ALGIE L. LANCE

Hughes Aircraft Company Microwave Standards Laboratory
Instructor, Santa Monica City College
Los Angeles Trade Technical College

INTRODUCTION TO MICROWAVE THEORY AND MEASUREMENTS

McGRAW-HILL BOOK COMPANY

New York
Toronto
London

INTRODUCTION TO MICROWAVE THEORY AND MEASUREMENTS

Library of Congress Catalog Card Number: 63-23468

36104

PREFACE

This book was written as a basic text for use in introductory courses in microwave theory and measurements and to provide a reference for engineers and technicians whose work is related to microwave measurements and microwave systems or components. There is an increasing need for a text which covers basic microwave theory and techniques and the applications of these techniques to measurement problems. This text was developed as a result of teaching microwave courses and it is based on practical experience in the development and applications of precision microwave measurement systems and techniques.

The book presents a compact, logical description of physical concepts, mathematical formulations, measurement systems, and illustrative examples of ideas and measurement procedures. In the organization of this material, particular attention has been given to fundamental principles and applications. The physical significance of the mathematical formulations is presented by use of pictorial diagrams and/or descriptive explanations. Even though this is basically a qualitative approach to the microwave technique, one must realize that an adequate treatment of fundamental principles requires certain mathematical details.

Most of the problems are used to illustrate or amplify points discussed in the text. Some problems outline details omitted from the text in order to shorten the presentation. In general, the problems are designed to help the student gain an understanding of the concepts and techniques and to help form the foundation for understanding more advanced microwave techniques.

The final judgment of relative emphasis on text material is based upon various discussions with my colleagues, especially J. M. Considine, whose suggestions were very helpful in organizing the original outline.

I wish to acknowledge my indebtedness to the many scientists and engineers upon whose work this text is based. Also, I would like to express my appreciation for the continued interest and cooperative efforts on the part of my colleagues. Special thanks are due T. Mukaihata, Head, Microwave Standards Laboratory, Hughes Aircraft Company, for his continued support and encouragement.

Algie L. Lance

CONTENTS

INTRODUCTION

Microwave theory deals with electromagnetic phenomena occurring in the wavelength range from 30 centimeters to a few millimeters. This is a general consideration since boundaries of the microwave region are not specifically defined.

The transition between the specialized point of view of lumped-constant circuits and the fundamental approach to electromagnetic theory occurs in the region between 50 and 100 Mc, which corresponds to wavelengths of 6 and 3 m. In this region the lumped-constant analysis is largely replaced by wave theory associated with the conventional transmission line. It is proper to regard transmission of power as taking place through the space between conductors and not through the conductors themselves. This concept seems to hold regardless of the frequency range. The power is specified by the intensities of the electric and magnetic fields and the velocity with which the configuration is propagated along the line.

It is possible to introduce some of the most fundamental concepts of electromagnetic wave theory without becoming involved in all the complications of vector field theory. This approach to the subject begins in Chap. 2.

The distribution of the electric and magnetic fields must be measured in order to explain the behavior of the given microwave system. Precise information concerning the configuration of these fields can be obtained by proper application of the impedance concept in one form or another.

The importance of measurement techniques in the microwave field cannot be overemphasized since progress in the application of microwaves demands the development of measurement techniques.

Microwaves have a broad range of application in the numerous forms of "communication of information." The overall trend in expansion is towards microwaves, thus placing increasing demands for knowledge of the essentials of this subject.

CHAPTER

1

INTRODUCTION TO ELEMENTARY FIELD THEORY

The *current* aspect of electricity meets most needs and requirements at low frequencies; only occasionally is there a need to discuss lines of electric and magnetic force. There is a tendency to regard the electric and magnetic fields as almost unrelated quantities since their roles are so different at the lower frequencies. At higher frequencies these fields are so intimately related that they may be regarded, at times, as different aspects of the same thing.

We are concerned with phenomena which are described in terms of the motions of charges. An introduction to the basic nature of charges leads to a description of the fundamental properties of electric and magnetic lines of force and serves as a background for the basic definitions of transmission line parameters. Therefore, an understanding of the basic nature of charges and of the relationship of these charges to electric and magnetic fields and to the basic definitions of transmission line parameters is necessary. The manner in which conductors and dielectrics affect the charge distribution and the electric and magnetic field distribution must be known.

1·1 Basic phenomena

A fundamental property of electricity is that every particle of it exerts a force on every other particle. The magnitude of this force depends on the electric charges of the particles, the medium in which the charges are placed, and their locations in space and time.

1·2 Electrostatic fields

Electricity has a dual aspect in that it consists of an electric charge q, which exists at a point (infinitesimal volume), and an electric flux ψ, which occupies the rest of space but with diminishing intensity with distance. *Force exists by the interaction of the flux of one particle upon the center of source of flux of another particle; the region in space in which the force can be detected is called the "field" of the charge.*

1·3 Force between charges

The unit charge of electricity, the coulomb, is defined in terms of the experimental law of Coulomb, which gives the force between electric charges and includes the following information:

1. Like charges repel and opposite charges attract.
2. Force is dependent upon the medium in which the charges are located and acts in a line joining the charges.
3. Force is proportional to the charge magnitudes.
4. Force is inversely proportional to the square of the distance between the charges.

$$F = \frac{q_1 q_2}{4\pi\epsilon r^2} \qquad \text{mks system of units} \tag{1·1}$$

F is the force in newtons (1 newton $= 10^5$ dynes or 102 g weight) that q_1 plus all the polarized particles of the dielectric exerts upon q_2. ϵ is a property of the medium which may be called the permittivity (farads per meter). q_1 and q_2 (coulombs per meter) represent the magnitude and sign of the charges. The distance between the charges is r. The dyne is the force necessary to accelerate a mass of one gram at the rate of one centimeter per second per second. The coulomb is a charge of 6.25×10^{18} electrons. The quantity $1/4\pi$ is a constant of proportionality peculiar to the mks system of units.

The advantage of the mks system of units is that the units of the electric quantities are those actually measured. *Length* is in *meters*, *mass* is in *kilograms*, *time* is in *seconds*, *current* is in *amperes*, *potential* is in *volts*, *impedance* is in *ohms*, and *power* is in *watts*.

1·4 Fields in dielectrics

If the charges $-q$ and $+q$ are immersed in a dielectric medium, the particles polarized by these charges will line up as shown in Fig. 1·1. The resultant effects cancel out in the medium, and, since the polarized particles are of opposite sign, they tend to reduce the effective magnitudes of the charges.

$$F_f = \epsilon' F \tag{1·2}$$

F_f is the force due to free charges, i.e., charges in free space with no polarized

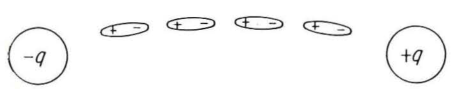


Fig. 1·1 Polarized particles in a dielectric between charges.

charges. ϵ' is the relative permittivity and is a function that F must be multiplied by to obtain F_f. The permittivity of the dielectric is defined by

$$\epsilon = \epsilon' \epsilon_0 \tag{1·3}$$

ϵ_0 is the *permittivity of free space* or *specific inductive capacity*. If the medium is free space, without polarized charges, then

$$\epsilon = \epsilon_0 \qquad \text{and} \qquad F = F_f$$

The nature of permittivity ϵ, or relative permittivity ϵ', depends upon the medium in which the field exists. If the polarization properties of the medium are such that, throughout the entire field, the medium is *linear* (ϵ' does not depend upon the magnitude of the flux density), *isotropic* (ϵ' does not depend upon the direction of flux density), and *homogeneous* (ϵ' does not depend upon the location of the flux density), then ϵ' is a scalar constant called the *dielectric constant*.

Free space, having no polarized particles, is linear, homogeneous, and isotropic. Therefore, the *permittivity of free space* ϵ_0 is a universal constant.

$$\epsilon_0 = \frac{1}{36\pi} \times 10^{-9} = 8.854 \times 10^{-12} \text{ farad per m} \tag{1·4}$$

1·5 Electric field strength

The *electric field strength* or *electric field intensity* $\mathbf{E}$ at a point is defined as the *force per unit charge* exerted upon a test charge placed at a point.

$$\mathbf{E} = \frac{F}{q} \tag{1·5}$$

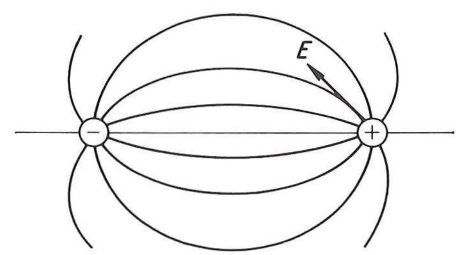
where F is the vector force acting upon the infinitesimal test charge q.

The electric field arising from the test charge is

$$\mathbf{E} = \frac{q}{2\pi\epsilon r^2}(\bar{a}_r) \qquad \text{volts per m} \tag{1·6}$$

Since $\bar{a}_r$ is the unit vector directed from the point in a direction away from the charge, the electric field vector points away from positive charges and toward negative charges. The electric field strength has magnitude and direction and

Fig. 1·2 Electrostatic lines of force in the region between two oppositely charged spheres.



is therefore a vector quantity. If the positive direction is taken, a positive test charge (proton) is displaced, as shown in Fig. 1·2. The electric field strength is measured in *volts per meter*.

1·6 Magnetic field strength

The force exerted on a unit magnetic pole is a measure of the *magnetic field intensity* **H**. It is a vector quantity measured in *amperes per meter*. The direction of force on the unit north pole is shown in Fig. 1·3. The force

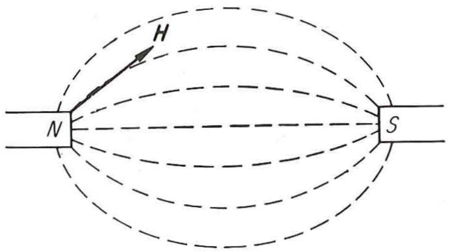


Fig. 1·3 Direction of force on a unit north pole in the region between two oppositely magnetized poles.

between magnetic charges (pole strength) m_1 and m_2, measured in webers, is expressed by

$$F = \frac{m_1 m_2}{4\pi\mu r^2} \quad \text{newtons} \tag{1·7}$$

B is the magnetic flux density in webers (1 weber is 10^8 maxwells or "lines"). This choice is convenient in time-varying fields since a rate of change of 1 weber per second generates an electromotive force of 1 volt. This measure of volt-seconds per square meter equals henrys per meter from the definition of inductance.

$$\mathbf{B} = \mu_0 \mathbf{H} \tag{1·8}$$

where μ_0 is the permeability of free space. Permeability is the unit of measurement which indicates the ease with which a magnetic field may be set up in a material. It represents the ability of the medium to support tubes of magnetic force.

The permeability of the medium

$$\mu = \mu' \mu_0 \tag{1·9}$$

where μ' is the relative permeability. μ_0 is the permeability of free space in *henrys per meter*.

$$\mu_0 = 4\pi \times 10^{-7} \quad \text{henry per m} \tag{1·10}$$

1·7 Resistance and conductance

From Ohm's law, the resistance $R = E/I$.

When two equipotential surfaces (for example, two conductors) have a potential difference between them, a certain amount of current will flow

because of the finite resistance of the insulation. This leakage is called *conductance* and is determined by the dielectric medium between the surfaces or conductors. *Conductance G* is the factor by which the potential between the two equipotential surfaces, at any instant, must be multiplied to give the current flowing between the two surfaces.

$$I = GE \quad \text{or} \quad G = \frac{I}{E}$$

from the definitions $R = 1/G$ and $G = 1/R$.

1·8 Capacitance and inductance

If a potential E exists between two equipotential surfaces (such as two conductors), an electric charge $+q$ will be set up on the surface at a positive potential. An equal charge of $-q$ will, from the principle of conservation, be set up on the other surface. The *charge per unit difference of potential* is called *capacitance C* and is measured in farads.

Inductance. Inductance is the property of a circuit by which it opposes any change in current. It manifests itself in a back emf (electromotive force) that is developed when current is changed. *The inductance of a circuit is the back emf induced in it by a unit time rate of change of current.* The unit of inductance is the *henry*. Inductance can also be defined as *flux linkages per unit current.*

1·9 Properties of electric and magnetic fields

The properties of electric and magnetic fields which explain numerous phenomena of electrical transmission are as follows:

1. When lines of electric force are displaced laterally, lines of magnetic force are induced in the immediate adjacent space. This resultant magnetic force has an intensity **H** which is proportional to the velocity **v** of displacement and to the intensity **E** of the electric force. The direction of the induced magnetic force is normal to the direction of the original electric force. This property of the electric and magnetic fields is expressed by the vector notation

$$\mathbf{H} = \epsilon(\mathbf{v} \times \mathbf{E}) \tag{1·11}$$

The above notation is pronounced "v cross E." The physical significance of this notation concerning the direction of **H** is given by the right-hand rule and is applied as follows:

> Let the curled fingers of the right hand lie in the plane of **v** and **E** and point in the direction from **v** to **E** through the smaller angle; the thumb will point in the direction of **H**.

This rule is used for the vector or cross product of any two vectors. It can be seen that in Fig. 1·4, **v** × **E** gives the direction of **H** to the left.

2. When lines of magnetic force are displaced laterally, lines of electric force are induced in the immediate adjacent space. This resultant electric force has an intensity **E** which is proportional to the velocity **v** of displacement and the intensity **H** of the magnetic force. The direction of the induced electric force is normal to the direction of motion and also normal to the direction of the original magnetic force. This property of the electric and magnetic fields is expressed by the vector notation

$$\mathbf{E} = -\mu(\mathbf{v} \times \mathbf{H}) \qquad (1\cdot12)$$

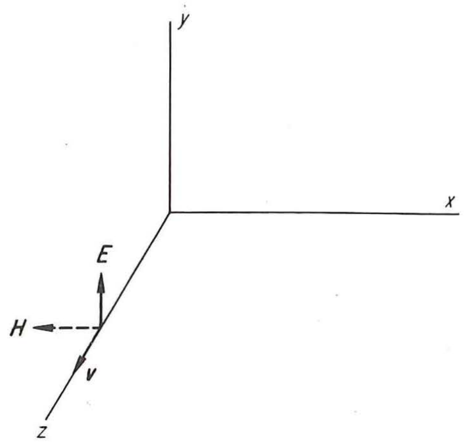


Fig. 1·4 Direction of the electric and magnetic vectors relative to the velocity.

In applying the right-hand rule to **v** × **H** in Fig. 1·4 it is noted that the thumb points in the direction *opposite* to the direction of **E** (downward). The minus (—) sign in the equation indicates that the direction of **E** is opposite the direction as determined by **v** × **H**.

1·10 Energy flow

The amount of energy transferred from one point to another depends upon the magnitudes, distribution, and phases of the electric and magnetic fields. The Poynting concept specifies that the magnitude of the energy flow per unit volume across a unit area measured perpendicular to **v** is proportional to the product of **E** and **H** and is in a direction normal to both **E** and **H**. The vector notation for the Poynting concept is

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} \qquad (1\cdot13)$$

The relative directions of **P**, **E**, and **H** are shown in Fig. 1·5. The energy moves at a velocity defined by

$$\mathrm{v} = \frac{1}{\sqrt{\mu\epsilon}} = \frac{1}{\sqrt{\mu'\mu_0\epsilon'\epsilon_0}} = \frac{1}{\sqrt{\epsilon'}}\frac{1}{\sqrt{\mu_0\epsilon_0}} \qquad (1\cdot14)$$

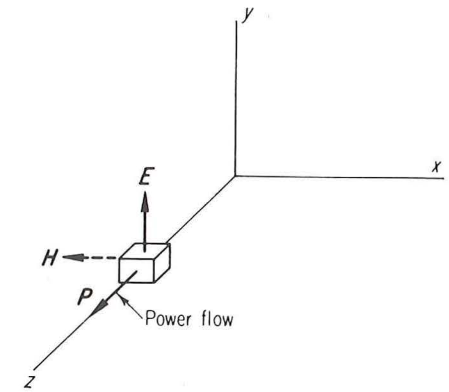


Fig. 1·5 Direction of the electric and magnetic vectors relative to the Poynting vector.

1·11 Electric current

The current per unit area is called the *current surface density* **J**, a vector measured in amperes per square meter. By vector notation, **J** = **n** × **H** and is illustrated in the diagram of Fig. 1·6. **n** represents a unit vector perpendicular to the surface, and **H** is the magnetic field tangent to the surface. It is noted that the direction of the current flow is normal to the magnetic field.

1·12 Boundary conditions

The tangential components of the electric and magnetic fields must be continuous in traversing the interface between physically real media. That is, the tangential components of the electric and magnetic fields must be equal on the two sides of the boundary. Therefore, the amplitudes of the tangential components of the incident and reflected waves at the interface must equal the amplitude of the transmitted components of the transmitted wave.

No conductor is perfect, but in many practical problems it is desirable to neglect the finite electric field along the conductor. *In the case of an ideal perfect conductor, there can be no component of electric field tangential to the surface and no component of time-varying magnetic field perpendicular to the surface.*

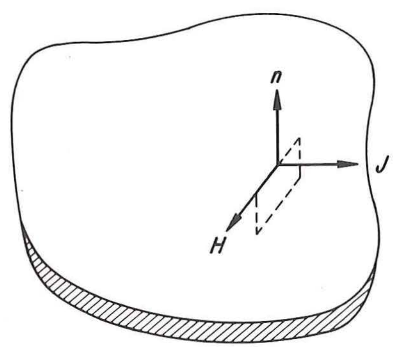


Fig. 1·6 Direction of current surface density relative to the magnetic field and the unit vector normal to the surface.

PROBLEMS

1·1 Calculate $\sqrt{\mu_0/\epsilon_0}$. What is the significance of this value?

1·2 Calculate $1/\sqrt{\mu_0\epsilon_0}$. What is the significance of this value?

1·3 If $q_1 = q_2 = \frac{1}{3} \times 10^{-9}$ coul per m, $F = 10^5$ dynes, and q_1 and q_2 are spaced 1 cm apart (10^{-2} m), calculate ϵ.

1·4 Calculate the velocity in a medium which has a dielectric constant of 2. Repeat for dielectric constants of 6 and 12.

1·5 Consider a sheet of paper as the plane of the following diagrams and indicate the direction of the third vector given in the parentheses.
a. **E** points up and **v** points left. Show the direction of (**H**).
b. **v** points to the left and **H** points up. Show the direction of (**E**).
c. **E** points left and **H** points down. Show the direction of (**P**).
d. **P** points left and **H** points up. Show the direction of (**E**).

1·6 Draw a diagram showing an electric and magnetic field approaching a perfect conducting metal sheet. Draw a diagram alongside this sheet to indicate the magnitude of the electric and magnetic fields at the surface of this perfect conductor.

1·7 Draw an end view of a single conductor showing concentric lines of magnetic field directed clockwise around the conductor. Show electric lines of force normal to the surface and directed away from the conductor.
a. What is the direction of power flow?
b. If there is a component of electric field at the surface of the conductor directed away from the observer (parallel to the wire), what is the direction of power flow due to this component of electric field?
c. What is indicated by part *b*?

CHAPTER

2

TRANSMISSION LINES

Introduction. A transmission line is any structure used to guide the flow of energy from one point to another. This general definition is required as evidenced by the cross sections of various types of guiding structures shown in Fig. 2·1.

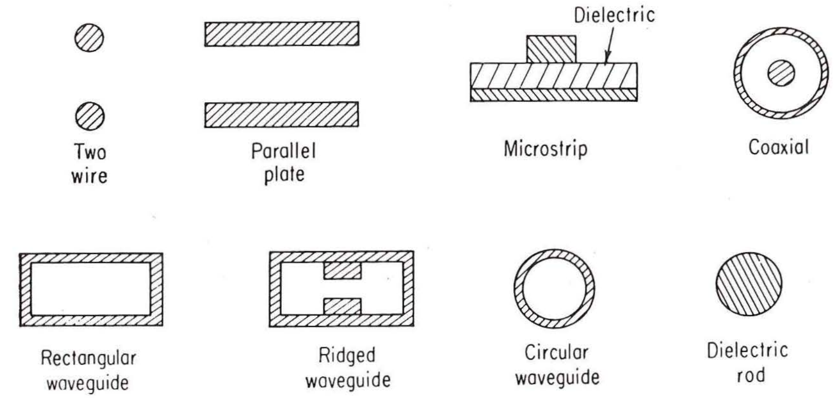


Fig. 2·1 Cross-sectional configurations of various types of guiding structures.

This chapter presents an introduction to basic transmission line theory involving the behavior of voltages and currents applied to the basic two-wire transmission line. The basic theory of reflections is the starting point for the major parts of this text. A treatment of the familiar two-wire transmission line is given, since it exhibits certain properties which are common to all types of transmission lines. The fundamental assumption is that the uniform spacing of the wires is so close that the effect in one wire of a change in current in the other wire is instantaneous. Therefore, the conductor spacing must be small compared to wavelength, and the length of the line must be long compared to the spacing when a description of the behavior of sinusoidal waves of voltage and current is considered.

2·1 The two-wire transmission line

The two-wire transmission line consists of two parallel conductors properly supported and insulated from each other.

The concept of movement of charge along a two-wire transmission line from a current-forcing source is illustrated in Fig. 2·2. For simplicity, it is customary to say that the charge flows into the line or that a charge moves

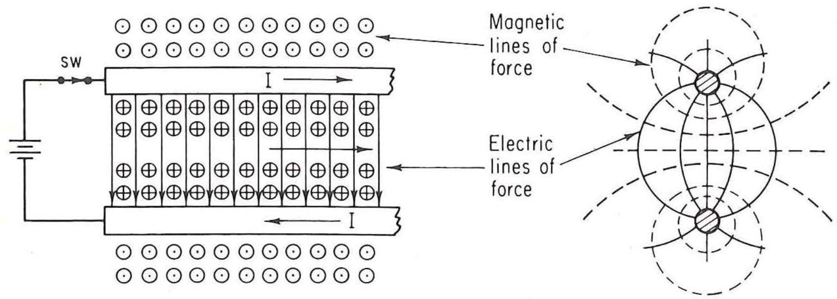


Fig. 2·2 Transmission of d-c power along a two-wire line.

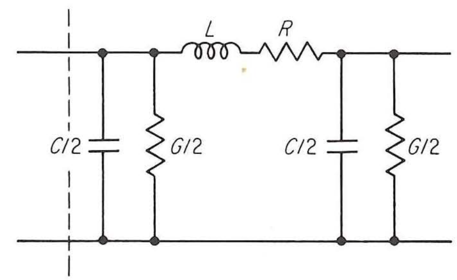
down the line. It should be noted, however, that individual charges within the conductor do not change their positions. The movement of electrons constitutes a current that produces an equivalent movement of charge. Also, no individual electron travels any great distance from its original position on the line.

It is necessary to describe the behavior of the charges on the two-conductor transmission line of Fig. 2·2. When the source voltage is connected to the two wires, a current flow exists because of the flow of charges into the line. There is a magnetic field, proportional to this current, surrounding the conductors. The associated flux linkages per unit current I are called *inductance L*. Therefore, there is an inductance per unit length of line when a unit current is flowing. The charge on the conductor is proportional to the potential difference (voltage). Therefore, the line has *shunt capacitance C*, previously defined as the charge per unit potential difference. If the dielectric medium between the conductors is not perfect, a conductive element must be assumed between the lines to account for this loss. This is the *conductance G* per unit length of line. In addition to the above parameters, there is a *series resistance R* associated with the conductors, since the perfect conductor does not exist in practice. The resistance depends upon the resistivity of the material, the length and cross section of the conductor, and the distribution of currents in the cross section.

If a given length of this transmission line were divided into more and more sections, the ultimate case would be an infinitesimal section of the basic elements, resistance R, conductance G, inductance L, and capacitance C. The equivalent circuit that results is shown in Fig. 2·3. When the source is applied to the line, it "sees" this first section of line which is made up of the basic circuit parameters.

The voltage generated by the flow of charge is referred to as a *voltage wave*, and the current induced in the line is referred to as the *current wave*. The relationship of the voltage wave to the current wave, the relationship of the electric field to the magnetic field, and the velocity with which these waves travel down the transmission line are determined by the values of the basic

Fig. 2·3 Equivalent circuit of a transmission line.



circuit parameters. This is a physical concept of the electromagnetic theory set forth in Chap. 1, where it was pointed out that the motion of electric field lines of force gives rise to magnetic lines of force in the immediate vicinity, and the two fields together give rise to component Poynting vectors representing power flow. The series resistance and inductance are shown in one line only, but they are actually in both lines and can be distributed either equally or unequally. If the basic transmission line parameters are spread evenly along the entire length of the line, the constants are said to be *distributed*. The capacitance of the line is spread out along the line; the effect of this capacitance is not the same that would be obtained if it were all centered at one point.

2·2 Displacement current

In Fig. 2·2 the current is shown as flowing to the right in the upper conductor and to the left in the lower conductor. The direction of force and current flow is given for conventional current flow; therefore electron flow is in the opposite direction. There seems to be an inconsistency since the current path is not complete. This apparent inconsistency can be resolved by taking into account the *displacement current*, which is defined as the *time rate of change of electric flux through a surface*. The electric field arises from the moving charge and is varying with time, thus producing a magnetic field proportional to the electric field. The current, other than leakage conductance current, from one conductor to the other is $C\,dv/dt$, where dv/dt is the change in voltage with respect to time. Another example of displacement current is the capacitor in an electronic circuit. The completed current path is by means of a displacement current between the capacitor plates.

2·3 Traveling waves on a lossless transmission line

In most microwave transmission lines the losses are extremely small, and in most practical cases the line can be considered lossless. If the loss parameters R and G are zero, then the equivalent circuit of Fig. 2·3 is reduced to series sections of inductance and shunt sections of capacitance.

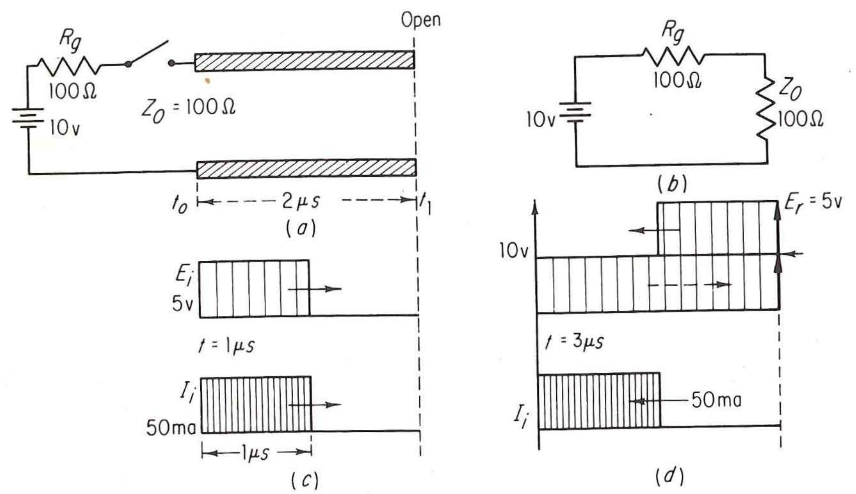


Fig. 2·4 Traveling waves on a transmission line terminated in an open circuit.

When the switch is closed (Fig. 2·2), the battery "sees" the first infinitesimal section of inductance and capacitance. The voltage cannot appear instantaneously at all points on the line because the series inductance of the line, which is associated with the magnetic flux, opposes a change in current, and the shunt capacitance, which is associated with the electric charge, opposes a change in voltage. A finite time is required to charge the capacitance of each small section of the line.

The voltage wave can progress down the line only as fast as the current can carry the necessary charge to the wavefront to produce the change in voltage. Also, the current can travel down the line only as fast as the voltage that is required, at the wavefront, to force the current through each short section of line inductance. *Therefore, the voltage and current must travel together along the line.* In Fig. 2·4 the current wave and voltage wave are shown in step at the wavefront with a current flowing away from the battery in the top conductor and a current flowing toward the battery in the lower conductor. The current path is completed by the displacement current between the conductors as explained in Sec. 2·2.

Define τ as the time required for waves of voltage and current to travel a unit length along the line. The velocity of travel is $v = 1/\tau$.

During the time interval τ, $I\tau$ is the charge that flows into the line, and CE is the charge accumulated on the line. Therefore,

$$I\tau = CE \tag{2·1}$$

LI is the increase in flux linkages encircling the conductor. $E\tau$ is the rate at which flux linkages are being produced at the wavefront. Therefore,

$$E\tau = LI \tag{2·2}$$

Multiply Eq. (2·1) by Eq. (2·2) and solve for τ.

$$\tau = \sqrt{LC} \qquad \text{and} \qquad v = \frac{1}{\sqrt{LC}}$$

Divide Eq. (2·2) by Eq. (2·1) and multiply by E/I.

$$\frac{E}{I} = \sqrt{\frac{L}{C}} \qquad \text{or} \qquad Z_0 = \sqrt{\frac{L}{C}} \tag{2·3}$$

The ratio E/I is called the *characteristic impedance* Z_0 of the transmission line. Z_0, *the characteristic impedance, is the ratio of the voltage to the current traveling in a particular direction.* This means that Z_0 is the *ratio of voltage to current traveling together in one direction or the other on the line.* By this definition it can be seen that any change in voltage and current on a transmission line has a constant of proportionality which is the characteristic impedance Z_0 of the line. This also indicates that regardless of the initial current and voltage conditions, if a wave of voltage and current is sent down the line, the voltage and current waves still travel at the velocity determined by the line parameters, and the ratio of the voltage to current is still Z_0 because the transmission line is a linear device.

Waves launched on a uniform two-conductor line that is infinitely long are assumed to be propagated to infinity. If the transmission line is terminated in a pure resistance equal to the characteristic impedance, a steady state would exist, since this value of resistance would exactly support the voltage and current of the traveling wave. This is, in effect, a transmission line of infinite length. The importance of this particular value of termination will become more apparent in subsequent discussions.

2·4 Reflections at an open-line termination

Reflections are caused by discontinuities in the transmission line structure. These discontinuities may be regarded as changes in the characteristic impedance of the transmission line or guiding structure. The circuit in Fig. 2·4 with a charging resistor R_g equal to the characteristic impedance Z_0 simplifies the initial discussion of the behavior of the open-circuit termination. When the switch is closed, the line appears as shown in the equivalent circuit of Fig. 2·4*b* because the battery sees Z_0 in series with the charging resistor R_g. Waves of voltage and current start from the source toward the open termination. These waves are designated *incident* waves and are labeled E_i (incident voltage wave) and I_i (incident current wave).

Assume that the velocity of propagation is such that a voltage wave and current wave can travel the length of the line in 2 μsec, as indicated in Fig. 2·4. At t_0, 5 volts is dropped across R_g and a 5-volt wave of voltage accompanied by a 50-ma current wave starts down the transmission line. These wavefronts travel *together* down the line at a velocity equal to $1/\sqrt{LC}$ and charge each element of line capacitance to 5 volts. No capacitance remains to be charged when the wave reaches the end of the line after 2 μsec of travel. The collapsing magnetic field, associated with the inductance element at the end of the line, tries to maintain the original current flow into the capacitance. This additional current flow charges the end capacitance to 10 volts ($2E_i$), and the current at the end of the line has decreased to zero. This action continues from section to section back toward the battery, as shown in Fig. 2·4*d*. This corresponds to a *reflected* wave of voltage E_r equal to the incident wave E_i, and a *reflected* wave of current I_r equal to the incident current I_i traveling back toward the battery. The reflected waves travel at the same velocity ($v = 1/\sqrt{LC}$) as the incident waves, and the ratio of E_r to I_r is equal to the characteristic impedance Z_0. The entire line is charged to 10 volts when the reflected wave arrives at the battery after a total travel time of 4 μsec. The steady-state condition is reached; a voltage of 10 volts exists between the conductors, and no current flows in the conductors.

Note: At an open-circuit termination, the *reflected voltage* E_r is *equal* in magnitude and is *in phase* with the incident voltage E_i. The reflected current I_r is *equal to* and 180° *out of phase* with the incident current I_i.

2·5 Reflections at a short-circuit termination

Another simple form of reflection occurs when the transmission line of Fig. 2·4 is terminated with a transverse sheet of metal which is assumed to be a perfect conductor (short circuit). At the instant the incident waves of voltage and current reach the termination, the boundary conditions of the perfect conductor indicate that there can be no tangential component of electric field, i.e., there can be no voltage drop across zero resistance.

The resultant zero voltage at the surface can be accounted for if it is assumed that the reflecting surface merely reverses the direction of the electric lines of force, thereby giving rise to a reflected wave which cancels the incident wave at the surface of the conducting plate or short circuit.

The behavior at the end of the line can also be described in terms of the inductance and capacitance of the line as considered for the open circuit. When the 5-volt, 50-ma incident waves reach the short circuit, a reflected voltage of 5 volts with reversed polarity starts back toward the battery. The line capacitance at the wavefront is discharged by a net current of 50 ma, and the voltage between the conductors drops from 5 volts to zero. This 50 ma of current adds to the 50-ma incident current, and the total line current flow is 100 ma at the wavefront. When the reflected wave reaches the battery, the voltage between the conductors is zero and the current flow is 100 ma. This is the steady state since the 100 ma of current is exactly the current drawn by the 100-ohm resistor R_g across a 10-volt battery. At a short circuit E_r is *equal to* E_i but has *opposite polarity*, while I_r is *equal to* and *in phase* with I_i.

2·6 Reflections from resistive terminations

If the transmission line is terminated in a resistance equal to Z_0, the line behaves as though it has infinite length. Since there are no reflected waves on the line, it is said to be match terminated.

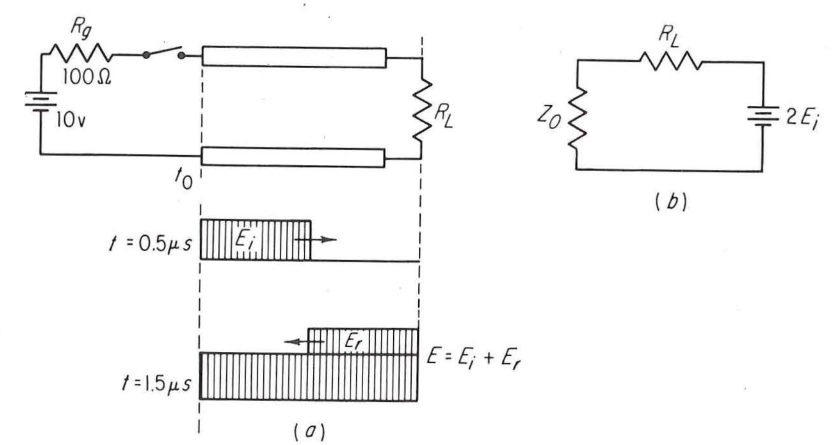


Fig. 2·5 Traveling waves on a transmission line terminated in a resistance greater than Z_0.

If the line is terminated in any resistance other than Z_0, there will be reflected waves from the termination. The amplitude and polarity of these reflected waves will be determined by the particular value of load and the value of characteristic impedance of the line.

Consider the line in Fig. 2·5 which is terminated with a pure resistance R_L somewhat greater than Z_0. After 1 μsec of time, E_i and I_i reach the resistive load termination R_L. The voltage cannot jump to $2E_i$ and it cannot go to zero as in the open-circuit or short-circuit considerations. The voltage at the load E_L can be determined by calculating E_L for an open circuit and then using Thévenin's theorem to find the value of E_L; then calculate the value of reflected voltage E_r.

Assume that at the instant the voltage on the open line jumps to $2E_i$, the load resistance Z_L is connected to the line. The equivalent circuit is shown in Fig. 2·5*b*. Since Thévenin's theorem for this circuit states that the current through Z_L is the ratio of the open-circuit voltage and the sum of Z_L and Z_0 of the line,

$$I_L = \frac{2E_i}{R_L + Z_0}$$

$$E_L = \frac{(R_L)2E_i}{R_L + Z_0} = 2E_i \frac{R_L}{R_L + Z_0} \tag{2·4}$$

E_L is the *total* voltage at the end of the line. The reflected voltage is obtained by subtracting the incident voltage E_i from this value.

$$E_r = 2E_i \frac{R_L}{R_L + Z_0} - E_i = E_i \frac{R_L - Z_0}{R_L + Z_0} \tag{2·5}$$

The ratio of the reflected voltage E_r to incident voltage E_i is called the *reflection coefficient* Γ (gamma).

$$\Gamma = \frac{E_r}{E_i} = \frac{R_L - Z_0}{R_L + Z_0} \tag{2·6}$$

The reflection coefficient is calculated from the above equation *provided the load is a pure resistance*. The reflection coefficient for sinusoidal waves and complex impedance is considered in a subsequent chapter.

The initial voltage and current wave for the lines discussed depends upon the value of R_g and Z_0 as shown by the equivalent circuit in Fig. 2·5*b*. In each of the previous discussions the steady-state condition was reached after the reflection from the load reached the battery. If the value of R_g had not been equal to the characteristic impedance, there would have been a re-reflection from the source end of the line. The re-reflected wave can be determined in the same way that the reflection from the load was determined. The resistance R_g is now the load for the reflected wave, and the reflection coefficient is calculated from the values of R_g and Z_0. Therefore, when there is a mismatched load and a mismatched source, there may be several round-trip reflections before the steady state is reached. As an example, 90 volts is applied to the line where $R_g = 2Z_0$. The voltage across R_g is 60 volts, and

a 30-volt wave of voltage travels down the line to the open circuit illustrated in Fig. 2·4. At the end of the line the 30 volts is reflected in phase and travels back toward the source. When the 30-volt wave arrives at R_g, there is 60 volts on the line. The 30-volt reflected wave encounters the mismatch of $2Z_0$, and part of the 30-volt wave is re-reflected. From the reflection coefficient equation,

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{2Z_0 - Z_0}{2Z_0 + Z_0} = \frac{1}{3}$$

Therefore a new incident wave of $(\frac{1}{3})30 = 10$ volts travels from the source toward the open end of the transmission line. This 10-volt wave is reflected from the open circuit, and upon arriving back at the input, one-third of this voltage is re-reflected. This process continues until there is 90 volts on the transmission line.

Conclusions: Any change on a transmission line takes the form of voltage and current waves traveling in one direction or the other or of pairs of voltage and current waves traveling in opposite directions on the transmission line. The characteristic impedance is the ratio of the voltage wave to the current wave traveling together in a particular direction. Z_0 is determined by the physical characteristics of the line which in turn determine $\sqrt{L/C}$. The waves of voltage and current travel at a velocity of $1/\sqrt{LC}$.

If a transmission line is terminated in a pure resistance Z_L *greater than* the characteristic impedance Z_0, the reflected voltage wave E_r is *in phase* with the incident voltage wave E_i, and the reflected current wave I_r is 180° *out of phase* with the incident current wave I_i.

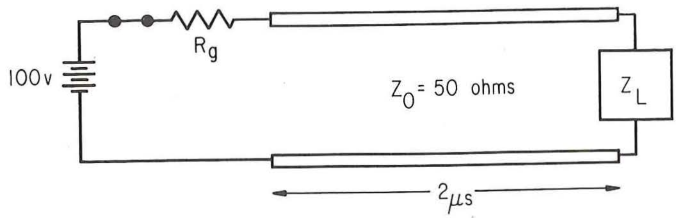
If a transmission line is terminated in a pure resistance Z_L *less than* the characteristic impedance Z_0, the reflected voltage wave E_r is 180° *out of phase* with the incident voltage wave E_i, and the reflected current wave is *in phase* with the incident current wave I_i.

If the source impedance (R_g in this discussion) is not equal to the characteristic impedance Z_0 of the line, there will be reflections from the source end of the line back toward the load.

PROBLEMS

2·1 An artificial transmission line can be formed using lumped L and C. Calculate the delay of an artificial line composed of eight sections of inductance $L = 4$ mh per section, and capacitance $C = 40$ $\mu\mu$f per section.

2·2 Draw a diagram illustrating the sinusoidal electric and magnetic fields on a two-wire line and show the complete loops of current flow including the displacement current. Refer to the following diagram when working Probs. 2·3 to 2·10.



2·3 In the above diagram, R_g is 50 ohms and Z_L is an open circuit. Calculate the total voltage at the points indicated by the following time intervals: (*a*) 1 μsec, (*b*) 2 μsec, (*c*) 3 μsec, (*d*) 4 μsec, and (*e*) 6 μsec.

2·4 Repeat Prob. 2·3 if Z_L is a short circuit.

2·5 Repeat the calculations in Prob. 2·3 if R_g is 450 ohms and Z_L is an open circuit.

2·6 Repeat the calculations in Prob. 2·3 if R_g is 450 ohms and Z_L is a short circuit.

2·7 Repeat the calculations in Prob. 2·3 if R_g and Z_L are 450 ohms.

2·8 Repeat the calculations in Prob. 2·3 if R_g is 25 ohms and Z_L is 50 ohms.

2·9 If R_g is 50 ohms and the reflected voltage is 20 volts, (*a*) What is the value of load resistance? (*b*) What is the value of load resistance if the reflected voltage is −20 volts?

2·10 If R_g is 50 ohms and Z_L is 150 ohms, calculate (*a*) reflected voltage, (*b*) incident power, and (*c*) reflected power. (*d*) Express the ratio of reflected to incident power in terms of the voltage reflection coefficient. (*e*) If the ratio calculated in *d* is multiplied by 100, what does the resulting value indicate?

CHAPTER

3

TRANSMISSION LINES AT MICROWAVE FREQUENCIES

Introduction. A variety of phenomena may take place when an alternating electromotive force is connected to the transmission line. The concepts of resonant lines, traveling waves, standing waves, propagation constant, and impedance are presented for dissipationless transmission lines.

The useful methods of analysis gained from this study of transmission lines propagating the *transverse electromagnetic* (TEM) wave can also be applied to problems involving such guiding structures as waveguides. In the TEM wave, both the electric and magnetic fields are transverse (at right angles) to the direction of wave travel. There are no electric or magnetic components in the direction of wave travel for lossless lines. TEM waves exist in free space and two-conductor transmission lines such as two-wire, parallel-plate, microstrip, and coaxial lines and resonators.

All microwave phenomena can ultimately be expressed in terms of frequency, wavelength, and power. This chapter is devoted to the development of analytical expressions involving these quantities, although power is not used directly in the analysis. Instead, the original electric and magnetic field problems are reduced to an impedance problem in which discontinuities are expressed in terms of impedances rather than field intensities. The generalized impedance concept (the complex ratio of voltage to current) forms the foundation of transmission line theory and will be considered in detail.

3·1 Transmission of power at high frequencies

If the frequency is very high, the initial lines of force sent out by the source will not travel far before the emf at the source reverses direction. Figure 3·1*b* shows that this second group of lines of force is exactly like the first group except that the lines are directed in the opposite direction. Alternate groups are observed to be identical. Figure 3·1*a* shows the lines of electric and magnetic force on the transmission line. Figure 3·1*b* shows the space relationship between the electric and magnetic fields. Most of the total power flow takes place in the immediate vicinity of the wire, as shown in the diagram

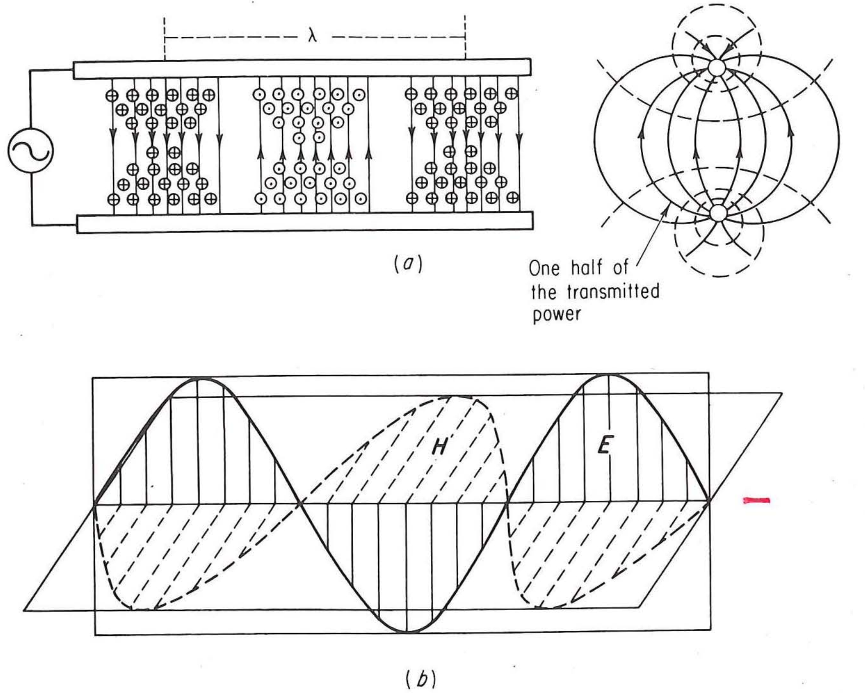


Fig. 3·1 (a) Electric and magnetic lines of force in the longitudinal and transverse sections of an infinitely long transmission line. (b) Space relationships of electric and magnetic vectors.

by the circle enclosing one-half of the transmitted power. The remaining half of the power extends to infinity.

