tubes in cascade, feeding the output of each of the tubes to the input of the following one. Be-
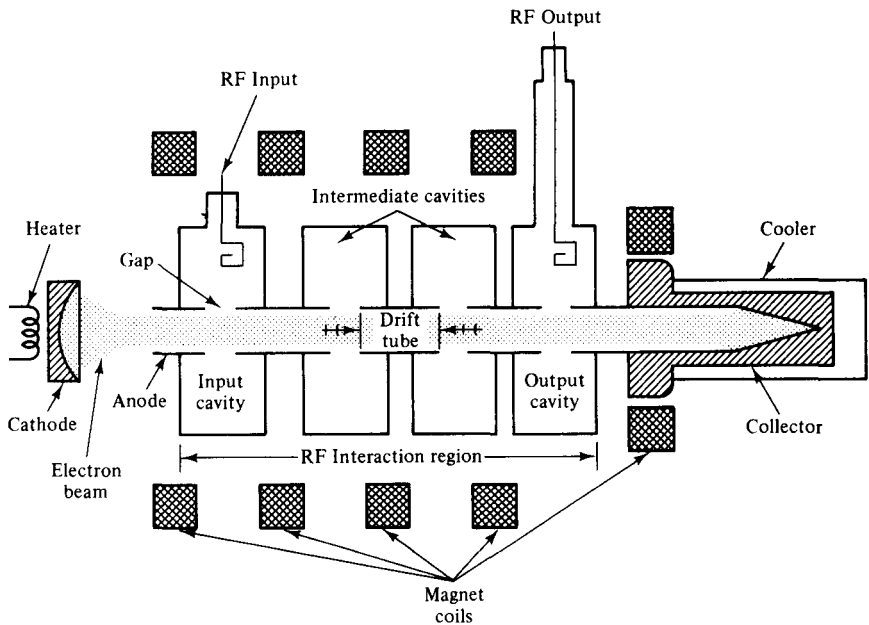
Figure 9-3-1   Schematic diagram of a four-cavity klystron amplifier. (*Courtesy of Varian Associates, Inc.*)

sides using the multistage techniques, the tube manufacturers have designed and produced multicavity klystron to serve the high-gain requirement as shown in Fig. 9-3-1.

In a multicavity klystron each of the intermediate cavities, placed at a distance of the bunching parameter X of 1.841 away from the previous cavity, acts as a buncher with the passing electron beam inducing a more enhanced RF voltage than the previous cavity, which in turn sets up an increased velocity modulation. The spacing between the consecutive cavities would therefore diminish because of the requirement of X being 1.841 and an increasing velocity modulating RF voltage as the beam progresses through the various cavities. Keeping the intercavity distance constant, an increasing beam voltage V_0 could be used in the subsequent cavities. Figure 9-3-2 shows the photograph of a four-cavity klystron amplifier.

9-3-1 Beam-Current Density

When the two-cavity klystron amplifier was discussed in Section 9-2, it was assumed that the space-charge effect was negligible, and it was not considered because of the assumed small density of electrons in the beam for low-power amplifier. However, when high-power klystron tubes are analyzed the electron density of the beam is large and the forces of mutual repulsion of the electrons must be considered. When the elecrons perturbate in the electron beam, the electron density consists of a dc part plus an RF perturbation caused by the electron bunches. The space-charge forces within electron bunches vary with the size and shape of an electron beam. In
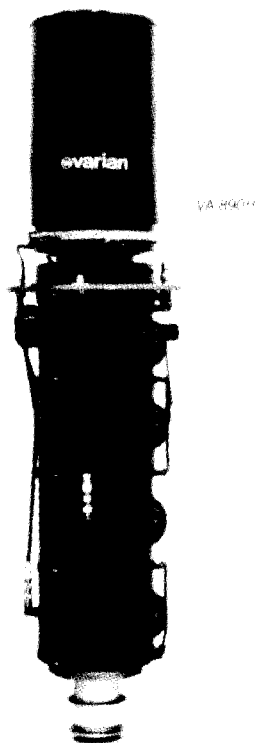
Figure 9-3-2 Photograph of a four-cavity klystron amplifier VA-890H. (*Courtesy of Varian Associates, Inc.*)

an infinitely wide beam, the electric fields (and, thus, the space-charge forces) are constrained to act only in the axial direction. In a finite beam, the electric fields are radial as well as axial with the result that the axial component is reduced in comparison to the infinite beam. With reduced axial space-charge force, the plasma frequency is reduced and the plasma wavelength is increased.

Mathematically, let the charge-density and velocity perturbations be simple sinusoidal variations in both time and position. They are

Charge density: $\qquad \rho = B \cos (\beta_e z) \cos (\omega_q t + \theta)$ $\qquad$ (9-3-1)

Velocity perturbation: $\qquad \mathcal{V} = -C \sin (\beta_e z) \sin (\omega_q t + \theta)$ $\qquad$ (9-3-2)

where B = constant of charge-density perturbation

$\qquad C$ = constant of velocity perturbation

$\qquad \beta_e = \dfrac{\omega}{\mathcal{V}_0}$ is the dc phase constant of the electron beam

$\qquad \omega_q = R\omega_p$ is the pertubation frequency or reduced plasma frequency

$\qquad R = \omega_q/\omega_p$ is the space-charge reduction factor and varies from 0 to 1

$\qquad \omega_p = \sqrt{\dfrac{e\rho_0}{m\epsilon_0}}$ is the plasma frequency and is a function of the electron-beam density

$\qquad \theta$ = phase angle of oscillation

The electron plasma frequency is the frequency at which the electrons will oscillate in the electron beam. This plasma frequency applies only to a beam of infinite diameter. Practical beams of finite diameter are characterized by plasma frequency that is less than ω_p. This lower plasma frequency is called the *reduced plasma frequency* and is designated ω_q. The space-charge reduction factor R is a function of the beam radius r and the ratio n of the beam-tunnel radius to the beam radius as shown in Fig. 9-3-3 [10].
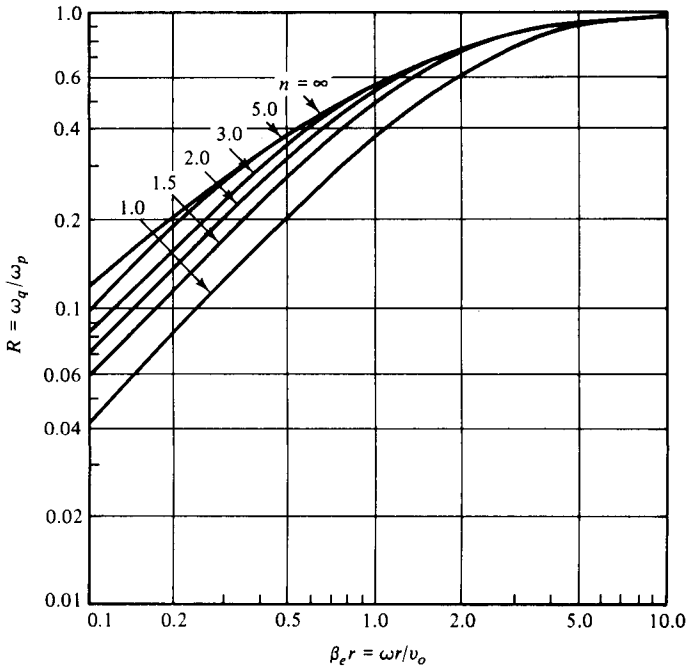
Figure 9-3-3   Plasma-frequency reduction factor R [10] (*Reprinted by permission of Pergamon Press.*)

As an example of the effect of the reduced space-charge forces, if $\beta_e r = \omega r / \mathcal{V}_0 = 0.85$ and $n = 2$, the reduced factor is $R = 0.5$. Therefore, $\omega_q = 2\omega_p$, which means that in a klystron the cavities would be placed twice as far apart as would be indicated by infinite-beam calculation. The total charge density and electron velocity are given by

$$\rho_{\text{tot}} = -\rho_0 + \rho \tag{9-3-3}$$

$$\mathcal{V}_{\text{tot}} = \mathcal{V}_0 + \mathcal{V} \tag{9-3-4}$$

where ρ_0 = dc electron charge density
  ρ = instantaneous RF charge density
  $\mathcal{V}_0$ = dc electron velocity
  $\mathcal{V}$ = instantaneous electron velocity perturbation

The total electron beam-current density can be written as

$$J_{\text{tot}} = -J_0 + J \tag{9-3-5}$$

where J_0 = dc beam-current density
$\quad\quad J$ = instantaneous RF beam-current perturbation

The instantaneous convection beam-current density at any point in the beam is expressed as

$$
\begin{aligned}
J_{\text{tot}} = \rho_{\text{tot}} \mathcal{V}_{\text{tot}} &= (-\rho_0 + \rho)(\mathcal{V}_0 + \mathcal{V}) \\
&= -\rho_0 \mathcal{V}_0 - \rho_0 \mathcal{V} + \rho \mathcal{V}_0 + \rho \mathcal{V} \\
&= -J_0 + J
\end{aligned}
\tag{9-3-6}
$$

where $J = \rho \mathcal{V}_0 - \rho_0 \mathcal{V}$ $\tag{9-3-7}$
$\quad\quad J_0 = \rho_0 \mathcal{V}_0$ is replaced
$\quad\quad \rho \mathcal{V}$ = very small is ignored

In accordance with the law of conservation of electric charge, the continuity equation can be written as

$$\boldsymbol{\nabla} \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \tag{9-3-8}$$

where $J = \rho \mathcal{V}_0 - \rho_0 \mathcal{V}$ is in positive z direction only. From

$$
\begin{aligned}
\frac{\partial J}{\partial z} = &-\omega B \sin\left(\beta_e z - \omega t\right) \cos\left(\omega_q t + \theta\right) \\
&+ \beta_e \rho_0 C \cos\left(\beta_e z - \omega t\right) \sin\left(\omega_q t + \theta\right)
\end{aligned}
\tag{9-3-9}
$$

and

$$
\begin{aligned}
-\frac{\partial \rho}{\partial t} = &-\omega B \sin\left(\beta_e z - \omega t\right) \cos\left(\omega_q t + \theta\right) \\
&+ \omega_q B \cos\left(\beta_e z - \omega t\right) \sin\left(\omega_q t + \theta\right)
\end{aligned}
\tag{9-3-10}
$$

then we have

$$\omega_q B = \beta_e \rho_0 C \tag{9-3-11}$$

The beam-current density and the modulated velocity are expressed as [10]

$$
\begin{aligned}
J = &\mathcal{V}_0 B \cos\left(\beta_e z - \omega t\right) \cos\left(\omega_q t + \theta\right) \\
&+ \frac{\omega_q}{\omega} \mathcal{V}_0 B \sin\left(\beta_e z - \omega t\right) \sin\left(\omega_q t + \theta\right)
\end{aligned}
\tag{9-3-12}
$$

and

$$\mathcal{V} = -\mathcal{V}_0 \frac{\beta_i V_1}{2V_0} \cos\left(\beta_q z\right) \sin\left(\beta_e z - \omega t\right) \tag{9-3-13}$$

In practical microwave tubes, the ratio of ω_q/ω is much smaller than unity and the second term in Eq. (9-3-12) may be neglected in comparison with the first. Then we obtain

$$J = \mathcal{V}_0 B \cos (\beta_e z - \omega t) \cos (\omega_q t + \theta) \qquad (9\text{-}3\text{-}14)$$

Example 9-3-1:   Four-Cavity Klystron

A four-cavity klystron VA-828 has the following parameters:

Beam voltage:	$V_0 = 14.5$ kV
Beam current:	$I_0 = 1.4$ A
Operating frequency:	$f = 10$ GHz
dc electron charge density:	$\rho_0 = 10^{-6}$ C/m³
RF charge density:	$\rho = 10^{-8}$ C/m³
Velocity perturbation:	$\mathcal{V} = 10^5$ m/s

Compute:

   a. The dc electron velocity
   b. The dc phase constant
   c. The plasma frequency
   d. The reduced plasma frequency for $R = 0.4$
   e. The dc beam current density
   f. The instantaneous beam current density

Solution

   a. The dc electron velocity is

$$\mathcal{V}_0 = 0.593 \times 10^6 \sqrt{14.5 \times 10^3} = 0.714 \times 10^8 \text{ m/s}$$

   b. The dc phase constant is

$$\beta_e = \frac{2\pi \times 10^{10}}{0.714 \times 10^8} = 8.80 \times 10^2 \text{ rads/m}$$

   c. The plasma frequency is

$$\omega_p = \left(1.759 \times 10^{11} \times \frac{10^{-6}}{8.854 \times 10^{-12}}\right)^{1/2} = 1.41 \times 10^8 \text{ rad/s}$$

   d. The reduced plasma frequency for $R = 0.4$ is

$$\omega_q = 0.4 \times 1.41 \times 10^8 = 0.564 \times 10^8 \text{ rad/s}$$

   e. The dc beam current density is

$$J_0 = 10^{-6} \times 0.714 \times 10^8 = 71.4 \text{ A/m}^2$$

   f. The instantaneous beam current density is

$$J = 10^{-8} \times 0.714 \times 10^8 - 10^{-6} \times 10^5 = 0.614 \text{ A/m}^2$$

The electrons leaving the input gap of a klystron amplifier have a velocity given by Eq. (9-2-17) at the exit grid as

$$\mathscr{V}(t_1) = \mathscr{V}_0 \left[1 + \frac{\beta_i V_1}{2V_0} \sin (\omega\tau) \right] \tag{9-3-15}$$

where V_1 = magnitude of the input signal voltage

$\tau = \dfrac{d}{\mathscr{V}_0} = t_1 - t_0$ is the transit time

d = gap distance

Since the electrons under the influence of the space-charge forces exhibit simple harmonic motion, the velocity at a later time t is given by

$$\mathscr{V}_{to\ell} = \mathscr{V}_0 \left[1 + \frac{\beta_i V_1}{2V_0} \sin (\omega\tau) \cos (\omega_p t - \omega_p \tau) \right] \tag{9-3-16}$$

where ω_p = plasma frequency. The current-density equation may be obtained from Eq. (9-3-14) as

$$J = -\frac{1}{2} \frac{J_0 \omega}{V_0 \omega_q} \beta_i V_1 \sin (\beta_q z) \cos (\beta_e z - \omega t) \tag{9-3-17}$$

where $\beta_q = \dfrac{\omega_q}{\mathscr{V}_0}$ is the plasma phase constant.

Example 9-3-2: Operation of a Four-Cavity Klystron

A four-cavity CW klystron amplifier VA-864 has the following parameters:

Beam voltage:	$V_0 = 18$ kV
Beam current:	$I_0 = 2.25$ A
Gap distance:	$d = 1$ cm
Operating frequency:	$f = 10$ GHz
Signal voltage:	$V_1 = 10$ V (rms)
Beam coupling coefficient:	$\beta_0 = \beta_i = 1$
dc electron beam current density:	$\rho_0 = 10^{-8}$ C/m^3

Determine:

a. The dc electron velocity
b. The dc electron phase constant
c. The plasma frequency
d. The reduced plasma frequency for $R = 0.5$
e. The reduced plasma phase constant
f. The transit time across the input gap
g. The electron velocity leaving the input gap

Solution

a. The dc electron velocity is
$$V_0 = 0.593 \times 10^6 \sqrt{18 \times 10^3} = 0.796 \times 10^8 \text{ m/s}$$

b. The dc electron phase constant is
$$\beta_e = 2\pi \times 10^{10}/(0.796 \times 10^8) = 7.89 \times 10^2 \text{ rad/m}$$

c. The plasma frequency is
$$\omega_p = [1.759 \times 10^{11} \times 10^{-8}/(8.854 \times 10^{-12})]^{1/2} = 1.41 \times 10^7 \text{ rad/s}$$

d. The reduced plasma frequency is
$$\omega_q = 0.5 \times 1.41 \times 10^7 = 0.705 \times 10^7 \text{ rad/s}$$

e. The reduced plasma phase constant is
$$\beta_q = 0.705 \times 10^7/(0.796 \times 10^8) = 0.088 \text{ rad/m}$$

f. The transit time across the gap is
$$\tau = 10^{-2}/(0.796 \times 10^8) = 0.1256 \text{ ns}$$

g. The electron velocity leaving the input gap is
$$V(t_1) = 0.796 \times 10^8[1 + 1 \times 10/(2 \times 18 \times 10^3) \sin (2\pi \times 10^{10} \times 1.256$$
$$\times 10^{-10})]$$
$$= 0.796 \times 10^8 + 2.21 \times 10^4 \text{ m/s}$$

9-3-2 Output Current and Output Power of Two-Cavity Klystron

If the two cavities of a two-cavity klystron amplifier are assumed to be identical and are placed at the point where the RF current modulation is a maximum, the magnitude of the RF convection current at the output cavity for a two-cavity klystron can be written from Eq. (9-3-17) as

$$|i_2| = \frac{1}{2} \frac{I_0 \omega}{V_0 \omega_q} \beta_i |V_1| \qquad (9\text{-}3\text{-}18)$$

where V_1 = magnitude of the input signal voltage. Then the magnitudes of the induced current and voltage in the output cavity are equal to

$$|I_2| = \beta_0 |i_2| = \frac{1}{2} \frac{I_0 \omega}{V_0 \omega_q} \beta_0^2 |V_1| \qquad (9\text{-}3\text{-}19)$$

and

$$|V_2| = |I_2| R_{shl} = \frac{1}{2} \frac{I_0 \omega}{V_0 \omega_q} \beta_0^2 |V_1| R_{shl} \qquad (9\text{-}3\text{-}20)$$

where $\beta_0 = \beta_i$ is the beam coupling coefficient

R_{shl} = total shunt resistance of the output cavity in a two-cavity klystron amplifier including the external load

The output power delivered to the load in a two-cavity klystron amplifier is given by

$$P_{\text{out}} = |I_2|^2 R_{shl} = \frac{1}{4}\left(\frac{I_0\omega}{V_0\omega_q}\right)^2 \beta_0^4 |V_1|^2 R_{shl} \tag{9-3-21}$$

The power gain of a two-cavity klystron amplifier is then expressed by

$$\text{Power gain} = \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{P_{\text{out}}}{|V_1|^2/R_{sh}} = \frac{1}{4}\left(\frac{I_0\omega}{V_0\omega_q}\right)^2 \beta_0^4 R_{sh} \cdot R_{shl} \tag{9-3-22}$$

where R_{sh} = total shunt resistance of the input cavity. The electronic efficiency of a two-cavity klystron amplifier is

$$\eta = \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{P_{\text{out}}}{I_0 V_0} = \frac{1}{4}\left(\frac{I_0}{V_0}\right)\left(\frac{|V_1|\omega}{V_0\omega_q}\right)^2 \beta_0^4 R_{shl} \tag{9-3-23}$$

Example 9-3-3: Characteristics of Two-Cavity Klystron

A two-cavity klystron has the following parameters:

Beam voltage:	$V_0 = 20$ kV
Beam current:	$I_0 = 2$ A
Operating frequency:	$f = 8$ GHz
Beam coupling coefficient:	$\beta_i = \beta_0 = 1$
dc electron beam current density:	$\rho_0 = 10^{-6}$ C/m^3
Signal voltage:	$V_1 = 10$ V (rms)
Shunt resistance of the cavity:	$R_{sh} = 10$ kΩ
Total shunt resistance including load:	$R = 30$ kΩ

Calculate:

  a. The plasma frequency
  b. The reduced plasma frequency for $R = 0.5$
  c. The induced current in the output cavity
  d. The induced voltage in the output cavity
  e. The output power delivered to the load
  f. The power gain
  g. The electronic efficiency

Solution

  a. The plasma frequency is

$$\omega_p = [1.759 \times 10^{11} \times 10^{-6}/(8.854 \times 10^{-12})]^{1/2} = 1.41 \times 10^8 \text{ rad/s}$$

b. The reduced plasma frequency is

$$\omega_q = 0.5 \times 1.41 \times 10^8 = 0.705 \times 10^8 \text{ rad/s}$$

$$\omega/\omega_q = 2\pi \times 8 \times 10^9/(0.705 \times 10^8) = 713$$

c. The induced current in the output cavity is

$$|I_2| = \frac{1}{2} \frac{2}{20 \times 10^3} \times 713 \times 1^2 \times |10| = 0.3565 \text{ A}$$

d. The induced voltage in the output cavity is

$$|V_2| = |I_2| R_{shl} = 0.357 \times 30 \times 10^3 = 10.71 \text{ kV}$$

e. The output power delivered to the load is

$$P_{out} = |I_2|^2 R_{shl} = 0.357^2 \times 30 \times 10^3 = 3.82 \text{ kW}$$

f. The power gain is

$$\text{Gain} = 1/4[2/(20 \times 10^3) \times 713]^2 \times 1^4 \times 10 \times 10^3 \times 30 \times 10^3$$

$$= 3.83 \times 10^5 = 55.8 \text{ dB}$$

g. The electronic efficiency is

$$\eta = \frac{P_{out}}{P_{in}} = \frac{3.82 \times 10^3}{2 \times 20 \times 10^3} = 9.6\%$$

9-3-3 Output Power of Four-Cavity Klystron

High power may be obtained by adding additional intermediate cavities in a two-cavity klystron. Multicavity klystrons with as many as seven cavities are commercially available, although the most frequently used number of cavities is four. Each of the intermediate cavities functions in the same manner as in the two-cavity klystron amplifier.

Let us carry out a simplified analysis of the four-cavity klystron amplifier. The four cavities are assumed to be identical and they have same unloaded Q and beam coupling coefficient $\beta_i = \beta_0$. The two intermediate cavities are not externally loaded, but the input and output cavities are matched to their transmission lines. If V_1 is the magnitude of the input cavity-gap voltage, the magnitude of the RF convection current density injected into the first intermediate cavity gap is given by Eq. (9-3-17). The induced current and voltage in the first intermediate cavity are given by Eqs. (9-3-19) and (9-3-20). This gap voltage of the first intermediate cavity produces a velocity modulation on the beam in the second intermediate cavity. The RF convection current in the second intermediate cavity can be written with $|V_1|$ replaced by $|V_2|$. That is,

$$|i_3| = \frac{1}{2} \frac{I_0 \omega}{V_0 \omega_q} \beta_0 |V_2|$$

$$= \frac{1}{4} \left(\frac{I_0 \omega}{V_0 \omega_q} \right)^2 \beta_0^3 |V_1| R_{sh}$$

(9-3-24)

The output voltage of the second intermediate cavity is then given by

$$|V_3| = \beta_0 |i_3| R_{sh}$$
$$= \frac{1}{4} \left(\frac{I_0 \omega}{V_0 \omega_q} \right)^2 \beta_0^4 |V_1| R_{sh}^2 \tag{9-3-25}$$

This voltage produces a velocity modulation again and is converted into an RF convection current at the output cavity for four-cavity klystron as

$$|i_4| = \frac{1}{2} \frac{I_0 \omega}{V_0 \omega_q} \beta_i |V_3|$$
$$= \frac{1}{8} \left(\frac{I_0 \omega}{V_0 \omega_q} \right)^3 \beta_0^5 |V_1| R_{sh}^2 \tag{9-3-26}$$

and

$$|I_4| = \beta_0 |i_4| = \frac{1}{8} \left(\frac{I_0 \omega}{V_0 \omega_q} \right)^3 \beta_0^6 |V_1| R_{sh}^2 \tag{9-3-27}$$

The output voltage is then

$$|V_4| = |I_4| R_{shl} = \frac{1}{8} \left(\frac{I_0 \omega}{V_0 \omega_q} \right)^3 \beta_0^6 |V_1| R_{sh}^2 R_{shl} \tag{9-3-28}$$

The output power from the output cavity in a four-cavity klystron amplifier can be expressed as

$$P_{out} = |I_4|^2 R_{shl} = \frac{1}{64} \left(\frac{I_0 \omega}{V_0 \omega_q} \right)^6 \beta_0^{12} |V_1|^2 R_{sh}^4 R_{shl}^2 \tag{9-3-29}$$

where R_{shl} = total shunt resistance of the output cavity including the external load.

The multicavity klystrons are often operated with their cavities stagger-tuned so as to obtain a greater bandwidth at some reduction in gain. In high-power klystrons, the cavity grids are omitted, because they would burn up during beam interception. High-power klystron amplifiers with a power gain of 40 to 50 dB and a bandwidth of several percent are commercially available.

Example 9-3-4:   Output Power of Four-Cavity Klystron

A four-cavity klystron has the following parameters:

Beam voltage:	$V_0 = 10$ kV
Beam current:	$I_0 = 0.7$ A
Operating frequency:	$f = 4$ GHz
Beam coupling coefficient:	$\beta_i = \beta_0 = 1$
dc electron beam current density:	$\rho_0 = 5 \times 10^{-5}$ C/m^3
Signal voltage:	$V_1 = 2$ V (rms)
Shunt resistance of cavity:	$R_{sh} = 10$ kΩ
Total shunt resistance including load:	$R_{shl} = 5$ kΩ

Determine:

  a. The plasma frequency
  b. The reduced plasma frequency for $R = 0.6$
  c. The induced current in the output cavity
  d. The induced voltage in the output cavity
  e. The output power delivered to the load

Solution

  a. The plasma frequency is

$$\omega_p = \left[1.759 \times 10^{11} \times \frac{5 \times 10^{-5}}{8.854 \times 10^{-12}}\right]^{1/2} = 0.997 \times 10^9 \text{ rad/s}$$

  b. The reduced plasma frequency is

$$\omega_q = 0.6 \times 0.997 \times 10^9 = 0.598 \times 10^9 \text{ rad/s}$$

$$\omega/\omega_q = 2\pi \times 4 \times 10^9/(0.598 \times 10^9) = 42.03$$

  c. The induced current in the output cavity is

$$|I_4| = \frac{1}{8}\left(\frac{0.7}{10^4} \times 42.03\right)^3 \times 1^6 \times |2| \times (10 \times 10^3)^2$$

$$= 0.6365 \text{ A}$$

  d. The induced voltage in the output cavity is

$$|V_4| = |I_4|R_{shl} = |0.6365| \times 5 \times 10^3$$

$$= 3.18 \text{ kV}$$

  e. The output power is

$$P_{\text{out}} = |I_4|^2 R_{shl} = |0.6365|^2 \times 5 \times 10^3 = 2.03 \text{ kW}$$

9-4 REFLEX KLYSTRONS

If a fraction of the output power is fed back to the input cavity and if the loop gain has a magnitude of unity with a phase shift of multiple 2π, the klystron will oscillate. However, a two-cavity klystron oscillator is usually not constructed because, when the oscillation frequency is varied, the resonant frequency of each cavity and the feedback path phase shift must be readjusted for a positive feedback. The reflex klystron is a single-cavity klystron that overcomes the disadvantages of the two-cavity klystron oscillator. It is a low-power generator of 10 to 500-mW output at a frequency range of 1 to 25 GHz. The efficiency is about 20 to 30%. This type is widely used in the laboratory for microwave measurements and in microwave receivers as local oscillators in commercial, military, and airborne Doppler radars as well as missiles. The theory of the two-cavity klystron can be applied to the analysis of

the reflex klystron with slight modification. A schematic diagram of the reflex klystron is shown in Fig. 9-4-1.

The electron beam injected from the cathode is first velocity-modulated by the cavity-gap voltage. Some electrons accelerated by the accelerating field enter the repeller space with greater velocity than those with unchanged velocity. Some electrons decelerated by the retarding field enter the repeller region with less velocity. All electrons turned around by the repeller voltage then pass through the cavity gap in bunches that occur once per cycle. On their return journey the bunched electrons pass through the gap during the retarding phase of the alternating field and give up their kinetic energy to the electromagnetic energy of the field in the cavity. Oscillator output energy is then taken from the cavity. The electrons are finally collected by the walls of the cavity or other grounded metal parts of the tube. Figure 9-4-2 shows an Applegate diagram for the $1\frac{3}{4}$ mode of a reflex klystron.

9-4-1 Velocity Modulation

The analysis of a reflex klystron is similar to that of a two-cavity klystron. For simplicity, the effect of space-charge forces on the electron motion will again be neglected. The electron entering the cavity gap from the cathode at $z = 0$ and time t_0 is assumed to have uniform velocity

$$v_0 = 0.593 \times 10^6 \sqrt{V_0} \tag{9-4-1}$$

The same electron leaves the cavity gap at $z = d$ at time t_1 with velocity

$$v(t_1) = v_0\left[1 + \frac{\beta_1 V_1}{2V_0} \sin\left(\omega t_1 - \frac{\theta_g}{2}\right)\right] \tag{9-4-2}$$

This expression is identical to Eq. (9-2-17), for the problems up to this point are identical to those of a two-cavity klystron amplifier. The same electron is forced back to the cavity $z = d$ and time t_2 by the retarding electric field E, which is given by

$$E = \frac{V_r + V_0 + V_1 \sin(\omega t)}{L} \tag{9-4-3}$$

This retarding field E is assumed to be constant in the z direction. The force equation for one electron in the repeller region is

$$m \frac{d^2z}{dt^2} = -eE = -e\frac{V_r + V_0}{L} \tag{9-4-4}$$

where $\mathbf{E} = -\nabla V$ is used in the z direction only, V_r is the magnitude of the repeller voltage, and $|V_1 \sin \omega t| \ll (V_r + V_0)$ is assumed.

Integration of Eq. (9-4-4) twice yields

$$\frac{dz}{dt} = \frac{-e(V_r + V_0)}{mL} \int_{t_1}^{t} dt = \frac{-e(V_r + V_0)}{mL}(t - t_1) + K_1 \tag{9-4-5}$$
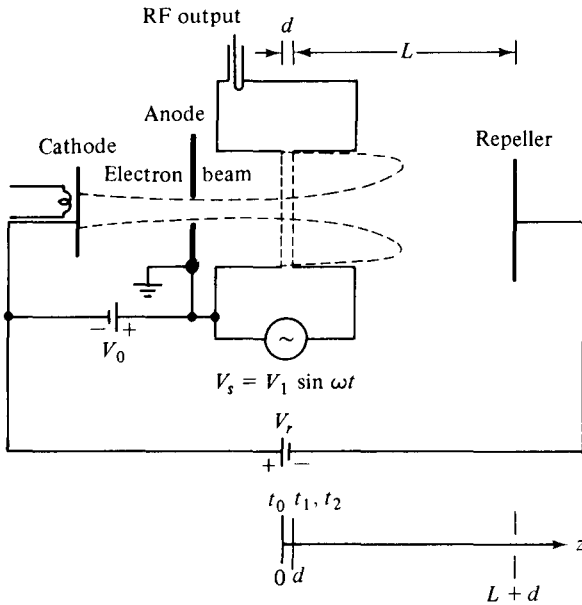
t_0 = time for electron entering cavity gap at $z = 0$

t_1 = time for same electron leaving cavity gap at $z = d$

t_2 = time for same electron returned by retarding field
$z = d$ and collected on walls of cavity

Figure 9-4-1   Schematic diagram of a reflex klystron.
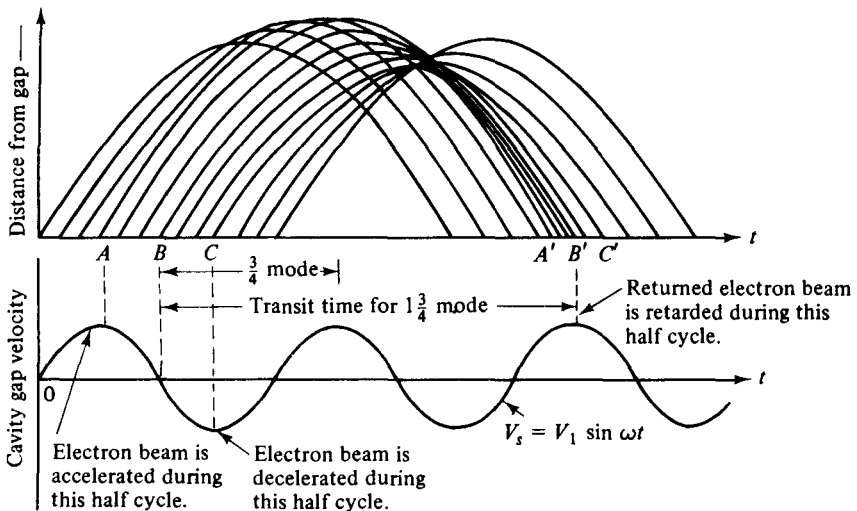
Figure 9-4-2   Applegate diagram with gap voltage for a reflex klystron.

at $t = t_1$, $dz/dt = v(t_1) = K_1$; then

$$z = \frac{-e(V_r + V_0)}{mL} \int_{t_1}^{t} (t - t_1)\, dt + v(t_1) \int_{t_1}^{t} dt$$

$$z = \frac{-e(V_r + V_0)}{2mL} (t - t_1)^2 + v(t_1)(t - t_1) + K_2$$

at $t = t_1$, $z = d = K_2$; then

$$z = \frac{-e(V_r + V_0)}{2mL} (t - t_1)^2 + v(t_1)(t - t_1) + d \qquad (9\text{-}4\text{-}6)$$

On the assumption that the electron leaves the cavity gap at $z = d$ and time t_1 with a velocity of $v(t_1)$ and returns to the gap at $z = d$ and time t_2, then, at $t = t_2$, $z = d$,

$$0 = \frac{-e(V_r + V_0)}{2mL} (t_2 - t_1)^2 + v(t_1)(t_2 - t_1)$$

The round-trip transit time in the repeller region is given by

$$T' = t_2 - t_1 = \frac{2mL}{e(V_r + V_0)} v(t_1) = T_0' \left[1 + \frac{\beta_i V_1}{2V_0} \sin \left(\omega t_1 - \frac{\theta_g}{2} \right) \right] \qquad (9\text{-}4\text{-}7)$$

where

$$T_0' = \frac{2mLv_0}{e(V_r + V_0)} \qquad (9\text{-}4\text{-}8)$$

is the round-trip dc transit time of the center-of-the-bunch electron.

Multiplication of Eq. (9-4-7) through by a radian frequency results in

$$\omega(t_2 - t_1) = \theta_0' + X' \sin \left(\omega t_1 - \frac{\theta_g}{2} \right) \qquad (9\text{-}4\text{-}9)$$

where

$$\theta_0' = \omega T_0' \qquad (9\text{-}4\text{-}10)$$

is the round-trip dc transit angle of the center-of-the-bunch electron and

$$X' \equiv \frac{\beta_i V_1}{2V_0} \theta_0' \qquad (9\text{-}4\text{-}11)$$

is the bunching parameter of the reflex klystron oscillator.

9-4-2 Power Output and Efficiency

In order for the electron beam to generate a maximum amount of energy to the oscillation, the returning electron beam must cross the cavity gap when the gap field is maximum retarding. In this way, a maximum amount of kinetic energy can be transferred from the returning electrons to the cavity walls. It can be seen from Fig.

9-4-2 that for a maximum energy transfer, the round-trip transit angle, referring to the center of the bunch, must be given by

$$\omega(t_2 - t_1) = \omega T_0' = \left(n - \frac{1}{4}\right)2\pi = N2\pi = 2\pi n - \frac{\pi}{2} \qquad (9\text{-}4\text{-}12)$$

where $V_1 \ll V_0$ is assumed, $n =$ any positive integer for cycle number, and $N = n - \frac{1}{4}$ is the number of modes.

The current modulation of the electron beam as it reenters the cavity from the repeller region can be determined in the same manner as in Section 9-2 for a two-cavity klystron amplifier. It can be seen from Eqs. (9-2-30) and (9-4-9) that the bunching parameter X' of a reflex klystron oscillator has a negative sign with respect to the bunching parameter X of a two-cavity klystron amplifier. Furthermore, the beam current injected into the cavity gap from the repeller region flows in the negative z direction. Consequently, the beam current of a reflex klystron oscillator can be written

$$i_{2t} = -I_0 - \sum_{n=1}^{\infty} 2I_0 J_n(nX') \cos\left[n(\omega t_2 - \theta_0' - \theta_g)\right] \qquad (9\text{-}4\text{-}13)$$

The fundamental component of the current induced in the cavity by the modulated electron beam is given by

$$i_2 = -\beta_i I_2 = 2I_0\beta_i J_1(X') \cos(\omega t_2 - \theta_0') \qquad (9\text{-}4\text{-}14)$$

in which θ_g has been neglected as a small quantity compared with θ_0'. The magnitude of the fundamental component is

$$I_2 = 2I_0\beta_i J_1(X') \qquad (9\text{-}4\text{-}15)$$

The dc power supplied by the beam voltage V_0 is

$$P_{dc} = V_0 I_0 \qquad (9\text{-}4\text{-}16)$$

and the ac power delivered to the load is given by

$$P_{ac} = \frac{V_1 I_2}{2} = V_1 I_0 \beta_i J_1(X') \qquad (9\text{-}4\text{-}17)$$

From Eqs.(9-4-10), (9-4-11), and (9-4-12) the ratio of V_1 over V_0 is expressed by

$$\frac{V_1}{V_0} = \frac{2X'}{\beta_i(2\pi n - \pi/2)} \qquad (9\text{-}4\text{-}18)$$

Substitution of Eq. (9-4-18) in Eq. (9-4-17) yields the power output as

$$P_{ac} = \frac{2V_0 I_0 X' J_1(X')}{2\pi n - \pi/2} \qquad (9\text{-}4\text{-}19)$$

Therefore the electronic efficiency of a reflex klystron oscillator is defined as

$$\text{Efficiency} \equiv \frac{P_{ac}}{P_{dc}} = \frac{2X' J_1(X')}{2\pi n - \pi/2} \qquad (9\text{-}4\text{-}20)$$

The factor $X'J_1(X')$ reaches a maximum value of 1.25 at $X' = 2.408$ and $J_1(X') = 0.52$. In practice, the mode of $n = 2$ has the most power output. If $n = 2$ or $1\frac{3}{4}$ mode, the maximum electronic efficiency becomes

$$\text{Efficiency}_{max} = \frac{2(2.408)J_1(2.408)}{2\pi(2) - \pi/2} = 22.7\% \qquad (9\text{-}4\text{-}21)$$

The maximum theoretical efficiency of a reflex klystron oscillator ranges from 20 to 30%. Figure 9-4-3 shows a curve of $X'J_1(X')$ versus X'.
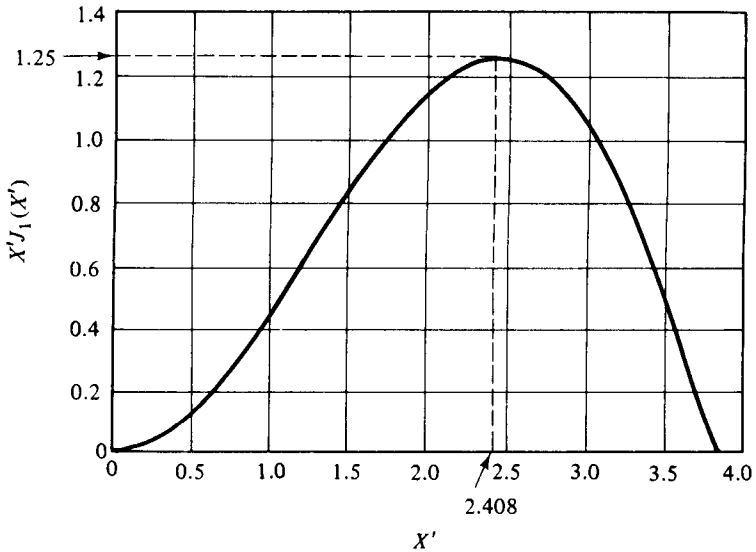
Figure 9-4-3 $X'J_1(X')$ versus X'.

For a given beam voltage V_0, the relationship between the repeller voltage and cycle number n required for oscillation is found by inserting Eqs. (9-4-12) and (9-4-1) into Eq. (9-4-8):

$$\frac{V_0}{(V_r + V_0)^2} = \frac{(2\pi n - \pi/2)^2}{8\omega^2 L^2}\frac{e}{m} \qquad (9\text{-}4\text{-}22)$$

The power output can be expressed in terms of the repeller voltage V_r. That is,

$$P_{ac} = \frac{V_0 I_0 X' J_1(X')(V_r + V_0)}{\omega L}\sqrt{\frac{e}{2mV_o}} \qquad (9\text{-}4\text{-}23)$$

It can be seen from Eq. (9-4-22) that, for a given beam voltage V_0 and cycle number n or mode number $\dot{N}$, the center repeller voltage V_r can be determined in terms of the center frequency. Then the power output at the center frequency can be calculated from Eq. (9-4-23). When the frequency varies from the center frequency and the repeller voltage about the center voltage, the power output will vary accordingly, assuming a bell shape (see Fig. 9-4-4).
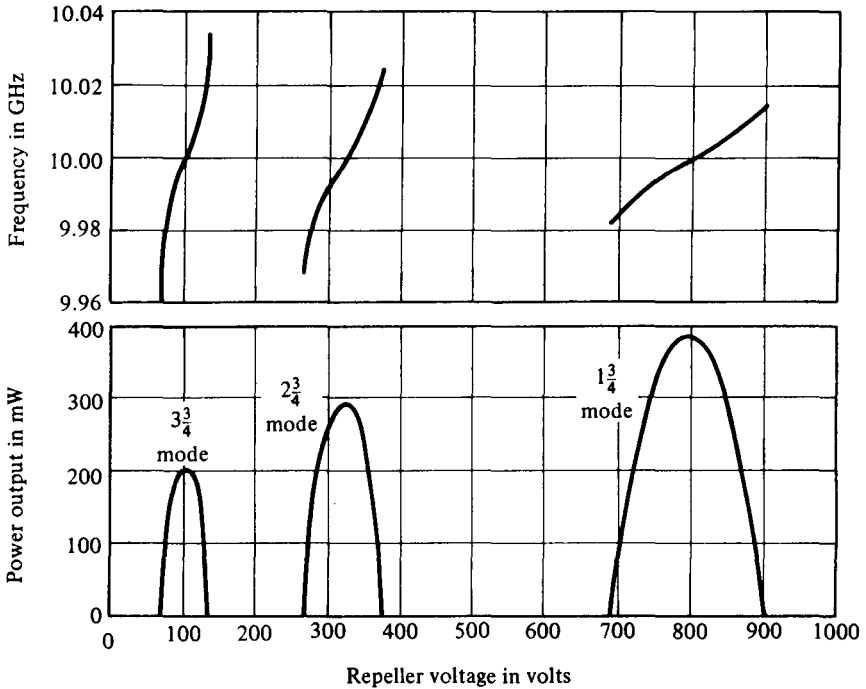
Figure 9-4-4   Power output and frequency characteristics of a reflex klystron.

9-4-3 Electronic Admittance

From Eq. (9-4-14) the induced current can be written in phasor form as

$$i_2 = 2I_0\beta_i J_1(X')e^{-j\theta_0'} \tag{9-4-24}$$

The voltage across the gap at time t_2 can also be written in phasor form:

$$V_2 = V_1 e^{-j\pi/2} \tag{9-4-25}$$

The ratio of i_2 to V_2 is defined as the electronic admittance of the reflex klystron. That is,

$$Y_e = \frac{I_0}{V_0} \frac{\beta_i^2 \theta_0'}{2} \frac{2J_1(X')}{X'} e^{j(\pi/2 - \theta_0')} \tag{9-4-26}$$

The amplitude of the phasor admittance indicates that the electronic admittance is a function of the dc beam admittance, the dc transit angle, and the second transit of the electron beam through the cavity gap. It is evident that the electronic admittance is nonlinear, since it is proportional to the factor $2J_1(X')/X'$, and X' is proportional to the signal voltage. This factor of proportionality is shown in Fig. 9-4-5. When the signal voltage goes to zero, the factor approaches unity.
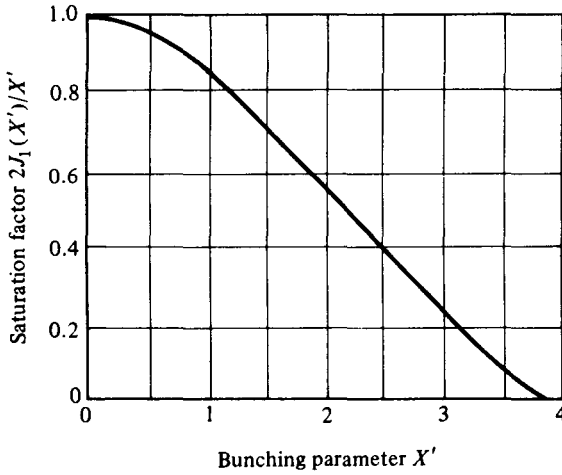
Figure 9-4-5 Reflex klystron saturation factor.

The equivalent ciruit of a reflex klystron is shown in Fig. 9-4-6. In this circuit L and C are the energy storage elements of the cavity; G_c represents the copper losses of the cavity, G_b the beam loading conductance, and G_ℓ the load conductance.

The necessary condition for oscillations is that the magnitude of the negative real part of the electronic admittance as given by Eq. (9-4-26) not be less than the total conductance of the cavity circuit. That is,

$$|-G_e| \geq G \qquad (9\text{-}4\text{-}27)$$

where $G = G_c + G_b + G_\ell = 1/R_{sh}$ and R_{sh} is the effective shunt resistance.
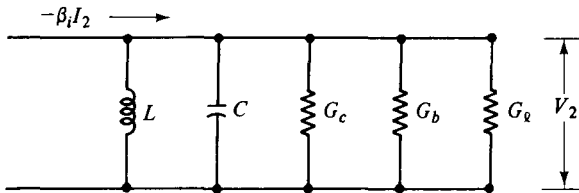
Figure 9-4-6 Equivalent circuit of a reflex klystron.

Equation (9-4-26) can be rewritten in rectangular form:

$$Y_e = G_e + jB_e \qquad (9\text{-}4\text{-}28)$$

Since the electronic admittance shown in Eq. (9-4-26) is in exponential form, its phase is $\pi/2$ when θ_0' is zero. The rectangular plot of the electron admittance Y_e is a spiral (see Fig. 9-4-7). Any value of θ_0' for which the spiral lies in the area to the left of line $(-G - jB)$ will yield oscillation. That is,

$$\theta_0' = \left(n - \frac{1}{4}\right)2\pi = N2\pi \qquad (9\text{-}4\text{-}29)$$

where N is the mode number as indicated in the plot, the phenomenon verifies the early analysis.
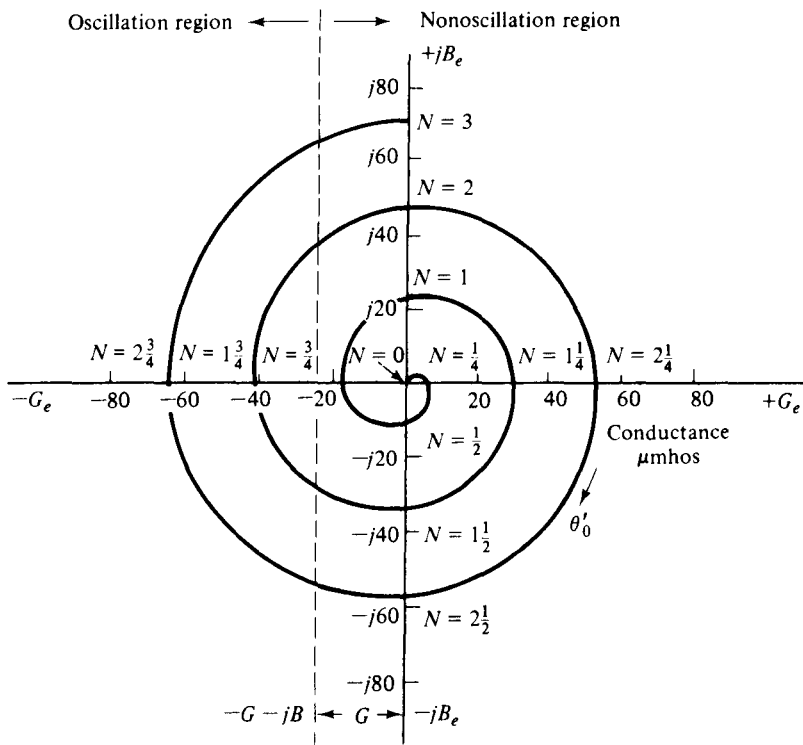
Figure 9-4-7 Electronic admittance spiral of a reflex klystron.

Example 9-4-1: Reflex Klystron

A reflex klystron operates under the following conditions:

$$V_0 = 600 \text{ V} \qquad L = 1 \text{ mm}$$

$$R_{sh} = 15\text{k}\Omega \qquad \frac{e}{m} = 1.759 \times 10^{11} \text{ (MKS system)}$$

$$f_r = 9 \text{ GHz}$$

The tube is oscillating at f_r at the peak of the $n = 2$ mode or $1\frac{3}{4}$ mode. Assume that the transit time through the gap and beam loading can be neglected.

a. Find the value of the repeller voltage V_r.

b. Find the direct current necessary to give a microwave gap voltage of 200 V.

c. What is the electronic efficiency under this condition?

Solution

a. From Eq. (9-4-22) we obtain

$$\frac{V_0}{(V_r + V_0)^2} = \left(\frac{e}{m}\right)\frac{(2\pi n - \pi/2)^2}{8\omega^2 L^2}$$

$$= (1.759 \times 10^{11})\frac{(2\pi 2 - \pi/2)^2}{8(2\pi \times 9 \times 10^9)^2(10^{-3})^2} = 0.832 \times 10^{-3}$$

381

$$(V_r + V_0)^2 = \frac{600}{0.832 \times 10^{-3}} = 0.721 \times 10^6$$

$$V_r = 250 \text{ V}$$

b. Assume that $\beta_0 = 1$. Since

$$V_2 = I_2 R_{sh} = 2 I_0 J_1(X')R_{sh}$$

the direct current I_0 is

$$I_0 = \frac{V_2}{2J_1(X')R_{sh}} = \frac{200}{2(0.582)(15 \times 10^3)} = 11.45 \text{ mA}$$

c. From Eqs. (9-4-11), (9-4-12), and (9-4-20) the electronic efficiency is

$$\text{Efficiency} = \frac{2X'J_1(X')}{2\pi n - \pi/2} = \frac{2(1.841)(0.582)}{2\pi(2) - \pi/2} = 19.49\%$$

9-5 HELIX TRAVELING-WAVE TUBES (TWTs)

Since Kompfner invented the helix traveling-wave tube (TWT) in 1944 [11], its basic circuit has changed little. For broadband applications, the helix TWTs are almost exclusively used, whereas for high-average-power purposes, such as radar transmitters, the coupled-cavity TWTs are commonly used.

In previous sections klystrons and reflex klystrons were analyzed in some detail. Before starting to describe the TWT, it seems appropriate to compare the basic operating principles of both the TWT and the klystron. In the case of the TWT, the microwave circuit is nonresonant and the wave propagates with the same speed as the electrons in the beam. The initial effect on the beam is a small amount of velocity modulation caused by the weak electric fields associated with the traveling wave. Just as in the klystron, this velocity modulation later translates to current modulation, which then induces an RF current in the circuit, causing amplification. However, there are some major differences between the TWT and the klystron:

1. The interaction of electron beam and RF field in the TWT is continuous over the entire length of the circuit, but the interaction in the klystron occurs only at the gaps of a few resonant cavities.
2. The wave in the TWT is a propagating wave; the wave in the klystron is not.
3. In the coupled-cavity TWT there is a coupling effect between the cavities, whereas each cavity in the klystron operates independently.

A helix traveling-wave tube consists of an electron beam and a slow-wave structure. The electron beam is focused by a constant magnetic field along the electron beam and the slow-wave structure. This is termed an O-type traveling-wave tube. The slow-wave structure is either the helical type or folded-back line. The applied signal propagates around the turns of the helix and produces an electric field at the

center of the helix, directed along the helix axis. The axial electric field progresses with a velocity that is very close to the velocity of light multiplied by the ratio of helix pitch to helix circumference. When the electrons enter the helix tube, an interaction takes place between the moving axial electric field and the moving electrons. On the average, the electrons transfer energy to the wave on the helix. This interaction causes the signal wave on the helix to become larger. The electrons entering the helix at zero field are not affected by the signal wave; those electrons entering the helix at the accelerating field are accelerated, and those at the retarding field are decelerated. As the electrons travel further along the helix, they bunch at the collector end. The bunching shifts the phase by $\pi/2$. Each electron in the bunch encounters a stronger retarding field. Then the microwave energy of the electrons is delivered by the electron bunch to the wave on the helix. The amplification of the signal wave is accomplished. The characteristics of the traveling-wave tube are:

Frequency range: 3 GHz and higher
Bandwidth: about 0.8 GHz
Efficiency: 20 to 40%
Power output: up to 10 kW average
Power gain: up to 60 dB

The present state of the art for U.S. high-power TWTs is shown in Fig. 9-5-1.
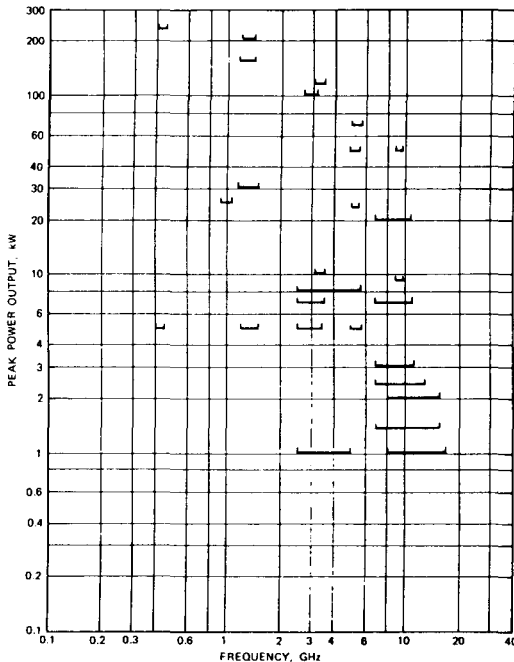
Figure 9-5-1   State of the art for U.S. high-power TWTs.