3·2 Traveling waves

The voltage is always accompanied by a current wave of similar shape on a uniform lossless transmission line, and, regardless of their shape, these waves will be propagated without any change in magnitude or shape.

Figure 3·2 shows a sinusoidal traveling wave of voltage on a transmission line. This line has a *physical length l* measured in meters. It also has *electrical length* which is measured in wavelengths λ. The *wavelength* λ is defined as the distance between successive points of the same electrical phase in a wave. It depends on the frequency f of alternation, the velocity of propagation, and the nature of the medium through which the wave travels. For free space, the velocity is substantially 300,000,000 m per sec (186,000 miles per

sec). If the medium has a relative dielectric constant, the wavelength will be reduced by the factor $1/\sqrt{\epsilon'}$.

$$\lambda = \frac{v}{f} = \frac{1}{\sqrt{\epsilon'}}\frac{v}{f} \tag{3·1}$$

Wavelength is also defined as the distance in which the phase changes by 2π rad (1 rad $= 180°/\pi = 57.3°$).

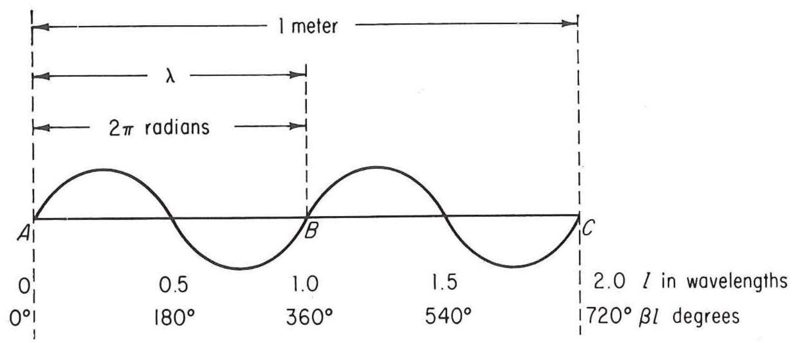


Fig. 3·2 Traveling wave of voltage on a lossless transmission line.

3·3 Propagation constant and characteristic impedance

The incident waves of voltage and current *decrease* in magnitude and vary in phase as one goes toward the receiving end of the transmission line which has losses. The *propagation constant* γ is a measure of the phase shift and attenuation along the line.

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{3·2}$$

α is the attenuation per unit length of line and is called the *attenuation constant*. The attenuation constant may be expressed in decibels per unit length or *nepers* per unit length, where 1 neper equals 8.686 db.

β is the phase shift per unit length, called the *phase constant*, and is measured in radians per unit length.

ω is $2\pi f$ and is called the *angular frequency*.

$(R + j\omega L)$ is the complex series impedance per unit length of transmission line, as shown in Fig. 2·3.

$(G + j\omega C)$ is the complex shunt admittance per unit length of line as shown in Fig. 2·3.

Since there is attenuation associated with the transmission line, the actual characteristic impedance of a transmission line (Fig. 2·2) is given by

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{3·3}$$

It is evident that the characteristic impedance Z_0 of a lossy line is a pure resistance only if $RC = GL$. As explained in Chap. 2, most microwave lines have very small losses and Z_0 can be considered to be a pure resistance $Z_0 = \sqrt{L/C}$. Also, the attenuation constant is zero, and the propagation constant becomes $\gamma = j\omega\sqrt{LC}$.

The characteristic impedance of the two-wire transmission line ranges from 100 to 300 ohms according to

$$Z_0 = 276 \log \frac{2b}{a} \tag{3·4}$$

where a is the diameter of the conductors and b is the distance between centers of conductors.

The *phase constant* is defined as the rate of change of phase with distance for fixed values of time and is given by

$$\beta = \omega\sqrt{LC} = \frac{2\pi f}{v} = \frac{2\pi}{\lambda} \tag{3·5}$$

Thus β determines the wavelength and for this reason is sometimes called the *wavelength constant*. At the end of each unit length of line, the voltage and current will lag the voltage and current at the beginning of that unit length by an angle of β rad. In other words, in traveling a distance l, in either direction on the line, the phase is retarded by βl rad. As an example, at the point B in Fig. 3·2,

$$\beta l = \frac{2\pi l}{\lambda} = \frac{2\pi(0.5)}{0.5} = 360°$$

At C,

$$\beta l = \frac{2\pi l}{\lambda} = \frac{2\pi(1)}{0.5} = 720°$$

The *phase velocity* is the *velocity of a point* denoting the location of a definite phase of the periodic disturbance *in space*. Phase velocity in meters per second is given by

$$v_p = f\lambda$$

where f is the frequency in cycles per second and λ is the wavelength in meters.

The time delay t_d of a transmission line is the time it takes *a point* to travel the length of the line. If the phase velocity is independent of frequency, this time is the time it takes a pulse or signal to travel the length of the line.

$$t_d = \frac{l}{v_p} = \frac{\beta l}{\omega} = \frac{lT}{\lambda} \tag{3·6}$$

The time delay in terms of the period T is equal to the number of wavelengths in the line.

3·4 Standing waves on the lossless transmission line

The basic theory of reflections as set forth in Chap. 2 will be applied to the case of sinusoidal waves of voltage and current, *in phase*, traveling along a transmission line which is terminated in an impedance different than the characteristic impedance of the line. *Reflected waves of voltage and current exist owing to the impedance mismatch.* The instantaneous total voltage or total current at any point on the line is the sum of the incident and reflected voltages or currents at that point.

The behavior of the sinusoidal voltage wave on a line terminated in a short circuit is considered for simplicity of explanation. The diagrams in Fig. 3·3 represent the traveling wave of voltage at different time intervals. The dotted sine waves to the right of the short circuit in each diagram indicate the position and distance that the wave would have traveled in the absence of the short circuit. With the short circuit placed at X, the wave travels the same distance back toward the generator. In order to satisfy the boundary conditions, the voltage at the short circuit must be zero at all times. This is accomplished by a reflected voltage wave which is equal in magnitude and reversed in polarity (shown by the superimposed reflected wave and the resultant total voltage on the line, as indicated by the dark line). In Fig. 3·3*a*, a time t_0 is selected after the wave has been reflected from the short circuit. The total voltage is $2E_i$ at a distance of one-quarter of a wavelength back toward the generator, and the total voltage is zero at a distance of one-half wavelength from the short. This point is labeled (*b*).

Figure 3·3*b* indicates a time t_1 when the wave has traveled (at the speed of light) one-quarter of a wavelength to the right. At this instant the reflected and incident waves cancel at all points on the line, resulting in zero voltage along the line.

At time t_2 the wave has traveled to the right a total of one-half wavelength. The incident and reflected waves add to $2E_i$ in a negative direction at a distance of one-quarter wavelength from the short. Again it is noted that the zero voltage point is at b and also at a, which is one-half wavelength from b.

Figure 3·3*d* shows the instantaneous voltage on the line as a function of position at successive intervals of time starting at t_0 which is labeled (1). As the wave moves to the right, the instantaneous voltages labeled (2) and (3) indicate a decrease in total voltage as the wave progresses to the right. When the wave has reached time t_1, which is one-quarter of a wavelength as shown in Fig. 3·3*b*, the voltage is zero along the line as indicated (4). As the wave travels from t_1 to t_2, a distance of one-quarter of a wavelength, the instantaneous voltages reach a total negative value of $2E_i$ in successive time intervals indicated by (5), (6), and (7). The corresponding current magnitude and the relative positions of maxima and minima points are indicated in Fig. 3·3*e*.

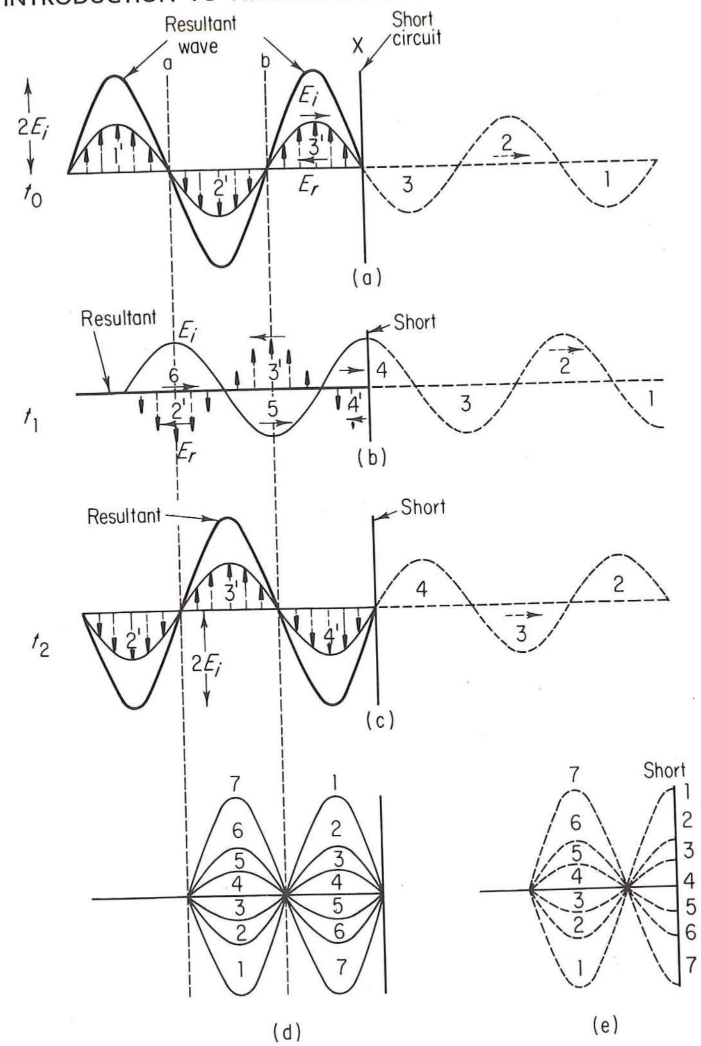


Fig. 3·3 Generation of standing waves on a shorted transmission line. Dotted lines to the right of the short circuit represent the distance the wave would have traveled in absence of the short. Dotted vectors represent the *reflected wave*. The heavy solid line represents the vector sum of the incident and reflected waves. (*d*) and (*e*) represent instantaneous voltages and currents at different intervals of time.

The total voltage pattern is called a *standing wave*. Standing waves exist as the result of two waves of the *same frequency* traveling in *opposite directions* on a transmission line.

If the signal source is assumed to have an internal impedance equal to Z_0, the steady-state conditions are reached when the reflected waves arrive at the input terminals. If the source is not matched, the re-reflections will result in an overall incident wave and an overall reflected wave with the same characteristics as previously discussed.

Figure 3·3*d* shows that the total voltage at any instant has a sine-wave distribution along the line with zero voltage at the short and zero points at half-wave intervals from the short circuit. The points of zero voltage are called voltage *nodes*, and the points of maximum voltage halfway between these nodes are called *antinodes*. By superimposing the current standing-wave diagram on the voltage standing-wave diagram, it can be seen that the current and voltage nodes are one-quarter of a wavelength apart and also that the antinode points are one-quarter of a wavelength apart.

At a distance of one-quarter of a wavelength from the short, the voltage is found to be $2E_i$, which is equivalent to an *open* circuit. Therefore, this same distribution would be obtained if an open circuit were placed $\lambda/4$ from the short. In this case, it is found that the zero points (nodes) are $\lambda/4$ from the open end of the line.

3·5 Standing-wave ratio

The *voltage-standing-wave ratio* is denoted by the symbol ρ and is defined as the ratio of the maximum voltage to the minimum voltage on a transmission line. This ratio is most frequently referred to as VSWR. The voltage-standing-wave ratio referred to in this text is as defined above. If the standing-wave ratio is measured in terms of the square of the voltage, the ratio is called the *power-standing-wave ratio* (PSWR) and is designated ρ^2.

$$\text{VSWR} = \rho = \frac{E_{\max}}{E_{\min}} = \frac{E_i + E_r}{E_i - E_r} = \frac{1 + \Gamma}{1 - \Gamma} \tag{3·7}$$

Also $$\text{VSWR} = \frac{I_{\max}}{I_{\min}}$$

If the equation for VSWR is solved for Γ, it is found that

$$\Gamma = \frac{\rho - 1}{\rho + 1} \tag{3·8}$$

If the transmission line is terminated in a short or open circuit, the reflected voltage E_r is equal to the incident voltage E_i. From the above equation, the reflection coefficient is 1.0, and the voltage-standing-wave ratio (VSWR) is infinite. If a matched termination is connected to the line, the reflected wave is zero, the reflection coefficient is zero, and the VSWR is 1.0.

3·6 Transmission line impedance

The *input impedance* Z_i of a transmission line is defined as the ratio of *total voltage* to *total current* at a point on the line looking toward the load.

$$Z_i = \frac{E_{\text{total}}}{I_{\text{total}}} \tag{3·9}$$

The incident and reflected voltage waves may be shown graphically as indicated in Fig. 3·4*b*. At the short circuit, E_r is equal to E_i and is reversed in polarity as shown. If the incident wave is rotated counterclockwise and the reflected wave is rotated clockwise as shown in the diagram, the resultant amplitude of $2E_i$ is the first $E_{\max}$ on the standing-wave diagram. Following this procedure of rotating the two waves, the complete standing-wave diagram can be plotted.

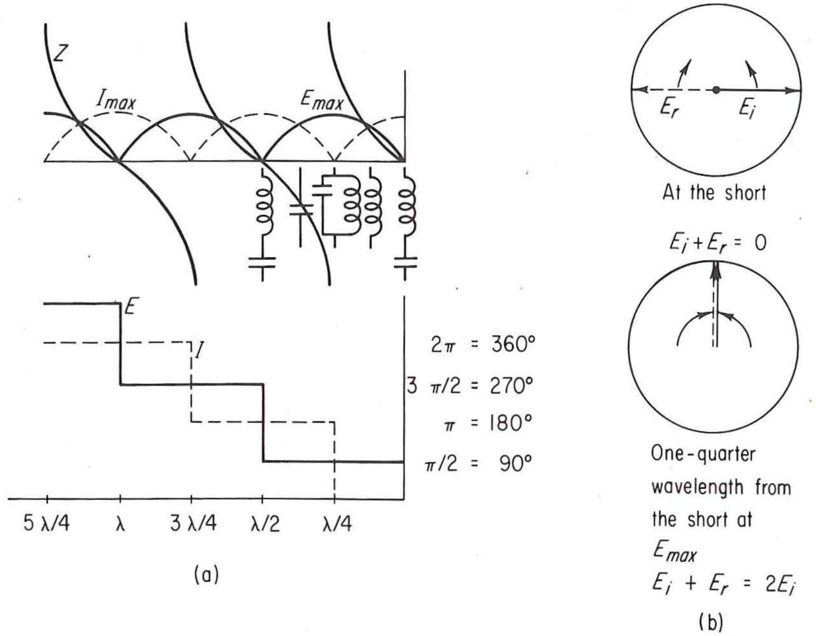


Fig. 3·4 Variations in impedance and the magnitude and phase of the voltage and current on a transmission line terminated in a short circuit.

During the first quarter wave of travel from the short toward the generator, the voltage leads the current by 90° and the transmission line input impedance is *inductive*, as shown on the diagram in Fig. 3·4*a*. The impedance is zero at the short circuit and increases, according to the tangent function, to positive infinity. At this point the impedance changes to a minus infinity value since here the current begins to lead the voltage, and the circuit becomes *capacitive*. It is noted that the voltage and current are always $\pm 90°$ apart when the incident and reflected voltages are equal. This condition exists when the transmission line is terminated in a pure open, short, inductance, or capacitance.

For the ideal dissipationless case, the input impedance of a short-circuited line is infinite for a line one-quarter of a wavelength long and is zero for a line one-half of a wavelength long. This is somewhat analogous to series and

parallel resonance in a dissipationless LC circuit, and the corresponding series and parallel points are shown on the diagram of Fig. 3·4*a*. The actual impedance at a short circuit is a low pure resistance. Also, the actual imput impedance of a low loss quarter-wave shorted line is not infinite but is a pure resistance in the range of 400 kilohms.

The equation for the input impedance of a transmission line is given by

$$Z_i = Z_0 \frac{Z_L + jZ_0 \tan \beta l}{Z_0 + jZ_L \tan \beta l} = Z_0 \frac{Z_L \cos \dfrac{2\pi l}{\lambda} + jZ_0 \sin \dfrac{2\pi l}{\lambda}}{Z_0 \cos \dfrac{2\pi l}{\lambda} + jZ_L \sin \dfrac{2\pi l}{\lambda}} \qquad (3·10)$$

3·7 Voltage and current relationships on the lossless transmission line

The properties of the voltage and current waves on the lossless transmission line are shown in Fig. 3·5 for the various terminating impedances. The angle of the reflection coefficient is denoted by ψ and represents the angle between the incident and reflected voltage waves. The angle of the reflection coefficient

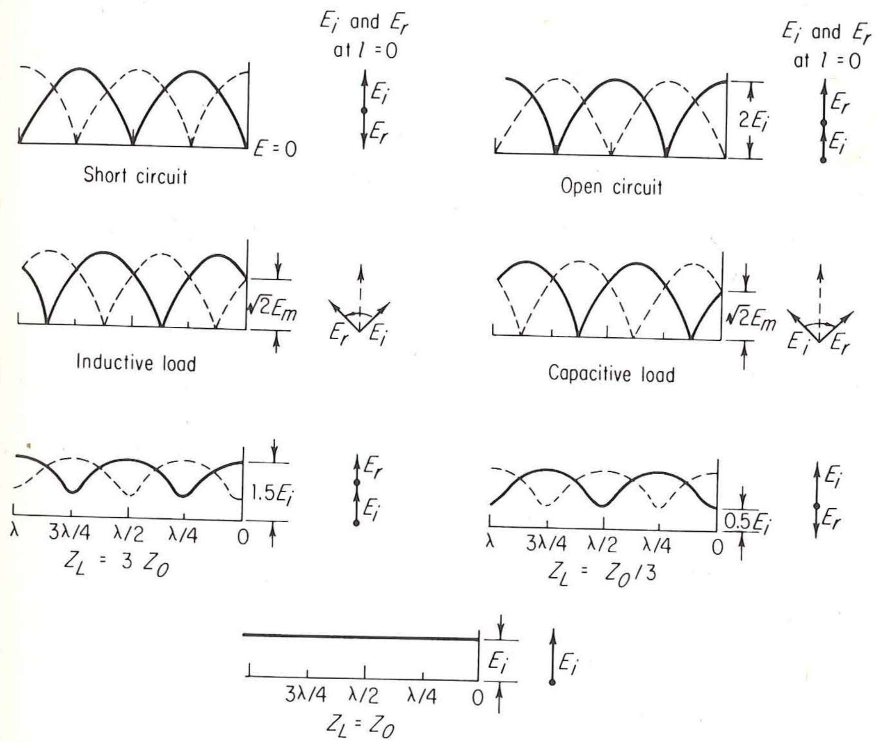


Fig. 3·5 Voltage and current distributions and relations on a lossless transmission line.

is zero when the incident and reflected voltage waves are in phase. Therefore, the angle of the reflection coefficient is always zero at E_{max} points on the standing wave.

Compare the following relationships and properties of voltage and current with the corresponding voltage and current distributions for the lossless line shown in Fig. 3·5.

Open-circuit Termination

1. The incident and reflected voltages are in phase at the open circuit and at one-half wavelength intervals from the open.
2. The angle of the reflection coefficient is zero at the open circuit and at intervals of one-half wavelength from the open.
3. The reflected current is equal in amplitude and 180° out of phase with the incident current at the open and at one-half wavelength intervals from the load.
4. The magnitude of the reflection coefficient is 1.0.
5. The VSWR is infinite.
6. The first voltage minimum along the line is located one-quarter of a wavelength from the open-circuit termination.
7. The first current minimum is located one-half wavelength from the open-circuit termination.

Short-circuit Termination

1. The incident and reflected currents are equal in amplitude and in phase at the short circuit and at one-half wavelength intervals along the line from the short circuit.
2. The VSWR is infinite.
3. The reflection coefficient is 1.0, and the angle of the reflection coefficient is 180°.
4. The first voltage minimum is located one-half wavelength along the line from the short circuit.
5. The first current minimum is located one-quarter wavelength from the short.
6. The input impedance of the line is a function of the line length.

Matched Termination

1. The reflected wave is zero.
2. There are no standing waves.
3. The VSWR is 1.0:1.
4. The reflection coefficient is zero.
5. The input impedance of the line is independent of the length of the line.

Pure Resistance Termination Greater Than Z_0

1. The incident and reflected waves of voltage are in phase at the load and at one-half wavelength intervals from the load.
2. A voltage maximum exists at the load and at one-half wavelength intervals from the load.
3. The angle of the reflection coefficient is zero at the load and at one-half wavelength intervals from the load.
4. The amplitude of the reflected wave, the magnitude of the reflection coefficient, and the VSWR depend upon the values of Z_0 and Z_L.
5. The reflected current is 180° out of phase with the incident current at the load and at one-half wavelength intervals from the load.
6. The wavelength locations of voltage and current maxima and minima follow the same pattern as for the open circuit except for amplitude variations.

Pure Resistance Termination Less Than Z_0

1. The incident and reflected currents are in phase at the load and at one-half wavelength intervals from the load.
2. The incident and reflected voltages are 180° out of phase at the load and at one-half wavelength intervals from the load.
3. A voltage minimum is located at the termination.

Pure Reactance Termination

1. The incident and reflected voltages are out of phase except at E_{max} and E_{min} points where they are either in phase or 180° out of phase, respectively.
2. The VSWR is infinite.
3. The reflection coefficient is 1.0.

3·8 Transmission line resistance

At high frequencies, current does not penetrate deeply into metal. When a conductor is carrying a current which is uniformly distributed throughout the cross section and this current begins to increase in value, a current path near the axis is encircled by more flux than is a current path near the surface. Therefore, there is a larger inductance associated with the internal path, and an easier path for current will lie toward the surface of the conductor. This phenomenon is called *skin effect*. An increase in current in the outer portions of the conductor will cause a greater heat loss than the reduction due to a decrease of current near the center. This redistribution of current results in an increase of total conductor resistance.

The current density decreases exponentially with distance beneath the surface, and the phase changes linearly with distance. At a depth δ, the skin depth, the current density decreases to $1/e$ times its surface value. The current at this depth lags the surface current by 1 rad. If the curvature of the conductor is large compared to the skin depth, the resistance of the conductor may be calculated by assuming that the current density is uniform and confined to a surface of thickness

$$\delta = \sqrt{\frac{\text{resistivity}}{\pi \mu_0 f}}$$

There is some current within the metal at a depth greater than the skin depth.

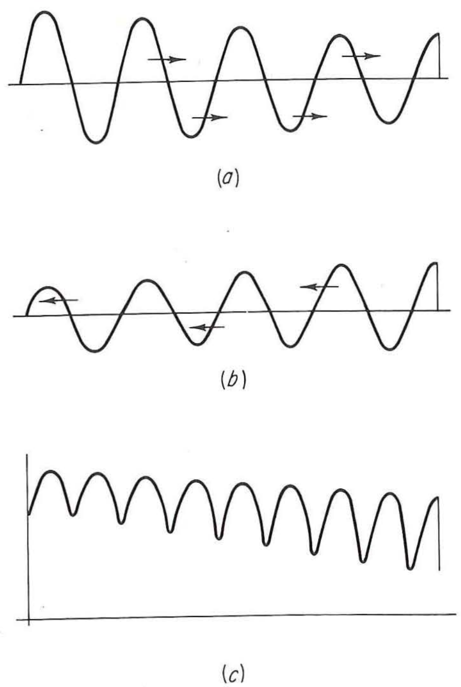


Fig. 3·6 Effects of losses on the incident and reflected voltage and the resulting standing-wave pattern.

For example, at a depth of three times the skin depth, the current density is 5 per cent of the value at the surface. The total loss, however, is the same as it would be if the current were uniformly distributed over a surface layer of depth equal to the skin depth.

Figure 3·6 represents a traveling wave of voltage on a lossy transmission line and the corresponding reflection from a termination greater than Z_0. The incident and reflected waves are attenuated as shown. The following facts concerning these waves are of special significance.

1. At the load, the ratio of the reflected to the incident wave is a maximum and indicates that the reflection coefficient and VSWR are greatest *at the mismatch.*

2. As the point of observation is moved from the mismatch toward the generator, the VSWR and reflection coefficient decrease. Therefore, the maxima and minima of the standing waves of voltage and current are not constant on the lossy transmission line. A detailed discussion of the properties of the lossy transmission line is given in the chapter dealing with the radial scaled parameters of the Smith chart.

PROBLEMS

3·1 The frequency of the signal applied to a two-wire transmission line is 3 Gc.
a. What is the wavelength if the dielectric in the medium is air?
b. What is the wavelength if the dielectric constant of the medium is 3.6?

3·2 A signal is applied to a quarter-wave section of transmission line which is terminated in a short circuit. Draw a diagram illustrating your answer to each of the following questions.
a. What is the theoretical value of input impedance?
b. What is Z_i if the frequency is lowered and the line length is not changed?
c. What is Z_i of the line if the section of line is shortened but the frequency of the applied wave remains fixed?
d. What is Z_i if the frequency is increased slightly and the line length is not changed?

3·3 Repeat Prob. 3·2 for a quarter-wave section of line terminated in an open circuit.

3·4 What is the characteristic impedance of a two-wire line if the distance between centers of the conductors is 0.8 cm and the diameter of the conductor is 0.15 cm?

3·5 Plot a graph of voltage-standing-wave ratio versus reflection coefficient in steps of 0.05 from zero to 0.2 and in steps of 0.1 from 0.2 to 1.0.

3·6 A transmission line has a characteristic impedance of 50 ohms and is terminated in an open circuit. Calculate the input impedance at the following wavelength intervals and compare your results with the diagrams in Fig. 3·5.
a. One-quarter wavelength from the load
b. One-half wavelength from the load
c. One-eighth wavelength from the load

3·7 Repeat Prob. 3·6 if the load impedance is a short circuit.

3·8 Repeat Prob. 3·6 if the load impedance is equal to the characteristic impedance of the line.

3·9 Explain the relationships of the various transmission line parameters for a lossy transmission line terminated in a mismatched load.

3·10 The resistivity of copper is 1.724×10^{-6} ohms per cm.

a. Calculate the skin effect at frequencies of 100 Mc, 1 Gc, and 3 Gc.

b. Silver has a conductivity of 6.17×10^7 mhos per m and aluminum has a conductivity of 3.72×10^7 mhos per m. Repeat part *a*.

3·11 If the incident voltage is 100 volts and the reflected voltage is 50 volts, calculate the reflected power in per cent, the voltage and power reflection coefficients, the voltage- and power-standing-wave ratios, and the transmitted power in per cent.

3·12 The characteristic impedance of a transmission line is 100 ohms and the load impedance is 200 ohms. If the incident voltage is 50 volts, calculate the reflected power in per cent, the VSWR, and the voltage reflection coefficient.

CHAPTER

4

GRAPHICAL REPRESENTATION OF TRANSMISSION LINE CHARACTERISTICS

Introduction. A graphical treatment of the impedance properties of the lossless transmission line is given in this chapter. The presentation consists of a description of incident and reflected waves in terms of complex numbers and is a step-by-step method of approach to the Smith chart.[1] The Smith chart is the most useful of the many graphical aids which have been presented for use in performing transmission line computations.

4·1 Application of complex exponentials

When a voltage is written as a complex exponential, a quantity is used which can be analyzed into real and imaginary parts. The real part is a cosine function, and the imaginary part is a sine function multiplied by $\sqrt{-1}$, written j. The voltage represented as a rotating vector may be written in the exponential form $Ee^{j\theta}$ where $e^{j\theta} = \cos\theta + j\sin\theta$.

In the study of transmission lines, complex numbers are frequently used to represent rms values and phase angles of sinusoidal functions of time.

If the instantaneous reference voltage at the input terminals of a line is represented by the complex number $E = (E/\sqrt{2})e^{j0°}$, a point 90° from the input terminals is represented by $(E/\sqrt{2})e^{j90°}$. The wave has the same amplitude but is delayed 90° in phase.

In the following analysis, the complex number notation is used to describe positions on a transmission line at a distance measured from the *receiving end* of the line.

Figure 4·1 is a circuit diagram of a shorted transmission line with a point P located a distance l from the short-circuit load termination. E_{iL} is defined as the incident voltage *at the load*, and E_{rL} is the reflected voltage *at the load*. The incident voltage E_i and the reflected voltage E_r are to be represented at P, measured from the load, in terms of the incident and reflected voltages at the load.

The incident wave is delayed in phase by βl in traveling from P to the short circuit. The incident wave at P can be expressed as

$$E_i = E_{iL} e^{j\beta l} \tag{4·1}$$

The reflected wave is also delayed in phase by βl in traveling from the

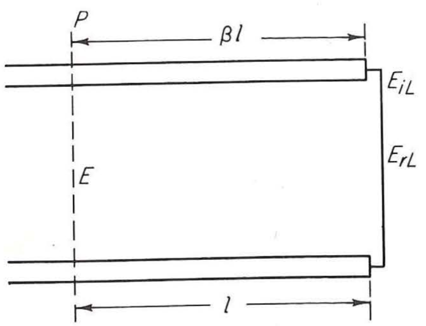


Fig. 4·1 The incident wave is delayed βl degrees in traveling from P to the short circuit while the reflected wave is delayed $-\beta l$ in traveling from the short to the point P.

short to the point P and is expressed as

$$E_r = E_{rL} e^{-j\beta l} \tag{4·2}$$

$$e^{j\beta l} = 1\underline{/\beta l} = \cos \beta l + j \sin \beta l$$

$$e^{-j\beta l} = 1\underline{/-\beta l} = \cos \beta l - j \sin \beta l$$

Since E_{iL} equals E_{rL} at the short circuit, the total voltage at P is

$$E_t = E_{iL} e^{j\beta l} + E_{rL} e^{-j\beta l} = E_{iL}(e^{j\beta l} - e^{-j\beta l})$$

$$E_t = j2E_{iL} \sin \beta l \tag{4·3}$$

This is true since the $\sin \beta l = (e^{j\beta l} - e^{-j\beta l})/2j$. The total current can be derived in the same manner and is found to be

$$I_t = 2\frac{E_{iL}}{Z_0} \cos \beta l \tag{4·4}$$

The input impedance is the ratio of these two complex numbers and is found to be

$$Z_i = \frac{E_t}{I_t} = jZ_0 \tan \beta l \tag{4·5}$$

The voltage and current for the open circuit can be derived in the same manner as for the shorted line, and the results are

$$E_t = 2E_{iL} \cos \beta l \qquad I_t = j2\frac{E_{iL}}{Z_0} \sin \beta l \qquad Z_i = jZ_0 \cot \beta l \tag{4·6}$$

Compare Eqs. (4·3) to (4·5) with Fig. 3·4, which shows the total voltage and current relationships.

4·2 Graphical representation of propagation characteristics

The circle diagram in Fig. 3·4 served its purpose for indicating the reflection coefficient and the plot of standing waves when the incident and reflected waves were the same amplitude. If the reflected wave is smaller than the incident wave, the resulting elliptical diagram would not be so simple or informative. In Fig. 3·4 the reflected and incident waves are rotated at the same speed but in opposite directions. A new diagram is to be constructed using E_i and I_i as unit vectors. These vectors are made to remain stationary while the vectors representing reflected voltage and current waves are rotated at *twice* the previous rate. The same results will be obtained with this diagram as were obtained with the diagram of Fig. 3·4.

The vector representation is illustrated in Fig. 4·2. The circle diagram represents the voltage on the transmission line terminated in a *resistive load* which has a reflection coefficient of 0.5. The unit vector representing the incident voltage wave remains fixed, and the vector representing the reflected wave is rotated *clockwise* an angle $-2\beta l$ or 45°, as shown at b. The reflected wave on the diagram is rotated 90° at c, and this point corresponds to 45° travel on the transmission line, as indicated at C on the standing-wave diagram. When the reflected wave has traveled 180° on the circle diagram as indicated at e, the actual distance traveled on the line is 90°, as indicated at E on the standing-wave diagram.

The diagram can be described by the following mathematical presentation. The equation for total voltage on the transmission line is divided by the incident voltage and modified to obtain the following results:

$$\frac{E}{E_i} = 1 + \frac{E_r}{E_i} \tag{4·7}$$

$$= 1 + \frac{E_{rL} e^{-j\beta l}}{E_{iL} e^{j\beta l}} = 1 + |\Gamma|\, e^{-j2\beta l}$$

$$= 1 + |\Gamma|\, \underline{/-2\beta l}$$

$$= 1 + |\Gamma|\, \underline{/\psi - 2\beta l} \tag{4·8}$$

$$\phi = \beta l - \psi/2$$

$$-2\phi = \psi - 2\beta l$$

$$E/E_i = 1 + |\Gamma|\, \underline{/-2\phi}$$

$$\frac{I}{I_i} = 1 - |\Gamma|\, \underline{/\psi - 2\beta l} \tag{4.9}$$

ψ is the angle of the reflection coefficient, and -2ϕ is the angle measured from the point where ψ equals zero. The angle of the reflection coefficient ψ is zero degrees when the load is a pure resistance and the incident and reflected voltages are *in phase* as shown at A.

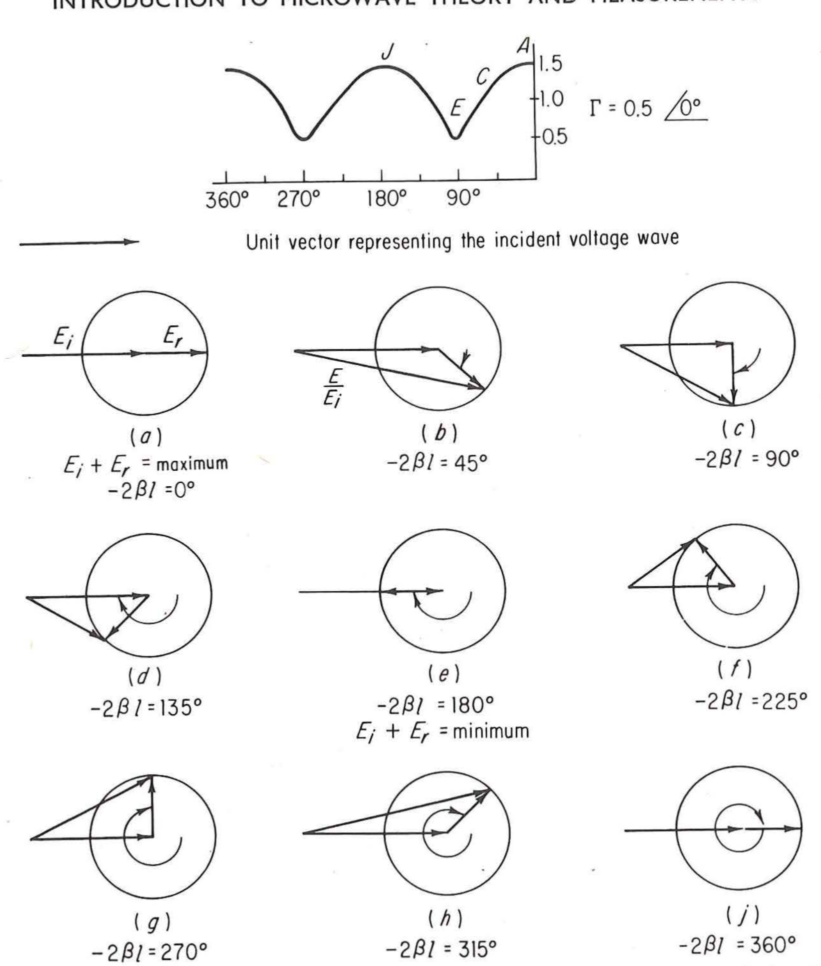


Fig. 4·2 Circle-diagram representation of the voltage on a transmission line terminated in a *resistive* load. The reflection coefficient is 0.5, which corresponds to a VSWR of 3.0:1.

A vector diagram representing Eqs. (4·8) and (4·9) is shown in Fig. 4·3 for a complex load impedance which has a reflection coefficient of 0.5 at an angle of 60°.

The vectors labeled E_i and I_i are the unit vectors representing the incident voltage and current and represent the first term on the right-hand side of Eqs. (4·8) and (4·9). The vector labeled E/E_i represents the total voltage on the line and is the left-hand term of Eq. (4·8). The vectors labeled $\Gamma\underline{/-2\beta l}$ and $-\Gamma\underline{/-2\beta l}$ are the remaining terms of Eqs. (4·8) and (4·9) and represent the reflected voltage and reflected current, respectively. I/I_i is the total current represented by the first term of Eq. (4·9).

The reflected voltage and current waves are drawn 180° out of phase on the circle diagram since all angles on the diagram are *twice* the angles on the transmission line.

Positive angles of reflection coefficient are measured *counterclockwise* from A, and negative angles are measured in a clockwise direction from A.

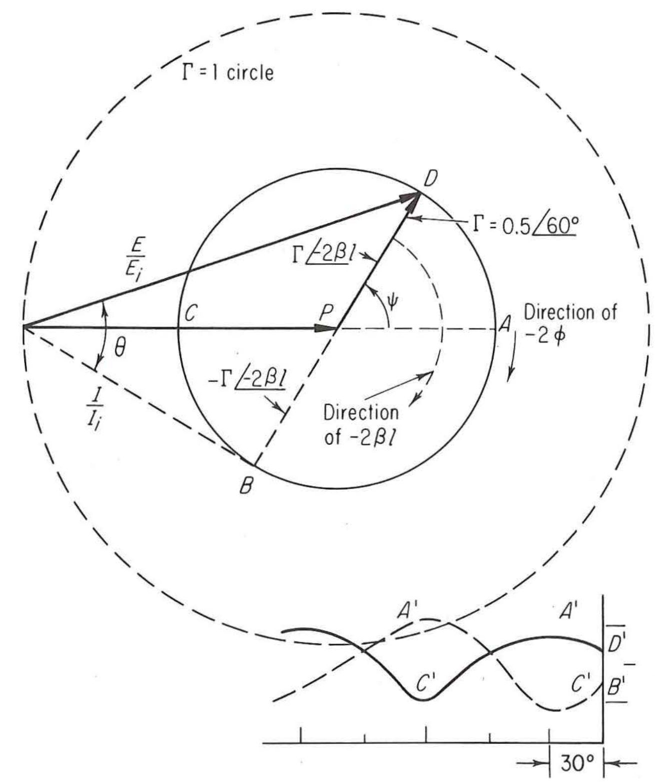


Fig. 4·3 Vector representation of voltage and current on a transmission line terminated in a complex impedance.

For this particular load, the angle of ψ is 60° on the circle diagram and corresponds to an angle of $2\beta l$ measured CCW from A. The corresponding point on the standing-wave diagram A' is found to be 30°, which is the distance of βl on the transmission line. In comparing the distance from A to C on the circle diagram to A' and C' on the standing-wave diagram, it is noted that a 90° distance on the transmission line is represented by 180° rotation of the reflected voltage and current vectors on the circle diagram. The point D on the diagram represents the voltage at the load, and the angle of the reflection coefficient is 60° as shown.

Traveling from D' to A' on the transmission line *toward the generator* corresponds to the distance D to A on the circle diagram. Therefore, clockwise rotation of the diagram represents moving along the transmission line *toward the generator* and *counterclockwise* rotation represents movement along the line *toward the load.*

The $\Gamma = 1$ circle is obtained when the incident and reflected waves are equal in amplitude. This circle is obtained when the line is terminated in a *short circuit*, *open circuit*, or *pure reactance*. The phase angle between E_t and I_t in this case is always $\pm 90°$. A family of reflection coefficient circles can be constructed with their centers at P. The reflection coefficient is zero at P and represents the matched condition. At this point the load impedance is equal to the characteristic impedance of the transmission line, and no standing waves exist on the line.

The angle θ represents the angle between the total current and total voltage on the transmission line. When the reflection coefficient approaches zero, the total voltage and total current approach the in-phase condition, which exists when only incident voltage and incident current appear on the line.

The transmission line characteristics which can be described on this diagram are referred to as the *propagation characteristics*. In addition to the family of reflection coefficient circles, a family of VSWR circles can be drawn on the diagram. The VSWR is *unity* at the center of the circle and *infinite* at the outside rim of the circle. Since 180° on the diagram represents 90° or one-quarter of a wavelength on the transmission line, it is convenient to label the 180° rotation in terms of fractional wavelength. In this case, one-half the distance around the diagram represents 0.25 wavelength. In this way, radii of constant phase in terms of fractional wavelengths (l/λ) are obtained. It was shown in Fig. 3·6 that the VSWR on a lossy line decreases as one travels toward the generator. This indicates that the radial distance on this diagram can also indicate attenuation along the line. A complete discussion of this characteristic will be covered in the discussion of the Smith chart.

4·3 Impedance and admittance coordinates

It is not necessary to understand the mathematical derivation in order to use the circle diagram. The approach to the *propagation grid* of the Smith chart as given in Sec. 4·2 is sufficient. Also, the mathematical proof that the loci of constant resistance and reactance are circles and the mathematical proof of the location of the centers of those circles will not be covered. However, a description of the construction of those circles will be presented.

The impedance or admittance at a point on a uniform lossless transmission line is defined by the amplitude and phase angle of the reflection coefficient. As previously shown, the angle of the reflection coefficient is directly related to the impedance and is indicated on the scale around the rim of the circle diagram.

The chart is constructed on a per unit basis. *The normalized or per unit load impedance is obtained by dividing the load impedance by the characteristic impedance* (Z_L/Z_0). In the case of complex impedances the *per unit* impedance is

$$z = \frac{R}{Z_0} \pm \frac{jX}{Z_0}$$

The construction line for the constant resistance and constant reactance circles is shown in Fig. 4·4 tangent to the circle at the point R/Z_0 = infinity,

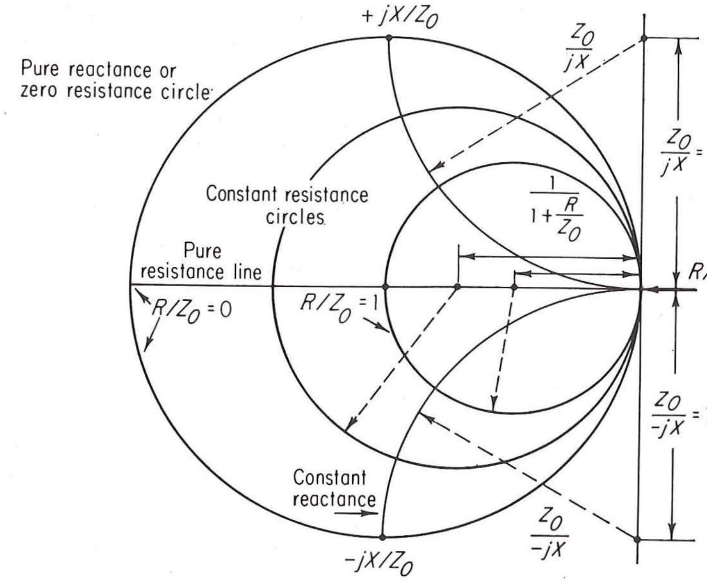


Fig. 4·4 Construction of constant resistance circles R/Z_0 and constant reactance curves $\pm jX/Z_0$.

located at the right-hand side of the diagram. The centers of the positive and negative reactance circles are located at a distance of Z_0/jX and $Z_0/-jX$ measured from the point (R/Z_0 = infinity), as shown on the diagram. The centers of the resistance circles are located on the pure resistance line R/Z_0 with the radii calculated from the formula shown on the diagram. Again it is important to note that the outside circle is the *pure reactance* circle and that the center line on the diagram represents *pure resistance.*

This diagram is intimately related to the diagram of Fig. 4·3. For example, the outside circle of Fig. 4·3 was generated when the incident and reflected voltages were equal in magnitude, a condition that exists when the line is terminated in a short, open, or pure reactance. The outside circle of Fig. 4·4 is also the pure reactance circle (zero resistance). The centers of the diagrams correspond since each represents the point where the line is terminated in its characteristic impedance and there are no reflections on the line. The extreme left in Fig. 4·3 is the zero voltage point, and the same point on Fig. 4·4 is R/Z_0, the zero resistance point which must also be the zero voltage point. The right side of Fig. 4·3 shows that E_i and E_r are in phase and a maximum when E_r equals E_i at an open circuit. This point corresponds to $R/Z_0 =$ infinity on the diagram of Fig. 4·4.

4·4 The Smith chart

By superimposing the propagation grid of Sec. 4·2, and the impedance or admittance coordinate grid of Sec. 4·3, the complete Smith chart of Fig. 4·5 is obtained. Several of the electrical parameters are not shown on the diagram proper because it would be quite confusing if all the properties of each coordinate system were included on one chart. Therefore, the parameters left off the chart proper are given in the form of radial-scaled parameters as shown in Fig. 4·5.

VSWR (*Voltage-Standing-wave Ratio*). The standing-wave ratio has been defined as the ratio of the maximum to minimum voltage or the ratio of maximum to minimum current on a transmission line.

$$\rho = \frac{E_{\max}}{E_{\min}} \quad \text{or} \quad \frac{I_{\max}}{I_{\min}} \tag{4·10}$$

From the above equation

$$\rho = \frac{I_{\max}}{I_{\min}} = \frac{I_i + I_r}{I_{\min}} = \frac{E_i/Z_0 + E_r/Z_0}{I_{\min}} = \frac{E_{\max}/Z_0}{I_{\min}} = \frac{Z_{\max}}{Z_0} \tag{4·11}$$

from which $Z_{\max} = Z_0\rho$ (a pure resistance). Also,

$$\rho = \frac{E_{\max}}{E_{\min}} = \frac{(I_i + I_r)Z_0}{E_{\min}} = \frac{Z_0}{Z_{\min}} \tag{4·12}$$

and $Z_{\min} = Z_0/\rho$ (a pure resistance).

The normalized value of $Z_{\max} = R + j0$ is the VSWR and is read on the right-hand axis of the chart as the normalized value of the constant resistance circles. As an example, a VSWR of 5.0:1 is read at point A, Fig. 4·5. The impedance at this point is 250 ohms for Z_0 of 50 ohms. The corresponding minimum impedance is indicated at B and is 10 ohms. *Note that the reciprocal of the normalized impedance is located diametrically opposite the given impedance value on the same VSWR circle.*

The VSWR circle is a circle with a radius equal to ρ drawn with the center at 1.0 on the chart as indicated in Fig. 4·5. For a normalized load impedance of $1 + j2$ located at C, the VSWR circle is drawn with a radius PC centered at P. The resulting VSWR is read as 5.9 where the VSWR circle crosses the $Z_{\max}$ axis at D. A load of $35 + j20$ ohms normalized on a 50-ohm line is $0.7 + j0.4$ located at E. The VSWR circle is drawn with the radius PE, and a VSWR of 1.8 is read where the circle crosses the R/Z_0 axis.

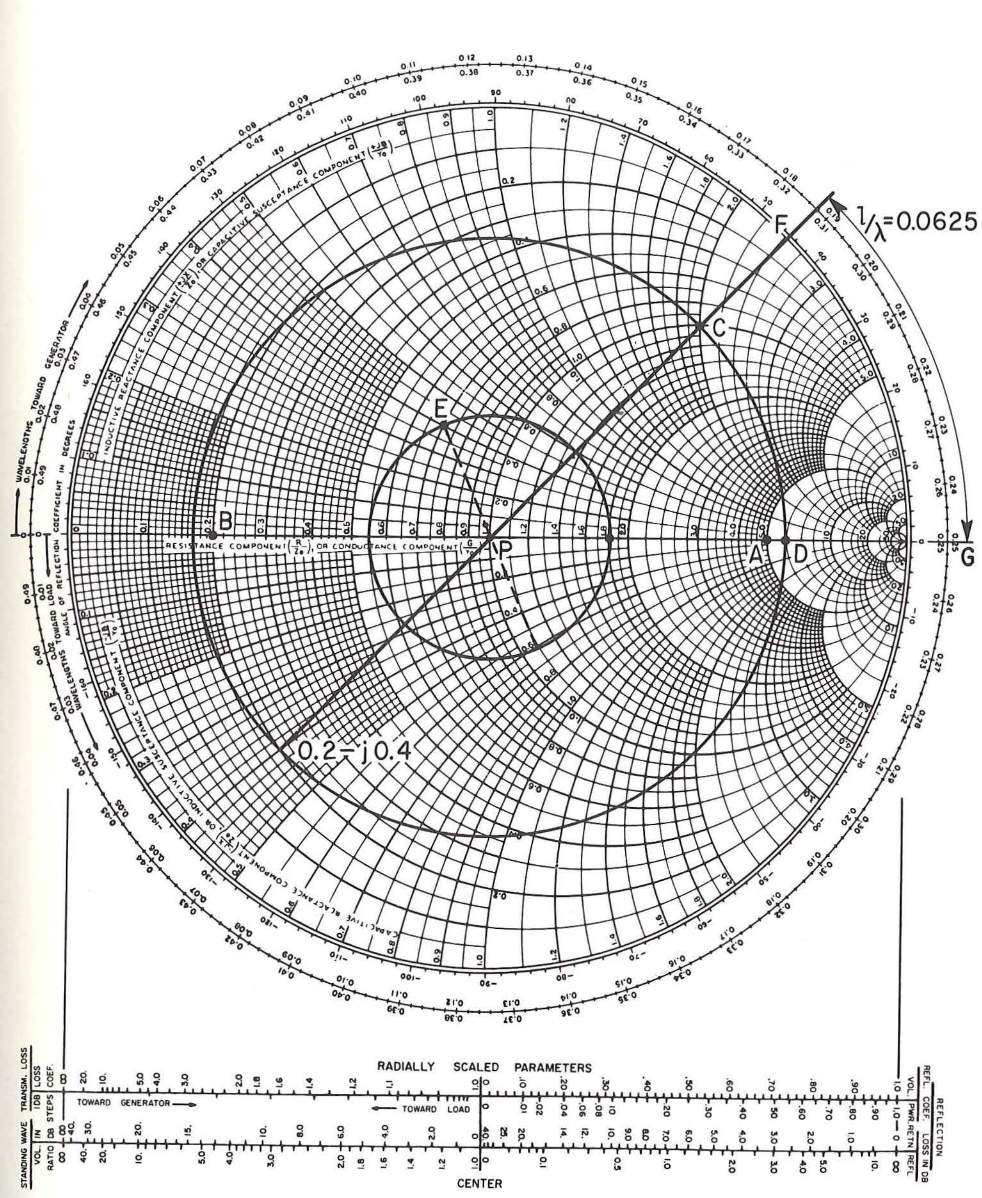


Fig. 4·5 Smith chart. (From Irving L. Kosow (ed.), "Microwave Theory and Measurements," by the Engineering Staff of the Microwave Division, Hewlett-Packard Company, © 1962, by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.)

Impedance and Admittance. Simplified schematic representations of the conventional forms of impedance and admittance are shown in Fig. 4·6. The impedance representation is a *series* one, as indicated at *a*, and the admittance is a *shunt* case, as indicated at *b*. The equation for impedance is

$$Z = R \pm jX \tag{4·13}$$

where R represents the resistive component and X represents the reactive

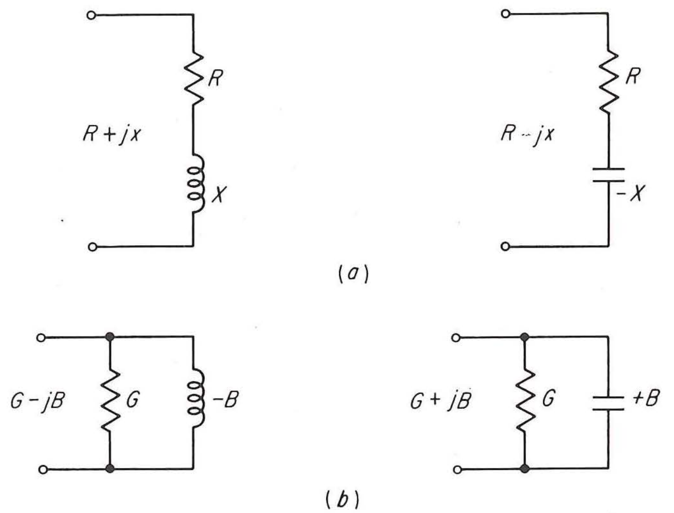


Fig. 4·6 Simplified schematic representation of the conventional forms of impedance and admittance.

component. The term j is the $\sqrt{-1}$ and indicates that R and X are perpendicular vectors and that the magnitude of Z is

$$Z = \sqrt{R^2 + X^2}$$

Admittance is expressed by

$$Y = G \pm jB \tag{4·14}$$

where G represents the conductance and B represents the susceptance. G is real and B is imaginary.

$$G = \frac{R}{R^2 + X^2} \qquad \text{and} \qquad B = \frac{X}{R^2 + X^2}$$

The chart provides a means of converting from impedance to admittance, and this conversion is complete even to the point of specifying resistance R in terms of conductance G and specifying reactance X in terms of susceptance B. This conversion is accomplished on the Smith chart by moving to a point diametrically opposite the known impedance or admittance and reading the

normalized value of the desired admittance or impedance. It was pointed out in a previous section that the *normalized resistance* value of 5.0 in Fig. 4·5 could be used to find the reciprocal of the impedance by obtaining the value of normalized impedance on the pure resistance line at a distance equal to PA. The located point was determined as 0.2 at B. This principle will be applied to normalized values of complex impedance and admittance since it is known that

$$Z = \frac{1}{Y} \qquad \text{and} \qquad Y = \frac{1}{Z}$$

Let r and x represent the normalized resistance and reactance, respectively. Then

$$\frac{Y}{Y_0} = \frac{1}{Z/Z_0} = \frac{1}{r + jx} = \frac{1}{r + jx}\,\frac{r - jx}{r - jx} = \frac{r - jx}{r^2 - jrx + jrx - (jx)^2}$$

$$\frac{Y}{Y_0} = \frac{r - jx}{r^2 + x^2} = \frac{r}{r^2 + x^2} - \frac{x}{r^2 + x^2}$$

As an example, assume that the impedance $Z = 50 + j100$ ohms. The normalized value becomes $1 + j2$. Substituting these values in the equation

$$\frac{Y}{Y_0} = \frac{1}{1 + (2)^2} - \frac{2}{1 + (2)^2} = 0.2 - j0.4$$

Compare the above results to the plotted points on the Smith chart of Fig. 4·5. The normalized impedance of $(1 + j2)$ is plotted at the point C, and the VSWR circle is drawn as shown. A straight line is drawn from the point C through the center P to the VSWR circle diametrically opposite, and the point of intersection with the VSWR circle is at $(0.2 - j0.4)$. This agrees with the calculated value. Therefore, the normalized admittance is read on the diagram at a point on the VSWR circle diametrically opposite the *normalized impedance* value for a lossless line. Also, the normalized impedance can be found by the same procedure when the normalized admittance is known.

The reciprocal properties at quarter-wave intervals are applied in quarter-wave impedance-matching problems. Suppose that it is desired to connect two lines which have characteristic impedance values Z_{01} and Z_{02}. This can be accomplished with a quarter-wave line as indicated by the schematic diagram of Fig. 4·7. Neglecting discontinuity capacitances at the steps

$$\frac{Z_{02}}{Z_0} = \frac{Z_0}{Z_{01}} \qquad \text{or} \qquad Z_{02} = \frac{Z_0^2}{Z_{01}}$$

From the above equation, the characteristic impedance of a quarter-wave line used to match two lines which have different characteristic impedances is

$$Z_0 = \sqrt{Z_{01}Z_{02}} \tag{4·15}$$

A numerical example is shown in Fig. 4·7*b*.

Line Length. The wavelength scale around the rim of the Smith chart calculator is linear. The values on the chart are fractional wavelengths l/λ. Fractional wavelengths on the chart correspond to fractional wavelengths on the transmission line. Wavelength measurements can be started from any point radially in line with any known impedance point on the coordinates and can be measured in either direction, clockwise toward the generator or counterclockwise toward the load.

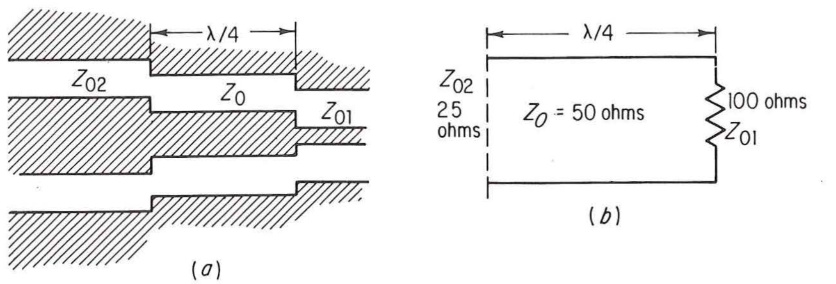


Fig. 4·7 A quarter-wavelength section of line with a characteristic impedance Z_0 used to connect two lines which have different values of characteristic impedance.

The radius vector *PC* of Fig. 4·5 is extended to *F*. The fractional wavelength readings are approximately 0.1875 on the outside scale and 0.3125 on the inside scale. The distance from *F* to *G* is 0.0625. Assuming an operating frequency of 3 Gc, the wavelength is 10 cm, and the distance on the transmission line from the load at *C* to E_{max} at *D* is 0.625 cm.

4·5 The decibel

The *decibel*, abbreviated db, is one-tenth of the international transmission unit known as the *bel*. The origin of the bel is the logarithm to the base 10 of the power ratio. The logarithm to the base 10 is the *common logarithm*. It is that power to which the number 10 must be raised in order to equal the given number. The number 10 is raised to the second power, or *squared*, in order to get 100. Therefore the log of 100 is 2.

The decibel is expressed mathematically by the equation

$$\text{db} = 10 \log \frac{P_2 \text{ (larger power)}}{P_1 \text{ (smaller power)}} \tag{4·16}$$

Throughout this text, the abbreviation log refers to the logarithm to the base 10 unless otherwise specified.

The complicated negative characteristics of the logarithm of the ratio can be avoided by always placing the larger power value in the numerator of the equation. In each problem it is known whether there is a gain or loss.

If the resistance level is the same at the points where both power levels are measured, the relative currents or voltages are expressed in decibels as

$$\text{db} = 20 \log R \tag{4·17}$$

where *R* is the ratio of voltages or currents.

The relationships of power ratio, decibels, and powers of 10 are given in Table 4·1. The table can be used to obtain approximate decibel or ratio values. By comparing the decibel column with the power of 10 column, it is noted that the number corresponding to the power of 10 also appears in the decibel column. The ratio of 1,000 is 10^3 or 30 db.

Table 4·1 Relationship of ratios to decibels

Power ratio	Power of 10	db	Power referred to 0 dbm = 1 mw
1	10^0	0 = 10 log 1	1 mw
1.259	$10^{0.1}$	1 = 10 log 1.259	1.259 mw
10	10^1	10 = 10 log 10	10 mw
100	10^2	20 = 10 log 100	100 mw
1,000	10^3	30 = 10 log 1,000	1 watt
10,000	10^4	40 = 10 log 10,000	10 watts
100,000	10^5	50 = 10 log 100,000	100 watts
0.1	10^{-1}	−10 = 10 log 0.1	100 μw
0.01	10^{-2}	−20 = 10 log 0.01	10 μw
0.001	10^{-3}	−30 = 10 log 0.001	1 μw

If it is desired to find the ratio corresponding to 27 db,

$$27 \text{ db} = 10 \log R$$

$$\frac{27}{10} = 2.7 = \log R$$

The ratio is found by taking the antilog of 2.7. The 2 represents the characteristic, and 0.7 is the mantissa. The number corresponding to the mantissa of 0.7 is obtained from a table of logarithms or from a slide rule. The number 2 indicates the number of zeros to add. From the table of mantissas, the number is 5.01. Two zeros are added to obtain the ratio of 501. From the table it is noted that the ratio must be between 100 and 1,000 or 20 db + *x* db.

The decibel value corresponding to a power ratio of 1,920 is db = 10 log 1,920. From Table 4·1 it is noted that the decibel value is 30 db + *x* db. The decibel value to be added is the mantissa of 1,920 multiplied by 10 or 10(0.283). The decibel value is 30 + 2.83 = 32.83 db.

The use of log tables can be avoided in practical applications where exact values of the power ratio are not required. A power ratio of 2 corresponds to 3.01 db. If the power ratio is expressed as 2^n,

$$\text{db} = n(3.01) \tag{4·18}$$

or

$$n = \frac{\text{db}}{3.01} \tag{4·19}$$

These equations can be used in conjunction with Table 4·1 to obtain any decibel value. For simplicity of calculations, the denominator in Eq. (4·19) is considered to be 3.0, that is, a change in power of ½ or 2:1 corresponds to 3 db.

If 3, 6, and 9 db values are substituted in Eqs. (4·18) and (4·19) and the power is expressed as 2^n, the following results are obtained:

$$3\text{ db} = 2^n = 2 \tag{4·20a}$$

$$6\text{ db} = 2^n = 4 \tag{4·20b}$$

$$9\text{ db} = 2^n = 8 \tag{4·20c}$$

This technique is based on the fact that 3, 6, and/or 9 db can be added or subtracted (in some combination) to the decibel values in Table 4·1 to obtain any decibel value.

Example

1. 17 db = 20 db − 3 db
   The ratio of 20 db is 100 (Table 4·1), and the ratio of 3 db expressed as 2^n is 2 [Eq. (4·20*a*)]. Therefore, 100/2 = 50.
2. 36 db = 30 db + 6 db = 1,000 (Table 4·1) × 4 [Eq. (4·20*b*)] = 1,000(4) = 4,000.
3. 25 db = 10 db + 9 db + 6 db = 10(8)(4) = 320.
4. 5 db = 30 db − 9 db − 10 db − 6 db = 1,000/(8)(10)(4) = 3.16.

An alternate unit called the *neper* is defined in terms of the logarithm to the base *e*. $e = 2.718$.

$$1\text{ neper} = 8.686\text{ db}$$
$$1\text{ db} = 0.1151\text{ neper}$$

The decibel is not a unit of power. The unit of power in our exponential or logarithmic system of numbers is represented by dbm, where the *m* is the *unit*, meaning *above or below one milliwatt*. Since 1 mw is neither above nor below 1 mw, 1 mw = 0 dbm. Relative values are tabulated in Table 4·1. It should be noted that

$$\text{db} \pm \text{db} = \text{db}$$
$$\text{dbm} \pm \text{db} = \text{dbm}$$
$$\text{dbm} \pm \text{dbm} = \text{db}$$

PROBLEMS

4·1 The angle of the load reflection coefficient is 115° and the VSWR is 6.4. Plot the following points on a Smith chart: Z_L, $Z_{\min}$, $Z_{\max}$, Y_L, and the point where $2\beta l - \psi = 0$.

4·2 The VSWR on a 50-ohm transmission line is 4.0. If the load impedance is a pure resistance greater than the characteristic impedance, what are the values of Z_L, $Z_{\min}$, $Z_{\max}$, Y_L, and the angle of the reflection coefficient? Record the distance in wavelengths to the first $E_{\max}$ and $E_{\min}$ points.

4·3 The VSWR on a 100-ohm transmission line is 2.5. If the load impedance is a pure resistance smaller than the characteristic impedance, what are the values of Z_L, $Z_{\min}$, $Z_{\max}$, Y_L, and the angle of the reflection coefficient? What is the input impedance at a distance 0.14 wavelength from the load?

4·4 A 50-ohm transmission line is terminated in a load impedance $Z_L = 50 - j60$ ohms. Use a Smith chart to obtain the VSWR, Y_L, the magnitude and angle of the reflection coefficient, and the distance in wavelengths to the first $E_{\max}$ point. Record the input impedance and input admittance at a point 6.18 wavelengths from the load.

4·5 The VSWR on a 50-ohm transmission line is 2.0 and the first $E_{\min}$ is located 0.2 wavelength from the load. Find Z_L, Y_L, $Z_{\min}$, $Z_{\max}$, the magnitude and angle of the reflection coefficient, and the reflected power in per cent.

4·6 If the frequency is 1,000 Mc in Prob. 4·5, what is the distance in centimeters from the load to the first $E_{\min}$?

4·7 The applied frequency is 3 Gc, Z_0 is 50 ohms, and the load impedance is $(120 + j85)$ ohms. Find the VSWR, the values of $Z_{\max}$ and $Z_{\min}$, and the distance in centimeters to the first $E_{\max}$ and $E_{\min}$ points.

4·8 The reflection coefficient is 0.6 and βl from the load to the first $E_{\max}$ is 0.1 wavelength. Find the VSWR, Z_L, Y_L, $Z_{\min}$, $Z_{\max}$, and the angle of the reflection coefficient if Z_0 is 50 ohms.

4·9 Z_L is 100 ohms and Z_0 is 50 ohms. Show mathematically and explain why the load admittance as read on the Smith chart is the reciprocal of the normalized impedance.

4·10 It is desired to use a quarter-wave matching section of transmission line between two lines which have characteristic impedance values of 600 and 50 ohms, respectively. What is the characteristic impedance of the matching section?

4·11 Calculate the power ratios for the following decibel values: 7, 24, 46, 32, 66, and 18 db.

4·12 Calculate the decibel values corresponding to the power ratios 1,620; 145; 1,840; 188; 16,700; 560; and 58,600.

4·13 The signal to a three-stage system is −35 dbm. The first stage has a gain of 28 db, the second stage has a loss of 3 db, and the third stage has a gain of 76 db. What is the power output at each stage?

4·14 Express the following powers in dbm: 160 $\mu\mu$w, 280 $\mu\mu$w, 25 mw, 56 watts, 20 kw, 2 megawatts, 720 $\mu\mu$w, and 18 mμw.

4·15 Convert the following dbm values to microwatts: −20, −28, −92, 35, 8, and 87 dbm.

REFERENCE

1. P. H. Smith, Transmission Line Calculator, *Electronics*, vol. 12, pp. 29–31, January, 1939. An Improved Transmission Line Calculator, *Electronics*, vol. 17, p. 130, January, 1944.

CHAPTER

5

USING THE SMITH CHART

Introduction. This chapter is concerned with examples which illustrate how the Smith chart can be used for practical microwave computations. A considerable amount of microwave technique is concerned with impedance transformation produced by a length of guiding structure. The usual time-consuming calculations are eliminated by using the Smith chart, and, for this reason, the chart is invaluable to the microwave engineer.

The discussions of the chart include basic impedance matching problems, a brief discussion of each of the radial scaled parameters, and sample problems which outline the methods used to evaluate the radial scaled parameters in practical applications. The chart is used to evaluate the unique relationships of impedance, reflection coefficient, voltage-standing-wave ratio, and the position of a voltage minimum on the line.

5·1 Determination of input impedance and admittance

In the example of Fig. 5·1 the load impedance of $40 + j25$ ohms terminates a 50-ohm transmission line. The normalized impedance $z_L = (40 + j25)/50 = 0.8 + j0.5$ is located at the point A on the diagram. The angle of the reflection coefficient is 96°, and the reference wavelength point is 0.116λ, as indicated on the wavelength toward generator scale. The VSWR read at point B is 1.79. The point B is also the maximum voltage and maximum impedance (pure resistance of 89.5 ohms) point and is located $0.25 - 0.116 = 0.134\lambda$ from the load. The normalized load admittance is read diametrically opposite the normalized impedance z_L and is located at point C, where $y_L = 0.89 - j0.56$.

Practical measurements of load impedance usually require the location of the load impedance point at a particular distance from a voltage minimum point when the VSWR and the position of the voltage minimum have been determined. As an example, suppose that the load is located a distance of

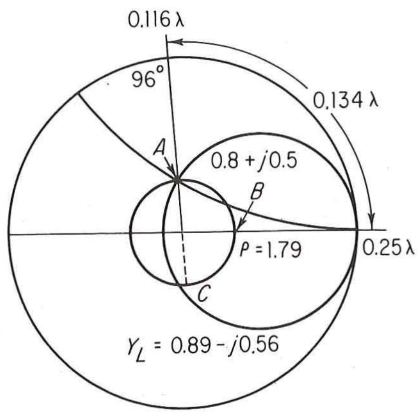


Fig. 5·1 Using the Smith chart to determine VSWR and admittance.

0.313λ from the first voltage minimum and the VSWR is 2.4. Figure 5·2 illustrates this problem in which the VSWR circle is shown, and the distance from $Z_{\min}$ to the load (0.313λ), measured CCW from 0 at the $E_{\min}$ point, is located at point A. A straight line from the center of the chart to A intersects the ρ circle at B, and the normalized impedance is read as $1.4 + j1.0$. The load impedance is therefore $50(1.4 + j1.0) = 70 + j50$ ohms. The normalized input admittance y_L is indicated at point C, where

$$y_L = (0.475 - j0.34)$$

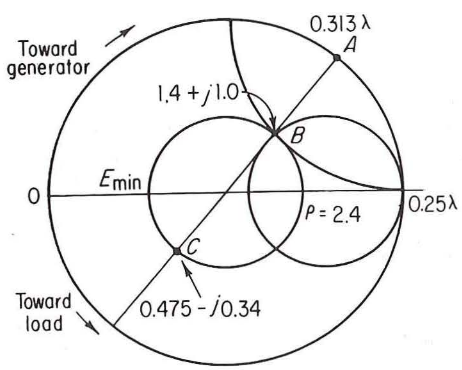


Fig. 5·2 Using the Smith chart to locate the load impedance when the VSWR and position of a voltage minimum have been determined.

5·2 Impedance matching

The principles of impedance matching are, in general, applicable to both waveguides and transmission lines, although the physical form and the behavior of a given matching structure as a function of frequency may be different.

Suppose it is desired to match a 50-ohm transmission line terminated with a load $Z_L = 100 + j50$ ohms. The normalized load impedance is $2 + j1$ ohms and is plotted as point A in Fig. 5·3. If the reactive component of the load is

eliminated by placing a capacitive reactance $-j50$ ohms in series with the load, only the resistive component remains. This is shown by moving along the constant resistance circle to point B. *There is still a VSWR of 2.0, so it is obvious that elimination of the reactive component of the load is not the only requirement for matching the transmission line.* For a matched condition to exist on a transmission line, the normalized input impedance must be $1 + j0$. Therefore, the reactive component must be eliminated on the $R/Z_0 = 1$ circle. A line is drawn from the center of the chart through the point A intersecting the wavelength scale at 0.213λ at D. The $R/Z_0 = 1$ circle is located by traveling clockwise around the constant VSWR circle to point C, where the normalized input impedance of the line is $1 - j1$. The distance along the line to this point is $0.338\lambda - 0.213\lambda = 0.125\lambda$, as indicated on the wavelengths toward generator scale. A *series* reactance of $+j50$ ohms placed at this point eliminates the capacitive reactance, and the input impedance point moves to the center of the chart, where $R/Z_0 = 1$ and $X/Z_0 = 0$. The transmission line is now matched. In other words, a VSWR equal in amplitude and opposite in phase has been introduced at point C.

The requirement that the input impedance for a match must be $1 + j0$ can also be obtained by traveling around the constant VSWR circle to the point P. The impedance is $1 + j1$, and a capacitive reactance must be added in *series* with the line to cancel the inductive reactance at this point. The distance from the load at A to this new point at P is 0.449λ as indicated.

5·3 Single-stub transformer

Since admittance values are additive at parallel junctions, transmission line problems are considerably simplified if they are considered in terms of the input admittance. A device which is often employed with parallel-wire and coaxial lines consists of a short section of transmission line connected in parallel with the transmission line and terminated in a short circuit. This impedance-transforming device is called a single-stub transformer.

The problem illustrated in Fig. 5·3 is now considered as an admittance (shunt) problem. The normalized input impedance is converted to normalized input admittance by $\lambda/4$ rotation of the diagram. This corresponds to location of the reciprocal of the normalized input impedance, as previously demonstrated, at a point diametrically opposite the point A. The normalized input impedance is located at point M, and the normalized value of input admittance y_i is read as $0.4 - j0.2$. For a match, $B/Y_0 \pm jB/Y_0$ must have the value $1 \pm j0$. The line from the center of the chart through M intersects the wavelengths toward generator scale at 0.4630. The distance measured on the wavelength scale by traveling clockwise from the load admittance point M to the point P, where $B/Y_0 = 1$, is 0.1990λ. The normalized input admittance at this point is $1 + j1$ (a capacitive susceptance). An inductive susceptance connected *across* the line at this point results in the input admittance point moving to the center of the chart, where $G/Y_0 = 1$.

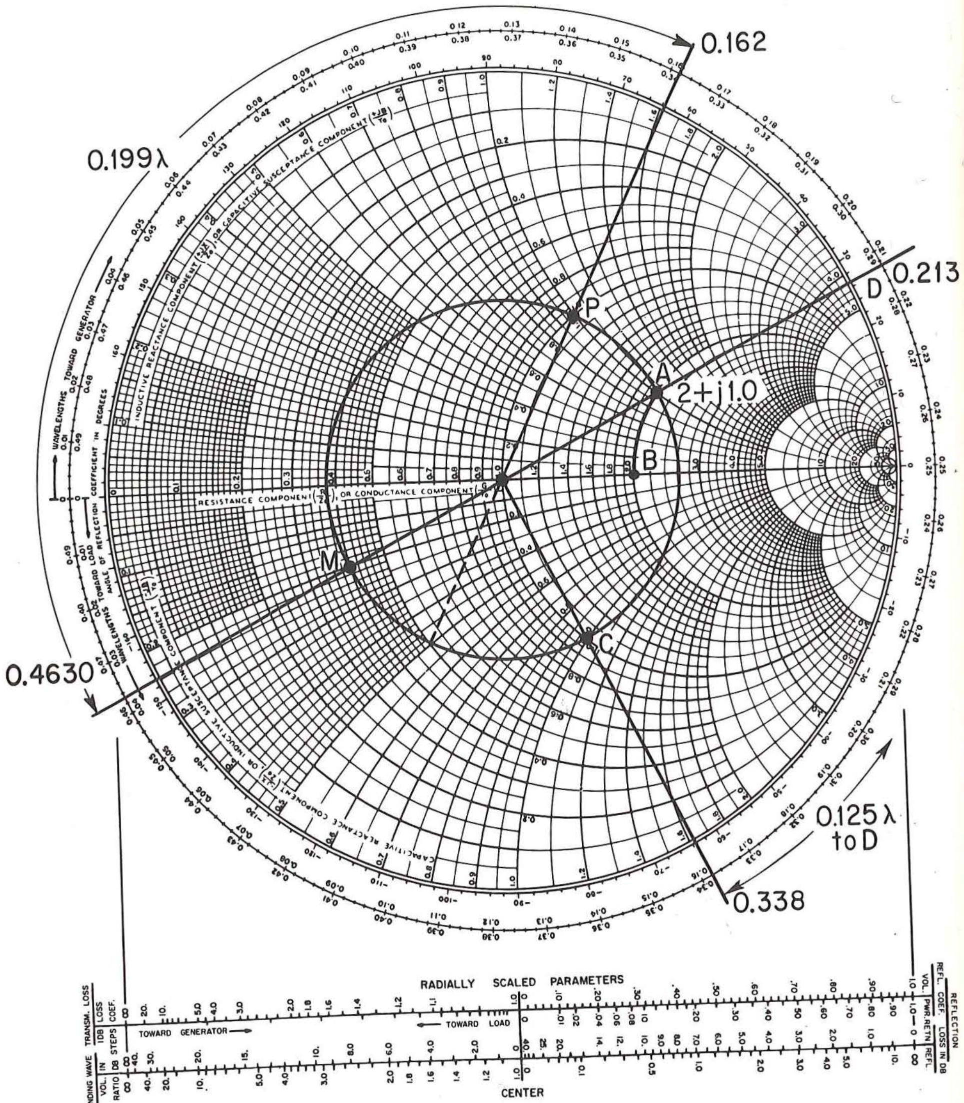


Fig. 5·3 Smith chart. (From Irving L. Kosow (ed.), *"Microwave Theory and Measurements,"* by the *Engineering Staff of the Microwave Division, Hewlett-Packard Company,* © 1962, by *permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.*)

If the frequency of the source in the above problem is considered to be 600 Mc, the distance along the line at which the series reactance or shunt susceptance must be placed can be calculated from $\lambda = v/f = 50$ cm. The series inductance used to match the line at C would be placed $0.125(50) =$ 6.25 cm from the load. The capacitive reactance to be inserted at a point on

the line corresponding to point P on the diagram is $0.499(50) = 24.95$ cm from the load. The shunt susceptance (inductive) required to match the line is placed across the line at a point corresponding to P located $0.199(5) = 9.95$ cm from the load position. As in the case of impedance matching, two points exist where the transmission line can be matched. The other point on the line where the shunt susceptance (capacitive) is placed for a match is point C and is located $(0.375)(50) = 18.75$ cm from the load.

5·4 Double-stub matching

In the previous section the matching procedure required that a point be located on the line where the reactance or susceptance element is placed either in series or in shunt with the line. The single-stub method of matching a line is not suitable for most transmission line structures because of the difficulty of building a single stub adjustable in position. Therefore, the double-stub transformer is used. It consists of two stubs separated by any distance less than one-half wavelength, and each stub is an adjustable short circuit. A diagram of the double-stub unit is shown in Fig. 5·4. The separation of the stubs is usually $\frac{3}{8}$ wavelength, as indicated on the diagram. The load admittance which is to be matched is the impedance as seen at the load side of stub 1, even though the actual load may be any distance from the stub. The separation between the load and tuner should not be great in terms of wavelengths, especially if the source exhibits frequency instability. If the tuner is many wavelengths from the load, a small change in frequency causes large changes in admittance (or impedance), as seen at the tuner. This is referred to as the *long line* effect. Even though a given spacing will not accommodate all impedances, the spacing can be chosen so that the range of impedances is sufficiently wide to allow for any slight variations in load or frequency that may be encountered.

The normalized input admittance y_i shown in Fig. 5·4*b* must be $1 + j0$ for a matched condition. Since the stubs can change susceptance only, the desired susceptance component is obtained by adjustment of stub 1. Stub 1 is used to add the proper amount of susceptance to Y_L so that the resulting admittance may be transformed by the $\frac{3}{8}$ wavelength line to a point on the $G/Y_0 = 1$ circle. This is illustrated by calculating the admittance values and lengths of stubs required to match the load shown in Fig. 5·4*b*.

The normalized (per unit) load admittance is $1.6 + j1.0$ (point A). The point representing the per unit admittance of stub 1 plus the load admittance $(y_L + y_1)$ must lie somewhere on the $G/Y_0 = 1.6$ circle because stub 1 only adds susceptance to the load admittance. Each point of the $G/Y_0 = 1.6$ circle is rotated through an angle corresponding to the distance between stubs 1 and 2, which is $\frac{3}{8}$ wavelength in this problem, and all possible values of the input admittance y_3 are located on this circle. This rotation of each point is obtained simply by rotating the G_L/Y_0 circle clockwise around the chart a

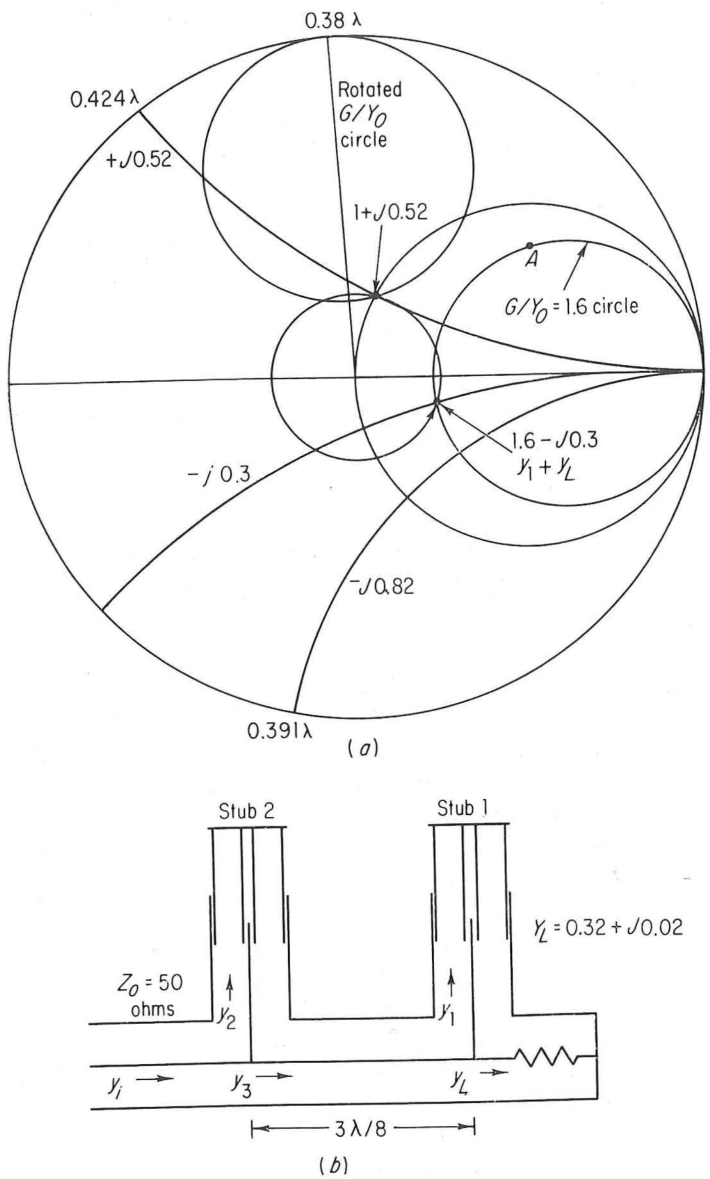


Fig. 5·4 Example illustrating the use of the double-stub transformer.

distance corresponding to the distance between the stubs. Since the normalized admittance $y_i = y_2 + y_3$ must be $1.0 + j0$ and since stub 2 can add susceptance only, it follows that the per unit admittance y_3 must equal $1 + jB$ and for the stated problem $y_3 = 1 + j0.52$. Stub 2 must be $-j0.52$, and the corresponding length is $0.424\lambda - 0.25\lambda = 0.174\lambda$.

The value of the admittance of stub 1 is determined by rotating the y_3 point $\frac{3}{8}$ wavelength CCW along the constant VSWR circle to $g_L = 1.6$ circle. This point represents the total admittance $(y_1 + y_L) = (1.6 - j0.3)$, and if y_L is subtracted from this value, the value of y_1 is found to be

$$y_1 = (y_1 + y_L) - y_L = (1.6 - j0.3) - (1.6 + j0.52) = -j0.82$$

The length of the stub 1 is determined as

$$0.391\lambda - 0.25\lambda = 0.141\lambda$$

The following examples illustrate the limitations of the double-stub transformer:

1. If a circle corresponding to a normalized conductance $g_L = 2$ is drawn at the 3.8 wavelength distance, it is noted that this circle is tangent to the $G/Y_0 = 1$. Since the rotated g_L circle must intersect the $G/Y_0 = 1$ circle, it is obvious that if the normalized conductance value is greater than 2, these circles will not intersect and the matched condition cannot be obtained.
2. If the distance between the stubs is $\lambda/4$ and the normalized conductance g_L is 1.0, it is noted that the $G/Y_0 = 1$ circle and the g_L circle are tangent at the center of the chart. This signifies that any normalized conductance greater than 1 cannot be tuned out.

The limitation of the stub is a function of the distance between the stubs. The range of operation is least if the stub separation is one-quarter wavelength and increases if the separation distance is increased toward one-half wavelength or is decreased toward zero.

The operation of the three-stub tuner is even more difficult to describe using this step-by-step method.

5·5 Radial scaled parameters

The various parameters which are uniquely related to one another are shown on the Smith chart diagrams of Figs. 4·5 and 5·5 plotted radially from the center of the chart and labeled *radially scaled parameters.* A circular transmission line calculator with separately rotatable wavelength scale around the rim and with a transparent arm engraved with all the radially scaled parameters can be constructed. An adjustable cross-hair index along the radial arm permits reading any or all of the several scales at the intersection of the slider index. The relationship of these parameters may be

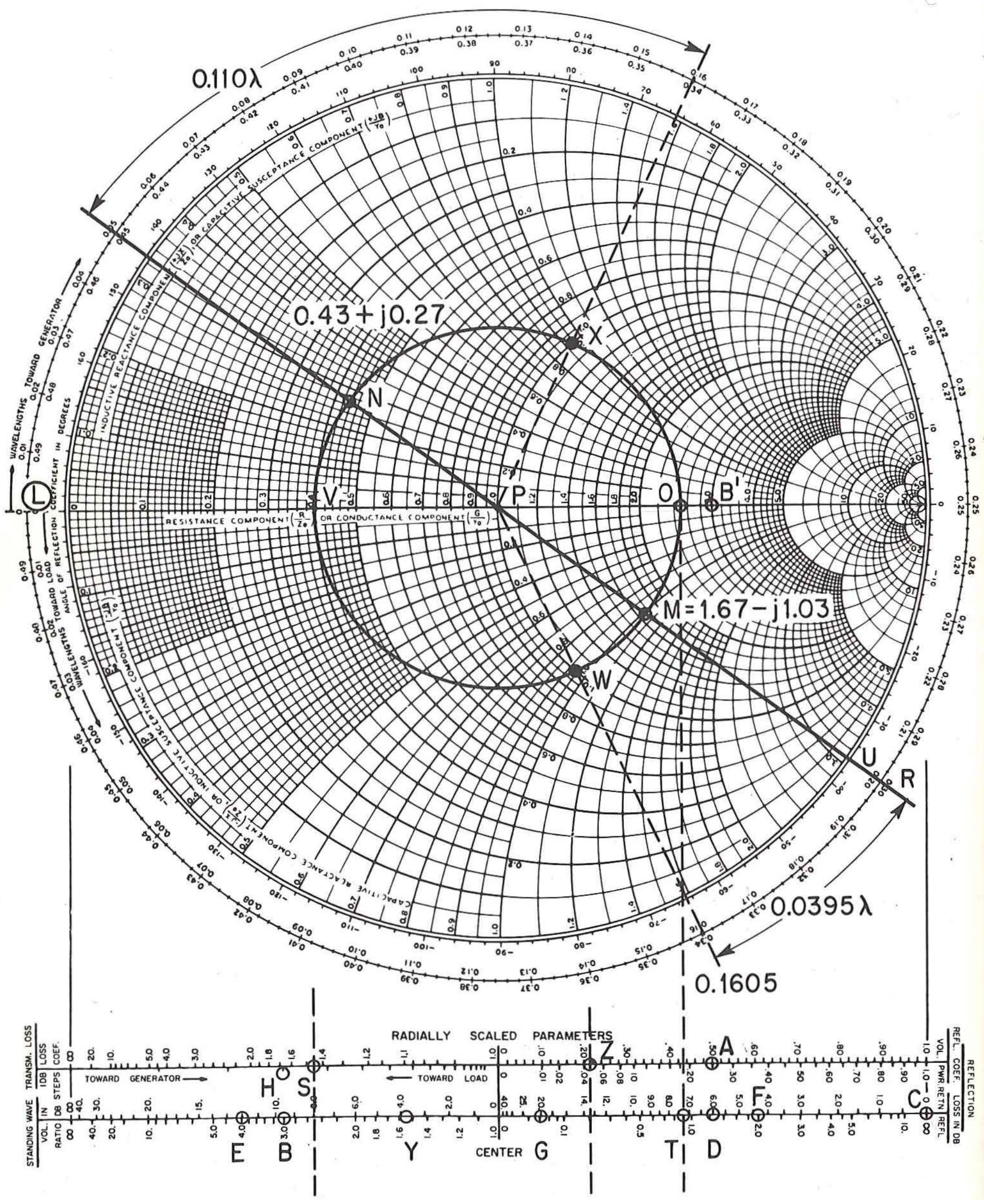


Fig. 5·5 Smith chart. ***(From Irving L. Kosow (ed.), "Microwave Theory and Measurements," by the Engineering Staff of the Microwave Division, Hewlett-Packard Company, © 1962, by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.)***

evaluated by using dividers or any other method for measuring the related radial distances on the chart and the radial scales. For convenience of explanation, the following discussions will consider the use of dividers.