9-5-1 Slow-Wave Structures

As the operating frequency is increased, both the inductance and capacitance of the resonant circuit must be decreased in order to maintain resonance at the operating frequency. Because the gain-bandwidth product is limited by the resonant circuit, the ordinary resonator cannot generate a large output. Several nonresonant periodic circuits or slow-wave structures (see Fig. 9-5-2) are designed for producing large gain over a wide bandwidth.

(a)

(b)

(c)

(d)

(e)
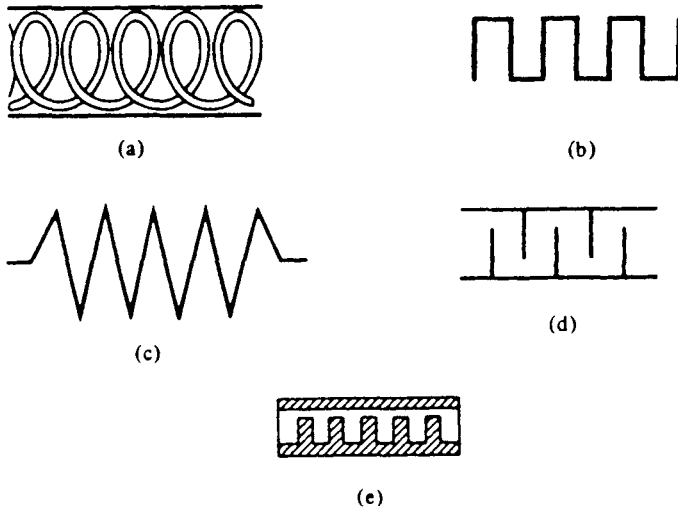
Figure 9-5-2 Slow-wave structures. (a) Helical line. (b) Folded-back line. (c) Zigzag line. (d) Interdigital line. (e) Corrugated waveguide.

Slow-wave structures are special circuits that are used in microwave tubes to reduce the wave velocity in a certain direction so that the electron beam and the signal wave can interact. The phase velocity of a wave in ordinary waveguides is greater than the velocity of light in a vacuum. In the operation of traveling-wave and magnetron-type devices, the electron beam must keep in step with the microwave signal. Since the electron beam can be accelerated only to velocities that are about a fraction of the velocity of light, a slow-wave structure must be incorporated in the microwave devices so that the phase velocity of the microwave signal can keep pace with that of the electron beam for effective interactions. Several types of slow-wave structures are shown in Fig. 9-5-2.

The commonly used slow-wave structure is a helical coil with a concentric conducting cylinder (see Fig. 9-5-3).

It can be shown that the ratio of the phase velocity v_p along the pitch to the phase velocity along the coil is given by

$$\frac{v_p}{c} = \frac{p}{\sqrt{p^2 + (\pi d)^2}} = \sin \psi \qquad (9\text{-}5\text{-}1)$$
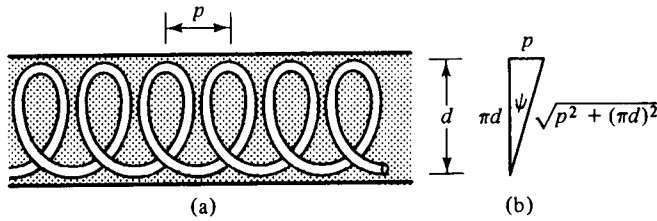
Figure 9-5-3 Helical slow-wave structure. (a) Helical coil. (b) One turn of helix.

where $c = 3 \times 10^8$ m/s is the velocity of light in free space

p = helix pitch

d = diameter of the helix

ψ = pitch angle

In general, the helical coil may be within a dielectric-filled cylinder. The phase velocity in the axial direction is expressed as

$$v_{p\epsilon} = \frac{p}{\sqrt{\mu\epsilon[p^2 + (\pi d)^2]}} \qquad (9\text{-}5\text{-}2)$$

If the dielectric constant is too large, however, the slow-wave structure may introduce considerable loss to the microwave devices, thereby reducing their efficiency. For a very small pitch angle, the phase velocity along the coil in free space is approximately represented by

$$v_p \approx \frac{pc}{\pi d} = \frac{\omega}{\beta} \qquad (9\text{-}5\text{-}3)$$

Figure 9-5-4 shows the ω-β (or Brillouin) diagram for a helical slow-wave structure. The helix ω-β diagram is very useful in designing a helix slow-wave structure. Once β is found, v_p can be computed from Eq. (9-5-3) for a given dimension of the helix. Furthermore, the group velocity of the wave is merely the slope of the curve as given by

$$v_{\text{gr}} = \frac{\partial \omega}{\partial \beta} \qquad (9\text{-}5\text{-}4)$$
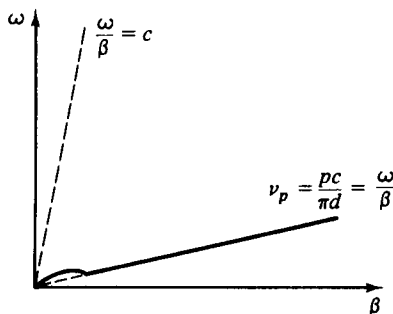
Figure 9-5-4 ω-β diagram for a helical structure.

In order for a circuit to be a slow-wave structure, it must have the property of periodicity in the axial direction. The phase velocity of some of the spatial harmonics in the axial direction obtained by Fourier analysis of the waveguide field may be smaller than the velocity of light. In the helical slow-wave structure a translation back or forth through a distance of one pitch length results in identically the same structure again. Thus the period of a helical slow-wave structure is its pitch.

In general, the field of the slow-wave structure must be distributed according to Floquet's theorem for periodic boundaries. Floquet's periodicity theorem states that:

> The steady-state solutions for the electromagnetic fields of a single propagating mode in a periodic structure have the property that fields in adjacent cells are related by a complex constant.

Mathematically, the theorem can be stated

$$E(x, y, z - L) = E(x, y, z)e^{j\beta_0 L} \tag{9-5-5}$$

where $E(x, y, z)$ is a periodic function of z with period L. Since β_0 is the phase constant in the axial direction, in a slow-wave structure β_0 is the phase constant of average electron velocity.

It is postulated that the solution to Maxwell's equations in a periodic structure can be written

$$E(x, y, z) = f(x, y, z)e^{-j\beta_0 z} \tag{9-5-6}$$

where $f(x, y, z)$ is a periodic function of z with period L that is the period of the slow-wave structure.

For a periodic structure, Eq. (9-5-6) can be rewritten with z replaced by $z - L$:

$$E(x, y, z - L) = f(x, y, z - L)e^{-j\beta_0(z-L)} \tag{9-5-7}$$

Since $f(x, y, z - L)$ is a periodic function with period L, then

$$f(x, y, z - L) = f(x, y, z) \tag{9-5-8}$$

Substitution of Eq. (9-5-8) in (9-5-7) results in

$$E(x, y, z - L) = f(x, y, z)e^{-j\beta_0 z}e^{j\beta_0 L} \tag{9-5-9}$$

and substitution of Eq. (9-5-6) in (9-5-9) gives

$$E(x, y, z - L) = E(x, y, z)e^{j\beta_0 L} \tag{9-5-10}$$

This expression is the mathematical statement of Floquet's theorem, Eq. (9-5-5). Therefore Eq. (9-5-6) does indeed satisfy Floquet's theorem.

From the theory of Fourier series, any function that is periodic, single-valued, finite, and continuous may be represented by a Fourier series. Hence the field distribution function $E(x, y, z)$ may be expanded into a Fourier series of fundamental period L as

$$E(x, y, z) = \sum_{n=-\infty}^{\infty} E_n(x, y)e^{-j(2\pi n/L)z}e^{-j\beta_0 z} = \sum_{n=-\infty}^{\infty} E_n(x, y)e^{-j\beta_n z} \tag{9-5-11}$$

where

$$E_n(x, y) = \frac{1}{L} \int_0^L E(x, y, z)e^{j(2\pi n/L)z} \, dz \qquad (9\text{-}5\text{-}12)$$

are the amplitudes of n harmonics and

$$\beta_n = B_0 + \frac{2\pi n}{L} \qquad (4\text{-}5\text{-}13)$$

is the phase constant of the nth modes, where $n = -\infty, \ldots, -2, -1, 0, 1, 2, 3, \ldots, \infty$.

The quantities $E_n(x, y)e^{-j\beta_n z}$ are known as spatial harmonics by analogy with time-domain Fourier series. The question is whether Eq. (9-5-11) can satisfy the electric wave equation, Eq. (2-1-20). Substitution of Eq. (9-5-11) into the wave equation results in

$$\nabla^2\left[\sum_{n=-\infty}^{\infty} E_n(x, y)e^{-j\beta_n z}\right] - \gamma^2\left[\sum_{n=-\infty}^{\infty} E_n(x, y)e^{-j\beta_n z}\right] = 0 \qquad (9\text{-}5\text{-}14)$$

Since the wave equation is linear, Eq. (9-5-14) can be rewritten as

$$\sum_{n=-\infty}^{\infty} [\nabla^2 E_n(x, y)e^{-j\beta_n z} - \gamma^2 E_n(x, y)e^{-j\beta_n z)}] = 0 \qquad (9\text{-}5\text{-}15)$$

It is evident from the preceding equation that if each spatial harmonic is itself a solution of the wave equation for each value of n, the summation of space harmonics also satisfies the wave equation of Eq. (9-5-14). This means that only the complete solution of Eq. (9-5-14) can satisfy the boundary conditions of a periodic structure.

Furthermore, Eq. (9-5-11) shows that the field in a periodic structure can be expanded as an infinite series of waves, all at the same frequency but with different phase velocities v_{pn}. That is

$$v_{pn} = \frac{\omega}{\beta_n} \equiv \frac{\omega}{\beta_0 + (2\pi n/L)} \qquad (9\text{-}5\text{-}16)$$

The group velocity v_{gr}, defined by $v_{gr} = \partial\omega/\partial\beta$, is then given as

$$v_{gr} = \left[\frac{d(\beta_0 + 2\pi n/L)}{d\omega}\right]^{-1} = \frac{\partial\omega}{\partial\beta_0} \qquad (9\text{-}5\text{-}17)$$

which is independent of n.

It is important to note that the phase velocity v_{pn} in the axial direction decreases for higher values of positive n and β_0. So it appears possible for a microwave of suitable n to have a phase velocity less than the velocity of light. It follows that interactions between the electron beam and microwave signal are possible and thus the amplification of active microwave devices can be achieved.

Figure 9-5-5 shows the ω-β (or Brillouin) diagram for a helix with several spatial harmonics. This ω-β diagram demonstrates some important properties needing more explanation. First, the second quadrant of the ω-β diagram indicates the negative phase velocity that corresponds to the negative n. This means that the electron beam moves in the positive z direction while the beam velocity coincides with
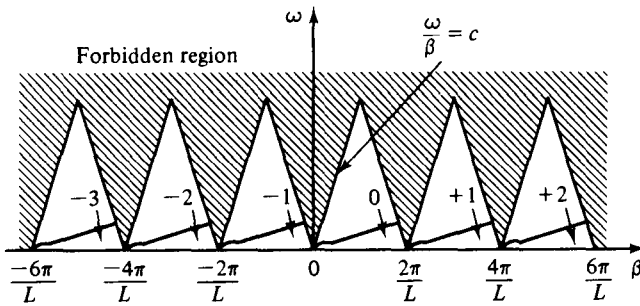
Figure 9-5-5 ω-β diagram of spatial harmonics for helical structure.

the negative spatial harmonic's phase velocity. This type of tube is called a *backward-wave oscillator*. Second, the shaded areas are the forbidden regions for propagation. This situation occurs because if the axial phase velocity of any spatial harmonic exceeds the velocity of light, the structure radiates energy. This property has been verified by experiments [10].

9-5-2 Amplification Process

A schematic diagram of a helix-type traveling-wave tube is shown in Fig. 9-5-6.

The slow-wave structure of the helix is characterized by the Brillouin diagram shown in Fig. 9-5-5. The phase shift per period of the fundamental wave on the structure is given by

$$\theta_1 = \beta_0 L \qquad (9\text{-}5\text{-}18)$$

where $\beta_0 = \omega/v_0$ is the phase constant of the average beam velocity and L is the period or pitch.

Since the dc transit time of an electron is given by

$$T_0 = \frac{L}{v_o} \qquad (9\text{-}5\text{-}19)$$

the phase constant of the nth space harmonic is

$$\beta_n = \frac{\omega}{v_0} = \frac{\theta_1 + 2\pi n}{v_0 T_0} = \beta_0 + \frac{2\pi n}{L} \qquad (9\text{-}5\text{-}20)$$

In Eq. (9-5-20) the axial space-harmonic phase velocity is assumed to be synchronized with the beam velocity for possible interactions between the electron beam and electric field. That is,

$$v_{np} = v_0 \qquad (9\text{-}5\text{-}21)$$

Equation (9-5-20) is identical to Eq. (9-5-13). In practice, the dc velocity of the electrons is adjusted to be slightly greater than the axial velocity of the electromagnetic wave for energy transfer. When a signal voltage is coupled into the helix, the axial electric field exerts a force on the electrons as a result of the following relationships:

$$\mathbf{F} = -e\mathbf{E} \qquad \text{and} \qquad \mathbf{E} = -\nabla V \qquad (9\text{-}5\text{-}22)$$
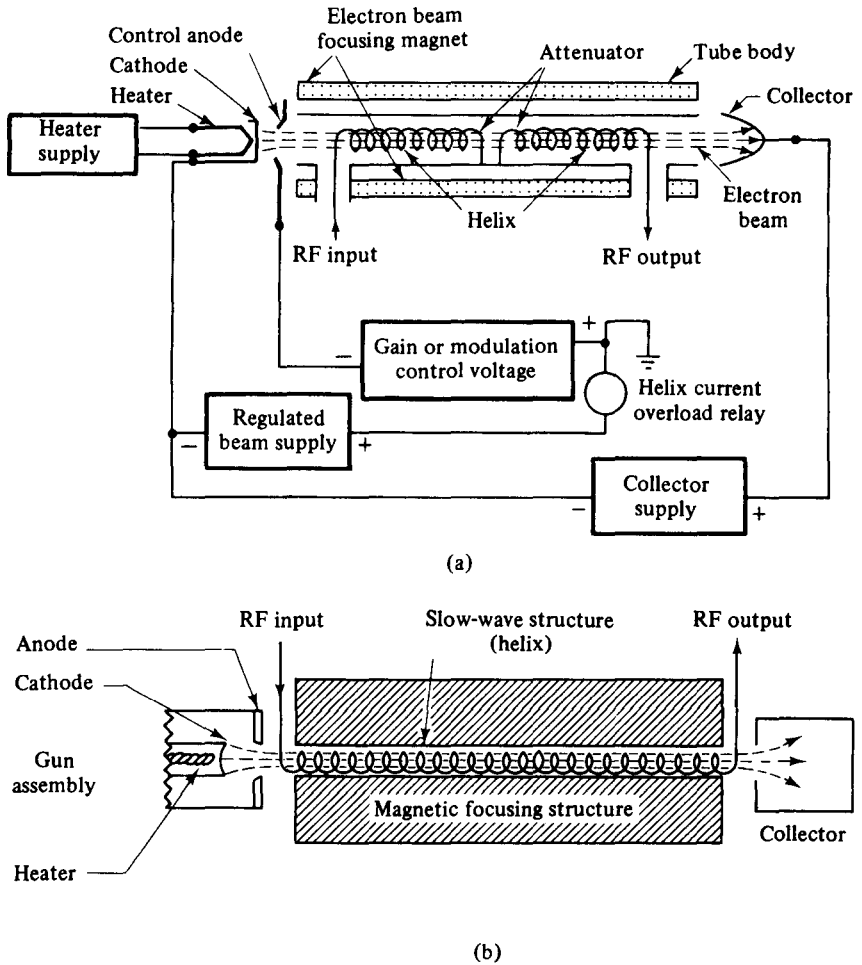
Figure 9-5-6   Diagram of helix traveling-wave tube: (a) schematic diagram of helix traveling-wave tube; (b) simplified circuit.

The electrons entering the retarding field are decelerated and those in the accelerating field are accelerated. They begin forming a bunch centered about those electrons that enter the helix during the zero field. This process is shown in Fig. 9-5-7.

Since the dc velocity of the electrons is slightly greater than the axial wave velocity, more electrons are in the retarding field than in the accelerating field, and a great amount of energy is transferred from the beam to the electromagnetic field. The microwave signal voltage is, in turn, amplified by the amplified field. The bunch continues to become more compact, and a larger amplification of the signal voltage occurs at the end of the helix. The magnet produces an axial magnetic field to prevent spreading of the electron beam as it travels down the tube. An attenuator placed near the center of the helix reduces all the waves traveling along the helix to nearly zero so that the reflected waves from the mismatched loads can be prevented from reaching the input and causing oscillation. The bunched electrons emerging
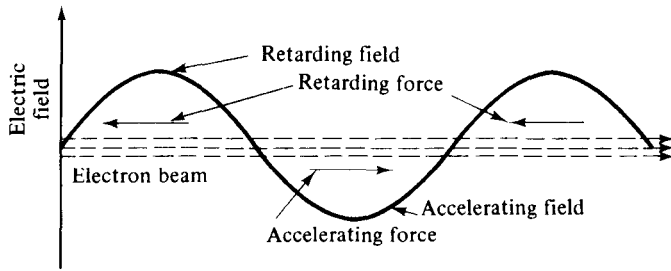
Figure 9-5-7 Interactions between electron beam and electric field.

from the attenuator induce a new electric field with the same frequency. This field, in turn, induces a new amplified microwave signal on the helix.

The motion of electrons in the helix-type traveling-wave tube can be quantitatively analyzed in terms of the axial electric field. If the traveling wave is propagating in the z direction, the z component of the electric field can be expressed as

$$E_z = E_1 \sin (\omega t - \beta_p z) \qquad (9\text{-}5\text{-}23)$$

where E_1 is the magnitude of the electric field in the z direction. If $t = t_0$ at $z = 0$, the electric field is assumed maximum. Note that $\beta_p = \omega/v_p$ is the axial phase constant of the microwave, and v_p is the axial phase velocity of the wave.

The equation of motion of the electron is given by

$$m \frac{dv}{dt} = -eE_1 \sin (\omega t - \beta_p z) \qquad (9\text{-}5\text{-}24)$$

Assume that the velocity of the electron is

$$v = v_0 + v_e \cos (\omega_e t + \theta_e) \qquad (9\text{-}5\text{-}25)$$

Then

$$\frac{dv}{dt} = -v_e \omega_e \sin (\omega_e t + \theta_e) \qquad (9\text{-}5\text{-}26)$$

where v_0 = dc electron velocity

v_e = magnitude of velocity fluctuation in the velocity-modulated electron beam

ω_e = angular frequency of velocity fluctuation

θ_e = phase angle of the fluctuation

Substitution of Eq. (9-5-26) in Eq. (9-5-24) yields

$$mv_e \omega_e \sin (\omega_e t + \omega_e) = eE_1 \sin (\omega t - \beta_p z) \qquad (9\text{-}5\text{-}27)$$

For interactions between the electrons and the electric field, the velocity of the velocity-modulated electron beam must be approximately equal to the dc electron velocity. This is

$$v \approx v_0 \qquad (9\text{-}5\text{-}28)$$

Hence the distance z traveled by the electrons is

$$z = v_0(t - t_0) \tag{9-5-29}$$

and

$$mv_e\omega_e \sin(\omega_e t + \theta_e) = eE_1 \sin[\omega t - \beta_p v_0(t - t_0)] \tag{9-5-30}$$

Comparison of the left and right-hand sides of Eq. (9-5-30) shows that

$$v_e = \frac{eE_1}{m\omega_e} \tag{9-5-31}$$

$$\omega_e = \beta_p(v_p - v_0)$$

$$\theta_e = \beta_p v_0 t_0$$

It can be seen that the magnitude of the velocity fluctuation of the electron beam is directly proportional to the magnitude of the axial electric field.

9-5-3 Convection Current

In order to determine the relationship between the circuit and electron beam quantities, the convection current induced in the electron beam by the axial electric field and the microwave axial field produced by the beam must first be developed. When the space-charge effect is considered, the electron velocity, the charge density, the current density, and the axial electric field will perturbate about their averages or dc values. Mathematically, these quantities can be expressed as

$$v = v_0 + v_1 e^{j\omega t - \gamma z} \tag{9-5-32}$$

$$\rho = \rho_0 + \rho_1 e^{j\omega t - \gamma z} \tag{9-5-33}$$

$$J = -J_0 + J_1 e^{j\omega t - \gamma z} \tag{9-5-34}$$

$$E_z = E_1 e^{j\omega t - \gamma z} \tag{9-5-35}$$

where $\gamma = \alpha_e + j\beta_e$ is the propagation constant of the axial waves. The minus sign is attached to J_0 so that J_0 may be a positive in the negative z direction. For a small signal, the electron beam-current density can be written

$$J = \rho v \approx -J_0 + J_1 e^{j\omega t - \gamma z} \tag{9-5-36}$$

where $-J_0 = \rho_0 v_0$, $J_1 = \rho_1 v_0 + \rho_0 v_1$, and $\rho_1 v_1 \approx 0$ have been replaced. If an axial electric field exists in the structure, it will perturbate the electron velocity according to the force equation. Hence the force equation can be written

$$\frac{dv}{dt} = -\frac{e}{m}E_1 e^{j\omega t - \gamma z} = \left(\frac{\partial}{\partial t} + \frac{dz}{dt}\frac{\partial}{\partial z}\right)v = (j\omega - \gamma v_0)v_1 e^{j\omega t - \gamma z} \tag{9-5-37}$$

where dz/dt has been replaced by v_0. Thus

$$v_1 = \frac{-e/m}{j\omega - \gamma v_0}E_1 \tag{9-5-38}$$

In accordance with the law of conservation of electric charge, the continuity equation can be written

$$\nabla \cdot J + \frac{\partial \rho}{\partial t} = (-\gamma J_1 + j\omega\rho_1)e^{j\omega t - \gamma z} = 0 \qquad (9\text{-}5\text{-}39)$$

It follows that

$$\rho_1 = -\frac{j\gamma J_1}{\omega} \qquad (9\text{-}5\text{-}40)$$

Substitution Eqs. (9-5-38) and (9-5-40) in

$$J_1 = \rho_1 v_0 + \rho_0 v_1 \qquad (9\text{-}5\text{-}44)$$

gives

$$J_1 = J\frac{\omega}{v_0}\frac{e}{m}\frac{J_0}{(j\omega - \gamma v_0)^2}E_1 \qquad (9\text{-}5\text{-}42)$$

where $-J_0 = \rho_0 v_0$ has been replaced. If the magnitude of the axial electric field is uniform over the cross-sectional area of the electron beam, the spatial ac current i will be proportional to the dc current I_0 with the same proportionality constant for J_1 and J_0. Therefore the convection current in the electron beam is given by

$$i = j\frac{\beta_e I_0}{2V_0(j\beta_e - \gamma)^2}E_1 \qquad (9\text{-}5\text{-}43)$$

where $\beta_e \equiv \omega/v_0$ is defined as the phase constant of the velocity-modulated electron beam and $v_0 = \sqrt{(2e/m)V_0}$ has been used. This equation is called the *electronic equation*, for it determines the convection current induced by the axial electric field. If the axial field and all parameters are known, the convection current can be found by means of Eq. (9-5-43).

9-5-4 Axial Electric Field

The convection current in the electron beam induces an electric field in the slow-wave circuit. This induced field adds to the field already present in the circuit and causes the circuit power to increase with distance. The coupling relationship between the electron beam and the slow-wave helix is shown in Fig. 9-5-8.
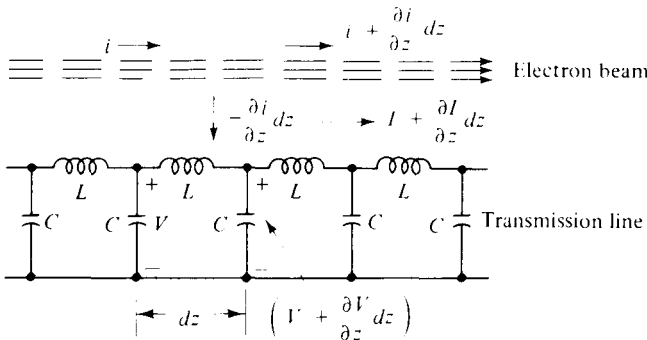
Figure 9-5-8 Electron beam coupled to equivalent circuit of a slow-wave helix.

For simplicity, the slow-wave helix is represented by a distributed lossless transmission line. The parameters are defined as follows:

L = inductance per unit length
C = capacitance per unit length
I = alternating current in transmission line
V = alternating voltage in transmission line
i = convection current

Since the transmission line is coupled to a convection-electron beam current, a current is then induced in the line. The current flowing into the left-end portion of the line of length dz is i, and the current flowing out of the right end of dz is $[i + (\partial i/\partial z)\, dz]$. Since the net change of current in the length dz must be zero, however, the current flowing out of the electron beam into the line must be $[-(\partial i/\partial z)\, dz]$. Application of transmission-line theory and Kirchhoff's current law to the electron beam results, after simplification, in

$$\frac{\partial I}{\partial z} = -C\frac{\partial V}{\partial t} - \frac{\partial i}{\partial z} \tag{9-5-44}$$

Then

$$-\gamma I = -j\omega CV + \gamma i \tag{9-5-45}$$

in which $\partial/\partial z = -\gamma$ and $\partial/\partial t = j\omega$ are replaced.

From Kirchhoff's voltage law the voltage equation, after simplification, is

$$\frac{\partial V}{\partial z} = -L\frac{\partial I}{\partial t} \tag{9-5-46}$$

Similarly,

$$-\gamma V = -j\omega LI \tag{9-5-47}$$

Elimination of the circuit current I from Eqs. (9-5-45) and (9-5-46) yields

$$\gamma^2 V = -V\omega^2 LC - \gamma ij\omega L \tag{9-5-48}$$

If the convection-elecron beam current is not present, Eq. (9-5-48) reduces to a typical wave equation of a transmission line. When $i = 0$, the propagation constant is defined from Eq. (9-5-48) as

$$\gamma_0 \equiv j\omega\sqrt{LC} \tag{9-5-49}$$

and the characteristic impedance of the line can be determined from Eqs. (9-5-45) and (9-5-47):

$$Z_0 = \sqrt{\frac{L}{C}} \tag{9-5-50}$$

When the electron beam current is present, Eq. (9-5-48) can be written in terms of Eqs. (9-5-49) and (9-5-50):

$$V = \frac{\gamma \gamma_0 Z_0}{\gamma^2 - \gamma_0^2} i \qquad (9\text{-}5\text{-}51)$$

Since $E_z = -\nabla V = -(\partial V / \partial z) = \gamma V$, the axial electric field is given by

$$E_1 = -\frac{\gamma^2 \gamma_0 Z_0}{\gamma^2 - \gamma_0^2} i \qquad (9\text{-}5\text{-}52)$$

This equation is called the *circuit equation* because it determines how the axial electric field of the slow-wave helix is affected by the spatial ac electron beam current.

9-5-5 Wave Modes

The wave modes of a helix-type traveling-wave tube can be determined by solving the electronic and circuit equations simultaneously for the propagation constants. Each solution for the propagation constants represents a mode of traveling wave in the tube. It can be seen from Eqs. (9-5-43) and (9-5-52) that there are four distinct solutions for the propagation constants. This means that there are four modes of traveling wave in the O-type traveling-wave tube. Substitution of Eq. (9-5-43) in Eq. (9-5-52) yields

$$(\gamma^2 - \gamma_0^2)(j\beta_e - \gamma)^2 = -j\frac{\gamma^2 \gamma_0 Z_0 \beta_e I_0}{2V_o} \qquad (9\text{-}5\text{-}53)$$

Equation (9-5-53) is of fourth order in γ and thus has four roots. Its exact solutions can be obtained with numerical methods and a digital computer. However, the approximate solutions may be found by equating the dc electron beam velocity to the axial phase velocity of the traveling wave, which is equivalent to setting

$$\gamma_0 = j\beta_e \qquad (9\text{-}5\text{-}54)$$

Then Eq. (9-5-53) reduces to

$$(\gamma - j\beta_e)^3(\gamma + j\beta_e) = 2C^3\beta_e^2\gamma^2 \qquad (9\text{-}5\text{-}55)$$

where C is the traveling-wave tube gain parameter and is defined as

$$C \equiv \left(\frac{I_0 Z_0}{4V_0}\right)^{1/3} \qquad (9\text{-}5\text{-}56)$$

It can be seen from Eq. (9-5-55) that there are three forward traveling waves corresponding to $e^{-j\beta_e z}$ and one backward traveling wave corresponding to $e^{+j\beta_e z}$. Let the propagation constant of the three forward traveling waves be

$$\gamma = j\beta_e - \beta_e C\delta \qquad (9\text{-}5\text{-}57)$$

where it is assumed that $C\delta \ll 1$.

Substitution of Eq. (9-5-57) in Eq. (9-5-55) results in

$$(-\beta_e C\delta)^3(j2\beta_e - \beta_e C\delta) = 2C^3\beta_e^2(-\beta_e^2 - 2j\beta_e^2 C\delta + \beta_e^2 C^2\delta^2) \qquad (9\text{-}5\text{-}58)$$

Since $C\delta \ll 1$, Eq. (9-5-58) is reduced to

$$\delta = (-j)^{1/3} \qquad (9\text{-}5\text{-}59)$$

From the theory of complex variables the three roots of $(-j)$ can be plotted in Fig. 9-5-9.
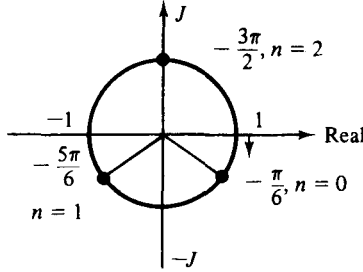
Figure 9-5-9   Three roots of $(-j)$.

Equation (9-5-59) can be written in exponential form as

$$\delta = (-j)^{1/3} = e^{-j[(\pi/2 + 2n\pi)/3]} \qquad (n = 0, 1, 2) \qquad (9\text{-}5\text{-}60)$$

The first root δ_1 at $n = 0$ is

$$\delta_1 = e^{-j\pi/6} = \frac{\sqrt{3}}{2} - j\frac{1}{2} \qquad (9\text{-}5\text{-}61)$$

The second root δ_2 at $n = 1$ is

$$\delta_2 = e^{-j5\pi/6} = -\frac{\sqrt{3}}{2} - j\frac{1}{2} \qquad (9\text{-}5\text{-}62)$$

The third root δ_3 at $n = 2$ is

$$\delta_3 = e^{-j3\pi/2} = j \qquad (9\text{-}5\text{-}63)$$

The fourth root δ_4 corresponding to the backward traveling wave can be obtained by setting

$$\gamma = -j\beta_e - \beta_e C\delta_4 \qquad (9\text{-}5\text{-}64)$$

Similarly,

$$\delta_4 = -j\frac{C^2}{4} \qquad (9\text{-}5\text{-}65)$$

Thus the values of the four propagation constants γ are given by

$$\gamma_1 = -\beta_e C\frac{\sqrt{3}}{2} + j\beta_e\left(1 + \frac{C}{2}\right) \qquad (9\text{-}5\text{-}66)$$

$$\gamma_2 = \beta_e C\frac{\sqrt{3}}{2} + j\beta_e\left(1 + \frac{C}{2}\right) \qquad (9\text{-}5\text{-}67)$$

$$\gamma_3 = j\beta_e(1 - C) \qquad (9\text{-}5\text{-}68)$$

$$\gamma_4 = -j\beta_e\left(1 - \frac{C^3}{4}\right) \qquad (9\text{-}5\text{-}69)$$

These four propagation constants represent four different modes of wave propagation in the O-type helical traveling-wave tube. It is concluded that the wave corresponding to γ_1 is a forward wave and that its amplitude grows exponentially with distance; the wave corresponding to γ_2 is also a forward wave, but its amplitude decays exponentially with distance; the wave corresponding to γ_3 is also a forward wave, but its amplitude remains constant; the fourth wave corresponding to γ_4 is a backward wave, and there is no change in amplitude. The growing wave propagates at a phase velocity slightly lower than the electron beam velocity, and the energy flows from the electron beam to the wave. The decaying wave propagates at the same velocity as that of the growing wave, but the energy flows from the wave to the electron beam. The constant-amplitude wave travels at a velocity slightly higher than the electron beam velocity, but no net energy exchange occurs between the wave and the electron beam. The backward wave progresses in the negative z direction with a velocity slightly higher than the velocity of the electron beam inasmuch as the typical value of C is about 0.02.

9-5-6 Gain Consideration

For simplicity, it is assumed that the structure is perfectly matched so that there is no backward traveling wave. Such is usually the case. Even though there is a reflected wave from the output end of the tube traveling backward toward the input end, the attenuator placed around the center of the tube subdues the reflected wave to a minimum or zero level. Thus the total circuit voltage is the sum of three forward voltages corresponding to the three forward traveling waves. This is equivalent to

$$V(z) = V_1 e^{-\gamma_1 z} + V_2 e^{-\gamma_2 z} + V_3 e^{-\gamma_3 z} = \sum_{n=1}^{3} V_n e^{-\gamma_n z} \qquad (9\text{-}5\text{-}70)$$

The input current can be found from Eq. (9-5-43) as

$$i(z) = -\sum_{n=1}^{3} \frac{I_0}{2V_0 C^2} \frac{V_n}{\delta_n^2} e^{-\gamma_n z} \qquad (9\text{-}5\text{-}71)$$

in which $C\delta \ll 1$, $E_1 = \gamma V$, and $\gamma = j\beta_e(1 - C\delta)$ have been used.

The input fluctuating component of velocity of the total wave may be found from Eq. (9-5-38) as

$$v_1(z) = \sum_{n=1}^{3} j \frac{v_0}{2V_0 C} \frac{V_n}{\delta_n} e^{-\gamma_n z} \qquad (9\text{-}5\text{-}72)$$

where $E_1 = \gamma V$, $C\delta \ll 1$, $\beta_e v_0 = \omega$, and $v_0 = \sqrt{(2e/m)V_0}$ have been used.

To determine the amplification of the growing wave, the input reference point is set at $z = 0$ and the output reference point is taken at $z = \ell$. It follows that at $z = 0$ the voltage, current, and velocity at the input point are given by

$$V(0) = V_1 + V_2 + V_3 \qquad (9\text{-}5\text{-}73)$$

$$i(0) = -\frac{I_0}{2V_0 C^2}\left(\frac{V_1}{\delta_1^2} + \frac{V_2}{\delta_2^2} + \frac{V_3}{\delta_3^2}\right) \qquad (9\text{-}5\text{-}74)$$

$$v_1(0) = -j \frac{v_0}{2V_0 C} \left(\frac{V_1}{\delta_1} + \frac{V_2}{\delta_2} + \frac{V_3}{\delta_3} \right) \tag{9-5-75}$$

The simultaneous solution of Eqs. (9-5-73), (9-5-74), and (9-5-75) with $i(0) = 0$ and $v_1(0) = 0$ is

$$V_1 = V_2 = V_3 = \frac{V(0)}{3} \tag{9-5-76}$$

Since the growing wave is increasing exponentially with distance, it will predominate over the total voltage along the circuit. When the length ℓ of the slow-wave structure is sufficiently large, the output voltage will be almost equal to the voltage of the growing wave. Substitution of Eqs. (9-5-66) and (9-5-76) in Eq. (9-5-70) yields the output voltage as

$$V(\ell) \approx \frac{V(0)}{3} \exp\left(\frac{\sqrt{3}}{2} \beta_e C \ell \right) \exp\left[-j\beta_e \left(1 + \frac{C}{2} \right) \ell \right] \tag{9-5-77}$$

The factor $\beta_e \ell$ is conventionally written $2\pi N$, where N is the circuit length in electronic wavelength—that is,

$$N = \frac{\ell}{\lambda_e} \quad \text{and} \quad \beta_e = \frac{2\pi}{\lambda_e} \tag{9-5-78}$$

The amplitude of the output voltage is then given by

$$V(\ell) = \frac{V(0)}{3} \exp\left(\sqrt{3}\, \pi N C \right) \tag{9-5-79}$$

The output power gain in decibels is defined as

$$A_p \equiv 10 \log \left| \frac{V(\ell)}{V(0)} \right|^2 = -9.54 + 47.3 NC \quad \text{dB} \tag{9-5-80}$$

where NC is a numerical number.

The output power gain shown in Eq. (9-5-80) indicates an initial loss at the circuit input of 9.54 dB. This loss results from the fact that the input voltage splits into three waves of equal magnitude and the growing wave voltage is only one-third the total input voltage. It can also be seen that the power gain is proportional to the length N in electronic wavelength of the slow-wave structure and the gain parameter C of the circuit.

Example 9-5-1:   Operation of Traveling-Wave Tube (TWT)

A traveling-wave tube (TWT) operates under the following parameters:

Beam voltage: $V_0 = 3$ kV
Beam current: $I_0 = 30$ mA
Characteristic impedance of helix: $Z_0 = 10\ \Omega$
Circuit length: $N = 50$
Frequency: $f = 10$ GHz

Determine: **(a)** the gain parameter C; **(b)** the output power gain A_p in decibels; and **(c)** all four propagation constants.

Solution

a. From Eq. (9-5-56) the gain parameter is

$$C = \left(\frac{I_0 Z_0}{4V_0}\right)^{1/3} = \left(\frac{30 \times 10^{-3} \times 10}{4 \times 3 \times 10^3}\right)^{1/3} = 2.92 \times 10^{-2}$$

b. From Eq. (9-5-80) the output power gain is

$$A_p = -9.54 + 4.73\, NC = -9.54 + 47.3 \times 50 \times 2.92 \times 10^{-2} = 59.52 \text{ dB}$$

c. The four propagation constants are

$$\beta_e = \frac{\omega}{v_0} = \left(\frac{2\pi \times 10^{10}}{0.593 \times 10^6 \sqrt{3 \times 10}}\right) = 1.93 \times 10^3 \text{ rad/m}$$

$$\gamma_1 = -\beta_e C \frac{\sqrt{3}}{2} + j\beta_e\left(1 + \frac{C}{2}\right)$$

$$= -1.93 \times 10^3 \times 2.92 \times 10^{-2} \times 0.87 + j1.93 \times 10^3\left(1 + \frac{2.92 \times 10^{-2}}{2}\right)$$

$$= -49.03 + j1952$$

$$\gamma_2 = \beta_e C \frac{\sqrt{3}}{2} + j\beta_e\left(1 + \frac{C}{2}\right) = 49.03 + j1952$$

$$\gamma_3 = j\beta_e(1 - C) = j1.93 \times 10^3(1 - 2.92 \times 10^{-2})$$

$$= j1872.25$$

$$\gamma_4 = -j\beta_e\left(1 - \frac{C^3}{4}\right) = -j1.93 \times 10^3\left[1 - \frac{(2.92 \times 10^{-2})^3}{4}\right]$$

$$= -j1930$$

9-6 COUPLED-CAVITY TRAVELING-WAVE TUBES

Helix traveling-wave tubes (TWTs) produce at most up to several kilowatts of average power output. Here we describe coupled-cavity traveling-wave tubes, which are used for high-power applications.

9-6-1 Physical Description

The term *coupled cavity* means that a coupling is provided by a long slot that strongly couples the magnetic component of the field in adjacent cavities in such a manner that the passband of the circuit is mainly a function of this one variable. Figure 9-6-1 shows two coupled-cavity circuits that are principally used in traveling-wave tubes.
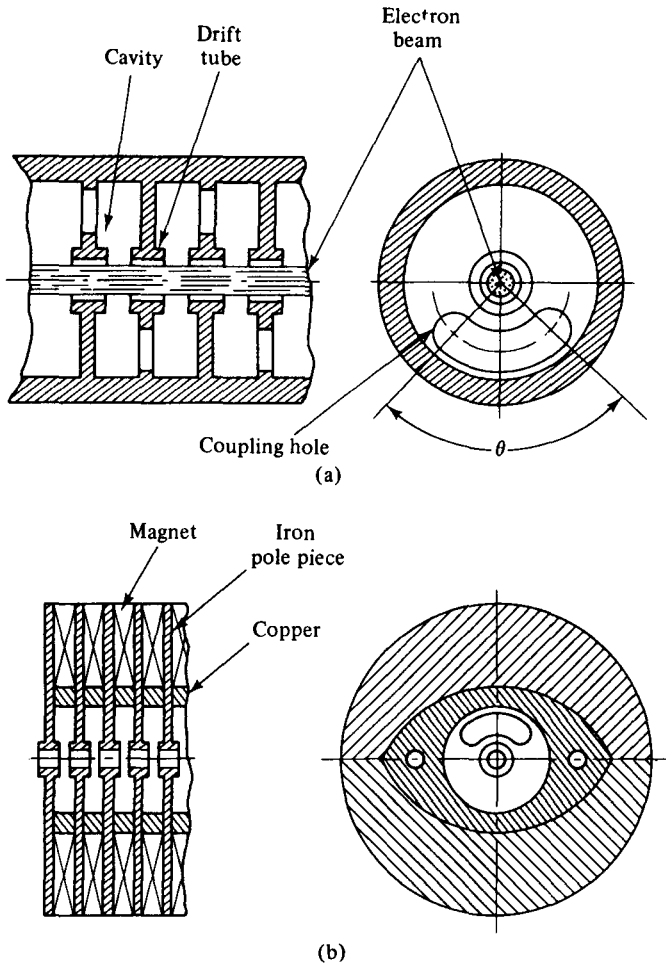
Figure 9-6-1   Coupled-cavity circuits in the TWTs. (a) Basic coupled-cavity circuit. (b) Coupled-cavity circuit with integral periodic-permanent-magnet (PPM) focusing. (*After J. T. Mendel [15]; reprinted by permission of IEEE.*)

As far as the coupling effect is concerned, there are two types of coupled-cavity circuits in traveling-wave tubes. The first type consists of the fundamentally forward-wave circuits that are normally used for pulse applications requiring at least half a megawatt of peak power. These coupled-cavity circuits exhibit negative mutual inductive coupling between the cavities and operate with the fundamental space harmonic. The cloverleaf [12] and centipede circuits [13] (see Fig. 9-6-2) belong to this type. The second type is the first space-harmonic circuit, which has positive mutual coupling between the cavities. These circuits operate with the first spatial harmonic and are commonly used for pulse or continuous wave (CW) applications from one to several hundred kilowatts of power output [14]. In addition, the long-slot circuit of the positive mutual coupling-cavity circuit operates at the fundamental spatial
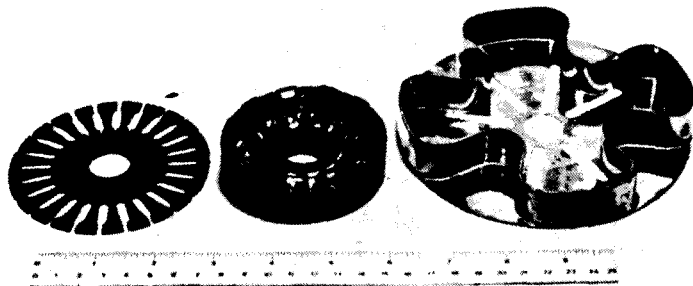
Figure 9-6-2　Centipede and cloverleaf coupled-cavity circuits. (*After A. Staprans et al. [7]; reprinted by permission of IEEE.*)

Figure 9-6-3　Space-harmonic coupled-cavity circuits. (*After A. Staprans et al. [7]; reprinted by permission of IEEE, Inc.*)

harmonic with a higher frequency mode. This circuit is suitable for megawatt power output. Figure 9-6-3 shows several space-harmonic coupled-cavity circuits.

9-6-2 Principles of Operation

Any repetitive series of lumped LC elements constitute a propagating filter-type circuit. The coupled cavities in the traveling-wave tube are usually highly over-coupled, resulting in a bandpass-filter-type characteristic. When the slot angle (θ) as shown in Fig. 9-6-1(a) is larger than 180°, the passband is close to its practical limits. The drift tube is formed by the reentrant part of the cavity, just as in the case of a klystron. During the interaction of the RF field and the electron beam in the traveling-wave tube a phase change occurs between the cavities as a function of frequency. A decreasing phase characteristic is reached if the mutual inductance of the coupling slot is positive, whereas an increasing phase characteristic is obtained if the mutual inductive coupling of the slot is negative [12].

　　The amplification of the traveling-wave tube interaction requires that the electron beam interact with a component of the circuit field that has an increasing phase characteristic with frequency. The circuit periodicity can give rise to field components that have phase characteristics [16] as shown in Fig. 9-6-4. In Fig. 9-6-4 the angular frequency (ω) is plotted as a function of the phase shift ($\beta \ell$) per cavity. The ratio of ω to β is equal to the phase velocity. For a circuit having positive mutual in-
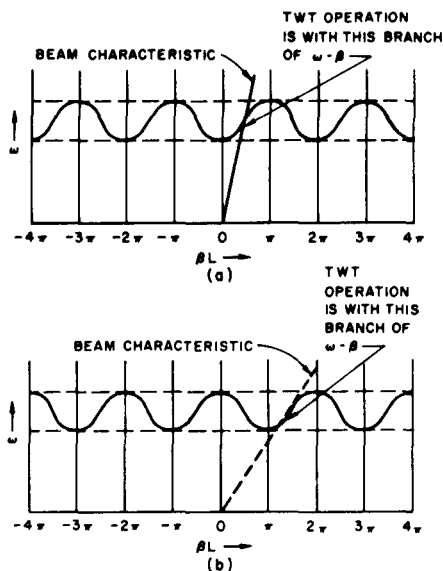
Figure 9-6-4 ω-β diagrams for coupled-cavity circuits. (a) Fundamental forward-wave circuit (negative mutual inductance coupling between cavities). (b) Fundamental backward-wave circuit (positive mutual inductive coupling between cavities). (*After A. Staprans et al. [7]; reprinted by permission of IEEE, Inc.*)

ductive coupling between the cavities, the electron beam velocity is adjusted to be approximately equal to the phase velocity of the first forward-wave spatial harmonic. For the circuits with negative mutual inductive coupling, the fundamental branch component of the circuit wave is suitable for synchronism with the electron beam and is normally used by the traveling-wave tube. The coupled-cavity equivalent circuit has been developed by Curnow [17] as shown in Fig. 9-6-5.
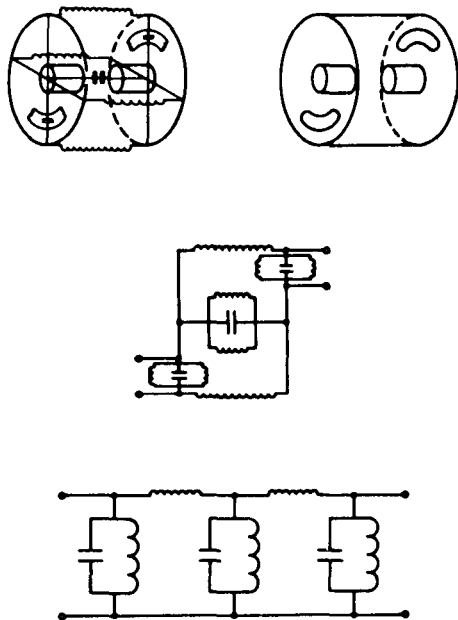
Figure 9-6-5   Equivalent circuits for a slot-coupled cavity. (*After A. Staprans et al. [7]; reprinted by permission of IEEE, Inc.*)

In Fig. 9-6-5 inductances are used to represent current flow and capacitors to represent the electric fields of the cavities. The circuit can be evolved into a fairly simple configuration. Loss in the cavities can be approximately calculated by adding resistance in series with the circuit inductance.

9-6-3 Microwave Characteristics

When discussing the power capability of traveling-wave tubes, it is important to make a clear distinction between the average and the peak power because these two figures are limited by totally different factors. The average power at a given frequency is almost always limited by thermal considerations relative to the RF propagating circuit. However, the peak RF power capability depends on the voltage for which the tube can be designed. The beam current varies as the three-half power of
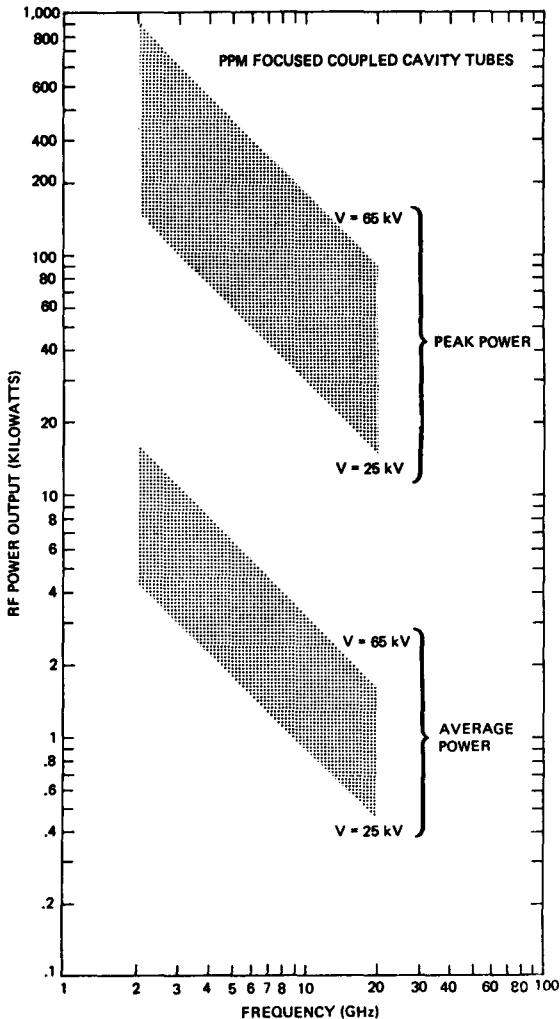
Figure 9-6-6 Peak and average power capability of typical TWTs in field use. (*After J. T. Mendel [15]; reprinted by permission of IEEE., Inc.*)

the voltage, and the product of the beam current and the voltage determines the total beam power. That is,

$$I_{\text{beam}} = KV_0^{3/2} \tag{9-6-1}$$

$$P_{\text{beam}} = KV_0^{5/2} \tag{9-6-2}$$

where V_0 is the beam voltage and K is the electron-gun perveance.

For a solid-beam electron gun with good optics, the perveance is generally considered to be between 1 to 2×10^{-6}. Once the perveance is fixed, the required voltage for a given peak beam power is then uniquely determined. Figures 9-6-6 and 9-6-7 demonstrate the difference between peak and average power capability and the difference between periodic-permanent-magnet (PPM) and solenoid-focused designs [15].

Coupled-cavity traveling-wave tubes are constructed with a limited amount of gain per section of cavities to ensure stability. Each cavity section is terminated by
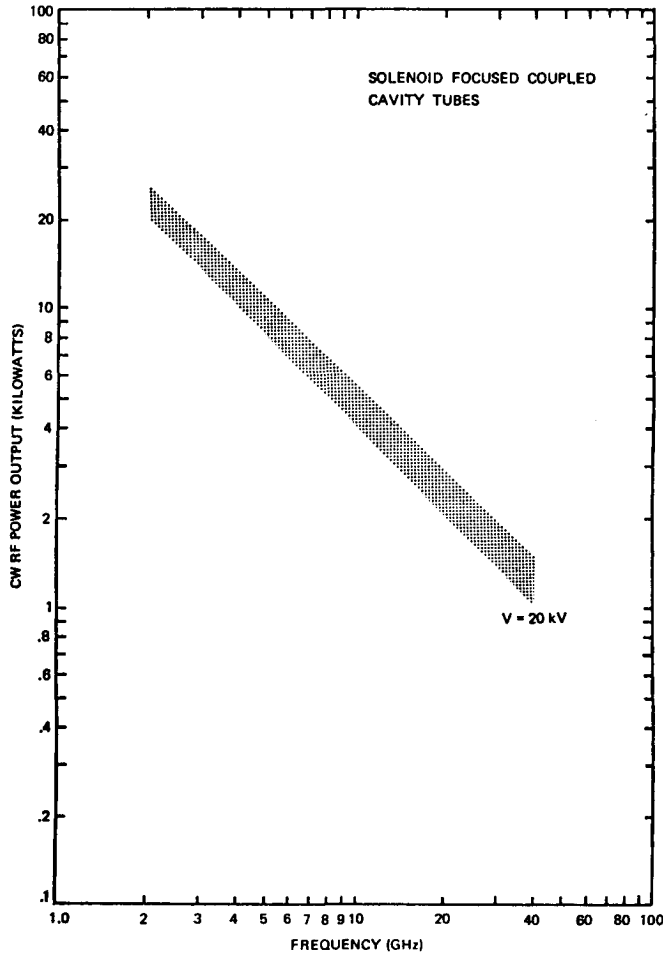
Figure 9-6-7 CW power capability of TWTs operating at nearly 20 kV. (*After J. T. Mendel [15]; reprinted by permission of IEEE., Inc.*)

either a matched load or an input or output line in order to reduce gain variations with frequency. Cavity sections are cascaded to achieve higher tube gain than can be tolerated in one section of cavities. Stable gain greater than 60 dB can be obtained over about 30% bandwidth by this method.