Reflection Coefficient. The reflection coefficient is read on the right-hand side of the radial scale at the bottom of Fig. 5·5. There are two reflection coefficient scales labeled *voltage* and *power*. The voltage reflection coefficient or scale is above the line and is a plot of E_r/E_i and varies from *zero* at the center of the chart (Z_0 point) to *one* at the outside rim of the chart, which is the zero resistance circle. The adjacent scale underneath the same line is the power reflection coefficient scale and is designated Γ^2. An examination of the values shows that each value on this scale is precisely the square of the value of Γ noted above the line.

The angle of the reflection coefficient is zero on the $Z_{\max}$ resistance line where the incident and reflected waves are in phase. The angle of Γ from the zero point is linearly related to the distance traveled, and the sign of the angle is indicated on the outside rim of the impedance coordinate system. The relationship of various parameters to the reflection coefficient can be illustrated on the chart. Use dividers to measure $\Gamma = 0.5$ on the radial scale (point A). Comparing the measured distance to the VSWR scale reading on the right-hand axis of the chart shows a VSWR of 3.0, as indicated by B and B' in Fig. 5·5. The power reflection coefficient Γ^2 at A is noted to be 0.25, and if this value is multiplied by 100, the per cent of reflected power is obtained. The reflected power is 25 per cent of the incident value. When performing measurements on the radial scales and the chart proper, the center of the radial scales corresponds to the center of the chart and the end of the radial scales corresponds to the outside rim of the chart (zero resistance). Therefore, dividers can be used to compare measured values either from the center out, or from the outside rim toward the center, as long as one is consistent as to the chosen reference. The relationships of the remaining parameters to reflection coefficient can be obtained by measuring the corresponding distances on the scales of interest.

Return Loss. The *return loss* is the ratio of the incident power to the reflected power at a point on the transmission line and is expressed in decibels. The reflected power from a discontinuity is expressed as a certain number of decibels below the incident power upon that discontinuity.

$$R_{\text{loss}}(\text{db}) = 10 \log \frac{P_i}{P_r} = 20 \log \frac{E_i}{E_r} = 20 \log \frac{1}{\Gamma} \tag{5·1}$$

The return loss scale is located on the right side of the radial scaled parameters on the line headed "Loss in db." The return loss values are indicated above the line. When the reflection coefficient is 1, the return loss is zero, as indicated at the extreme right which is on the outside rim of the impedance coordinate system (point C). This return loss value indicates that no signal was lost and that all of the signal incident upon the discontinuity was returned toward the source. As the reflection coefficient approaches zero, the return loss approaches infinity. That is, the more perfect the load, the less the reflection from that load. Practical applications of this parameter are all important to the microwave engineer or technician.

Assume that the 3-db attenuator in Fig. 5·6 is perfectly matched (input and output VSWR = 1.0). The indicated input power of 100 mw is decreased to 50 mw at the output of the 3-db attenuator. This 50 mw is reflected from the short circuit back through the attenuator in the reverse direction, and one-half of this reflected power is lost in the 3-db attenuator. The reflected power

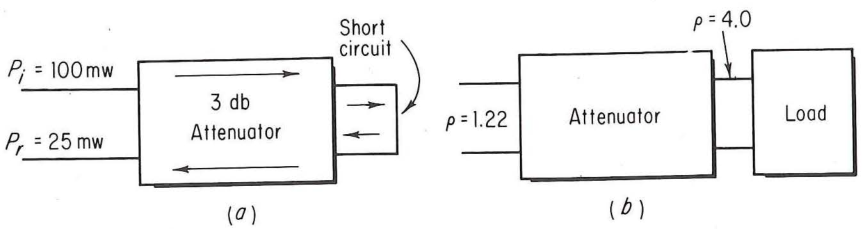


Fig. 5·6 Schematic representation of practical applications of return loss theory.

at the input is 25 mw. The definition of return loss signifies that return loss is the total *round-trip loss* of the signal, and in this case the round-trip loss is 6 db. Use a set of dividers and measure the distance from zero on the return loss scale to the 6-db line on the return loss scale. This point is labeled D on the return loss scale. Use this setting to mark the distance from the right-hand rim of the impedance coordinate system toward the center of the chart. The VSWR as read from the chart is 3.0 at B'. *This example is of special significance since it shows that the VSWR is decreased when attenuation exists on the line and also that a high VSWR can be decreased by placing attenuation in the line.* In many cases, it is not desirable to place attenuation in the line in order to reduce VSWR because of the loss of power in the forward direction. Special components of a unidirectional nature are required to provide the necessary attenuation and to dissipate minimum power in the desired direction of power flow. Using the same setting on the dividers, measure the distance from 1.0 on the voltage and power reflection coefficient scales to obtain the voltage reflection coefficient of 0.5 and power reflection coefficient of 0.25. This verifies that the reflected power is 25 per cent of the incident power. The voltage reflection coefficient is squared and multiplied by 100 in order to obtain the reflected power in per cent.

The second example is illustrated in Fig. 5·6*b*. In order to find the proper value of attenuator required to reduce the VSWR from 4.0 to 1.22, use the Smith chart of Fig. 5·5 and measure the distance on the VSWR scale from the left end of the scale (infinity) to the point on the scale which indicates

a VSWR of 4.0 (point E). This distance can also be measured on the chart from the outside rim of the impedance coordinate system to the normalized resistance circle of 4.0. This same distance on the return loss scale indicates a return loss of 4.4 db (point F). Measure the scale length for a VSWR of 1.22 and mark off the same distance on the return loss scale. The return loss is found to be 20 db (point G). Therefore, 20 db − 4.4 db = 15.6 db, which is the value of return loss to be added. The attenuator value required to obtain this value of return loss is 7.8 db, since the round-trip loss must be 15.6 db.

In the above examples the distances were measured from the outside rim of the impedance coordinate system on the chart and the right end of the radial scales. The measurements can also be made from the center of the chart and the center line of the radial scales; one need only be consistent as to the reference point chosen.

Transmission Loss Scale. The transmission loss in 1-db steps is presented on the left-hand side of the radial scaled parameters. It is the equivalent *one-way* loss in the signal on the transmission line. By measuring the distance to D on the return loss scale and measuring off the same distance to H on the transmission loss scale, it is noted that the attenuation reading is one-half the reading on the return loss scale.

Standing-wave Ratio in Decibels. In many practical applications the standing-wave ratio is measured in decibels, and the corresponding VSWR and other parameters are obtained from this measurement. The standing-wave ratio in decibels is expressed as

$$\text{SWR (db)} = 20 \log \text{VSWR} \tag{5·2}$$

A radial scale plot of the standing-wave ratio in decibels is shown in Fig. 5·5, and values of SWR (db) can be compared to the various parameters by measuring equidistant points on the desired scales as demonstrated in previous examples.

Mismatch Loss (Reflected Loss). The mismatch loss is a measure of the loss caused by reflection. It is the ratio of incident power to the difference between incident and reflected power and is expressed in decibels as follows:

$$\text{Mismatch loss (db)} = 10 \log \frac{P_i}{P_i - P_r} = 10 \log \frac{E_i^2}{E_i^2 - E_r^2}$$

$$= 10 \log \frac{1}{1 - \Gamma^2} = 10 \log \frac{(1 + \rho)^2}{4\rho} \tag{5·3}$$

As an example, a VSWR of 3.0 represents a mismatch loss of 1.25 db. This point is shown, opposite the return loss of 6 db, at B in Fig. 5·5.

Example: A signal frequency of 3 Gc is applied to a transmission line which has a characteristic impedance of 50 ohms. The voltage-standing-wave ratio is 2.5:1, and the first voltage minimum is located 0.2 wavelength from the load. It is desired to find the value of load impedance, load admittance, reflection coefficient, angle of reflection coefficient, the maximum and minimum impedance on the line, mismatch loss, return loss, and the distance to the first voltage minimum in centimeters. Also, it is desired to find the point where a *series* reactance and the point where a *shunt* susceptance would be added in order to match the line.

A matched 3-db attenuator is inserted in the line in front of the load; it is desired to know the new values of VSWR, reflection coefficient, and return loss.

The problem is solved using the Smith chart of Fig. 5·5.

$$\text{Wavelength } \lambda = \frac{v}{f} = 10 \text{ cm}$$

The VSWR circle is drawn with a radius of 2.5. The load impedance is found by traveling from the point L on the fractional wavelength scale at the left side of the chart CCW toward the load 0.2λ to the point R. Draw a line from R to the point P at the center of the chart. Record the normalized load impedance at M $(1.67 - j1.03)$. Multiply this value by the characteristic impedance of 50 ohms to obtain the load impedance of $(83.5 - j51.5)$ ohms.

The load admittance is found by extending the line RP to intersect the VSWR circle at N and reading the normalized admittance value $(0.43 + j0.27)$. $Y_L = (0.43 + j0.27)/50 = (0.0086 + j0.0054)$ mhos.

Use dividers to measure the distance PO, the 2.5 VSWR point on the Z_{max} axis. Use this value to obtain the mismatch loss of 0.88 db on right side lower radial scale. Also, the return loss is found at point T and is 7.3 db.

The angle of reflection coefficient is $-36°$ at point U.

The maximum impedance at O is $(\text{VSWR})(Z_0) = 125$ ohms, and the minimum impedance at V is $Z_0/\text{VSWR} = 20$ ohms.

The distance from the load to the first voltage minimum is 2 cm since the wavelength is 10 cm.

The series reactance required to match the line is located by traveling around the VSWR circle from the load at M to the point W. The impedance at this point is $1 - j0.95$. A series reactance of $+j0.95$ is required at this point to cancel the $-j0.95$ value. This reactance is located a distance of $0.20\lambda - 0.1605\lambda = 0.0395\lambda$, as indicated on the fractional wavelength scale. Since $\lambda = 10$ cm, the series reactance is located $0.0395(10) = 0.395$ cm from the load.

The point where the shunt susceptance is added is located by traveling around the constant VSWR circle to the load admittance value of $1 + j0.095$ at X. A shunt susceptance of $-j0.95$ is placed at this point to obtain a match. The distance from the load is obtained by noting the fractional wavelength change indicated at the intersection of the lines PN and PX extended to the outside rim of the chart. The values of $0.45 - 0.34 = 0.110\lambda$ from the load. The distance in centimeters is therefore $(0.11)(10) = 1.1$ cm.

The VSWR of 2.5 resulted in a return loss of 7.3 db, as indicated at point T. Insertion of the 3-db matched attenuator corresponds to the addition of a return loss of 6 db. The total return loss is now 13.3 db. The corresponding VSWR is 1.55 at Y, and the reflection coefficient is 0.215 at Z on the radial scale.

PROBLEMS

Use the Smith chart to work the following problems.

5·1 The load impedance of a 100-ohm transmission line operating at a frequency of 1.5 Gc is located 0.16λ from the first voltage minimum. The VSWR is 3.6:1.

a. Find Z_L, Y_L, Z_{min}, Z_{max}, the angle and magnitude of the reflection coefficient, and the distance in centimeters to the first voltage maximum.

b. What are the values of return loss and equivalent attenuation?

c. What is the standing-wave ratio in decibels?

d. If a 6-db attenuator (three wavelengths long) is placed in front of the load, what are the new values of the parameters listed in (*a*)?

5·2 An antenna is located 4.6 m from the source. The characteristic impedance of the lossless line is 50 ohms, the VSWR is 3:1, and the antenna presents a pure resistance load greater than the characteristic impedance. If the operating frequency is 50 Mc:

a. Find Z_L, Z_{min}, Z_{max}, the angle and magnitude of the reflection coefficient, Y_L, and the per cent of reflected power.

b. Find the values of return loss, mismatch loss, equivalent attenuation, and SWR in decibels.

c. Find the distance in centimeters to the first E_{max} and E_{min} points.

d. Find the input impedance and the input admittance at the generator.

e. Locate the point on the line and find the value of series reactance required to match the antenna.

f. Find the point on the line and the value of shunt susceptance required to match the antenna.

g. If the line had an attenuation loss of 0.2 db per meter, calculate the input impedance, VSWR, input admittance, and the return loss presented to the generator.

5·3 The first voltage minimum is located 0.23λ in front of the load, and the VSWR is 6:1 on a lossless transmission line which has a characteristic impedance of 100 ohms. The operating frequency is 100 Mc.

a. Find Z_L, Z_{max}, Z_{min}, return loss, and per cent of reflected power.

b. A 6-db attenuator (five wavelengths long) is placed in the line. What are the new values of the parameters listed in (*a*)?

5·4 What is the value of attenuation which is required to reduce a VSWR of 8:1 to 1.08:1? How much power is lost because of the attenuator?

CHAPTER

6

COAXIAL TRANSMISSION LINES AND MEASURING EQUIPMENT

Introduction. This chapter deals with coaxial transmission lines and measuring instruments used at ultrahigh frequencies. The study is concerned with the same general principles previously applied to unconfined fields. An introduction to basic measurement equipment and measurement techniques is presented in order to acquaint the student with practical applications of microwave techniques.

6·1 The coaxial line

The coaxial line is a special form of confined space composed of two concentric conductors separated by an insulating material. The coaxial line illustrated in Fig. 6·1 can be constructed in both rigid and flexible forms.

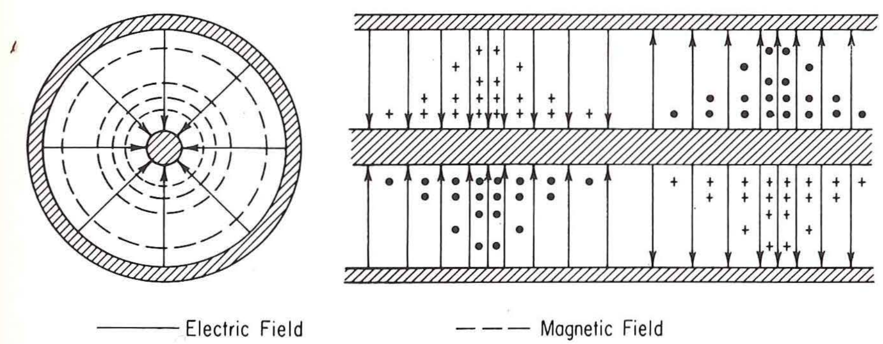


Fig. 6·1 The coaxial transmission line.

The dielectric in the rigid coaxial line is usually air, and the center conductor is concentrically located within the outer conductor by means of dielectric insulating supports called *beads*. In the case of the flexible cable, the center conductor is surrounded throughout its length by a flexible dielectric material such as polyethylene. The characteristic impedance of the coaxial line is in the order of 30 to 100 ohms as given by

$$Z_0 = \frac{138}{\sqrt{\epsilon'}} \log \frac{b}{a} \qquad (6\cdot1)$$

where b is the outer diameter of the inner conductor and a is the inner diameter of the outer conductor, or a and b can be the corresponding radii.

The electromagnetic field is restricted to the region between the inner and outer conductors, and this results in practically perfect shielding between fields inside and outside of the line. The dominant mode in coaxial lines is the TEM (transverse electromagnetic) in which the electric lines of force are radial and the magnetic lines of force are concentric circles as shown in the diagram. The strength of the electric field is proportional to the voltage difference between conductors and is inversely proportional to the distance between the axis of the line and the point in question. The magnetic field strength is proportional to the current flow and varies inversely with the distance from the axis of the line.

Both inner and outer conductors can supply charge and current distributions so that any rate of variation of the electric and magnetic fields can take place along the line. That is, the currents adjust themselves so that the field is transmitted for any applied frequency. Therefore, the coaxial line is not frequency sensitive, and it is a broad band device having no cutoff frequency. The electric and magnetic vectors for the dominant TEM mode are exclusively perpendicular to the direction of motion so that the *velocity of propagation is the same as for a wave in the same insulating medium without any conductors.*

Higher order modes of propagation may be present unless the conductor dimensions are chosen properly when operating in the microwave range. The higher order modes which are most likely to be encountered are the TE_{11} and TM_{01} waveguide-type modes, which are discussed in Chap. 7, Sec. 7·9. The cutoff wavelength for TE modes, which have variations around the circumference, is given by the approximate relation

$$\lambda_c \cong \frac{2\pi}{n} \frac{b + a}{2} \qquad n = 1, 2, 3, \ldots \qquad (6\cdot2)$$

where a and b are the inner and outer radii, respectively.

The approximate cutoff for TM modes is given by

$$\lambda_c = \frac{2}{p}(b - a) \qquad p = 1, 2, 3, \ldots \qquad (6.3)$$

where a and b have been defined and p is the half wavelength spacing between conductors.

6·2 The slotted line

The slotted line is one of the important measuring instruments used at microwave frequencies. It is designed to measure the standing-wave pattern of the electric field intensity which is a function of the longitudinal position in the guiding structure. A probe is mounted on a carriage which slides along the outside of the section of coaxial line or waveguide which has a longitudinal slot. The probe extends into the slot and is provided with an adjustment for varying the probe penetration into the slot and with a tuning adjustment (usually a stub) used to cancel the reactive component of probe impedance. The probe is connected to a barretter or crystal detector which detects the r-f voltage. This voltage is amplified and applied to the appropriate indicating meter.

A slotted section used over the frequency range from about 300 to 5,000 Mc is shown in Fig. 6·2*a*. The slotted section in Fig. 6·2*b* is usable over the frequency range of 500 to 4,000 Mc.

The standing-wave ratio is measured by sliding the probe along the line for a maximum and minimum indication on the output meter. The standing-wave ratio is calculated from the above data or read directly from the indicating device if the indicator is calibrated in SWR.

The wavelength of the signal frequency can be measured by obtaining the distance between minima, since it was previously shown that the distance between successive minima or maxima is one-half wavelength.

6·3 Errors in slotted-line techniques

The possible sources of error associated with slotted-line measurements must be carefully evaluated and proper operating techniques must be applied in order to eliminate or minimize these errors. The accuracy of standing-wave ratio measurements is limited by the connectors on the coaxial slotted section since the slotted section probe responds to the combined reflections from the connector and the load beyond the connector. The uncertainty in the standing-wave ratio measurements can be evaluated when the inherent SWR of the slotted section is known.

Variation of the maxima and minima at different points on the line is referred to as the *slope* error. It can be caused by variation of the probe depth as the probe carriage is varied and also by energy leakage through the slot. This error can be adjusted to a minimum value in some slotted sections.

Probe Tuning Errors. One of the major sources of error in standing-wave measurements is excessive probe penetration. The presence of the probe affects the VSWR because it is essentially an admittance shunting the line.

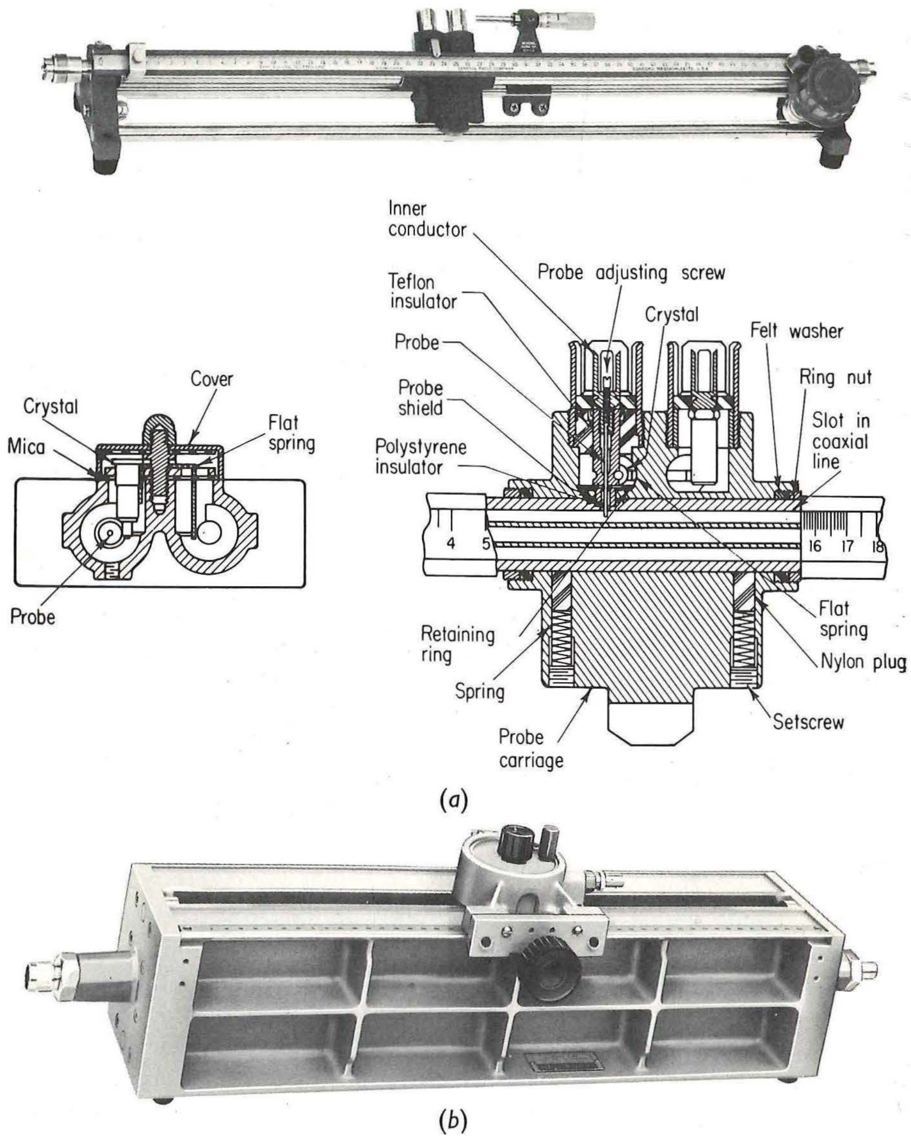


Fig. 6·2 Coaxial slotted sections. (*a*) General Radio Type 874 LBA slotted line. (*General Radio Company.*) (*b*) Model 805. (*Hewlett-Packard Company.*)

Excessive coupling to the line causes a shift in the maxima and minima and also causes the measured VSWR to be lower than the true VSWR. In addition to the distortion of the field pattern, reflections from the probe vary when the probe is moved. Errors in measurement of low VSWR arise when these reflections are re-reflected from a mismatched source. Therefore, the probe

coupling should be kept as small as possible except in cases where it is only desired to examine the minimum point on the standing-wave pattern. Excessive probe penetration can be minimized by using a high sensitivity detector, assuming that there is adequate signal source power available.

Harmonics and Spurious Signals. Harmonics are usually present in signal sources that have coaxial outputs. Errors are possible if the probe is tuned to a harmonic of the fundamental. The frequency to which the probe is tuned can be easily checked by measuring the half-wavelength distance between two voltage minima on the line. Harmonics are usually reduced to a negligible value in coaxial systems by the use of low-pass filters.

Spurious signals usually arise from improper adjustment of modulating voltages used to square-wave-modulate a signal source such as a klystron which has several modes of oscillation. An illustration of improper modulation of a klystron is shown in Fig. 12·4.

Frequency Modulation. Variations of the instantaneous signal source frequency are referred to as frequency modulation (f-m). The minima of a standing-wave pattern are obscured in the presence of frequency modulation since the minima of the standing-wave patterns at the different frequencies do not appear at the same position on the line. If the f-m becomes excessive, it is possible that other portions of the standing-wave pattern can be distorted. Poor regulation of potentials applied to the oscillator is often responsible for f-m problems. In order to prevent frequency modulation of modulated signal sources, square-wave modulation is used. The presence of f-m on modulated sources is usually displayed on a scope using a frequency meter (frequency discriminating device). The f-m can also be detected by investigating the minimum of the standing-wave pattern when the slotted section is terminated with a short circuit.

Detector Characteristics. The characteristics of barretters and crystals determine the power levels in a measurement system. These detector elements are usually used with equipment which is calibrated in terms of the square-law response. In order to prevent a departure from the square-law response, the barretter should be operated at power levels less than 200 μw and the crystal should be operated at power levels less than 10 μw (Sec. 8·10). In either case, the departure from square law can be checked by noting the detector response on the standing-wave indicator for different levels of input power. This check is performed by measuring a fixed mismatch at different power inputs to the detecting element.

6·4 Standing-wave measurements

A typical setup for performing measurements with the slotted line in the frequency range below 5 Gc using an unmodulated source and a superheterodyne detector or receiver is shown in Fig. 6·3. It consists of a signal source, local oscillator, linear mixer, 30-Mc intermediate frequency (i-f)

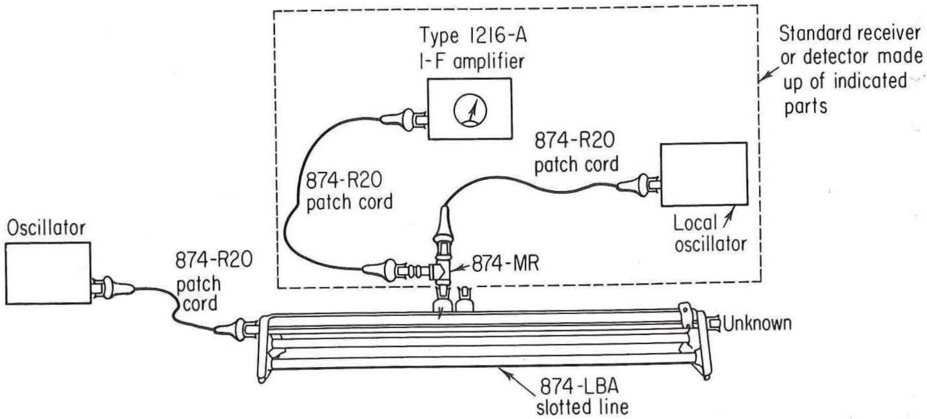


Fig. 6·3 A typical setup for measurements with the Type 874 LBA slotted line using an unmodulated source and superheterodyne detector or receiver. (*General Radio Company*.)

amplifier, and the slotted section. The continuous-wave (c-w) source frequency is sampled by the probe and mixed with the local oscillator frequency in the linear mixer. The resulting 30-Mc i-f signal is applied to the 30-Mc amplifier which incorporates a calibrated attenuator and output meter. The indicating meter has a voltage scale and decibel scale. The VSWR is calculated from the ratio of the maximum to minimum voltage as measured using the voltage scale, or it can be calculated by taking the antilog of the decibel variation since the SWR in decibels is given by

$$\text{SWR (db)} = 20 \log \text{VSWR} \tag{6·4}$$

When using this system, care must be taken to tune the local oscillator to beat with the desired signal and not with one of its harmonics. An amplifier output is produced by any two signals which beat together to produce a 30-Mc difference frequency. As an example, the local oscillator must be tuned to 470 or 530 Mc in order to produce the 30-Mc difference frequency when the source frequency is 500 Mc. If the local oscillator is tuned to 485 or

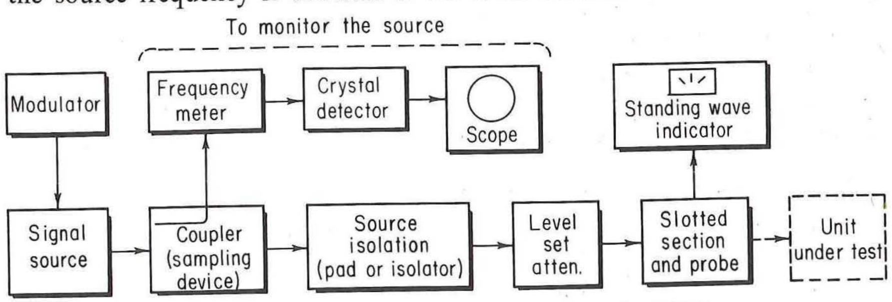
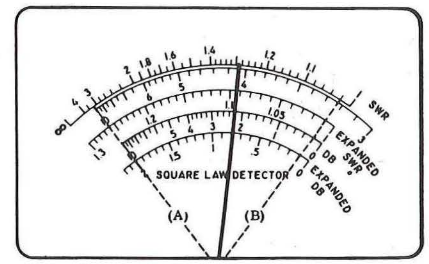


Fig. 6·4 Basic standing-wave measurement system.

515 Mc, it is noted that the second harmonics are 970 and 1,030 Mc, respectively. Either of these frequencies can mix with the second harmonic of 500 Mc to produce a 30-Mc difference frequency. In this case, the system would be operating at 1,000 Mc instead of the indicated 500 Mc.

A block diagram of the most commonly used standing-wave ratio measurement system is shown in Fig. 6·4. This basic system is used in all frequency ranges where the necessary system components are available. The source frequency is modulated at an audio frequency, usually 1,000 cycles per second. The standing-wave indicator is a selective tuned amplifier (tuned to the modulation frequency) which has a calibrated range switch and an output meter which is calibrated in terms of the square-law response of the detector. The standing-wave indicator also provides the necessary bias for barretters. The meter face of a standing-wave indicator is shown in Fig. 6·5. The VSWR measurement is made by moving the slotted section probe along the line to

Fig. 6·5 Detail of meter face which is calibrated in terms of square-law detection. (*Hewlett-Packard Company*.)



obtain maximum deflection as indicated on the output meter. The amplifier gain is adjusted to place the meter pointer to one (1) on the SWR scale. The VSWR is read from the scale when the probe is adjusted for minimum deflection on the output meter. Standing-wave ratios up to 10:1 can be measured by switching downrange and reading VSWR values greater than 3:1 on the second scale. The expanded scale can be used to measure ratios less than 1.3:1. The SWR in decibels can also be obtained using this indicating meter.

Measurement of VSWR values greater than 10 : 1 requires special techniques. Accurate measurements can be made using a method referred to as the width-of-minimum or twice-minimum-power method. In using this method, it is necessary to establish the electrical distance Δl between points on the line at which the r-f voltage is $\sqrt{2}$ times the voltage at the minimum, as illustrated in Fig. 6·6. The validity of this method is based upon the assumption that the standing-wave pattern approximates a parabola in the vicinity of the minimum. If the output meter is calibrated for use with a square-wave detector, the ratio of 1.4:1 corresponds to 3 db, as shown on the meter dial in Fig. 6·5. The VSWR is related to the spacing (Δl in centimeters) and the waveguide wavelength (λ_g in centimeters) by the expression

$$\text{VSWR} \cong \frac{\lambda_g}{\pi \Delta l} \tag{6·5}$$

Measurements of high standing-wave ratios can be performed by measuring the standing-wave ratio in decibels. A precision calibrated attenuator is placed in the line with the level set attenuator (Fig. 6·4). With the precision

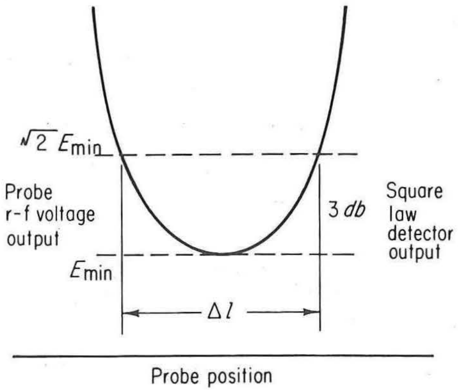


Fig. 6·6 Width of voltage minimum for determination of high VSWR.

attenuator set to zero db, the slotted-line probe is moved to a minimum of the standing-wave pattern, and a convenient reference is set on the standing-wave indicator meter. The slotted-line probe is then moved to the maximum of the standing-wave pattern, and the precision attenuator is used to decrease the standing-wave indicator reading to the original reference. The decibel change in attenuation is used to calculate the VSWR using Eq. (6·4).

6·5 Impedance measurements

A very substantial part of microwave techniques is centered around impedance concepts which explain the nature of transmission lines. Therefore, the measurement of impedance is important in many microwave applications. It has been shown that the input impedance varies with position along a transmission line which has standing waves. Therefore, the impedance measurement must be referred to some reference plane.

The load impedance can be determined on a lossless transmission line from the VSWR and the position of a voltage minimum with a load connected and with the line shorted. The VSWR of the load is measured, and the position of the minimum of the standing-wave pattern is obtained. The load is replaced with the short circuit and the shift in the minimum (less than $\lambda/4$) is obtained. The impedance is a function of the VSWR, the shift in the minimum, and the direction in which the minimum shifts. The reference plane is the plane of the short circuit. The voltage variations with a load connected and with the short connected are illustrated in Fig. 6·7. The load impedance is determined as follows:

1. Connect the load to the slotted section, measure the VSWR, and record the position of a voltage minimum as read on the slotted section scale.
2. Remove the load and connect a short circuit in its place. Record the shift in the minimum of the standing wave. The shift cannot be greater than a quarter of a wavelength. Note whether the shift in minimum is *toward the load* or *toward the generator*.

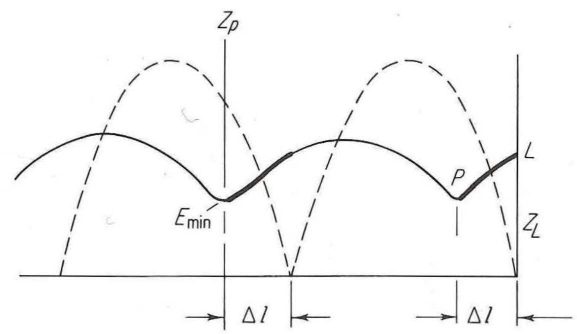


Fig. 6·7 Voltage variation along a transmission line with a load connected (solid lines) and with a short circuit connected at the load position (dotted line).

3. The normalized impedance can be computed from the equation

$$Z_L = Z_0 \frac{Z_p - jZ_0 \tan \beta(\Delta l)}{Z_0 - jZ_p \tan \beta(\Delta l)}$$

$$z_L = \frac{1 - j(\text{VSWR}) \tan \beta(\Delta l)}{\text{VSWR} - j \tan \beta(\Delta l)} \tag{6·6}$$

since $Z_p = Z_{\min}$ in the measurement. $\beta(\Delta l) = \frac{2\pi}{\lambda}(\pm \Delta l)$ in which case (Δl) is *positive* when the minimum shifts *toward the load* and is *negative* when the minimum shifts *toward the generator*.

The calculation of impedance transformation using the equations above can be time-consuming. The Smith chart can be used to solve the same problem as illustrated by the following example which is plotted on the Smith chart of Fig. 6·8.

1. The load VSWR is 2.4 and the VSWR circle is drawn through this point (A) on the chart.

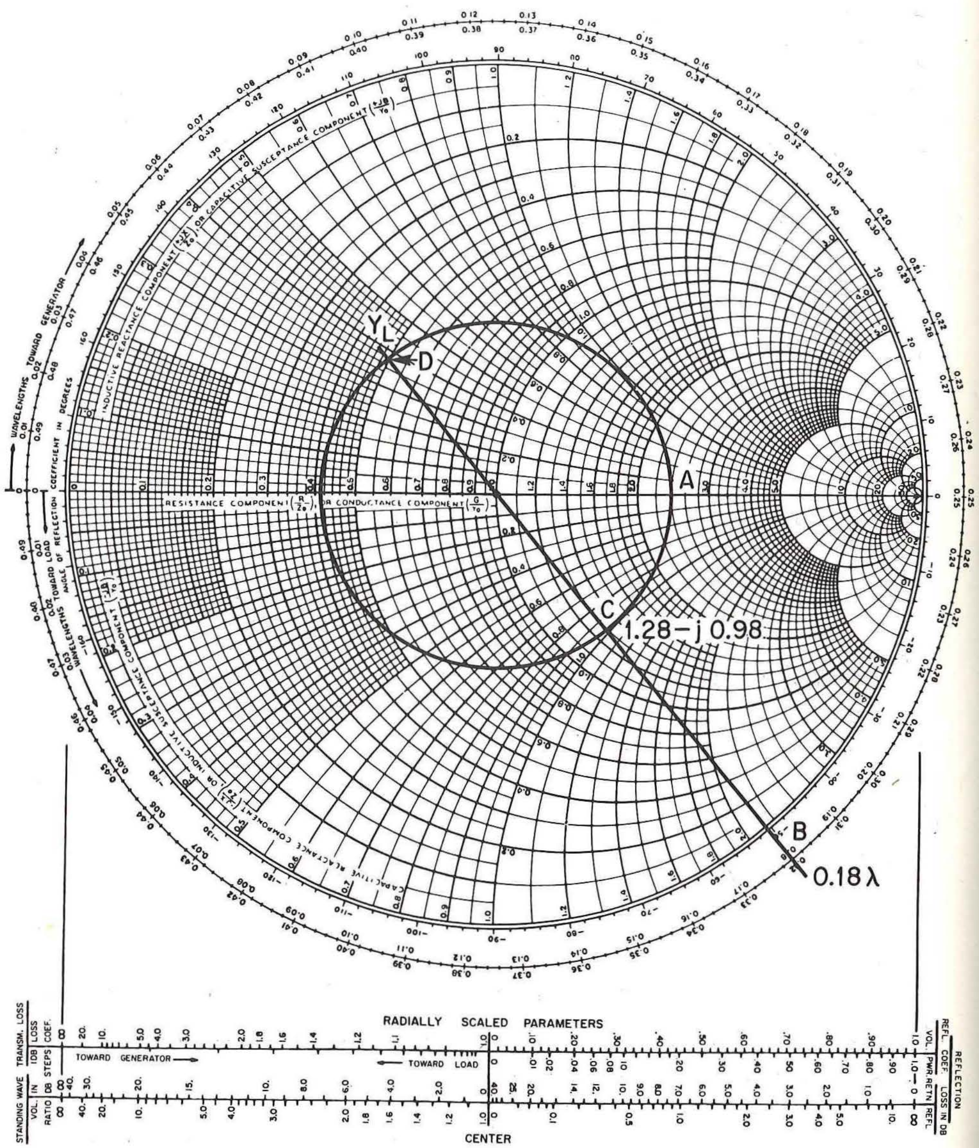


Fig. 6·8 Smith chart. (From Irving L. Kosow (ed.), *"Microwave Theory and Measurements," by the Engineering Staff of the Microwave Division, Hewlett-Packard Company,* © 1962, *by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.*)

2. The centimeter scale reading on the slotted section at the voltage minimum was 26 cm.
3. The minimum shifted to 15 cm when the short was connected to the line, and the shift in minimum was *toward the load.*
4. $\Delta l = 26 - 15 = 9$ cm.
5. The wavelength was found to be 50 cm, therefore $\Delta l/\lambda = 0.18$ wavelengths.
6. Proceed from the short-circuit impedance point on the Smith chart *toward the load* (CCW) to the 0.18 wavelength on the inside wavelength scale and draw a line from this point (B) to the center of the chart and extend the line to intersect the VSWR circle on the opposite side of the chart.
7. The normalized load impedance is located at (C) and is approximately $1.28 - j0.98$. The normalized load admittance is located diametrically opposite point C at D.

It can be seen in Fig. 6·7 that the shift of Δl represents the actual distance of $E_{\min}$ from the load, and, as one proceeds around the VSWR circle from $E_{\min}$ to point C on the Smith chart, the same distance is represented by the points P and L in Fig. 6·7.

If the minimum shifts *toward the load* when the short circuit is connected, the load is *capacitive*. If the shift in minimum is *toward the generator*, the load is *inductive*.

6·6 Residual VSWR of a slotted section

The VSWR caused by reflection from one or more discontinuities in the slotted section system is called the *residual* VSWR. The reflections which are predominant in causing the residual VSWR are fixed phase reflections. Therefore, the discontinuities can be considered as a *single lumped discontinuity*. This residual VSWR must be known in order to determine the possible effects of its interaction with the VSWR being measured.

In the ideal case, the residual VSWR could be determined by terminating the slotted section with a load having no reflection. The "perfect load" does not exist because all loads have some reflection. The effect of a perfect load can be obtained by using a measurement technique in which the slotted section is terminated with a low-loss sliding load. The reflections from the sliding load can be distinguished from those caused by the discontinuities of the slotted section. It should be noted that *any discontinuity in the connector of the load is combined with the residual of the slotted section and that only the VSWR of the load can be isolated from the rest of the system.*

The technique for determining the residual VSWR is illustrated by the relationships of the incident and reflected voltages in Fig. 6·9. The phasor combination of the two voltages is also discussed in Chap. 12.

The sliding load is connected to the slotted section in place of the usual load, and the vector addition of the voltage reflected from the load E_L and the voltage reflected from the slotted section discontinuity E_D add at some random phase, as indicated in Fig. 6·9*a*. The position of the slotted section probe and the position of the sliding load are varied for maximum output on the standing-wave indicator connected to the slotted section probe. This is the in-phase addition of E_L and E_D as shown in Fig. 6·9*b*, which represents the highest obtainable VSWR. The adjustments require some care because they are interdependent.

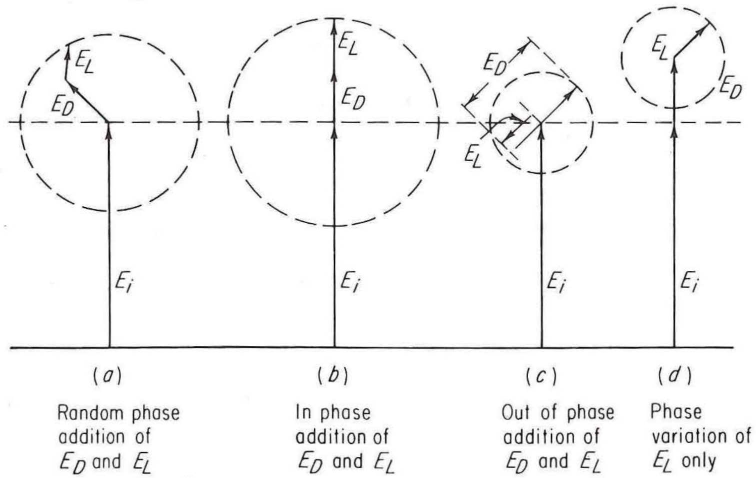


Fig. 6·9 Vector relations of incident and reflected voltages associated with the residual VSWR measurement.

The VSWR is measured with the slotted section, and the value obtained is

$$\rho_{\max} = \frac{E_i + (E_D + E_L)}{E_i - (E_D + E_L)} \tag{6·7}$$

which is represented in Fig. 6·9*b*.

The slotted section probe is returned to the maximum of the standing-wave pattern, and the sliding load is adjusted to obtain a minimum indication on the output at the standing-wave indicator. E_L and E_D are out of phase, and the result is indicated in Fig. 6·9*c*. The VSWR is measured again, and the new value obtained is

$$\rho_{\min} = \frac{E_i + |E_D - E_L|}{E_i - |E_D - E_L|} \tag{6·8}$$

The maximum and minimum reflection coefficients are calculated from

$$\Gamma_{\max} = \frac{\rho_{\max} - 1}{\rho_{\max} + 1} \tag{6·9}$$

$$\Gamma_{\min} = \frac{\rho_{\min} - 1}{\rho_{\min} + 1} \tag{6·10}$$

The reflection coefficient of the discontinuity Γ_D combines with the load reflection coefficient Γ_L to obtain

$$\Gamma_{\max} = \Gamma_L + \Gamma_D \tag{6·11}$$

$$\Gamma_{\min} = \Gamma_L - \Gamma_D \quad \text{or} \quad \Gamma_D - \Gamma_L \tag{6·12}$$

The load and discontinuity reflection coefficients are determined from the above equations.

$$\Gamma_L \quad \text{or} \quad \Gamma_D = \frac{\Gamma_{\max} + \Gamma_{\min}}{2} \tag{6·13}$$

$$\Gamma_L \quad \text{or} \quad \Gamma_D = \frac{\Gamma_{\max} - \Gamma_{\min}}{2} \tag{6·14}$$

The load and discontinuities can be evaluated, and it is only necessary to determine Γ_L or Γ_D. The equivalent VSWR values are obtained and the load can be varied in order to establish its VSWR when the slotted section probe is kept fixed in position. This variation of the load is shown in Fig. 6·9*d*. If the two VSWR values are almost the same value, it may be necessary to measure the residual with another load, or to deliberately introduce an additional reflection from the load (such as placing a small piece of metal foil on the load) and then perform the measurements again.

Another set of equations will be considered in order to avoid the previous time-consuming calculations. The interaction of two mismatches results in maximum and minimum mismatches given by the approximations

$$\rho_{\max} = \rho_1 \rho_2 \tag{6·15}$$

$$\rho_{\min} = \frac{\rho_1}{\rho_2} \tag{6·16}$$

assuming that ρ_1 is greater than ρ_2.

Solving for ρ_1 and ρ_2 it is found that

$$\rho_1 = \sqrt{\rho_{\max}\rho_{\min}} \tag{6·17}$$

$$\rho_2 = \sqrt{\frac{\rho_{\max}}{\rho_{\min}}} \tag{6·18}$$

It is only necessary to measure the maximum and minimum VSWR values and calculate ρ_L and ρ_D from Eqs. (6·17) and (6·18). ρ_L and ρ_D are uniquely determined as related previously.

6·7 Admittance meter

The General Radio Type 1602-B uhf admittance meter is used to measure the admittance and impedance of coaxial circuits in the frequency range from 40 to 1,500 Mc.

The schematic diagram of the circuit is shown in Fig. 6·10. Three adjustable loops couple to the magnetic fields in the coaxial lines. The input voltage is the same for each line since all lines are fed from a common source at a common junction point. The device balances in the same manner as a bridge, and zero output is obtained when the loops are properly oriented.

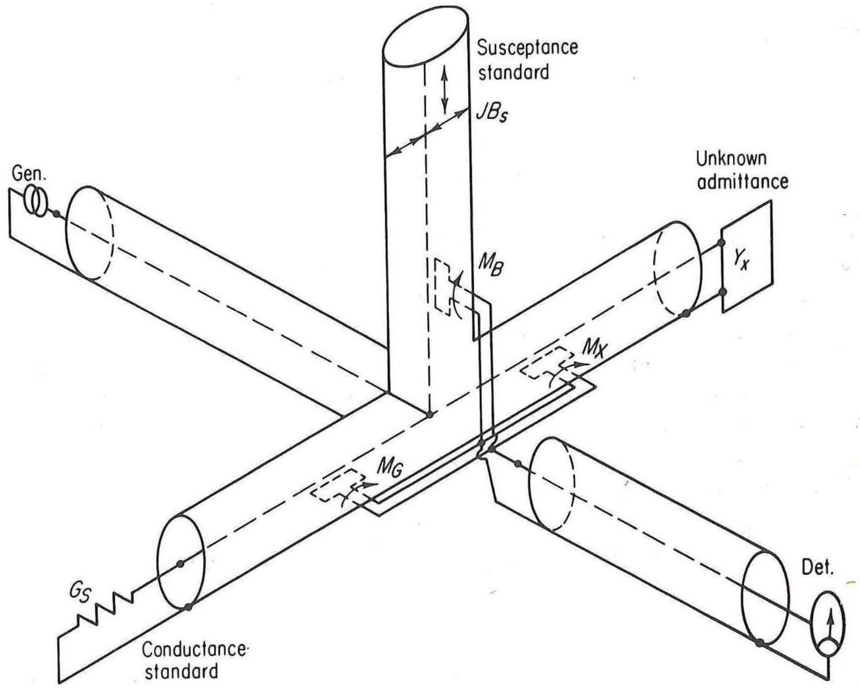


Fig. 6·10 Schematic diagram of the admittance meter. (*General Radio Company.*)

The admittance meter actually measures the admittance at a point inside the junction block directly under the center of the loop coupling to the unknown line. Therefore, a line-length correction is necessary in order to obtain the admittance or impedance at the point of connection to the line. The line-length correction can be eliminated if a low-loss adjustable constant-impedance line is used to obtain an integer multiple of a half wavelength between the admittance meter and the unknown, so that the unknown admittance is the same as the measured admittance. The meter balance is obtained by adjusting the line length when the adjustable constant impedance line is terminated with an open circuit. The unknown is then connected in place of the open circuit and causes the unbalance at a multiple of one-half wavelength from the load. When the meter is rebalanced, the unknown admittance is indicated by the position of the conductance and susceptance indicator arms.

The impedance measurement is performed in a similar way except that the adjustable constant impedance line is terminated with a short circuit prior to initial balance. The short circuit is removed, the load is connected, and the necessary adjustments are performed to obtain a balance or null. The conductance and susceptance scale readings are multiplied by the appropriate multiplying factors in order to obtain the impedance value.

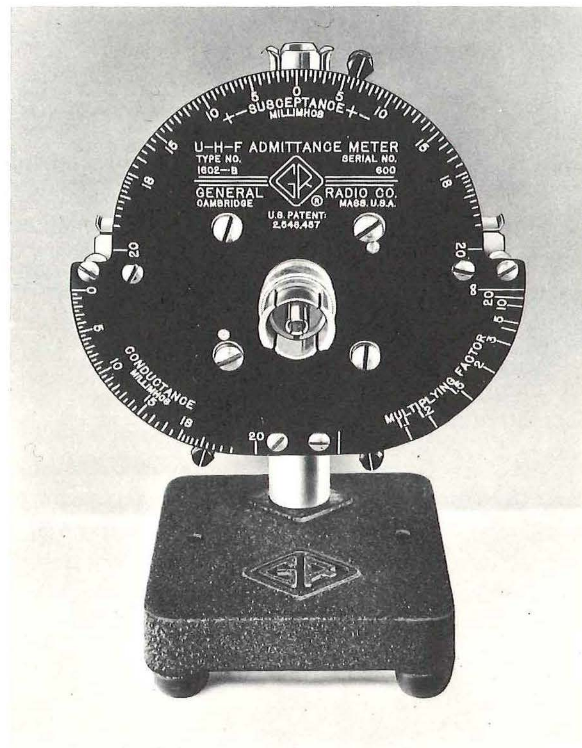


Fig. 6·11 The admittance meter. (*General Radio Company.*)

6·8 Strip transmission lines

Strip transmission lines are essentially modifications of the two-wire line and coaxial lines. The form of the line called *microstrip*, which corresponds to the two-wire line, is shown in Fig. 6·12*a*. It is composed of a flat strip of transmission line, which may be photoetched, and a ground plane using only one copper-clad dielectric sheet. This type of line is more economical to manufacture compared to other types, but since it is incompletely shielded, it exhibits substantial radiation loss and stray coupling effects. Also, the phase velocity and characteristic impedance are unpredictable due to changes in thickness and dielectric constant of the copper-clad dielectric material.

The flat symmetrical-plane strip line of Fig. 6·12*b* can be compared to the coaxial line. The magnetic field lines circle the center conductor in conformance with its geometric shape. All electric field lines are in the transverse plane and terminate at the ground planes in the region of the center strip. The symmetrical strip line is also an open line but not to the same degree as microstrip. There are several types of strip transmission line structures in addition to the examples in Fig. 6·12.

Fig. 6·12 Strip transmission lines. (*a*) Microstrip. (*b*) Symmetrical-strip line.

Strip lines possess the wide band characteristics of coaxial lines and have dominant modes with zero cutoff frequencies. They are well suited for use in complete microwave circuits where economy, light weight, and compactness are of foremost importance. Strip lines have been used extensively in many types of couplers, attenuators, and filters.

CHAPTER 7

WAVEGUIDES

Introduction. The definition of a transmission line which was given in Chap. 2 did not distinguish between the various forms of guiding structures. Any type of transmission line is a waveguide since it is designed primarily to "guide" or conduct energy from one point to another. The term "waveguide" as used in this text refers to guiding structures which enclose the electric and magnetic fields.

The most commonly used waveguide takes the form of a rectangular hollow metal pipe. Therefore, the waveguide theory developed in this chapter will be primarily centered around the rectangular waveguide, and a less detailed examination of transmission through circular pipes and dielectric rods will follow.

7·1 Review of basic concepts

It might be well to review some of the concepts of the electromagnetic wave and the electromagnetic properties of different media before going further into this subject. We do not know the exact nature of the electromagnetic field, as with gravity, but laws that it obeys have been devised by observation of its effects and by hypothesis. Some of the more important concepts are the following:

1. The source of electromagnetic waves is time-varying currents.
2. The electromagnetic wave may be pictured as a vector field consisting of two interrelated components, the electric vector and the magnetic vector.
3. Propagation of the fields is radial from the source in wave fashion except as affected by a change in media of propagation.
4. The properties of media affecting transmission are permittivity, permeability, and conductivity.
5. At the boundary between media of different electrical properties, the

electromagnetic wave will, in the general case, be partially transmitted and partially reflected.

a. The tangential components of the electric and magnetic fields are equal on the two sides of any boundary between physically real media.

b. The tangential component of electric field vanishes at the surface of a perfect conductor.

c. The normal component of a time-varying magnetic field vanishes at the surface of a perfect conductor.

6. Electric field lines may begin and end on charges. If an electric field ends on a conductor, it must represent a charge induced on the conductor.
7. Magnetic field lines can never end, since magnetic charges are not known physically. Magnetic fields must always form continuous closed loops, surrounding either a conduction current or a changing electric field (displacement current).
8. Electric field lines may form continuous closed paths, surrounding a changing magnetic field.
9. Electromagnetic waves are classified according to their equiphase surfaces or wavefronts, and the polarization is defined by the direction of the electric field.

Rays Lines indicating the direction of propagation.

Plane waves Wavefronts of the electric and magnetic fields are parallel to the plane of the source.

Transverse waves Electric and/or magnetic fields lie in a plane perpendicular to the direction of propagation.

1. TEM (transverse-electromagnetic)—Both the electric and magnetic fields are transverse to the direction of propagation.
2. TE (transverse-electric)—The electric field is transverse to the direction of propagation.
3. TM (transverse-magnetic)—The magnetic field is transverse to the direction of propagation.

Linear polarized waves The components of the electric and magnetic fields in the transverse plane do not change in direction from instant to instant or point to point.

Vertical polarization The electric field is in the vertical direction.

Horizontal polarization The electric field lies in the horizontal plane.

Circular polarization Resultant electric field when equal amplitude horizontal and vertical polarized waves 90° out of phase are combined. The direction of this wave varies with time and distance. The wave is elliptically polarized if the horizontal and vertical waves are not equal in amplitude.

7·2 Engineering aspects of guiding systems

The exact solution of the fields and currents of a guiding system involves several vector quantities varying in both magnitude and direction as a

function of time and three-dimensional space. As would be expected, such a solution is at least a long and arduous process. Fortunately, however, it is rarely necessary to obtain an exact solution for field and current distribution. The factors of most interest in the transfer and distribution of electromagnetic energy in waveguides are the efficiency (losses), reflections (impedance), power-handling capacity, band width, network and special purpose application, physical aspects, and cost. To analyze these factors, or any other engineering problems, it is first wise to reduce the problem to the most simple form for solution. This is especially true in waveguide work due to the complexity of field theory. Actually, a large percentage of waveguide work concerns the application of transmission line theory which may be applied to waveguides almost without reservations. A reduction to simple circuit theory is possible in a number of instances. Although it is seldom necessary to actually derive field equations in waveguide work, it is quite important to have knowledge of the possible field and current distributions in the guiding system of interest in order to fully understand the problems that arise.

7·3 Elemental concept of the waveguide

Since it is difficult to attach physical significance to the mathematical procedure involved in determining the possible field and current distributions along the waveguide structure, we will attempt to justify the resulting field in

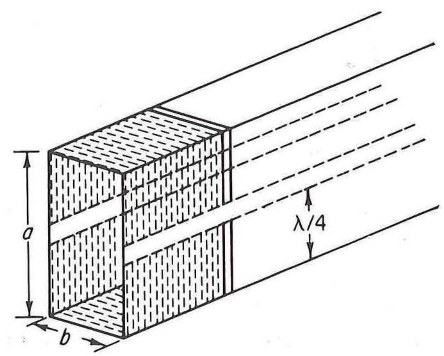


Fig. 7·1 Section of waveguide developed from an infinite number of elementary quarter-wave sections.

other ways. One method commonly used to derive the rectangular waveguide from the two-wire transmission line is illustrated in Fig. 7·1. The transmission line is supported by two quarter-wavelength sections, and since the input impedance of each section is theoretically infinite, they have no effect on the transmission of power. If the number of stubs is increased to infinity, the rectangular waveguide is formed as illustrated. It can be seen that the *a* dimension of the waveguide cannot be less than one-half wavelength. In fact, it must be slightly more than one-half wavelength in order to completely accommodate the transmission line function and at the same time preserve the insulating properties of the quarter-wave sections. Any frequency lower than that which makes the a dimension less than one-half wavelength will cause the circuit to become an inductive shunt and there is no propagation. The frequency at which the a dimension is one-half wavelength (free-space wavelength) is called the *cutoff* frequency and is designated f_c. The free-space wavelength associated with this cutoff frequency is the cutoff wavelength designated λ_c; ($\lambda_c = 2a$).

7·4 Advantages of hollow waveguides

There is no power loss by radiation from metal pipes of any type, including coaxial lines, if the ends are closed. Hollow waveguides are superior to coaxial lines in their ability to handle large concentrations of charges since these concentrations can be kept farther apart so that the electric field is less intense.

The construction of the waveguide is more simple than that of the coaxial cable since the inner conductor and its supports are eliminated. It is also more rugged and less susceptible to vibration and shock. Elimination of the insulating supports also results in a decrease of attenuation. Waveguides are usually air-filled, and for practical purposes they are considered to have no dielectric loss.

The current-carrying capacity is greater since, in practice, the waveguide is likely to have much greater conducting surface than a coaxial line. The overall power lost as heat in the walls of the waveguide is lower than the heat dissipated in the conductors of conventional size coaxial lines.

7·5 Reflections from a metal surface

The uniform plane wave is a highly idealized field pattern in which the electric and magnetic field wavefronts are parallel to the source. This idealized wavefront is never quite attained in practice but is assumed here in order to simplify the explanations. A great variety of complex field patterns is possible and may be classified in accordance with the shape of their wavefronts or equiphase surfaces. The locus of points at which the fields are in the same phase of variation is called the equiphase surface, and the direction of propagation is perpendicular to this surface. Consequently, the field varies from point to point except in the case of the idealized uniform plane wave.

Assume a plane wave incident upon a perfect conducting surface placed at an oblique angle as shown in Fig. 7·2a. The line which indicates the direction of propagation is referred to as the *incident ray* and is perpendicular to the equiphase surface. It intersects the conducting surface at the point P and makes an angle θ with the perpendicular at P as shown. Assume that the particular wavefront used for illustration represents a maximum equiphase surface of the wave, in which case, the lines of electric field intensity are

pointing out of the page. The *ray* of the reflected wave lies in the plane of the incident ray and is a normal to the boundary. The corresponding angles of reflection and incidence are equal. These equal angles θ are indicated in Fig. 7·2a. The ray of the incident and reflected waves is the free-space velocity of the electric and magnetic fields and is labeled v on the diagram. The angle between the incident wave and the conducting surface is also θ. The direction of the magnetic fields is shown, and all vectors of v, **E**, and **H** are properly directed to agree with the right-hand rule (Poynting's theorem).

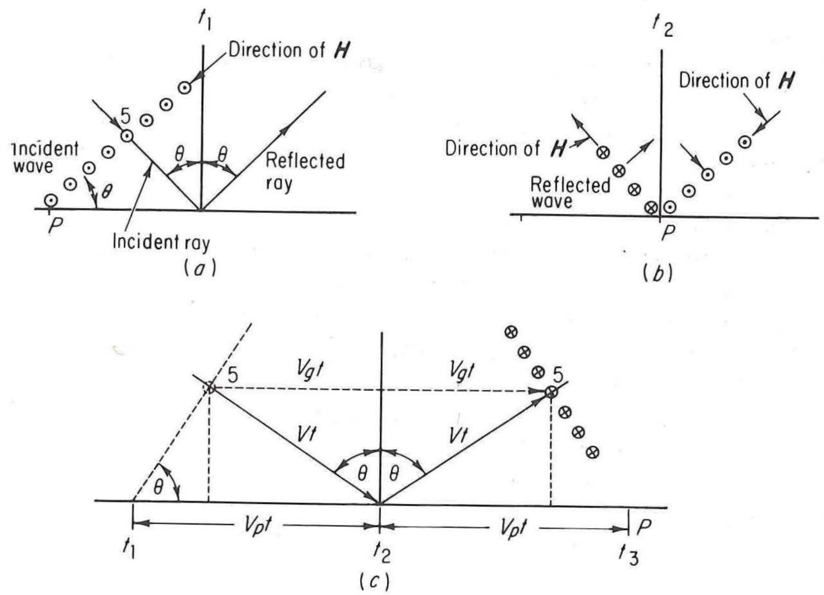


Fig. 7·2 Reflections from a conducting surface.

In order to satisfy the boundary conditions at the interface of a perfect conductor, the electric field is reflected without change in amplitude but with a reversal of phase. The tangential component of the magnetic field is reflected without change in amplitude or phase. This results in an electric force of zero everywhere along the interface and a tangential component of magnetic force equal to 2**H** at the interface.

The properties of this wave can be considered from two points of view. Assume that a nearsighted observer is located at P in Fig. 7·2a and that he wishes to single out the line of electric force labeled 5. This observer cannot see far beyond the point P, therefore he is unable to distinguish one line of force from another until the line of force is in the vicinity of some point P near the interface. The observer would have to travel parallel to the interface from the point P at t_1 in Fig. 7·2a to the point P at t_2 in Fig. 7·2b in order to single out line 5. The distance traveled is labeled $v_p t$ in Fig. 7·2c. The velocity parallel to the interface is called the *phase velocity* and is given by the relation

$$v_p = \frac{v}{\sin \theta} \tag{7·1}$$

This is the velocity at which the point of incidence of the wavefront moves along the surface. It is an *apparent velocity* since the actual velocity of the electric line of force parallel to the interface is less than the free-space velocity.

If the angle θ becomes smaller, this apparent velocity increases, and, when the wave approaches the surface at perpendicular incidence, the phase velocity approaches infinity. The most common analogy is that of a long ocean wave which approaches the shore at an angle θ with the shore. The point at which the wave strikes the shore moves parallel to the shore at a velocity considerably greater than the velocity with which the wave actually advances perpendicular to the wavefront. If the wavefront approaches the shore at a perpendicular angle, θ is zero and the velocity of the point of wave contact is infinite.

If a second observer is capable of singling out particular lines of force, a different view is obtained. If he observes line 5 from time t_1 in Fig. 7·2*a* to time t_2 in Fig. 7·2*b*, he notes that the line of force travels a distance v_t (Fig. 7·2*c*) and that the actual progress parallel to the interface is $v_g t$ as shown in Fig. 7·2*c*. v_g is the effective velocity with which energy is propagated parallel to the metal surface and is called *group velocity*.

$$v_g = v \sin \theta \tag{7·2}$$

This velocity approaches zero when the wavefront approaches the surface at perpendicular incidence.

7·6 Field patterns obtained by oblique reflections

The field patterns in front of the plane conductor placed at an oblique angle to the direction of wave propagation may be determined through use of principles set forth in the previous section. Figure 7·3*a* shows the structure of the incident and reflected waves in terms of the electric field. The (+) and (−) lines are at right angles to the direction of power flow and represent wavefronts or regions of constant electric field strength and, for the purpose of explanation, are assumed to be straight and parallel. The velocity of the wavefront is the free-space velocity. The dark solid lines represent zero voltage points of the incident and reflected waves. The incident wave is reflected from the reflecting surface with the same amplitude but with a reversal of phase. The total electric field is zero over the entire surface, thus satisfying the boundary conditions.

Along the $E_{\max}$ line the (+) incident and reflected waves are in phase and the (−) incident and reflected waves are in phase. These waves combine and produce a resultant wave which is *longer* than the applied wave and which

travels parallel to the surface of the conductor; along the $E_{\min}$ line the (+) and (−) incident and reflected waves are out of phase, and zero electric field exists at all points along the line. This process continues as indicated by the next $E_{\max}$ line, and it can be seen that a series of zero and maximum lines exists parallel to the metal surface.

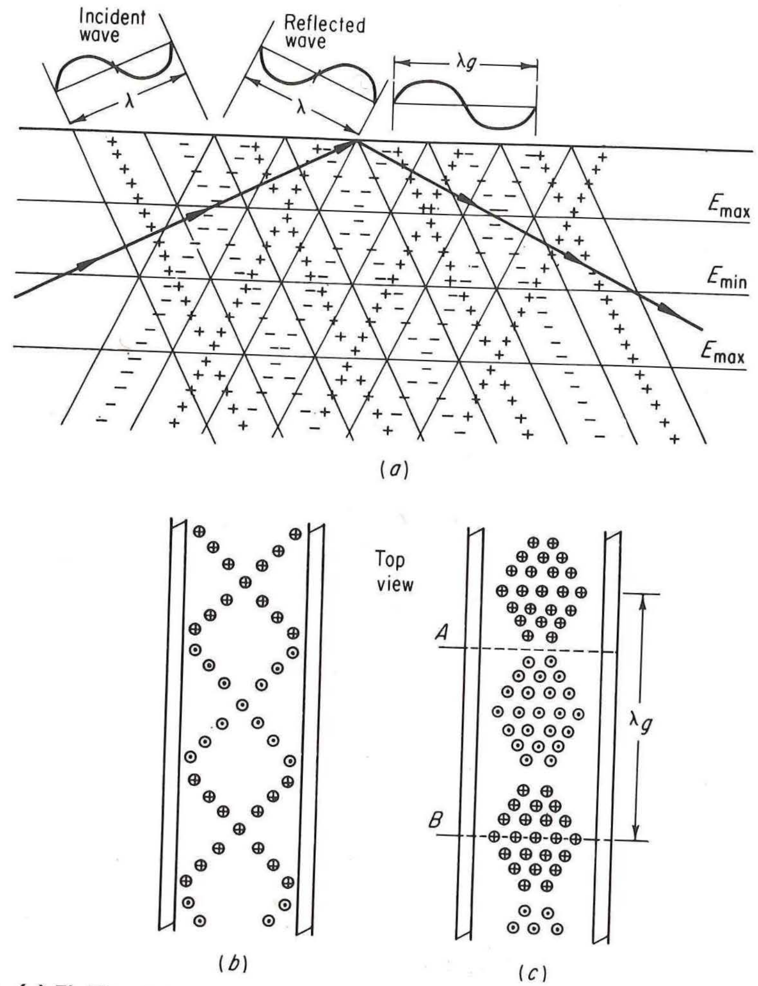


Fig. 7·3 (a) Field patterns obtained by oblique reflections from a plane conductor. (b) and (c) indicate the electric field distributions inside a waveguide.

Since a metal surface represents the zero voltage condition, it follows that wherever a zero voltage line occurs, a second metal-reflecting surface could be placed parallel to the first with no distortion of the wavefront pattern. It is apparent that many possible locations are available for the second metal surface, and each different location would result in a different field configuration. Each particular set of field configurations is referred to as a *mode*.

7·7 Waveguide transmission

Figure 7·3*b* represents the incident and reflected electric fields in a waveguide when viewed from the top. This is the same distribution between the surface and the first $E_{\min}$ as previously pointed out. The lines indicate the electric field distribution for the fundamental mode of operation. Each of the various sets of configurations can be considered as the resultant of a series of plane waves traveling with a velocity characteristic of the medium inside the guide, and all multiply reflected between opposite walls.

Figure 7·3*c* represents the electric field distribution in the waveguide. If we pass laterally across the guide at *A*, the instantaneous value of the electric field is everywhere zero. If we cross the guide at *B*, the electric field varies sinusoidally beginning at zero at either wall and reaching a maximum in the middle of the waveguide. If we travel along the *Z* axis of the guide, the electric field at any instant varies sinusoidally with distance. Thus, a *standing wave* is produced across the guide, and as time goes on, this standing wave moves down the guide. The peak value of this traveling wave is a function of the normal distance *x* from the reflecting surface and is given by

$$E = 2E_{\text{im}} \sin [(\beta \cos \theta)x] \tag{7·3}$$

($\beta \cos \theta$) is the phase constant of the standing wave in the *x* direction (across the waveguide). The peak value therefore varies between zero and $2E_{\text{im}}$ in accordance with variations of $\sin (\beta \cos \theta)x$.

The nature of the field patterns and the relationship of the free-space wavelength λ to the waveguide wavelength λ_g are illustrated in Fig. 7·4*a*, *b*, and *c*. The distance between the sides of the waveguide is labeled *a*. In Fig. 7·4*a* the free-space wavelength λ is long (low in frequency) and approaches the *a* dimension of the guide. The incident and reflected wavefronts make contact with the sides of the waveguide at a small angle θ. The rays (normal to the wavefronts) indicate that the wavefronts bounce back and forth across the waveguide many times in traveling down the guide. Therefore, power progresses down the guide at a velocity far below the free-space velocity with which the waves travel zigzag across the guide. From a previous discussion of a wavefront reflected from a metal surface, it was noted that the phase velocity of the wave became great as the angle θ became smaller. One-half of a waveguide wavelength $\lambda_g/2$ is indicated on the diagram by the intersections of the incident and reflected wavefronts and can be compared to $\lambda_g/2$ on the diagram of Fig. 7·3.

If the angle θ approaches zero, it can be seen that the phase velocity v_p approaches infinity and the waveguide wavelength λ_g approaches infinity. At the same time, the rays are normal to the two surfaces and the wavefronts

are bouncing back and forth without making any progress down the waveguide (the group velocity v_g becomes zero). The free-space wavelength λ is equal to $2a$ for this particular case. The frequency at which this condition exists is called the *cutoff frequency*, and the corresponding free-space wavelength is called the *critical* or *cutoff wavelength*. At wavelengths greater than cutoff no appreciable amount of power is propagated through the guide.

Figure 7·4*b* and *c* illustrates the configurations when the frequency is increased. The wavefronts approach the sides of the guide at a greater angle,

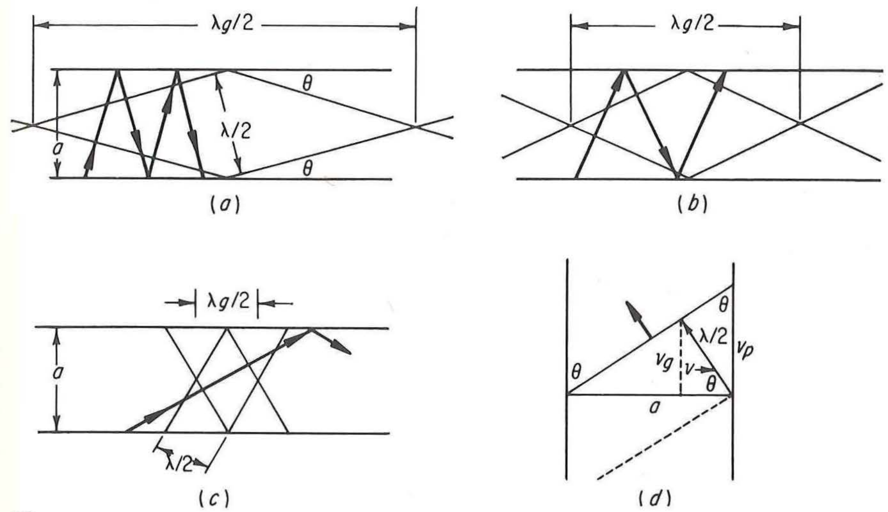


Fig. 7·4 Relationship of free-space wavelength and waveguide wavelength. Low frequency and the corresponding long wavelength is represented at (*a*). At (*b*), the frequency is higher than at (*a*) and lower than at (*c*). (*d*) Velocity and wavelength relationships.

the phase velocity decreases, λ_g decreases, the group velocity v_g increases, and there are less reflections from the sides of the guide. As the frequency is continually increased, the group velocity v_g and the phase velocity v_p approach the speed of light. The corresponding wavelengths approach the free-space wavelength.

The mathematical relationships related to waveguide transmission can be obtained from the geometrical relation between the free-space wavelength and the width (*a* dimension) of the waveguide as shown in Fig. 7·4*d*.

A right triangle is constructed with *a* and $\lambda/2$ as the sides. This illustration shows that

$$\cos \theta = \frac{\lambda}{2a} \tag{7·4}$$

The velocity at which any point of incidence of the wavefront moves along the guide is given by the relation

$$v_p = \frac{v}{\sin \theta} \tag{7·5}$$

and since

$$\sin \theta = \sqrt{1 - \cos^2 \theta}$$

$$\sin \theta = \sqrt{1 - \left(\frac{\lambda}{2a}\right)^2} \tag{7·6}$$

By substitution into Eq. (7.5) the phase velocity is

$$v_p = \frac{v}{\sqrt{1 - (\lambda/2a)^2}} \tag{7·7}$$

Also

$$v_p = \frac{v}{\sqrt{1 - (\lambda/\lambda_c)^2}} = \frac{v}{\sqrt{1 - (f_c/f)^2}} \tag{7·8}$$

The waveguide-wavelength λ_g equation is obtained from the triangle of Fig, 7.4*d* as follows

$$\sin \theta = \frac{v}{v_p} = \frac{\lambda}{\lambda_g}$$

and

$$\lambda_g = \frac{\lambda}{\sin \theta} = \frac{\lambda}{\sqrt{1 - (\lambda/\lambda_c)^2}} = \frac{\lambda}{\sqrt{1 - (f_c/f)^2}} \tag{7·9}$$

In the particular configuration just described, only the electric field was considered and this field was *everywhere transverse*. A complete account of transmission must also include a consideration of the lines of magnetic force. The magnetic component may be either longitudinal or transverse, depending on the point in the guide at which the observations are made.

Along the line *A* in Fig. 7·3*c*, the magnetic vector is a maximum near each wall and decreases cosinusoidally to zero at the middle of the guide. In this type of wave, the magnetic lines of force form closed loops, and the electric field extends from the upper to the lower walls of the guide. Along the line *B* in Fig. 7·3*c*, the magnetic force is zero at the walls and increases to a maximum at the center of the guide. The field configurations are shown in Fig. 7·5.

A wave of this type is of great importance, and it is convenient to think of it as a new type of wave rather than a combination of two plane transverse waves. This wave can take one of two forms, designated as TE or TM modes.

In the TE (transverse-electric) mode, the electric field is transverse to the direction of propagation. In the TM (transverse-magnetic) mode, the magnetic field is transverse to the direction of propagation.

In addition, subscripts are used to describe the electric and magnetic field configurations. The general symbol will be TE_{mn} or TM_{mn} where the subscript *m* indicates the number of half-wave variations of the electric field intensity along the *a* (wide) dimension of the guide. The second subscript *n* indicates the number of half-wave variations of the electric field in the *b* (narrow) dimension of the guide. The field distributions for the TE_{10} mode are shown in Fig. 7·5. The TE_{10} mode has the *longest operating wavelength* and is designated as the *dominant mode*. It is the mode for the lowest frequency that can be propagated in a waveguide. When the *a* dimension is less than one-half wavelength, there is no propagation down the guide, therefore the

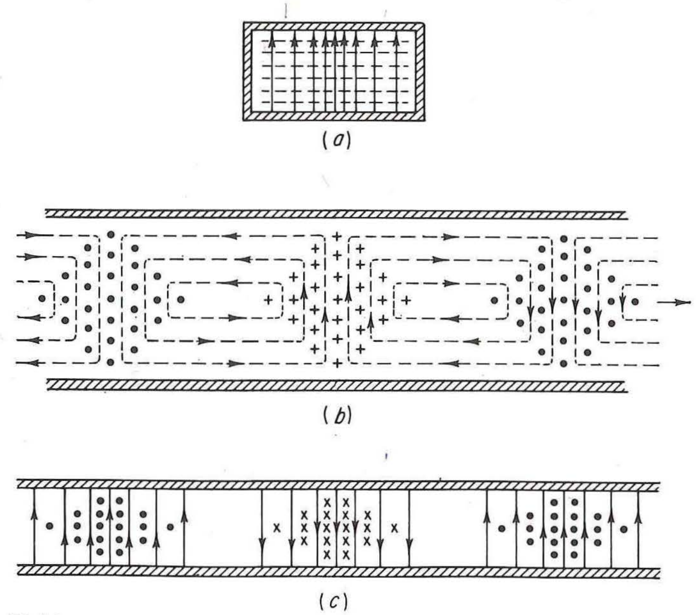


Fig. 7·5 Field configurations of the TE_{10} mode in the rectangular waveguide. (*a*) End view. (*b*) Top view. (*c*) Side view.

waveguide acts as a *high-pass filter* in that it passes all frequencies above a critical or cutoff frequency. For the standard rectangular waveguide the cutoff frequency is given as

$$f_c = \frac{2.998 \times 10^{10}}{2\sqrt{\mu'\epsilon'}} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} \tag{7·10}$$

$$\lambda_c = \frac{2}{\sqrt{(m/a)^2 + (n/b)^2}} \tag{7·11}$$

where *a* and *b* are measured in centimeters.

The magnetic field has a definite relationship to the currents in the waveguide walls. If the walls are perfect conductors, the current is confined to sheets of near zero thickness at the inside surface of the walls. The current lines are everywhere perpendicular to the magnetic lines at the conductor surface. The magnetic field at the surface of the waveguide is longitudinal, as indicated in previous considerations, and the flow of current is parallel to the b dimension as shown in Fig. 7·6.

This current pattern propagates as a wave in the direction of propagation. The components of current in the top and bottom of the guide are related to the longitudinal components of the magnetic field and can be accounted for by the fact that the charge concentration that must be present to allow the beginning and termination of the electric field lines must travel along the guide. The distribution of the wall currents is important since it is often necessary to cut small slots in the walls to sample the fields inside the guide.

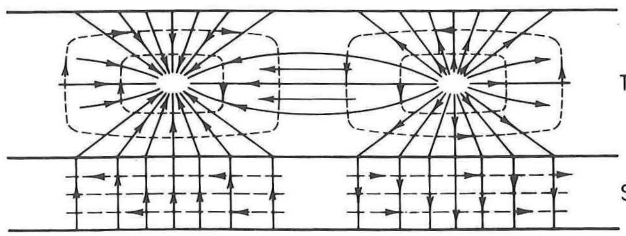


Fig. 7·6 Current distribution in the rectangular waveguide. The *dominant* (TE_{10}) *mode*.

These slots must be small in width and must be parallel to the direction of current flow in order to avoid disturbing the field inside the waveguide. The current path is completed by displacement current from top to bottom of the guide in the central regions shown in the illustration. The displacement current in the transverse plane is maximum where the electric field intensity is zero.

7·8 Higher order modes

A double infinity of higher order modes is possible in the rectangular pipe, as indicated by the integral number of half waves that may exist between the a and b dimensions. Practical use of waveguides is centered around the dominant mode, and these higher order modes are mainly of academic interest. In general, these waves are more highly attenuated and more difficult to recover from the waveguide system than the corresponding dominant wave.

If the source frequency is increased so that the a dimension is greater than one wavelength, the TE_{20} mode can exist. Also, half-wave distributions can be obtained in the b dimension resulting in, for example, the TE_{11} mode. The wave tends to remain in the dominant mode so that even though the source frequency may be high enough for higher modes to exist, they do not necessarily exist unless through deliberate or accidental distortion of the fields.

Higher order modes are present in the vicinity of a discontinuity where the fields are distorted. Since the waveguide dimensions, for the particular operating frequency, are such that the higher order mode cannot be propagated, these waves exist as fringing effects which become negligible at relatively short distances from their source.

Higher order modes are frequently given relative cutoff frequency values. These numbers are obtained by normalizing the cutoff frequency of the mode in question to the cutoff frequency of the TE_{10} mode in a rectangular waveguide which has a diameter equal to the a dimension of the rectangular waveguide. As an example, for the TE_{11} mode in a circular waveguide

$$\text{Relative cutoff frequency} = \frac{f_c(TE_{11})}{f_c(TE_{10})} = \frac{7{,}700}{6{,}557} = 1.17$$

Several of the higher order modes are shown in Fig. 7·8.

7·9 Practical operating range

The dominant mode propagation without the presence of higher order modes is usually desirable since, for a given excitation frequency, it has

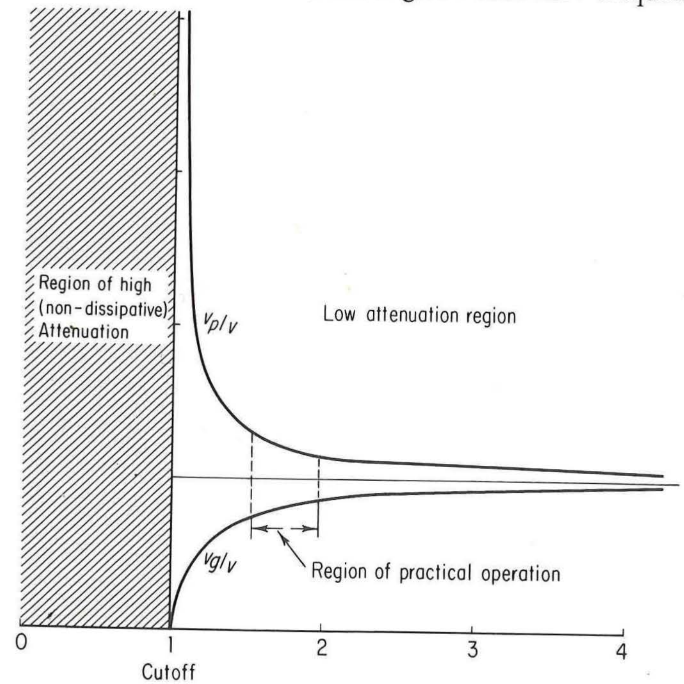


Fig. 7·7 Variation of phase velocity and group velocity with frequency.

lower power dissipation, requires smaller, lighter, and cheaper guiding structures, and requires simpler associated components.

The practical operating range of the TE_{10} mode in a rectangular waveguide where $b/a = 0.5$ is from 62 to 95 per cent of cutoff of the TE_{20} mode. Thus, the a dimension must be *greater* than $\lambda/2$ and *less* than λ.

The region of practical operation is shown in Fig. 7·7, which shows the relative phase velocity and group velocity for various conditions of waveguide operation.

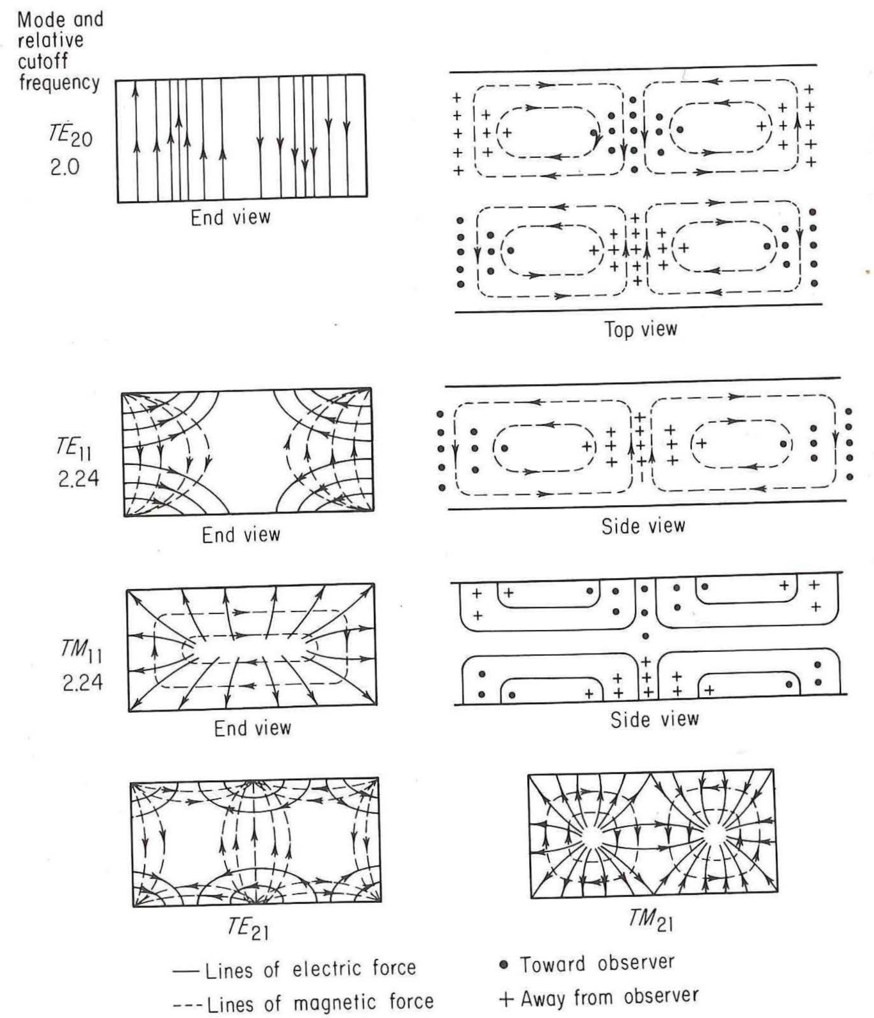


Fig. 7·8 Field configurations of higher order modes in the rectangular waveguide.

Reviewing again the previous analysis, we find that below cutoff the waveguide acts as a *nondissipative attenuator*, the group velocity is zero, and the phase velocity is infinite. As the operating frequency is increased far above the cutoff frequency, the phase and group velocities approach the free-space velocity.

7·10 Waveguide dimensions

Rectangular waveguides are usually chosen so that only the dominant mode exists over a certain frequency range. The above consideration determines the a dimension. The b dimension is important because of the following considerations:

a. The attenuation loss is greater as the b dimension is made smaller.

b. The b dimension determines the voltage breakdown characteristics and therefore determines the maximum power capacity.

The power transmitted by a waveguide has been determined from the Poynting vector evaluated at points on a cross section of guide. The total power transmitted is **P**.

$$\mathbf{P} = \tfrac{1}{4}\mathbf{E}_m\mathbf{H}_{xm}ab$$

and

$$\mathbf{H}_{xm} = \frac{\mathbf{E}_m}{Z_i}\frac{\lambda}{\lambda_g} = \frac{\mathbf{E}_m}{Z_i}\sqrt{1 - \left(\frac{\lambda}{\lambda_c}\right)^2}$$

therefore

$$\mathbf{P} = \tfrac{1}{4}\frac{\mathbf{E}_m{}^2}{Z_i}\sqrt{1 - \left(\frac{\lambda}{\lambda_c}\right)^2}\,ab \qquad (7.12)$$

The power capacity depends on the maximum electric field intensity, a, and b. Therefore, a and b should be made as large as possible for higher power capacity. In practice, the dimension b is usually chosen to be about one-half of a. The ratio of b/a is $\tfrac{1}{2}$. Standard waveguide dimensions are shown in Table 7·1. The accepted waveguide designations are given in Col. 1. The MDL letter designations are given in Col. 2. The WR-90 waveguide is usually referred to as *X band* as shown. However, the letter designations have not been standardized among the various manufacturers of microwave waveguide components.

7·11 Waveguide wave impedance

The *characteristic wave impedance* is analogous to the characteristic impedance of the two-wire and coaxial lines. The wave impedance represents the ratio of the electric to the magnetic fields, in which case, the electric field is the analogue of voltage and the magnetic field is the analogue of current.