The overall efficiency of coupled-cavity traveling-wave tubes is determined by the amount of energy converted to RF energy and the energy dissipated by the collector. Interaction efficiencies from 10 to 40% have been achieved from coupled-cavity traveling-wave tubes. Overall efficiencies of 20 to 55% have been obtained [7].

9-7 HIGH-POWER AND GRIDDED-CONTROL TRAVELING-WAVE TUBES

Coupled-cavity traveling-wave tubes are the most versatile devices used for amplification at microwave frequencies with high gain, high power, high efficiency, and wide bandwidth. The present state of the art for U.S. high-power gridded tubes is shown in Fig. 9-7-1. High-power TWTs have four main sections: electron gun for electron emission, slow-wave structure for effective beam interaction, magnetic circuit for beam focusing, and collector structure for collecting electron beams and dissipating heat energy. Specifically, the physical components of a coupled-cavity traveling-wave tube consist of an electron emitter, a shadow grid, a control grid, a modulating anode, a coupled-cavity circuit, a solenoid magnetic circuit, and a collector depression structure (see Fig. 9-7-2).
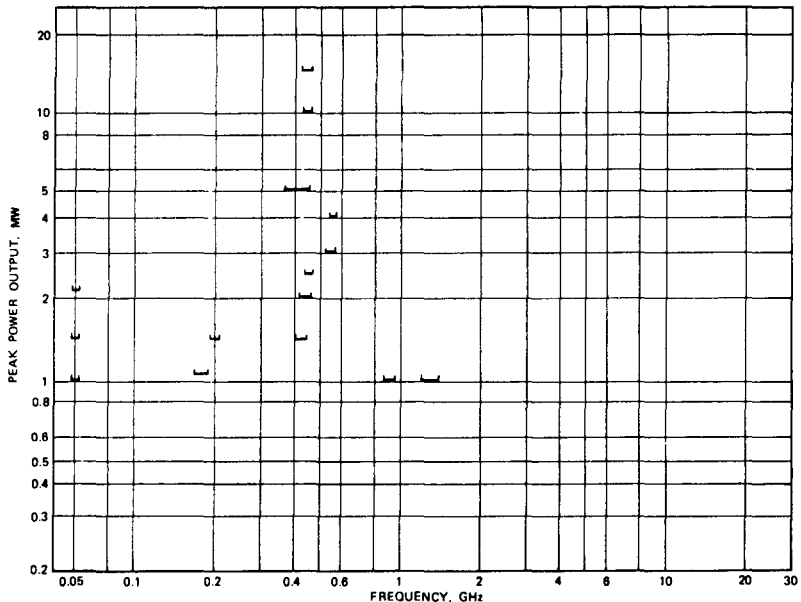
Figure 9-7-1 State of the art for U.S. high-power gridded tubes.

Figure 9-7-2 Cutaway of a coupled-cavity traveling-wave tube. (*Courtesy of Hughes Aircraft Company, Electron Dynamics Division.*)

After electrons are emitted from the cathode, the electron beam has a tendency to spread out because of the electron-repelling force. On the other hand, the electron beam must be small enough for effective interaction with the slow-wave circuit. Usually the diameter of the electron beam is smaller than one-tenth wavelength of the signal. Coupled-cavity traveling-wave high-power tubes utilize a shadow-grid technique to control the electron beam; so the device is called a *gridded traveling-wave tube* (GTWT). As shown in Fig. 9-7-3 [18], the electron emitter of a gridded traveling-wave tube has two control electrodes: one shadow grid near the cathode and one control grid slightly away from the cathode. The shadow grid, which is at cathode potential and interposed between the cathode and the control grid, suppresses electron emission from those portions of the cathode that would give rise to interception at the control grid. The control grid, which is at a positive potential, controls the electron beam. These grids can control far greater beam power than would otherwise be possible.
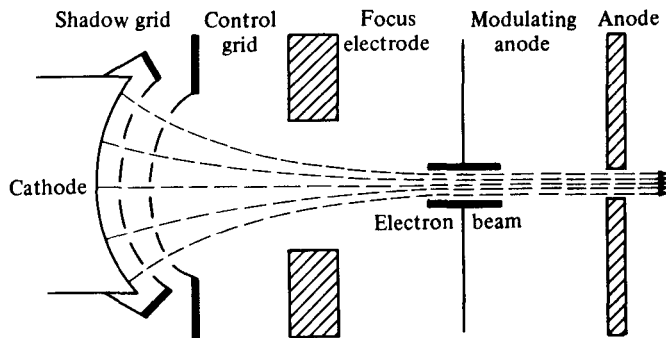
Figure 9-7-3   Electron emitter with control electrodes for gridded high-power TWT.

In general, an anode modulation technique is frequently used in traveling-wave tubes to eliminate voltage pulsing through the lower unstable beam voltage and to reduce modulator power requirements for high-power pulse output. In gridded traveling-wave tubes the modulator applies a highly regulated positive grid drive voltage with respect to the cathode to turn the electron beam on for RF amplification. An unregulated negative grid bias voltage with respect to the cathode is used to cut the electron beam off. Thus the anode modulator acts as a pulse switch for the electron beam of the gridded traveling-wave tube.

The anode of the electron gun is operated at a voltage higher than that of the slow-wave structure to prevent positive ions, formed by the electron beam in the region of the slow-wave structure, from draining toward the cathode and bombarding it.

9-7-1 High Efficiency and Collector Voltage Depression

After passing through the output cavity, the electron beam strikes a collector electrode. The function of the collector electrode could be performed by replacing the second grid of the output cavity with a solid piece of metal. However, using a sepa-

rate electrode may have two advantages. First, the collector can be made as large as desired in order to collect the electron beam at a lower density, thus minimizing localized heating. If the collector were part of the slow-wave circuit, its size would be limited by the maximum gap capacitance consistent with good high-frequency performance. Second, using a separate collector can reduce its potential considerably below the beam voltage in the RF interaction region, thereby reducing the power dissipated in the collector and increasing the overall efficiency of the device. Gridded traveling-wave high-power tubes have a separate collector that dissipates the electrons in the form of heat. A cooling mechanism absorbs the heat by thermal conduction to a cooler surface.

The efficiency of a gridded traveling-wave high-power tube is the ratio of the RF power output to the product of cathode voltage (beam voltage) and cathode current (beam current). It may be expressed in terms of the product of the electronic efficiency and the circuit efficiency. The electronic efficiency expresses the percentage of the dc or pulsed input power that is converted into RF power on the slow-wave structure. The circuit efficiency, on the other hand, determines the percentage of dc input power that is delivered to the load exterior to the tube. The electron beam does not extract energy from any dc power supply unless the electrons are actually collected by an electrode connected to that power supply. If a separate power supply is connected between cathode and collector and if the cavity grids intercept a negligible part of the electron beam, the power supply between the cathode and collector will be the only one supplying power to the tube. For a gridded traveling-wave tube, the collector voltage is normally operated at about 40% of the cathode voltage. Thus the overall efficiency of conversion of dc to RF power is almost twice the electronic efficiency. Under this condition the tube is operating with collector voltage depression.

9-7-2 Normal Depression and Overdepression of Collector Voltage

Most gridded traveling-wave tubes are very sensitive to variations of collector depression voltages below normal depression level, since the tubes operate close to the knee of the electron spent beam curves. Figure 9-7-4 [18] shows the spent beam curves for a typical gridded traveling-wave tube.

Under normal collector depression voltage V_c at -7.5 kV with full saturated power output, the spent beam electrons are collected by the collector and returned to the cathode. Thus the collector current I_c is about 2.09 A. A small amount of electrons intercepted by the beam scraper or slow-wave circuit contribute the tube body current for about 0.178 A. Very few electrons with lower kinetic energy reverse the direction of their velocity inside the collector and fall back onto the output pole piece. These returning electrons yield a current I_r of 0.041 A, which is only a small fraction of the body current I_b. These values are shown in Fig. 9-7-4.

When the collector voltage is overdepressed from the normal level of -7.5 kV to the worst case of about -11.5 kV, a greater number of the spent electrons inside the collector reverse the direction of their velocity by a highly negative collector voltage and fall back onto the grounded output pole piece because the potential of the pole piece is 11.5 kV higher than the collector voltage. It can be seen from Fig.
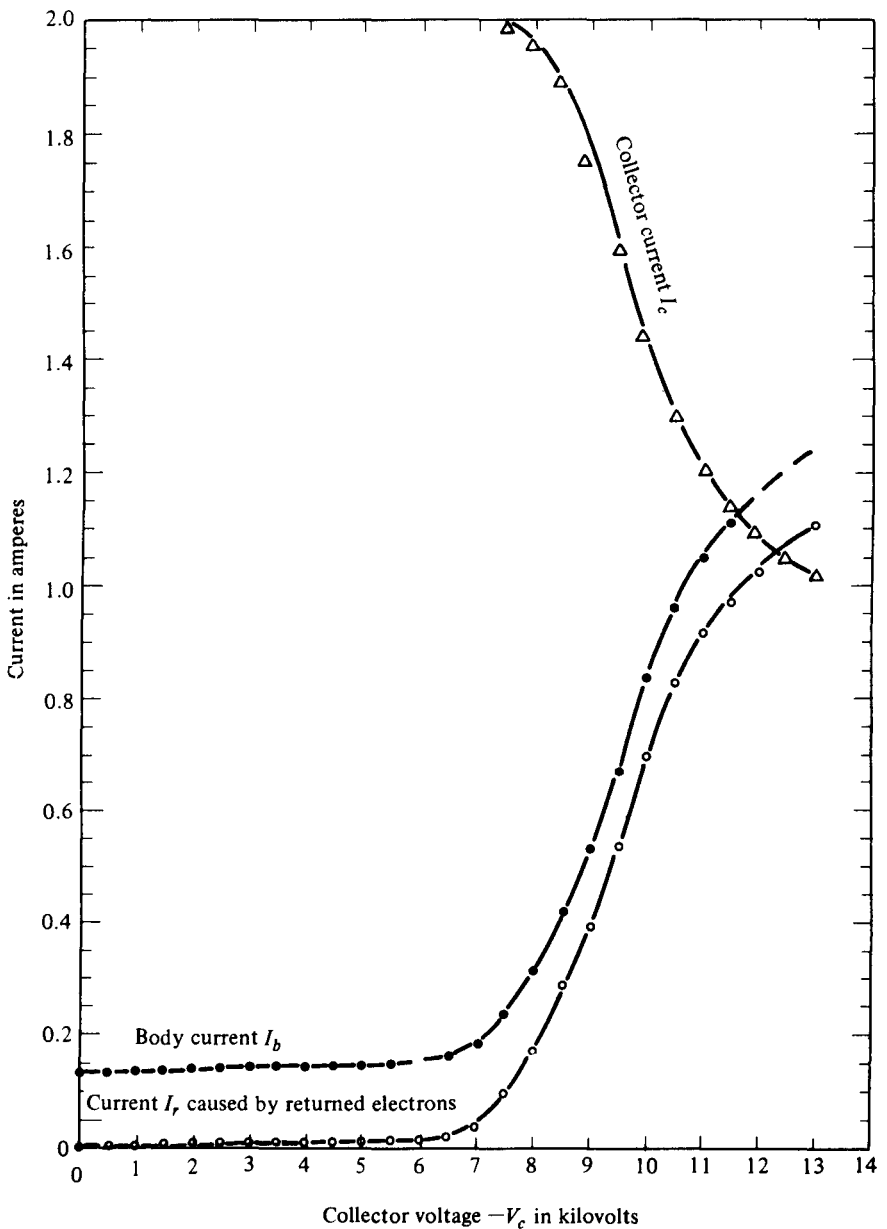
Figure 9-7-4 Spent beam curves for a typical gridded TWT.

9-7-4 that when the collector voltage is overdepressed from -7.5 to -11.5 kV, the collector current is decreased sharply from 2.01 to 1.14 A and the body current is increased rapidly from 0.237 to 1.110 A. The body current consists of two parts: One part is the current caused by the electrons intercepted by the circuit or the beam scrapers; another part is the current caused by the electrons returned by the overdepressed collector voltage. Figure 9-7-5 [18] shows the impact probability of returned electrons by certain overdepressed collector voltage.
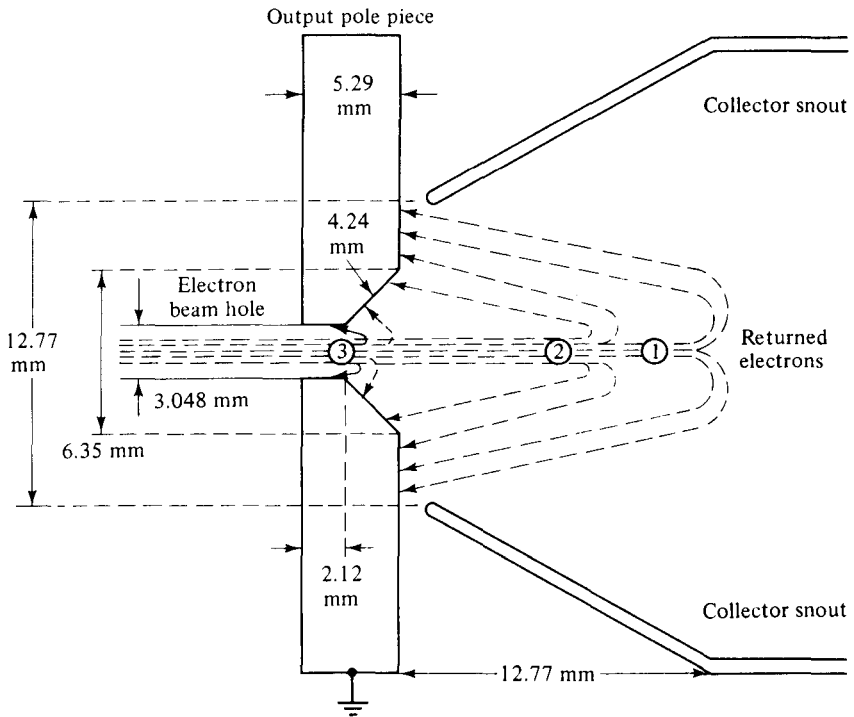
Figure 9-7-5 Impact probability of returned electrons by overdepressed collector voltage.

Example 9-7-1:   Gridded Traveling-Wave Tube (GTWT)

A gridded traveling-wave tube is operated with overdepression of collector voltage as follows:

Overdepression collector voltage:	$V_c = -11$ kV
Returned current:	$I_r = 0.85$ A
Mass of heated iron pole piece:	mass $= 250$ mg
Specific heat H of iron at 20°C:	$H = 0.108$ calories/g-°C

Determine:

a. The number of electrons returned per second
b. The energy (in eV) associated with these returning electrons in 20 ms

c. The power in watts for the returning electrons

d. The heat in calories associated with the returning electrons (a factor for converting joules to calories is 0.238)

e. The temperature T in degrees Celsius for the output iron pole piece [*Hint:* $T = 0.238VIt/(\text{mass} \times \text{specific heat})$]

f. Whether the output iron pole piece is melted

Solution

a. The number of electrons returned per second is

$$I_r = \frac{0.85}{1.6 \times 10^{-19}} = 5.31 \times 10^{18} \text{ electrons per second}$$

b. The energy is

$$W = Pt = VI_r t = 11 \times 10^3 \times 5.31 \times 10^{18} \times 20 \times 10^{-3}$$
$$= 1.168 \times 10^{21} \text{ eV}$$

c. The power is

$$P = VI_r = 11 \times 10^3 \times 0.85 = 9.35 \text{ kW}$$

d. The created heat is

$$H \text{ (heat)} = 0.238Pt = 0.238VI_r t$$
$$= 0.238 \times 11 \times 10^3 \times 0.85 \times 20 \times 10^{-3}$$
$$= 44.51 \text{ calories}$$

e. The temperature is

$$T = \frac{0.238VI_r t}{\text{mass} \times \text{specific heat}}$$
$$= \frac{44.51}{250 \times 10^{-3} \times 0.108} = 1648.52°C$$

f. The output iron pole piece is melted.

9-7-3 Two-Stage Collector Voltage Depression Technique

When the spent electron beam arrives in the collector, the kinetic energies of each electron are different. Under the normal operation at a collector voltage of about 40% of the cathode voltage, very few electrons will be returned by the negative collector voltage. Consequently, the tube body current is very small and negligible because the returned electrons are the only ones intercepted by the cavity grids and the slow-wave circuit. When the collector is more negative, however, more electrons with lower energy will reverse their direction of velocity and fall onto the output

pole piece. Thus the tube body current will increase sharply. Since electrons of various energy classes exist inside the collector, two-stage collector voltage depression may be utilized. Each stage is biased at a different voltage. Specifically, the main collector may be biased at 40% depression of the cathode voltage for normal operation, but the collector snout may be grounded to the output pole piece for overdepression operation. As a result, the returned electrons will be collected by the collector snout and returned to the cathode even though the collector voltage is overdepressed to be more negative. Since the collector is cooled by a cooling mechanism, the overheating problem for overdepression is eased. Figure 9-7-6 shows a structure of two-stage collector voltage depression, and Fig. 9-7-7 depicts a basic interconnection of a gridded traveling-wave tube with its power supplies [18].
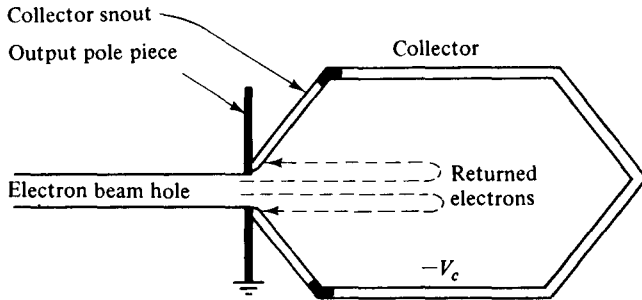
Figure 9-7-6   Diagram for two-stage collector depression.
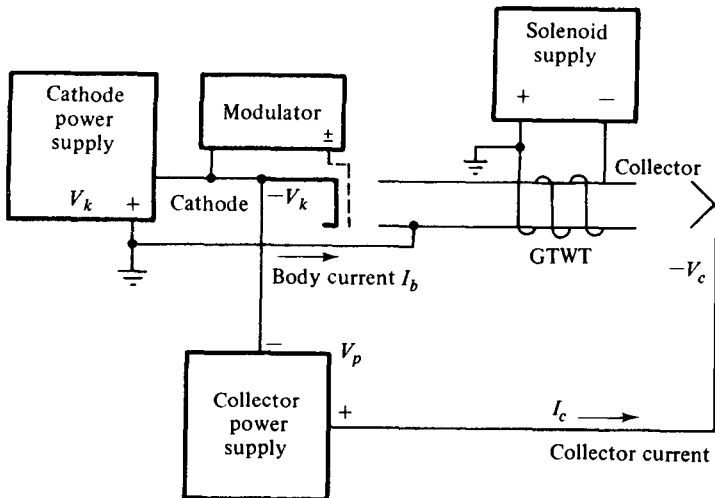
Figure 9-7-7   Basic interconnection of a gridded traveling-wave tube with its power supplies.

9-7-4 Stabilization of Cathode and Collector Voltages

The cathode voltage of a gridded traveling-wave tube is negative with respect to ground so that the electrons can be emitted from the cathode. In order to maintain a constant beam power for a uniform gain, the cathode voltage must be constant. In addition, the phase shift through the tube is directly related to the beam velocity; thus high resolution and low ripple are required in the cathode voltage power supply to avoid undesirable phase-shift variations. Consequently, the cathode power supply of the gridded traveling-wave high-power tube is usually regulated for better than 1% over line and load changes and is also well filtered because of the critical requirements on the cathode voltage with respect to ground. The cathode power supply provides the tube body current. Under normal operation the body current is very small in comparison with the collector current. Figure 9-7-8 shows a basic interconnection for a gridded traveling-wave tube with two regulated power supplies.
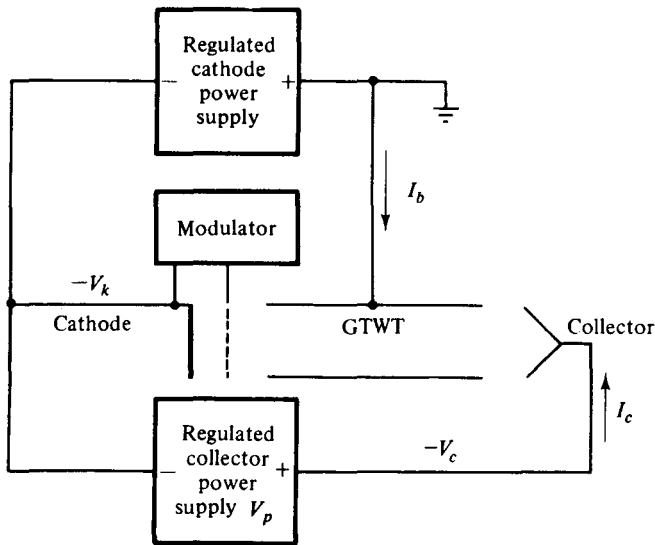
Figure 9-7-8 Interconnections for a gridded traveling-wave high-power tube.

Figure 9-7-9 illustrates a voltage-regulator circuit for a cathode power supply. The voltage regulator indicated in the circuit consists of two devices: one differential amplifier and one tetrode tube. The solid-state differential amplifier amplifies the difference between the preset reference voltage and the voltage that is one-thousandth of the output voltage. The reference voltage is adjustable to a preset level. The output voltage of the differential amplifier drives the control grid of the regulator tube between cutoff and saturation in order to nullify the difference voltage.

As shown in Fig. 9-7-8 the negative terminal of the collector power supply is connected to the cathode and the positive terminal to the collector electrode. The
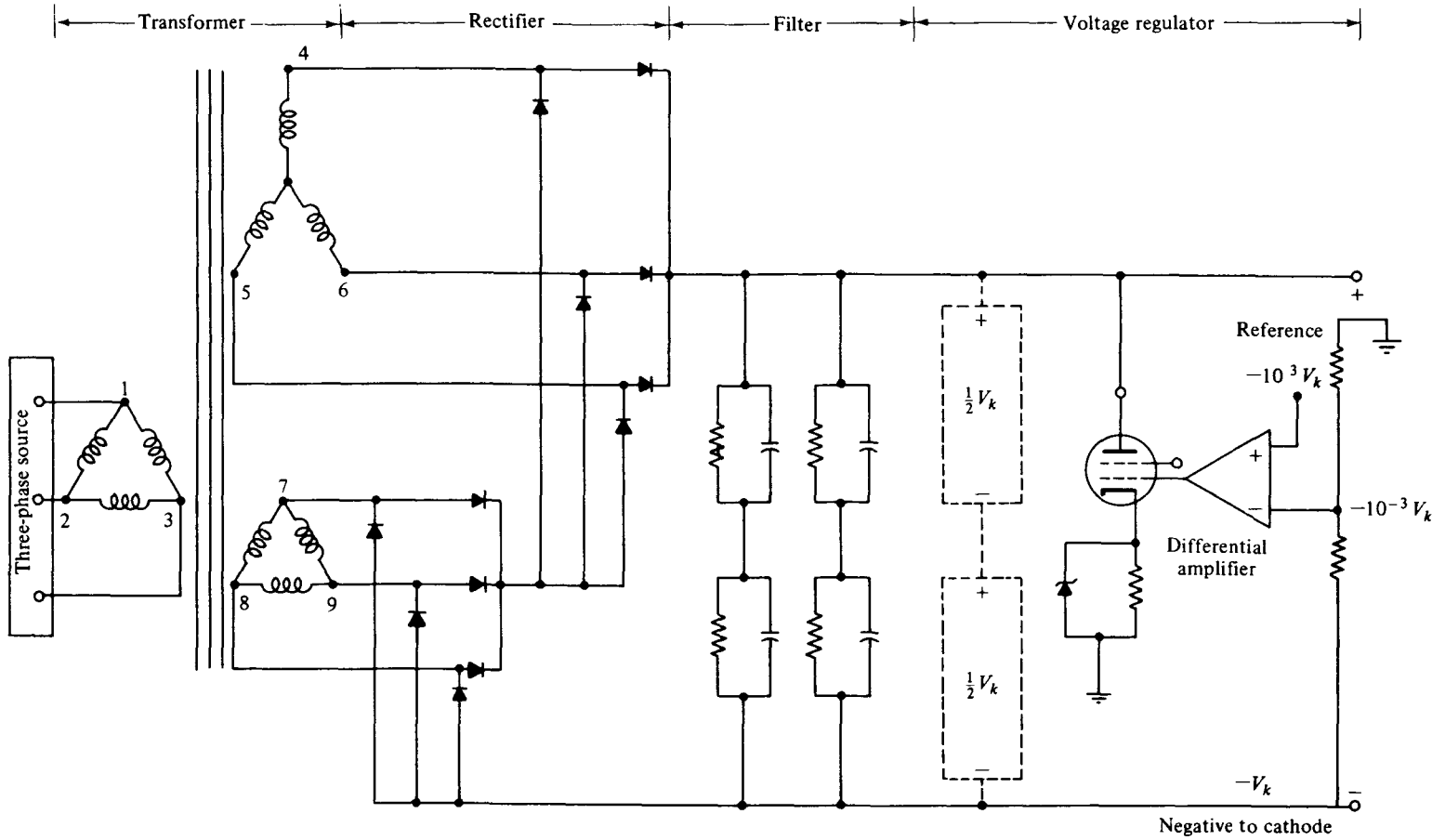
Figure 9-7-9 Voltage-regulator circuit for cathode power supply.

collector depression voltage is the difference of the two regulated supply voltages. The cathode supply provides the tube body current and the collector supply yields the collector current. The ratio of the collector current over the body current is about 10 for an operation of 40% voltage depression. Thus the power delivered to the tube by the collector supply is about four times larger than the power furnished by the cathode supply. An electrical transient may occur in the circuit when power supplies are just being switched on or off. The reasons for an electrical transient may be caused by two factors:

1. **Load changes:** When the load of a generator is suddenly increased, a larger current is needed. Since the generator cannot meet the demand instantaneously, the output voltage of the generator drops momentarily. Conversely, when the load of a generator is suddenly decreased, the output voltage drops accordingly.

2. **Switching on or off:** When the switch of a generator is just turned on or off, the armature current in the armature conductors produces armature reaction. The nature of armature reaction reduces the terminal voltage for lagging loads.

When an electrical transient is created in the circuit, the collector voltage is overdepressed. As a result, the spent electrons inside the collector reverse the direction of their velocity by the highly negative collector voltage and fall back onto the grounded output pole piece. The tube body current is sharply increased and the collector current is greatly decreased. When the returned electrons impact the output pole piece, the pole piece will be damaged by high heat. The damage of the output pole piece creates a mismatch in the interaction circuit and degrades the performance of the tube. In particular, the tube gain, efficiency, bandwidth, and power output are affected accordingly by the circuit mismatch when the collector voltage is overdepressed below the normal depression level. Furthermore, the large body current may burnout the solid-state differential amplifier of the cathode voltage regulator and vary the electron beam of the gridded tube. If the damage is beyond the tolerance of the gridded traveling-wave high-power tube, the tube may cease to function.

In order to maintain a constant collector depression voltage, the collector voltage must be regulated. There are three possible ways to do so [18]:

1. **Regulator in series with the collector power supply:** In this method a voltage regulator is incorporated in series with the collector supply as shown in Fig. 9-7-10 so that the output voltage of the collector supply may be regulated at a certain level with respect to ground. Since the output voltage of the cathode supply is highly regulated at a certain level, the difference between the two regulated voltages will produce a well-regulated voltage with respect to ground at the collector electrode.

2. **Regulator in parallel with the collector supply:** In this method a voltage regulator is inserted in parallel with the collector supply as shown in Fig. 9-7-11 so that a regulated voltage with respect to ground at normal depression may be achieved at the collector terminal.
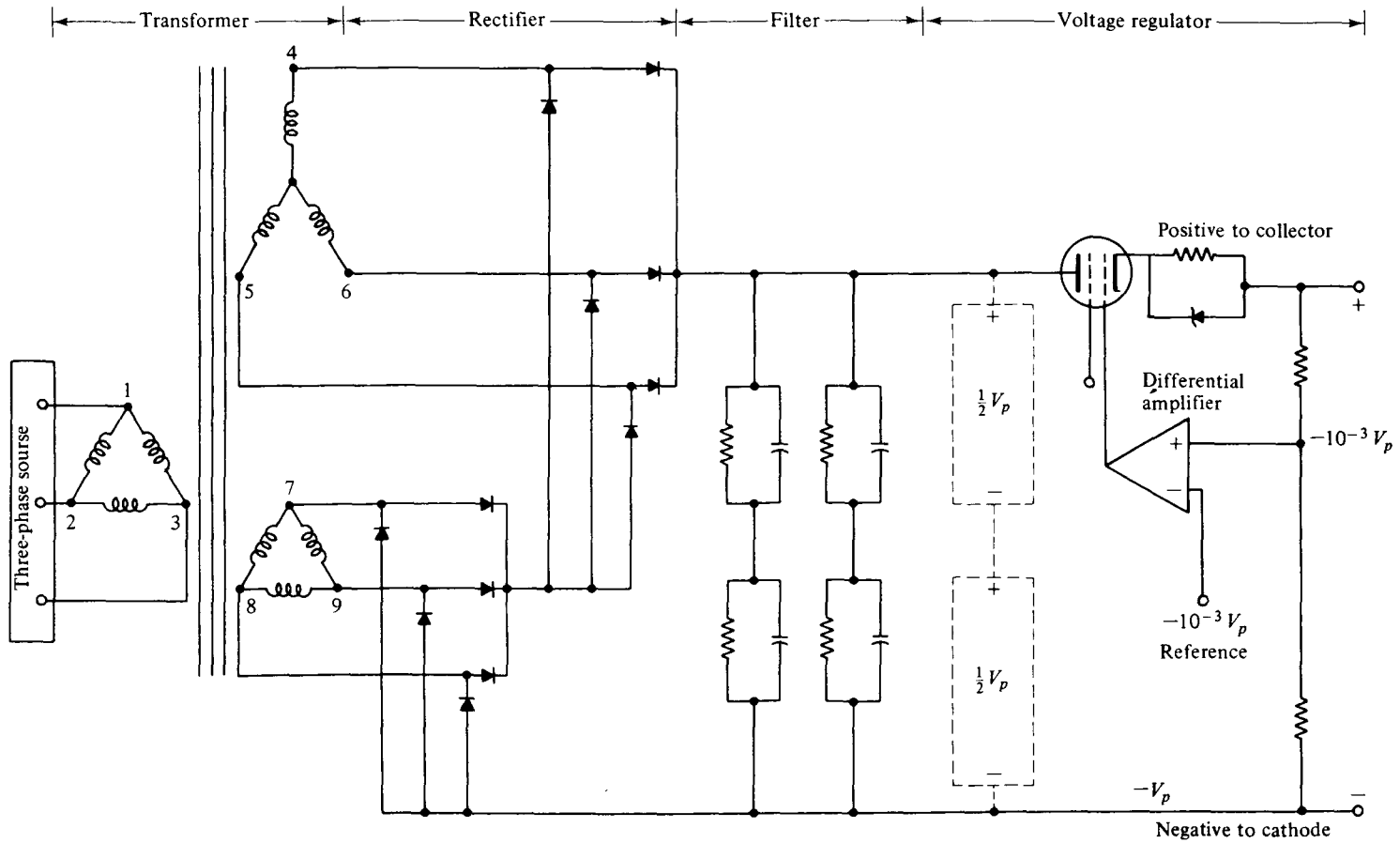
Figure 9-7-10 Regulator in series with collector supply.
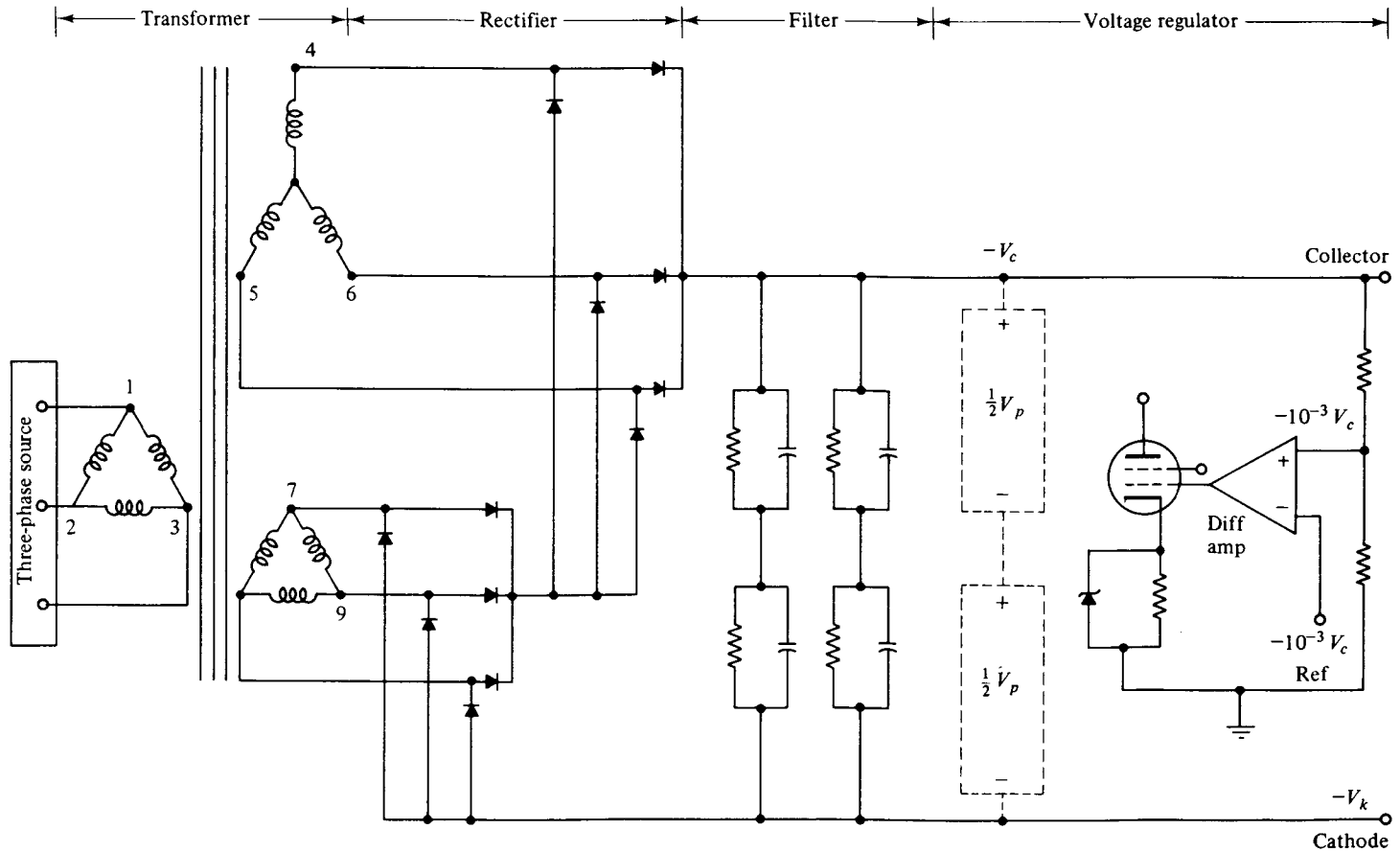
Figure 9-7-11 Regulator in parallel with collector supply.

3. Regulator between the cathode voltage and the collector voltage: In this method the collector depression voltage is regulated with respect to the cathode voltage as shown in Fig. 9-7-12. If the collector voltage is overdepressed above the normal depression value (absolute value), differential amplifier 2 tends to adjust the cathode voltage below its fixed level (absolute value). When the cathode voltage is dropped, the collector voltage is readjusted to its normal depression level with respect to ground.
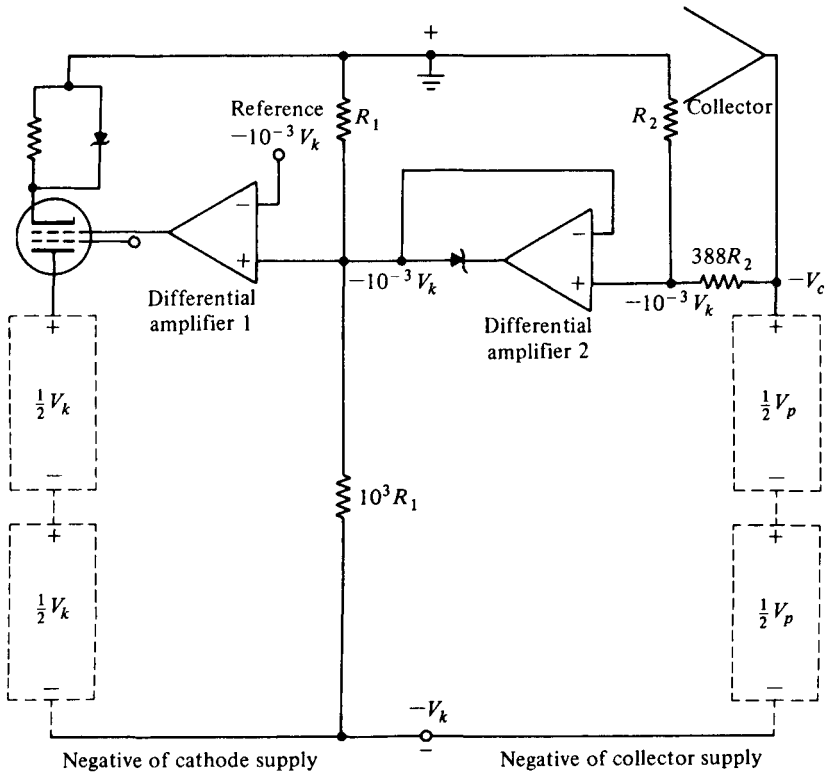
Figure 9-7-12   Regulator between cathode voltage and collector voltage.

REFERENCES

[1] WARNECKE, R. R., et al., Velocity modulated tubes. In *Advances in Electronics,* Vol. 3. Academic Press, New York, 1951.

[2] CHODOROW, M., and C. SUSSKIND, *Fundamentals of Microwave Electronics.* McGraw-Hill Book Company, New York, 1964.

[3] CHODOROW, M., and T. WESSEL-BERG, A high-efficiency klystron with distributed interaction. *IRE Trans. Electron Devices,* **ED-8,** 44–55, January 1961.

[4] LaRUE, A. D., and R. R. RUBERT, Multi-Megawatt Hybrid TWT's at *S*-band and *C*-band. Presented to the IEEE Electron Devices Meeting, Washington, D.C., October 1964.

[5] HIESLMAIR, H., et al., State of the art of solid-state and tube transmitters. *Microwave J.,* **26,** No. 10, 46–48, October 1983.

[6] FEENBERG, E., Notes on velocity modulation. *Sperry Gyroscope Laboratories Report. 5521–1043,* Chapter I, 41–44.

[7] STAPRANS, A., et al., High-power linear beam tubes. *Proc. IEEE,* **61,** No. 3, 299–330. March 1973.

[8] LIEN, E. L., High efficiency klystron amplifier. In *Conv. Rec. MOGA 70* (8th Int. Conf., Amsterdam, September 1970).

[9] MERDINIAN, G., and J. V. LEBACQZ, High power, permanent magnet focused, *S*-band klystron for linear accelerator use. *Proc. 5th Int. Conf. of Hyper-frequency Tubes* (Paris, September 1964).

[10] BECK, A. H. W., Space charge wave and slow electromagnetic waves. p. 106, Pergamon Press, New York, 1985.

[11] KOMPFNER, R., The traveling-wave tube as amplifier at microwaves. *Proc. IRE,* **35,** 124–127, February 1947.

[12] CHODOROW, M., and R. A. CRAIG, Some new circuits for high-power traveling-wave tubes. *Proc. IRE,* **45,** 1106–1118, August 1957.

[13] ROUMBANIS, T., et al., A megawatt *X*-band TWT amplifier with 18% bandwidth. *Proc. High-Power Microwave Tubes Symp.,* Vol. 1 (The Hexagon, Fort Monmouth, N.J., September 25–26, 1962).

[14] RUETZ, A. J., and W. H. YOCOM, High-power traveling-wave tubes for radar systems. *IRE Trans. Mil. Electron.,* **MIL-5,** 39–45, April 1961.

[15] MENDEL, J. T., Helix and coupled-cavity traveling-wave tubes. *Proc. IEEE,* **61,** No. 3, 280–298, March 1973.

[16] BRILLOUIN, L., *Wave Propagation in Periodic Structures,* 2nd ed. Dover, New York, 1953.

[17] CURNOW, H. J., A general equivalent circuit for coupled-cavity slow-wave structures. *IEEE Trans. on Microwave Theory and Tech.,* **MTT-13,** 671–675, September 1965.

[18] LIAO, S. Y., The effect of collector voltage overdepression on tube performance of the gridded traveling-wave tubes. Report for Hughes Aircraft Company, El Segundo, Calif., August 1977.

SUGGESTED READINGS

COLLIN, R. E., *Foundations for Microwave Engineering,* Chapter 9. McGraw-Hill Book Company, New York, 1966.

GANDHI, O. P., *Microwave Engineering and Applications,* Chapters 9, 10, and 11. Pergamon Press, New York, 1981.

GEWARTOWSKI, J. W., and H. A. WATSON, *Principles of Electron Tubes,* Chapters 5, 6, and 10 to 13. D. Van Nostrand Company, Princeton, N.J., 1965.

GILMOUR, A. S., JR., *Microwave Tubes.* Artech House, Dedham, Mass., 1986.

IEEE Proceedings, **61,** No. 3, March 1973. Special issue on high-power microwave tubes.

LIAO, S. Y., *Microwave Electron Tubes.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.

PROBLEMS

Vacuum Tubes

9-1. A vacuum pentode tube has five grids: a cathode, a control grid, a screen grid, a suppressor grid, and an anode plate as shown in Fig. P9-1.
   a. Sketch the equivalent circuit.
   b. Derive an expression for the input impedance Z_{in} in terms of the angular frequency ω and the circuit parameters.
   c. Determine the transit-angle effect.
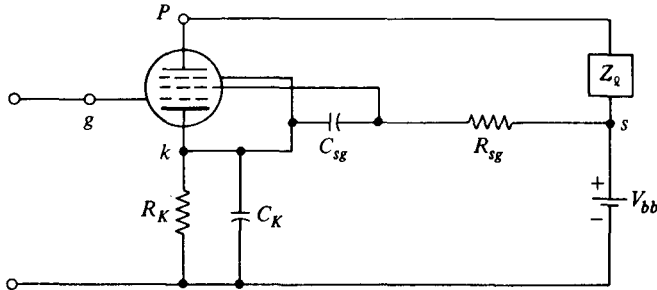
Figure P9-1

Klystrons

9-2. The parameters of a two-cavity amplifier klystron are as follows:

Beam voltage:	$V_0 = 1200$ V
Beam current:	$I_0 = 28$ mA
Frequency:	$f = 8$ GHz
Gap spacing in either cavity:	$d = 1$ mm
Spacing between the two cavities:	$L = 4$ cm
Effective shunt resistance:	$R_{sh} = 40$ kΩ (excluding beam loading)

   a. Find the input microwave voltage V_1 in order to generate a maximum output voltage V_2 (including the finite transit-time effect through the cavities).
   b. Determine the voltage gain (neglecting the beam loading in the output cavity).
   c. Calculate the efficiency of the amplifier (neglecting the beam loading).
   d. Compute the beam loading conductance and show that one may neglect it in the preceding calculations.

9-3. A two-cavity amplifier klystron has the following characteristics:

Voltage gain:	15 dB
Input power:	5 mW
Total shunt impedance of input cavity R_{sh}:	30 kΩ
Total shunt impedance of output cavity R_{sh}:	40 kΩ
Load impedance at output cavity R_ℓ:	40 kΩ

Determine:
a. The input voltage (rms)
b. The output voltage (rms)
c. The power delivered to the load in watts

9-4. A two-cavity amplifier klystron has the following parameters:

Beam voltage:	$V_0 = 900$ V
Beam current:	$I_0 = 30$ mA
Frequency:	$f = 8$ GHz
Gap spacing in either cavity:	$d = 1$ mm
Spacing between centers of cavities:	$L = 4$ cm
Effective shunt impedance:	$R_{sh} = 40$ kΩ

Determine:
a. The electron velocity
b. The dc electron transit time
c. The input voltage for maximum output voltage
d. The voltage gain in decibels

9-5. Derive Eq. (9-2-49).

Multicavity Klystrons

9-6. A four-cavity klystron amplifier has the following parameters:

Beam voltage:	$V_0 = 20$ kV
Beam current:	$I_0 = 2$ A
Operating frequency:	$f = 9$ GHz
dc charge density:	$\rho_0 = 10^{-6}$ C/m^3
RF charge density:	$\rho = 10^{-8}$ C/m^3
Velocity perturbation:	$\mathcal{V} = 10^5$ m/s

Determine:
a. The dc electron velocity
b. The dc phase constant
c. The plasma frequency
d. The reduced plasma frequency for $R = 0.5$
e. The beam current density
f. The instantaneous beam current density

9-7. A four-cavity CW klystron amplifier has the following parameters:

Beam voltage:	$V_0 = 30$ kV
Beam current:	$I_0 = 3$ A
Gap distance:	$d = 1$ cm
Operating frequency:	$f = 8$ GHz
Signal voltage:	$V_1 = 15$ V(rms)
Beam coupling coefficient:	$\beta_i = \beta_0 = 1$
dc electron charge density:	$\rho_0 = 10^{-7}$ C/m^3

Compute:
a. The dc electron velocity
b. The dc electron phase constant
c. The plasma frequency
d. The reduced plasma frequency for $R = 0.4$
e. The reduced plasma phase constant
f. The transit time across the input gap
g. The modulated electron velocity leaving the input gap

9-8. A two-cavity klystron amplifier has the following parameters:

Beam voltage:	$V_0 = 30$ kV
Beam current:	$I_0 = 3$ A
Operating frequency:	$f = 10$ GHz
Beam coupling coefficient:	$\beta_i = \beta_0 = 1$
dc electron charge density:	$\rho_0 = 10^{-7}$ C/m^3
Signal voltage:	$V_1 = 15$ V(rms)
Cavity shunt resistance:	$R_{sh} = 1$ kΩ
Total shunt resistance including load:	$R_{sht} = 10$ kΩ

Calculate:
a. The plasma frequency
b. The reduced plasma frequency for $R = 0.4$
c. The induced current in the output cavity
d. The induced voltage in the output cavity
e. The output power delivered to the load
f. The power gain
g. The electronic efficiency

9-9. A four-cavity klystron amplifier has the following parameters:

Beam voltage:	$V_0 = 20$ kV
Beam current:	$I_0 = 1.5$ A
Operating frequency:	$f = 2$ GHz
Beam coupling coefficient:	$\beta_i = \beta_0 = 1$
dc electron charge density:	$\rho_0 = 10^{-6}$ C/m^3
Signal voltage:	$V_1 = 2$ V(rms)
Cavity shunt resistance:	$R_{sh} = 2$ kΩ
Total shunt resistance including load:	$R_{sht} = 1$ kΩ

Determine:
a. The plasma frequency
b. The reduced plasma frequency for $R = 0.4$
c. The induced current in the output cavity
d. The induced voltage in the output cavity
e. The output power delivered to the load
f. The electronic efficiency

Reflex Klystrons

9-10. A reflex klystron operates at the peak mode of $n = 2$ with

$$\begin{aligned}
\text{Beam voltage:} \quad & V_0 = 300 \text{ V} \\
\text{Beam current:} \quad & I_0 = 20 \text{ mA} \\
\text{Signal voltage:} \quad & V_1 = 40 \text{ V}
\end{aligned}$$

Determine:
a. The input power in watts
b. The output power in watts
c. The effficiency

9-11. A reflex klystron operates under the following conditions:

$$\begin{aligned}
V_0 &= 500 \text{ V} \\
R_{sh} &= 20 \text{ k}\Omega \\
f_r &= 8 \text{ GHz} \\
L &= 1 \text{ mm is the spacing between repeller and cavity}
\end{aligned}$$

The tube is oscillating at f_r at the peak of the $n = 2$ mode or $1\frac{3}{4}$ mode. Assume that the transit time through the gap and the beam loading effect can be neglected.
a. Find the value of repeller voltage V_r.
b. Find the direct current necessary to give microwave gap voltage of 200 V.
c. Calculate the electronic efficiency.

9-12. A reflex klystron operates at the peak of the $n = 2$ mode. The dc power input is 40 mW and $V_1/V_0 = 0.278$. If 20% of the power delivered by the beam is dissipated in the cavity walls, find the power delivered to the load.

9-13. A reflex klystron operates at the peak of the $n = 1$ or $\frac{3}{4}$ mode. The dc power input is 40 mW and the ratio of V_1 over V_0 is 0.278.
a. Determine the efficiency of the reflex klystron.
b. Find the total output power in milliwatts.
c. If 20% of the power delivered by the electron beam is dissipated in the cavity walls, find the power delivered to the load.

Traveling-Wave Tubes (TWTs)

9-14. A traveling-wave tube (TWT) has the following characteristics:

$$\begin{aligned}
\text{Beam voltage:} \quad & V_0 = 2 \text{ kV} \\
\text{Beam current:} \quad & I_0 = 4 \text{ mA} \\
\text{Frequency:} \quad & f = 8 \text{ GHz} \\
\text{Circuit length:} \quad & N = 50 \\
\text{Characteristic impedance:} \quad & Z_0 = 20 \text{ }\Omega
\end{aligned}$$

Determine:
a. The gain parameter C
b. The power gain in decibels

9-15. A TWT operates under the following parameters:

Beam current: $I_0 = 50$ mA
Beam voltage: $V_0 = 2.5$ kV
Characteristic impedance of helix: $Z_0 = 6.75$ Ω
Circuit length: $N = 45$
Frequency: $f = 8$ GHz

Determine:
a. The gain parameter C
b. The output power gain A_p in decibels
c. All four propagation constants
d. The wave equations for all four modes in exponential form

9-16. An O-type traveling-wave tube operates at 2 GHz. The slow-wave structure has a pitch angle of 5.7°. Determine the propagation constant of the traveling wave in the tube. It is assumed that the tube is lossless.

9-17. An O-type helix traveling-wave tube operates at 8 GHz. The slow-wave structure has a pitch angle of 4.4° and an attenuation constant of 2 Np/m. Determine the propagation constant γ of the traveling wave in the tube.

9-18. In an O-type traveling-wave tube, the acceleration voltage (beam voltage) is 3000 V. The characteristic impedance is 10 Ω. The operating frequency is 10 GHz and the beam current is 20 mA. Determine the propagation constants of the four modes of the traveling waves.

9-19. Describe the structure of an O-type traveling-wave tube and its characteristics; then explain how it works.

9-20. In an O-type traveling-wave tube, the acceleration voltage is 4000 V and the magnitude of the axial electric field is 4 V/m. The phase velocity on the slow-wave structure is 1.10 times the average electron beam velocity. The operating frequency is 2 GHz. Determine the magnitude of velocity fluctuation.

Gridded Traveling-Wave Tubes (GTWTs)

9-21. The current I_r caused by the returning electrons at an overdepression voltage of -11.5 kV in a GTWT is about 0.973 A as shown by the spent beam curve in Fig. 9-7-4.
a. Calculate the number of electrons returned per second.
b. Determine the energy in electron volts associated with these returning electrons in 1 ms for part (a).
c. Find the power in watts for the returning electrons.

9-22. The output iron pole piece of a GTWT has the following characteristics:

1. The specific heat H at 20°C is 0.108 cal/g°C.
2. A factor for converting joules to calories is 0.238.
3. The mass of the heated iron pole piece is assumed to be 203.05 mg.
4. The duration time t of the collector depression transient voltage of -11.5 kV is 15 ms.
5. The melting point of iron is 1535°C.

Using these characteristics,

a. Calculate the heat in calories associated with the returning electrons at an overdepression voltage of -11.5 kV.

b. Compute the temperature T in degrees Celsius for the output iron pole piece. [*Hint:* $T = 0.238VIt/(\text{mass} \times \text{specific heat})$.]

c. Determine whether the output iron pole piece is melted.

9-23. The efficiency of a GTWT is expressed as

$$\eta = \frac{RFP_{ac}}{P_{dc}} = \frac{RFP_{ac}}{V_0 I_0}$$

If the cathode voltage is -18 kV and the collector voltage is depressed to -7.5 kV determine the efficiency of the GTWT.

Chapter 10

Microwave Crossed-Field Tubes (*M* Type)

10-0 INTRODUCTION

In the previous chapter, several commonly used linear-beam tubes were described in detail. In these tubes, the dc magnetic field that is in parallel with the dc electric field is used merely to focus the electron beam. In crossed-field devices, however, the dc magnetic field and the dc electric field are perpendicular to each other. In all crossed-field tubes, the dc magnetic field plays a direct role in the RF interaction process.