The actual wave impedance is of little use and is seldom determined. There is no advantage in extending calculations beyond the determination of the

Table 7·1 Reference table of rigid rectangular waveguide data and fittings

Waveguide										
EIA designation WR( )	MDL designation ( ) band	JAN designation RG( )/U	Material alloy	Dimensions (in.)					Recommended operating range for TE_{10} mode	
				Inside	Tol.	Outside	Tol.	Wall thickness nominal	Frequency (kmc/sec)	Wavelength (cm)
2300	2300	...	Alum.	23.000–11.500	±0.020	23.250–11.750	±0.020	0.125	0.32–0.49	93.68–61.18
2100	2100	...	Alum.	21.000–10.500	±0.020	21.250–10.750	±0.020	0.125	0.35–0.53	85.65–56.56
1800	1800	201	Alum.	18.000–9.000	±0.020	18.250–9.250	±0.020	0.125	0.41–0.625	73.11–47.96
1500	1500	202	Alum.	15.000–7.500	±0.015	15.250–7.750	±0.015	0.125	0.49–0.75	61.18–39.97
1150	1150	203	Alum.	11.500–5.750	±0.015	11.750–6.000	±0.015	0.125	0.64–0.96	46.84–31.23
975	975	204	Alum.	9.750–4.875	±0.010	10.000–5.125	±0.010	0.125	0.75–1.12	39.95–26.76
770	770	205	Alum.	7.700–3.850	±0.005	7.950–4.100	±0.005	0.125	0.96–1.45	31.23–20.67
650	L	69 103	Copper Aluminum	6.500–3.250	±0.005	6.660–3.410	±0.005	0.080	1.12–1.70	26.76–17.63
510	510	...	...	5.100–2.550	±0.005	5.260–2.710	±0.005	0.080	1.45–2.20	20.67–13.62
430	W	104 105	Copper Aluminum	4.300–2.150	±0.005	4.460–2.310	±0.005	0.080	1.70–2.60	17.63–11.53
340	340	112 113	Copper Aluminum	3.400–1.700	±0.005	3.560–1.860	±0.005	0.080	2.20–3.30	13.63–9.08
284	S	48 75	Copper Aluminum	2.840–1.340	±0.005	3.000–1.500	±0.005	0.080	2.60–3.95	11.53–7.59
229	229	...	...	2.290–1.145	±0.005	2.418–1.273	±0.005	0.064	3.30–4.90	9.08–6.12
187	C	49 95	Copper Aluminum	1.872–0.872	±0.005	2.000–1.000	±0.005	0.064	3.95–5.85	7.59–5.12
159	159	...	...	1.590–0.795	±0.004	1.718–0.923	±0.004	0.064	4.90–7.05	6.12–4.25
137	X_B	50 106	Copper Aluminum	1.372–0.622	±0.004	1.500–0.750	±0.004	0.064	5.85–8.20	5.12–3.66
112	X_L	51 68	Copper Aluminum	1.122–0.497	±0.004	1.250–0.625	±0.004	0.064	7.05–10.00	4.25–2.99
90	X	52 67	Copper Aluminum	0.900–0.400	±0.003	1.000–0.500	±0.003	0.050	8.20–12.40	3.66–2.42
75	75	...	...	0.750–0.375	±0.003	0.850–0.475	±0.003	0.050	10.00–15.00	2.99–2.00
62	K_U	91 — 107	Copper Aluminum Silver	0.622–0.311	±0.0025	0.702–0.391	±0.003	0.040	12.4–18.00	2.42–1.66
51	51	...	...	0.510–0.255	±0.0025	0.590–0.335	±0.003	0.040	15.00–22.00	2.00–1.36
42	K	53 121 66	Copper Aluminum Silver	0.420–0.170	±0.0020	0.500–0.250	±0.003	0.040	18.00–26.50	1.66–1.13
34	34	...	...	0.340–0.170	±0.0020	0.420–0.250	±0.003	0.040	22.00–33.00	1.36–0.91
28	K_A	— — 96	Copper Aluminum Silver	0.280–0.140	±0.0015	0.360–0.220	±0.002	0.040	26.50–40.00	1.13–0.75
22	Q	— 97	Copper Silver	0.224–0.112	±0.0010	0.304–0.192	±0.002	0.040	33.00–50.00	0.91–0.60
19	19	...	...	0.188–0.094	±0.0010	0.268–0.174	±0.002	0.040	40.00–60.00	0.75–0.50
15	V	— 98	Copper Silver	0.148–0.074	±0.0010	0.228–0.154	±0.002	0.040	50.00–75.00	0.60–0.40
12	12	— 99	Copper Silver	0.122–0.061	±0.0005	0.202–0.141	±0.002	0.040	60.00–90.00	0.50–0.33
10	10	...	...	0.100–0.050	±0.0005	0.180–0.130	±0.002	0.040	75.00–110.00	0.40–0.27

* This is an MDL flange number.
SOURCE: Microwave Development Laboratories, Inc.

Cutoff for TE_{10} mode		Range in $\frac{2\lambda}{\lambda c}$	Range in $\frac{\lambda g}{\lambda}$	Theoretical attenuation lowest to highest frequency (db/100 ft)	Theoretical c-w power rating lowest to highest frequency (mw)	Fittings: Flange		EIA designation WR( )
Frequency (kmc/sec)	Wavelength (cm)					Choke UG( )/U	Cover UG( )/U	
0.256	116.84	1.60–1.05	1.68–1.17	0.051–0.031	153.0–212.0	...		2300
0.281	106.68	1.62–1.06	1.68–1.18	0.054–0.034	120.0–173.0	...	FA168A*	2100
0.328	91.44	1.60–1.05	1.67–1.18	0.056–0.038	93.4–131.9	...		1800
0.393	76.20	1.61–1.05	1.62–1.17	0.069–0.050	67.6–93.3	...		1500
0.513	58.42	1.60–1.07	1.82–1.18	0.128–0.075	35.0–53.8	...		1150
0.605	49.53	1.61–1.08	1.70–1.19	0.137–0.095	27.0–38.5	...		975
0.766	39.12	1.60–1.06	1.66–1.18	0.201–0.136	17.2–24.1	...		770
0.908	33.02	1.62–1.07	1.70–1.18	0.317–0.212 0.269–0.178	11.9–17.2	... ...	417A 418A	650
1.157	25.91	1.60–1.05	1.67–1.18	...	...	...		510
1.372	21.84	1.61–1.06	1.70–1.18	0.588–0.385 0.501–0.330	5.2–7.5	... ...	435A 437A	430
1.736	17.27	1.58–1.05	1.78–1.22	0.877–0.572 0.751–0.492	3.1–4.5	... ...	553 554	340
2.078	14.43	1.60–1.05	1.67–1.17	1.102–0.752 0.940–0.641	2.2–3.2	54A 585	53 584	284
2.577	11.63	1.56–1.05	1.62–1.17	...	...	...		229
3.152	9.510	1.60–1.08	1.67–1.19	2.08–1.44 1.77–1.12	1.4–2.0	148B 406A	149A 407	187
3.711	8.078	1.51–1.05	1.52–1.19	...	...	...		159
4.301	6.970	1.47–1.05	1.48–1.17	2.87–2.30 2.45–1.94	0.56–0.71	343A 440A	344 441	137
5.259	5.700	1.49–1.05	1.51–1.17	4.12–3.21 3.50–2.74	0.35–0.46	52A 137A	51 138	112
6.557	4.572	1.60–1.06	1.68–1.18	6.45–4.48 5.49–3.83	0.20–0.29	40A 136A	39 135	90
7.868	3.810	1.57–1.05	1.64–1.17	...	...	...		75
9.486	3.160	1.53–1.05	1.55–1.18	9.51–8.31 ... 6.14–5.36	0.12–0.16	541 FA190A* ...	419 FA191A* ...	62
11.574	2.590	1.54–1.05	1.58–1.18	...	...	...		51
14.047	2.134	1.56–1.06	1.60–1.18	20.7–14.8 17.6–12.6 13.3–9.5	0.043–0.058	596 598 ...	595 597 ...	42
17.328	1.730	1.57–1.05	1.62–1.18	...	...	...		34
21.081	1.422	1.59–1.05	1.65–1.17	... ... 21.9–15.0	0.022–0.031	600 FA1241A* ...	599 FA1242A* ...	28
26.342	1.138	1.60–1.05	1.67–1.17	... 31.0–20.9	0.014–0.020	... ...	383 ...	22
31.357	0.956	1.57–1.05	1.63–1.16	...	...	...		19
39.863	0.752	1.60–1.06	1.67–1.17	... 52.9–39.1	0.0063–0.0090	... ...	385 ...	15
48.350	0.620	1.61–1.06	1.68–1.18	... 93.3–52.2	0.0042–0.060	... ...	387 ...	12
59.010	0.508	1.57–1.06	1.61–1.18	...	...	...		10

per-unit impedance since the actual measurement of impedance in a waveguide results in a value of normalized impedance. The wave impedance equations for TE and TM modes are

$$Z_{\mathrm{TE}} = \frac{\eta}{\sqrt{1-(\lambda/\lambda_c)^2}} = \frac{\eta}{\sqrt{1-(f_c/f)^2}} \tag{7·13}$$

$$Z_{\mathrm{TM}} = \eta\sqrt{1-\left(\frac{\lambda}{\lambda_c}\right)^2} = \eta\sqrt{1-\left(\frac{f_c}{f}\right)^2} \tag{7·14}$$

where η is the intrinsic impedance of the medium (377 ohms for free space).

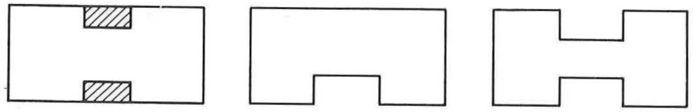
7·12 Comparison and summary of modes of propagation

1. The characteristic impedance and phase velocity are essentially independent of frequency in the TEM mode of propagation. Lossless transmission lines carrying TEM waves are nondispersive.
2. Propagation occurs in a waveguide only above a critical frequency called the *cutoff frequency*. Thus, a waveguide carrying a TE or TM wave may be considered to be a high pass filter as far as the particular mode is concerned. At frequencies far above cutoff, the propagation characteristics approach that of free space, i.e., the waveguide wavelength approaches the free unbounded wavelength in the medium contained inside the waveguide.
3. In the waveguide, the phase and group velocities and the wave impedance of TE and TM waves are functions of frequency. Therefore, the waveguide is a *dispersive* medium in that it tends to spread out waves of different frequencies which may be applied simultaneously at a common point. This dispersion phenomenon would cause distortion in waves which have modulation components extending over a wide frequency range.
4. TEM waves cannot exist in closed waveguides without inner conductors, but TE and TM waves may exist as higher order modes on two-conductor or coaxial lines if the operating frequency is sufficiently high.

7·13 Ridged waveguide

A ridged waveguide may be constructed by adding a longitudinal metal strip to the top and/or bottom of a standard rectangular guide as shown in Fig. 7·9*a*. The ridge acts as a uniform distributed loading and reduces the characteristic impedance of the waveguide and lowers the phase velocity. The reduction in phase velocity is accompanied by a lowering of the cutoff frequency of the TE_{10} mode by a factor that may be more than 5 to 1. At the same time, the cutoff frequency of the TE_{20} and TE_{30} modes is higher, depending on the ridge width. In this case, the ridge should be less than half the total a dimension of the guide for the TE_{20} mode and between one-third

and two-thirds of the total dimension. This increase in bandwidth is accompanied by an increase in the losses in the boundary walls and a decrease in power-handling capacity. By suitable variations of ridge dimensions, it is possible to vary the characteristic impedance of the guide by a factor of 25 or more and increase the attenuation by several hundred. The ridge waveguide is useful in certain coupling and matching requirements since the



(*a*) (*b*) (*c*)

Fig. 7·9 Single-ridge (*b*) and double-ridge waveguide (*a*) and (*c*).

characteristic impedance can be changed easily by gradually tapering the ridge.

7·14 Dielectric rod or slab guides

If the dielectric constant of a material is substantially higher than the surrounding space, the interface between the two media acts as the guiding discontinuity, and waves very similar to those propagated through metal pipes may be propagated through the rod or slab of dielectric material. This follows from the concept of total reflection by which a wave traveling in a dense dielectric can strike a boundary of less dense dielectric at an angle of incidence greater than a certain critical angle where all energy is reflected. At frequencies lower than this critical frequency the dielectric does not act as a perfect waveguide. For large diameters, the velocities in dielectric wires and metal pipes are substantially the same. There is no definite cutoff and power is propagated even when the diameter of the guide is relatively small. The necessary guiding discontinuity is not, for most materials, as definite as that provided by a metal surface, and, as a result, the fields associated with the transmitted power reside partly inside and partly outside the guiding medium. One of the most important practical uses of dielectric rod guides has been for radiation where the leakage along the rod is permitted in order to form an end-fire antenna array.

7·15 Circular waveguides

The circular waveguide is used in many special applications in microwave techniques. The circular guide has the advantage of greater power-handling capacity and lower attenuation for a given cutoff wavelength, but it has the disadvantage of somewhat greater size and weight. Also, the polarization of the transmitted wave can be altered due to minor irregularities of the wall surface of the circular guide, whereas the rectangular cross section definitely fixes the polarization.

The wave of lowest frequency or the *dominant mode* in the circular waveguide is the TE_{11} mode. The subscripts which describe the modes in the circular waveguide are different than for the rectangular waveguide. For the circular waveguide, *the first subscript m indicates the number of full-wave variations of the radial component of the electric field around the circumference of the guide. The second subscript n indicates the number of half-wave variations across a diameter.* Also, the second subscript indicates the number of diameters that can be drawn perpendicular to all electric field lines, and, in the case of TE_{0n} waves, it indicates the half-wave variation of the electric field across a radius of the guide. Illustrations of various circular waveguide modes are shown in Fig. 7·10.

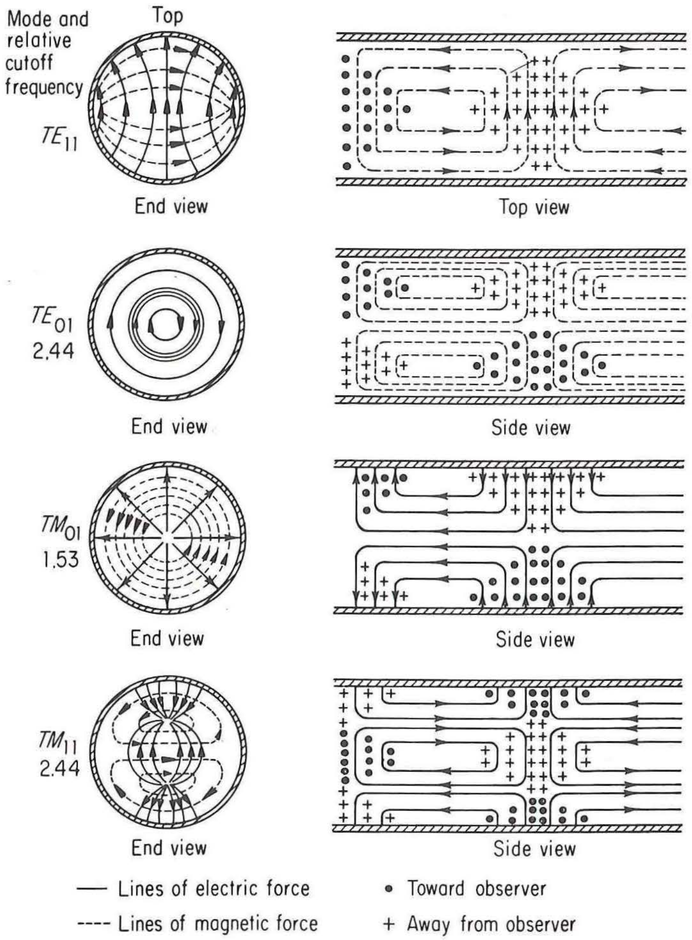


Fig. 7·10 Field configurations of modes in the circular waveguide.

The higher order modes which have properties of importance are the TM_{01} and TE_{01} modes. The TM_{01} mode has a circular symmetry which makes it adaptable for use in rotating joints. The TE_{01} mode has the property of decreasing attenuation as the frequency of the source is increased relative to cutoff. This mode is used for high-Q frequency meters.

The electromagnetic equations for circular waveguides usually appear as Bessel's functions. The theory of Bessel's functions is beyond the scope of this book but the functions may be used in our present requirement without any knowledge of the functions themselves. The cutoff wavelength and also the conditions for higher order modes are related to the waveguide radius a by way of the zeros of the particular Bessel function as designated by the subscripts of the mode being considered.

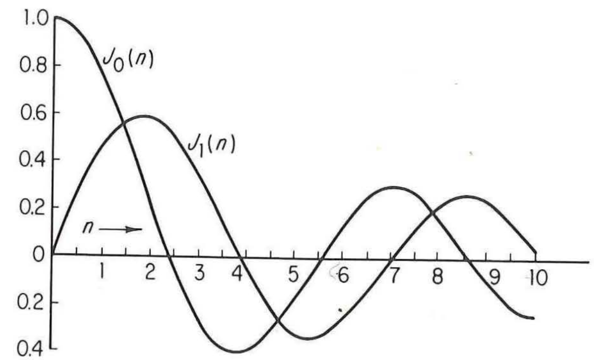


Fig. 7·11 The Bessel functions.

The approximate curves of the Bessel functions $J_0(n)$ and $J_1(n)$ are shown in Fig. 7·11. The TM_{mn} waves or modes may be supported when

$$\frac{2\pi a}{\lambda} = p_{mn}$$

where $J_m(n) = 0$ and p_{mn} is the nth root of $J_m(n) = 0$.

For TM_{0n} modes (the subscript m is zero)

$$\frac{2\pi a}{\lambda} = 2.405,\ 5.52,\ 8.65, \text{ etc.}$$

These values correspond respectively to the subscripts $n = 1, 2, 3$, etc., as previously defined in terms of the field configurations inside the guide. As an example, consider the TM_{02} mode in which case p_{mn} equals p_{02} equals the second root of $J_0(2) = 0$. This means that the Bessel function under consideration is $J_0(n)$ and that the value of $2\pi a/\lambda$ is found where this curve crosses the x axis the *second* time. The value is 5.52, which can be closely read from the curve.

For TM_{1n} waves (the subscript $m = 1$)

$$\frac{2\pi a}{\lambda} = p_{mn} \qquad \text{where} \qquad J_1(n) = 0$$

$$\frac{2\pi a}{\lambda} = 3.83,\ 7.02,\ 10.17,\ \ldots$$

$$n = 1, 2, 3, \text{ etc.}$$

For TE_{0n} modes or waves

$$\frac{2\pi a}{\lambda} = p'_{mn} \qquad \text{where} \qquad J_0'(n) = 0$$

$J_0'(n)$ is the first derivative of $J_0(n)$ and $J_0'(n) = J_1(n)$. Therefore, the values of $2\pi a/\lambda$ are the same for TE_{0n} and TM_{1n} modes. For TE_{1n} modes or waves

$$\frac{2\pi a}{\lambda} = p_1'(n) = 0$$

$$\frac{2\pi a}{\lambda} = 1.84,\ 5.33,\ 8.54,\ \ldots$$

and in Fig. 7·11 the approximate values are read where the slope of the $J_1(n)$ curve is zero.

The cutoff frequency and cutoff wavelength for circular waveguide modes are given by

$$\lambda_c(TM_{mn}) = \frac{2\pi a}{p_{mn}} \tag{7·15}$$

$$f_c(TM_{mn}) = \frac{p_{mn}}{2\pi a\sqrt{\mu\epsilon}} \tag{7·16}$$

$$\lambda_c(TE_{mn}) = \frac{2\pi a}{p'_{mn}} \tag{7·17}$$

$$f_c(TE_{mn}) = \frac{p'_{mn}}{2\pi a\sqrt{\mu\epsilon}} \tag{7·18}$$

7·16 Coupling to transmission lines

Several methods which can be used to couple energy from one transmission structure to another are shown in Fig. 7·12.

Magnetic coupling into a waveguide by means of a loop is shown at *a*. The loop may be placed at any position as long as it links with the magnetic lines of force. The degree of coupling can be altered by rotating the loop relative to the direction of the magnetic lines. The probe can sometimes be positioned where the magnetic field is weaker.

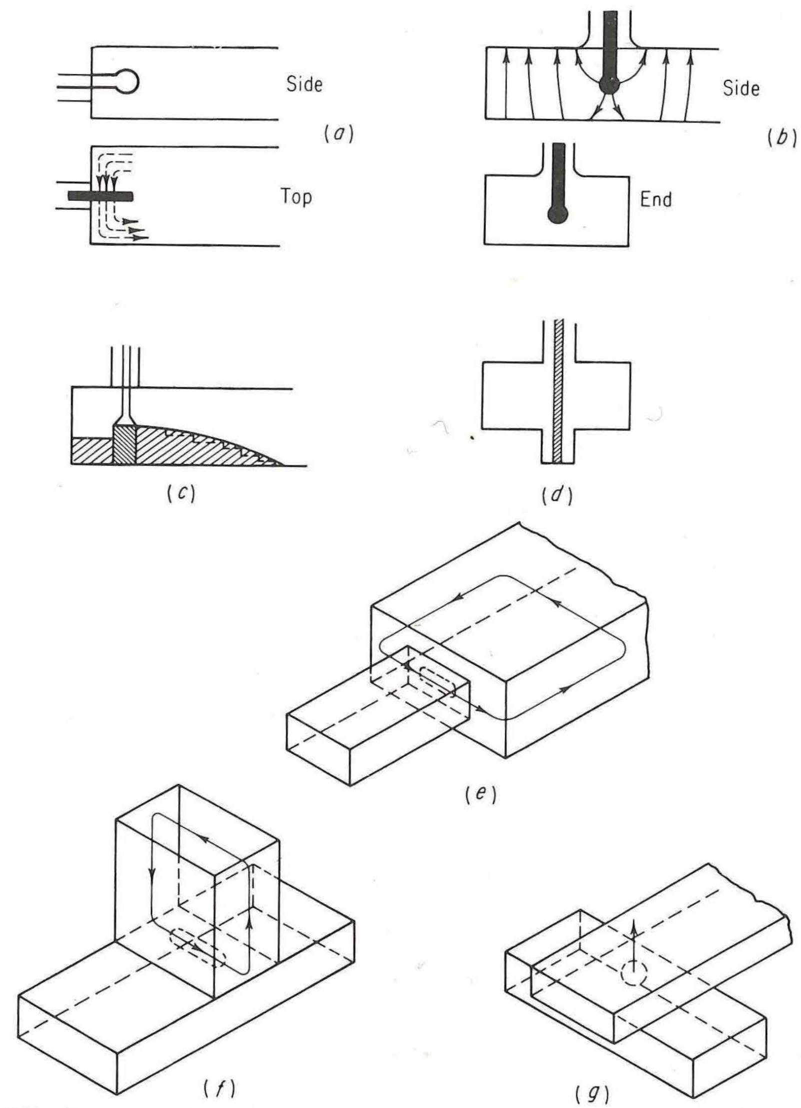


Fig. 7·12 Examples of coupling between transmission lines. (*a*) Side and top views. (*b*) Side and end views. (*c*) Tapered waveguide to coaxial transition. (*d*) Coaxial to waveguide transition. (*e*) Slot coupling. (*f*) Series slot coupling. (*g*) Transverse electric coupling.

The coaxial line may be coupled to the waveguide by extending the center conductor of the coaxial line into the waveguide parallel to the electric field at or near a point where the electric field has its maximum value. The conductor is located in the center of the guide as shown in Fig. 7·12*b*. The outer conductor terminates at the wall of the guide. The waveguide is usually closed at one end, and the probe is located one-quarter wavelength from the closed end so that the waves from the probe are reinforced by waves reflected from the closed end of the waveguide.

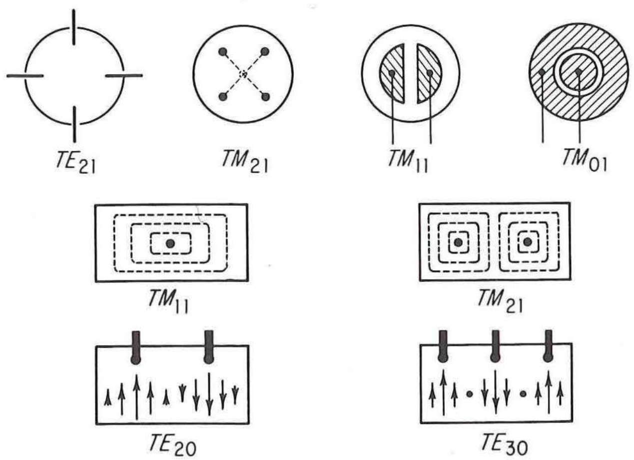


Fig. 7·13 Excitation of higher order modes in waveguides.

The coupling probe can be extended across the waveguide as shown at *d*. An adjustable short circuit can be connected to the center conductor beyond the junction so that the transition can be adjusted for best match over an appreciable frequency band.

The tapered transition at *c* is used when it is necessary to lower the waveguide impedance appreciably. The taper can be either a continuous or step type, as indicated by the dotted lines. This type transition is therefore useful for broadband applications.

There will be electric coupling through a round hole when there are components of electric field normal to the common surface as shown at *g*. There can be magnetic coupling if the magnetic field has a component parallel to the magnetic field at the adjoining surface. These conditions are sometimes determined by field distributions set up by certain types of restricted enclosures connected to the waveguide. Resonant cavities fall in this category.

Magnetic coupling occurs through narrow slots. The electric coupling is usually small, but magnetic coupling is appreciable.

There are many variations and combinations of the different types of coupling structures. Other types of coupling will be encountered in subsequent discussions.

Several methods[1] which have been used to excite higher order modes are shown in Fig. 7·13.

PROBLEMS

7·1 The *a* dimension of a rectangular waveguide is 0·9 in. and the *b* dimension is 0.4 in.

a. Calculate the cutoff frequency and cutoff wavelength of the waveguide

b. Calculate the wave impedance, phase velocity, and group velocity at 8.2, 9.0, and 12.4 Gc.

7·2 Plot curves of the wave impedance for TE and TM modes. Use $Z_{TE}/377$ and $Z_{TM}/377$ as the ordinate (wave impedance normalized to 377 ohms) and plot f/f_c as the abscissa. Compare the curves with Fig. 7·7.

7·3 Calculate the relative cutoff frequency for the TE_{21} and TM_{21} modes in rectangular waveguides. (Use WR-90 waveguide dimensions.)

7·4 Calculate the relative cutoff frequency for TE_{02} and TE_{12} modes in circular waveguides (relative to WR-90 waveguide).

7·5 Draw illustrations of TE_{01} and TE_{32} modes in a rectangular waveguide.

7·6 Draw illustrations of TE_{22} and TE_{02} modes in a circular waveguide.

7·7 Illustrate loop coupling and the associated magnetic fields in a rectangular waveguide with the loop located in the wide dimension. Repeat for a loop located in the narrow dimension. Show top and side views.

7·8 Make a complete comparison chart of the significant characteristics of TEM, TE, and TM modes. Six columns of the chart are as follows: Type of mode, cutoff frequency, type of line, velocity of propagation (v_p and v_g), attenuation (for a line of finite losses), and variation of characteristic impedance with frequency for lossless lines.

7·9 A K_u-band (WR-62) waveguide is filled with a material which has a dielectric constant of 2.4. Calculate the cutoff frequency and cutoff wavelength. Calculate the wave impedance at 12.4, 15.0, and 18.0 Gc.

7·10 Calculate the cutoff wavelength for the following circular waveguide modes in terms of *a*, where *a* is the radius of the circular guide.

a. TE_{02}, TE_{12}, TE_{11}, TM_{11}, TE_{03}, TE_{01}, TM_{01}, and TM_{12}.

b. Arrange the modes in order of increasing cutoff *frequency*.

Note: The relative cutoff frequencies of circular waveguide modes can be calculated from the cutoff wavelength values obtained in this problem. However, since relative cutoff wavelength and frequency require that the diameter of the circular waveguide be equal to the broad dimension of the rectangular guide (relative cutoff to TE_{10} mode in a rectangular guide), it is necessary to divide each λ_c calculation by 2 in order to obtain an equation in terms of the diameter.

Example:

$$\frac{\lambda_c(\mathrm{TE}_{10})\ \text{rectangular}}{\lambda_c(\mathrm{TE}_{11})\ \text{circular}} = \frac{2a}{3.142a/2} = 1.17$$

REFERENCES

1. G. C. Southworth, Some Fundamental Experiments with Waveguides, *Proc. IRE*, vol. 25, no. 7, July, 1937.

L. J. Chu, and W. L. Barrow, Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section, *Proc. IRE*, vol. 26, December, 1938.

CHAPTER

8

WAVEGUIDE ELEMENTS AND WAVEGUIDE COMPONENTS

8·1 Waveguide impedance-matching elements

Impedance-changing devices are introduced into the waveguide near the sources of reflected waves in order to eliminate standing waves. If diaphragms of good conductivity and of thickness small compared to wavelength are extended into the waveguide, the necessary susceptance can be introduced to reduce the standing waves to nearly zero as desired. Such diaphragms therefore play an important part in microwave techniques. The element illustrated in Fig. 8·1*a* adds inductive susceptance across the waveguide and is called an *inductive window* or *inductive iris*. The *capacitive window* or *iris* is shown at *b* and can be explained by the fact that the electric field in the vicinity of the diaphragm is mainly electric. The inductive iris has an advantage over the capacitive iris in high-power waveguide installations. If the combined effects of the inductive and capacitive windows are employed, the composite type illustrated in *c* and *d* is obtained and is referred to as the *resonant window* or *iris*. The circular iris is usually preferred in the circular waveguide. The circular iris, which provides inductive susceptance, is shown at *e*.

The tuning screw shown in *f* is capacitive, but if the screw is long enough and is inserted on into the guide, it becomes series resonant when the depth of penetration equals one-quarter of a wavelength and becomes inductive if the depth is made greater. When the screw is inserted across the guide, the inductive post in *g* is obtained. If the diameter of the post or screw is decreased, or if they are moved away from the center of the waveguide, the susceptance of the equivalent inductor or capacitor decreases. A considerable range of susceptance can be obtained by variation of the probe insertion and distance along the line by the structure shown in *h*. The illustrations in Fig. 8·1 are only representative of the many possible configurations and combinations of these elements.

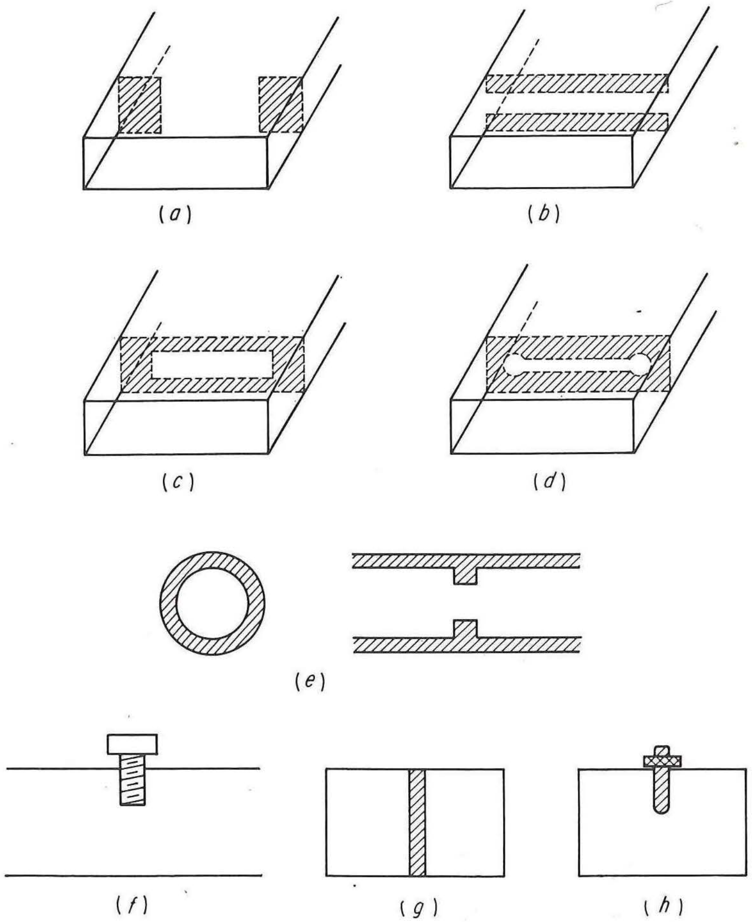


Fig. 8·1 Waveguide impedance-changing devices. (*a*) Symmetrical inductive iris. (*b*) Symmetrical capacitive iris. (*c*) and (*d*) Resonant windows. (*e*) Circular iris. (*f*) Capacitive tuning screw. (*g*) Inductive post. (*h*) Probe adjustable along the line with variable probe insertion.

8·2 Waveguide short circuit

One of the most useful waveguide elements is the waveguide closed at one end. It may be fixed closed at the end or an adjustable short-circuiting plunger may be provided. The adjustable short is more convenient and finds many applications in microwave techniques. A metal piston with contacting fingers riding against the sides of the waveguide is used in many applications, but a design which finds wide application at present is shown in Fig. 8·2 and is referred to as the *dumbbell* type.

8·3 Waveguide tees

Waveguide junctions play an important part in waveguide techniques; two of the common forms are shown in Fig. 8·3. The interesting properties

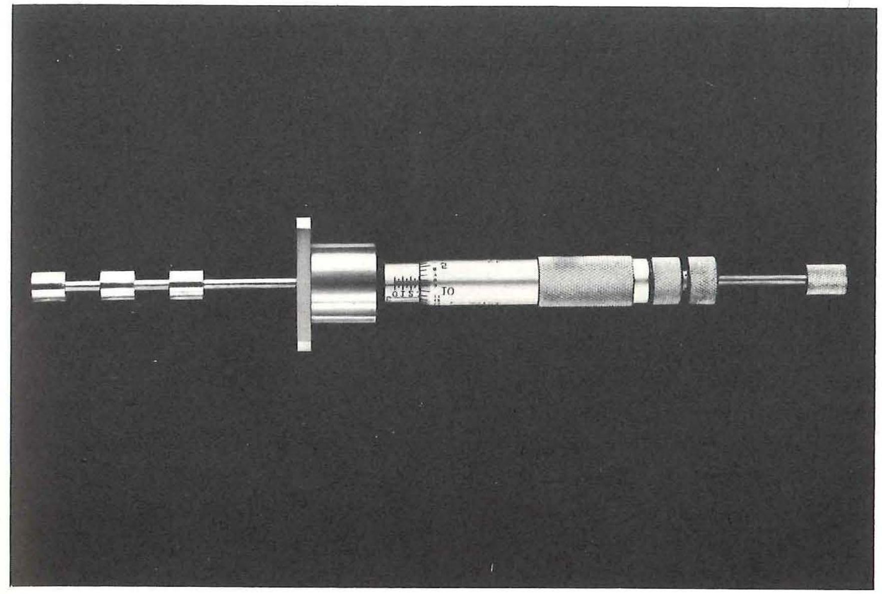


Fig. 8·2 Dumbbell type waveguide short circuit (quick-sliding and precisely adjustable). (*Hughes Aircraft Company, Microwave Standards Laboratory.*)

and characteristics of these junctions can be examined by considering the behavior of the electromagnetic fields in passing through the junctions. These junctions are referred to as waveguide *tees*, and they differ with regard to the plane in which the branch lies. The tee in Fig. 8·3*a* is called an **E** *plane* or *series tee* because the axis of the side arm is parallel to the E field of the main transmission line. The tee at b is called an **H** *plane* or *shunt tee* because the axis of the side arm is parallel to the planes of the magnetic field of the main transmission line.

The progress of the representative line of electric force of the wavefront is shown in the series tee. The line of electric force bends as it leaves the side

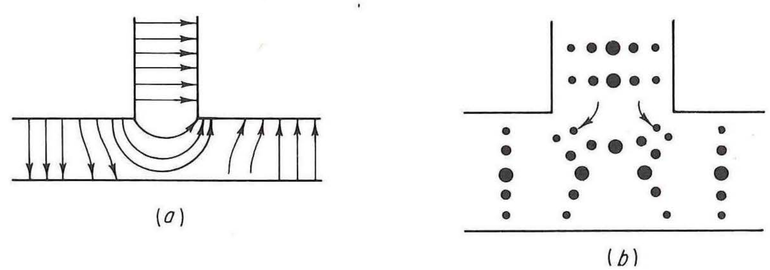


Fig. 8·3 Waveguide tee junctions. The E *plane* or *series tee* is shown at (*a*). (*b*) The H *plane* or *shunt tee*.

arm and causes the electric fields in the side arms to be of opposite polarity (out of phase). These waves set up in the side arms are equal in magnitude if the junction is completely symmetrical.

If the shunt tee is completely symmetrical, a similar analysis shows that a wave enters the **H** side arm and leaves the side arms equal in magnitude and in phase, as shown at *b* where a single cross section of electric lines of force is followed through the junction.

8·4 The magic tee

A combination of the **E**-plane tee and the **H**-plane tee forms a hybrid waveguide junction called the *magic tee*. The magic tee is illustrated in

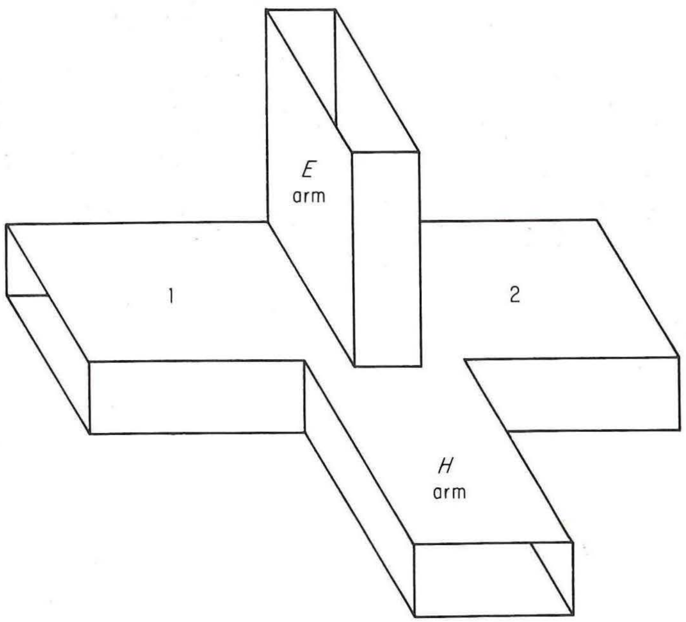


Fig. 8·4 The magic tee.

Fig. 8·4. The term magic tee is sometimes reserved for the hybrid junction in which matching structures have been introduced to improve the match of the **E** and **H** arms to the junction.

The characteristics of the hybrid in which the **E** and **H** arms are symmetrically placed are such that energy applied to either the **E** or **H** arm divides equally between arms 1 and 2, and no energy emerges from the opposite arm. If the fields entering arms 1 and 2 are equal in amplitude and of the same polarity, the net field in the **E** arm is zero, and the total energy emerges from the **H** arm.

At the junction each arm is effectively terminated by two other arms of equal impedance. Therefore, discontinuities will be inevitable unless special precautions are observed. Several forms of compensated junctions have been developed. One manner in which the magic tee can be matched is by the use of tuning rods. One rod is placed normal to the **E** field in the series arm, and one is placed normal to the **E** field in the shunt arm. The precise location and configuration of the various matching structures are usually determined empirically.

The magic tee can be used as a phase shifter when the series and shunt arms are terminated with adjustable short circuits. An extensive use of the magic tee has been in connection with microwave receivers where crystal detectors are placed in arms 1 and 2. The signal frequency enters the **H** arm, and the local oscillator signal is fed into the **E** arm. The tee provides isolation between sources of the two signals which are mixed in the crystals. The magic tee can be used in microwave phase-measuring systems and various microwave bridge circuits.

8·5 Phase shifters

The difference in phase shift between any two points is determined by the velocity of propagation and is therefore a function of the medium. There are several methods by which the effective velocity of propagation may be modified to introduce varying amounts of phase shift between two points. The phase may be changed by restricting the cross section of the guide, introducing inductive or capacitive irises into the guide, or by placing diametral rods across a circular guide. Our attention is mainly centered on commercial phase shifters of the variable type.

One method of obtaining a variable phase shift is to insert a dielectric vane into the waveguide. The vane may be attached to supporting rods and moved across the *a* dimension of the guide or the dielectric vane may be inserted through a slot in the top of the guide. The dielectric vane is properly tapered to give minimum reflection over the desired frequency range. This type of phase shifter usually has a dial gauge indicator and does not read phase shift directly.

A precision waveguide phase shifter which is direct reading in phase shift is the Hewlett-Packard Type 885A. It consists basically of three sections of round waveguide, each of which contains a plate of dielectric material. Rectangular-to-round transition sections provide the required waveguide input and output.

A functional drawing of the phase shifter is shown in Fig. 8·5*a*. The input and output differential phase sections are fixed in position at an angle of 45° in the waveguide. The center section is free to rotate. The input and output sections are called quarter-wave plates since an electric vector in the plane of the dielectric vane is delayed in phase by 90°. The central rotary vane is

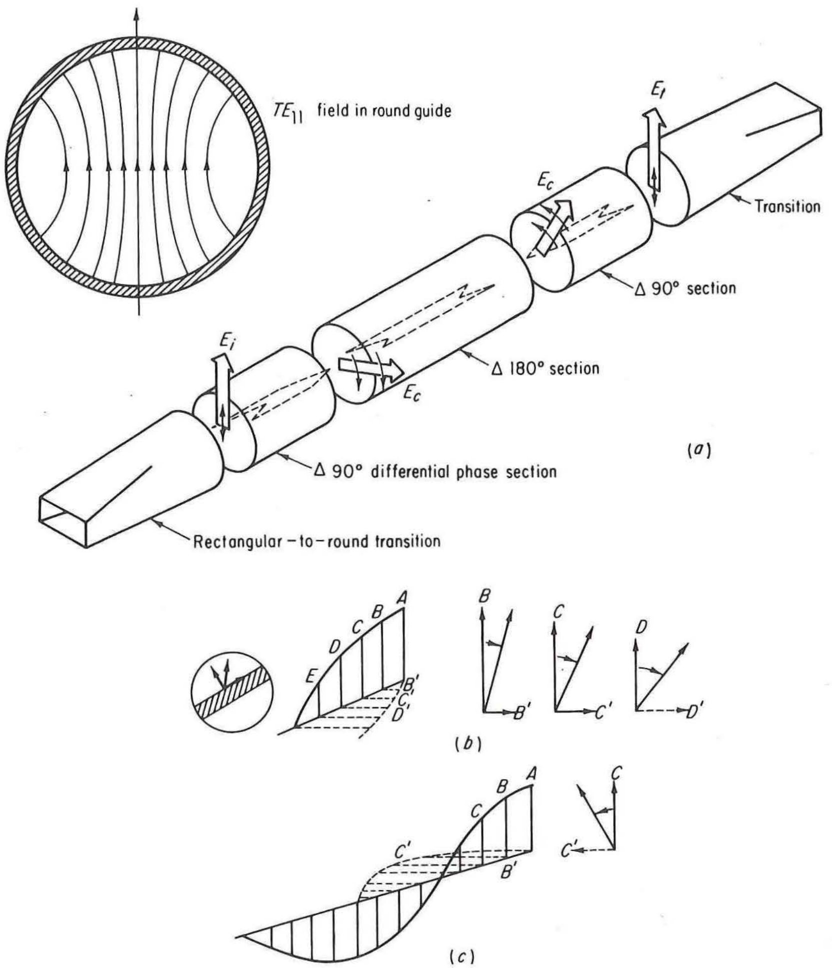


Fig. 8·5 Functional drawing of the Type 885A phase shifter. (*Hewlett-Packard Company.*) Clockwise rotation of the electric field due to the quarter-wave section is shown at (b). Counterclockwise rotation due to the half-wave section is illustrated at (c).

referred to as a half-wave plate since the electric vector in the plane of the dielectric vane is delayed by 180°.

The TE_{10} dominant mode in the rectangular waveguide is converted to the dominant TE_{11} mode in the circular waveguide. The linearly polarized electric vector is resolved into two mutually perpendicular vectors, one in the plane of the dielectric vane and the other perpendicular to the vane, as shown in Fig. 8·5*b*.

The representative illustration in Fig. 8·5*b* indicates the relative magnitudes

of the two waves emerging from the end plane of the quarter-wave plate. *A* indicates maximum amplitude in the vertical plane. *B-B'* indicates the magnitudes of the two emergent waves an instant later. The orientation at the output plane of the quarter-wave plate indicated that the original linearly polarized wave is now a *circularly polarized* wave rotating in a clockwise direction.

Figure 8·5*c* indicates a vector which has traveled through a half-wave plate, in which case the electric vector in the plane of the vane has been delayed by 180°. Applying the previous method of analysis at the output plane of the half-wave section, it is noted that the circularly polarized wave is now rotating in a counterclockwise direction.

The output quarter-wave plate converts the counterclockwise rotating wave back to a linearly polarized wave.

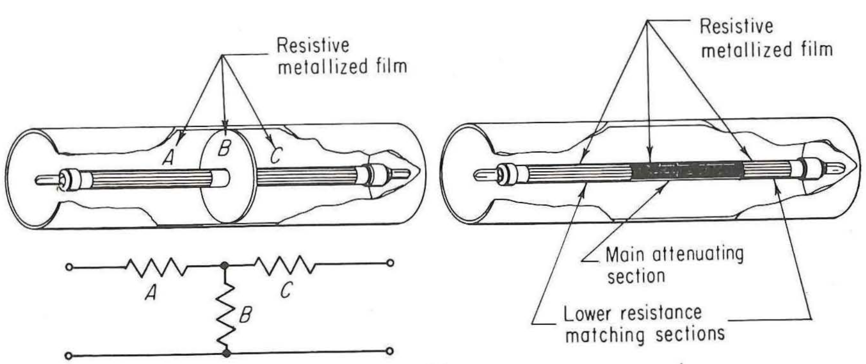
Rotation of the 180° section through an angle results in an angular displacement of the electric vector at the output of 2θ. Therefore, 180° mechanical rotation results in 360 electrical degrees phase shift between the input and output electric vectors.

8·6 Attenuators

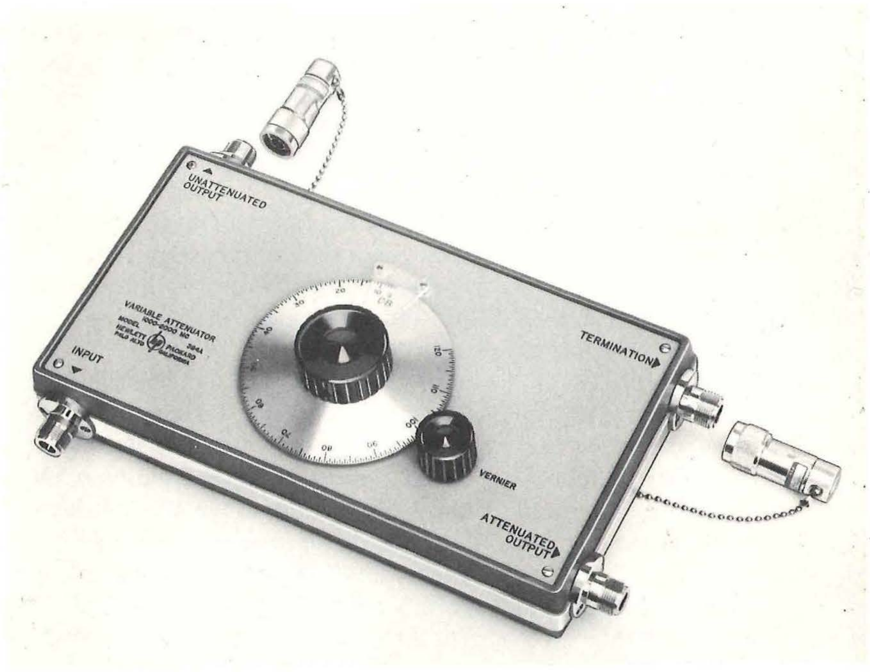
Attenuators are used to control the amount of power transferred between points on a transmission line by absorbing and/or reflecting some of the microwave power. Attenuators that operate on the reflection principle employ sections of waveguide-below-cutoff, and attenuators that operate on the power absorption principle use dissipative elements.

The attenuator placed in the transmission line should present a good impedance match at each terminal in order to decrease the dependence of the attenuator value upon the circuit in which the attenuator is used. Various techniques are used to obtain the desired impedance-match conditions. Other desirable properties have been obtained by selection of the materials used in the dissipative element and by the particular construction details of the dissipative element. One class of attenuators consists of thin metallic films coated on glass. A baked-on metallic film combining platinum and palladium and also an evaporated film of chromium or nichrome with a protective film of magnesium fluoride have been satisfactorily used. Glass was chosen as the base material because it does not react with the film, its surface is very smooth, it will not warp or change shape, and its melting point makes it usable for most processes.

Coaxial Attenuators. Two basic types of fixed coaxial resistive-film attenuators in common use are shown in Fig. 8·6*a*. The T-section attenuator is most useful in the frequency range below 4 Gc, and the distributed type is used for the upper frequencies. The center conductor is a glass tube which has two resistive metalized film sections on either side of a metalized circular disk. The distributed type attenuator consists of a higher resistive main



(a)

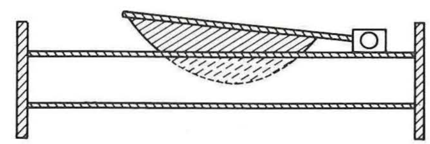


(b)

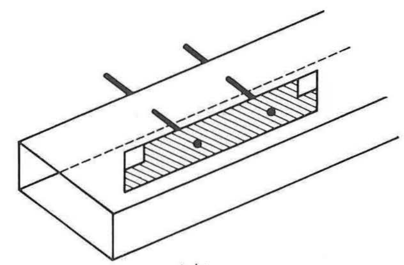
Fig. 8·6 Coaxial attenuators. (a) Basic types of resistive-film fixed attenuators. (*PRD Electronics.*) (b) Variable attenuator. (*Hewlett-Packard Company.*)

attenuation section in the center and lower resistance matching sections on each end. The resistance of the inner conductor is made high by maintaining the thickness of the film less than the skin depth of the film material. The attenuation value of this type of unit changes as a function of frequency. Therefore, if the precise value of the attenuator is required, the attenuator has to be calibrated at the particular operating frequency in question.

The variable coaxial attenuator in Fig. 8·6*b* has an attenuation range from 6 to 120 db. The instrument can be used as an attenuator or directional coupler or for mixing two signals. It is a type of waveguide-below-cutoff attenuator. The fields in a waveguide operating below cutoff attenuate exponentially with distance along the axis. The amount of power is varied by means of a movable element that is similar to the input coupling element. The plot of attenuation in decibels versus the displacement of the coupling



(a)



(b)

Fig. 8·7 Waveguide attenuators. (a) Flap attenuator. (b) Glass-vane attenuator.

elements is, except for close spacing, a straight line. If a waveguide-below-cutoff attenuator is designed to have a cutoff well above the frequencies at which it is to be used, the variation of attenuation with frequency will be negligible.

Waveguide Attenuators. The waveguide attenuator consists of an attenuation plate supported within the guide by means of metal rods or by moving the plate in and out of the guide at a fixed location. As an example of the latter, a resistance card is inserted through a longitudinal slot in the top (center) of the waveguide. This attenuator is referred to as a *flap* attenuator. Absorbing material is necessary next to the slot in order to lower the leakage.

Cross sections of the two types of waveguide attenuators are shown in Fig. 8·7. A drive mechanism is attached to the struts in order to move the vane in and out of the electric field to obtain the variable attenuation characteristic. Also, a dial mechanism is attached to provide a means of accurately calibrating the attenuator. Various methods are used to match the waveguide attenuators. The resistive transformer shown in the diagram is only one of

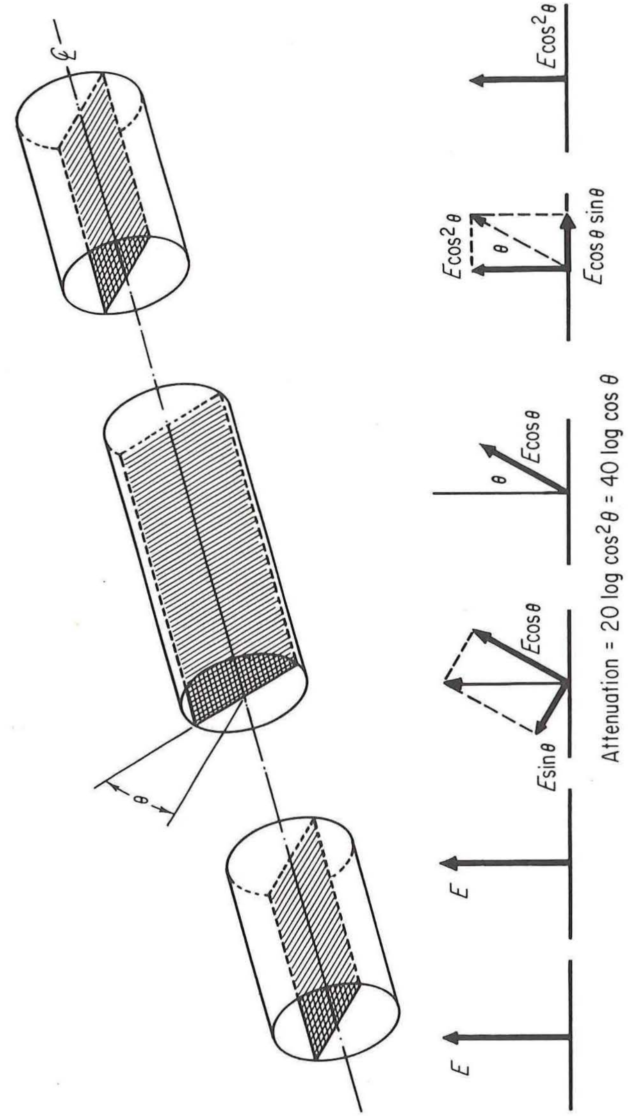


Fig. 8·8 Functional drawing including operating principle of the rotary-vane type attenuator. *(Hewlett-Packard Company.)*

several of its type, and, in addition to this method, a match can be obtained by tapering the vane at each end. A one-half-wavelength taper provides a good match. In addition to the frequency sensitivity of the variable glass-vane attenuator, considerable phase shift is encountered since the moving vane is a dielectric material. Typical curves for a complete line of glass-vane attenuators are found in the PRD catalog of microwave components.

Rotary-vane Attenuator. A direct-reading precision waveguide attenuator which obeys a mathematical law is the rotary-vane attenuator. The attenuator has a calibrated range of 50 db which is accurate within 2 per cent of the decibel reading at any frequency in the waveguide band. The phase-shift variation of the attenuator is less than one degree between 0- and 40-db variations. A functional drawing indicating the operating principle of this type attenuator is shown in Fig. 8·8.

Basically, the attenuator consists of three sections of waveguide in tandem. In each section, a resistive film is placed across the guide. The middle section is a short length of round guide which is free to rotate axially with respect to the two fixed end sections. The end sections are rectangular-to-round waveguide transitions in which the resistive films are normal to the field of the applied wave. The construction is symmetrical and the device is bidirectional.

When all films are aligned, the **E** field of the applied wave is normal to the films. No current then flows in the films, and no attenuation occurs. If the center film is now rotated to some angle θ, the **E** field can be considered to be split into two components: **E** $\sin\theta$ in the plane of the film, and **E** $\cos\theta$ at right angles to the film. The **E** $\sin\theta$ component will be absorbed by the film, while the **E** $\cos\theta$ component, oriented at an angle θ with respect to the original wave, will be passed unattenuated to the third section. When it encounters the third film, the **E** $\cos\theta$ component will be split into two components. The **E** $\cos\theta \sin\theta$ component will be absorbed, and the **E** $\cos^2\theta$ component will emerge at the same orientation as the original wave. The attenuation is thus ideally proportional only to the angle at which the center film is rotated and is completely independent of frequency. The attenuation, in decibels, is $40 \log \cos\theta$. The attenuation of the device is limited by the attenuation of the center rotating vane which normally has an attenuation value of 70 db or more.

8·7 Directional couplers

A useful form of hybrid known as a directional coupler is constructed by placing an auxiliary section of uniform waveguide along the narrow or wide dimension of the main guide with appropriately located apertures connecting the two.

The *Bethe hole* coupler consists of a single coupling aperture in the wide dimension of the guide. The electric field which leaks into the auxiliary guide is a combination of two fields, the fringing electric field through the hole

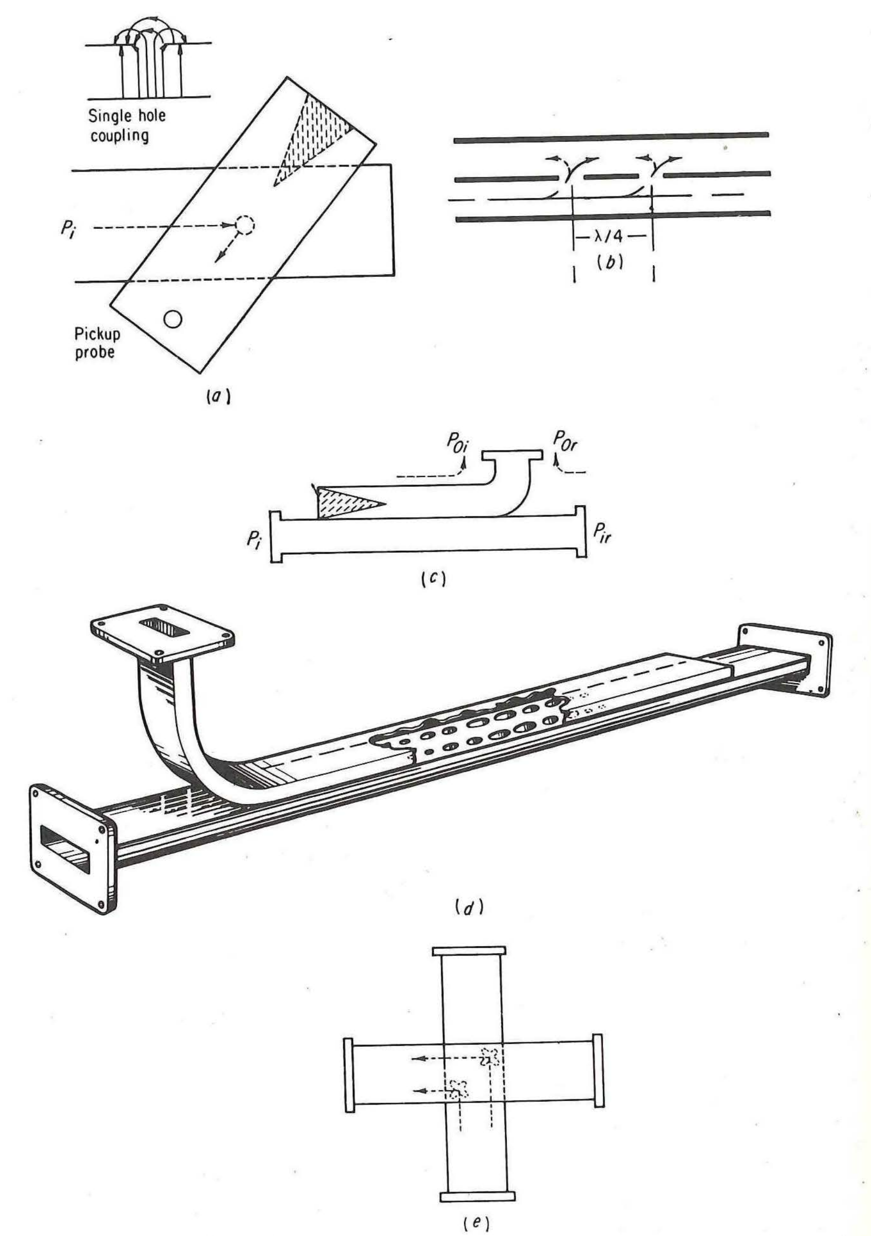


Fig. 8·9 Illustrations of waveguide couplers. (a) Bethe hole coupler. (b) Cross section of a two-hole coupler. (*Hewlett-Packard Company.*) (c) Side view of the multihole coupler. (d) Construction details of the multihole coupler. (*Hewlett-Packard Company.*) (e) Crossguide coupler.

and the electric field across the hole caused by the flow of charge out through the hole. This flow of charge is caused by the transverse component of the magnetic field of the wave. The directions of the component electric fields in the auxiliary guide are such that the fields tend to cancel in the forward direction from the aperture and reinforce in the opposite direction. The magnetic coupling is a function of the angle between the guide axis, whereas the electric coupling is essentially independent of the angle. Therefore, the maximum directional property can be obtained by an optimum setting of the angle between the auxiliary guide and the main guide, as shown in Fig. 8·9*a*. The coupling and directivity are sensitive to changes in frequency.

The two-hole directional coupler is illustrated in the cross-section diagram in Fig. 8·9*b*. The two auxiliary holes are placed one-quarter wavelength apart. The signal which travels back to the first hole from the second hole is 180° out of phase, and the two signals tend to cancel in the reverse direction. The signals in the forward direction are in phase and reinforce each other. The coupling and directional properties are impaired by off-frequency operation when the distance between the holes is no longer one-quarter wavelength.

The coupling factor is defined as the ratio, expressed in decibels, of the power entering the main line input to the power output of the auxiliary guide.

$$C = 10 \log \frac{P_i}{P_0}$$

From the definition and illustration of Fig. 8·9*c*, the directivity is given by

$$D = 10 \log \frac{P_{0i}}{P_{0r}}$$

Directivity is defined as the ratio, expressed in decibels, of the powers out of the auxiliary guide when a given amount of power is alternately applied in the forward and reverse directions in the main guide.

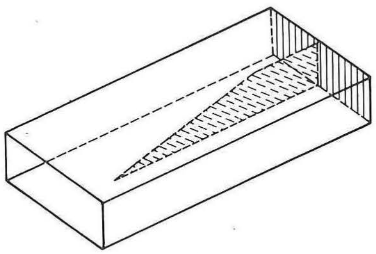
The leakage signal due to the intrinsic directivity of the coupling holes, the reflected signals from the internal termination, and any discontinuities at the output flange of the coupler determine the total directivity signal.

Multihole couplers operate on the same basic principle as the two-hole coupler. The coupling array is illustrated in Fig. 8·9*d*; it provides high directivity. The coupling versus frequency variations are ± 0.5 db over the entire waveguide frequency range. Coupling is obtained through a series of graduated holes which have been accurately machined along the broad or narrow face of the waveguide. Couplers designed for high-power operation usually have the holes in the narrow dimension of the waveguide.

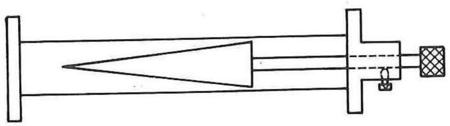
In applications where a multihole coupler is not required, the less expensive and more compact crossguide coupler shown in Fig. 8·9*e* is used.

8·8 Terminations (matched loads)

Matched loads or terminations are special devices designed to absorb incident energy without appreciable reflection. Terminations are constructed by mounting power-absorbing material in the space near the end of a closed section of waveguide. A list of low-power-absorbing materials includes powdered iron and a binder, carbon mixed with a binder and deposited on a strip of dielectric, and porcelain containing silicon carbide. At higher powers, graphite mixed with cement, and Aquadag-coated sand have been used.



(a)



(b)

Fig. 8·10 Low-power waveguide terminations. (*a*) Tapered resistance strip. (*b*) Movable polyiron termination.

Radiating fins are added to the waveguide casing to aid in removal of heat. Noncirculating and circulating water loads have been used extensively. Space is not available for illustrations of the various configurations and material forms such as the symmetrical taper in circular waveguides, the tapers in the electric and magnetic plane in rectangular waveguides, etc. Figure 8·10 illustrates two of the most commonly used low-power loads, the tapered resistance strip and the polyiron pyramid. The polyiron load is shown on a sliding-load assembly which finds many applications in microwave measurement techniques. The sliding termination is variable over at least one-half waveguide wavelength at the lowest waveguide operating frequency.

Mismatched loads or *standard reflections* are precision loads used to set up exact reflections for standardizing standing-wave and reflection-coefficient measuring setups. The *a* dimension of the precision casing is the same as the standard waveguide, but the *b* dimension is reduced the proper distance to

establish the required reflection at the junction. The precision-tapered load is movable so that reflections caused by the load can be isolated from the standard calibrated discontinuity.

8·9 Tuners

Tuners are used primarily for correcting discontinuities in a microwave system. The double-stub coaxial tuner was discussed in a previous chapter, and its limitations were pointed out. Tuning out mismatches with waveguide irises or tuning screws inserted into the guide was also discussed.

Three types of waveguide tuners are shown in Fig. 8·11. The slide-screw tuner consists of a slotted section of waveguide with a carriage on which is mounted an adjustable probe. The position of the probe along the line and the penetration into the waveguide can be adjusted to set up a reflection of the proper phase and amplitude to cancel out existing reflections in a system. VSWR values of 20:1 can be tuned out with this type tuner, and the insertion loss of a 20:1 mismatch is usually less than 2 db.

The **E-H** tuner is particularly useful where power leakage is undesirable. VSWR values of 20:1 result in insertion loss of usually less than 3 db. The series and shunt arms are terminated with adjustable short circuits which reflect the proper susceptance to the junction of the main guide.

The broadband five-stub tuner shown in Fig. 8·11 assures a high degree of measurement repeatability and does not have energy leakage.

8·10 Detectors (bolometers and crystals)

A *bolometer* is a temperature-sensitive, resistive element. The useful property of this device is that its resistance changes when it absorbs electromagnetic radiation (power), and the change in resistance is a function of the power dissipated. Bolometers are usually classed as *barretters* or *thermistors*. There is no strict adherence to the above classification since the most common usage of the term "bolometer" refers to the barretter class only. The barretter has a *positive* temperature coefficient (the resistance increases with an increase in temperature), whereas the thermistor has a *negative* temperature coefficient. The bolometer element is small compared to the wavelength of the microwave signal, and its resistance change is large enough to be measured accurately when small power changes occur. The change in barretter resistance is not the same for d-c and r-f heating because of lengthwise temperature distribution in the r-f case, but since the barretter and thermistor elements are small, the discrepancy between the r-f and d-c heating is negligible.

Barretters. The barretter consists of a thin Wollaston wire mounted in structures such as the Sperry Type 821 cartridge and the threaded Type 560 and 825 constructions shown in Fig. 8·12*a* and *b*. The coaxial type element holders are represented in Fig. 8·12*c*. The metal used in Wollaston wires is platinum. Barretters are normally biased to an operating resistance of from

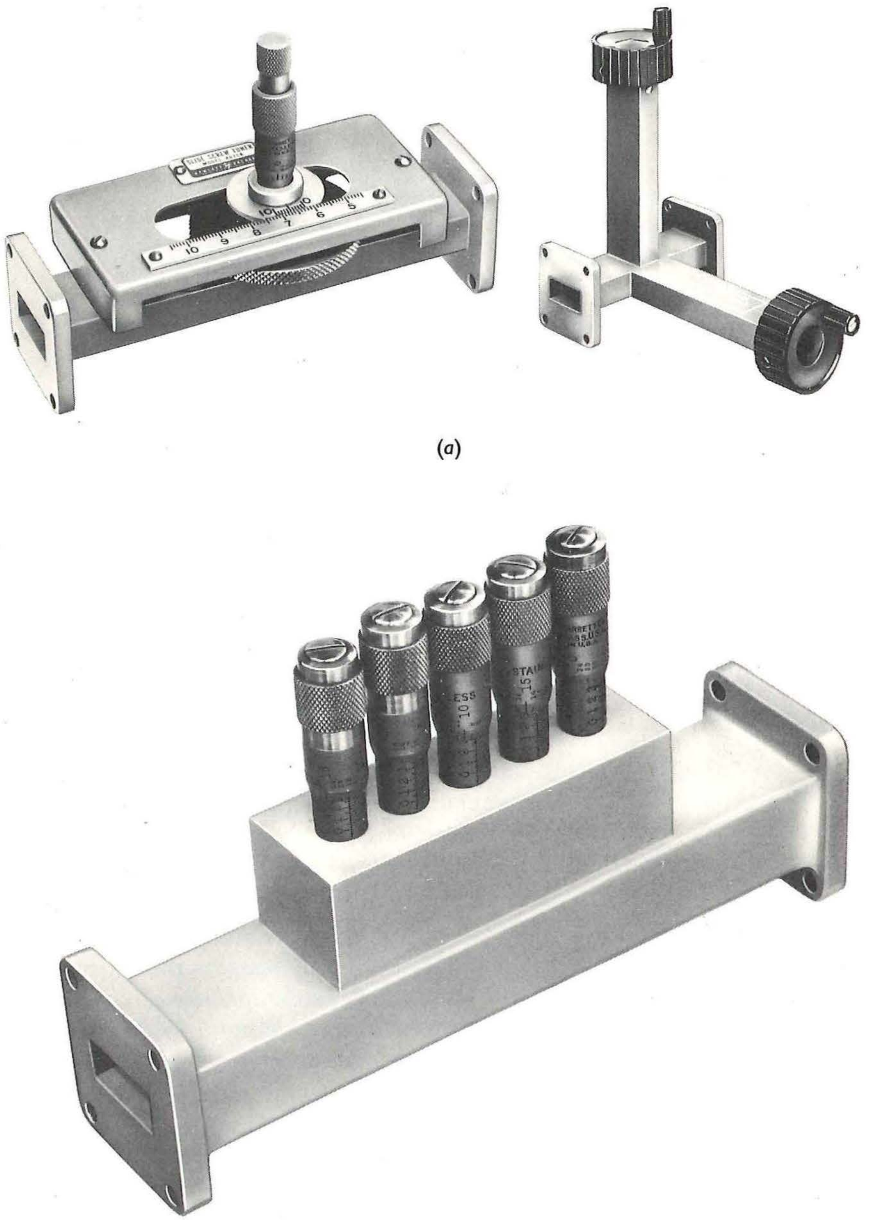


Fig. 8·11 Waveguide tuners. (*a*) Slide-screw tuner and E-H tuner. (*Hewlett-Packard Company.*) (*b*) Five-stub tuner. (*Hughes Aircraft Company, Primary Standards Laboratory.*)

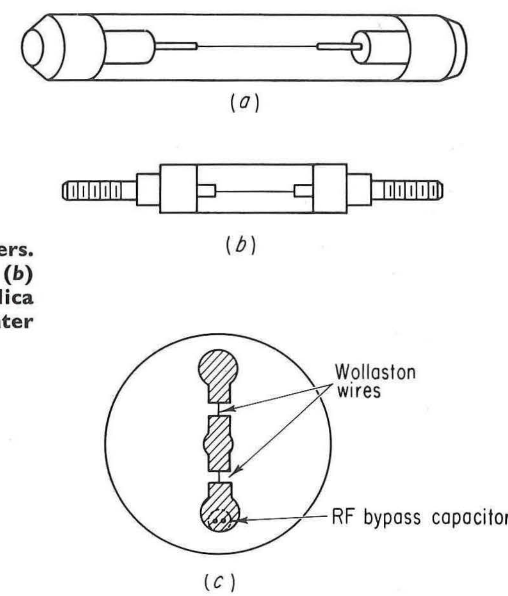


Fig. 8·12 Wollaston wire barretters. (*a*) Sperry Type 821 Barretter. (*b*) Sperry Type 560 and 825. (*c*) Mica disk (Series r-f bypass under center section).

50 to 400 ohms with 200 ohms operation being most common. A biasing circuit is shown in Fig. 8·13 where the barretter acts as a transfer device to convert r-f power change to audio-voltage changes when a constant bias current flows through the barretter. This is a common application in which power-ratio measurements are made with the r-f source 100 per cent modulated at an audio rate.

If the output voltage is proportional to the r-f power input, the detector is said to be *square law*. Since audio-voltage ratios can be measured with a high

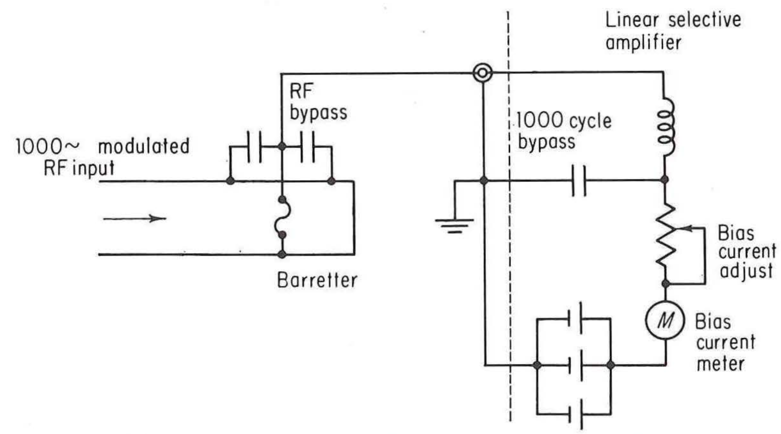


Fig. 8·13 Basic circuit of linear selective amplifier.

degree of accuracy, the accuracy of such a system depends mostly on the deviation of these devices from square law. If the bias-adjust resistance is very large compared to the 200 ohms of the barretter, the current through the barretter is relatively constant. Therefore, since $P = I^2R$, and in this case I is constant, the device is linear and the change in output voltage (due to change in barretter resistance with change in input power) is proportional to the change in the input r-f power.

The per cent deviation from square law is given by

$$\left(\frac{R_1 + \Delta R}{R_1} \times \frac{P_1}{P_1 + \Delta P} - 1\right)100$$

where R_1 is the barretter operating resistance, ΔR is the change in barretter resistance when microwave power is applied, P_1 is the d-c bias power, and ΔP is the microwave power.

The square-law device must be able to follow the r-f power change sufficiently fast in order to obtain a usable audio output. This ability is limited by the thermal time constant of the element. Typical time constants for Wollaston wires are from 80 to 300 μsec; this limits the modulation frequencies to values below 2,000 cps.

The range of r-f power ratio which can be accurately measured is limited at the upper end by the deviation of the detecting element from the square-law characteristic, and at the lower end by the noise which is produced by the detector itself and by the following amplifying system. By rule of thumb, the deviation of the barretter from square law is *one* per cent per 100 μw of applied microwave power. For accurate attenuation measurements, the ratio of minimum signal to noise should be at least 10:1, in which case the decibel range to 2 per cent deviation from square law is 46 db for the barretter when used with an amplifier which has a 25-cycle bandwidth.

Measurements on several types of barretters have indicated that the output voltage, across the element, ranges from 20 to 28 mv per microwatt of r-f power.

The sensitivity of the Wollaston wire barretter is usually in the range of 4.5 ohms per milliwatt with the sensitivity of some types approaching 10 ohms per milliwatt of applied power.

The characteristics of the barretter do not change appreciably up to the burnout point. The burnout power for the Sperry 821 barretter is approximately 32 mw and includes 15 mw bias power. The lower current PRD 610 barretter has a power rating of 7 mw, which includes approximately 4.5 mw bias power.

Thermistors. The *thermistor* is a temperature-sensitive resistor with a large *negative* temperature coefficient. It consists of a small bead of semiconducting material supported between two wires, as illustrated in Fig. 8·14. Thermistors

are made of complex metallic-oxide compounds using oxides of manganese, nickel, copper, cobalt, and sometimes other metals. The bead which is formed by these metallic-oxide mixtures is coated with a thin film of glass, which makes the assembly strong, heat resistant, and stable. The thermistor element is sometimes in the form of a thin film deposited on mica.

The thermistor is far more sensitive than the barretter and requires higher bias current. The thermistor is biased to 100 or 200 ohms. The sensitivity ranges from 40 to 140 ohms per milliwatt of applied power. The sensitivity decreases as the power dissipation is increased, and a large mismatch exists if the applied power is excessive. This high VSWR protects the thermistor when overloading occurs because a large portion of the applied power never

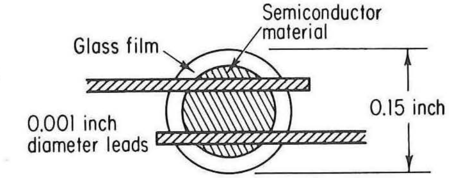


Fig. 8·14 Basic form of microwave thermistor bead.

reaches the thermistor element. The high sensitivity creates difficulties at low levels since the element responds to ambient temperature variations.

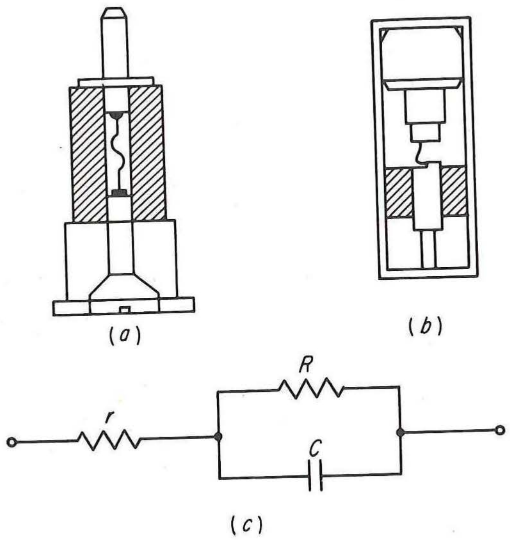
The thermal time constant of the thermistor is in the order of seconds. This can be an advantage or disadvantage depending upon the specific application.

Crystals. The crystal rectifier is the most sensitive and the simplest of all rectifying devices. It consists of a fine gold-plated tungsten wire (cat whisker) which is carefully pointed and brought into contact with a suitable semiconductor such as silicon or germanium. The volume surrounding the contact is usually filled with wax for mechanical stability and to prevent moisture from accumulating on the elements. The two types of physical structures shown in Fig. 8·15 illustrate the placement of the semiconductor and cat whisker between the base and contact prong. The cat whisker and semiconductor are reversed in position for reversed polarity types. The simplified equivalent circuit of the crystal is shown in Fig. 8·15*c*.

Microwave crystal diodes are designed as nonlinear circuit elements for frequency conversion, rectification, modulation, detection, and harmonic generation.

When crystals are used in mixer applications, the degree of crystal performance is defined by the noise figure, noise ratio, i-f impedance, r-f impedance, and conversion loss. The crystal noise is a combination of thermal noise, barrier noise caused by the diode rectification action, and noise known as fluctuation noise. The noise is high in the audio range and approaches thermal noise above 500 kc. The noise ratio of a crystal excited with 0.5 to 1 mw of c-w power may be in the range of 3 to 1 (referenced to the noise power generated by a 300-ohm resistor excited by the same amount of power).

The *i-f impedance* is the impedance of the diode and holder when looking from the output terminals while the diode is excited at a microwave frequency. The *r-f impedance* is the impedance looking into the r-f terminals of the mixer at the local oscillator frequency and power level. The i-f and r-f impedance are functions of the crystal geometry, the crystal holder, and the impedance



R = nonlinear resistance of the barrier (about 5,000 ohms for very low levels)
r = ohmic spreading resistance caused by the restricted current flow between the metal whisker and the semiconductor (10 to 20 ohms)
C = barrier capacitance (0.1 to 0.5 $\mu\mu$f)

Fig. 8·15 Microwave crystals. (*a*) Ceramic type cartridge. (*b*) Coaxial type structure. (*c*) Simplified equivalent diode circuit.

under excitation. Therefore, the impedance of a crystal rectifier is of significance only in terms of the circuit in which it is measured.

The conversion loss is defined as the ratio of available r-f power input to the measured i-f power output at the mixer. The conversion loss equation is

$$L = 10 \log \frac{(\Delta P)^2}{(\Delta I)^2 2PR_L}$$

where L = conversion loss, db
ΔP = change in power level, mw
ΔI = change in crystal current, μa
P = average power level, mw
R_L = load resistance

Crystal harmonic generators are useful for a variety of reasons. They are widely used in frequency-multiplying chains in the generation of standard frequency signals. The source power is introduced into the crystal mounted in a waveguide structure appropriate to the input frequency. The output waveguide is provided with tunable shorting plungers which provide accurate tuning of the desired harmonic and minimization of other harmonics which are simultaneously present in the output. The amplitude of the harmonic is much smaller than that of the fundamental and is a function of the fundamental power level and the increasing order of the harmonic used. The conversion loss increases rapidly with harmonic number.

One of the major applications of microwave crystals is video detection. The performance criteria used in determining the suitability of a detector are simplicity and response time, video impedance or video resistance, tangential signal sensitivity, and square-law response.

The following material is devoted to the characteristics of the crystal as a low-level detector since we are mainly concerned with the problems associated with the detection of microwave power.

The video impedance is the dynamic impedance of the crystal diode considered as a constant-current generator in the microwatt region and is determined by measuring the current flow when 5 mv is applied. If the video impedance is too high, the pulse shape is distorted because of splitting of the high-frequency components.

The sensitivity depends upon the nature of the semiconductor material and the contact area. Crystals are frequency sensitive, and the magnitude of this variation in output signal is further related to the applied power. The sensitivity decreases at high power levels. Typical video characteristic curves for type 1N31 and 1N32 crystals indicate that the open-circuit output is in the range of 4,000 to 5,000 μv output per microwatt input. The sensitivity decreases with increasing frequency and, in order to obtain the output level at a constant value in the range of 3,000 to 10,000 Mc, may result in input power changes as great as 100:1.

The d-c output resistance R of the crystal is not constant but decreases from a value of several thousand ohms at very low levels to a few hundred ohms at higher levels. Therefore, the matching of a crystal to a line is difficult to maintain. The crystal can suffer damage from overload, and the resulting changes in characteristics can cause inaccuracies of measurements.

The *minimum detectable signal* and *tangential sensitivity* are used as measurement levels for video diodes. The minimum detectable signal is the amount of signal power, below a 1-mw reference level, which exists when its presence is barely discernible in the noise. The tangential sensitivity is the amount of signal power, below a 1-mw reference level, required to produce an output pulse whose amplitude is sufficient to raise the noise fluctuation by an amount equal to the average noise level and is approximately 4 db above the minimum detectable signal.

The pulse response of a crystal video detector is a function of the crystal itself and the video output circuit. The rise time is controlled by the applied microwave power level, and the decay time is controlled by the external load resistance.

The burnout power is a complicated function of the particular crystal, its impedance match—whether c-w or pulsed, duty cycle, pulse width, and the particular load circuit. With the appropriate mixer circuit, pulse width, and duty cycle, some units can be subjected to as much as 25 watts peak power. For most video detection the c-w power level should be limited to 300 mw maximum.

The forward response of the crystal diode obeys the square law (rectified current is proportional to input power) in the region below approximately 10 μw. In some crystals the deviation from square law may take place at levels as low as 1 μw. The termination impedance is significant in determining the operational characteristics of the crystal. A certain amount of control may be exercised over the deviation from square law by proper selection of the load resistance. If the crystal is used with an amplifier which has a 25-cycle bandwidth, the decibel range from 10 db above noise level to 2 per cent deviation from square law is about 34 db.

For accuracies in the order of 0.2 db, the crystal diodes are capable of measuring r-f power ratios over a 38-db range with a maximum r-f power input of 0.1 μw average below approximately 5 Gc. This can be compared to the barretter which is capable of covering a range of 53 db with the same accuracy but requires a maximum input power of 1 mw average.

Crystals are not too well suited for high precision measurements. In general, high accuracy requirements call for alternate detection techniques. When a crystal is employed, it becomes of utmost importance that the crystal be operated in a closely controlled microwave and video environment in order to minimize errors. Also, operation must be checked as a part of the measurement.

8·11 Detector mounts

The usefulness of the microwave-detecting elements is intimately related to the design of the mounting structure. The mounting structure should be arranged to transform the r-f impedance of the element to equal the impedance of the reflections. For greatest usefulness the structure should maintain this condition (low VSWR) over as broad a frequency range as possible. The importance of obtaining a match is indicated by the fact that approximately one per cent of the input power is reflected from a VSWR of 1:2.

If the detecting element is connected directly across the open end of the waveguide, it turns out that there is a conductance component associated with radiation into open space and a susceptance component associated with reflections from the open end of the waveguide. The radiation is eliminated and the open circuit is obtained at the detecting element by attaching a closed quarter-wave section beyond the detecting element.

A broadband thermistor mount is illustrated in Fig. 8·16*a*. The transmission line impedance is matched to the resistive component of the

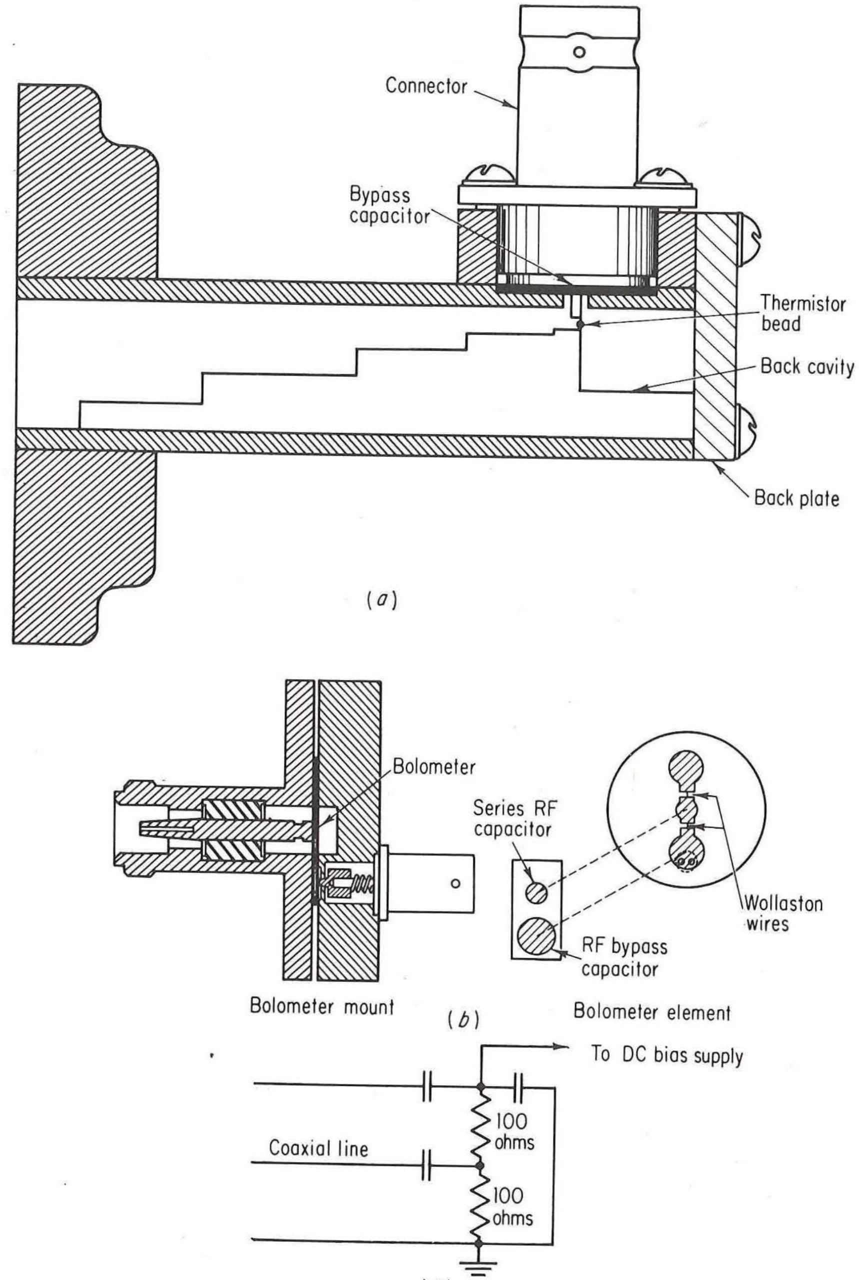


Fig. 8·16 Bolometer mounts. (*a*) The PRD broadband thermistor mount. (*b*) Essential details of the PRD Type 627-A coaxial bolometer mount and bolometer element. (*c*) The equivalent d-c circuit of the Type 627-A bolometer mount. The d-c bolometer resistance is 200 ohms, and its r-f resistance is 50 ohms. (*PRD Electronics.*)

thermistor by the ridged waveguide in the form of stepped transformers. The impedance can also be decreased to the appropriate value by use of a tapered ridged waveguide. A coaxial barretter mount is shown in Fig. 8·16*b* along with the equivalent circuit. It consists of the Wollaston wires and bypass capacitors mounted on a thin mica disk. The tunable detector mount in Fig. 8·17 accommodates the Sperry 821 or Narda 810-B barretter and crystals such as the type 1N23.

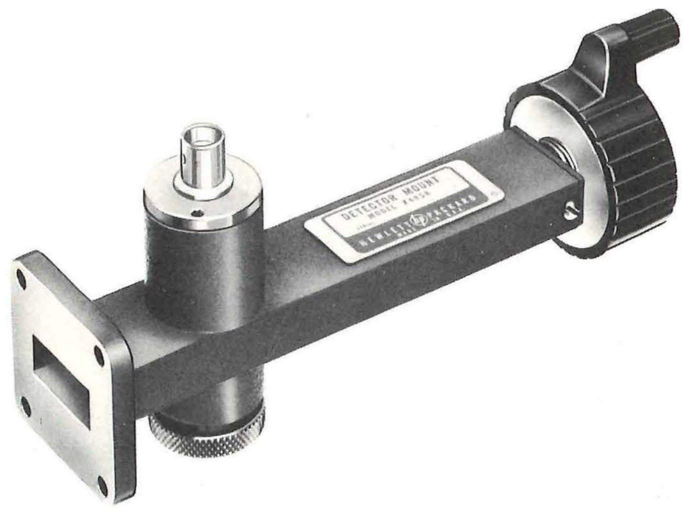


Fig. 8·17 Tunable detector mount. (*Hewlett-Packard Company*.)

In addition to evaluating the error due to loss of power by reflection in thermistor and barretter mounts, it is also necessary to determine to what extent the available power is absorbed by the detecting element and what fraction of the available power is dissipated by stray losses in the mount. This is a measure of the *efficiency* of the mount and is determined by comparing the power values obtained when a known amount of d-c power and a known amount of r-f power are alternately applied to the mount. An accurately calibrated d-c bridge is used to measure the d-c power entering the detecting element. The efficiency would be 100 per cent if all the power entering the mount were used in heating the bolometric element. The efficiency is usually given as

$$\eta = \frac{\text{power absorbed by the bolometer}}{\text{power delivered to the bolometer mount}}$$

The efficiency of commercial bolometer mounts may range from 95 to 98 per cent.