Crossed-field tubes derive their name from the fact that the dc electric field and the dc magnetic field are perpendicular to each other. They are also called *M*-type tubes after the French TPOM (*tubes à propagation des ondes à champs magnétique: tubes for propagation of waves in a magnetic field*). In a crossed-field tube, the electrons emitted by the cathode are accelerated by the electric field and gain velocity, but the greater their velocity, the more their path is bent by the magnetic field. If an RF field is applied to the anode circuit, those electrons entering the circuit during the retarding field are decelerated and give up some of their energy to the RF field. Consequently, their velocity is decreased, and these slower electrons will then travel the dc electric field far enough to regain essentially the same velocity as before. Because of the crossed-field interactions, only those electrons that have given up sufficient energy to the RF field can travel all the way to the anode. This phenomenon would make the *M*-type devices relatively efficient. Those electrons entering the circuit during the accelerating field are accelerated by means of receiving enough energy from the RF field and are returned back toward the cathode. This back-bombardment of the cathode produces heat in the cathode and decreases the operational efficiency.

In this chapter, several commonly used crossed-field tubes such as magnetrons, forward-wave crossed-field amplifiers (FWCFAs), backward-wave crossed-field amplifiers (BWCFAs or amplitrons), and backward-wave crossed-field oscillators (BWCFOs or carcinotrons) are studied.

Cylindrical magnetron: The cylindrical magnetron was developed by Boot and Randall in early 1940.

Coaxial magnetron: The coaxial magnetron introduced the principle of integrating a stabilizing cavity into the magnetron geometry.

Voltage-tunable magnetron: The voltage-tunable magnetron has the cathode-anode geometry of the conventional magnetron, but its anode can be tuned easily.

Inverted magnetron: The inverted magnetron has the inverted geometry of the conventional magnetron with the cathode placed on the outside surrounding the anode and microwave circuit.

Forward-wave crossed-field amplifier (FWCFA): The forward-wave crossed-field amplifier is also called the *M*-type forward-wave amplifier.

Backward-wave crossed-field amplifier (BWCFA): The backward-wave crossed-field amplifier was developed by Raytheon Company in 1960 and its trade name is Amplitron. It is a broadband, high-power, high-gain, and high-efficiency microwave tube, and it has many applications such as in airborne radar systems and spaceborne communications systems.

Backward-wave crossed-field oscillator (BWCFO): In 1960, the French Company developed a new type of crossed-field device in which an injection gun replaced the conventional cylindrical cathode of the magnetron. Its trade name is Carcinotron, and it is also called the *M-type backward-wave oscillator*.

All of these crossed-field electron tubes are tabulated in Table 10-0-1.
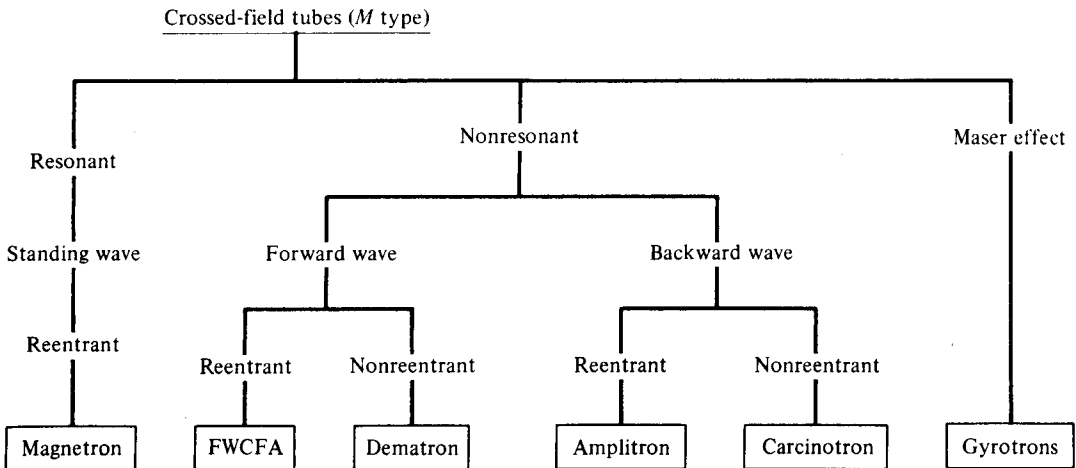
10-1 MAGNETRON OSCILLATORS

Hull invented the magnetron in 1921 [1], but it was only an interesting laboratory device until about 1940. During World War II, an urgent need for high-power microwave generators for radar transmitters led to the rapid development of the magnetron to its present state.

All magnetrons consist of some form of anode and cathode operated in a dc magnetic field normal to a dc electric field between the cathode and anode. Because of the crossed field between the cathode and anode, the electrons emitted from the cathode are influenced by the crossed field to move in curved paths. If the dc magnetic field is strong enough, the electrons will not arrive in the anode but return instead to the cathode. Consequently, the anode current is cut off. Magnetrons can be classified into three types:

1. *Split-anode magnetron:* This type of magnetron uses a static negative resistance between two anode segments.
2. *Cyclotron-frequency magnetrons:* This type operates under the influence of synchronism between an alternating component of electric field and a periodic oscillation of electrons in a direction parallel to the field.
3. *Traveling-wave magnetrons:* This type depends on the interaction of electrons with a traveling electromagnetic field of linear velocity. They are customarily referred to simply as *magnetrons.*

Negative-resistance magnetrons ordinarily operate at frequencies below the microwave region. Although cyclotron-frequency magnetrons operate at frequencies in microwave range, their power output is very small (about 1 W at 3 GHz), and their efficiency is very low (about 10% in the split-anode type and 1% in the single-anode type). Thus, the first two types of magnetrons are not considered in this text. In this section, only the traveling-wave magnetrons such as the cylindrical magnetron, linear (or planar) magnetron, coaxial magnetron, voltage-tunable magnetron, inverted coaxial magnetron, and the frequency-agile magnetron will be discussed.

10-1-1 Cylindrical Magnetron

A schematic diagram of a cylindrical magnetron oscillator is shown in Fig. 10-1-1. This type of magnetron is also called a *conventional magnetron.*

In a cylindrical magnetron, several reentrant cavities are connected to the gaps. The dc voltage V_0 is applied between the cathode and the anode. The magnetic flux density B_0 is in the positive z direction. When the dc voltage and the magnetic flux are adjusted properly, the electrons will follow cycloidal paths in the cathode-anode space under the combined force of both electric and magnetic fields as shown in Fig. 10-1-2.

Equations of electron motion.   The equations of motion for electrons in a cylindrical magnetron can be written with the aid of Eqs.(1-2-5a) and (1-2-5b) as
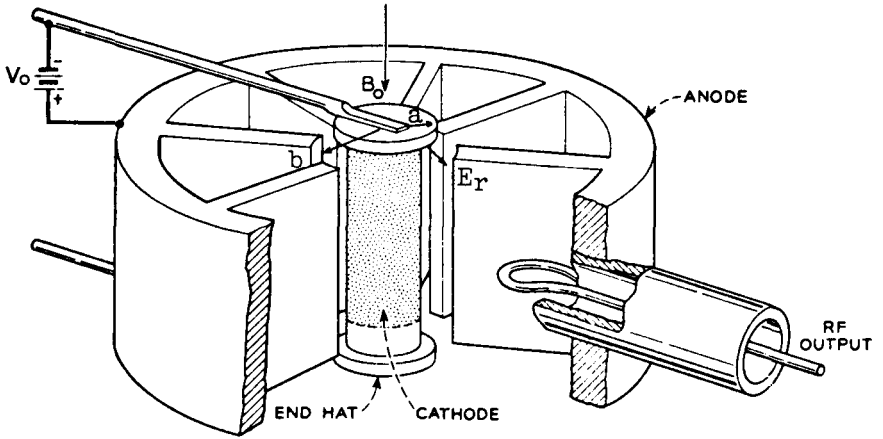
Figure 10-1-1 Schematic diagram of a cylindrical magnetron.
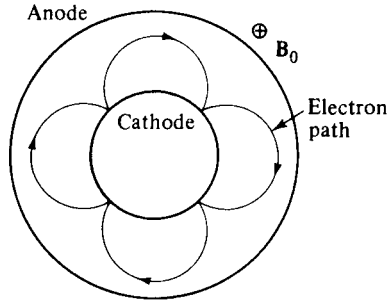
Figure 10-1-2 Electron path in a cylindrical magnetron.

$$\frac{d^2 r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 = \frac{e}{m}E_r - \frac{e}{m}rB_z\frac{d\phi}{dt} \qquad (10\text{-}1\text{-}1)$$

$$\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = \frac{e}{m}B_z\frac{dr}{dt} \qquad (10\text{-}1\text{-}2)$$

where $\frac{e}{m} = 1.759 \times 10^{11}$ C/kg is the charge-to-mass ratio of the electron and $B_0 = B_z$ is assumed in the positive z direction.

Rearrangement of Eq. (10-1-2) results in the following form

$$\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = \frac{e}{m}B_z r\frac{dr}{dt} = \frac{1}{2}\omega_c\frac{d}{dt}(r^2) \qquad (10\text{-}1\text{-}3)$$

where $\omega_c = \frac{e}{m}B_z$ is the cyclotron angular frequency. Integration of Eq. (10-1-3) yields

$$r^2\frac{d\phi}{dt} = \frac{1}{2}\omega_c r^2 + \text{constant} \qquad (10\text{-}1\text{-}4)$$

at $r = a$, where a is the radius of the cathode cylinder, and $\frac{d\phi}{dt} = 0$, constant $= -\frac{1}{2}\omega_c a^2$. The angular velocity is expressed by

$$\frac{d\phi}{dt} = \frac{1}{2}\omega_c\left(1 - \frac{a^2}{r^2}\right) \tag{10-1-5}$$

Since the magnetic field does no work on the electrons, the kinetic energy of the electron is given by

$$\frac{1}{2}m\mathcal{V}^2 = eV \tag{10-1-6}$$

However, the electron velocity has r and ϕ components such as

$$\mathcal{V}^2 = \frac{2e}{m}V = \mathcal{V}_r^2 + \mathcal{V}_\phi^2 = \left(\frac{dr}{dt}\right)^2 + \left(r\frac{d\phi}{dt}\right)^2 \tag{10-1-7}$$

at $r = b$, where b is the radius from the center of the cathode to the edge of the anode, $V = V_0$, and $dr/dt = 0$, when the electrons just graze the anode, Eqs. (10-1-5) and (10-1-7) become

$$\frac{d\phi}{dt} = \frac{1}{2}\omega_c\left(1 - \frac{a^2}{b^2}\right) \tag{10-1-8}$$

$$b^2\left(\frac{d\phi}{dt}\right)^2 = \frac{2e}{m}V_0 \tag{10-1-9}$$

Substitution of Eq. (10-1-8) into Eq. (10-1-9) results in

$$b^2\left[\frac{1}{2}\omega_c\left(1 - \frac{a^2}{b^2}\right)\right]^2 = \frac{2e}{m}V_0 \tag{10-1-10}$$

The electron will acquire a tangential as well as a radial velocity. Whether the electron will just graze the anode and return toward the cathode depends on the relative magnitudes of V_0 and B_0. The *Hull cutoff magnetic equation* is obtained from Eq. (10-1-10) as

$$B_{0c} = \frac{\left(8V_0\frac{m}{e}\right)^{1/2}}{b\left(1 - \frac{a^2}{b^2}\right)} \tag{10-1-11}$$

This means that if $B_0 > B_{0c}$ for a given V_0, the electrons will not reach the anode. Conversely, the cutoff voltage is given by

$$V_{0c} = \frac{e}{8m}B_0^2 b^2\left(1 - \frac{a^2}{b^2}\right)^2 \tag{10-1-12}$$

This means that if $V_0 < V_{0c}$ for a given B_0, the electrons will not reach the anode. Equation (10-1-12) is often called the *Hull cutoff voltage equation*.

Example 10-1-1: Conventional Magnetron

An X-band pulsed cylindrical magnetron has the following operating parameters:

Anode voltage: $V_0 = 26$ kV
Beam current: $I_0 = 27$ A
Magnetic flux density: $B_0 = 0.336$ Wb/m^2
Radius of cathode cylinder: $a = 5$ cm
Radius of vane edge to center: $b = 10$ cm

Compute:

a. The cyclotron angular frequency
b. The cutoff voltage for a fixed B_0
c. The cutoff magnetic flux density for a fixed V_0

Solution

a. The cyclotron angular frequency is

$$\omega_c = \frac{e}{m} B_0 = 1.759 \times 10^{11} \times 0.336 = 5.91 \times 10^{10} \text{ rad}$$

b. The cutoff voltage for a fixed B_0 is

$$V_{0c} = \frac{1}{8} \times 1.759 \times 10^{11} (0.336)^2 (10 \times 10^{-2})^2 \left(1 - \frac{5^2}{10^2}\right)^2$$

$$= 139.50 \text{ kV}$$

c. The cutoff magnetic flux density for a fixed V_0 is

$$B_{0c} = \left(8 \times 26 \times 10^3 \times \frac{1}{1.759 \times 10^{11}}\right)^{1/2} \left[10 \times 10^{-2}\left(1 - \frac{5^2}{10^2}\right)\right]^{-1}$$

$$= 14.495 \text{ mWb/m}^2$$

Cyclotron angular frequency.   Since the magnetic field is normal to the motion of electrons that travel in a cycloidal path, the outward centrifugal force is equal to the pulling force. Hence

$$\frac{m \mathcal{V}^2}{R} = e \mathcal{V} B \tag{10-1-13}$$

where R = radius of the cycloidal path
$\mathcal{V}$ = tangential velocity of the electron

The cyclotron angular frequency of the circular motion of the electron is then given by

$$\omega_c = \frac{\mathcal{V}}{R} = \frac{eB}{m} \tag{10-1-14}$$

The period of one complete revolution can be expressed as

$$T = \frac{2\pi}{\omega} = \frac{2\pi m}{eB} \qquad (10\text{-}1\text{-}15)$$

Since the slow-wave structure is closed on itself, or "reentrant," oscillations are possible only if the total phase shift around the structure is an integral multiple of 2π radians. Thus, if there are N reentrant cavities in the anode structure, the phase shift between two adjacent cavities can be expressed as

$$\phi_n = \frac{2\pi m}{N} \qquad (10\text{-}1\text{-}16)$$

where n is an integer indicating the nth mode of oscillation.

In order for oscillations to be produced in the structure, the anode dc voltage must be adjusted so that the average rotational velocity of the electrons corresponds to the phase velocity of the field in the slow-wave structure. Magnetron oscillators are ordinarily operated in the π mode. That is,

$$\phi_n = \pi \qquad (\pi \text{ mode}) \qquad (10\text{-}1\text{-}17)$$

Figure 10-1-3 shows the lines of force in the π mode of an eight-cavity magnetron. It is evident that in the π mode the excitation is largely in the cavities, having opposite phase in successive cavities. The successive rise and fall of adjacent anode-cavity fields may be regarded as a traveling wave along the surface of the slow-wave structure. For the energy to be transferred from the moving electrons to the traveling field, the electrons must be decelerated by a retarding field when they pass through each anode cavity. If L is the mean separation between cavities, the phase constant of the fundamental-mode field is given by

$$\beta_0 = \frac{2\pi n}{NL} \qquad (10\text{-}1\text{-}18)$$

The traveling-wave field of the slow-wave structure may be obtained by solving
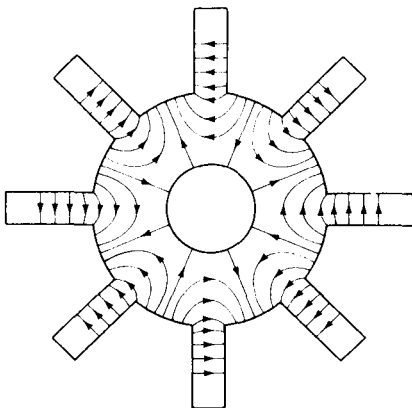
Figure 10-1-3   Lines of force in π mode of eight-cavity magnetron.

Maxwell's equations subject to the boundary conditions. The solution for the fundamental ϕ component of the electric field has the form [1]

$$E_{\phi 0} = jE_1 e^{j(\omega t - \beta_0 \phi)} \tag{10-1-19}$$

where E_1 is a constant and β_0 is given in Eq. (10-1-18). Thus, the traveling field of the fundamental mode travels around the structure with angular velocity

$$\frac{d\phi}{dt} = \frac{\omega}{\beta_0} \tag{10-1-20}$$

where $\frac{d\phi}{dt}$ can be found from Eq. (10-1-19).

When the cyclotron frequency of the electrons is equal to the angular frequency of the field, the interactions between the field and electron occurs and the energy is transferred. That is,

$$\omega_c = \beta_0 \frac{d\phi}{dt} \tag{10-1-21}$$

Power output and efficiency.   The efficiency and power output of a magnetron depend on the resonant structure and the dc power supply. Figure 10-1-4 shows an equivalent circuit for a resonator of a magnetron.

where Y_e = electronic admittance
  V = RF voltage across the vane tips
  C = capacitance at the vane tips
  L = inductance of the resonator
  G_r = conductance of the resonator
  G = load conductance per resonator
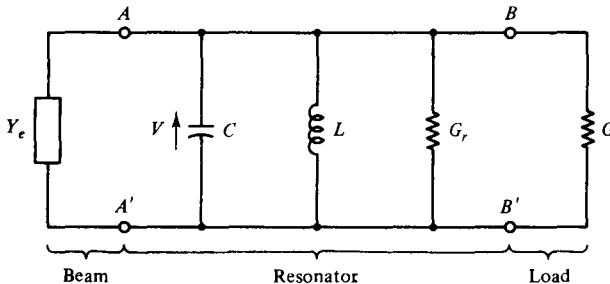
Figure 10-1-4   Equivalent circuit for one resonator of a magnetron.

Each resonator of the slow-wave structure is taken to comprise a separate resonant circuit as shown in Fig. 10-1-4. The unloaded quality factor of the resonator is given by

$$Q_{un} = \frac{\omega_0 C}{G_r} \tag{10-1-22}$$

where $\omega_0 = 2\pi f_0$ is the angular resonant frequency. The external quality factor of the load ciruit is

$$Q_{ex} = \frac{\omega_0 C}{G_\ell} \qquad (10\text{-}1\text{-}23)$$

Then the loaded Q_ℓ of the resonant circuit is expressed by

$$Q_\ell = \frac{\omega_0 C}{G_r + G_\ell} \qquad (10\text{-}1\text{-}24)$$

The circuit efficiency is defined as

$$\eta_c = \frac{G_\ell}{G_\ell + G_r}$$

$$= \frac{G_\ell}{G_{ex}} = \frac{1}{1 + Q_{ex}/Q_{un}} \qquad (10\text{-}1\text{-}25)$$

The maximum circuit efficiency is obtained when the magnetron is heavily loaded, that is, for $G_\ell \gg G_r$. Heavy loading, however, makes the tube quite sensitive to the load, which is undesirable in some cases. Therefore, the ratio of Q_ℓ/Q_{ex} is often chosen as a compromise between the conflicting requirements for high circuit efficiency and frequency stability.

The electronic efficiency is defined as

$$\eta_e = \frac{P_{gen}}{P_{dc}} = \frac{V_0 I_0 - P_{lost}}{V_0 I_0} \qquad (10\text{-}1\text{-}26)$$

where P_{gen} = RF power induced into the anode circuit
$\quad\ P_{dc} = V_0 I_0$ power from the dc power supply
$\quad\ V_0$ = anode voltage
$\quad\ I_0$ = anode current
$\quad\ P_{lost}$ = power lost in the anode circuit

The RF power generated by the electrons can be written as

$$P_{gen} = V_0 I_0 - P_{lost}$$

$$= V_0 I_0 - I_0 \frac{m}{2e} \frac{\omega_0^2}{\beta^2} + \frac{E_{max}^2}{B_z^2} \qquad (10\text{-}1\text{-}27)$$

$$= \frac{1}{2} N \, |V|^2 \frac{\omega_0 C}{Q_\ell}$$

where N = total number of resonators
$\quad\ V$ = RF voltage across the resonator gap
$\quad\ E_{max} = M_1 |V|/L$ is the maximum electric field
$\quad\ M_1 = \sin\left(\beta_n \frac{\delta}{2}\right)\Big/\left(\beta_n \frac{\delta}{2}\right) = 1$ for small δ is the gap factor for the π-mode
$\qquad\quad$ operation
$\quad\ \beta$ = phase constant
$\quad\ B_z$ = magnetic flux density
$\quad\ L$ = center-to-center spacing of the vane tips

The power generated may be simplified to

$$P_{gen} = \frac{NL^2\omega_0 C}{2M_1^2 Q_\ell} E_{max}^2 \tag{10-1-28}$$

The electronic efficiency may be rewritten as

$$\eta_e = \frac{P_{gen}}{V_0 I_0} = \frac{1 - \dfrac{m\omega_0^2}{2eV_0\beta^2}}{1 + \dfrac{I_0 m M_1^2 Q_\ell}{B_z e N L^2 \omega_0 C}} \tag{10-1-29}$$

Example 10-1-1A: Pulsed Magnetron

An X-band pulsed conventional magnetron has the following operating parameters:

Anode voltage:	$V_0 = 5.5$ kV
Beam current:	$I_0 = 4.5$ A
Operating frequency:	$f = 9 \times 10^9$ Hz
Resonator conductance:	$G_r = 2 \times 10^{-4}$ mho
Loaded conductance:	$G_\ell = 2.5 \times 10^{-5}$ mho
Vane capacitance:	$C = 2.5$ pF
Duty cycle:	$DC = 0.002$
Power loss:	$P_{loss} = 18.50$ kW

Compute:

a. The angular resonant frequency
b. The unloaded quality factor
c. The loaded quality factor
d. The external quality factor
e. The circuit efficiency
f. The electronic efficiency

Solution

a. The angular resonant frequency is

$$\omega_r = 2 \times 9 \times 10^9 = 56.55 \times 10^9 \text{ rad}$$

b. The unloaded quality factor is

$$Q_{un} = \frac{56.55 \times 10^9 \times 2.5 \times 10^{-12}}{2 \times 10^{-4}} = 707$$

c. The loaded quality factor is

$$Q_l = \frac{56.55 \times 10^9 \times 2.5 \times 10^{-12}}{2 \times 10^{-4} + 2.5 \times 10^{-5}} = 628$$

d. The external quality factor is

$$Q_{ex} = \frac{56.55 \times 10^9 \times 2.5 \times 10^{-12}}{2.5 \times 10^{-5}} = 5655$$

e. The circuit efficiency is

$$\eta_c = \frac{1}{1 + 5655/707} = 11.11\%$$

f. The electronic efficiency is

$$\eta_e = \frac{5.5 \times 10^3 \times 4.5 - 18.5 \times 10^3}{5.5 \times 10^3 \times 4.5} = 25.25\%$$

State of the art.   For many years, magnetrons have been the high-power sources in operating frequencies as high as 70 GHz. Military radar relies on conventional traveling-wave magnetrons to generate high-peak-power RF pulses. No other microwave devices can perform the same function with the same size, weight, voltage, and efficiency-range advantage as can the conventional magnetrons. At the present state of the art, a magnetron can deliver a peak power output of up to 40 MW with the dc voltage in the order of 50 kV at the frequency of 10 GHz. The average power outputs are up to 800 kW. Its efficiency is very high, ranging from 40 to 70%. Figure 10-1-5 shows the state of the art for U.S. high-power magnetrons.
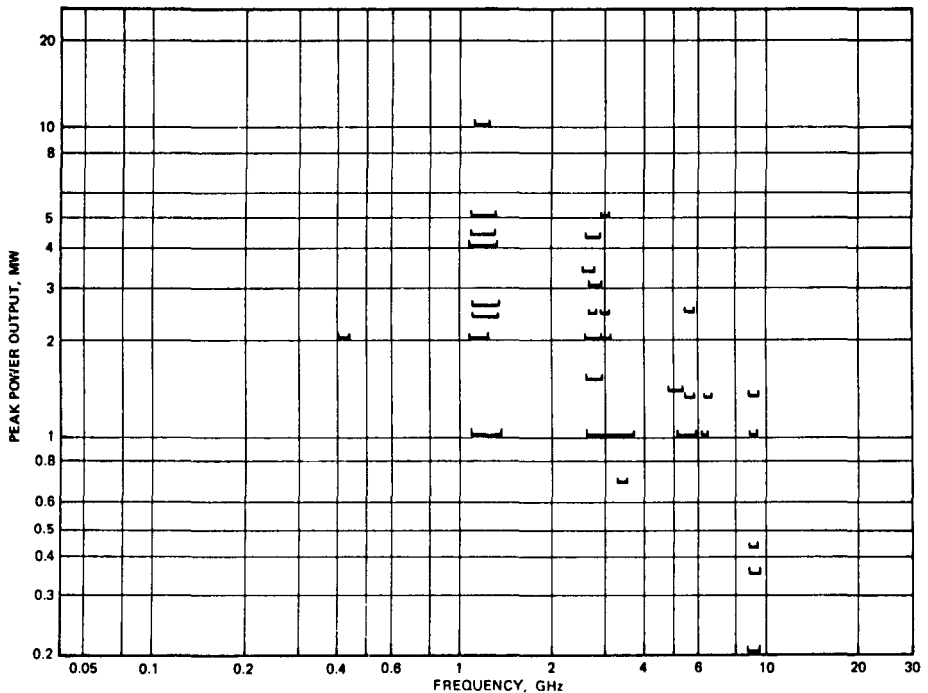
Figure 10-1-5   State of the art for U.S. high-power magnetrons.

Beacon magnetrons—miniature conventional magnetrons—deliver peak outputs as high as 3.5 kW, yet weigh less than two pounds. These devices are ideal for use where a very compact, low-voltage source of pulsed power is required, such as in airborne, missile, satellite, or Doppler systems. Most of the beacon magnetrons exhibit negligible frequency shift and provide long-life performance under the most severe environmental and temperature conditions.

The Litton L-5080 pulse magnetron, shown in Fig. 10-1-6, is a typical vane-strap magnetron oscillator. It has a maximum peak output power of 250 kW at a frequency range from 5.45 to 5.825 Ghz. Its duty cycle is 0.0012.
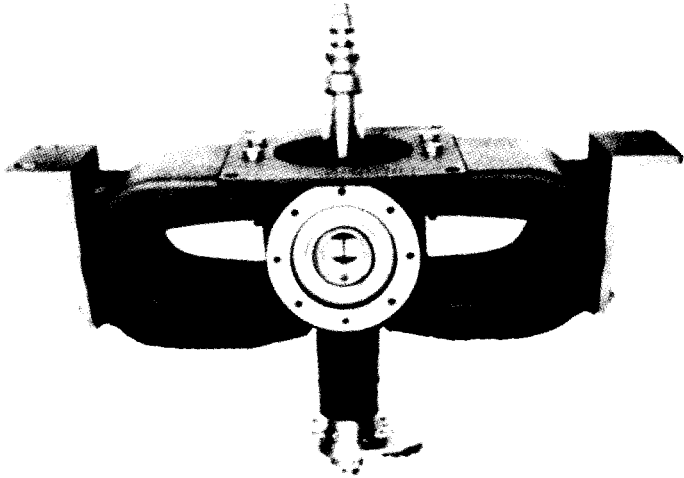
Figure 10-1-6   Photograph of Litton L-5080 magnetron. (*Courtesy of Litton Electron Tube Division.*)

10-1-2 Linear Magnetron

The schematic diagram of a linear magnetron is shown in Fig. 10-1-7. In the linear magnetron as shown in Fig. 10-1-7, the electric field E_x is assumed in the positive x direction and the magnetic flux density B_z in the positive z direction. The differential equations of motion of electrons in the crossed-electric and magnetic fields can be written from Eqs. (1-3-2a, b, c) as

$$\frac{d^2x}{dt^2} = -\frac{e}{m}\left(E_x + B_z \frac{dy}{dt}\right) \qquad (10\text{-}1\text{-}30)$$
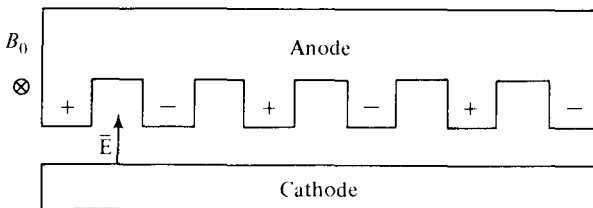
Figure 10-1-7   Schematic diagram of a linear magnetron.

$$\frac{d^2y}{dt^2} = \frac{e}{m}B_z\frac{dx}{dt} \tag{10-1-31}$$

$$\frac{d^2z}{dt^2} = 0 \tag{10-1-32}$$

where $\dfrac{e}{m} = 1.759 \times 10^{11}$ C/kg is the charge-to-mass of an electron

B_z = magnetic flux density in positive z direction

E_x = electric field in positive x direction

In general, the presence of space charges causes the field to be a nonlinear function of the distance x, and the complete solution of Eqs. (10-1-30) through (10-1-32) is not simple. Equation (10-1-31), however, can be integrated directly. Under the assumptions that the electrons emit from the cathode surface with zero initial velocity and that the origin is the cathode surface, Eq. (10-1-31) becomes

$$\frac{dy}{dt} = \frac{e}{m}B_z x \tag{10-1-33}$$

Equation (10-1-33) shows that, regardless of space charges, the electron velocity parallel to the electrode surface is proportional to the distance of the electron from the cathode and to the magnetic flux density B_z. How far the electron moves from the cathode depends on B_z and on the manner in which the potential V varies with x, which in turn depends on the space-charge distribution, anode potential, and electrode spacing.

If the space charge is assumed to be negligible, the cathode potential zero, and the anode potential V_0, the differential electric field becomes

$$\frac{dV}{dx} = \frac{V_0}{d} \tag{10-1-34}$$

where V_0 = anode potential in volts

d = distance between cathode and anode in meters

Substitution of Eq. (10-1-34) into Eq. (10-1-30) yields

$$\frac{d^2x}{dt^2} = \frac{e}{m}\left(\frac{V_0}{d} - B_z\frac{dy}{dt}\right) \tag{10-1-35}$$

Combination of Eqs. (10-1-33) and (10-1-35) results in

$$\frac{d^2x}{dt^2} + \left(\frac{e}{m}B_z\right)^2 x - \frac{e}{m}\frac{V_0}{d} = 0 \tag{10-1-36}$$

Solution of Eq. (10-1-36) and substitution of the solution into Eq. (10-1-33) yield the following equations for the path of an electron with zero velocity at cathode (origin point) as

$$x = \frac{V_0}{B_z\omega_c d}[1 - \cos(\omega_c t)] \tag{10-1-37}$$

$$y = \frac{V_0}{B_z \omega_c d}[\omega_c t - \sin(\omega_c t)] \tag{10-1-38}$$

$$z = 0 \tag{10-1-39}$$

where $\omega_c = \frac{e}{m}B_z$ is the cyclotron angular frequency

$f_c = 2.8 \times 10^6 B_z$ is the cyclotron frequency in Hz

Equations (10-1-37) through (10-1-39) are those of a cycloid generated by a point on a circle of radius $V_0/(B_z \omega_c d)$ rolling on the plane of the cathode with angular frequency ω_c. The maximum distance to which the electron moves in a direction normal to the cathode is $2V_0 m/(B_z^2 ed)$. When this distance is just equal to the anode-cathode distnce d, the electrons graze the anode surface, and the anode current is cut off. Then the cutoff condition is

$$\frac{2V_0 m}{B_z^2 ed} = d \tag{10-1-40}$$

Let a constant K equal to

$$K = \frac{d^2 B_z^2}{V_0} = \frac{2m}{e} = 1.14 \times 10^{-11} \tag{10-1-41}$$

When the value of K is less than 1.14×10^{-11}, electrons strike the anode; when the value is larger than 1.14×10^{-11}, they return to the cathode. Figure 10-1-8 shows the electron path.
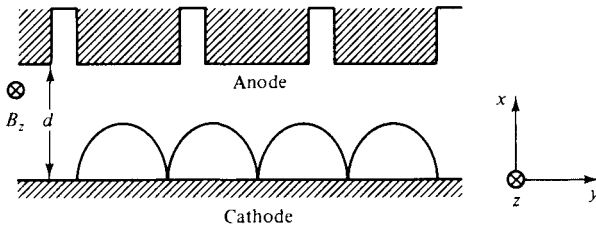
Figure 10-1-8   Electron path in a linear magnetron.

From Eq. (10-1-40), the Hull cutoff voltage for a linear magnetron is given by

$$V_{0c} = \frac{1}{2}\frac{e}{m}B_0^2 d^2 \tag{10-1-42}$$

where $B_0 = B_z$ is the magnetic flux density in the positive z direction. This means that if $V_0 < V_{0c}$ for a given B_0, the electrons will not reach the anode.

Similarly, the Hull cutoff magnetic flux density for a linear magnetron is expressed as

$$B_{0c} = \frac{1}{d}\sqrt{2\frac{m}{e}V_0} \tag{10-1-43}$$

This means that if $B_0 > B_{0c}$ for a given V_0, the electrons will not reach the anode.

Example 10-1-2:   Linear Magnetron

A linear magnetron has the following operating parameters:

Anode voltage:	$V_0 = 10$ kV
Cathode current:	$I_0 = 1$ A
Magnetic flux density:	$B_0 = 0.01$ Wb/m^2
Distance between cathode and anode:	$d = 5$ cm

Compute:

    a. The Hull cutoff voltage for a fixed B_0
    b. The Hull cutoff mangetic flux density for a fixed V_0

Solution

    a. The Hull cutoff voltage is

$$V_{0c} = \tfrac{1}{2} \times 1.759 \times 10^{11} \times (0.01)^2 \times (5 \times 10^{-2})^2$$

$$= 22.00 \text{ kV}$$

    b. The Hull magnetic flux density is

$$B_{0c} = \frac{1}{5 \times 10^{-2}} \times \left(\frac{2 \times 10 \times 10^3}{1.759 \times 10^{11}} \right)^{1/2}$$

$$= 6.74 \text{ mWb/m}^2$$

Hartree condition.   The Hull cutoff condition determines the anode voltage or magnetic field necessary to obtain nonzero anode current as a function of the magnetic field or anode voltage in the absence of an electromagnetic field. The Hartree condition can be derived as follows and as shown in Fig. 10-1-9.

The electron beam lies within a region extending a distance h from the cathode, where h is known as the hub thickness. The spacing between the cathode and anode is d. The electron motion is assumed to be in the positive y direction with a velocity

$$\mathcal{V}_y = -\frac{E_x}{B_0} = \frac{1}{B_0} \frac{dV}{dx} \tag{10-1-44}$$
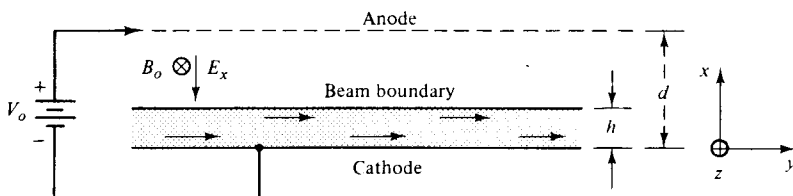
Figure 10-1-9   Linear model of a magnetron.

where $B_0 = B_z$ is the magnetic flux density in the positive z direction
  V = potential

From the principle of energy conservation, we have

$$\tfrac{1}{2} m \, \mathcal{V}_y^2 = eV \tag{10-1-45}$$

Combining Eqs. (10-1-44) and (10-1-45) yields

$$\left(\frac{dV}{dx}\right)^2 = \frac{2eV}{m} B_0^2 \tag{10-1-46}$$

This differential equation may be rearranged as

$$\left(\frac{m}{2eB_0}\right)^{1/2} \frac{dV}{\sqrt{V}} = dx \tag{10-1-47}$$

Integration of Eq. (10-1-47) yields the potential within the electron beam as

$$V = \frac{eB_0^2}{2m} x^2 \tag{10-1-48}$$

where the constant of integration has been eliminated for $V = 0$ at $x = 0$. The potential and electric field at the hub surface are given by

$$V(h) = \frac{e}{2m} B_0^2 h^2 \tag{10-1-49}$$

and

$$E_x = -\frac{dV}{dx} = -\frac{\epsilon}{m} B_0^2 h \tag{10-1-50}$$

The potential at the anode is thus obtained from Eq. (10-1-50) as

$$V_0 = -\int_0^d E_x \, dx$$

$$= -\int_0^h E_x \, dx - \int_h^d E_x \, dx$$

$$= V(h) + \frac{e}{m} B_0^2 h (d - h)$$

$$= \frac{e}{m} B_0^2 h (d - h/2) \tag{10-1-51}$$

The electron velocity at the hub surface is obtained from Eqs. (10-1-44) and (10-1-50) as

$$\mathcal{V}_y(h) = \frac{e}{m} B_0 h \tag{10-1-52}$$

For synchronism, this electron velocity is equal to the phase velocity of the slow-wave structure. That is,

$$\frac{\omega}{\beta} = \frac{e}{m} B_0 h \tag{10-1-53}$$

For the π-mode operation, the anode potential is finally given by

$$V_{0h} = \frac{\omega B_0 d}{\beta} - \frac{m}{2e} \frac{\omega^2}{\beta^2} \tag{10-1-54}$$

This is the Hartree anode voltage equation that is a function of the magnetic flux density and the spacing beween the cathode and anode.

Example 10-1-2a: Linear Magnetron

A linear magnetron has the following operation parameters:

Anode voltage:	$V_0 = 15$ kV
Cathode current:	$I_0 = 1.2$ A
Operating frequency:	$f = 8$ GHz
Magnetric flux density:	$B_0 = 0.015$ Wb/m^2
Hub thickness:	$h = 2.77$ cm
Distance between anode and cathode:	$d = 5$ cm

Calculate:

a. The electron velocity at the hub surface
b. The phase velocity for synchronism
c. The Hartree anode voltage

Solution

a. The electron velocity is

$$\mathcal{V} = 1.759 \times 10^{11} \times 0.015 \times 2.77 \times 10^{-2}$$

$$= 0.73 \times 10^3 \text{ m/s}$$

b. The phase velocity is

$$\mathcal{V}_{ph} = \frac{\omega}{\beta} = 0.73 \times 10^8 \text{ m/s}$$

c. The Hartree anode voltage is

$$V_{0h} = 0.73 \times 10^8 \times 0.015 \times 5 \times 10^{-2} - \frac{1}{2 \times 1.759 \times 10^{11}} \times (0.73 \times 10^8)^2$$

$$= 5.475 \times 10^4 - 1.515 \times 10^4$$

$$= 39.60 \text{ kV}$$

10-1-3 Coaxial Magnetron

The coaxial magnetron is composed of an anode resonator structure surrounded by an inner-single, high-Q cavity operating in the TE_{011} mode as shown in Fig. 10-1-10.

The slots in the back walls of alternate cavities of the anode resonator structure tightly couple the electric fields in these resonators to the surrounding cavity. In the

(a) Cross section

(b) Cutaway view
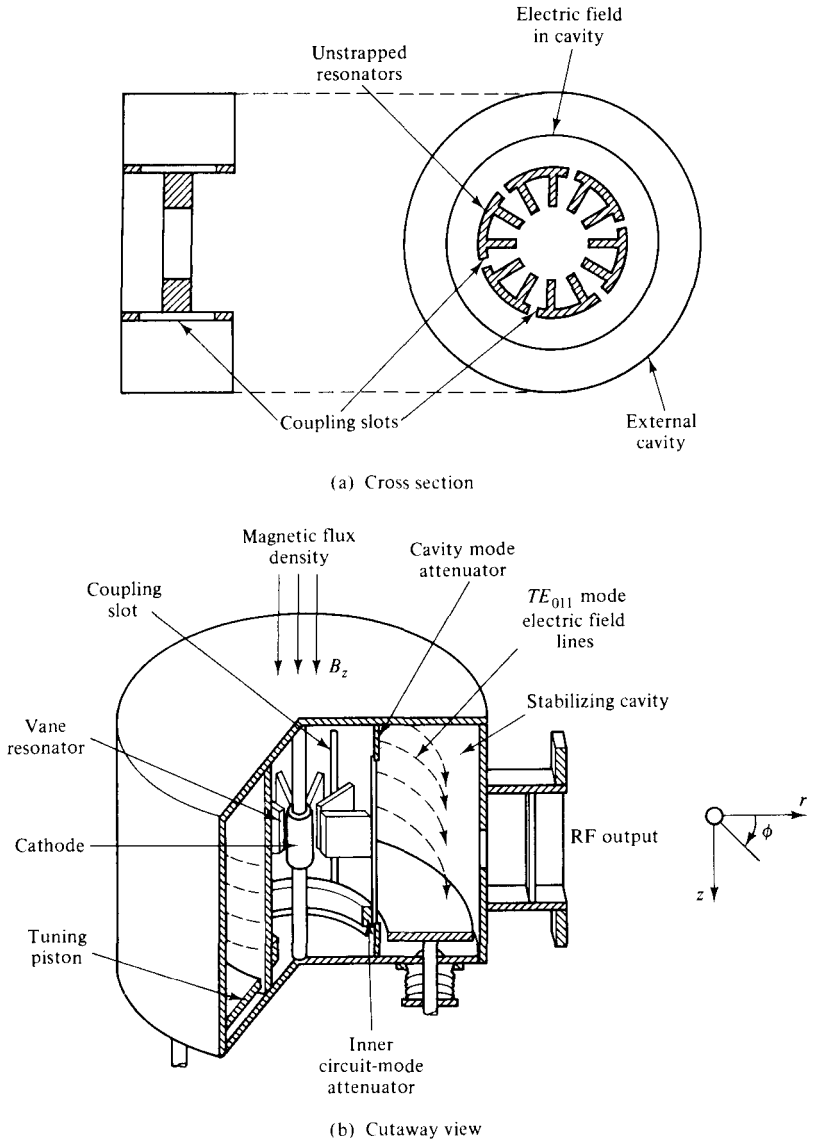
Figure 10-1-10   Schematic diagram of a coaxial magnetron. (*Courtesy of Varian Associates, Inc.*)

π-mode operation, the electric fields in every other cavity are in phase, and so they couple in the same direction into the surrounding cavity. As a result, the surrounding coaxial cavity stabilizes the magnetron in the desired π-mode oepration.

In the desired TE_{011} mode, the electric fields follow a circular path within the cavity and reduce to zero at the walls of the cavity. Current flow in the TE_{011} mode is in the walls of the cavity in circular paths about the axis of the tube. The undesired modes are damped out by the attenuator within the inner slotted cylinder near the ends of the coupling slots. The tuning mechanism is simple and reliable. As the straps are not required, the anode resonator for the coaxial magnetron can be larger and less complex than for the conventional strapped magnetron. Thus cathode loading is lower, and voltage gradients are reduced.

The Varian SFD-333TM magnetron, as shown in Fig. 10-1-11 is a typical X-band coaxial magnetron. It has a minimum peak power of 400 kW at a frequency range from 8.9 to 9.6 GHz. Its duty cycle is 0.0013. The nominal anode voltage is 32 kV, and the peak anode current is 32 A.
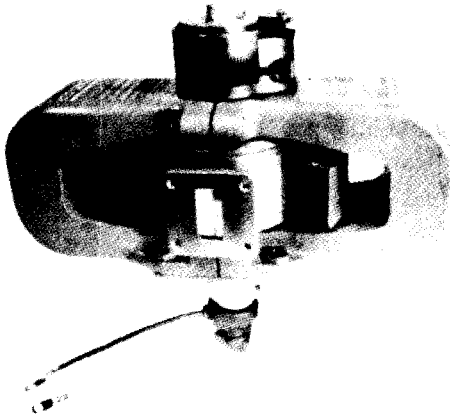
Figure 10-1-11   Photograph of Varian
SFD-333TM coaxial magnetron
(*Courtesy of Varian Associates, Inc.*)

10-1-4 Voltage-Tunable Magnetron

The voltage-tunable magnetron is a broadband oscillator with frequency changed by varying the applied voltage between the anode and sole. As shown in Fig. 10-1-12, the electric beam is emitted from a short cylindrical cathode at one end of the device. Electrons are formed into a hollow beam by the electric and magnetic forces near the cathode and then are accelerated radically outward from the cathode. The electron beam is then injected into the region between the sole and the anode. The beam rotates about the sole at the rate controlled by the axial magnetic field and the dc voltage applied between the anode and the sole.

The voltage-tunable magnetron uses a low-Q resonator, and its bandwidth may exceed 50% at low-power levels. In the π-mode operation, the bunch process of the hollow beam occurs in the resonator, and the frequency of oscillation is determined by the rotational velocity of the electron beam. In other words, the oscillation frequency can be controlled by varying the applied dc voltage between the anode and
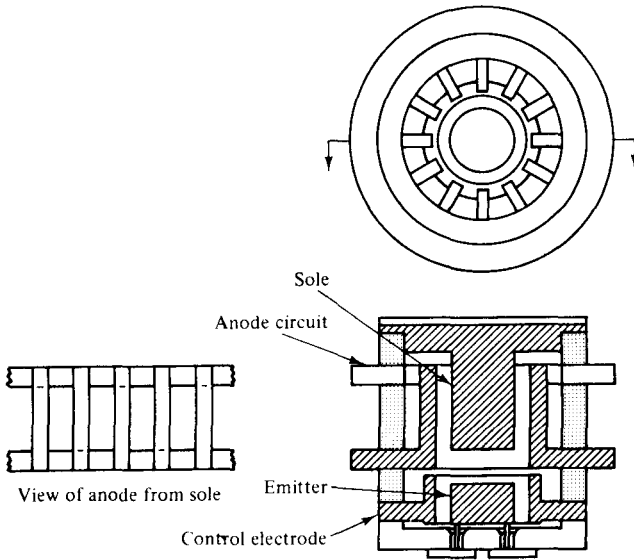
Figure 10-1-12 Cross-section view of a voltage-tunable magnetron. (*Courtesy of Varian Associates, Inc.*)

sole. Power output can be adjusted to some extent through the use of the control electrode in the electron gun. At high-power levels and high frequencies, the bandwidth percentage is limited. However, at low-power levels and low frequencies, the bandwidth may approach 70%.

10-1-5 Inverted Coaxial Magnetron

A magnetron can be built with the anode and cathode inverted—that is, with the cathode surrounding the anode. For some time, the basic problem of mode suppression prevented its use. In an inverted coaxial magnetron, the cavity is located inside a slotted cylinder, and a resonator vane array is arranged on the outside. The cathode is built as a ring around the anode. Figure 10-1-13 shows the schematic diagram of an inverted coaxial magnetron.
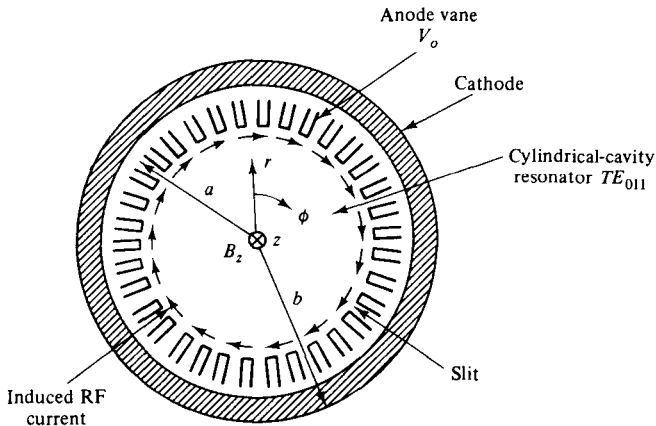
Figure 10-1-13 Schematic diagram of an inverted coaxial magnetron. (*Courtesy of Varian Associates, Inc.*)

Mathematically, the motion equations of the electrons in an inverted coaxial magnetron can be written from Eqs. (1-2-5a, b, c) as

$$\frac{d^2r}{dt^2} - r\left(\frac{d\phi}{dt}\right)^2 = \frac{e}{m}E_r - \frac{e}{m}rB_z\frac{d\phi}{dt} \qquad (10\text{-}1\text{-}55)$$

$$\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = \frac{e}{m}B_z\frac{d\phi}{dt} \qquad (10\text{-}1\text{-}56)$$

where $\dfrac{e}{m} = 1.759 \times 10^{11}$ C/kg is the charge-to-mass ratio of electron

$B_0 = B_z$ is assumed in the positive z direction

Rearrangement of Eq. (10-1-56) results in the following form

$$\frac{d}{dt}\left(r^2\frac{d\phi}{dt}\right) = \frac{e}{m}B_z r\frac{dr}{dt} = \frac{1}{2}\omega_c\frac{d}{dt}(r^2) \qquad (10\text{-}1\text{-}57)$$

where $\omega_c = \dfrac{e}{m}B_z$ is the cyclotron angular frequency. Integration of Eq. (10-1-57) yields

$$r^2\frac{d\phi}{dt} = \frac{1}{2}\omega_c r^2 + \text{constant} \qquad (10\text{-}1\text{-}58)$$

at $r = b$, where b is the radius of the cathode cylinder, and $\dfrac{d\phi}{dt} = 0$, constant $= -\dfrac{1}{2}\omega_c b^2$. The angular velocity is expressed by

$$\frac{d\phi}{dt} = \frac{1}{2}\omega_c\left(1 - \frac{b^2}{r^2}\right) \qquad (10\text{-}1\text{-}59)$$

Since the magnetic field does no work on the electrons, the kinetic energy of the electron is given by

$$\tfrac{1}{2}m\mathcal{V}^2 = eV \qquad (10\text{-}1\text{-}60)$$

However, the electron velocity has r and ϕ components such as

$$\mathcal{V}^2 = \frac{2e}{m}V = \mathcal{V}_r^2 + \mathcal{V}_\phi^2 = \left(\frac{dr}{dt}\right)^2 + \left(r\frac{d\phi}{dt}\right)^2 \qquad (10\text{-}1\text{-}61)$$

at $r = a$, where a is the radius from the center of the cylinder to the edge of the anode, $V = V_0$, and $dr/dt = 0$.

When the electrons just graze the anode, Eqs. (10-1-60) and (10-1-61) become

$$\frac{d\phi}{dt} = \frac{1}{2}\omega_c\left(1 - \frac{b^2}{a^2}\right) \qquad (10\text{-}1\text{-}62)$$

$$a^2\left(\frac{d\phi}{dt}\right)^2 = \frac{2e}{m}V_0 \qquad (10\text{-}1\text{-}63)$$

Substitution of Eq. (10-1-62) into Eq. (10-1-63) results in

$$a^2 \left[\frac{1}{2} \omega_c \left(1 - \frac{b^2}{a^2} \right) \right]^2 = \frac{2e}{m} V_0 \qquad (10\text{-}1\text{-}64)$$

The electron will acquire a tangential as well as a radial velocity. Whether the electron will just graze the anode and return back toward the cathode depends on the relative magnitudes of the anode voltage V_0 and the magnetic flux density B_0. The cutoff condition can be obtained from Eq. (10-1-64) as

$$V_{0c} = \frac{e}{8m} B_0^2 a^2 \left(1 - \frac{b^2}{a^2} \right)^2 \qquad (10\text{-}1\text{-}65)$$

This means that if $V_0 < V_{0c}$ for a given B_0, the electrons will not reach the anode. Equation (10-1-65) is often called the *Hull cutoff voltage equation*. Similarly, the magnetic cutoff condition is expressed by

$$B_{0c} = \frac{-\left(8V_0 \frac{m}{e} \right)^{1/2}}{a \left(1 - \frac{b^2}{a^2} \right)} \qquad (10\text{-}1\text{-}66)$$

This means that if $B_0 > B_{0c}$ for a given V_0, the electrons will not reach the anode. Equation (10-1-66) is called the *Hull cutoff magnetic equation*.

The advantage of an inverted coaxial magnetron design is that the cathode current density can be reduced to one-tenth of that used in cathode-centered magnetrons. Thus the millimeter magnetron is a practical and long-life device. The output waveguide can be in the circular electric mode that has extremely low transmission loss. Figure 10-1-14 compares the inverted coaxial magnetron with a conventional magnetron designed for the same frequency. It should be noted that the cathode sizes are quite different.
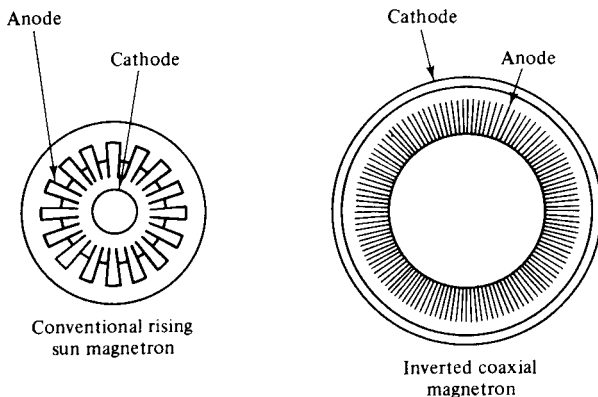
Figure 10-1-14 Comparison of a conventional magnetron with an inverted coaxial magnetron. (*Courtesy of Varian Associates, Inc.*)

Example 10-1-5:   Inverted Coaxial Magnetron

An inverted coaxial magnetron has the following parameters:

Anode voltage:	$V_0 = 10$ kV
Cathode current:	$I_0 = 2$ A
Anode radius:	$a = 3$ cm
Cathode radius:	$b = 4$ cm
Magnetic flux density:	$B_0 = 0.01$ Wb/m^2

Determine:

a. The cutoff voltage for a fixed B_0

b. The cutoff magnetic flux density for a fixed V_0

Solution

a. The cutoff voltage is

$$V_{0c} = \frac{1}{8} \times 1.759 \times 10^{11} \times (0.01)^2 \times (3 \times 10^{-2})^2$$

$$\times \left(1 - \frac{4^2}{3^2}\right)^2$$

$$= 1.20 \text{ kV}$$

b. The cutoff magnetic flux density is

$$B_{0c} = -\left(\frac{8 \times 10 \times 10^3}{1.759 \times 10^{11}}\right)^{1/2}\left[3 \times 10^{-2}\left(1 - \frac{4^2}{3^2}\right)\right]^{-1}$$

$$= 0.0289 \text{ Wb/m}^2$$

10-1-6 Frequency-Agile Coaxial Magnetron

A frequency-agile coaxial magnetron differs from a standard tunable magnetron. The frequency agility (FA) of a coaxial magnetron is defined as the capability to tune the output frequency of the radar with sufficiently high speed to produce a pulse-to-pulse frequency change greater than the amount required effectively to obtain decorrelation of adjacent radar echoes. The frequency-agile magnetron, together with appropriate receiver integration circuits, can reduce target scintillation, increase the detectability of target in a clutter environment, and improve resistance to electronic countermeasures (ECM). The increase of the pulse-to-pulse frequency separation will improve the radar system performance. Furthermore, the greater the pulse-to-pulse frequency separation, the more difficult it will be to center a jamming transmitter on the radar frequency for effective interference with system operation.

The frequency-agile coaxial magnetrons are classified into three types:

1. *Dither magnetrons:* The output RF frequency varies periodically with a constant excursion, constant rate, and a fixed center frequency.
2. *Tunable/dither magnetrons:* The output RF frequency varies periodically with a constant excursion and constant rate, but the center frequency can be manually tuned by hand or mechanically tuned by a servomotor.
3. *Accutune magnetrons:* The output RF frequency variations are determined by the waveforms of an externally generated, low-level voltage signal. With proper selection of a tuning waveform, the accutune magnetron combines with features of dither and tunable/dither magnetrons, together with a capability for varying the excursion, rate, and tuning waveform. Figure 10-1-15 shows an accutune magnetron.
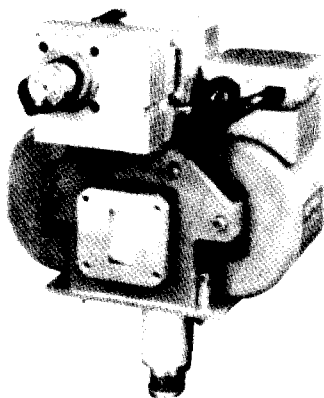
Figure 10-1-15   X-band accutune magnetron VMX-1430. (*Courtesy of Varian Associates, Inc.*)

The X-band frequency-agile coaxial magnetron VMX-1430 is a typical agile magnetron. Its pulse voltage is 15 kV, and pulse current is 15 A. Its maximum duty cycle is 0.0011, and accutune range is 1 GHz. Its center frequency is 9.10 GHz, and peak output power is 90 kW. The agile rate and agile excursion are shown in Fig. 10-1-16.

From Fig. 10-1-16, it can be seen that the agile rate is the number of times per second that the transmitter frequency traverses the agile excursion and returns to its starting frequency. Similarly, the agile excursion is defined as the total frequency variation of the transmitter during agile operation.

The number of pulses that can be effectively integrated cannot be greater than the number of pulses placed on the target during one scan of the antenna. Therefore, the antenna beamwidth and scan rate become factors that must also be considered in determining the integration period of the radar. Consequently, a design value for agile excursion can now be expressed in terms of radar operating parameters as

$$\text{Agile excursion} = \frac{N}{\tau} \qquad (10\text{-}1\text{-}67)$$
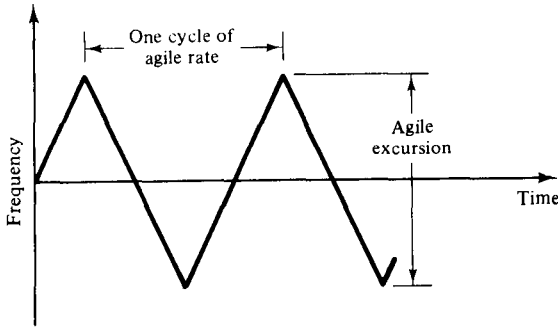
Figure 10-1-16   Agile rate and agile excursion.

where N = number of pulses placed on the target during one radar scan, say, 20, whichever is smaller

τ = shortest pulse duration used in the system

The frequency, or *pulse repetition rate (PRR)*, is given by

$$f = \frac{DC}{\tau} \tag{10-1-68}$$

where DC = duty cycle is the ratio of the pulse duration over the repetition period for a pulse. The duty cycle is defined as

$$\text{Duty cycle} = \frac{\text{Pulse duration}}{\text{Pulse repetition period}} = \frac{\tau}{T} = \tau f \tag{10-1-69}$$

Hence, the agile rate can be written as

$$\text{Agile rate} = \frac{1}{2T} \tag{10-1-70}$$

where the 2 in the denominator is counted for the fact that two excursions through the agile frequency range occur during each cycle of agile rate.

Example 10-1-6:   Frequency-Agile Magnetron

A frequency-agile coaxial magnetron has the following operating parameters:

Pulse duration: $\tau = 0.20, 0.40, 0.80 \ \mu s$

Duty cycle: $DC = 0.001$

Pulse rate on target: $N = 14$ per scan

Determine:

a. The agile excursion
b. The pulse-to-pulse frequency separation
c. The signal frequency
d. The time for N pulses
e. The agile rate

Solution

a. The agile excursion is

$$\text{Agile excursion} = \frac{14}{0.2 \times 10^{-6}} = 70 \text{ MHz}$$

b. The pulse-to-pulse frequency separation is

$$f_p = \frac{1}{\tau} = \frac{1}{0.20 \times 10^{-6}} = 5 \text{ MHz}$$

c. The signal frequency is

$$f = \frac{\text{DC}}{\tau} = \frac{0.001}{0.20 \times 10^{-6}} = 5 \text{ kHz}$$

d. The time for 14 pulses per second is

$$\text{Time} = \frac{\text{N}}{f} = \frac{14}{5000} = 2.8 \text{ ms}$$

e. The agile rate is

$$\text{Agile rate} = \frac{1}{2 \times 0.0028} = 178.57 \text{ Hz}$$

10-2 FORWARD-WAVE CROSSED-FIELD AMPLIFIER (FWCFA OR CFA)

The crossed-field amplifier (CFA) is an outgrowth of the magnetron. CFAs can be grouped by their mode of operation as forward-wave or backward-wave types and by their electron stream source as emitting-sole or injected-beam types. The first group concerns the direction of the phase and group velocity of the energy on the microwave circuit. This can be seen from the ω-β diagram in Fig. 9-5-5 as discussed in Sec. 9-5. Since the electron stream reacts to the RF electric field forces, the behavior of the phase velocity with frequency is of prime concern. The second group emphasizes the method by which electrons reach the interaction region and how they are controlled. This can be seen in the schematic diagrams of Fig. 10-2-1.

In the forward-wave mode, the helix-type slow-wave structure is often selected as the microwave circuit for the crossed-field amplifier; in the backward-wave mode, the strapped bar line represents a satisfactory choice. A structure of strapped crossed-field amplifier is shown in Fig. 10-2-2.

10-2-1 Principles of Operation

In the emitting-sole tube, the current emanated from the cathode is in response to the electric field forces in the space between the cathode and anode. The amount of

(a)

(b)
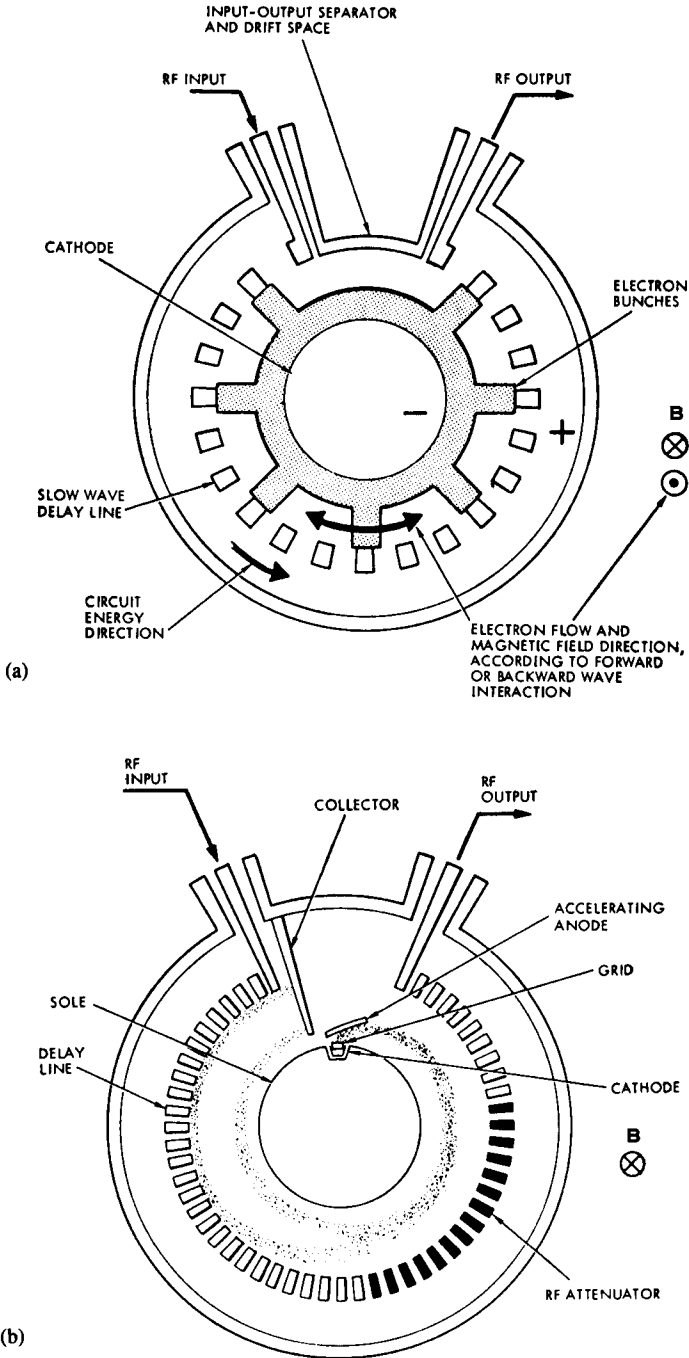
Figure 10-2-1   Schematic diagrams of CFAs. (*After J. F. Skowron [2], reprinted by permission of IEEE, Inc.*)
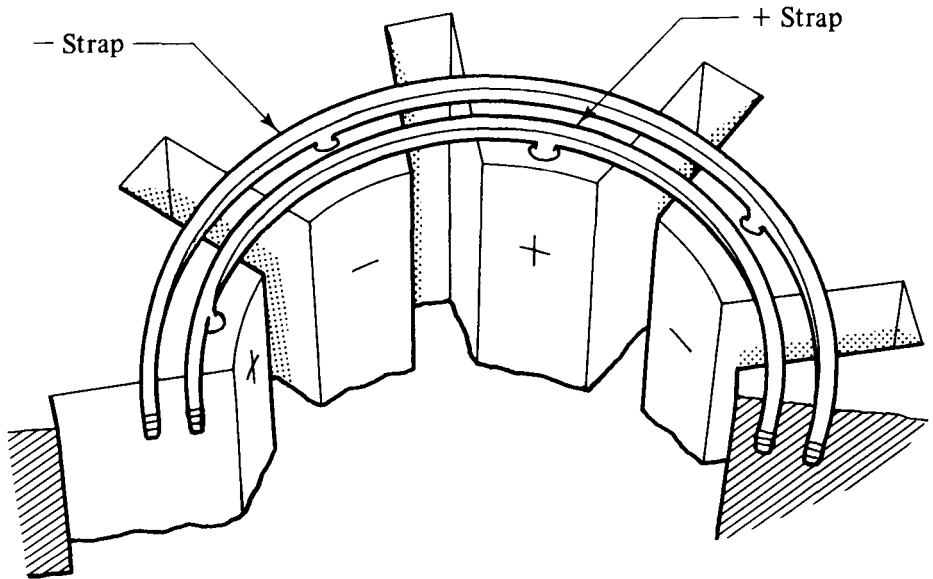
Figure 10-2-2   Diagram of a strapped CFA. (*After J. F. Skowron [2], reprinted by permission of IEEE, Inc.*)

current is a function of the dimension, the applied voltage, and the emission properties of the cathode. The perveance of the interaction geometry tends to be quite high, about 5 to 10 $\times$ 10^{-10}, which results in a high-current and high-power capability at relative low voltage. In the injected-beam tube the electron beam is produced in a separate gun assembly and is injected into the interaction region.

The beam-circuit interaction features are similar in both the emitting-sole and the injected-beam tubes. Favorably phased electrons continue toward the positively polarized anode and are ultimately collected, whereas unfavorably phased electrons are directed toward the negative polarized electrode.

In linear-beam interaction, as discussed for traveling-wave tubes in Sec. 9-5, the electron stream is first accelerated by an electric gun to the full dc velocity; the dc velocity is approximately equal to the axial phase velocity of the RF field on the slow-wave structure. After interaction occurs, the spent electron beam leaves the interaction region with a low-average velocity. The difference in velocity is accounted for by the RF energy created on the microwave circuit. In the CFA, the electron is exposed to the dc electric field force, magnetic field force, and the electric field force of the RF field, and even to the space-charge force from other electrons. The last force is normally not considered in analytic approaches because of its complexity. Under the influence of the three forces, the electrons travel in spiral trajectories in a direction tending along equipotentials. The exact motion has been subject to much analysis by means of a computer. Figure 10-2-3 shows the pattern of the electron flow in the CFA by computerized techniques [2]. It can be seen that when the spoke is positively polarized or the RF field is in the positive half cycle, the electron speeds up toward the anode; while the spoke is negatively polarized or the RF field is
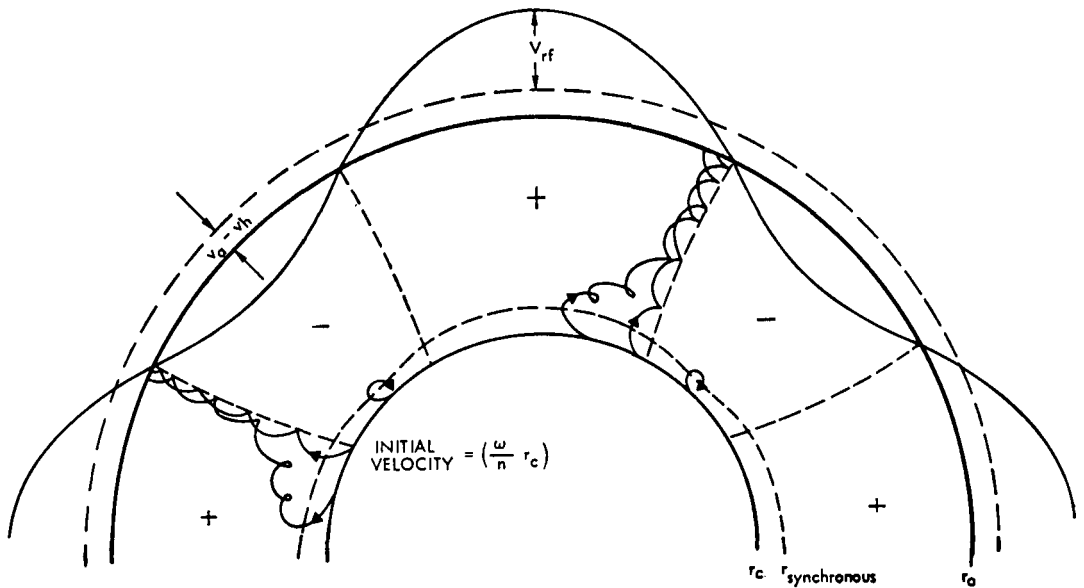
Figure 10-2-3   Motion of electrons in CFA. (*After J. F. Skowron [2], reprinted by permission of IEEE, Inc.*)

in the negative half cycle, the electrons are returned toward the cathode. Consequently, the electron beam moves in a spiral path in the interaction region.

The total power generated in a given CFA is independent of the RF input power, as long as the input power exceeds the threshold value for spoke stability at the input. The power generated can be increased only by increasing the anode voltage and current. Neglecting circuit attenuation, the output power of the CFA is equal to the sum of the input power and the power generated in the interaction region. That is, the power gain of a CFA is given by

$$g = \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{P_{\text{in}} + P_{\text{gen}}}{P_{\text{in}}} = 1 + \frac{P_{\text{gen}}}{P_{\text{in}}} \qquad (10\text{-}2\text{-}1)$$

where $P_{\text{out}} = P_{\text{in}} + P_{\text{gen}}$

P_{in} = RF input power

P_{gen} = RF power induced into the anode circuit by electrons

Therefore, the CFA is not a linear amplifier but rather is termed a *saturated amplifier*.

The efficiency of a CFA is defined as the product of the electronic efficiency η_e and the circuit efficiency η_c. The electronic efficiency η_e is defined as in Eq. 10-1-29). The overall efficiency is then expressed as

$$\eta = \eta_c \eta_e = \frac{P_{out} - P_{in}}{V_{a0} I_{a0}} \qquad (10\text{-}2\text{-}2)$$

where P_{out} = RF output power
  P_{in} = RF input power
  η_e = P_{gen}/P_{dc}
  P_{dc} = $V_{a0}I_{a0}$ is dc power
  V_{a0} = anode dc voltage
  I_{a0} = anode dc current

The circuit efficiency is defined as

$$\eta_c = \frac{\eta}{\eta_e} = \frac{P_{out} - P_{in}}{P_{gen}} \tag{10-2-3}$$

where $P_{gen} = \eta_e V_{a0} I_{a0}$

Since the power generated per unit length is constant, the output power is given by

$$P_{out} = P_{in}e^{-2\alpha\ell} + \int_0^\ell \frac{P_{gen}}{\ell} e^{-2\alpha(\ell-\phi)}\, d\phi$$

$$= P_{in}e^{-2\alpha\ell} + \frac{P_{gen}}{2a\ell}(1 - e^{-2\alpha\ell}) \tag{10-2-4}$$

where α = circuit attenuation constant
  ℓ = circuit length in ϕ direction

Substitution of Eq. (10-2-4) in Eq. (10-2-3) results in the circuit-efficiency equation as

$$\eta_c = \left(\frac{1}{2\alpha\ell} - \frac{P_{in}}{P_{gen}}\right)(1 - e^{-2\alpha\ell}) \tag{10-2-5}$$

The term P_{in}/P_{gen} becomes negligible for high-gain CFA.
  Assume that the input signal is sufficiently strong for spoke stability, that the RF power grows linearly with distance along the circuit, that the dc current for the spoke is constant, and that the back-bombardment loss is not considered, then the electronic-efficiency equation can be derived as follows. The average drift electron velocity at any position is given by

$$\mathcal{V} = \frac{E}{B} \tag{10-2-6}$$

where E = totall electric field at the position under consideration
  B = magnetic flux density at the position

The power flow at any position is related to the RF field and the beam-coupling impedance of the circuit by

$$P = \frac{E_{max}^2}{2\beta^2 Z_c} \tag{10-2-7}$$

where E_{max} = peak electric field
  $\beta = \omega/\mathcal{V}$ is phase constant
  Z_c = beam-coupling impedance

The power loss per spoke due to the electron motion toward the anode at any position is derived from Eq. (10-2-6) as

$$P_s = V_{s0}I_{s0} = I_{s0}\frac{m}{e}\frac{\beta^2 Z_c P}{B^2} \tag{10-2-8}$$

where V_{s0} = dc voltage per spoke
  I_{s0} = dc current per spoke

Since the power varies linearly with position, the average power loss over the entire circuit length is

$$P_{s,\text{avg}} = I_{s0}\frac{m}{2e}\frac{\beta^2 Z_c}{B^2}(P_{\text{in}} + P_{\text{out}}) \tag{10-2-9}$$

By using Eqs. (10-2-1), (10-2-9), and $\mathcal{V}\beta = \omega$, the total power loss for all the spokes is given by

$$P_{\text{lost}} = I_{a0}\frac{m}{2e}\left(\frac{\omega}{\beta}\right)^2 + I_{a0}\frac{m}{2e}\frac{\beta^2 Z_c}{B^2}\left(\frac{g+1}{g-1}\right)P_{\text{gen}} \tag{10-2-10}$$

where g = power gain.

Finally, the electronic efficiency is given by

$$\eta_e = \frac{P_{\text{gen}}}{V_{a0}I_{a0}} = \frac{1 - \dfrac{m\omega^2}{2eV_{a0}\beta^2}}{1 + \dfrac{I_{a0}}{B^2}\dfrac{m\beta^2 Z_c}{2e}\left(\dfrac{g+1}{g-1}\right)} \tag{10-2-11}$$

10-2-2 Microwave Characteristics

The crossed-field amplifier (CFA) is characterized by its low or moderate power gain, moderate bandwidth, high efficiency, saturated amplification, small size, low weight, and high perveance. These features have allowed the CFA to be used in a variety of electronic systems ranging from low-power and high-reliability space communications to multimegawatt, high average power, coherent pulsed radar. Figure 10-2-4 shows the present state of the art for the U.S. high-power forward-wave crossed-field amplifiers.

The Raytheon QKS-1541 amplifier, shown in Fig. 10-2-5, is a typical forward-wave CFA with an average power of 14 kW at frequency range from 2.9 to 3.1 GHz. Its pulse width is 28 s. Its typical peak anode voltage is 44 to 56 kV and peak anode current is 58 A.

Example 10-2-1:   Crossed-Field Amplifier

A CFA operates under the following parameters:

Anode dc voltage: $V_{a0} = 2$ kV
Anode dc current: $I_{a0} = 1.5$ A

Electronic efficiency: $\eta_e = 20\%$
RF input power: $P_{in} = 80$ W

Calculate:

a. The induced RF power
b. The total RF output power
c. The power gain in decibels

Solution

a. The induced RF power is

$$P_{gen} = 0.20 \times 2 \times 10^3 \times 1.5 = 600 \text{ W}$$

b. The RF output power is

$$P_{out} = P_{in} - P_{gen} = 80 + 600 = 680 \text{ W}$$

c. The power gain is

$$g = \frac{P_{out}}{P_{in}} = \frac{680}{80} = 8.50 = 9.3 \text{ dB}$$
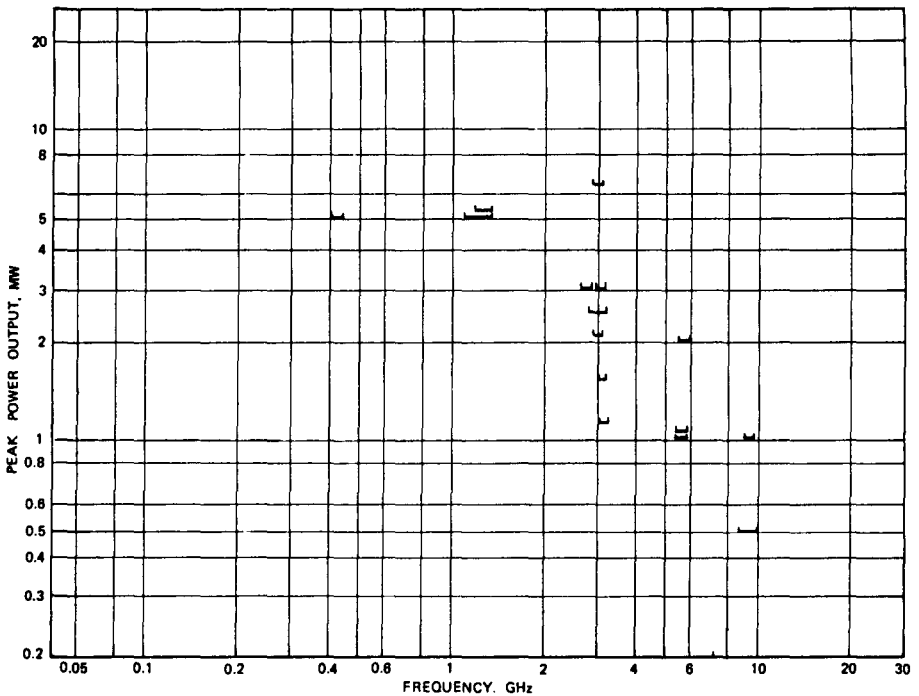
Figure 10-2-4   State of the art for U.S. high-power CFAs.
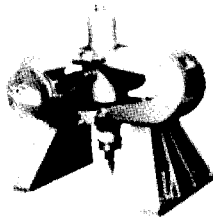
Figure 10-2-5   Photograph of Raytheon QKS-1541 CFA *(Courtesy of Raytheon Company, Microwave Tube Operation.)*

10-3 BACKWARD-WAVE CROSSED-FIELD AMPLIFIER (AMPLITRON)

The trade name of the backward-wave crossed-field amplifier (BWCFA) is Amplitron, and its schematic diagram is shown in Fig. 10-3-1.
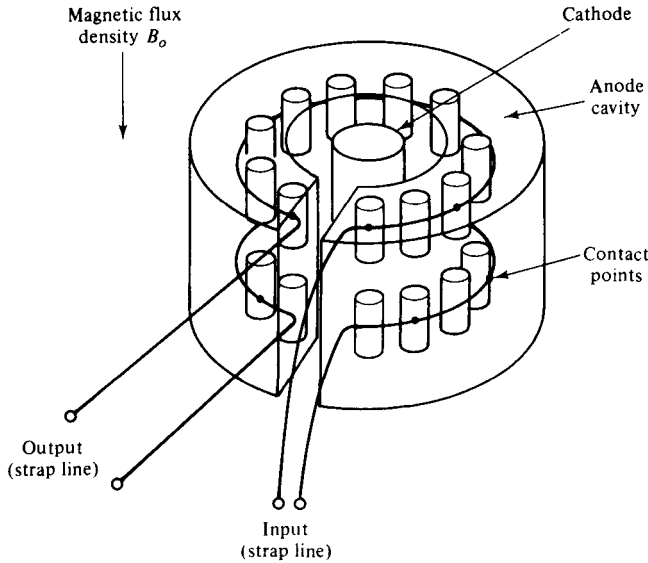
Figure 10-3-1   Schematic diagram of Amplitron.

The anode cavity and pins comprise the resonator circuits. A pair of pins and the cavity are excited in opposite phase by the strap line. The electron beam and the electromagnetic waves interact in the resonant circuits. The BWCFA can deliver 3-MW pulse with 10-μs duration at S-band, and the tube gain reaches 8 dB.

The highly successful QK434 Amplitron produced by Raytheon exhibited stable gain as high as 16 dB, power output levels ranging from a few hundred kilowatts to 3 MW, and efficiencies ranging from 60% for normal power levels to 76% for high-power, low-gain operation. The tube is commonly used in air surveillance radar and military pulsed radar. Figure 10-3-2 pictures the QK434 Amplitron.

The two-stage superpower Amplitron, also manufactured by Raytheon, generated 425 kW of CW power at an efficiency of 76%. The gain was 9 dB and the band-
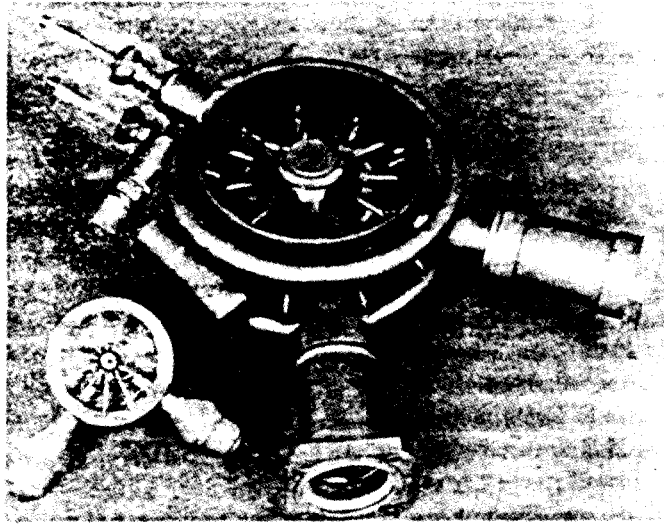
Figure 10-3-2 Photograph of QK-434 Amplitron. (*From W. C. Brown [3]; reprinted by permission of IEEE, Inc.*)

width was 5% at a mean frequency of 3 GHz. The tube was used for all high-data-rate transmission in the Apollo program. Figure 10-3-3 shows the two-stage superpower Amplitron.
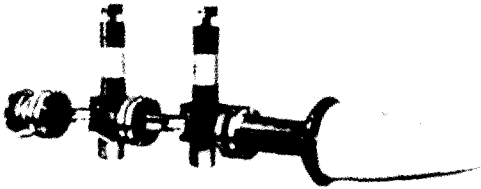
Figure 10-3-3 Photograph of a two-stage superpower Amplitron. (*From W. C. Brown [3]; reprinted by permission of IEEE, Inc.*)

The circuit and electronic equations for the *M*-type amplifier and oscillator were developed by several authors [4, 5]. The basic secular equation including the space-charge effect is given by

$$(\gamma^2 - \gamma_0^2)(j\beta_e - \gamma)\left[(j\beta_e - \gamma)^2 + \beta_m^2\right]$$

$$= -j\beta_e\gamma_0\gamma^2\left[(j\beta_e - \gamma) + j\frac{2\alpha}{1 + \alpha^2}B_m\right]H^2 \qquad (10\text{-}3\text{-}1)$$

where γ_0 = circuit propagation constant

γ = harmonic wave propagation constant

$\beta_e = \dfrac{\omega}{\mathcal{V}_0}$ is the electron-beam phase constant

$\mathcal{V}_0 = \sqrt{\dfrac{2e}{m}V_0}$ is the dc electron-beam velocity

$$\beta_m = \frac{\omega_c}{\mathcal{V}_0} = \frac{e}{\mathcal{V}_0 m} B_0 \text{ is the cyclotron phase constant}$$

$$\omega_c = \frac{e}{m} B_0 \text{ is the cyclotron angular frequency}$$

$B_0 = $ crossed-magnetic flux density

$$H_2 = 2(1 + \alpha^2)\phi^2 C^3$$

$$C = \left(\frac{I_0 Z_0}{4V_0}\right)^{1/3} \text{ is the gain parameter}$$

$$\phi = A \exp(-j\gamma y) + B \exp(j\gamma y) \text{ is the wave equation}$$

$$\alpha = \frac{A \exp(-j\gamma y) - B \exp(j\gamma y)}{A \exp(-j\gamma y) + B \exp(j\gamma y)} = j \frac{1}{\gamma\phi} \frac{d\phi}{dy} \text{ is a factor}$$

In general, there are five solutions of γ from Eq. (10-3-1). Let

$$\gamma_0 = j\beta \tag{10-3-2}$$

$$\gamma = j\beta(1 + p) \tag{10-3-3}$$

where p is a very small constant ($p \ll 1$).

Substituting Eqs. (10-3-2) and (10-3-3) in Eq. (10-3-1) we have

$$p\left(\frac{\beta_e}{\beta} - 1 - p\right)\left[\left(\frac{\beta_e}{\beta} - 1 - p\right)^2 - \left(\frac{\beta_m}{\beta}\right)^2\right]$$
$$= \frac{-B_e}{2\beta}\left[\left(\frac{\beta_e}{\beta} - 1 - P\right) + \frac{2\alpha\beta_m}{(1 + \alpha^2)\beta}\right]H^2 \tag{10-3-4}$$

The right-hand term is small, as it is not much different from C, so that if either one of the left-hand terms is also small, Eq. (10-3-4) will be satisfied.

If $\beta_e \doteq \beta$ is assumed, the first factor is small and the solution for p becomes

$$p = \pm j\left(\frac{\alpha}{1 + \alpha^2}\right)^{1/2}\left(\frac{\beta}{\beta_m}\right)^{1/2} H \tag{10-3-5}$$

If $\beta_e \pm \beta_m \doteq \beta$ is assumed, the second factor is small and the solution for p is given by

$$p^2 = \pm \frac{1}{4}\frac{(1 \pm \alpha)^2}{1 + \alpha^2}\left(\frac{\beta}{\beta_m} \pm 1\right)H^2 \tag{10-3-6}$$

From the definition of p, gain is only obtained when p is imaginary. From Eq. (10-3-5), this only happens when α is positive. In Eq. (10-3-6), p is imaginary when $\beta_e + \beta_m = \beta$ and α is less than unity. This last condition comes from rewriting Eq. (10-3-6) as

$$p = \pm \frac{1}{2}\frac{1 - \alpha}{(1 + \alpha)^{1/2}}\left(\frac{\beta_e}{\beta_m}\right)^{1/2} H \tag{10-3-7}$$

The solution for an *M*-type backward-wave amplifier can be obtained by setting

$$\gamma = -j\beta(1 + p) \tag{10-3-8}$$

The only possible solution for $\beta + \beta_e = \beta_m$ is

$$p = \pm j\frac{1}{2}\frac{1 + \alpha}{(1 + \alpha^2)^{1/2}}\left(\frac{\beta_e}{\beta_m}\right)^{1/2} H \tag{10-3-9}$$

This gives rise to increasing backward waves for all values of α except $\alpha = -1$. Because $H \propto C^{3/2}$, Eqs. (10-3-5), (10-3-7), and (10-3-9) show that the gain per unit length in *M*-type devices is lower than in *O*-type devices using similar circuits because $C^{3/2} < C$.

Example 10-3-1:   Amplitron Characteristics

An Amplitron has the following operating parameters:

Anode voltage:	$V_0 = 15$ kV
Anode current:	$I_0 = 3$ A
Magnetic flux density:	$B_0 = 0.2$ Wb/m^2
Operating frequency:	$f = 8$ GHz
Characteristic impedance:	$Z_0 = 50\ \Omega$

Determine:

a. The dc electron-beam velocity
b. The electron-beam phase constant
c. The cyclotron angular frequency
d. The cyclotron phase constant
e. The gain parameter

Solution

a. The dc electron-beam velocity is

$$\mathcal{V}_0 = 0.593 \times 10^6 \times (15 \times 10^3)^{1/2} = 0.762 \times 10^8 \text{ m/s}$$

b. The electron-beam phase constant is

$$\beta_e = \frac{\omega}{\mathcal{V}_0} = \frac{2\pi \times 8 \times 10^9}{0.762 \times 10^5} = 692.36 \text{ rad/m}$$

c. The cyclotron angular frequency is

$$\omega_c = \frac{e}{m}B_0 = 1.759 \times 10^{11} \times 0.2 = 35.18 \times 10^9 \text{ rad/s}$$

d. The cyclotron phase constant is

$$\beta_m = \frac{\omega_c}{\mathcal{V}_0} = \frac{35.18 \times 10^9}{0.726 \times 10^3} = 484.57 \qquad \text{rad/m}$$

e. The gain parameter is

$$C = \left(\frac{I_0 Z_0}{4V_0}\right)^{1/3} = \left(\frac{3 \times 50}{4 \times 15 \times 10^3}\right)^{1/3} = 0.136$$

10-4 BACKWARD-WAVE CROSSED-FIELD OSCILLATOR (CARCINOTRON)

The backward-wave crossed-field oscillator of M-Carcinotron has two configurations: linear M-carcinotron and circular M-carcinotron.

Linear M-Carcinotron. The M-Carcinotron oscillator is an M-type backward-wave oscillator. The interaction between the electrons and the slow-wave structure takes place in a space of crossed field. A linear model of the M-Carcinotron oscillator is shown in Fig. 10-4-1.
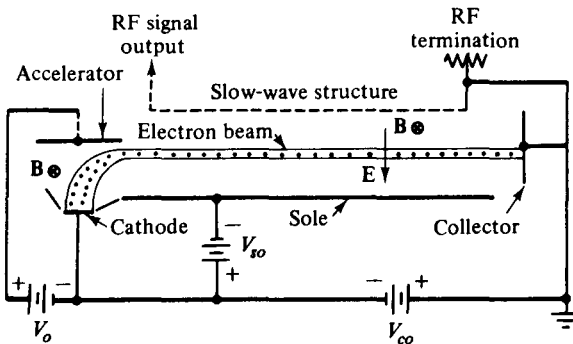
Figure 10-4-1 Linear model of an M-Carcinotron oscillator. (*From J. V. Gewartowski and H. A. Watson [6]; reprinted by permission of Van Nostrand Company.*)

The slow-wave structure is in parallel with an electrode known as the sole. A dc electric field is maintained between the grounded slow-wave structure and the negative sole. A dc magnetic field is directed into the page. The electrons emitted from the cathode are bent through a 90° angle by the magnetic field. The electrons interact with a backward-wave space harmonic of the circuit, and the energy in the circuit flows opposite to the direction of the electron motion. The slow-wave structure is terminated at the collector end, and the RF signal output is removed at the electron-gun end. Since the M-Carcinotron is a crossed-field device, its efficiency is very high, ranging from 30 to 60%.

The perturbed electrons moving in synchronism with the wave in a linear M-Carcinotron are shown in Fig. 10-4-2.

Electrons at position A near the beginning of the circuit are moving toward the circuit, whereas electrons at position B are moving toward the sole. Farther down
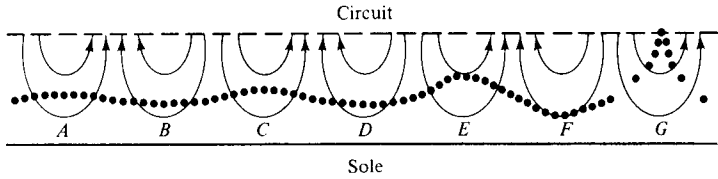
Figure 10-4-2   Beam electrons and electric field lines in an M-Carcinotron.

the circuit, electrons at position *C* are closer to the circuit, and electrons at position *D* are closer to the sole. However, electrons at position *C* have departed a greater distance from the unperturbed path than have electrons at position *D*. Thus, the electrons have lost a net amount of potential energy, this energy having been transferred to the RF field. The reason for the greater displacement of the electrons moving toward the circuit is that these electrons are in stronger RF fields, since they are closer to the circuit. Electrons at position *G* have moved so far from the unperturbed position that some of them are being intercepted on the circuit. The length from position *A* through position *G* is a half cycle of the electron motion.

Circular *M*-Carcinotron.   The *M*-Carcinotrons are generally constructed in the circular reentrant form as shown in Fig. 10-4-3. The slow-wave structure and sole are circular and nearly reentrant to conserve magnet weight. The sole has the appearance of the cathode in a magnetron.

The Litton L-3721 M-BWO, shown in Fig. 10-4-4, is a typical *M*-type backward-wave oscillator (*M*-BWO or Carcinotron) with a minimum power of 200 W at frequency range from 1.0 to 1.4 GHz.
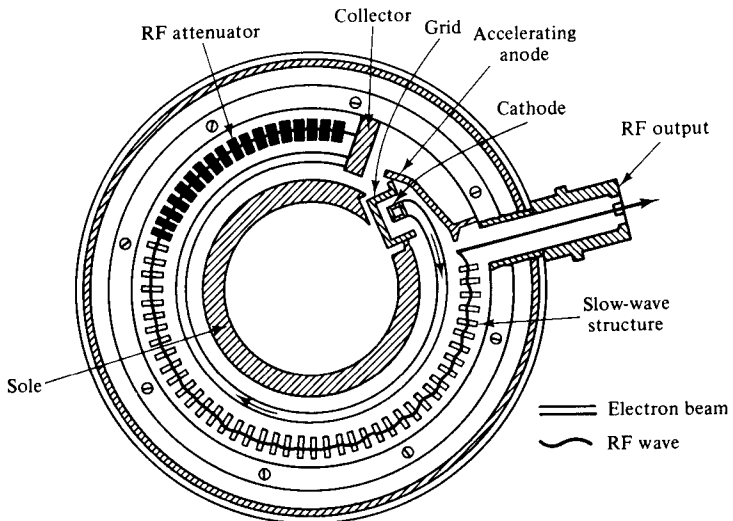
Figure 10-4-3   Schematic diagram of a circular M-Carcinotron. (*Courtesy of Raytheon Company, Microwave Tube Operation.*)
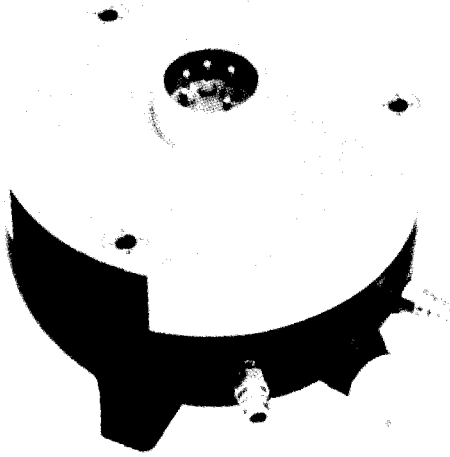
Figure 10-4-4   Photograph of Litton L-3721 BWO. (*Courtesy of Litton Company, Electron Tube Division.*)

L-3721 7⅝″ Wide

In the circular configurations, the delay line is terminated at the collector end by spraying attenuating material on the surfaces of the conductors. The output is taken from the gun end of the delay line which is an interdigital line. Clearly, in this case, the electron drift velocity has to be in synchronism with a backward-space harmonic.

As in the case of O-type devices, the only modification in the secular equation is a change of sign in the circuit equation. If this change is made in Eq. (10-3-1), we write

$$\gamma_0 = j\beta \tag{10-4-1}$$

$$\gamma = jk + \epsilon \tag{10-4-2}$$

By eliminating negligible terms, we obtain for the Carcinotron

$$(\beta^2 - k^2 + j2k\epsilon)\left[j(\beta_e - k) - \epsilon\right]\left[\beta_m^2 - (\beta_e - k)^2 - j2(\beta_e - k)\epsilon\right]$$

$$= j\beta\beta_e k^2\left(\beta_e - k + \frac{2\alpha}{1 + \alpha^2}\beta_m\right)H_2 \tag{10-4-3}$$

A solution of Eq. (10-4-3) for synchronism can be obtained by setting $\beta = \beta_e$ and $\beta_e - k = \beta_e b'$, where b' is a small number so that terms like b'^2 and $b'\epsilon$ may be neglected. This yields

$$2\epsilon(j\beta_e b' - \epsilon) = \beta\beta_e k\left(\frac{b'}{\beta_m^2} + \frac{2\alpha}{1 + \alpha^2}\frac{1}{\beta_m}\right)H^2$$

$$\doteq \beta_e k\frac{2\alpha}{1 + \alpha^2}\frac{\beta}{\beta_m}H^2 \tag{10-4-4}$$

$$= 2\beta_e kD^2$$

where

$$D^2 = \frac{\alpha}{1 + \alpha^2} \frac{\beta}{\beta_m} H^2 \tag{10-4-5}$$

$$\epsilon = \beta_e D \delta \tag{10-4-6}$$

$$b' = bD \tag{10-4-7}$$

$$\delta(\delta - jb) = -1 \text{ or } \delta^2 - jb\delta + 1 = 0 \tag{10-4-8}$$

As a result, we have reduced the number of waves to two, with propagation constants given by

$$\gamma_1 = j(\beta_e + b) + \beta_e D \delta_1 \tag{10-4-9}$$

$$\gamma_2 = j(\beta_e + b) + \beta_e D \delta_2 \tag{10-4-10}$$

where the δ's are the roots of Eq. (10-4-8) and they are

$$\delta_1 = j \frac{b - \sqrt{b^2 + 4}}{2}$$

$$\delta_2 = j \frac{b + \sqrt{b^2 + 4}}{2}$$

To determine the amplification of the growing waves, the input reference point is set at $y = 0$, and the output reference point is taken at $y = \ell$. It follows that at $y = 0$, the voltage at the input point can be computed as follows:

$$V_1(0) + V_2(0) = V(0) \tag{10-4-11}$$

$$\frac{V_1(0)}{\delta_1} + \frac{V_2(0)}{\delta_2} = 0 \tag{10-4-12}$$

Solving Eqs. (10-4-11) and (10-4-12) simultaneously we have

$$V_1(0) = \frac{V(0)}{1 - \delta_2/\delta_1} = \frac{\delta_1 V(0)}{\delta_2} \tag{10-4-13}$$

$$V_2(0) = \frac{-V(0)}{1 - \delta_1/\delta_2} = \frac{-\delta_2 V(0)}{\delta_1} \tag{10-4-14}$$

Then the voltage at the output point $y = \ell$ is given by

$$V(0) = V_1(0) \exp(-\gamma_1 \ell) + V_2(0) \exp(-\gamma_2 \ell) \tag{10-4-15}$$

$$= V(0)[\delta_1 \exp(-\gamma_1 \ell) - \delta_2 \exp(-\gamma_2 \ell)]/(\delta_1 - \delta_2)$$

The term in the square bracket is the inverse of the voltage gain of the device. Oscillation takes place when this is zero. That is,

$$\delta_1 \exp(-\gamma_1 \ell) = \delta_2 \exp(-\gamma_2 \ell) \tag{10-4-16}$$

or

$$\delta_1/\delta_2 = \exp(\gamma_1\ell - \gamma_2\ell) \tag{10-4-17}$$

$$= \exp[-\beta_e D\,\ell(\delta_2 - \delta_2)]$$

From Eq. (10-4-8) we have

$$\delta_1/\delta_2 = \frac{b - \sqrt{b^2 + 4}}{b + \sqrt{b^2 + 4}} \tag{10-4-18}$$

and

$$\delta_2 - \delta_1 = j\sqrt{b^2 + 4} \tag{10-4-19}$$

Then

$$\delta_1/\delta_2 = \exp\left(-j\beta_e D\,\ell\sqrt{b^2 + 4}\right) \tag{10-4-20}$$

Equations (10-4-18) and (10-4-20) can only be satisfied simultaneously if $b = 0$ and $\delta_1 = -\delta_2$. Then

$$2\beta_e D\,\ell = (2n + 1)\pi \tag{10-4-21}$$

where n = any integer numbers. If we introduce N as usually defined through $\beta_e\ell = 2\pi N$, the oscillation condition becomes

$$DN = \frac{2n + 1}{4} \tag{10-4-22}$$

Example 10-4-1:   Carcinotron Characteristics

A circular carcinotron has the operating parameters:

Anode voltage:	$V_0 = 20$ kV
Anode current:	$I_0 = 3.5$ A
Magnetic flux density:	$B_0 = 0.3$ Wb/m^2
Operating frequency:	$f = 4$ GHz
Characteristic impedance:	$Z_0 = 50\ \Omega$
D factor:	$D = 0.8$
b factor:	$b = 0.5$

Compute:

    a. The dc electron velocity

    b. The electron-beam phase constant

    c. The delta differentials

    d. The propagation constants

    e. The oscillation condition

Solution

a. The dc electron velocity is

$$\mathcal{V}_0 = 0.593 \times 10^6 \times (20 \times 10^3)^{1/2} = 0.8386 \times 10^8 \text{ m/s}$$

b. The electron-beam phase constant is

$$\beta_e = \frac{\omega}{\mathcal{V}_0} = \frac{2\pi \times 4 \times 10^9}{0.8386 \times 10^8} = 300 \text{ rad/m}$$

c. The delta differentials are

$$\delta_1 = j\frac{0.5 - \sqrt{(0.5)^2 + 4}}{2} = -j0.78$$

$$\delta_2 = j\frac{0.5 + \sqrt{(0.5)^2 + 4}}{2} = j1.28$$

d. The propagation constants are

$$\gamma_1 = j(\beta_e + b) + \beta_e D\delta_1$$

$$= j(300 + 0.5) + 0.5 \times 0.8 \times (-j0.78) = j300.20$$

$$\gamma_2 = j(300 + 0.5) + 0.5 \times 0.8 \times (-j1.28) = j301.00$$

e. The oscillation occurs at

$$DN = 1.25 \qquad \text{for } n = 1$$

then $$N = 1.5625$$

and $$\ell = \frac{2\pi N}{\beta_e} = \frac{2\pi \times 1.5625}{300} = 3.27 \text{ cm}$$

REFERENCES

[1] HUTTER, R. G. E., *Beam and Wave Electrons in Microwave Tubes*. D. Van Nostrand Company, Princeton, N.J., 1960.

[2] SKOWRON, J. F., The Continuous-Cathode (Emitting-Sole) Crossed-Field Amplifier. *Proc. IEEE*, **61**, No. 3, March 1973, 330–56.

[3] BROWN, W. C., The microwave magnetron and its derivatives. *IEEE Trans. on Electron Devices*, **ED-31**, No. 11, November 1984, 1595–1605.

[4] PIERCE, J. R., *Traveling-Wave Tubes*, P. 210. D. Van Nostrand Company, Princeton, N.J., 1950.

[5] BECK, A. H. W., *Space-Charge Waves and Slow Electromagnetic Waves*, P. 250. Pergamon Press, New York, 1958.

[6] GEWARTOWSKI, J. V., and H. A. WATSON, *Principles of Electron Tubes*, P. 391. D. Van Nostrand Company, Princeton, N.J., 1965.

PROBLEMS

Magnetrons

10-1. Describe the principle of operation for a normal cylindrical magnetron and its characteristics.

10-2. A normal cylindrical magnetron has the following parameters:

Inner radius:	$R_a = 0.15$ meter
Outer radius:	$R_b = 0.45$ meter
Magnetic flux density:	$B_0 = 1.2$ milliwebers/m²

   a. Determine the Hull cutoff voltage.
   b. Determine the cutoff magnetic flux density if the beam voltage V_0 is 6000 V.

10-3. It is assumed that in a normal cylindrical magnetron the inner cylinder of radius a carries a current of I_0 in the z direction (i.e., $I = I_0 u_z$) and the anode voltage is V_0. The outer radius is b. Determine the differential equation in terms of the anode voltage V_0 and the current I_0.

10-4. Compare the cutoff conditions for an inverted cylindrical magnetron (i.e., the inner cathode voltage is V_0 and the outer anode is grounded) with a normal cylindrical magnetron. It is assumed that the magnetic field does no work on the electrons.

10-5. It is assumed that electrons in an inverted cylindrical magnetron leave the interior of the coaxial cathode with initial velocity caused by thermal voltage V_t in volts. Find the initial velocity required for the electrons to just hit the anode at the center conductor.

10-6. It is assumed that electrons in an inverted cylindrical magnetron leave the interior of the coaxial cathode with zero initial velocity. Find the minimum velocity for an electron to just graze the anode at the center conductor.

10-7. In a linear magnetron the electric and magnetic field intensities as shown in Fig. P10-7 are given by

$$\mathbf{E} = E_2 \mathbf{u}_z = -\frac{V_0}{d} \mathbf{u}_z$$

$$\mathbf{B} = B_y \mathbf{u}_y$$

Determine the trajectory of an electron with an initial velocity $\mathcal{V}_0$ in the z direction.

$$\mathbf{E} = E_2 \mathbf{u}_z = -\frac{V_0}{d} \mathbf{u}_z$$

$$\mathbf{B} = B_y \mathbf{u}_y$$
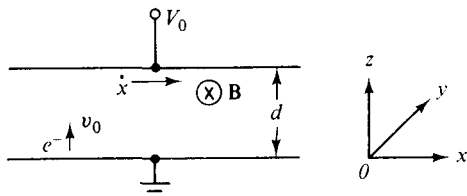
Figure P10-7

—

10-8. An X-band pulsed cylindrical magnetron has the following parameters:

$$Anode\ voltage: \qquad V_0 = 32\ kV$$

Anode voltage:	$V_0 = 32$ kV
Anode current:	$I_0 = 84$ A
Magnetic flux density:	$B_0 = 0.01$ Wb/m^2
Radius of cathode cylinder:	$a = 6$ cm
Radius of vane edge to center:	$b = 12$ cm

Compute:
a. The cyclotron angular frequency
b. The cutoff voltage for a fixed B_0
c. The cutoff magnetic flux density for a fixed V_0

10-9. An X-band pulsed conventional magnetron has the following parameters:

Anode voltage:	$V_0 = 22$ kV
Anode current:	$I_0 = 28$ A
Operating frequency:	$f = 10$ GHz
Resonator conductance:	$G_r = 3 \times 10^{-4}\ \mho$
Loaded conductance:	$G_\ell = 3 \times 10^{-5}\ \mho$
Vane capacitance:	$C = 3$ pF
Duty cycle:	$DC = 0.001$
Power loss:	$P_{\text{loss}} = 200$ kW

Compute:
a. The angular resonant frequency
b. The unloaded quality factor Q_{un}
c. The loaded quality factor Q_ℓ
d. The external quality factor Q_{ex}
e. The circuit efficiency
f. The electronic efficiency

10-10. A linear magnetron has the following parameters:

Anode voltage:	$V_0 = 20$ kV
Anode current:	$I_0 = 17$ A
Magnetic flux density:	$B_0 = 0.01$ Wb/m^2
Distance between cathode and anode:	$d = 6$ cm

Calculate:
a. The Hull cutoff voltage for a fixed B_0
b. The Hull cutoff magnetic flux density for a fixed V_0

10-11. A linear magnetron has the following parameters:

Anode voltage:	$V_0 = 32$ kV
Anode current:	$I_0 = 60$ A
Operating frequency:	$f = 10$ GHz

Magnetic flux density: $B_0 = 0.01$ Wb/m^2

Hub thickness: $h = 3$ cm

Distance between anode and cathode: $d = 6$ cm

Compute:
a. The electron velocity at the hub surface
b. The phase velocity for synchronism
c. The Hartree anode voltage

10-12. An inverted coaxial magnetron has the following parameters:

Anode voltage: $V_0 = 30$ kV

Cathode current: $I_0 = 25$ A

Anode radius: $a = 2.5$ cm

Cathode radius: $b = 5$ cm

Magnetic flux density: $B_0 = 0.01$ Wb/m^2

Determine:
a. The cutoff voltage for a fixed B_0
b. The cutoff magnetic flux density for a fixed V_0

10-13. A frequency-agile coaxial magnetron has the following parameters:

Pulse duration: $\tau = 0.30, 0.60, 0.90 \ \mu$s

Duty cycle: $DC = 0.0011$

Pulse rate on target: $N = 15$ per scan

Determine:
a. The agile excursion
b. The pulse-to-pulse frequency separation
c. The signal frequency
d. The time for N pulse
e. The agile rate

Crossed-Field Amplifiers (CFAs)

10-14. Derive Eq. (10-2-4).