8·12 Short-slot hybrid[1]

The short-slot hybrid junction is a four-port device in which the power input at one port divides equally between two other terminals. If these two

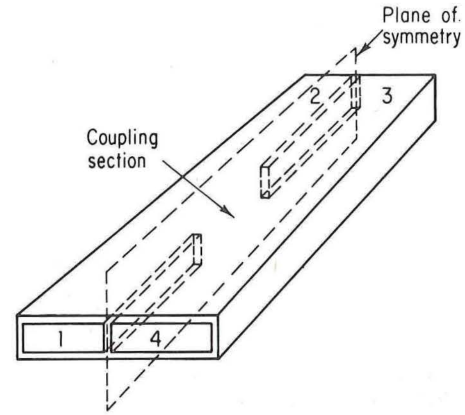


Fig. 8·18 The short-slot hybrid.

output terminals are perfectly matched, the energy reaching the fourth terminal (port) will be zero, and the structure becomes an ideal hybrid junction.

A schematic diagram of the hybrid is shown in Fig. 8·18 in which a plane of symmetry is indicated for the full length of the device. Two waveguide sections are placed side by side with a portion of the common wall removed to permit coupling between the two sections. Under suitable conditions, the power entering port 1 will have divided equally toward ports 2 and 3 by the time the energy reaches the end of the coupling section. The output voltages at port 3 lead the voltage at port 2 by 90°. The output voltage at port 3 leads by 45° what it would be if there were no slot in the waveguide. The short-slot hybrid has many applications in specialized waveguide circuits such as power splitters, phase splitters, phase shifters, couplers, balanced mixers, antenna feeds, bridge circuits, and duplexers.

8·13 Solid-state microwave switches

There are many applications which require rapid switching or modulation of microwave power. The solid-state switching diodes[2] can be used to obtain switching functions at nanosecond speeds. The switches are compact and lightweight and require only milliwatts of driving power. The control of microwave power is obtained by variation of a d-c bias which changes the impedance of the crystal diode. The present low-power applications (less than 1.5 watts) include ON-OFF switching of microwave power, amplitude modulation, electronic controlled attenuators, pulse shaping, and clipping. With improved application techniques and power-handling capabilities, the diode can be employed in phased antenna array systems, transmit-receive devices, etc.

The equivalent circuit of the crystal diode switch is shown in Fig. 8·19. The switching action is achieved by alternately applying forward and reverse bias to the diode. The explanation of switching action is given for the resonant frequency operation with the diode designed so that $X_L = X_C = X_{C_s}$.

Forward Bias. Consider the crystal diode connected across the waveguide. When the diode is forward biased, the nonlinear barrier resistance approaches a very low value and shunts the barrier capacitance. The inductive reactance of the lead is much greater than the spreading resistance. At the operating frequency, $X_L = X_{C_s}$ and the diode appears as a parallel resonant circuit. The crystal diode is in the ON condition, and nearly all the incident power is delivered to the load.

Reverse Bias. When reverse bias is applied to the diode, $X_{C_s} \gg r$ and the nonlinear barrier resistance is much greater than the barrier capacitance ($R \gg X_C$). The barrier capacitance shunts the large resistance and resonates with the lead inductance. The diode impedance is the small spreading resistance r shunted across the waveguide. Most of the incident power is reflected, and nearly all the power that is not reflected is absorbed in r. Very little power reaches the load, and the switch is in the OFF or isolation condition.

In the reverse bias condition OFF, the diode is represented by the spreading resistance r. In the forward bias condition ON, the diode is represented by $r(1 + Q^2)$. $Q = X_L/r$.

The shunt-connected diode is the most convenient waveguide configuration. The approximate insertion loss and isolation for the shunt-connected diode are

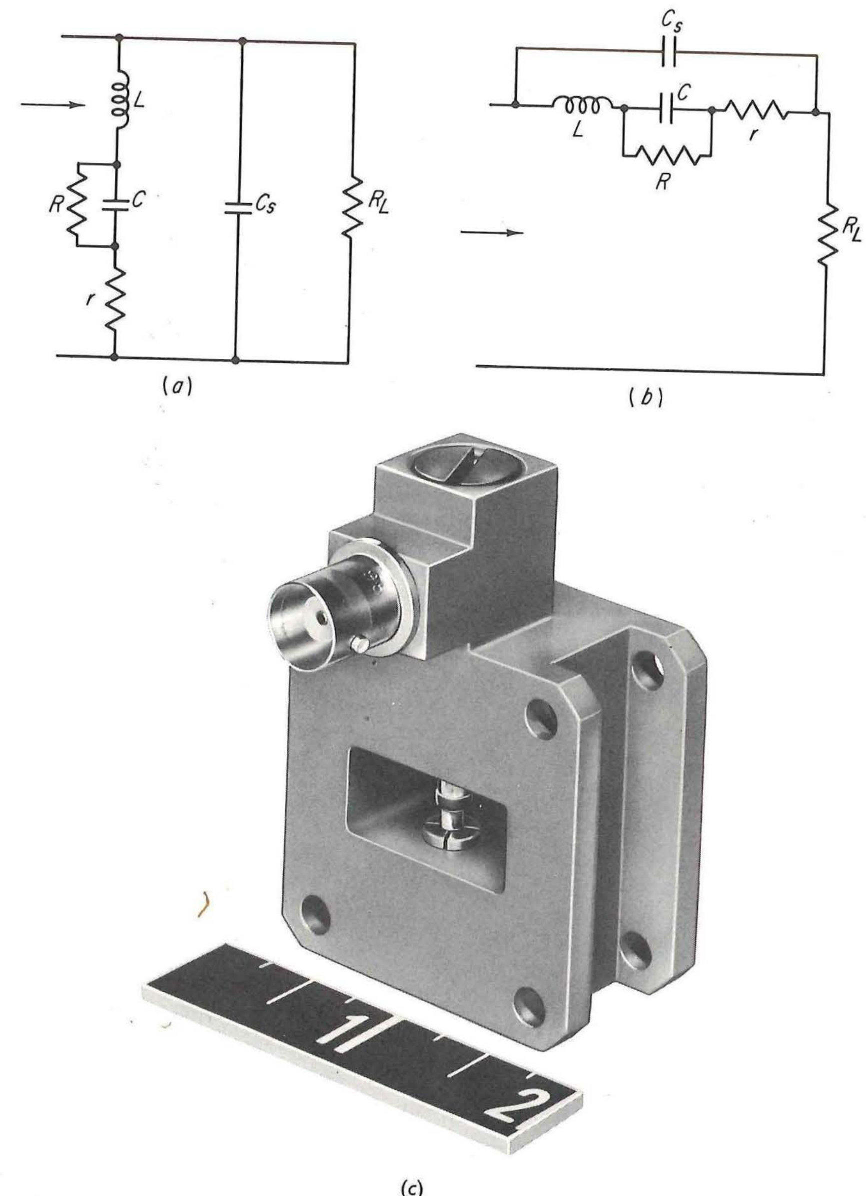
$$\text{Insertion loss} = 20 \log \left(1 + \frac{R_L}{2r(1 + Q^2)}\right)$$

$$\text{Isolation} = 20 \log \left(1 + \frac{R_L}{2r}\right)$$

The series-connected diode is the most convenient coaxial configuration. The insertion loss and isolation for the series-connected diode are

$$\text{Insertion loss} = 20 \log \left(1 + \frac{r}{2R_L}\right)$$

$$\text{Isolation} = 20 \log \left(1 + \frac{r(1 + Q^2)}{2R_L}\right)$$



L = inductance of the diode lead (whisker)
r = spreading resistance (fixed series resistance of the diode)
R = diode junction or barrier resistance (nonlinear resistance of the point contact which varies with amplitude and polarity of the bias voltage)
C = diode junction or barrier capacitance (a function of the applied bias voltage)
C_s = shunt capacity of the diode package

Fig. 8·19 Resonant switch operation. (*a*) Shunt switch. (*b*) Series-connected switch. (*c*) Diode switch. (*Hughes Aircraft Company, Solid-state Products, Aerospace Group.*)

8·14 Waveguide joints

A useful form of coupling called a choke joint is shown in schematic form in Fig. 8·20*c*. Choke joints enable two line sections to be joined electrically despite lack of good mechanical contact between sections. The L-shaped channel or slot is a half-wave transmission line. The minimum impedance at

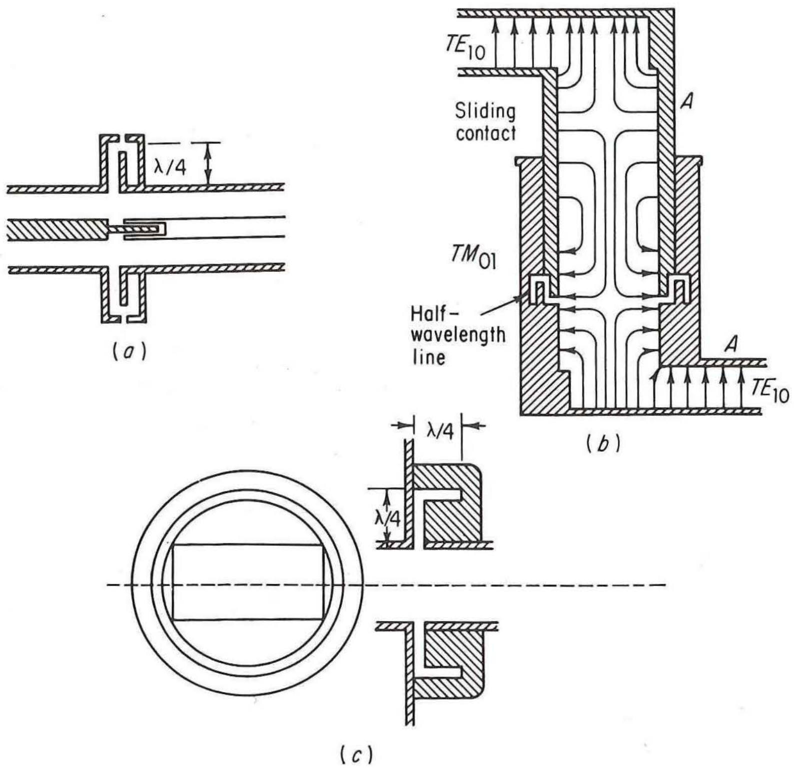


Fig. 8·20 Waveguide joints. (*a*) Coaxial rotating choke joint. (*b*) Waveguide rotating joint. (*c*) Waveguide choke joint.

the shorted end of the slot is transformed to the junction at the waveguide, and the junction behaves as though the adjacent walls were continuous.

A coaxial-line rotating choke joint is shown in Fig. 8·20*a*. Waveguide rotating joints find frequent applications in waveguide systems. A waveguide rotary joint is shown in Fig. 8·20*b*. The rotation is accomplished by using the radial symmetrical TM_{01} mode. The two sections of rectangular waveguide are joined through a section of circular waveguide operating in the TM_{01} mode. This mode is excited directly from the rectangular waveguide mode. Matching partitions, such as inductive windows, are placed in the rectangular waveguide sections to reduce reflections at the junctions of the

two guides. Resonant rings may be placed in the circular guide to prevent the formation of the rectangular mode in the circular waveguide.

8·15 Waveguide slotted section

The waveguide slotted section is shown in Fig. 8·21*a*. The slotted section is an accurately machined section of waveguide in which a small longitudinal slot is cut. The section is mounted on a probe-carriage assembly. Since the waveguide sections can be interchanged, measurements can be performed over a number of waveguide bands by interchanging the waveguide slotted sections.

The probe shown in Fig. 8·21*b* is mounted on the carriage and samples the variations of the electric field throughout the length of probe travel along the slot. A crystal detector or barretter is mounted in the probe housing. The detector output is available at the BNC connector and is connected to a standing-wave amplifier. The probe is a thin silver wire which is inserted into the slot by the screw-on knob at the top of the unit. The probe is connected to the inner conductor of the two concentric coaxial lines. Each line is terminated by a variable short circuit. The inner conductor tuner is precisely adjusted by rotation of the inner knob shown in the open slot. The second tuner is a coarse adjustment controlled by the external cylindrical nut. In this way, the standing wave in the probe assembly is varied to obtain maximum output from the detector.

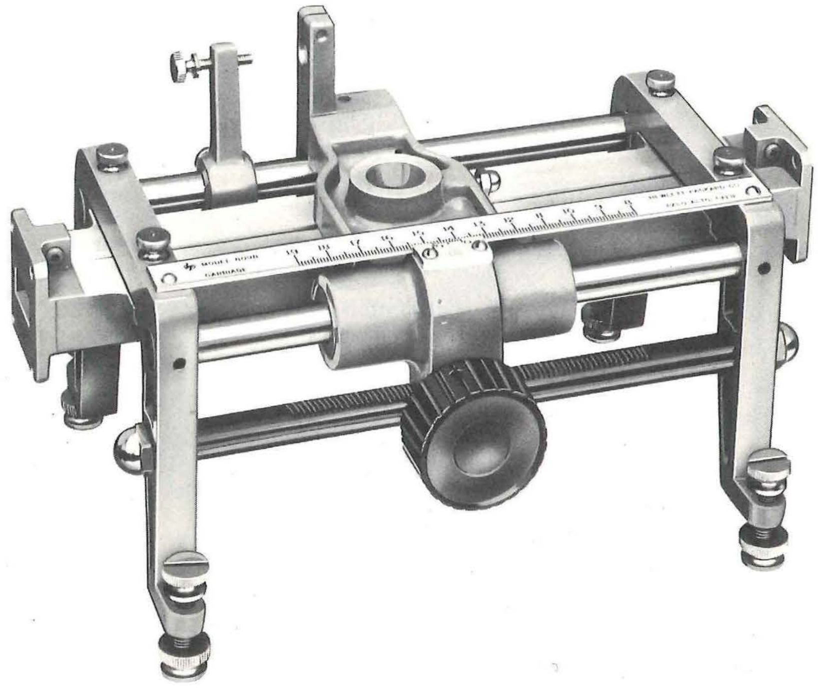
8·16 Antennas

The phenomenon known as *radiation* is the property of a transmission line structure by which wave power detaches itself and continues into space upon reaching the end of the line. An *antenna* or *radiator* is the coupling structure between the guided wave and the free-space wave or vice versa. It is used for either radiating electromagnetic energy into space or collecting electromagnetic energy from space.

Two types of sources are to be considered in a discussion of radiation patterns. The ideal point source radiates equal power in all directions. It is referred to as an *isotropic radiator* or as a *spherical radiator*. The hypothetical isotropic radiator is convenient in theory but is not a physically realizable source. The performance of the ideal source can be calculated and used as a base upon which the performance of a real antenna can be calculated. The power per unit area as a function of the distance from the point source is

$$P_r = \frac{P_t}{4\pi d^2}$$

where P_r is the radial component of average Poynting vector in watts per square meter, P_t is the total radiated power in watts, and d is the distance from the source in meters.



(a)

(b)

Fig. 8·21 (a) Probe carriage and waveguide slotted section. (*Hewlett-Packard Company.*) (b) Adjustable broadband probe. (*Sperry Microwave Electronics Company.*)

The *elementary source*, sometimes referred to as an *oscillating doublet*, may be regarded as an infinitely short linear current element. Ideally, it consists of two closely spaced charges of opposite sign, both oscillating in the same phase. The elementary source differs from the ideal source in that it has an axis. This source can be approximated but cannot be completely realized in practice. The radiation pattern of the elementary source is of considerable interest because it is the pattern produced by a short dipole coincident with

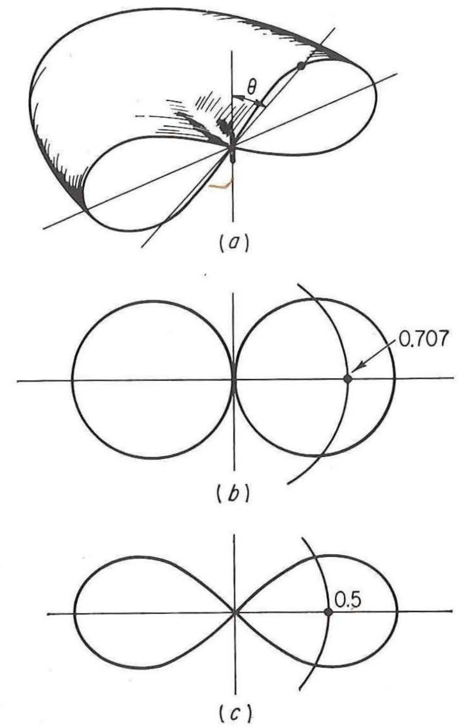


Fig. 8·22 (a) Vertical plot of the elementary source pattern. (b) Relative signal level. (c) Relative power level.

the polar axis. The source radiates but in directions perpendicular to the axis, and there is no radiation along the axis.

There are two kinds of field patterns associated with radiation from the elementary source, the field-intensity patterns and the power-density patterns. The *field-intensity pattern*, sometimes called the *relative field pattern* or *relative signal level*, is a plot of the electric field intensity at a fixed distance from the source as a function of the angle θ formed by the axis and the line joining the source to the point of measurement, as indicated in Fig. 8·22*a*. The relative signal level is proportional to $(\sin \theta)/d$, where d is the distance from the source to the point of measurement. A polar plot of the relative signal level pattern is shown in Fig. 8·22*b*.

The *power-density pattern,* called the *relative power pattern* or *relative power level,* is a plot of the power density at a fixed distance from the source as a function of the angle θ measured from the axis to the line joining the source to the point of measurement. The relative power level is proportional to $(\sin^2 \theta)/d^2$. A plot of relative power level is shown in Fig. 8·22*c*. The power level drops to one-half of its maximum value in directions that are 45° from the direction of maximum power level. The beam width, measured between the two half-power directions, is therefore 90° in a plane containing the axis. The signal level in the half-power directions is 0.707 of its maximum since 0.707 is the square root of 0.5.

The behavior of real antennas can be obtained by considering the line source formed by a large number of ideal isotropic sources which are not directional. The pattern of any antenna can be regarded as having been produced by an array of point sources. A radiating system which consists of a number of point sources arranged in a line is called a *line array* or *linear array*. The simplest linear array consists of two isotropic point sources which are excited in phase. The radiation pattern formed by two isotropic point sources of equal amplitude and the same phase is shown in Fig. 8·23*a*. The vector addition of the fields indicates that the field pattern is a maximum at P_1 and P_2, which are equidistant from the point sources. The path lengths from the sources are different at other points on the pattern such as P_3. The fields from the sources therefore differ in phase, thus producing the field pattern as indicated. This pattern is revolved around the y axis to form the space pattern (doughnut-shaped). The relative field pattern for two isotropic point sources of the same amplitude but opposite phase and spaced one-half wavelength apart is indicated in Fig. 8·23*b*. The low-level beams shown in *c*, *d*, *e*, and *f* are called *side-lobes*. The number and positions of these lobes are determined by the spacing and phase of the sources.

The field patterns in Fig. 8·23*a* and *c* are of the *broadside* type since the maximum radiation is perpendicular to the line joining the sources. The *end-fire* types of array are indicated in *b*, *d*, *e*, and *f*, in which maximum radiation is in the same direction as the line joining the sources. The directive end-fire arrays are illustrated in *e* and *f*.

The Half-wave Antenna. A short linear conductor is often called a short dipole. It is always of finite length even though it may be very short. Current oscillating in the dipole generates electromagnetic waves which travel out into free space at the velocity of light. The dipole has properties that are characteristic of a transmission line, a resonant circuit, and an antenna. The dipole exhibits characteristics of a resonator since energy concentrations oscillate from entirely electric energy to entirely magnetic energy and back twice per cycle.

A common form of the fundamental half-wave radiating element is a conductor of essentially uniform diameter, one-half wavelength long,

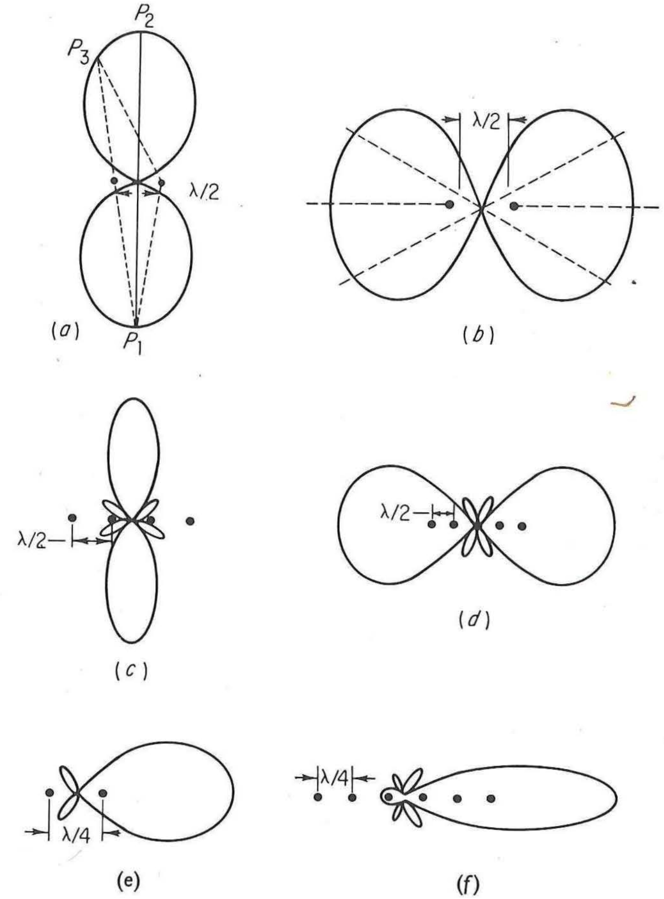


Fig. 8·23 Relative field patterns. (*a*) Two isotropic point sources of the same amplitude and phase spaced one-half wavelength apart. (*b*) Two isotropic point sources of the same amplitude but opposite phase. (*c*) Array of four isotropic point sources of equal amplitude and the same phase spaced one-half wavelength apart. (*d*) Array of four point sources of equal amplitude and opposite phase spaced one-half wavelength apart. (e) Sources equal in amplitude, 90° out of phase. (*f*) Directive end-fire type array.

connected at its midpoint to a two-wire transmission line. This type of connection presents a low impedance to the transmission line and is referred to as *center-fed* or *current-fed*. The *voltage-fed* or *end-fed* method of exciting half-wave antennas presents a high impedance. Both types are illustrated in Fig. 8·24.

The fields which are associated with the stored energy (periodic buildup and collapse of the electric and magnetic fields around the antenna) are called the *induction fields*. The induction fields are principally responsible for

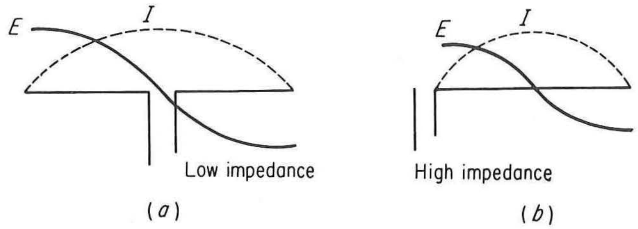


Fig. 8·24 Half-wave antenna feeds.

the behavior of the antenna as a resonant circuit element. The induction field components are large at the antenna and are not detectable beyond a distance of about two wavelengths from the antenna. The resonant behavior of the antenna also includes the movement of charge up and down the length of the antenna at the frequency of the applied wave. At a particular instant of time, maximum negative charge exists on one end of the antenna and maximum positive charge exists on the opposite end. Maximum intensity of electric flux lines exists at this instant (zero current). The flux lines associated with the charges collapse, and the unlike charges come together at another instant of the cycle (maximum current flow). The original lines of force try to return rapidly to the antenna. However, all of the lines of force in the fields cannot return before the new induction field, of opposite polarity, starts to move outward at the beginning of the next half-cycle. The new induction field components encounter the returning fields and force them back away from the antenna. This periodic action around the antenna produces a steady flow of energy into space, as indicated for a particular instant of time in Fig. 8·25.

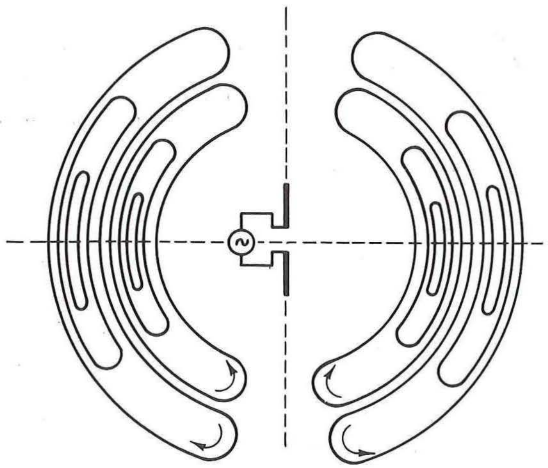


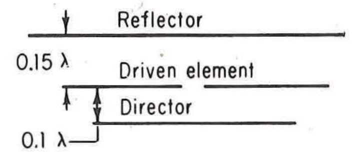
Fig. 8·25 Idealized representation of radiation (electric field) from a dipole.

The portion of the radiated energy which does not return to the antenna is called the *radiation field.* The flux lines expand radially with the velocity of light, and new flux lines are created at the antenna to replace the radiated lines. The frequency of the sinusoidal oscillating fields is the same as the frequency of the current in the antenna. The radiation fields are smaller than the induction fields near the antenna but are much larger in regions removed from the antenna.

The field patterns of the dipole are the same as those shown in Fig. 8·22 for the elementary source. The intensity of the radiated wave depends primarily upon the radiation resistance and the current flow in the antenna. In turn, the radiation resistance depends upon the antenna shape, size, length, height above ground, and the operating frequency.

Antenna Arrays. The limited directive property of the half-wave antenna makes it necessary to use other types of antennas in order to produce a concentration of radiated energy in a specific direction. Directional arrays can be constructed with the aid of elements in which currents are induced by the fields of the driven element. These elements have no transmission line connection to the transmitter or receiver and are usually referred to as *parasitic elements.* A parasitic element can act as a *reflector* or *director.* A parasitic array is perhaps the simplest of all directional arrays. A parasitic element longer than the driven element is called a *reflector* and reinforces radiation in the direction of a line pointing away from itself toward the driven element. This element is usually spaced about 0.15 wavelength from the driven element, as indicated in Fig. 8·26. If the parasitic element is shorter

Fig. 8·26 Physical arrangement of a three-element parasitic array.



than the driven element, it is called a *director.* The sharpness of resonance of the multielement array is determined by the lengths of the parasitic elements, and the gain or directional property is chiefly determined by the element spacing. The input resistance of the driven element is usually lowered by the parasitic element since current in the parasitic element causes a voltage to be induced in the driven element. A wide frequency range is obtained by adjusting the director length to resonate at the highest frequency and by adjusting the reflector to resonate at the lowest frequency.

Collinear elements lie in the same plane or axis and are excited in phase. If collinear elements are stacked above and below another set of similar elements, the result is a *broadside array.* The connections of a broadside array consisting of a number of one-half wave antennas are shown in Fig. 8·27*a*. The radiation pattern minus the side lobes is shown in Fig. 8·27*b*. One of the two main lobes can be eliminated by placing a metal surface or metal screen approximately one-quarter wavelength behind the array. One lobe can also be eliminated using a second array of antennas placed behind

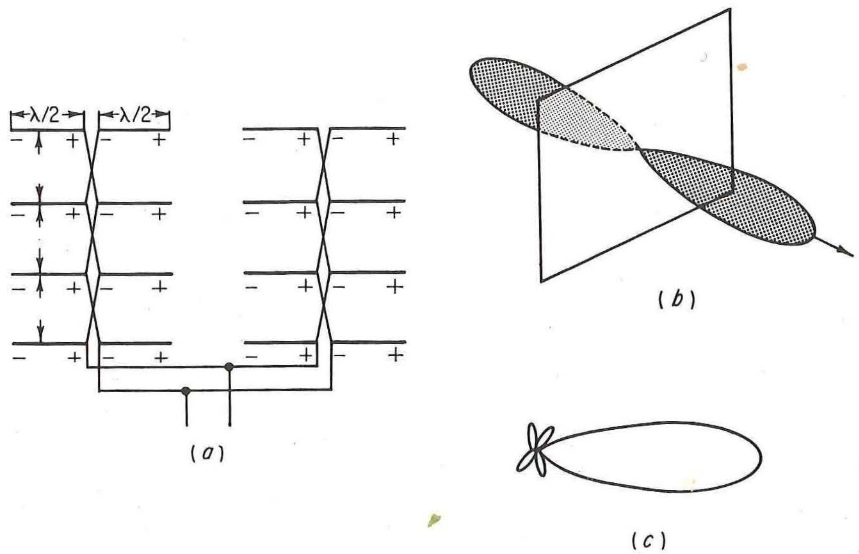


Fig. 8·27 (*a*) Broadside array. (*b*) Radiation pattern of the broadside array. (*c*) Field pattern for broadside array with reflector.

the first array. The resulting pattern is shown at *c*. Additional rows of dipoles in the horizontal and vertical directions provide more narrow beams.

Parabolic Reflector. The parabola is a curve that passes through all points which are equidistant from a point called the focus and a fixed line called the *directrix*. The distance from any point must be measured perpendicular to the directrix, as indicated in Fig. 8·28*a*. Constant-phase surfaces determine the directional characteristics of antennas. The field reflected from the parabola has a single time phase in a plane across the mouth of the parabola as illustrated. Energy from the focal point P strikes the parabola surface at points such as A, B, and C and is reflected in a direction parallel to the axis (the angle of incidence is equal to the angle of reflection as shown). The sum of the distance PA and AE equals the sum of the distances PB and BD.

Sources of illumination for parabolic antennas are classified as *front feed* and *rear feed*. The feed system may be coaxial or waveguide, as illustrated in Fig. 8·28. The antenna at *b* is a *paraboloid of revolution*, which is sometimes referred to as a *dish*. The antenna at *c* is a *parabolic cylinder*. A

schematic diagram of a rear-fed dish antenna is shown at *d*. The parabola-of-revolution antenna is a high-gain antenna used in tracking radar systems where sharp, well-defined beams are required. The parabolic cylinder antenna is used in high-power, search-radar systems applications. High directive

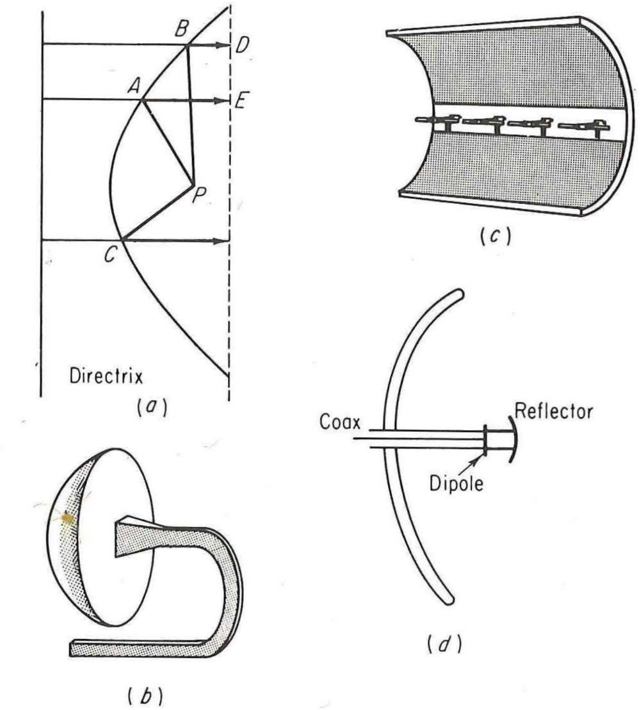


Fig. 8·28 Parabolic antennas. (*b*) Paraboloid of revolution. (*c*) Parabolic cylinder. (*d*) Rear-fed dish.

properties are obtained because of the large dimensions in wavelengths of the plane constant-phase surface.

Electromagnetic Horn. A rectangular waveguide may be flared in either or both of its two dimensions to form a rectangular horn. The radiation pattern is highly directive. The beam is sharp in the plane of the longer sides and comparatively broad in the plane of the shorter sides. The actual beam directivity is determined by horn length and flare angle and is generally increased by an increased horn length. The length and flare angle provide a gradual decrease in waveguide wavelength. As the waveguide wavelength approaches the free-space wavelength, the wave impedance approaches the free-space impedance, and a near perfect impedance match is obtained. If the guide is flared symmetrically in the *a* dimension only, the structure is known as an H-plane *sectoral horn*. The E-plane sectoral horn is flared symmetrically in the b dimension. The *pyramidal horn* is formed by flaring the waveguide in both planes, as shown in Fig. 8·29. Power gain is a function of the flare in both planes. As shown, the constant-phase wavefront is not a plane wave.

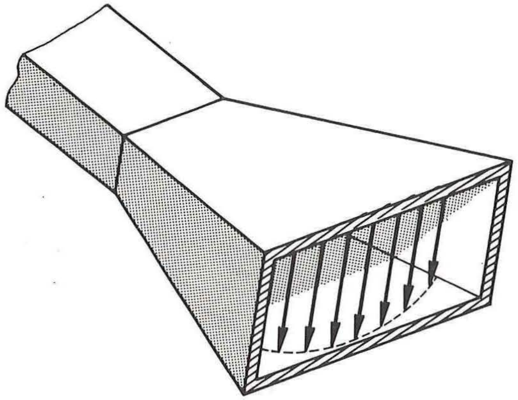


Fig. 8·29 The pyramidal horn.

Metal-plate Lens Antennas. A waveguide lens makes use of the optical properties of microwaves. A constant-phase wavefront can be obtained from a spherical wavefront, as shown in Fig. 8·30. Waves which pass through the thick center of the plane-convex dielectric lens have a lower phase velocity and therefore suffer a decrease in phase shift. Waves which pass through the

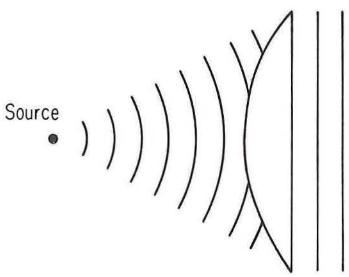


Fig. 8·30 Dielectric lens.

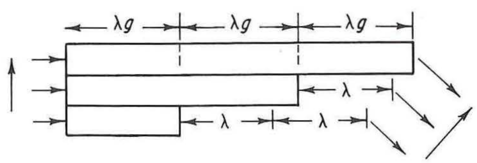


Fig. 8·31 Illustration of phase focusing.

thin portions of dielectric material have less phase shift and therefore emerge with a relative leading phase. With proper shaping, the lens produces a plane wave at the output side.

The parallel plates of a waveguide lens increase the phase velocity of the electromagnetic wave. Therefore, this type of lens must be concave so that energy at the edges of the lens experiences an increased phase velocity in passing through the longer waveguide section. As an example of focusing, Fig. 8·31 shows a plane-wave incident upon three sections of waveguide. The

waveguide wavelength is greater than the free-space wavelength, and the corresponding phase velocity is greater than the free-space velocity. Because of the different velocities, the wavefront is staggered as shown. The waves can be made convergent by proper lengths of waveguides. Alternately, waves originating from a focal point will emerge as a beam of parallel waves, as shown in Fig. 8·32*a*. In the example, it was assumed that waveguide sections

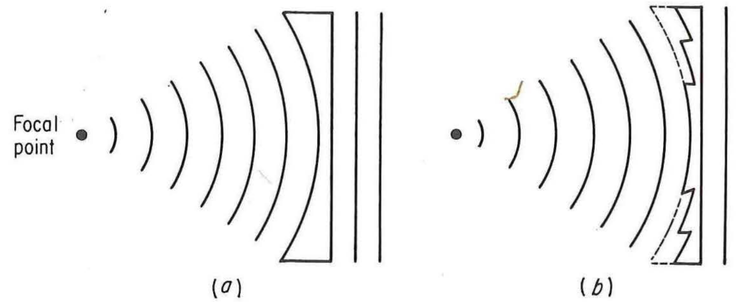
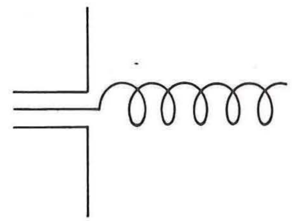


Fig. 8·32 (*a*) E-plane metal-plate lens. (*b*) Stepped or zoned lens.

were placed one on top of the other. However, the phase velocity in a waveguide depends only on the width or a dimension of a waveguide propagating the TE_{10} mode. The tops and bottoms can be removed, and the lens is therefore constructed with thin sheets of metal. One form of lens antenna is the stepped waveguide lens shown in Fig. 8·32*b*. It is stepped so that a full-wave path length difference is removed whenever the thickness is increased by a full wave in going from the center of the lens outward.

Fig. 8·33 Helical antenna.



The *helical antenna* is shown in Fig. 8·33. This type antenna is useful where a circularly polarized radiation field is desired. Radiation from helices with circumferences of the order of one wavelength and a number of turns is usually a well-defined beam with a maximum radiation in the direction of the helix axis. A wide variety of nonuniform or tapered helices is also possible. The terminal impedance of a helical antenna operating in the axial mode is a pure resistance in the range of 100 to 200 ohms.

Slot Antennas. A linear array antenna which has excellent electrical characteristics and which has mechanical advantages over other types of antennas can be formed by positioning a series of slots along a length of waveguide. The slots radiate electromagnetic energy, and the shape of the radiation pattern depends upon the orientation of the slot with respect to the edges and faces of the guide. The slots may be classified as *resonant*, or *near resonant*, and *nonresonant*. The majority of practical arrays consists of arrays of slots cut in the same face of the waveguide structure. If the field pattern, at a distance from the radiator, is similar for each element or slot, the array is called a *parallel array*.

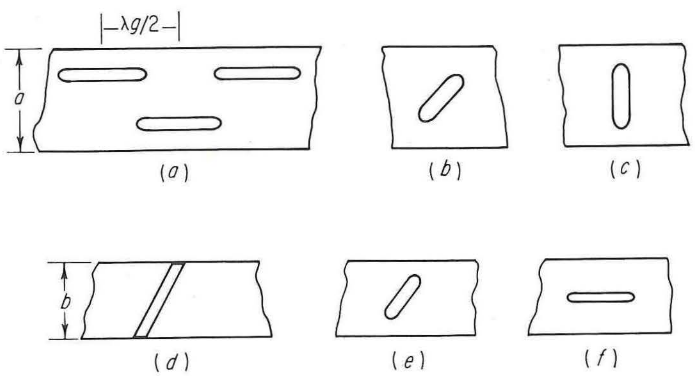


Fig. 8·34 Resonant slots. (*a*) Longitudinal shunt slot array. (*b*) Inclined series slot. (c) Series slot. (*d*) Edge slot (shunt). (e) Shunt inclined slot. (*f*) Longitudinal shunt slot in *b* dimension.

Radiating slots can be located at a number of different positions along the rectangular waveguide, as indicated in Fig. 8·34. The *shunt* type looks like a series resonant circuit in parallel with the line. When this type of slot is arranged in an array, a short circuit is placed an odd multiple of one-quarter waveguide wavelengths from the last slot. The *series* type slot looks like a parallel resonant circuit in series with the line. In a series slot array, a short circuit is placed a multiple of one-half wavelengths from the last slot.

The longitudinal shunt slot and the edge or shunt-inclined slots are most widely used. The longitudinal slot produces polarization transverse to the axis of the guide. The edge slot gives longitudinal polarization, which is the normal requirement. The edge slot can be of the "dumbbell" type; however, the simple edge cut is easier to fabricate.

Nonresonant or traveling-wave arrays have slot spacings other than one-half or one wavelength. The beam emerges at an angle to the array, and the angle varies with frequency. The nonresonant array is matched along its whole length because of the random phase additions of the reflections from the slots.

Antenna Properties. The radiating antenna acts like a network containing resistance and reactance. Some of the resistance loss is due to the heat loss in the antenna, and some is associated with the loss by radiation. The loss due to radiation is referred to as *radiation resistance.* Radiation resistance is defined as the ratio of total power radiated to the square of the effective value of the maximum current in the radiating system. The radiation resistance of a linear center-fed one-half wavelength antenna with sinusoidal current distribution is 73 ohms.

The *directive gain* of an antenna is a measure of its ability to transmit a signal to a distant point. A measure of the ability of an antenna to collect power is called the *effective area.* The *effective area* is the ratio of the power available at the antenna terminals (terminating impedance) to the power per unit area (power density) of the polarized incident wave. That is, the received power is equal to the power flow through an area that is equal to the effective area of the antenna.

From the law of reciprocity, a good transmitting antenna is also a good receiving antenna.

The *gain* of an antenna is the ratio of the maximum radiation intensity to the maximum radiation intensity from a reference antenna with the same power. If the reference antenna is an isotropic point source, then the gain is the *absolute gain.* A simple expression for antenna gain is

$$G = \frac{P}{P_a}$$

where P is the power flow per unit area in the plane, linearly polarized wave which the antenna causes in a distant region and P_a is the power flow per unit area which would have been produced by a reference antenna transmitting the same amount of power. Antenna gain may also be expressed by

$$G = \frac{4\pi A}{\lambda^2}$$

where A is the effective area. From the above formula,

$$A = \frac{\lambda^2 G}{4\pi}$$

Since the gain of the isotropic source is *one*, the effective area is $\lambda^2/4\pi = 0.079\lambda^2$. The simple oscillating doublet has an effective area of $3\lambda^2/8\pi = 0.119\lambda^2$ and a corresponding gain of 1.5. The half-wave antenna has an effective area of $1.31\lambda^2$ and a gain of 1.64.

Other terms which are commonly used to describe the properties of antennas are the effectiveness, radiation efficiency, radiation intensity, and aperture. The *effectiveness* of an antenna is the ratio of the effective area to the actual area. The *radiation efficiency* is defined as the ratio of the power radiated to the total power supplied to the antenna at a given frequency.[3] *Radiation intensity* is defined as the power radiated from an antenna per unit solid angle in that direction.[3] The *aperture* of an array is defined as the portion of a plane surface near the antenna, perpendicular to the direction of maximum radiation, through which a major portion of the radiation passes.[3]

PROBLEMS

8·1 The magic tee is used in a microwave bridge circuit. Two signals of equal amplitude are applied to arms 1 and 2.
   a. Illustrate the electric field orientations in arms 1 and 2 and the **H** arm for a zero output (null) at the output of the **H** arm.
   b. Repeat *a* for a null in the **E** arm.
   c. Illustrate the electric field orientation in arms 1 and 2 and the **E** arm for a zero output (null) at the output of the **E** arm.

8·2 Plot the attenuation characteristic of the type 382A attenuator in 10° increments from 0 to 90°.

8·3 What is the directivity of the coupler in Fig. 8·9*c* if P_{0i} is -10 dbm and P_{0r} is -42 dbm? If P_i is 0 dbm, what is the coupling value?

8·4 The directivity of the coupler in Fig. 8·9*c* is 45 db. With a matched load connected to the main line input, the power output P_{0r} is 63 db below the input power P_{ir}. What is the coupling value of the coupler? If P_{ir} is 1 watt, what is P_{0r}?

8·5 If a short circuit is placed on the normal input line of the coupler in Prob. 8·4, what is the output power P_{0i}, assuming no loss in the main line?

8·6 Illustrate the expected detector output waveforms if a 1,000-cps square-wave signal is alternately applied to a crystal, barretter, and thermistor.

8·7 Make a comparison summary between crystal and barretter characteristics and applications.

8·8 Make a comparison summary between barretter and thermistor characteristics and applications.

8·9 Typical values for the Philco 1N3482 diode at 9 Gc are $L = 3$ nanohenrys, $C = 0.1$ picofarads, $r = 10$ ohms, and $R_L = 400$ ohms. Calculate the insertion loss and isolation for shunt switch operation.

8·10 What are the insertion loss and isolation values of the same diode used in a series 50-ohm coaxial line application?

8·11 The power input to the normal output terminal of a 20-db directional coupler is 300 mw. The directivity is greater than 60 db and can be ignored in the calculations.
   a. How much power appears at the auxiliary arm output if the main line input is terminated with a short circuit?
   b. How much power arrives at the auxiliary arm when the main line input is terminated with loads having the following VSWR values: 1.5, 2.0, 3.6, and 6.0?

REFERENCES

1. Henry J. Riblett, The Short-slot Hybrid Junction, *Proc. IRE*, vol. 40, no. 2, February, 1952.
2. M. R. Millet, Microwave Switching by Crystal Diodes, *IRE MTT*-6, no. 3, pp. 284–290, July, 1958.
   R. V. Garver, E. G. Spencer, and M. A. Harper, Microwave Semiconductor Switching Techniques, *IRE MTT*-6, no. 4, pp. 378–383, October, 1958.
   R. V. Garver, Theory of TEM Diode Switching, *IRE MTT*-9, no. 3, pp. 224–238, May, 1961.
3. "Standards on Antennas," Definitions of Terms, The Institute of Radio Engineers, New York, 1948.