10-15. Derive the circuit-efficiency Eq. (10-2-5).

10-16. Derive Eqs. (10-2-8), (10-2-10), and (10-2-11).

10-17. A CFA has the following operating parameters:

Anode dc voltage: $V_{a0} = 3$ kV

Anode dc current: $I_{a0} = 3$ A

Electronic efficiency: $\eta_e = 25\%$

RF input power: $P_{in} = 100$ W

Compute:
a. The induced RF power

b. The total RF output power

c. The power gain in dB

10-18. A CFA operates under the following parameters:

Anode dc voltage:	$V_{a0} = 1.80$ kV
Anode dc current:	$I_{a0} = 1.30$ A
Electronic efficiency:	$\eta_e = 22\%$
RF input power:	$P_{in} = 70$ W

Calculate:

a. The induced RF power

b. The total RF output power

c. The power gain in dB

Amplitrons

10-19. An Amplitron operates under the following parameters:

Operating frequency:	$f = 9$ GHz
Anode voltage:	$V_0 = 20$ kV
Anode current:	$I_0 = 3.5$ A
Magnetic flux density:	$B_0 = 0.3$ Wb/m^2
Characteristic impedance:	$Z_0 = 50 \ \Omega$

Compute:

a. The dc electron-beam velocity

b. The electron-beam phase constant

c. The cyclotron angular frequency

d. The cyclotron phase constant

e. The gain parameter

10-20. An Amplitron has the following operating parameters:

Operating frequency:	$f = 10$ GHz
Anode voltage:	$V_0 = 25$ kV
Anode current:	$I_0 = 4$ A
Magnetic flux density:	$B_0 = 0.35$ Wb/m^2
Characteristic impedance:	$Z_0 = 50 \ \Omega$

Calculate:

a. The dc electron-beam velocity

b. The electron-beam phase constant

c. The cyclotron angular frequency

d. The cyclotron phase constant

e. The gain parameter

Carcinotrons

10-21. A circular Carcinotron operates under the following parameters:

Operating frequency:	$f = 8$ GHz
Anode voltage:	$V_0 = 25$ kV
Anode current:	$I_0 = 4$ A
Magnetic flux density:	$B_0 = 0.35$ Wb/m^2
Characteristic impedance:	$Z_0 = 50$ ohms
D factor:	$D = 0.75$
b factor:	$b = 0.50$

Calculate:
a. The dc electron velocity
b. The electron-beam phase constant
c. The delta differentials
d. The propagation constants
e. The oscillation condition

Chapter 11

Strip Lines

11-0 INTRODUCTION

Prior to 1965 nearly all microwave equipment utilized coaxial, waveguide, or parallel strip-line circuits. In recent years—with the introduction of monolithic microwave integrated circuits (MMICs)—microstrip lines and coplanar strip lines have been used extensively, because they provide one free and accessible surface on which solid-state devices can be placed. In this chapter parallel, coplanar, and shielded strip lines and microstrip lines, which are shown in Fig. 11-0-1 [1], are described.

(a) Schematic diagram   (b) Field lines
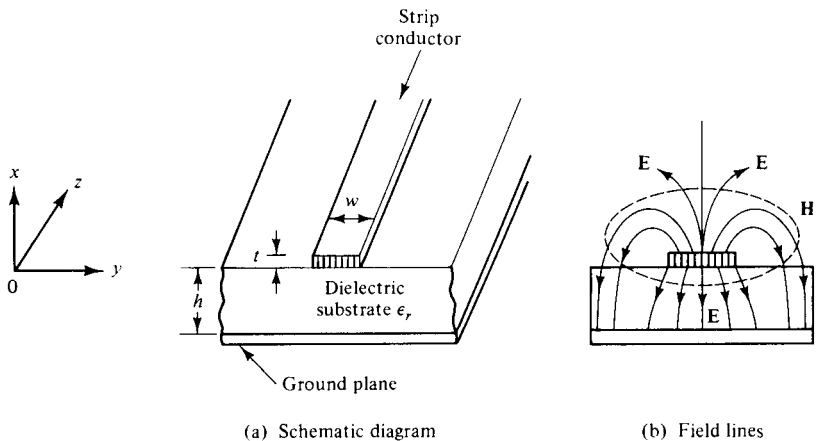
Figure 11-0-1   Schematic diagrams of strip lines.

11-1 *MICROSTRIP LINES*

Chapter 3 described and discussed conventional transmission lines in detail. All electrical and electronic devices with high-power output commonly use conventional lines, such as coaxial lines or waveguides, for power transmission. However, the microwave solid-state device is usually fabricated as a semiconducting chip with a volume on the order of $0.008-0.08$ mm^3. The method of applying signals to the chips and extracting output power from them is entirely different from that used for vacuum-tube devices. Microwave integrated circuits with microstrip lines are commonly used with the chips. The microstrip line is also called an *open-strip line*. In engineering applications, MKS units have not been universally adopted for use in designing the microstrip line. In this section we use either English units or MKS units, depending on the application, for practical purposes.

Modes on microstrip line are only quasi-transverse electric and magnetic (TEM). Thus the theory of TEM-coupled lines applies only approximately. Radiation loss in microstrip lines is a problem, particularly at such discontinuities as short-circuit posts, corners, and so on. However, the use of thin, high-dielectric materials considerably reduces the radiation loss of the open strip. A microstrip line has an advantage over the balanced-strip line because the open strip has better interconnection features and easier fabrication. Several researchers have analyzed the circuit of a microstrip line mounted on an infinite dielectric substrate over an infinite ground plane [2 to 5]. Numerical analysis of microstrip lines, however, requires large digital computers, whereas microstrip-line problems can generally be solved by conformal transformations without requiring complete numerical calculations.

11-1-1 *Characteristic Impedance of Microstrip Lines*

Microstrip lines are used extensively to interconnect high-speed logic circuits in digital computers because they can be fabricated by automated techniques and they provide the required uniform signal paths. Figure 11-1-1 shows cross sections of a microstrip line and a wire-over-ground line for purposes of comparison.

In Fig. 11-1-1(a) you can see that the characteristic impedance of a microstrip

(a)   (b)
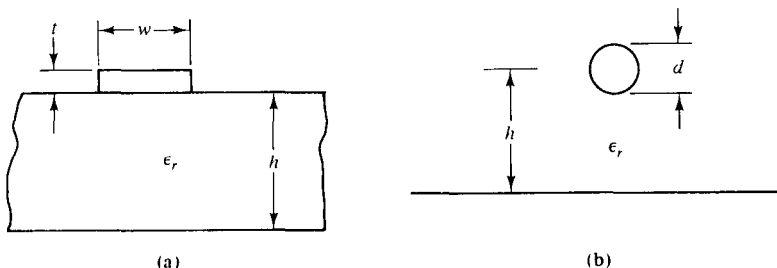
Figure 11-1-1   Cross sections of (a) a microstrip line and (b) a wire-over-ground line.

line is a function of the strip-line width, the strip-line thickness, the distance between the line and the ground plane, and the homogeneous dielectric constant of the board material. Several different methods for determining the characteristic impedance of a microstrip line have been developed. The field-equation method was employed by several authors for calculating an accurate value of the characteristic impedance [3 to 5]. However, it requires the use of a large digital computer and is extremely complicated. Another method is to derive the characteristic-impedance equation of a microstrip line from a well-known equation and make some changes [2]. This method is called a *comparative,* or an *indirect,* method. The well-known equation of the characteristic impedance of a wire-over-ground transmission line, as shown in Fig. 11-1-1(b), is given by

$$Z_0 = \frac{60}{\sqrt{\epsilon_r}} \ln \frac{4h}{d} \qquad \text{for } h \gg d \qquad (11\text{-}1\text{-}1)$$

where ϵ_r = dielectric constant of the ambient medium
$\quad h$ = the height from the center of the wire to the ground plane
$\quad d$ = diameter of the wire

If the effective or equivalent values of the relative dielectric constant ϵ_r of the ambient medium and the diameter d of the wire can be determined for the microstrip line, the characteristic impedance of the microstrip line can be calculated.

Effective dielectric constant ϵ_{re}.   For a homogeneous dielectric medium, the propagation-delay time per unit length is

$$T_d = \sqrt{\mu \epsilon} \qquad (11\text{-}1\text{-}2)$$

where μ is the permeability of the medium and ϵ is the permittivity of the medium. In free space, the propagation-delay time is

$$T_{df} = \sqrt{\mu_0 \epsilon_0} = 3.333 \quad \text{ns/m} \quad \text{or} \quad 1.016 \quad \text{ns/ft} \qquad (11\text{-}1\text{-}3)$$

where

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m, or } 3.83 \times 10^{-7} \text{ H/ft}$$

$$\epsilon_0 = 8.854 \times 10^{-12} \text{ F/m, or } 2.69 \times 10^{-12} \text{ F/ft}$$

In transmission lines used for interconnections, the relative permeability is 1. Consequently, the propagation-delay time for a line in a nonmagnetic medium is

$$T_d = 1.106 \sqrt{\epsilon_r} \qquad \text{ns/ft} \qquad (11\text{-}1\text{-}4)$$

The effective relative dielectric constant for a microstrip line can be related to the relative dielectric constant of the board material. DiGiacomo and his coworkers discovered an empirical equation for the effective relative dielectric constant of a microstrip line by measuring the propagation-delay time and the relative dielectric constant of several board materials, such as fiberglass-epoxy and nylon phenolic [6].

The empirical equation, as shown in Fig. 11-1-2, is expressed as

$$\epsilon_{re} = 0.475\epsilon_r + 0.67 \tag{11-1-5}$$

where ϵ_r is the relative dielectric constant of the board material and ϵ_{re} is the effective relative dielectric constant for a microstrip line.

Effective dielectric constant, ϵ_{re}

$\epsilon_{re} = 0.475\epsilon_r + 0.67$

Relative dielectric constant, ϵ_r
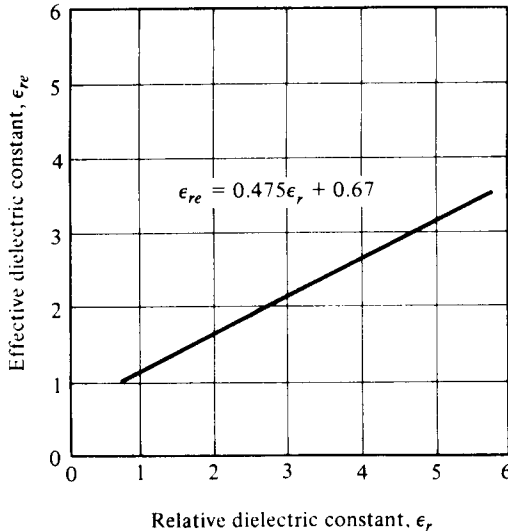
Figure 11-1-2 Effective dielectric constant as a function of relative dielectric constant for a microstrip line. (*After H. R., Kaupp [2]; reprinted by permission of IEEE, Inc.*)

Transformation of a rectangular conductor into an equivalent circular conductor.   The cross-section of a microstrip line is rectangular, so the rectangular conductor must be transformed into an equivalent circular conductor. Springfield discovered an empirical equation for the transformation [7]. His equation is

$$d = 0.67w\left(0.8 + \frac{t}{w}\right) \tag{11-1-6}$$

where d = diameter of the wire over ground
$\quad\quad w$ = width of the microstrip line
$\quad\quad t$ = thickness of the microstrip line

The limitation of the ratio of thickness to width is between 0.1 and 0.8, as indicated in Fig. 11-1-3.

Characteristic impedance equation.   Substituting Eq. (11-1-5) for the dielectric constant and Eq. (11-1-6) for the equivalent diameter in Eq. (11-1-1) yields

$$Z_0 = \frac{87}{\sqrt{\epsilon_r + 1.41}}\ln\left[\frac{5.98h}{0.8w + t}\right] \quad\quad \text{for } (h < 0.8w) \tag{11-1-7}$$

Figure 11-1-13 graph: y-axis "Ratio of conductor diameter to width, d/w" ranging 0 to 1.3; x-axis "Ratio of strip thickness to width, t/w" ranging 0 to 1.0. Line labeled:

$$\frac{d}{w} = 0.67 \left(0.8 + \frac{t}{w}\right)$$
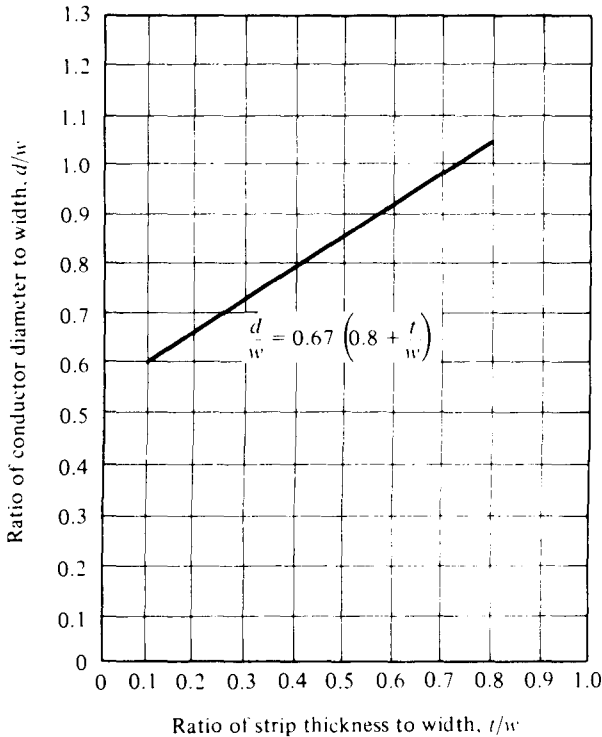
Figure 11-1-13 Relationship between a round conductor and a rectangular conductor far from its ground plane. (*After H. R. Kaupp [2]; reprinted by permission of IEEE, Inc.*)

where ϵ_r = relative dielectric constant of the board material
h = height from the microstrip line to the ground
w = width of the microstrip line
t = thickness of the microstrip line

Equation (11-1-7) is the equation of characteristic impedance for a narrow microstrip line. The velocity of propagation is

$$v = \frac{c}{\sqrt{\epsilon_{re}}} = \frac{3 \times 10^8}{\sqrt{\epsilon_{re}}} \qquad \text{m/s} \qquad (11\text{-}1\text{-}8)$$

The characteristic impedance for a wide microstrip line was derived by Assadourian and others [8] and is expressed by

$$Z_0 = \frac{h}{w} \sqrt{\frac{\mu}{\epsilon}} = \frac{377}{\sqrt{\epsilon_r}} \frac{h}{w} \qquad \text{for } (w \gg h) \qquad (11\text{-}1\text{-}9)$$

Limitations of Equation (11-1-7).   Most microstrip lines are made from boards of copper with a thickness of 1.4 or 2.8 mils (1 or 2 ounces of copper per square foot). The narrowest widths of lines in production are about 0.005–0.010 in. Line widths are usually less than 0.020 in.; consequently, ratios of thickness to width of less than 0.1 are uncommon. The straight-line approximation from Eq. (11-1-6) is an accurate value of characteristic impedance, or the ratio of thickness to width between 0.1 and 0.8.

Since the dielectric constant of the materials used does not vary excessively with frequency, the dielectric constant of a microstrip line can be considered independent of frequency. The validity of Eq. (11-1-7) is doubtful for values of dielectric thickness h that are greater than 80% of the line width w. Typical values for the characteristic impedance of a microstrip line vary from 50 Ω to 150 Ω, if the values of the parameters vary from $\epsilon_r = 5.23$, $t = 2.8$ mils, $w = 10$ mils, and $h = 8$ mils to $\epsilon_r = 2.9$, $t = 2.8$ mils, $w = 10$ mils, and $h = 67$ mils [2].

Example 11-1-1:   Characteristic Impedance of Microstrip Line

A certain microstrip line has the following parameters:

$\epsilon_r = 5.23$
$h = 7$ mils
$t = 2.8$ mils
$w = 10$ mils

Calculate the characteristic impedance Z_0 of the line.

Solution

$$Z_0 = \frac{87}{\sqrt{\epsilon_r + 1.41}} \ln \left[\frac{5.98h}{0.8w + t} \right]$$

$$= \frac{87}{\sqrt{5.23 + 1.41}} \ln \left[\frac{5.98 \times 7}{0.8 \times 10 + 2.8} \right]$$

$$= 45.78 \ \Omega$$

11-1-2 Losses in Microstrip Lines

Microstrip transmission lines consisting of a conductive ribbon attached to a dielectric sheet with conductive backing (see Fig. 11-1-4) are widely used in both microwave and computer technology. Because such lines are easily fabricated by printed-circuit manufacturing techniques, they have economic and technical merit.
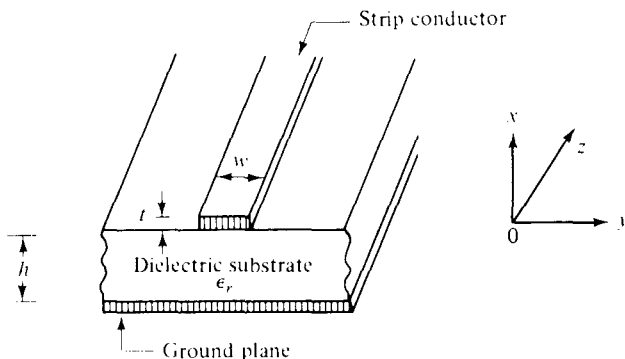
Figure 11-1-4   Schematic diagram of a microstrip line.

The characteristic impedance and wave-propagation velocity of a microstrip line was analyzed in Section 11-1-1. The other characteristic of the microstrip line is its attenuation. The attenuation constant of the dominant microstrip mode depends on geometric factors, electrical properties of the substrate and conductors, and on the frequency. For a nonmagnetic dielectric substrate, two types of losses occur in the dominant microstrip mode: (1) dielectric loss in the substrate and (2) ohmic skin loss in the strip conductor and the ground plane. The sum of these two losses may be expressed as losses per unit length in terms of an attenuation factor α. From ordinary transmission-line theory, the power carried by a wave traveling in the positive z direction is given by

$$P = \tfrac{1}{2} VI^* = \tfrac{1}{2}(V_+ e^{-\alpha z} I_+ e^{-\alpha z}) = \tfrac{1}{2}\frac{|V_+|^2}{Z_0} e^{-2\alpha z} = P_0 e^{-2\alpha z} \qquad (11\text{-}1\text{-}10)$$

where $P_0 = |V_+|^2/(2Z_0)$ is the power at $z = 0$.

The attenuation constant α can be expressed as

$$\alpha = -\frac{dP/dz}{2P(z)} = \alpha_d + \alpha_c \qquad (11\text{-}1\text{-}11)$$

where α_d is the dielectric attenuation constant and α_c is the ohmic attenuation constant.

The gradient of power in the z direction in Eq. (11-1-11) can be further expressed in terms of the power loss per unit length dissipated by the resistance and the power loss per unit length in the dielectric. That is,

$$-\frac{dP(z)}{dz} = -\frac{d}{dz}(\tfrac{1}{2} VI^*)$$

$$= \tfrac{1}{2}\left(-\frac{dV}{dz}\right) I^* + \tfrac{1}{2}\left(-\frac{dI^*}{dz}\right) V$$

$$= \tfrac{1}{2}(RI)I^* + \tfrac{1}{2}\sigma V^* V$$

$$= \tfrac{1}{2}|I|^2 R + \tfrac{1}{2}|V|^2 \sigma = P_c + P_d \qquad (11\text{-}1\text{-}12)$$

where σ is the conductivity of the dielectric substrate board.

Substitution of Eq. (11-1-12) into Eq. (11-1-11) results in

$$\alpha_d \simeq \frac{P_d}{2P(z)} \qquad \text{Np/cm} \qquad (11\text{-}1\text{-}13)$$

and

$$\alpha_c \simeq \frac{P_c}{2P(z)} \qquad \text{Np/cm} \qquad (11\text{-}1\text{-}14)$$

Dielectric losses. As stated in Section 2-5-3, when the conductivity of a dielectric cannot be neglected, the electric and magnetic fields in the dielectric are no longer in time phase. In that case the dielectric attenuation constant, as expressed in Eq. (2-5-20), is given by

$$\alpha_d = \frac{\sigma}{2} \sqrt{\frac{\mu}{\epsilon}} \qquad \text{Np/cm} \qquad (11\text{-}1\text{-}15)$$

where σ is the conductivity of the dielectric substrate board in $\mho$/cm. This dielectric constant can be expressed in terms of dielectric loss tangent as shown in Eq. (2-5-17):

$$\tan \theta = \frac{\sigma}{\omega\epsilon} \qquad (11\text{-}1\text{-}16)$$

Then the dielectric attenuation constant is expressed by

$$\alpha_d = \frac{\omega}{2} \sqrt{\mu\epsilon} \tan \theta \qquad \text{Np/cm} \qquad (11\text{-}1\text{-}17)$$

Since the microstrip line is a nonmagnetic mixed dielectric system, the upper dielectric above the microstrip ribbon is air, in which no loss occurs. Welch and Pratt [9] derived an expression for the attenuation constant of a dielectric substrate. Later on, Pucel and his coworkers [10] modified Welch's equation [9]. The result is

$$\alpha_d = 4.34 \frac{q\sigma}{\sqrt{\epsilon_{re}}} \sqrt{\frac{\mu_0}{\epsilon_0}}$$

$$= 1.634 \times 10^3 \frac{q\sigma}{\sqrt{\epsilon_{re}}} \qquad \text{dB/cm} \qquad (11\text{-}1\text{-}18)$$

In Eq. (11-1-18) the conversion factor of 1 Np = 8.686 dB is used, ϵ_{re} is the effective dielectric constant of the substrate, as expressed in Eq. (11-1-5), and q denotes the dielectric filling factor, defined by Wheeler [3] as

$$q = \frac{\epsilon_{re} - 1}{\epsilon_r - 1} \qquad (11\text{-}1\text{-}19)$$

We usually express the attenuation constant per wavelength as

$$\alpha_d = 27.3 \left(\frac{q\epsilon_r}{\epsilon_{re}}\right) \frac{\tan \theta}{\lambda_g} \qquad \text{dB/}\lambda_g \qquad (11\text{-}1\text{-}20)$$

where $\lambda_g = \dfrac{\lambda_0}{\sqrt{\epsilon_{re}}}$ and λ_0 is the wavelength in free space, or

$\lambda_g = \dfrac{c}{f\sqrt{\epsilon_{re}}}$ and c is the velocity of light in vacuum.

If the loss tangent, $\tan \theta$, is independent of frequency, the dielectric attenuation per wavelength is also independent of frequency. Moreover, if the substrate conductivity is independent of frequency, as for a semiconductor, the dielectric attenuation per unit is also independent of frequency. Since q is a function of ϵ_r and w/h, the filling factors for the loss tangent $q\epsilon_n/\epsilon_{re}$ and for the conductivity $q/\sqrt{\epsilon_{re}}$ are also functions of these quantities. Figure 11-1-5 shows the loss-tangent filling factor against w/h for a range of dielectric constants suitable for microwave inte-
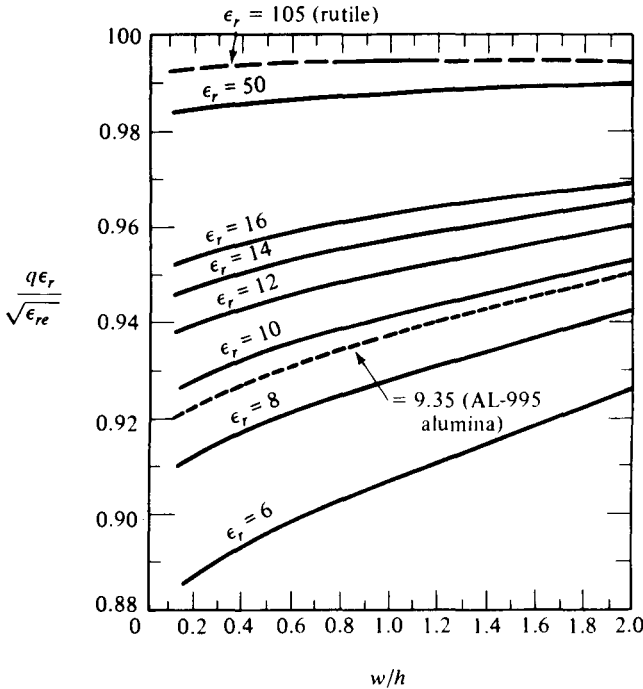
Figure 11-1-5 Filling factor for loss tangent of microstrip substrate as a function of w/h. (*After R. A. Pucel et al.* [10]; *reprinted by permission of IEEE, Inc.*)

grated circuits. For most practical purposes, this factor is considered to be 1. Figure 11-1-6 illustrates the product $\alpha_d \rho$ against w/h for two semiconducting substrates, silicon and gallium arsenide, that are used for integrated microwave circuits. For design purposes, the conductivity filling factor, which exhibits only a mild dependence on w/h, can be ignored.

Ohmic losses. In a microstrip line over a low-loss dielectric substrate, the predominant sources of losses at microwave frequencies are the nonperfect conductors. The current density in the conductors of a microstrip line is concentrated in a sheet that is approximately a skin depth thick inside the conductor surface and exposed to the electric field. Both the strip conductor thickness and the ground plane thickness are assumed to be at least three or four skin depths thick. The current density in the strip conductor and the ground conductor is not uniform in the transverse plane. The microstrip conductor contributes the major part of the ohmic loss. A diagram of the current density J for a microstrip line is shown in Fig. 11-1-7.

Because of mathematical complexity, exact expressions for the current density of a microstrip line with nonzero thickness have never been derived [10]. Several researchers [8] have assumed, for simplicity, that the current distribution is uniform and equal to I/w in both conductors and confined to the region $|x| < w/2$. With this assumption, the conducting attenuation constant of a wide microstrip line is given by

$$\alpha_c \simeq \frac{8.686 R_s}{Z_0 w} \qquad \text{dB/cm} \quad \text{for } \frac{w}{h} > 1 \qquad (11\text{-}1\text{-}21)$$
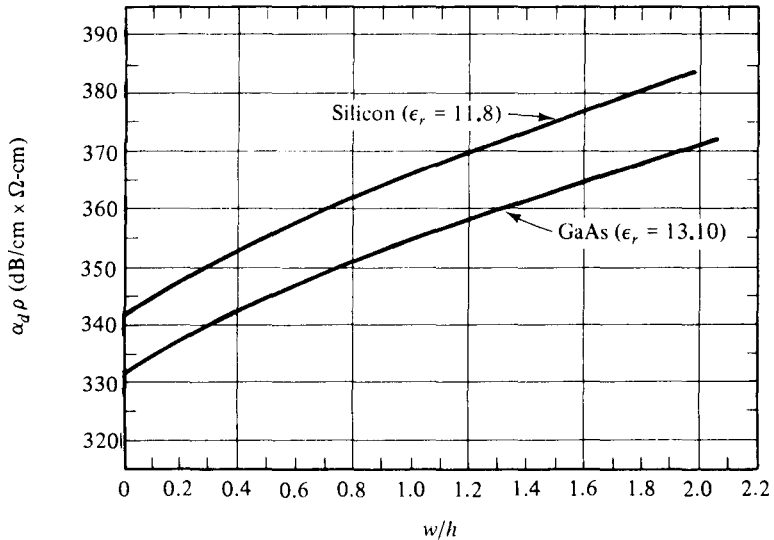
Figure 11-1-6 Dielectric attenuation factor of microstrip as a function of w/h for silicon and gallium arsenide substrates. (*After R. A. Pucel et al. [10]; reprinted by permission of IEEE, Inc.*)
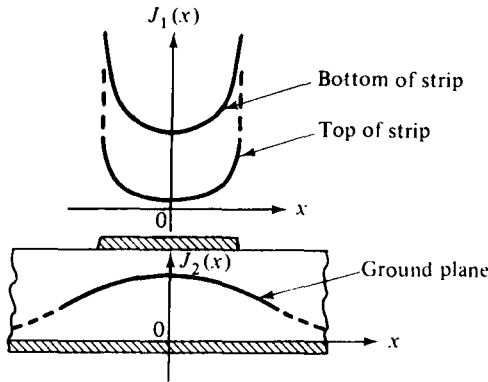
Figure 11-1-7 Current distribution on microstrip conductors. (*After R. A. Pucel et al. [10]; reprinted by permission of IEEE, Inc.*)

where $R_s = \sqrt{\dfrac{\pi f \mu}{\sigma}}$ is the surface skin resistance in Ω/square,

$R_s = \dfrac{1}{\delta \sigma}$ is Ω/square

$\delta = \sqrt{\dfrac{1}{\pi f \mu \sigma}}$ is the skin depth in cm

For a narrow microstrip line with $w/h < 1$, however, Eq. (11-1-21) is not applicable. The reason is that the current distribution in the conductor is not uniform, as assumed. Pucel and his coworkers [10, 11] derived the following three formulas from the results of Wheeler's work [3]:

$$\frac{\alpha_c Z_0 h}{R_s} = \frac{8.68}{2\pi} \left[1 - \left(\frac{w'}{4h}\right)^2 \right] \left[1 + \frac{h}{w'} + \frac{h}{\pi w'} \left(\ln \frac{4\pi w}{t} + \frac{t}{w} \right) \right]$$

$$\text{for } \frac{w}{h} \leq \frac{1}{2\pi} \quad (11\text{-}1\text{-}22)$$

$$\frac{a_c Z_0 h}{R_s} = \frac{8.68}{2\pi} \left[1 - \left(\frac{w'}{4h}\right)^2 \right] \left[1 + \frac{h}{w'} + \frac{h}{w'} \left(\ln \frac{2h}{t} - \frac{t}{h} \right) \right]$$

$$\text{for } \frac{1}{2\pi} < \frac{w}{h} \leq 2 \quad (11\text{-}1\text{-}23)$$

and

$$\frac{\alpha_c Z_0 h}{R_s} = \frac{8.68}{\left\{ \frac{w'}{h} + \frac{2}{\pi} \ln \left[2\pi e \left(\frac{w'}{2h} + 0.94 \right) \right] \right\}^2} \left[\frac{w'}{h} + \frac{w'/(\pi h)}{\frac{w'}{2h} + 0.94} \right]$$

$$\times \left[1 + \frac{h}{w'} + \frac{h}{\pi w'} \left(\ln \frac{2h}{t} - \frac{t}{h} \right) \right] \qquad \text{for } 2 \leq \frac{w}{h} \quad (11\text{-}1\text{-}24)$$

where α_c is expressed in dB/cm and

$$e = 2.718$$

$$w' = w + \Delta w \qquad (11\text{-}1\text{-}25)$$

$$\Delta w = \frac{t}{\pi} \left(\ln \frac{4\pi w}{t} + 1 \right) \qquad \text{for } \frac{2t}{h} < \frac{w}{h} \leq \frac{\pi}{2} \quad (11\text{-}1\text{-}26)$$

$$\Delta w = \frac{t}{\pi} \left(\ln \frac{2h}{t} + 1 \right) \qquad \text{for } \frac{w}{h} \geq \frac{\pi}{2} \quad (11\text{-}1\text{-}27)$$

The values of α_c obtained from solving Eqs. (11-1-22) through (11-1-24) are plotted in Fig. 11-1-8. For purposes of comparison, values of α_c based on Assadourian and Rimai's Eq. (11-1-21) are also shown.

Radiation losses. In addition to the conductor and dielectric losses, microstrip line also has radiation losses. The radiation loss depends on the substrate's thickness and dielectric constant, as well as its geometry. Lewin [12] has calculated the radiation loss for several discontinuities using the following approximations:

1. TEM transmission
2. Uniform dielectric in the neighborhood of the strip, equal in magnitude to an effective value
3. Neglect of radiation from the transverse electric (TE) field component parallel to the strip
4. Substrate thickness much less than the free-space wavelength
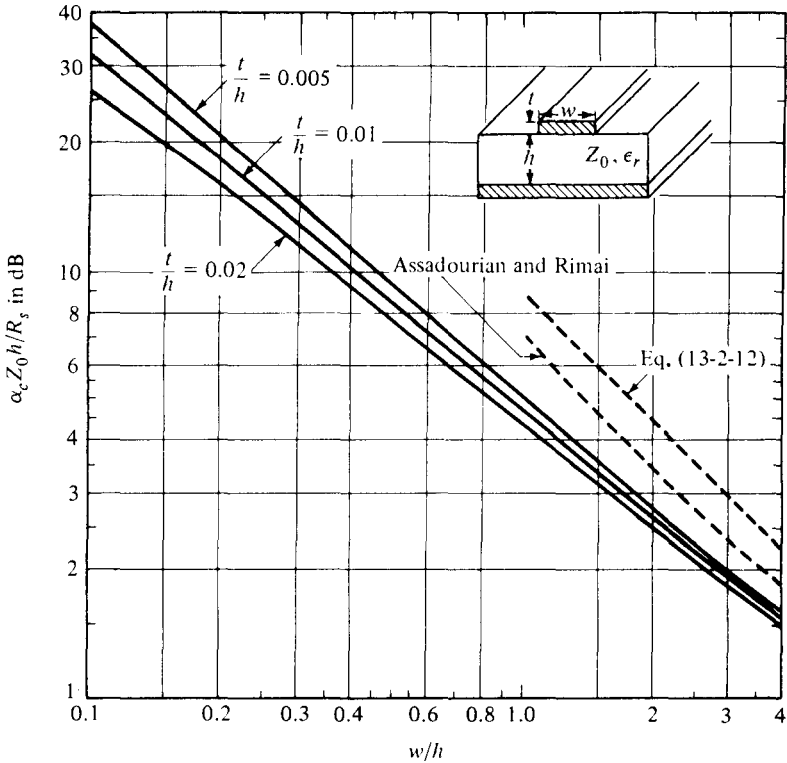
Figure 11-1-8   Theoretical conductor attenuation factor of microstrip as a function of w/h. (*After R. A. Pucel et al. [10]; reprinted by permission of IEEE, Inc.*)

Lewin's results show that the ratio of radiated power to total dissipated power for an open-circuited microstrip line is

$$\frac{P_{\text{rad}}}{P_t} = 240\pi^2 \left(\frac{h}{\lambda_0}\right)^2 \frac{F(\epsilon_{re})}{Z_0} \tag{11-1-28}$$

where $F(\epsilon_{re})$ is a radiation factor given by

$$F(\epsilon_{re}) = \frac{\epsilon_{re} + 1}{\epsilon_{re}} - \frac{\epsilon_{re} - 1}{2\epsilon_{re}\sqrt{\epsilon_{re}}} \ln \frac{\sqrt{\epsilon_{re}} + 1}{\sqrt{\epsilon_{re}} - 1} \tag{11-1-29}$$

in which ϵ_{re} is the effective dielectric constant and $\lambda_0 = c/f$ is the free-space wavelength.

The radiation factor decreases with increasing substrate dielectric constant. So, alternatively, Eq. (11-1-28) can be expressed as

$$\frac{P_{\text{rad}}}{P_t} = \frac{R_r}{Z_0} \tag{11-1-30}$$

where R_r is the radiation resistance of an open-circuited microstrip and is given by

$$R_r = 240\pi^2 \left(\frac{h}{\lambda_0}\right)^2 F(\epsilon_{re}) \tag{11-1-31}$$

The ratio of the radiation resistance R_r to the real part of the characteristic impedance Z_0 of the microstrip line is equal to a small fraction of the power radiated from a single open-circuit discontinuity. In view of Eq. (11-1-28), the radiation loss decreases when the characteristic impedance increases. For lower dielectric-constant substrates, radiation is significant at higher impedance levels. For higher dielectric-constant substrates, radiation becomes significant until very low impedance levels are reached.

11-1-3 Quality Factor Q of Microstrip Lines

Many microwave integrated circuits require very high quality resonant circuits. The quality factor Q of a microstrip line is very high, but it is limited by the radiation losses of the substrates and with low dielectric constant. Recall that for uniform current distribution in the microstrip line, the ohmic attenuation constant of a wide microstrip line is given by Eq. (11-1-21) as

$$\alpha_c = \frac{8.686R_s}{Z_0 w} \qquad \text{dB/cm}$$

and that the characteristic impedance of a wide microstrip line, as shown in Eq. (11-1-9), is

$$Z_0 = \frac{h}{w}\sqrt{\frac{\mu}{\epsilon}} = \frac{377}{\sqrt{\epsilon_r}}\frac{h}{w} \qquad \Omega$$

The wavelength in the microstrip line is

$$\lambda_g = \frac{30}{f\sqrt{\epsilon_r}} \qquad \text{cm} \tag{11-1-32}$$

where f is the frequency in GHz.

Since Q_c is related to the conductor attenuation constant by

$$Q_c = \frac{27.3}{\alpha_c} \tag{11-1-33}$$

where α_c is in dB/λ_g, Q_c of a wide microstrip line is expressed as

$$Q_c = 39.5 \left(\frac{h}{R_s}\right) f_{\text{GHz}} \tag{11-1-34}$$

where h is measured in cm and R_s is expressed as

$$R_s = \sqrt{\frac{\pi f \mu}{\sigma}} = 2\pi\sqrt{\frac{f_{\text{GHz}}}{\sigma}} \qquad \Omega/\text{square.} \tag{11-1-35}$$

Finally, the quality factor Q_c of a wide microstrip line is

$$Q_c = 0.63h \sqrt{\sigma f_{GHz}} \tag{11-1-36}$$

where α is the conductivity of the dielectric substrate board in $\mho/m$.

For a copper strip, $\alpha = 5.8 \times 10^7 \, \mho/m$ and Q_c becomes

$$Q_{Cu} = 4780h \sqrt{f_{GHz}} \tag{11-1-37}$$

For 25-mil alumina at 10 GHz, the maximum Q_c achievable from wide microstrip lines is 954 [13].

Similarly, a quality factor Q_d is related to the dielectric attenuation constant:

$$Q_d = \frac{27.3}{\alpha_d} \tag{11-1-38}$$

where α_d is in dB/λ_g.

Substituting Eq. (11-1-20) into Eq. (11-1-38) yields

$$Q_d = \frac{\lambda_0}{\sqrt{\epsilon_{re}} \tan \theta} \simeq \frac{1}{\tan \theta} \tag{11-1-39}$$

where λ_0 is the free-space wavelength in cm. Note that the Q_d for the dielectric attenuation constant of a microstrip line is approximately the reciprocal of the dielectric loss tangent θ and is relatively constant with frequency.

11-2 PARALLEL STRIP LINES

A parallel strip line consists of two perfectly parallel strips separated by a perfect dielectric slab of uniform thickness, as shown in Fig. 11-2-1. The plate width is w, the separation distance is d, and the relative dielectric constant of the slab is ϵ_{rd}.
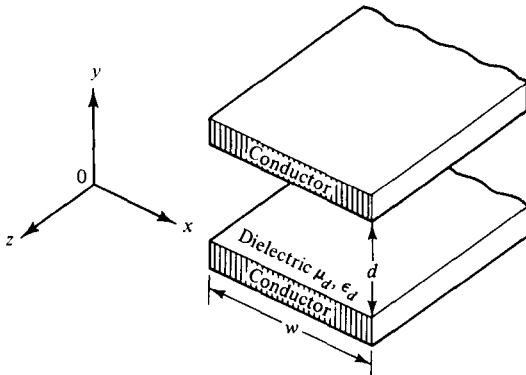
Figure 11-2-1   Schematic diagram of a parallel strip line.

11-2-1 Distributed Parameters

In a microwave integrated circuit a strip line can be easily fabricated on a dielectric substrate by using printed-circuit techniques. A parallel stripline is similar to a two-conductor transmission line, so it can support a quasi-TEM mode. Consider a TEM-mode wave propagating in the positive z direction in a lossless strip line ($R = G = 0$). The electric field is in the y direction, and the magnetic field is in the x direction. If the width w is much larger than the separation distance d, the fringing capacitance is negligible. Thus the equation for the inductance along the two conducting strips can be written as

$$L = \frac{\mu_c d}{w} \qquad \text{H/m} \qquad (11\text{-}2\text{-}1)$$

where μ_c is the permeability of the conductor. The capacitance between the two conducting strips can be expressed as

$$C = \frac{\epsilon_d w}{d} \qquad \text{F/m} \qquad (11\text{-}2\text{-}2)$$

where ϵ_d is the permittivity of the dielectric slab.
    If the two parallel strips have some surface resistance and the dielectric substrate has some shunt conductance, however, the parallel stripline would have some losses. The series resistance for both strips is given by

$$R = \frac{2R_s}{w} = \frac{2}{w}\sqrt{\frac{\pi f \mu_c}{\sigma_c}} \qquad \Omega/\text{m} \qquad (11\text{-}2\text{-}3)$$

where $R_s = \sqrt{(\pi f \mu_c)/\sigma_c}$ is the conductor surface resistance in Ω/square and σ_c is the conductor conductivity in $\mho$/m. The shunt conductance of the strip line is

$$G = \frac{\sigma_d w}{d} \qquad \mho/\text{m} \qquad (11\text{-}2\text{-}4)$$

where σ_d is the conductivity of the dielectric substrate.

11-2-2 Characteristic Impedance

The characteristic impedance of a lossless parallel strip line is

$$Z_0 = \sqrt{\frac{L}{C}} = \frac{d}{w}\sqrt{\frac{\mu_d}{\epsilon_d}} = \frac{377}{\sqrt{\epsilon_{rd}}}\frac{d}{w} \qquad \text{for } w \gg d \qquad (11\text{-}2\text{-}5)$$

The phase velocity along a parallel strip line is

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\mu_d \epsilon_d}} = \frac{c}{\sqrt{\epsilon_{rd}}} \qquad \text{m/s} \qquad \text{for } \mu_c = \mu_0 \quad (11\text{-}2\text{-}6)$$

The characteristic impedance of a lossy parallel strip line at microwave frequencies ($R \ll \omega L$ and $G \ll \omega C$) can be approximated as

$$Z_0 \simeq \sqrt{\frac{L}{C}} = \frac{377}{\sqrt{\epsilon_{rd}}} \frac{d}{w} \qquad \text{for } w \gg d \qquad (11\text{-}2\text{-}7)$$

11-2-3 Attenuation Losses

The propagation constant of a parallel strip line at microwave frequencies can be expressed by

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \qquad \text{for } R \ll \omega L \quad \text{and} \quad G \ll \omega C$$

$$\simeq \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) + j\omega\sqrt{LC} \qquad (11\text{-}2\text{-}8)$$

Thus the attenuation and phase constants are

$$\alpha = \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \qquad \text{Np/m} \qquad (11\text{-}2\text{-}9)$$

and

$$\beta = \omega\sqrt{LC} \qquad \text{rad/m} \qquad (11\text{-}2\text{-}10)$$

Substitution of the distributed parameters of a parallel strip line into Eq.(11-2-9) yields the attenuation constants for the conductor and dielectric losses:

$$\alpha_c = \frac{1}{2}R\sqrt{\frac{C}{L}} = \frac{1}{d}\sqrt{\frac{\pi f \epsilon_d}{\sigma_c}} \qquad \text{Np/m} \qquad (11\text{-}2\text{-}11)$$

and

$$\alpha_d = \frac{1}{2}G\sqrt{\frac{L}{C}} = \frac{188\sigma_d}{\sqrt{\epsilon_{rd}}} \qquad \text{Np/m} \qquad (11\text{-}2\text{-}12)$$

Example 11-2-1:   Characteristics of a Parallel Strip Line

A lossless parallel strip line has a conducting strip width w. The substrate dielectric separating the two conducting strips has a relative dielectric constant ϵ_{rd} of 6 (beryllia or beryllium oxide BeO) and a thickness d of 4 mm.

Calculate:

  a. The required width w of the conducting strip in order to have a characteristic impedance of 50 Ω
  b. The strip-line capacitance
  c. The strip-line inductance
  d. The phase velocity of the wave in the parallel strip line

Solution

  a. From Eq. (11-2-5) the width of the conducting strip is

$$w = \frac{377}{\sqrt{\epsilon_{rd}}} \frac{d}{Z_0} = \frac{377}{\sqrt{6}} \frac{4 \times 10^{-3}}{50}$$

$$= 12.31 \times 10^{-3} \text{ m}$$

b. The strip-line capacitance is

$$C = \frac{\epsilon_d w}{d} = \frac{8.854 \times 10^{-12} \times 6 \times 12.31 \times 101^{-3}}{4 \times 10^{-3}}$$

$$= 163.50 \text{ pF/m}$$

c. The strip-line inductance is

$$L = \frac{\mu_c d}{w} = \frac{4\pi \times 10^{-7} \times 4 \times 10^{-3}}{12.31 \times 10^{-3}}$$

$$= 0.41 \ \mu\text{H/m}$$

d. The phase velocity is

$$v_p = \frac{c}{\sqrt{\epsilon_{rd}}} = \frac{3 \times 10^8}{\sqrt{6}}$$

$$= 1.22 \times 10^8 \text{ m/s}$$

11-3 COPLANAR STRIP LINES

A coplanar trip line consists of two conducting strips on one substrate surface with one strip grounded, as shown in Fig. 11-3-1. The coplanar strip line has advantages over the conventional parallel strip line (see Section 11-2) because its two strips are on the same substrate surface for convenient connections. In microwave integrated circuits (MICs) the wire bonds have always presented reliability and reproducibility problems. The coplanar strip lines eliminate the difficulties involved in connecting the shunt elements between the hot and ground strips. As a result, reliability is increased, reproducibility is enhanced, and production cost is decreased.

The characteristic impedance of a coplanar strip line is

$$Z_0 = \frac{2 P_{\text{avg}}}{I_0^2} \tag{11-3-1}$$

where I_0 is the total peak current in one strip and P_{avg} is the average power flowing in the positive z direction. The average flowing power can be expressed as

$$P_{\text{avg}} = \tfrac{1}{2} \text{Re} \iint (\mathbf{E} \times \mathbf{H}^*) \cdot \mathbf{u}_z \, dx \, dy \tag{11-3-2}$$

where $\mathbf{E}_x$ = electric field intensity in the positive x direction
$\quad \mathbf{H}_y$ = magnetic field intensity in the positive y direction
$\quad *$ = conjugate
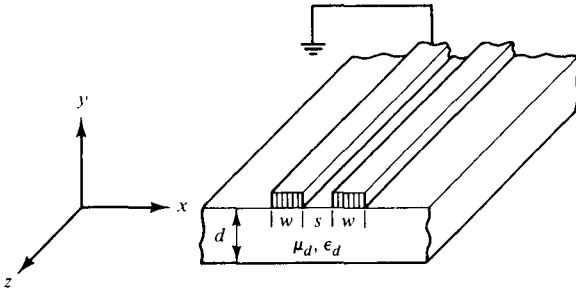
Figure 11-3-1 Schematic diagram of a coplanar strip line.

Example 11-3-1:   Characteristic Impedance of a Coplanar Strip Line

A coplanar strip line carries an average power of 250 mW and a peak current of 100 mA. Determine the characteristic impedance of the coplanar strip line.

Solution   From Eq. (11-3-1), the characteristic impedance of the coplanar strip line is

$$Z_0 = \frac{2 \times 250 \times 10^{-3}}{(100 \times 10^{-3})^2} = 50 \ \Omega$$

11-4 SHIELDED STRIP LINES

A partially shielded strip line has its strip conductor embedded in a dielectric medium, and its top and bottom ground planes have no connection, as shown in Fig. 11-4-1.
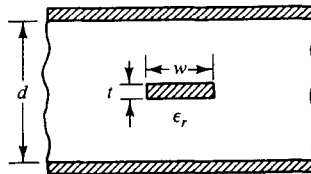
Figure 11-4-1 Partially shielded strip line.

The characteristic impedance for a wide strip $(w/d \gg 0.35)$ [14] is

$$Z_0 = \frac{94.15}{\sqrt{\epsilon_r}} \left(\frac{w}{d} K + \frac{C_f}{8.854\epsilon_r} \right)^{-1} \tag{11-4-1}$$

where $K = \dfrac{1}{1 - t/d}$

t = the strip thickness
d = the distance between the two ground planes
$C_f = \dfrac{8.854\epsilon_r}{\pi} [2K \ln (K + 1) - (K - 1) \ln (K^2 - 1)]$ and is the fringe capacitance in pF/m

Figure 11-4-2 shows the characteristic impedance Z_0 for a partially shielded strip line, with the t/d radio as a parameter.

Figure 11-4-2 Characteristic impedance Z_0 of a partially shielded strip line with the t/d ratio as a parameter. (*After S. Cohn [14]; reprinted by permission of IEEE, Inc.*)

Example 11-4-1:   Characteristic Impedance of a Shielded Strip Line

A shielded strip line has the following parameters:

Dielectric constant of the insulator
  (polystyrene): $\epsilon_r = 2.56$
Strip width: $w = 25$ mils
Strip thickness: $t = 14$ mils
Shield depth: $d = 70$ mils

Calculate:

a. The K factor
b. The fringe capacitance
c. The characteristic impedance of the line.

Solution

a. Using Eq. (11-4-1), the K factor is obtained:

$$K = \left(1 - \frac{t}{d}\right)^{-1} = \left(\frac{1 - 14}{70}\right)^{-1} = 1.25$$

b. From Eq. (11-4-1), the fringe capacitance is

$$C_f = \frac{8.854 \times 2.56}{3.1416}[2 \times 1.25 \ln (1.25 + 1) - (1.25 - 1) \ln (1.25^2 - 1)]$$

$$= 15.61 \text{ pF/m.}$$

c. The characteristic impedance from Eq. (11-4-1) is

$$Z_0 = \frac{94.15}{\sqrt{2.56}}\left[\frac{25}{70}(1.25) + \frac{15.61}{8.854 \times 2.56}\right]^{-1}$$

$$= 50.29 \ \Omega$$

REFERENCES

[1] LIAO, S. Y., *Engineering Applications of Electromagnetic Theory*, Chapter 3. West Publishing Co., St. Paul, Minn, 1988.

[2] KAUPP, H. R., Characteristics of microstrip transmission lines. *IEEE Trans. on Electronic Computers*, **EC-16,** No. 2, 185–193, April 1967.

[3] WHEELER, H. A., Transmission-line properties of parallel strips separated by a dielectric sheet, *IEEE Trans. on Microwave Theory and Techniques*, **MTT-3,** No. 3, 172–185, March 1965.

[4] BRYANT, T. G., and J. A. WEISS, Parameters of microstrip transmission lines and of coupled pairs of microstrip lines, *IEEE Trans. on Microwave Theory and Techniques*, **MTT-6,** No. 12, 1021–1027, December 1968.

[5] STINEHELFER, H. E., An accurate calculation of uniform microstrip transmission lines. *IEEE Trans. on Microwave Theory and Techniques*, **MTT-16,** No. 7, 439–443, July 1968.

[6] DiGIACOMO, J. J., et al., "Design and Fabrication of Nanosecond Digital Equipment," RCA, March 1965.

[7] SPRINGFIELD, A., *Simplified Theory of Microwave Transmission Systems*, F. Assodourian and E. Rimol. pp. 1651–1657, IRE Proceeding, December 1952.

[8] ASSODOURIAN, F., and E. RIMOL, Simplified theory of mirowave transmission systems. *Proc. IRE*, **40,** 1651–1657, December 1952.

[9] WELCH, J. D., and H. J. PRATT, Losses in microstrip transmission systems for integrated microwave circuits, *NEREM Rec.*, **8,** 100–101, (1966).

[10] PUCEL, R. A., D. J. MASSE, and C. P. HARTWIG, Losses in microstrip. *IEEE Trans. on Microwave Theory and Techniques*, **MTT-16,** No. 6, 342–350, June 1968.

[11] PUCEL, R. A., D. J. MASSE, and C. P. HARTWIG, Correction to "Losses in microstrip." *IEEE Trans. on Microwave Theory and Techniques,* **MTT-16,** No. 12, 1064, December 1968.

[12] LEWIN, L., Radiation from Discontinuities in Strip-Line, *IEEE Monograph No. 358E,* February 1960.

[13] VENDELINE, G. D., Limitations on stripline *Q. Microwave J.,* 63–69, May 1970.

[14] COHN, S., Characteristic impedance of the shielded-strip transmission line. *IRE Trans. on Microwave Theory and Techniques,* **MTT-2,** No. 7, 52, July 1954.

PROBLEMS

Microstrip Lines

11-1. A microstrip line has the following parameters:

ϵ_r = 5.23 and is the relative dielectric constant of the fiberglass board material

h = 0.8 mils

t = 2.8 mils

w = 10 mils

Write a FORTRAN program to complete the characteristic impedance Z_0 of the line. Use a READ statement to read in the input values, the F10.5 format for numerical outputs, and the Hollerith format for character outputs.

11-2. Since modes on microstrip lines are only quasi-transverse electric and magnetic (TEM), the theory of TEM-coupled lines applies only approximately. From the basic theory of a lossless line, show that the inductance L and capacitance C of a microstrip line are

$$L = \frac{Z_0}{v} = \frac{Z_0 \sqrt{\epsilon_r}}{c}$$

and

$$C = \frac{1}{Z_0 v} = \frac{\sqrt{\epsilon_r}}{Z_0 c}$$

where Z_0 = characteristic impedance of the microstrip line

v = wave velocity in the microstrip line

c = 3×10^8 m/s, the velocity of light in vacuum

ϵ_r = relative dielectric constant of the board material

11-3. A microstrip line is constructed of a perfect conductor and a lossless dielectric board. The relative dielectric constant of the fiberglass-epoxy board is 5.23, and the line characteristic impedance is 50 Ω. Calculate the line inductance and the line capacitance.

11-4. A microstrip line is constructed of a copper conductor and nylon phenolic board. The relative dielectric constant of the board material is 4.19, measured at 25 GHz, and its thickness is 0.4836 mm (19 mils). The line width is 0.635 mm (25 mils), and the line thickness is 0.071 mm (2.8 mils). Calculate the

    a. Characteristic impedance Z_0 of the microstrip line
    b. Dielectric filling factor q
    c. Dielectric attenuation constant α_d
    d. Surface skin resistivity R_s of the copper conductor at 25 GHz
    e. Conductor attenuation constant α_c

11-5. A microstrip line is made of a copper conductor 0.254 mm (10 mils) wide on a G-10 fiberglass-epoxy board 0.20 mm (8 mils) in height. The relative dielectric constant ϵ_r of the board material is 4.8, measured at 25 GHz. The microstrip line 0.035-mm (1.4 mils) thick is to be used for 10 GHz. Determine the
    a. Characteristic impedance Z_0 of the microstrip line
    b. Surface resistivity R_s of the copper conductor
    c. Conductor attenuation constant α_c
    d. Dielectric attenuation constant α_d
    e. Quality factors Q_c and Q_d

Parallel Striplines

11-6. A gold parallel stripline has the following parameters:

Relative dielectric constant of teflon:	$\epsilon_{rd} = 2.1$
Strip width:	$w = 26$ mm
Separation distance:	$d = 5$ mm
Conductivity of gold:	$\sigma_c = 4.1 \times 10^7$ ℧/m
Frequency:	$f = 10$ GHz

Determine the
    a. Surface resistance of the gold strip
    b. Characteristic impedance of the strip line
    c. Phase velocity

11-7. A gold parallel strip line has the following parameters:

Relative dielectric constant of polyethylene:	$\epsilon_{rd} = 2.25$
Strip width:	$w = 25$ mm
Separation distance:	$d = 5$ mm

Calculate the
    a. Characteristic impedance of the strip line
    b. Strip-line capacitance
    c. Strip-line inductance
    d. Phase velocity

Coplanar Strip Lines

11-8. A 50-Ω coplanar strip line has the following parameters:

Relative dielectric constant of alumina:	$\epsilon_{rd} = 10$
Strip width:	$w = 4$ mm
Strip thickness:	$t = 1$ mm

TEM-mode field intensities:

$$E_y = 3.16 \times 10^3 \sin\left(\frac{\pi x}{w}\right) e^{-j\beta z}$$

$$H_x = 63.20 \sin\left(\frac{\pi x}{w}\right) e^{-j\beta z}$$

Find the
  a. Average power flow
  b. Peak current in one strip

11-9. A shielded stripline has the following parameters:

Relative dielectric constant of the
  insulator polyethylene: $\epsilon_{rd} = 2.25$
Strip width: $w = 2$ mm
Strip thickness: $t = 0.5$ mm
Shield depth: $d = 4$ mm

Calculate the
  a. K factor
  b. Fringe capacitance
  c. Characteristic impedance

11-10. A shielded strip line is made of a gold strip in a polystyrene dielectric insulator and has the following parameters:

Relative dielectric constant of
  polystyrene: $\epsilon_{rd} = 2.56$
Strip width: $w = 0.7$ mm
Strip thickness: $t = 1.4$ mm
Shield depth: $d = 3.5$ mm

Determine the
  a. K factor
  b. Fringe capacitance
  c. Characteristic impedance

Chapter 12

Monolithic Microwave Integrated Circuits

12-0 INTRODUCTION

Integrated circuits are a combination of active and passive elements that are manufactured by successive diffusion or ion implantation processes on a semiconductor substrate. The active elements are generally silicon planar chips. The passive elements are either thin or thick film components. In thin films, a thin film of conducting (resistor) or nonconducting (capacitor) material is deposited on a passive insulated substrate, such as ceramic, glass, or silicon dioxide, by vacuum deposition. Thick film refers to films more than several thousand angstroms ($\mathring{A}$) thick. Such films are used almost exclusively to form resistors, and the pattern is usually defined by silk-screening.

The integrated-circuit (IC) complexity has advanced from small-scale integration (SSI) for up to 100 components per chip, to medium-scale integration (MSI) for up to 1000 components per chip, to large-scale integration (LSI) for up to 10^5 components per chip, and finally the very large-scale integration (VLSI) for more than 1 million components per chip. Recently, the integrated circuit has advanced to the ultralarge-scale integration (ULSI) stage. For example, a 32-bit microprocessor chip contains more than 0.15×10^6 components, and a 1-megabit dynamic random-access memory (DRAM) chip contains more than 2.2×10^6 components. In this chapter, we discuss the integrated circuit devices.

Electronic circuits can be classified into three categories according to circuit technology as shown in Fig. 12-0-1.

1. **Discrete circuit (DC):** The conventional electrical or electronic circuit, in which the elements are separately manufactured and then connected together
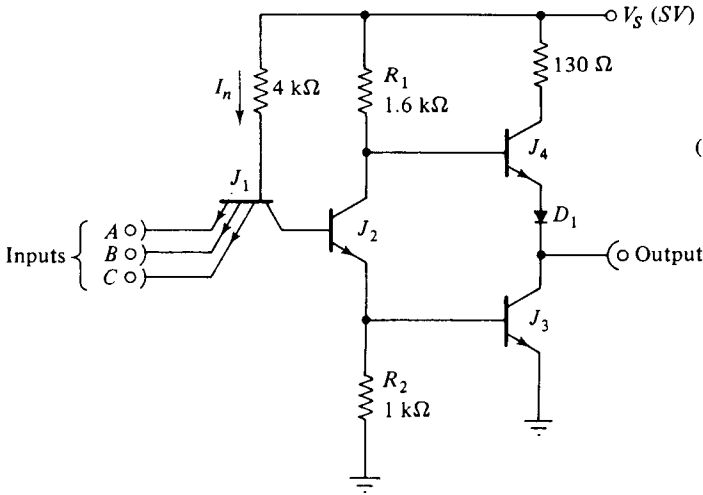
(1) Discrete circuit where elements are separately made

(2) Integrated circuit for J_1 of (1)
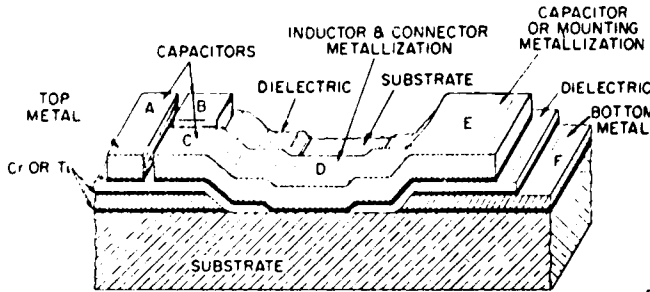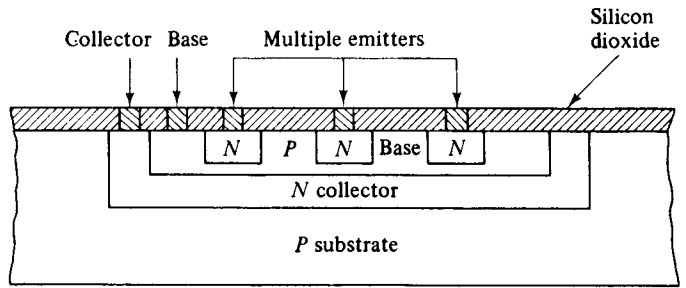
(3) Microwave integrated circuit
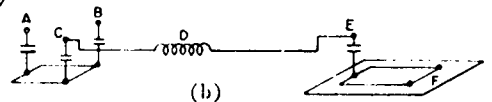
Figure 12-0-1 Discrete circuit, integrated circuit and microwave integrated circuit. (*From M. Caulton et al. [1]; reprinted by permission of IEEE, Inc.*)

by conducting wires, is now referred to as a *discrete circuit*. The word *discrete* means separately distinct.

2. Integrated circuit (IC): The integrated circuit consists of a single-crystal chip of semiconductor, typically 50 × 50 mils in cross section, containing both active and passive elements and their interconnections.

3. Monolithic microwave integrated circuit (MMIC): The word *monolithic* is derived from the Greek *monos* (single) and *lithos* (stone). Thus a monolithic integrated circuit is built on a single crystal. Such circuits are produced by the processes of epitaxial growth, masked impurity diffusion, oxidation growth, and oxide etching. Monolithic integrated circuits, like conventional integrated circuits, can be made in monolithic or hybrid form. However, MMICs are quite different from the conventional ICs. The conventional ICs contain very high packing densities, whereas the packing density of a typical MMIC is quite low. An MMIC whose elements are formed on an insulating substrate, such as glass or ceramic, is called a *film integrated circuit*. An MMIC, which consists of a combination of two or more integrated circuit types, such as monolithic or film, or one IC type together with discrete elements, is referred to as a *hybrid integrated circuit*.

Monolithic microwave integrated circuits offer the following advantages over discrete circuits:

1. Low cost (because of the large quantities processed)
2. Small size
3. Light weight
4. High reliability (all components are fabricated simultaneously, and there are no soldered joints)
5. Improved reproducibility
6. Improved performance

MMICs are suitable for space and military applications because they meet the requirements for shock, temperature conditions, and severe vibration. A major factor in the success of MMICs has been the advances in the development of microwave solid-state devices as described previously. In this chapter the basic materials and processes necessary for fabrication of MMICs are described. Three general types of circuits can be utilized for hybrid MMICs: distributed microstrip lines, lumped-element (inductors and capacitors) circuits, and thin-film circuits. These three types are discussed in Section 4.

12-1 MATERIALS

The basic materials for monolithic microwave integrated circuits, in general, are subdivided into four categories:

1. Substrate materials—alumina, beryllia, ferrite/garnet, GaAs, glass, rutile, and sapphire
2. Conductor materials—aluminum, copper, gold, and silver
3. Dielectric films—Al_2O_3, SiO, SiO_2, Si_3N_4, and Ta_2O_5
4. Resistive films—Cr, Cr-SiO, NiCr, Ta, and Ti

12-1-1 Substrate Materials

A substrate of monolithic microwave integrated circuits is a piece of substance on which electronic devices are built. The ideal substrate materials should have the following characteristics [2]:

1. High dielectric constant (9 or higher)
2. Low dissipation factor or loss tangent
3. Dielectric constant should remain constant over the frequency range of interest and over the temperature range of interest
4. High purity and constant thickness
5. High surface smoothness
6. High resistivity and dielectric strength
7. High thermal conductivity

Table 12-1-1 lists the properties of some popular substrates that have been used for MMICs [3,4]. The selection of a substrate material also depends on the expected circuit dissipation, the circuit function, and the type of circuit to be used.

TABLE 12-1-1 PROPERTIES OF SUBSTRATES

Material	$\tan \theta \times 10^4$ at 10 GHz	Relative dielectric constant (ϵ_r)	Thermal conductivity K (W/cm°C)	Applications
Alumina	2	10	0.3	Microstrip, suspended substrate
Beryllia	1	6	2.5	Compound substrated
Ferrite/garnet	2	13–16	0.03	Microstrip, coplanar, compound substrate
GaAs	16	13	0.03	High frequency, microstrip, monolithic MIC
Glass	4	5	0.01	Lumped element
Rutile	4	100	0.02	Microstrip
Sapphire	1	9.3–11.7	0.4	Microstrip, lumped element

Source: From H. Sobol [4]; reprinted by permission of IEEE, Inc.

12-1-2 Conductor Materials

The ideal conductor materials for monolithic microwave integrated circuits should have the following properties [2]:

1. High conductivity
2. Low temperature coefficient of resistance
3. Good adhesion to the substrate
4. Good etchability and solderability
5. Easily deposited or electroplated

Table 12-1-2 shows the properties of some widely used conductor materials for microcircuits [4]. These materials not only have excellent conductivity, but they can also be deposited by a number of methods and are capable of being photoetched. They are used to form both the conductor pattern and the bottom ground plane. The conductor thickness should be equal to at least four skin depths, to include 98% of the current density. It can be seen from Table 12-1-2 that good electrical conductors have poor substrate adhesion, whereas poor electrical conductors have good substrate adhesion. Aluminum has relatively good conductivity and good adhesion. It is possible to obtain good adhesion with high-conductivity materials by using a very thin film of one of the poorer conductors between the substrate and the good conductor. Some typical combinations are Cr-Au, Cr-Cu, and Ta-Au. A typical adhesion layer may have a surface resistivity ranging from 500 to 1000 Ω/square without loss. The choice of conductors is usually determined by compatibility with other materials required in the circuit and the processes required. For small losses, the conductors should be of the order of three to five skin depths in thickness. That is, thick films of

TABLE 12-1-2   PROPERTIES OF CONDUCTORS

Material	Skin depth δ at GHZ (μm)	Surface resistivity (Ω/sq $\times$ 10^{-7} $\sqrt{f}$)	Coefficient of thermal expansion (α_t/°C $\times$ 10^6)	Adherence to dielectric film or substrate	Method of deposition
Ag	1.4	2.5	21	Poor	Evaporation, screening
Cu	1.5	2.6	18	Very poor	Evaporation, plating
Au	1.7	3.0	15	Very poor	Evaporation, plating
Al	1.9	3.3	26	Very good	Evaporation
W	2.6	4.7	4.6	Good	sputtering, vapor phase, electron-beam evaporation
Mo	2.7	4.7	6.0	Good	Electron-beam evaporation, sputtering
Cr	2.7	4.7	9.0	Good	Evaporation
Ta	4.0	7.2	6.6	Very good	Electron-beam sputtering

Source: After H. Sobol [4]; reprinted by permission of IEEE, Inc.

the good conductor (about 10 μm thick) are required. Films of this thickness can be achieved by evaporation or plating or by any of the standard thick-film processes.

12-1-3 Dielectric Materials

Dielectric materials are used in monolithic microwave integrated circuits for blockers, capacitors, and some couple-line structures. The properties of dielectric materials should be

1. Reproducibility
2. Capability of withstanding high voltages
3. Ability to undergo processes without developing pin holes
4. Low RF dielectric loss

Some of the dielectrics used in microcircuits are shown in Table 12-1-3; SiO, SiO_2, and Ta_2O_5 are the most commonly used. Thin-film SiO_2 with high-dielectric Q can be obtained by growing the pyrolitic deposition of SiO_2 from silane and then densifying it by heat treatment. SiO_2 can also be deposited by sputtering. With proper processing, SiO_2 capacitors with Qs in excess of 100 have been fabricated with good success. Capacitors fabricated with SiO_2 films have capacitances in the range of 0.02 to 0.05 pF/square mil. Thin-film SiO is not very stable and can be used only in noncritical applications, such as bypass capacitors. In power microwave integrated circuits, capacitors may require breakdown voltages in excess of 200 volts. Such capacitors can be achieved with films on the order of 0.5 to 1.0 μm thick with low probability of pin holes or shorts.

TABLE 12-1-3 PROPERTIES OF DIELECTRIC MATERIALS

Material	Method of deposition	Relative dielectric constant (ϵ_r)	Dielectric strength (V/cm)	Microwave Q
SiO	Evaporation	6–8	4×10^5	30
SiO_2	Deposition	4	10^7	100–1000
Si_3N_4	Vapor phase, sputtering	7.6 6.5	10^7 10^7	
Al_2O_3	Anodization, evaporation	7–10	4×10^6	
Ta_2O_5	Anodization, sputtering	22–25	6×10^6	100

Source: After H. Sobol [4]; reprinted by permission of IEEE Inc.

12-1-4 Resistive Materials

Resistive materials are used in monolithic microwave integrated circuits for bias networks, terminations, and attenuators. The properties required for a good microwave resistor are similar to those required for low-frequency resistors and should be [5]

1. Good stability
2. Low temperature coefficient of resistance (TCR)
3. Adequate dissipation capability
4. Sheet resistivities in the range of 10 to 1000 Ω per square

Table 12-1-4 lists some of the thin-film resistive materials used in monolithic integrated circuits. Evaporated nichrome and tantalum nitride are the most widely used materials. The exact temperature coefficient of resistance achieved depends on film formation conditions. Thick-film resistors may be utilized in circuits incorporating chip components. The thickness of the thick film is in the range of 1 to 500 μm. The term *thick film* refers to the process used, not to the film thickness. Thick-film techniques involve silk-screening through a mask, such as the printing and screening of silver or gold in a glass frit, which is applied on the ceramic and fired at 850° C. Microwave thick-film metals are sometimes several micrometers thick, thicker than those of low-frequency integrated circuits.

TABLE 12-1-4   PROPERTIES OF RESISTIVE MATERIALS

Material	Method of deposition	Resistivity (Ω/square)	TCR (%/°C)	Stability
Cr	Evaporation	10–1000	−0.100–+0.10	Poor
NiCr	Evaporation	40–400	+0.002–+0.10	Good
Ta	Sputtering	5–100	−0.010–+0.01	Excellent
Cr-SiO	Evaporation or cermet	−600	−0.005––0.02	Fair
Ti	Evaporation	5–2000	−0.100–+0.10	Fair

Source: From H. Sobol [4]; reprinted by permission of IEEE, Inc.

12-2 MONOLITHIC MICROWAVE INTEGRATED-CIRCUIT GROWTH

Like lower-frequency integrated circuits, monolithic microwave integrated circuits (MMICs) can be made in monolithic or hybrid form. In a monolithic circuit, active devices are grown on or in a semiconducting substrate, and passive elements are either deposited on the substrate or grown in it. In the hybrid circuit active devices are attached to a glass, ceramic, or substrate, which contains the passive circuitry. Monolithic integrated circuits have been successful in digital and linear applications in which all required circuit components can be simultaneously fabricated. In most cases, the same device, such as bipolar or metal-oxide-semiconductor (MOS) transistors, can be used for amplifiers, diodes, resistors, and capacitors with no loss in performance. Many digital circuits used in computers require large arrays of identical devices. Thus the conventional ICs contain very high packing densities. On the other hand, very few applications of microwave integrated circuits require densely packed arrays of identical devices, and there is little opportunity to utilize active devices for passive components.

Monolithic technology is not well suited to microwave integrated circuits because the processing difficulties, low yields, and poor performance have seriously limited their applications. To date, the hybrid form of technology is used almost exclusively for microwave integrated circuits in the frequency range of 1 to 15 GHz. Hybrid MMICs are fabricated on a high-quality ceramic, glass, or ferrite substrate. The passive circuit elements are deposited on the substrate, and active devices are mounted on the substrate and connected to the passive circuit. The active devices may be utilized in chip form, in chip carriers, or in small plastic packages. The resistivity of microwave integrated circuits should be much greater than 1000 Ω–cm for good circuit performance.

12-2-1 MMIC Fabrication Techniques

Monolithic microwave integrated circuits (MMICs) can be fabricated by using different techniques such as diffusion and ion implantation, oxidation and film deposition, epitaxial growth, lithography, etching and photoresist, and deposition.

Diffusion and ion implantation.   Diffusion and ion implantation are the two processes used in controlling amounts of dopants in semiconductor device fabrications. The process of diffusion consists of diffusing impurities into a pure material in order to alter the basic electronic characteristics of the pure material. Ion implantation is used to dope the substrate crystal with high-energy ion impurities. Both processes are used to dope selectively the semiconductor substrate to produce either an n- or p-type layer. Until 1970, selective doping was performed mainly by the diffusion method at elevated temperatures. Since 1970, many doping operations have been conducted by ion implantation. In this process the dopant ions are implanted into the semiconductor by using a high-energy ion beam. The advantages of the ion-implantation method are precise control of the total amount of dopants, the improvement of reproducibility, and reduced processing temperature. Both diffusion and ion implantation can be used for fabricating discrete and integrated devices because these processes are generally complementary to one another.

Oxidation and film deposition.   To fabricate discrete and integrated devices or circuits many different types of thin films are used. There are four groups of thin films:

1. Thermal oxides
2. Dielectric layers
3. Polycrystalline silicon
4. Metal films

Epitaxial growth.   In epitaxy technology, single-crystal semiconductor layers grow on a single-crystal semiconductor substrate. The word *epitaxy* comes from the Greek *epi* (on) and *taxis* (arrangement). The epitaxial process offers an important means of controlling the doping profiles so that device and circuit performances can be optimized. There are three types of epitaxy.

1. Vapor-phase epitaxy (VPE) is the most important technique for silicon and GaAs devices.
2. Molecular-beam epitaxy (MBE) is a process involving the reaction of one or more thermal beams of atoms or molecules with a crystalline surface under ultrahigh vacuum conditions ($\sim 10^{-1}$ torr). MBE can achieve precise control in both chemical composition and doping profiles. Single-crystal multilayer structures with dimensions of the order of atomic layers can be made by the MBE method.
3. Liquid-phase epitaxy (LPE) is the growth of epitaxial layers on crystalline substrates by direct precipitation from the liquid phase. This process is particularly useful for growing GaAs and related III–V compounds. LPE is suited to grow thin epitaxial layers (≥ 0.2 μm) because it has a slow growth rate. It is also useful to grow multilayered structures in which precise doping and composition controls are required.

Lithography.   Lithography is the process of transferring patterns of geometric shapes on a mask to a thin layer of radiation-sensitive material, which is known as *resist*, for covering the surface of a semiconductor wafer. The resist patterns defined by the lithographic process are not permanent elements of the final device but only replicas of circuit features. There are four types of lithography technology:

1. Electron-beam lithography
2. Ion-beam lithography
3. Optical lithography
4. X-ray lithography

Etching and photoresist.   In the processes of making MICs, a selective removal of SiO_2 is required in order to form openings through which impurities can be diffused. The photoetching method used for this removal is shown in Fig. 12-2-1.

(a) Masking and exposure to ultraviolet radiation   (b) Photoresist
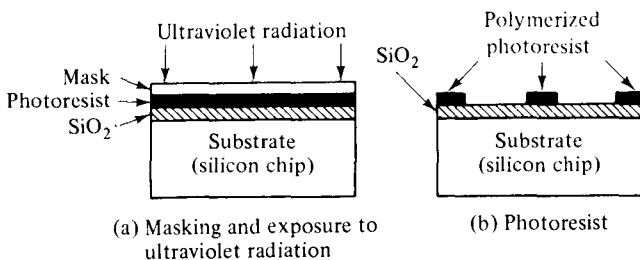
Figure 12-2-1   Photoetching process.

During the photolithographic process the substrate is coated with a uniform film of Kodak photoresist (KPR), which is a photosensitive emulsion. A mask for the desired openings is placed over the photoresist, and ultraviolet light exposes the photoresist through the mask as shown in Fig. 12-2-1(a). A polymerized photoresist is developed, and the unpolymerized portions are dissolved by using trichloroethylene after the mask is removed; see Fig. 12-2-1(b). The SiO_2, which is not covered by the photoresist, can be removed by hydrofluoric acid. The thick-film

process usually involves the printing and silk-screening of silver or gold through a metal mask in a glass frit, which is applied on the ceramic and fired at 850° C. After firing, the initial layer may be covered with gold.

Deposition. Three methods—vacuum evaporation, electron-beam evaporation, and dc sputtering—are commonly used for making MMICs.

Vacuum Evaporation. Here the impurity material to be evaporated is placed in a metallic boat through which a high current is passed. The substrate with a mask on it and the heated boat are located in a glass tube in which a high vacuum at a pressure of 10^{-6} to 10^{-8} torr is maintained. The substrate is heated slightly while the heat is evaporating the impurities, and the impurity vapor deposits itself on the substrate, forming a polycrystalline layer on it.

Electron-Beam Evaporation. In another method of evaporating the impurity a narrow beam of electrons is generated to scan the substrate in the boat in order to vaporize the impurity.

dc Sputtering. The third method of vacuum deposition is known as dc sputtering or cathode sputtering. In a vacuum, the crucible containing the impurity is used as the cathode and the substrate as the anode of a diode. A slight trace of argon gas is introduced into the vacuum. When the applied voltage between cathode and anode is high enough, a glow discharge of argon gas is formed. The positive argon ions are accelerated toward the cathode, where they dislodge atoms of the impurity. The impurity atoms have enough energy to reach the substrate and adhere to it.

12-2-2 Fabrication Example

For example, the photoresist technique can be used to remove the oxide layer in related areas. As shown in Fig. 12-2-2, the fabrication procedures include the following:

1. Deposition. An oxide layer is deposited on the semiconductor material, and then a photoresist layer is deposited to cover the oxide on top of the semiconductor chip.
2. Mask. Ultraviolet light is used to shine through a precision photographic mask to the photoresist.
3. Chemical etching. Chemical etching with hydrofluoric acid is used to remove the selected oxide region.
4. Etching. The photoresist is finally dissolved with an organic solvent in the oxide leaving the desired opening.

12-3 MOSFET FABRICATION

In recent years, the metal-oxide-semiconductor field-effect transistor (MOSFET) has superseded the bipolar junction transistor in many electronic applications. This is because the structure of the MOSFET is simple and its cost is low. The MOSFET is the

(a) Deposition

(b) Mask

(c) Chemical etching

(d) Etching
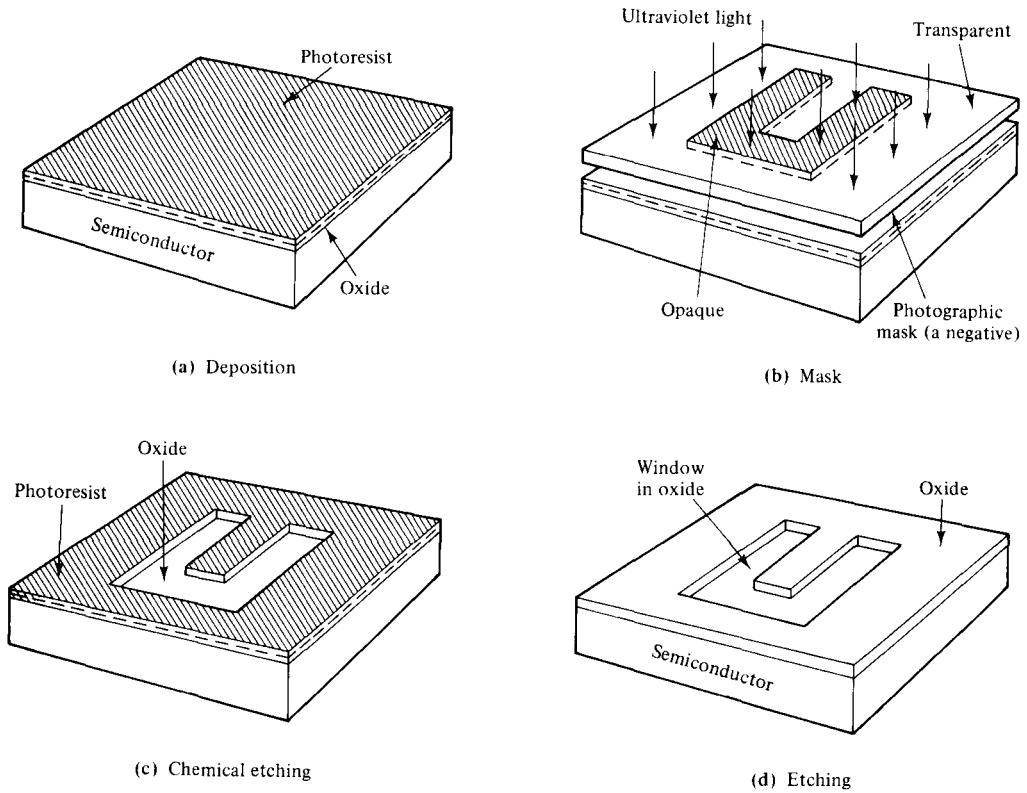
Figure 12-2-2   Illustration of photoresist technique. (*After P. R. Nanavati [6]; courtesy of Intext Educational Publishers.*)

most important device for very large-scale integrated circuits (VLSICs) such as microprocessors and semiconductor memories. Its basic fabrication processes can be described in three areas: MOSFET formation, NMOS growth, and CMOS development.

12-3-1 MOSFET Formation

MOSFETs can be fabricated by using the following steps as shown in Fig. 12-3-1.

1. **Oxidation:** Select the p-type substrate and form a SiO_2 layer on the surface.
2. **Diffusion:** Open two windows by using the photoresist technique and diffuse an n^+-layer through the windows.
3. **Etching:** Remove the center oxide region by the photoetching technique.
4. **Oxidation:** Again expose the entire surface to dry oxygen so that the SiO_2 covers the top surface.
5. **Deposition:** Deposit phosphorous glass over the surface to cover the oxide layer.
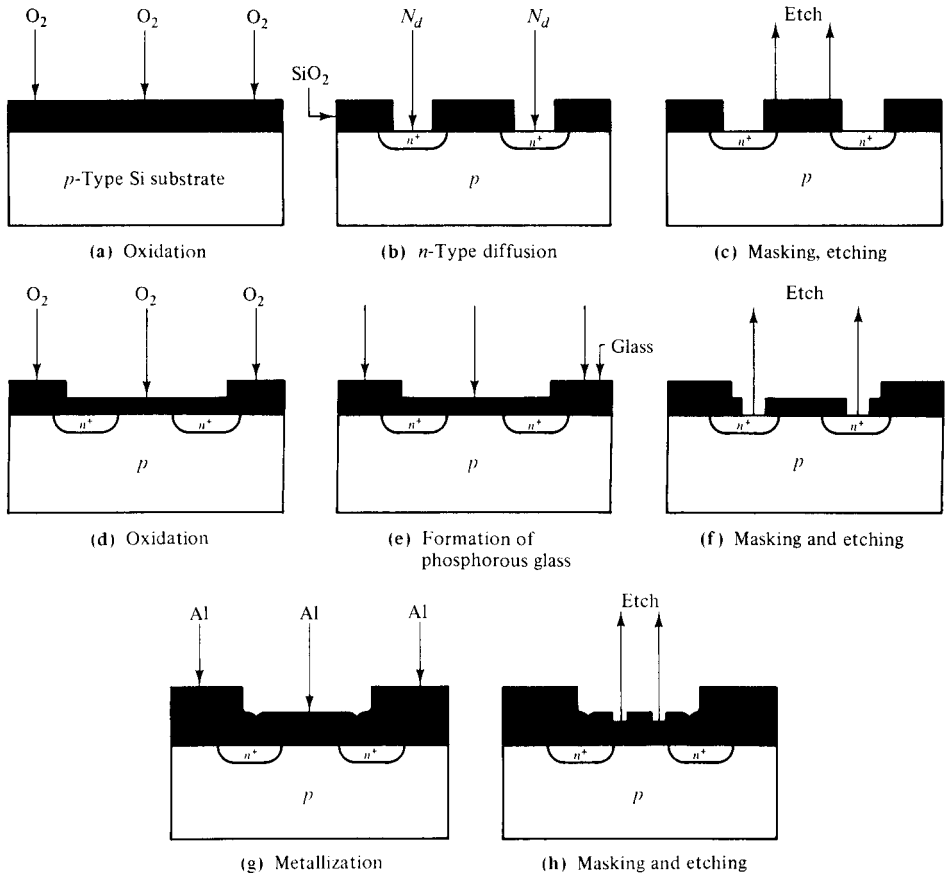
Figure 12-3-1 Fabrication steps for making MOSFETs. (*After P. R. Nanavati [6]; courtesy of Intext Educational Publishers.*)

6. Etching: Open two windows above the two n^+-type diffused regions by using the photoetching method.

7. Metallization: Now see that aluminum metallization is carried out over the entire surface of the device.

8. Etching: Finally, etch away the unwanted metal and attach the metal contacts to the diffused gate, drain, and source regions.

It can be seen that there is only one diffusion process in the fabrication of a MOSFET compared to the three required for the bipolar junction transistor. Therefore, MOSFET fabrication is more efficient and less expensive than the BJT. These attributes make MOSFET integrated circuits attractive.

12-3-2 NMOS Growth

The n-channel MOS (NMOS) logic gate was discussed in Section 6-5-1, and its fabrication processes are described as shown in Fig. 12-3-2.
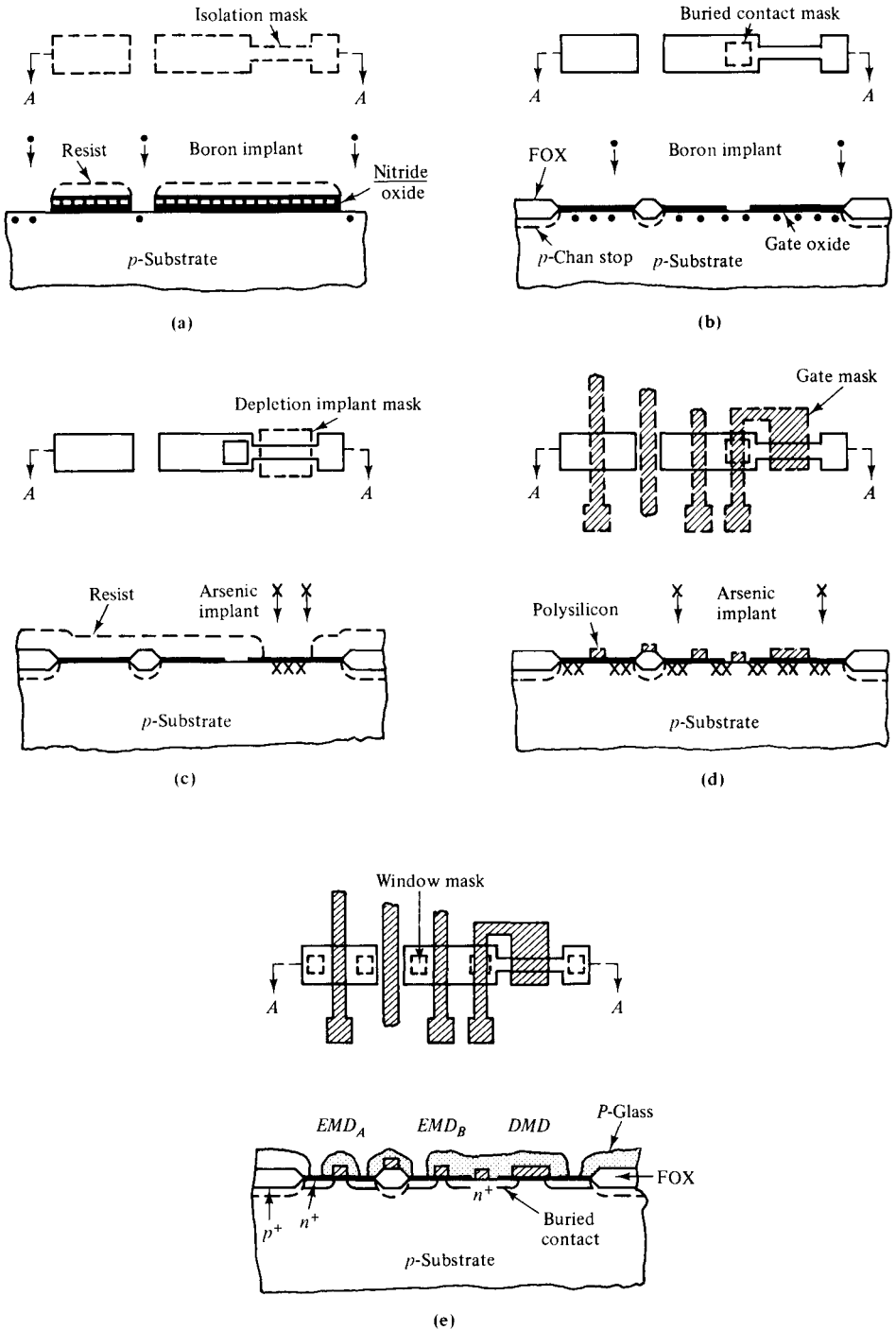
Figure 12-3-2   NMOS fabrication processes. (*After L. C. Parrillo [7]; reprinted by permission of AT&T Bell Laboratories.*)

1. **Deposition and implantation:** The starting p-type substrate is lightly doped and then an oxide (SiO_2) layer is grown on the top of the substrate. A silicon-nitride (Si_3N_4) layer is deposited on the oxide surface. An isolation mask is used to define the active areas covered by $SiO_2 - Si_3N_4$, and the isolation or field areas are etched by plasma or reactive ion etching.

2. **Implantation:** Boron ions are implanted as the channel stop to prevent inversion under the field oxide (FOX). The wafer is then put in an oxidation furnace to grow a thick layer of FOX.

3. **Implantation:** After the nitride-oxide layers are cleaned, a thin gate oxide (about 20 nm thick) is formed. Using the photoresist to mask the enhancement-mode device (EMD), an n-channel implant is made to form the depletion-mode device (DMD).

4. **Deposition:** The polysilicon is deposited and patterned as the gates. The gates are also used as the self-aligned mask for source and drain arsenic implantation.

5. **Metallization:** Metal films are evaporated and etched to produce the electrode contacts.

12-3-3 CMOS Development

The CMOS device was described in Section 6-5-2, and its fabrication processes are explained as follows (see Fig. 12-3-3).

1. **Epitaxy:** The starting material is a lightly doped n epitaxy over a heavily doped n^+ substrate.
   Deposition: A composite layer of SiO_2 (pad) and SiN_4 (nitride) is defined, and silicon is exposed over the intended n-tub region. Phosphorus is implanted as the n-tub dopant at low energy, and enters the exposed silicon, but is masked from the adjacent region by the Si_3N_4 layer.

2. **Implantation:** The wafers are then selectively oxidized over the n-tub regions. The nitride is stripped and boron is implanted for the p-tub.

3. **Oxidation:** The boron enters the silicon through the thin pad oxide but is masked from the n tub by the thick SiO_2 layer there. All oxides are then stripped and the two tubs are driven in.

4. **Deposition:** n^+ polysilicon is deposited and defined, and the source and drain regions are implanted.

5. **Implantation:** Phosphorus is selectively implanted into the n-channel source and drain regions at a higher dose so that it overcompensates the existing boron.

6. **Deposition and metallization:** A phosphorus glass layer is then deposited. After windows are dry-etched in the P-glass, aluminum metallization is defined using dry etching.

(a) Epitaxy and diposition

(b) Implantation

(c) Oxidation

(d) Deposition

(e) Implantation

(f) Deposition and metallization
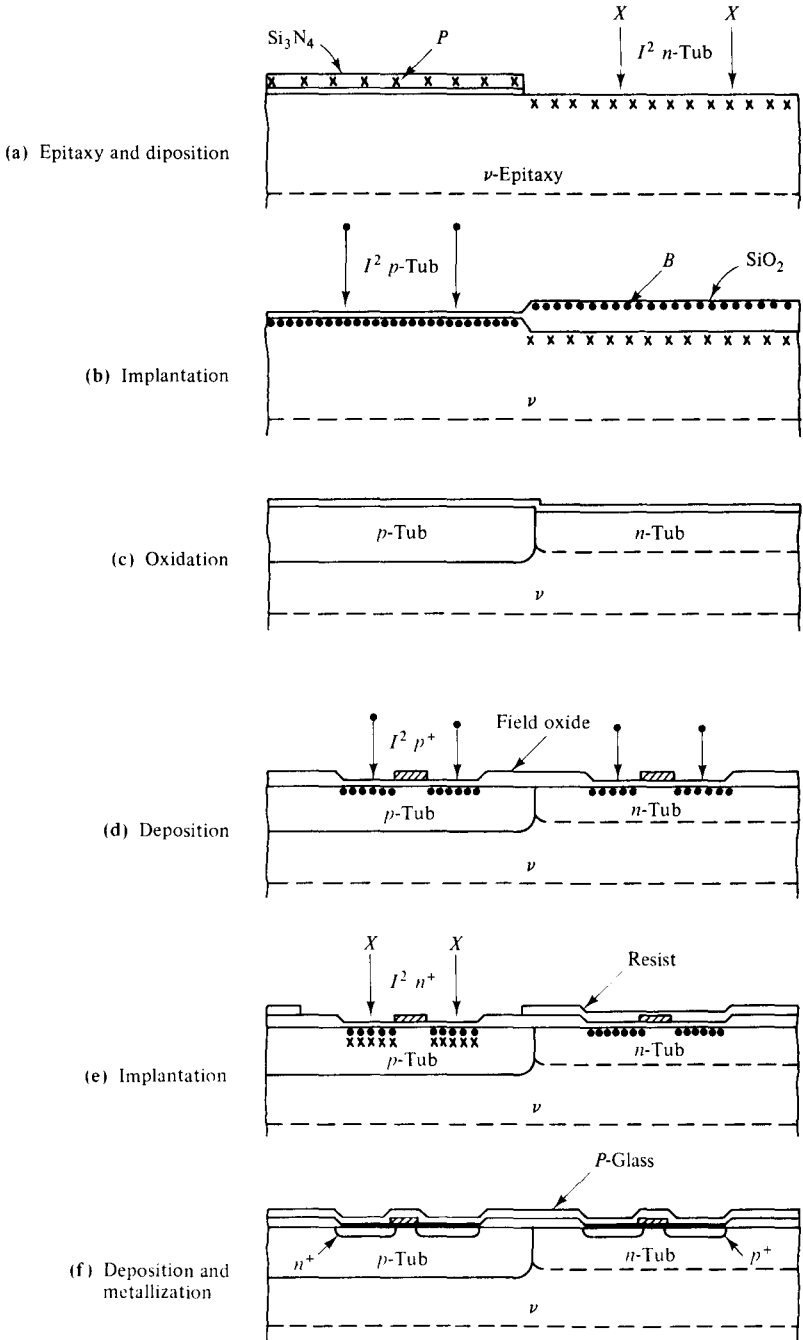
Figure 12-3-3   CMOS fabrication. (*After L. C. Parrillo [7]; reprinted by permission of the Bell Laboratories.*)

12-3-4 Memory Construction

As discussed in Section 6-5-3, there are seven types of memory devices depending on their different structures. Their basic functions are to store in, read out, and write out data. The fabrication processes of a memory are complicated. In this section, some basic fabrication techniques are described.

The *random access memory* (RAM) device is the most basic memory, and it can be subdivided into *static RAM* (SRAM) and *dynamic RAM* (DRAM). Among memory chips, the RAM device has the highest component density per chip. In a RAM any bit of information in a matrix of bits can be accessed independently. Individual rows of memory bits are accessed by a conductive word line which may be a diffusion, polysilicon, or metal line. Similarly, individual columns of bits in the matrix are accessed by a bit line. The acronym RAM is generally used to refer to randomly addressable memories into which data can be written and retrieved indefinitely. The *read-only-memory* (ROM), however, stores data permanently but cannot accept new information. Static RAMs retain their data indefinitely unless the power to the circuit is interrupted. Dynamic RAMs require that the charge (data) stored in each memory cell be "refleshed" periodically to retain the stored data. The fabrication processes of memory devices are described as follows:

1. SRAM cell with transistor load
2. SRAM cell with resistor load
3. DRAM cell with storage capacitor
4. DRAM cell with double-level polysilicon layer

SRAM cell with transistor load. Figure 12-3-4 shows the structure for a single static RAM cell. Figure 12-3-4(b) shows a layout for the SRAM cell. The width-to-channel-length ratio of the depletion-mode load transistor (1/5) is adjusted to provide enough current drive to meet the speed requirements of the cell without causing excessive steady-state (quiescent) power consumption. To minimize the cell area, buried contacts (diffusion to polysilicon contact) are required. Figure 12-3-4(a) shows a six-transistor (n-channel) cell, which uses a cross-coupled inverter pair (flip-flop, T_1 to T_4) to store 1 bit of information. The numbers next to the transistors indicate the relative width-to-channel-length ratios. A pair of access transistors (T_4 and T_5) transmits data into and out of the cell when the word and bit lines are simultaneously activated. The loads for the flip-flop are depletion-mode transistors (T_1 and T_2) with their sources and gates tied together as a NAND circuit. The data (logic 1 or 0) are retained in the cell by the positive feedback existing in the flip-flop circuit. When the gate of T_4 is at a high potential, its drain voltage is low ($\ll V_T$). This voltage, in return, is fed to the gate of T_3, and keeps T_3 off. The drain of T_3 is then tied to the high potential of T_1 (which is always on), so is the gate of T_4. This state of the cell defines a logic 1, which is retained unless new data are entered.

SRAM cell with resistor load. When the depletion-mode load transistors are replaced by high-valued resistors, the structure becomes a *resistor MOS* (RMOS) as shown in Fig. 12-3-5.
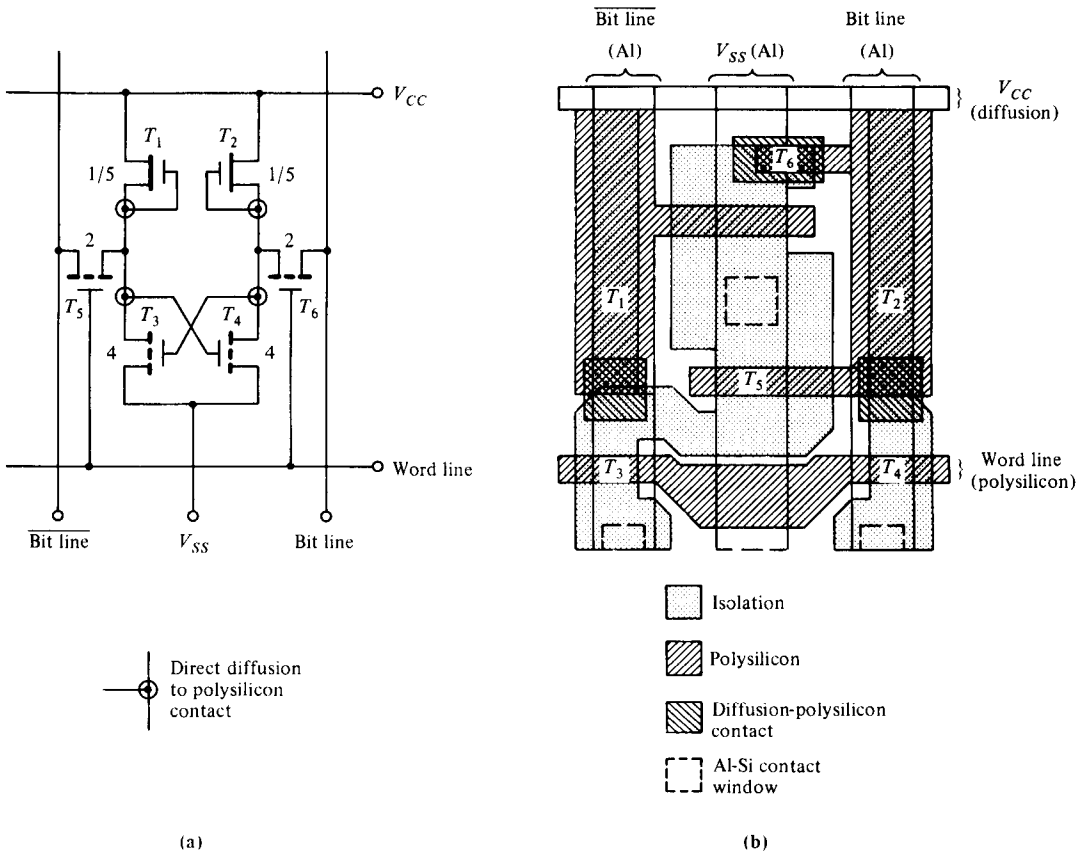
Figure 12-3-4   Structure of a single SRAM cell. (*After R. W. Hunt [8]; reprinted by permission of John Wiley & Sons.*)
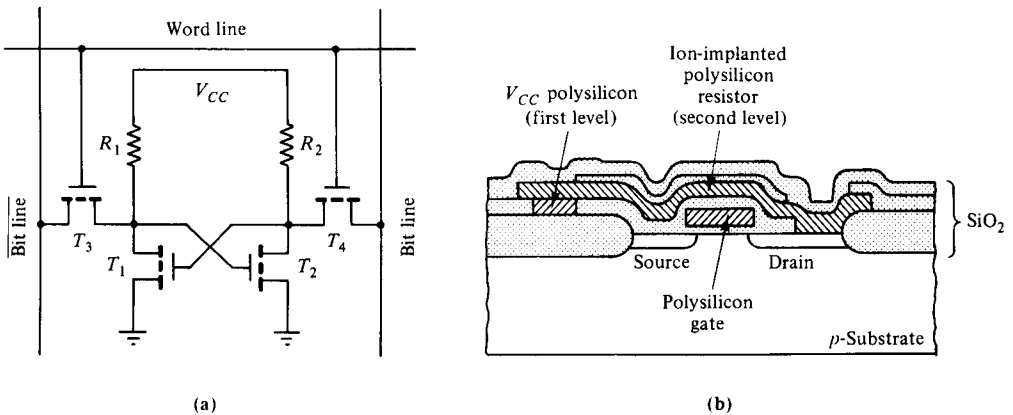
Figure 12-3-5   SRAM cell with resistor load. (*After L. C. Parrillo [7] and T. Ohzone et al. [9]; reprinted by permission of the Bell Lab.*)

High-valued resistors reduce the power consumption, and they can be made in a relatively small space by using polysilicon which has been ion-implanted to provide the proper resistance. The polysilicon resistors (R_1 and R_2) can be made in the same single layer of polysilicon (gate and interconnect) by masking the polysilicon resistor regions from the high-impurity doping used in the gate and interconnect portions of the polysilicon level. Using this type of fabrication process, SRAM-cell areas can be reduced to half the cell area required in conventional transistor load cells. Figure 12-3-5(a) shows the circuit diagram of a SRAM cell with polysilicon resistor load (R_1 and R_2), and (b) the device' cross section. The first-level polysilicon is used for gate and routing power supply V_{CC}. The second-level polysilicon is used for resistor load directly over an active transistor. The connection to the drain and V_{CC} is made directly from an implanted polysilicon resistor.

DRAM cell with storage capacitor.   Since a SRAM device consists of a large number of static RAM cells, it can dissipate a great deal of electrical power. For these reasons large memory chips use dynamic RAM cells that require only one transistor and one storage capacitor per bit of information. Figure 12-3-6 shows the diagrams of a basic SRAM cell with storage capacitor.
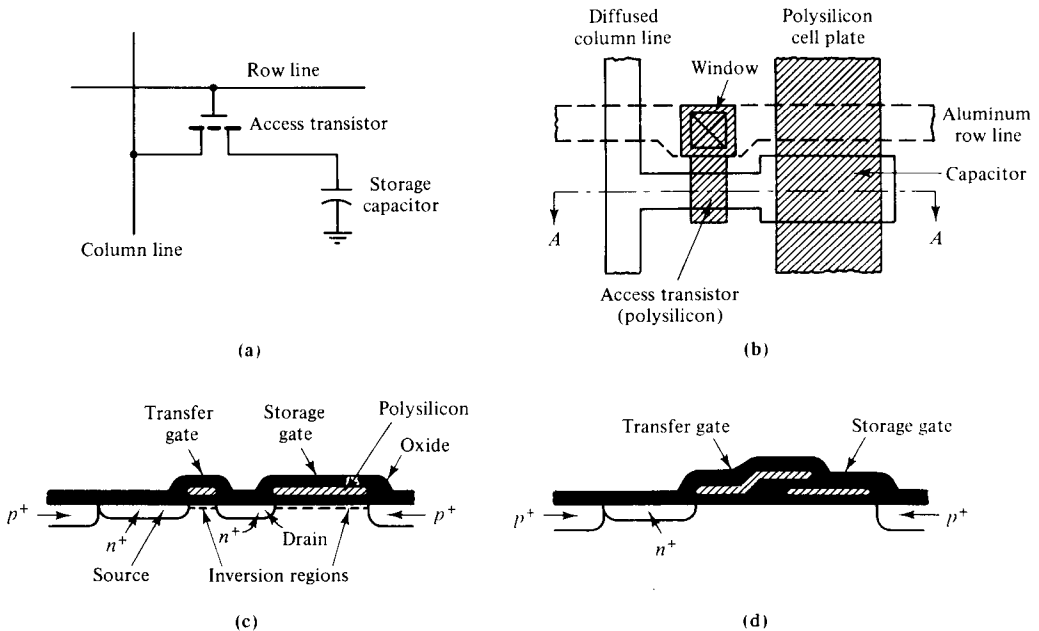
Figure 12-3-6   Diagrams of DRAM cell with storage capacitor. (*After L. C. Parrillo [7] and R. W. Hunt [8]; reprinted by permission of the Bell Laboratories.*)

Figure 12-3-6(a) shows the DRAM cell layout. A diffusion process forms the bit line (source/drain) and also the source of the access transistor. The capacitance of the diffused bit line (or the junction capacitance) and its resistance can be limiting factors for the DRAM performance. In order to minimize these parasitic effects, it is necessary to use $MoSi_2$ for word lines and Al for bit lines in fabricating advanced

memory chips such as the 256-kilobit DRAM. To increase the charge-storage capacity, the use of thinner gate insulators with higher dielectric contacts (e.g., Si_3N_4 and Ta_2O_5 with dielectric constants of 8 and 22, respectively) are being explored. When word and bit lines are simultaneously addressed (or activated to a high voltage), the access transistor is turned on and the charge is transferred into the storage capacitor if it had no initial charge (stored "zero"); little charge is transferred to the storage capacitor if it had been fully charged initially (stored "one"). The amount of charge that the bit line must supply to the storage capacitor is measured by the sensing circuitry, and this information is used to interpret whether a "zero" or "one" has been stored in the cell. The sense circuitry then stores full charge in the capacitor if the charge was there originally, or it fully depletes the capacitor if little charge existed originally. The information in the cell is thus "refleshed" after it is read.
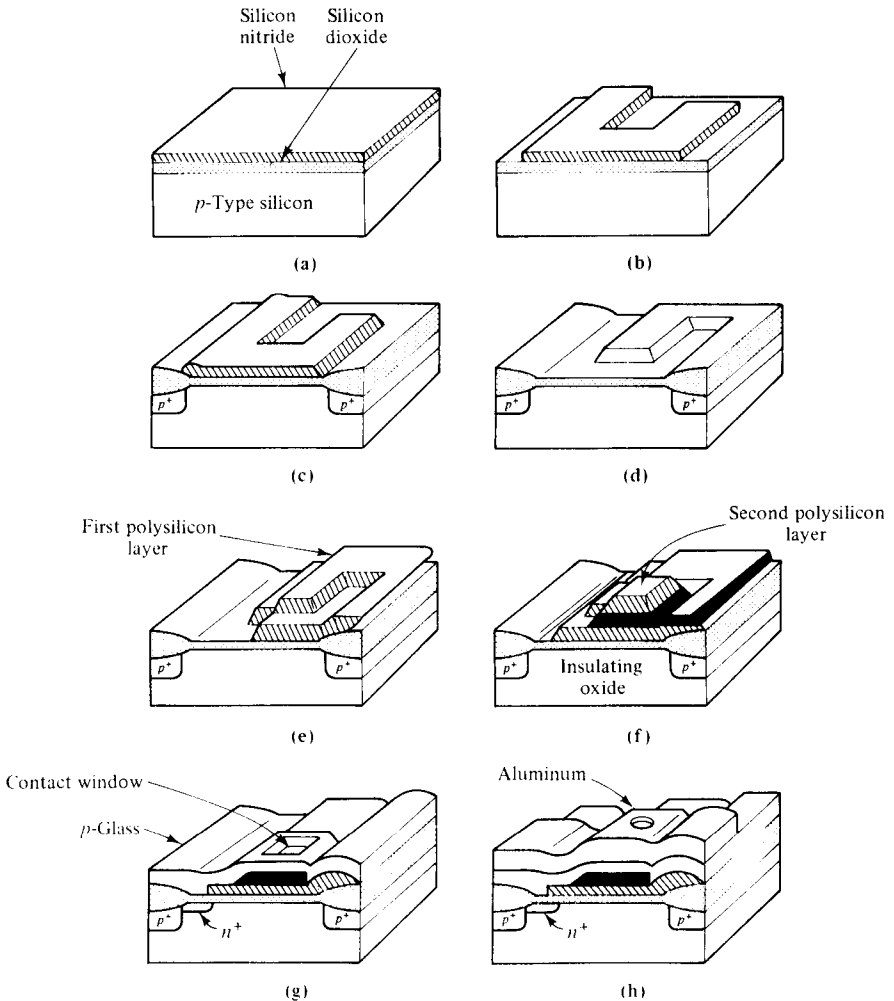
Figure 12-3-7 Diagrams of double-level polysilicon DRAM cell. (*After L. C. Parrillo [7] and W. G. Oldham [10] ; reprinted by permission of the Bell Lab.*)

DRAM cell with double-level polysilicon layers. Figure 12-3-7 shows three-dimensional diagrams of the fabrication processes for the double-level polysilicon DRAM cell. Fabrication processes for the double-level polysilicon DRAM cell follow:

1. **Oxidation:** The first step is to perform selective oxidation of silicon using silicon nitride-pad oxide layers as the oxidation mask in Fig. 12-3-7(a), (b), and (c).

2. **Etching:** The silicon nitride-pad oxide layers are then removed in a selective etchant that does not attack silicon, and the first gate oxide is grown [Fig. 12-3-7(d)].

3. **Deposition:** The first-level polysilicon layer is deposited and patterned as shown in Fig. 12-3-7(e).

4. **Oxidation and deposition:** The second gate oxide is grown and the second-level polysilicon is deposited and patterned as shown in Fig. 12-3-7(f).

5. **Implantation:** The exposed gate oxide region is implanted with an n-type dopant. A thick layer of silicon dioxide (P-glass) is deposited next, and the contact windows are opened in the oxide to reach the second-level polysilicon as shown in Fig 12-3-7(g).

6. **Deposition:** Finally, a layer of aluminum is deposited and patterned as shown in Fig. 12-3-7(h). A protective coating of silicon nitride can be deposited on the wafer to seal it from contaminations.

12-4 THIN-FILM FORMATION

The choice of lumped or distributed elements for amplifier matching networks depends on the operating frequency. When the frequency is up to X band, its wavelength is very short, and a smaller lumped element exhibits a negligible phase shift. Because of the advanced thin-film technology, the size of lumped elements can be greatly reduced and their operating frequencies can reach up to 20 GHz. Beyond that distributed elements are preferred. In monolithic microwave integrated circuits (MMICs), lumped resistors are very useful in thin-film resistive terminations for couplers, lumped capacitors are absolutely essential for bias bypass applications, and planar inductors are extremely useful for matching purposes, especially at lower microwave frequencies where stub inductors are physically too large [11].

12-4-1 Planar Resistor Film

A planar resistor consists of a thin resistive film deposited on an insulating substrate. Thin-film resistor materials are aluminum, copper, gold, nichrome, titanium, tantalium, and so forth, and their resistivity ranges from 30 to 1000 ohms per square. The planar resistors are essential for terminations for hybrid couplers, power combiners or dividers, and bias-voltage circuits. Some design considerations should be aimed to include the following:

1. the sheet resistivity available
2. the thermal stability or temperature coefficient of the resistive material
3. the thermal resistance of the load
4. the frequency bandwidth

Planar resistors can be grouped into semiconductor films, deposited metal films, and cermets.

Planar resistors based on semiconductors can be fabricated by forming an isolated band of conducting epitaxial film on the substrate by mesa etching or by isolation implant of the surrounding conducting film. Another way is by implanting a high-resistivity region within the semi-insulating substrate. Metal-film resistors are formed by evaporating a metal layer over the substrate and forming the desired pattern by photolithography. Cermet resistors are formed from films consisting of a mixture of metal and a dielectric. Figure 12-4-1 shows several examples of planar resistor design.

Resistor

Metal $\rho_s > 20$

Semi-insulating substrate

(a) Implanted resistor

Metal $\rho_s > 3$

Semi-insulating substrate

(b) Mesa resistor

Metal $3 < \rho_s < 100$

Semi-insulating substrate

(c) Deposited resistor
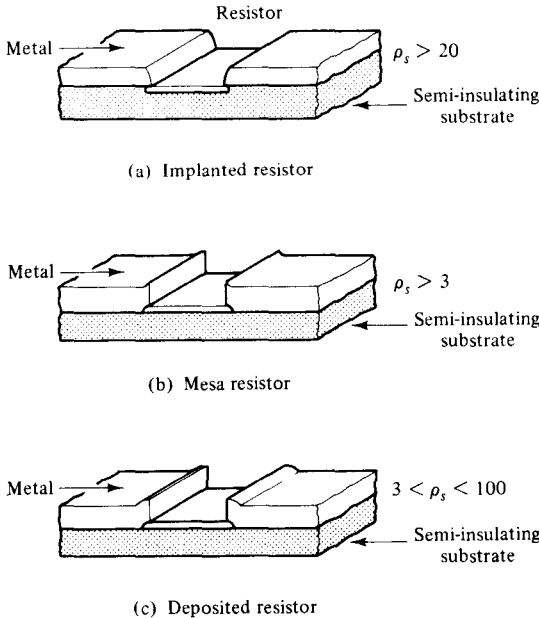
Figure 12-4-1   Configurations of planar resistors.

The resistance of a planar resistor, as shown in Fig. 12-4-2, can be expressed as

$$R = \frac{\ell \rho_s}{wt} \qquad \text{ohms} \qquad (12\text{-}4\text{-}1)$$

where ℓ = length of the resistive film
w = width of the film
ρ_s = sheet resistivity of the film in ohm-meter
t = film thickness in meters
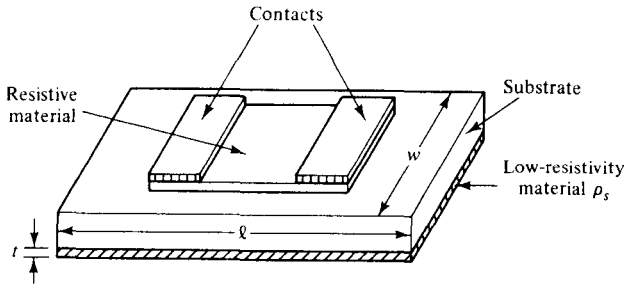
Figure 12-4-2   Thin film resistor.

When units of length ℓ and width w are chosen to have equal magnitude, the result is in a square. Therefore, the resistance R in ohms per square is independent of the dimension of the square.

Example 12-4-1:   Resistance of a Planar Resistor

A planar resistor has the following parameters:

Resistive film thickness:	$t = 0.1~\mu$m
Resistive film length:	$\ell = 10$ mm
Resistive film width:	$w = 10$ mm
Sheet resistivity of gold film:	$\rho_s = 2.44 \times 10^{-8}~\Omega$–m

Calculate the planar resistance.

Solution   The planar resistance is

$$R = \frac{10 \times 2.44 \times 10^{-8}}{10 \times 1 \times 10^{-7}} = 0.244~\Omega/\text{square}$$

12-4-2 Planar Inductor Film

Planar inductors for monolithic circuits can be realized in a number of configurations some of which are shown in Fig. 12-4-3.

Typical inductance values for monolithic circuits range from 0.5 to 10 nH. The equations of inductance are different for different configurations.

(a) Meander line

(b) S line

(c) Square spiral

(d) Circular spiral
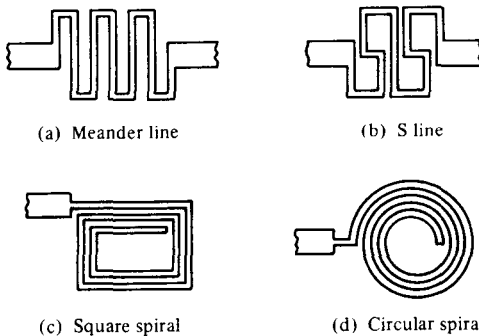
Figure 12-4-3   Configurations of planar inductors.

Ribbon inductor.   The inductance of a ribbon inductor can be expressed as [12]:

$$L = 5.08 \times 10^{-3} \ell \left[\ell n \left(\frac{\ell}{w + t} \right) + 1.19 + 0.022 \left(\frac{w + t}{\ell} \right) \right] \quad nH/mil \qquad (12\text{-}4\text{-}2)$$

where ℓ = ribbon length in mils
    t = ribbon thickness in mils
    w = ribbon width in mils

Round-wire inductor.   The inductance of a round-wire inductor is given by

$$L = 5.08 \times 10^{-3} \ell [\ell n(\ell/d) + 0.386] \quad nH/mil \qquad (12\text{-}4\text{-}3)$$

where d = wire diameter in mils
    ℓ = wire length in mils

Circular spiral inductor.   The inductance of a circular spiral inductor is expressed as

$$L = 0.03125 \, n^2 d_o \quad nH/mil \qquad (12\text{-}4\text{-}4)$$

where $d_o = 5d_i = 2.5n(w + s)$ in mils
    n = number of turns
    s = separation in mils
    w = film width in mils

Figure 12-4-4 shows the schematic diagram of a circular spiral inductor.
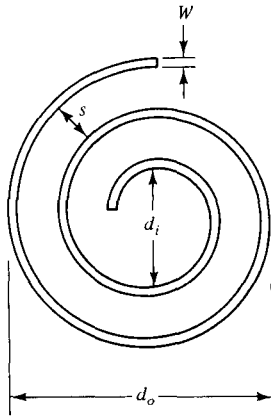
Figure 12-4-4   Circular spiral inductor.

Circular loop inductor.   The inductance of a single-turn flat circular loop inductor is given by

$$L = 5.08 \times 10^{-3} \ell \left[\ell n \left(\frac{t}{w + t} \right) - 1.76 \right] \quad nH/mil \qquad (12\text{-}4\text{-}5)$$

Square spiral inductor.   The inductance of a square spiral inductor can be written as

$$L = 8.5 \, A^{\frac{1}{2}} \, n^{5/3} \qquad \text{nH} \qquad (12\text{-}4\text{-}6)$$

where A = surface area in cm^2
   n = number of turns

Example 12-4-2:   Calculation of a Planar Circular Spiral Inductor

A circular spiral inductor has the following parameters:

Number of turns:	$n = 5$
Separation:	$s = 100$ mils
Film width:	$w = 50$ mils

Compute the inductance.

Solution   The inductance is

$$L = 0.03125 \, (5)^2 \times 2.5 \, (5)(50 + 100)$$
$$= 1464.84 \quad \text{nH/mil}$$

12-4-3 Planar Capacitor Film

Two types of planar capacitors commonly used for MMICs are the metal-oxide-metal capacitor and the interdigitated capacitor, as shown in Fig. 12-4-5.

(a) Metal–oxide–metal capacitor   (b) Interdigitated capacitor
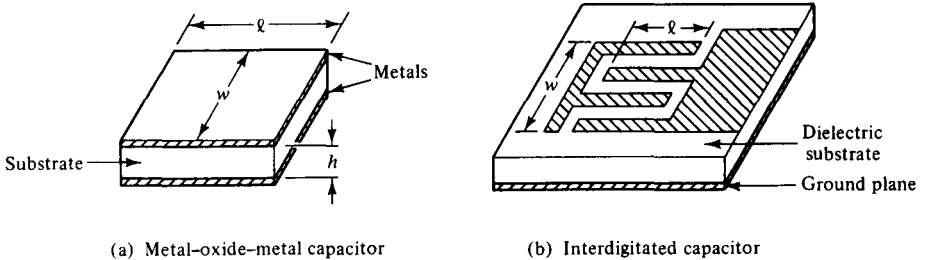
Figure 12-4-5   Schematic diagrams of planar capacitors.

Metal-oxide-metal capacitor.   The metal-oxide-metal capacitor has three layers; the middle dielectric layer is sandwiched by the top and bottom electrode layers as shown in Fig. 12-4-5(a). The capacitance can be expressed as

$$C = \epsilon_o \epsilon_r \frac{\ell w}{h} \qquad \text{farads} \qquad (12\text{-}4\text{-}7)$$

where $\epsilon_o = 8.854 \times 10^{-12}$ F/m
   ϵ_r = relative dielectric constant of the dielectric material
   ℓ = metal length
   w = metal width
   h = height of the dielectric material

Interdigitated capacitor.   The interdigitated capacitor consists of a single-layer structure, and it can be fabricated easily on substrates as microstrip lines with values between 0.1 and 15 pF as shown in Fig. 12-4-5(b). The capacitance can be approximated as [13]

$$C = \frac{\epsilon_r + 1}{w} \ell [(N - 3)A_1 + A_2] \qquad \text{pF/cm} \qquad (12\text{-}4\text{-}8)$$

where N = number of fingers
   A_1 = 0.089 pF/cm is the contribution of the interior finger for $h > w$
   A_2 = 0.10 pF/cm is the contribution of the two external fingers for $h > w$
   ℓ = finger length in cm
   w = finger-base width in cm

Example 12-4-3:   Computations of a Planar Capacitor

An interdigitated capacitor fabricated on a GaAs substrate has the following parameters:

Number of fingers:	$N = 8$
Relative dielectric constant of GaAs:	$\epsilon_r = 13.10$
Substrate height:	$h = 0.254$ cm
Finger length:	$\ell = 0.00254$ cm
Finger-base width:	$w = 0.051$ cm

Compute the capacitance.

Solution   The capacitance is

$$C = \frac{13.10 + 1}{0.051} \times 0.00254 \, x [(8 - 3) \times 0.089 + 0.10]$$

$$= 0.380 \text{ pF/cm}$$

12-5 HYBRID INTEGRATED-CIRCUIT FABRICATION

In the hybrid integrated circuit (HIC), semiconductor devices and passive circuit elements are formed on a dielectric substrate. The passive circuits are either distributed elements or lumped elements or a combination of both. The distributed and lumped elements are formed by using a thin- or thick-film process. The distributed circuit elements are generally single-layer metallization. The lumped elements are either fabricated by using multilevel deposition and plating techniques or are attached to the substrate in chip form. Hybrid integrated circuits have been used almost exclusively in the frequency range of 1 to 20 GHz for satellite communications, phased-array radar systems, electronic warfare measures, and other commercial or military electronic systems because they can offer higher reliability, greater reproducibility, better performance, smaller size, and lower cost than conventional electronic circuits or systems.

Hybrid integrated circuits can be classified into two types: hybrid IC and miniature hybrid IC.

1. **Hybrid IC:** This type uses the distributed circuit elements that are fabricated on a substrate using a single-layer metallization technique. Other circuit elements such as resistors, inductors, capacitors, and semiconductor devices are added to the substrate.

2. **Miniature hybrid IC:** The miniature hybrid IC uses multilevel elements. All passive elements such as resistors, inductors, and capacitors are deposited on the substrate; the semiconductor devices are attached to the substrate.

Hybrid ICs use a single-layer metallization technique to form the circuit components on a substrate. There are two techniques—plate-through and etchback—for fabricating the hybrid integrated circuits.

Plate-through technique.   The plate-through technique begins with a substrate coated with a thin layer of evaporated metal. The second step is to form a thick photoresist as shown in Fig. 12-5-1(a).

Figure 12-5-1   Fabrication of hybrid ICs. (*After M. Kumar and I. J. Bahl [13];*
reprinted by permission of John Wiley & Sons.)

The thickness of the photoresist is similar to the thickness of the final metal film required. After defining a pattern in the photoresist, the second layer is plated up to the desired thickness with precise definition, and only in the areas where metal is required. Finally, the photoresist layer is removed, and the thin seed metal is etched with very little undercut from the undesired areas.

Etchback technique.   The etchback technique utilizes a thick metal layer obtained completely by evaporation or by a combination of a thin evaporated layer and a thick plated layer as shown in Fig. 12-5-1(b). A thin photoresist layer is used as a mask to define the circuit pattern. Finally, the unwanted areas of the metal are etched away. This technique results in thicker conductors than the plate-through technique.

REFERENCES

[1] CAULTON, M., et al., Status of lumped elements in microwave integrated circuits: present and future. *IEEE Trans. on Microwave Theory and Techniques*, **MTT-19,** No. 7, July 1971.

[2] KEISTER, F. Z., An evaluation of materials and processes for integrated microwave circuits. *IEEE Trans. on Microwave Theory and Techniques*, **MTT-16,** No. 7, 469–475, July 1968.

[3] CAULTON, M., and H. SOBOL, Microwave integrated circuit technology: A survey. *IEEE J. Solid-State Circuits*, **SC-5,** No. 6, 292–303, December 1970.

[4] SOBOL, H., Applications of integrated circuit technology to microwave frequencies. *Proc. IEEE*, **59,** No. 8, 1200–1211, August 1971.

[5] SOBOL, H., Technology and design of hybrid microwave integrated circuits. *Solid State Technology*, **13,** No. 2, 49–59, February 1970.

[6] NANAVATI, R. P., *Semiconductor Devices*, p. 415, fig. 11-10. Intext Education Publishers, Scranton, PA, 1975.

[7] PARRILLO, L. C., VLSI process integration. Chapter 11, p. 464, fig. 14; p. 484, fig. 29 in S. M. Sze (Ed.) *VLSI Technology*. John Wiley & Sons, New York, 1983.

[8] HUNT, R. W., Memory design and technology. Chapter 11, p. 474, fig. 20 in S. M. Sze (Ed.) *VLSI Technology*. John Wiley & Sons, New York, 1983.

[9] OHZONE, T., et al., A 2k × 8 static RAM. P. 475, fig. 21 in S. M. Sze (Ed.) *VLSI Technology*. John Wiley & Sons, New York, 1983.

[10] OLDHAM, W. G., The fabrication of microelectronic circuits. P. 488, fig. 19 in S. M. Sze (Ed.) *VLSI Technology*. John Wiley & Sons, New York, 1983.

[11] PUCEL, ROBERT A., Design considerations for monolithic microwave circuits. *IEEE Trans. on Microwave Theory and Techniques*, **MTT-29,** No. 6, 513–534, July 1981.

[12] KUMAR, M., and I. J. BAHL, Microwave integrated circuits. Chapter 15, p. 794, fig. 15-4 in I. Bahl (Ed.) *Microwave Solid-State Circuit Design*. John Wiley & Sons, New York, 1988.

[13] YOUNG, L., *Advances in Microwaves*, pp. 148–158. Academic Press, New York, 1974.

SUGGESTED READINGS

BAHL, I., ed. *Microwave Solid-State Circuit Design*. John Wiley & Sons, New York, 1988.

HESS, K., *Advanced Theory of Semiconductor Devices*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.

LIAO, S. Y., *Semiconductor Electronic Devices*. Prentice-Hall, Inc., Englewood Cliifs, N.J., 1990.

STREETMAN, B. G., *Solid-State Electronic Devices*, 3rd ed. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1989.

SZE, S. M., ed., *VLSI Technology*. John Wiley & Sons, New York, 1983.

PROBLEMS

Materials

12-1. List the basic materials for MMICs.

12-2. List the basic characteristics required for an ideal substrate material.

12-3. List the basic properties provided by ideal conductor, dielectric, and resistive materials used in MMICs.

MMIC Growth

12-4. Describe the MMIC techniques.

12-5. Explain the photoresist process.

MOSFET Fabrication

12-6. Describe the basic fabrication processes for MOSFETs.

12-7. Explain the NMOS growth.

12-8. Analyze the CMOS formation.

12-9. Describe the memory construction.

Thin-Film Formation

12-10. Describe the resistor-film growth.

12-11. Explain the inductor-film formation.

12-12. Discuss the capacitor-film development.

Hybrid MMICs

12-13. Discuss the discrete, integrated, and monolithic microwave integrated circuits.

12-14. Analyze the hybrid MMICs.

12-15. Describe the hybrid IC techniques.

Appendix A

TABLE A-1 CONDUCTIVITY σ IN MHOS PER METER

Conductor	σ	Insulator	σ
Silver	6.17×10^7	Quartz	10^{-17}
Copper	5.80×10^7	Polystyrene	10^{-16}
Gold	4.10×10^7	Rubber (hard)	10^{-15}
Aluminum	3.82×10^7	Mica	10^{-14}
Tungsten	1.82×10^7	Porcelain	10^{-13}
Zinc	1.67×10^7	Diamond	10^{-13}
Brass	1.50×10^7	Glass	10^{-12}
Nickel	1.45×10^7	Bakelite	10^{-9}
Iron	1.03×10^7	Marble	10^{-8}
Bronze	1.00×10^7	Soil (sandy)	10^{-5}
Solder	0.70×10^7	Sands (dry)	2×10^{-4}
Steel (stainless)	0.11×10^7	Clay	10^{-4}
Nichrome	0.10×10^7	Ground (dry)	$10^{-4} - 10^{-5}$
Graphite	7.00×10^4	Ground (wet)	$10^{-2} - 10^{-3}$
Silicon	1.20×10^3	Water (distilled)	2×10^{-4}
Water (sea)	$3-5$	Water (fresh)	10^{-3}
		Ferrite (typical)	10^{-2}

TABLE A-2   DIELECTRIC CONSTANT—RELATIVE PERMITTIVITY ϵ_r

Material	ϵ_r	Material	ϵ_r
Air	1	Sands (dry)	4
Alcohol (ethyl)	25	Silica (fused)	3.8
Bakelite	4.8	Snow	3.3
Glass	4–7	Sodium chloride	5.9
Ice	4.2	Soil (dry)	2.8
Mica (ruby)	5.4	Styrofoam	1.03
Nylon	4	Teflon	2.1
Paper	2–4	Water (distilled)	80
Plexiglass	2.6–3.5	Water (sea)	20
Polyethylene	2.25	Water (dehydrated)	1
Polystyrene	2.55	Wood (dry)	1.5–4
Porcelain (dry process)	6	Ground (wet)	5–30
Quartz (fused)	3.80	Ground (dry)	2–5
Rubber	2.5–4	Water (fresh)	80

TABLE A-3   RELATIVE PERMEABILITY μ_r

Diamagnetic material	μ_r	Ferromagnetic material	μ_r
Bismuth	0.99999860	Nickel	50
Paraffin	0.99999942	Cast iron	60
Wood	0.99999950	Cobalt	60
Silver	0.99999981	Machine steel	300
Paramagnetic material	μ_r	Ferrite (typical)	1,000
		Transformer iron	3,000
Aluminum	1.00000065	Silicon iron	4,000
Beryllinum	1.00000079	Iron (pure)	4,000
Nickel chloride	1.00004	Mumetal	20,000
Manganese sulphate	1.0001	Supermalloy	100,000

TABLE A-4   PROPERTIES OF FREE SPACE

Velocity of light in vacuum c	2.997925×10^8 meter per second
Permittivity ϵ_0	8.854×10^{-12} farad per meter
Permeability μ_0	$4\pi \times 10^{-7}$ henry per meter
Intrinsic impedance η_0	377 or 120π ohms

TABLE A-5   PHYSICAL CONSTANTS

Boltzmann constant	k	1.381×10^{-23} J/°K
Electronvolt	eV	1.602×10^{-19} J
Electron charge	q	1.602×10^{-19} C
Electron mass	m	9.109×10^{-31} kg
Ratio of charge to mass of an electron	e/m	1.759×10^{-11} C/kg
Planck's constant	h	6.626×10^{-34} J $-$ s

TABLE A-6 ENERGY GAP STRUCTURES OF SEMICONDUCTORS

Semi-conductor	Energy gap (eV)	Lattice constants (Å)	Energy gap struc-ture	Expansion coefficient at 300°K ($\times 10^{-6}$ °C^{-1})	Hetero-junction preferred doping	Typical dopants	Electron affinity (eV)
Ge$_{0.9}$Si$_{0.1}$	0.77	(5.63)	Indirect	—	n	P, As, Sb	(4.1)
Ge	0.66	5.658	Indirect	5.7	p	Al, Ga, In	4.13
GaAs	1.43	5.654	Direct	5.8	n	Se, Te	4.07
Ge	0.66	5.658	Indirect	5.7	p	Al, Ga, In	4.13
ZnSe	2.67	5.667	Direct	7.0	n	Al, Ga, In	4.09
Ge	0.66	5.658	Indirect	5.7	p	Al, Ga, In	4.13
ZnSe	2.67	5.667	Direct	7.0	n	Al, Ga, In	4.09
GaAs	1.43	5.654	Direct	5.8	p	Zn, Cd	4.07
AlAs	2.15	5.661	Indirect	5.2	p	Zn	3.5
GaAs	1.43	5.654	Direct	5.8	n	Se, Te	4.07
GaP	2.25	5.451	Indirect	5.3	n	Se, Te	4.3
Si	1.11	5.431	Indirect	2.33	p	Al, Ga, In	4.01
AlSb	1.6	6.136	Indirect	3.7	n/p	Se, Te/Zn, Cd	3.65
GaSb	0.68	6.095	Direct	6.9	p/n	Zn, Cd/Se, Te	4.06
GaSb	0.68	6.095	Direct	6.9	n	Se, Te	4.06
InAs	0.36	6.058	Direct	4.5(5.3)	p	Zn, Cd	4.9
ZnTe	2.26	6.103	Direct	8.2	p	Cu	3.5
GaSb	0.68	6.095	Direct	6.9	n	Se, Te	4.06
ZnTe	2.26	6.103	Direct	8.2	p	Cu	3.5
InAs	0.36	6.058	Direct	4.5(5.3)	n	Se, Te	4.9
ZnTe	2.26	6.103	Direct	8.2	p	Cu	3.5
AlSb	1.6	6.136	Indirect	3.7	n	Se, Te	3.65
CdTe	1.44	6.477	Direct	5	p/n	Li, Sb, P/I	4.28
PbTe	0.29	6.52	Indirect	19.8	n/p	Cl, Br/Na, K	4.6
CdTe	1.44	6.477	Direct	5	p	Li, Sb	4.28
InSb	0.17	6.479	Direct	4.9	n	Se, Te	4.59
ZnTe	2.26	6.103	Direct	8.2	p	Cu	3.5
CeSe(hex)	1.7	4.3($\sqrt{2}$) (6.05)	Direct	2.45–4.4	n	Cl, Br, I	4.95

Source: After Milnes & Freucht, 116. Reprinted with permission from *Heterojunctions and Metal-Semiconductor Junctions,* Academic Press, N.Y., 1972.

TABLE A-7 PROPERTIES OF SEMICONDUCTOR MATERIALS

	E_g (eV)	μ_n (cm²/V-sec)	μ_p (cm²/V-sec)	ρ (Ω-cm)	Transition	Doping	Lattice	a (Å)	ϵ_r	Density (g/cm³)	Melting point (°C)
Si	1.11	1350	480	$2.5 \times 10^{5*}$	i	n, p	D	5.43	11.8	2.33	1415
Ge	0.67	3900	1900	43	i	n, p	D	5.66	16.	5.32	936
SiC(α)	2.86	500		10^{10}	i	n, p	W	3.08	10.2	3.21	2830
AlP	2.45	80		10^{-5}	i	n, p	Z	5.46		2.40	2000
AlAs	2.16	180		0.1	i	n, p	Z	5.66	10.9	3.60	1740
AlSb	1.6	200	300	5	i	n, p	Z	6.14	11	4.26	1080
GaP	2.26	300	150	1	i	n, p	Z	5.45	11.1	4.13	1467
GaAs	1.43	8500	400	$4 \times 10^{8*}$	d	n p	Z	5.65	13.2	5.31	1238
GaSb	0.7	5000	1000	0.04	d	n, p	Z	6.09	15.7	5.61	712
InP	1.28	4000	100	8×10^{-3}	d	n, p	Z	5.87	12.4	4.79	1070
InAs	0.36	22,600	200	0.03	d	n, p	Z	6.06	14.6	5.67	943
InSb	0.18	10^5	1700	0.06	d	n, p	Z	6.48	17.7	5.78	525
ZnS	3.6	110		10^{10}	d	n	Z, W	5.409	8.9	4.09	1650†
ZnSe	2.7	600		10^9	d	n	Z	5.671	9.2	5.65	1100†
ZnTe	2.25		100	100	d	p	Z	6.101	10.4	5.51	1238†
CdS	2.42	250	15		d	n	W, Z	4.137	8.9	4.82	1475
CdSe	1.73	650		10^5	d	n	W	4.30	10.2	5.81	1258
CdTe	1.58	1050	100	10^{10}	d	n, p	Z	6.482	10.2	6.20	1098
PbS	0.37	575	200	5×10^{-3}	i	n, p	H	5.936	161	7.6	1119
PbSe	0.27	1000	1000	10^{-3}	i	n, p	H	6.147	280	8.73	1081
PbTe	0.29	1600	700	10^{-2}	i	n, p	H	6.452	360	8.16	925

TABLE A-8 FIRST-ORDER BESSEL FUNCTION VALUES

x	$J_1(x)$	x	$J_1(x)$	x	$J_1(x)$	x	$J_1(x)$	x	$J_1(x)$
0.00	0.000	0.92	0.413	1.86	0.582	2.86	0.389	3.84	−0.003
.02	+.010	.94	.420	1.88	.5815	2.88	.3825	3.86	.011
.04	.020	.96	.427	1.90	.581	2.90	.375	3.88	.019
.06	.030	.98	.4335	1.92	.5805	2.92	.368	3.90	.027
.08	.040	1.00	.440	1.94	.580	2.94	.361	3.92	.035
.10	.050	1.02	.4465	1.96	.579	2.96	.354	3.94	.043
.12	.060	1.04	.453	1.98	.578	2.98	.3465	3.96	.051
.14	.070	1.06	.459	2.00	.577	3.00	.339	3.98	.058
.16	.080	1.08	.465	2.02	.575	3.02	.3315	4.00	.066
.18	.090	1.10	.471	2.04	.574	3.04	.324	4.10	.103
.20	.0995	1.12	.477	2.06	.572	3.06	.316	4.20	.139
.22	.109	1.14	.482	2.08	.570	3.08	.309	4.30	.172
.24	.119	1.16	.488	2.10	.568	3.10	.301	4.40	.203
.26	.129	1.18	.493	2.12	.566	3.12	.293	4.50	.231
.28	.139	1.20	.498	2.14	.564	3.14	.285	4.60	.2565
.30	.148	1.22	.503	2.16	.561	3.16	.277	4.70	.279
.32	.158	1.24	.508	2.18	.559	3.18	.269	4.80	.2985
.34	.1675	1.26	.513	2.20	.556	3.20	.261	4.90	.315
.36	.177	1.28	.5175	2.22	.553	3.22	.253	5.00	.3275
.38	.187	1.30	.522	2.24	.550	3.24	.245	5.05	.334
.40	.196	1.32	.526	2.26	.547	3.26	.237	5.10	.337
.42	.205	1.34	.5305	2.28	.543	3.28	.229	5.16	.341
.44	.215	1.36	.534	2.30	.540	3.30	.221	5.20	.343
.46	.224	1.38	.538	2.32	.536	3.32	.212	5.26	.345
.48	.233	1.40	.542	2.34	.532	3.34	.204	5.30	.346
.50	.242	1.42	.5455	2.36	.5285	3.36	.196	5.32	.346
.52	.251	1.44	.549	2.38	.524	3.38	.1865	5.34	.346
.54	.260	1.46	.552	2.40	.520	3.40	.179	5.36	.346
.56	.269	1.48	.555	2.42	.516	3.42	.171	5.38	.346
.58	.278	1.50	.558	2.44	.511	3.44	.1625	5.40	.345
.60	.287	1.52	.561	2.46	.507	3.46	.154	5.47	.343
.62	.295	1.54	.563	2.48	.502	3.48	.146	5.50	.341
.64	.304	1.56	.566	2.50	.497	3.50	.137	5.56	.3375
.66	.312	1.58	.568	2.52	.492	3.52	.129	5.60	.334
.68	.321	1.60	.570	2.54	.487	3.54	.121	5.66	.3285
.70	.329	1.62	.572	2.56	.482	3.56	.112	5.70	.324
.72	.337	1.64	.5735	2.58	.476	3.58	.104	5.80	.311
.74	.345	1.66	.575	2.60	.471	3.60	.0955	5.90	.295
.76	.353	1.68	.5765	2.62	.465	3.62	.087	6.00	.277
.78	.361	1.70	.578	2.64	.4595	3.64	.079	6.10	.256
.80	.369	1.72	.579	2.66	.454	3.66	.070	6.20	.233
.82	.3765	1.74	.580	2.68	.448	3.68	.062	6.30	.208
.84	.384	1.76	.5805	2.70	.442	3.70	.054	6.40	.182
.86	.3915	1.78	.581	2.72	.435	3.72	.0455	6.60	.125
.88	.399	1.80	.5815	2.74	.429	3.74	.037	6.70	.095
.90	.406	1.82	.582	2.76	.423	3.76	.029	6.80	.065
		1.84	.582	2.78	.416	3.78	.021	6.90	.035
				2.80	.410	3.80	.013	7.00	.005
				2.82	.403	3.82	.005	7.01	.000
				2.84	.396	3.83	.000		
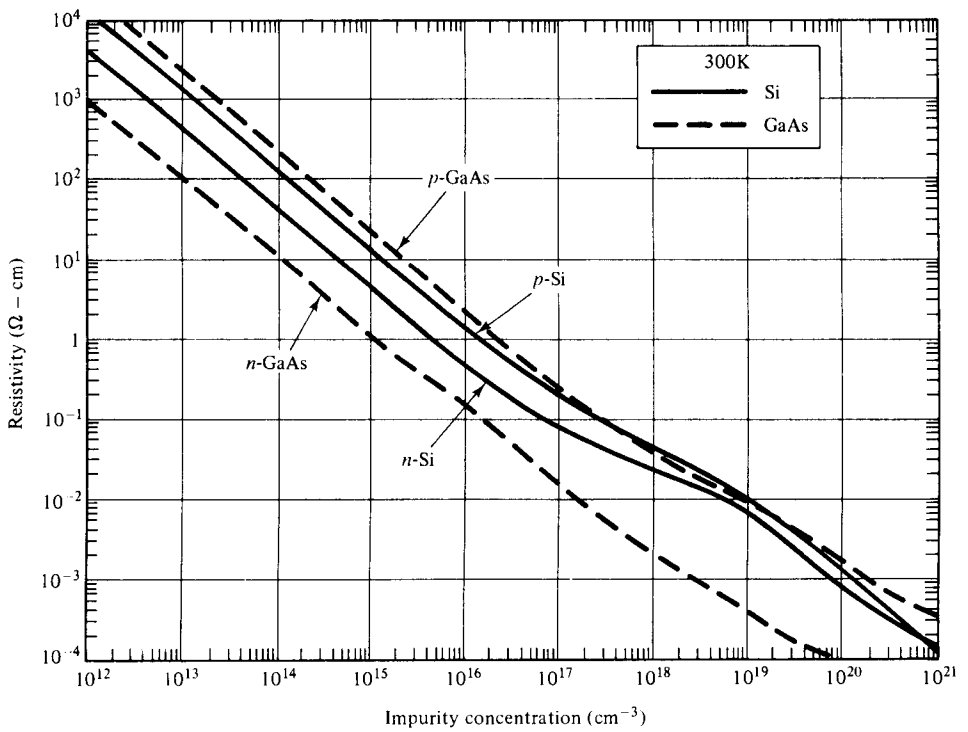
Figure A-1 Resistivity versus impurity concentration for Si and GaAs.
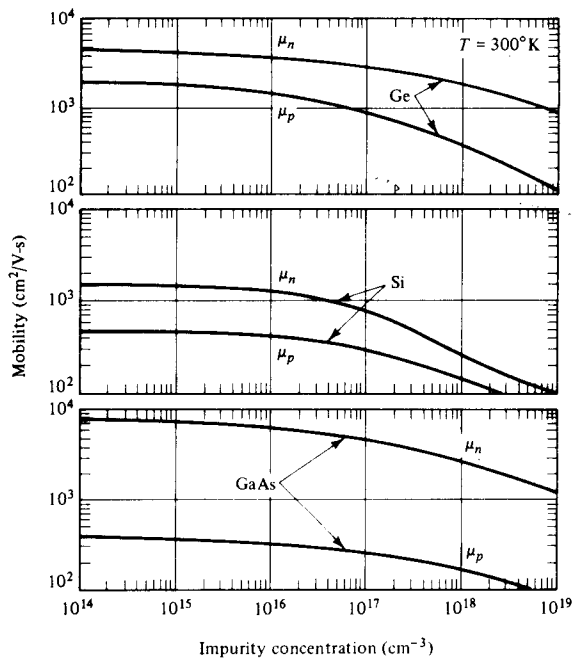
(Reprinted from Sze's *Physics of Semiconductor Devices*, 1981, by permission of John Wiley & Sons, Inc.)

Figure A-2 Mobility for Ge, Si, and GaAs.

Appendix B

UNITS OF MEASUREMENT

In field intensity measurements the units of measure and the conversion from one unit to another are the essential parts of the process. A few widely used units are described here:

1. dB—The decibel (dB) is a dimensionless number that expresses the ratio of two power levels. It is defined as

$$dB \equiv 10 \log_{10}\left(\frac{P_2}{P_1}\right) \tag{B-1}$$

The two power levels are relative to each other. If power level P_2 is higher than power level P_1, dB is positive and vice versa. Since $P = V^2/R$, when their voltages are measured across the same or equal resistors, the number of dB is given by

$$dB \equiv 20 \log_{10}\left(\frac{V_2}{V_1}\right) \tag{B-2}$$

The voltage definition of dB has no meaning at all unless the two voltages under consideration appear across equal impedances. Above 10 GHz the impedance of waveguides varies with frequency, and the dB calibration is limited to power levels only. Table B-1 shows the conversion of voltage and power ratios to dB.

2. dBW—The decibel above 1 watt (dBW) is another useful measure for expressing power level P_2 with respect to a reference power level P_1 of 1 W. Similarly, if the power level P_2 is lower than 1 W, the dBW is negative.

TABLE B-1   CONVERSION OF VOLTAGE AND POWER RATIOS TO DECIBELS

Voltage ratio	Power ratio	−dB+	Voltage ratio	Power ratio	Voltage ratio	Power ratio	−dB+	Voltage ratio	Power ratio
1.000	1.000	0.0	1.000	1.000	0.596	0.355	4.5	1.679	2.818
0.989	0.977	.1	1.012	1.023	.589	.347	4.6	1.698	2.884
.977	.955	.2	1.023	1.047	.582	.339	4.7	1.718	2.951
.966	.933	.3	1.035	1.072	.575	.331	4.8	1.738	3.020
.955	.912	.4	1.047	1.096	.569	.324	4.9	1.758	3.090
.944	.891	.5	1.059	1.122	.562	.316	5.0	1.778	3.162
.933	.871	.6	1.072	1.148	.556	.309	5.1	1.799	3.236
.923	.851	.7	1.084	1.175	.550	.302	5.2	1.820	3.311
.912	.832	.8	1.095	1.202	.543	.295	5.3	1.841	3.388
.902	.813	.9	1.109	1.230	.537	.288	5.4	1.862	3.467
.891	.947	1.0	1.122	1.259	.531	.282	5.5	1.884	3.548
.881	.776	1.1	1.135	1.288	.525	.275	5.6	1.905	3.631
.871	.759	1.2	1.148	1.318	.519	.269	5.7	1.928	3.715
.861	.741	1.3	1.161	1.349	.513	.263	5.8	1.950	3.802
.851	.724	1.4	1.175	1.380	.507	.257	5.9	1.972	3.890
.841	.708	1.5	1.189	1.413	.501	.251	6.0	1.995	3.981
.832	.692	1.6	1.202	1.445	.496	.246	6.1	2.018	4.074
.822	.676	1.7	1.216	1.479	.490	.240	6.2	2.042	4.159
.813	.661	1.8	1.230	1.514	.484	.234	6.3	2.065	4.265
.804	.646	1.9	1.245	1.549	.479	.229	6.4	2.089	4.365
.794	.631	2.0	1.259	1.585	.473	.224	6.5	2.113	4.467
.785	.617	2.1	1.274	1.622	.468	.219	6.6	2.138	4.571
.776	.603	2.2	1.288	1.660	.462	.214	6.7	2.163	4.677
.767	.589	2.3	1.303	1.698	.457.	209	6.8	2.188	4.786
.759	.575	2.4	1.318	1.738	.452	.204	6.9	2.215	4.898
.750	.562	2.5	1.334	1.778	.447	.200	7.0	2.239	5.012
.741	.550	2.6	1.349	1.820	.442	.195	7.1	2.265	5.129
.733	.537	2.7	1.365	1.862	.437	.191	7.2	2.291	5.248
.724	.525	2.8	1.380	1.905	.432	.186	7.3	2.317	5.370
.716	.513	2.9	1.390	1.950	.427	.182	7.4	2.344	5.495
.708	.501	3.0	1.413	1.995	.422	.178	7.5	2.371	5.623
.700	.490	3.1	1.429	2.042	.417	.174	7.6	2.399	5.754
.692	.479	3.2	1.445	2.089	.412	.170	7.7	2.427	5.888
.684	.468	3.3	1.462	2.138	.407	.166	7.8	2.455	6.026
.676	.457	3.4	1.479	2.188	.403	.162	7.9	2.483	6.166
.668	.447	3.5	1.496	2.239	.398	.159	8.0	2.512	6.310
.661	.437	3.6	1.514	2.291	.394	.155	8.1	2.541	6.457
.653	.427	3.7	1.531	2.344	.389	.151	8.2	2.570	6.607
.646	.417	3.8	1.549	2.399	.385	.148	8.3	2.600	6.761
.638	.407	3.9	1.567	2.455	.380	.145	8.4	2.630	6.918
.631	.398	4.0	1.585	2.512	.376	.141	8.5	2.661	7.079
.624	.389	4.1	1.603	2.570	.372	.138	8.6	2.692	7.244
.617	.380	4.2	1.622	2.630	.367	.135	8.7	2.723	7.413
.610	.372	4.3	1.641	2.692	.363	.132	8.8	2.754	7.586
.603	.363	4.4	1.660	2.754	.359	.129	8.9	2.786	7.762

TABLE B-1 (CONTINUED)

Voltage ratio	Power ratio	−dB+	Voltage ratio	Power ratio	Voltage ratio	Power ratio	−dB+	Voltage ratio	Power ratio
0.355	0.126	9.0	2.818	7.943	0.211	0.0447	13.5	4.732	22.39
.351	.123	9.1	2.851	8.128	.209	.0437	13.6	4.786	22.91
.347	.120	9.2	2.884	8.318	.207	.0427	13.7	4.842	23.44
.343	.118	9.3	2.917	8.511	.204	.0417	13.8	4.898	23.99
.339	.115	9.4	2.951	8.710	.202	.0407	13.9	4.955	24.55
.335	.112	9.5	2.985	8.913	.200	.0398	14.0	5.012	25.12
.331	.110	9.6	3.020	9.120	.197	.0389	14.1	5.070	25.70
.327	.107	9.7	3.055	9.333	.195	.0380	14.2	5.129	26.30
.324	.105	9.8	3.090	9.550	.193	.0372	14.3	5.188	26.92
.320	.102	9.9	3.126	9.772	.191	.0363	14.4	5.248	27.54
.316	.100	10.0	3.162	10.000	.188	.0355	14.5	5.309	28.18
.313	.0977	10.1	3.199	10.23	.186	.0347	14.6	5.370	28.84
.309	.0955	10.2	3.236	10.47	.184	.0339	14.7	5.433	29.51
.306	.0933	10.3	3.273	10.72	.182	.0331	14.8	5.495	30.20
.302	.0912	10.4	3.311	10.96	.180	.0324	14.9	5.559	30.90
.299	.0891	10.5	3.350	11.22	.178	.0316	15.0	5.623	31.62
.295	.0871	10.6	3.388	11.48	.176	.0309	15.1	5.689	32.36
.292	.0851	10.7	3.428	11.75	.174	.0302	15.2	5.754	33.11
.288	.0832	10.8	3.467	12.02	.172	.0295	15.3	5.821	33.88
.283	.0813	10.9	3.508	12.30	.170	.0288	15.4	5.888	34.67
.282	.0794	11.0	3.548	12.59	.168	.0282	15.5	5.957	35.48
.279	.0776	11.1	3.589	12.88	.166	.0275	15.6	6.026	36.31
.275	.0759	11.2	3.631	13.18	.164	.0269	15.7	6.095	37.15
.272	.0741	11.3	3.673	13.49	.162	.0263	15.8	6.166	38.02
.269	.0724	11.4	3.715	13.80	.160	.0257	15.9	6.237	38.90
.266	.0708	11.5	3.758	14.13	.159	.0251	16.0	6.310	39.81
.263	.0692	11.6	3.802	14.45	.157	.0246	16.1	6.383	40.74
.260	.0676	11.7	3.846	14.79	.155	.0240	16.2	6.457	41.69
.257	.0661	11.8	3.890	15.14	.153	.0234	16.3	6.531	42.66
.254	.0646	11.9	3.936	15.49	.151	.0229	16.4	6.607	43.65
.251	.0631	12.0	3.981	15.85	.150	.0224	16.5	6.683	44.67
.248	.0617	12.1	4.027	16.22	.148	.0219	16.6	6.761	45.71
.246	.0603	12.2	4.074	16.60	.146	.0214	16.7	6.839	46.77
.243	.0589	12.3	4.121	16.98	.145	.0209	16.8	6.918	47.86
.240	.0575	12.4	4.169	17.38	.143	.0204	16.9	6.998	48.98
.237	.0562	12.5	4.217	17.78	.141	.0200	17.0	7.079	50.12
.234	.0550	12.6	4.266	18.20	.140	.0195	17.1	7.161	51.29
.232	.0537	12.7	4.315	18.62	.138	.0191	17.2	7.244	52.48
.229	.0525	12.8	4.365	19.05	.137	.0186	17.3	7.328	53.70
.227	.0513	12.9	4.416	19.50	.135	.0182	17.4	7.413	54.95
.224	.0501	13.0	4.467	19.95	.133	.0178	17.5	7.499	56.23
.221	.0490	13.1	4.519	20.42	.132	.0174	17.6	7.586	57.54
.219	.0479	13.2	4.571	20.89	.130	.0170	17.7	7.674	58.88
.216	.0468	13.3	4.624	21.38	.129	.0166	17.8	7.762	60.26
.214	.0457	13.4	4.677	21.88	.127	.0162	17.9	7.852	61.66

TABLE B-1   (CONTINUED)

Voltage ratio	Power ratio	−dB+	Voltage ratio	Power ratio	Voltage ratio	Power ratio	−dB+	Voltage ratio	Power ratio
0.126	0.0159	18.0	7.943	63.10	0.106	0.0112	19.5	9.441	89.13
.125	.0155	18.1	8.035	64.57	.103	.0110	19.6	9.550	91.20
.123	.0151	18.2	8.128	66.07	.104	.0107	19.7	9.661	93.33
.122	.0148	18.3	8.222	67.61	.102	.0105	19.8	9.772	95.50
.120	.0145	18.4	8.318	69.18	.101	.0102	19.9	9.886	97.72
.119	.0141	18.5	8.414	70.79	.100	.0100	20.0	10.000	100.00
.118	.0138	18.6	8.511	72.44		10^{-3}	30		10^3
.116	.0135	18.7	8.610	74.13	10^{-2}	10^{-4}	40	10^2	10^4
.115	.0132	18.8	8.710	75.86		10^{-5}	50		10^5
.114	.0129	18.9	8.811	77.62	10^{-3}	10^{-6}	60	10^3	10^6
.112	.0126	19.0	8.913	79.43		10^{-7}	70		10^7
.111	.0123	19.1	9.016	81.28	10^{-4}	10^{-8}	80	10^4	10^8
.110	.0120	19.2	9.120	83.18		10^{-9}	90		10^9
.108	.0118	19.3	9.226	85.11	10^{-5}	10^{-10}	100	10^5	10^{10}
.107	.0115	19.4	9.333	87.10		10^{-11}	110		10^{11}
					10^{-6}	10^{-12}	120	10^6	10^{12}

3. **dBm**—The decibel above 1 milliwatt (dBm) is also a useful measure of expressing power level P_2 with respect to a reference power level P_1 of 1 milliwatt (mW). Since the power level in the microwave region is quite low, the dBm unit is very useful in that frequency range. It is customary to designate milli by a lowercase letter m and mega by an uppercase letter M.

4. **dBV**—The decibel above 1 volt (dBV) is a dimensionless voltage ratio in dB referred to a reference voltage of 1 V.

5. **dBµV**—The decibel above 1 microvolt (dBµV) is another dimensionless voltage ratio in dB referred to a reference voltage of 1 µV. The field intensity meters used for the measurements in the microwave region often have a scale in dBµV, since the power levels to be measured are usually extremely low.

6. **µV/m**—Microvolts per meter (µV/m) are units of 10^{-6} V/m expressing the electric field intensity.

7. **dBµV/m**—The decibel above 1 microvolt per meter (dBµV/m) is a dimensionless electric field intensity ratio in dB relative to 1µV/m. This unit is also often used for field intensity measurements in the microwave region.

8. **µV/m/MHz**—Microvolts per meter per megahertz (µV/m/MHz) are units of 10^{-6} V/m per broadband electric field intensity distribution. This is a two-dimensional distribution, in space and in frequency.

9. **dBµV/m/MHz**—The decibel above 1 microvolt per meter per megahertz (dBµV/m/MHz) is a dimensionless broadband electric field intensity distribution ratio with respect to 1 µV/m/MHz.

10. **µV/MHz**—Microvolts per megahertz per second of bandwidth (µV/MHz) are units of 10^{-6} V · s of broadband voltage distribution in the frequency do-

main. The use of this unit is based on the assumption that the voltage is even-lly distributed over the bandwidth of interest.

Figure B-1 shows the conversion of dB scales in power and voltage:

$$\text{dBW} = -30 + \text{dBm} \qquad \text{dBV} = -60 + \text{dBmV}$$
$$= -60 + \text{dB}\mu\text{W} \qquad = -120 + \text{dB}\mu\text{V} \qquad \text{(B-1)}$$

$$\text{dBW} = -30 + \text{dBm}$$
$$= -60 + \text{dB}\mu\text{W}$$

(a) Power scale

$$\text{dBV} = -60 + \text{dBmV}$$
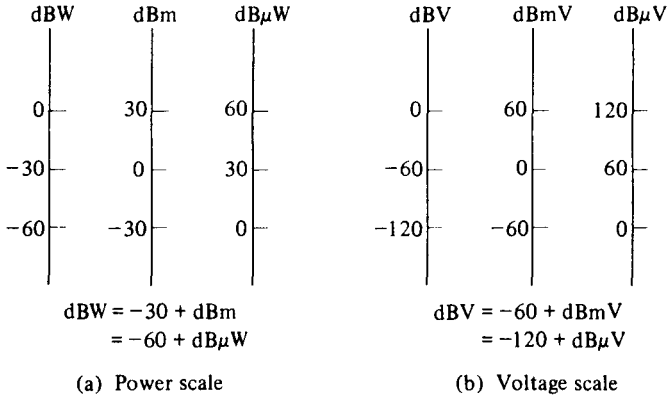$$= -120 + \text{dB}\mu\text{V}$$

(b) Voltage scale

Figure B-1   Conversion of dB scale in power and in voltage.