FIELDS AND WAVES IN COMMUNICATION ELECTRONICS

SECOND EDITION

SIMON RAMO
TRW Inc.

JOHN R. WHINNERY
University of California, Berkeley

THEODORE VAN DUZER
University of California, Berkeley

JOHN WILEY & SONS
NEW YORK CHICHESTER BRISBANE TORONTO SINGAPORE

5

TRANSMISSION LINES

5.1 Introduction

In Chapter 3, we saw that the interchange of electric and magnetic energy gives rise to the propagation of electromagnetic waves in space. More specifically, the magnetic fields that change with time induce electric fields as explained by Faraday's law, and the time-varying electric fields induce magnetic fields, as explained by the generalized Ampère's law. This interrelationship also occurs along conducting or dielectric boundaries, and can give rise to waves that are guided by such boundaries. These waves are of paramount importance in guiding electromagnetic energy from a source to a device or system in which it is to be used. Dielectric guides, hollow-pipe waveguides, and surface guides are all important for such purposes, but one of the simplest systems to understand—and one very important in its own right—is the two-conductor transmission line. This system may be considered a distributed circuit and so is useful in establishing a relation between circuit theory and the more general electromagnetic theory expressed in Maxwell's equations. The concepts of energy propagation, reflections at discontinuities, standing versus traveling waves and the resonance properties of standing waves, phase and group velocity, and the effects of losses upon wave properties are easily extended from these transmission-line results to the more general classes of guiding structures.

A parallel two-wire system is a typical and important example of the transmission lines to be studied in this chapter. In any transverse plane, electric field lines pass from one conductor to the other, defining a voltage between conductors for that plane. Magnetic field lines surround the conductors, corresponding to current flow in one conductor and an equal but oppositely directed current

flow in the other. Both voltage and current (and of course the fields from which they are derived) are functions of distance along the line. In the two following sections we set down the transmission-line equations from distributed-circuit theory, but then discuss its relation to field theory.

Transmission-line effects are not always desirable ones. A cable interconnecting two high-speed computers may be intended as a direct connection, but will at the very least introduce a time delay (around 5 ns/m in typical dielectric-filled cable). Moreover, if the interconnections are not impedance matched at the two ends there will be reflections of the waves (as we shall see later in the chapter). These "echoes" of pulses representing the digits could introduce serious errors. A still further complication is *dispersion*. In a real transmission line the velocity of propagation varies to some degree with frequency, so the frequency components which represent the pulse (by Fourier analysis) travel at different velocities and the pulse distorts as it travels. If dispersion is excessive, the pulses may be blurred enough so that individual digits cannot be clearly distinguished. All of these effects occur also in the interconnections of elements in printed circuits and even in semiconductor integrated circuits, but the close spacings in the last case limit performance only for extremely short pulses.

Transmission-line analysis is useful, by analogy, in studying a variety of wave phenomena, such as the propagation of acoustic waves and their reflection from materials with different acoustic properties. An especially interesting analog is that of the propagation of signals along a nerve of the human body.

TIME AND SPACE DEPENDENCE OF SIGNALS ON IDEAL TRANSMISSION LINES

5.2 Voltage and Current Variations Along an Ideal Transmission Line

We begin by considering the transmission line as a distributed circuit. In Sec. 2.5 we identified an inductance per unit length associated with the flux produced by the oppositely directed currents in a pair of parallel conductors. When the currents vary with time, there is a voltage change along the line. Likewise, the distributed capacitance produces displacement current between the conductors when the voltage is time varying and leads to change of the current flowing along the conductors. The interrelationship leads to the wave equation for voltage and current along an ideal lossless transmission line.

Figure 5.2 shows a representative two-conductor line and the circuit model for a differential length. It should be kept in mind that the external inductance per unit length of a parallel-conductor line is not associated with one conductor or the other. Also, the circuit model is simply a representation of a differential length of line; there is not a one-for-one identification of the two sides of the circuit with the two conductors of the modeled line.

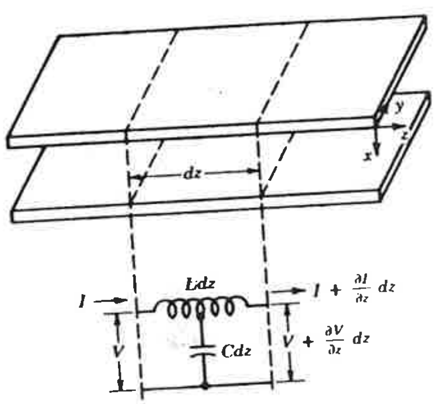


Fig. 5.2 Section of a representative transmission line and the equivalent circuit for a differential length.

Consider a differential length of line dz, having the distributed inductance, L per unit length, and the distributed capacitance, C per unit length. The length dz then has inductance $L\,dz$ and capacitance $C\,dz$. The change in voltage across this length is then equal to the product of this inductance and the time rate of change of current. For such a differential length, the voltage change along it at any instant may be written as the length multiplied by the rate of change of voltage with respect to length. Then

$$\text{voltage change} = \frac{\partial V}{\partial z}\,dz = -(L\,dz)\frac{\partial I}{\partial t} \tag{1}$$

Note that time and space derivatives are written as partial derivatives, since the reference point may be changed in space or time in independent fashion.

Similarly, the change in current along the line at any instant is merely the current that is shunted across the distributed capacitance. The rate of decrease of current with distance is given by the capacitance multiplied by time rate of change of voltage. Partial derivatives are again called for:

$$\text{current change} = \frac{\partial I}{\partial z}\,dz = -(C\,dz)\frac{\partial V}{\partial t} \tag{2}$$

The length dz may be canceled in (1) and (2):

$$\frac{\partial V}{\partial z} = -L\frac{\partial I}{\partial t} \tag{3}$$

$$\frac{\partial I}{\partial z} = -C\frac{\partial V}{\partial t} \tag{4}$$

Equations (3) and (4) are the fundamental differential equations for the analysis of the ideal transmission line. Note that they are identical in form with the pairs,

Eqs. 3.9(5) and 3.9(9) or Eqs. 3.9(6) and 3.9(8) found from Maxwell's equations for plane electromagnetic waves. As was done there, (3) and (4) can be combined to form a wave equation for either of the variables. For example, one can differentiate (3) partially with respect to distance and (4) with respect to time:

$$\frac{\partial^2 V}{\partial z^2} = -L\frac{\partial^2 I}{\partial z\,\partial t} \tag{5}$$

$$\frac{\partial^2 I}{\partial t\,\partial z} = -C\frac{\partial^2 V}{\partial t^2} \tag{6}$$

Partial derivatives are the same taken in either order (assuming continuous functions) so (6) may be substituted directly in (5):

$$\frac{\partial^2 V}{\partial z^2} = LC\frac{\partial^2 V}{\partial t^2} = \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2} \tag{7}$$

where

$$v = (LC)^{-1/2} \tag{8}$$

All real signals are continuous functions, as required for (7) to apply. The discontinuous step waves used later as examples are to be understood as approximations to real signals. Equations (3) and (4) are known as the *telegraphist's equations*, and the differential equation (7) is the one-dimensional wave equation. A similar equation may be obtained in terms of current by differentiating (4) with respect to z, (3) with respect to t, and combining the results:

$$\frac{\partial^2 I}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 I}{\partial t^2} \tag{9}$$

We saw in Sec. 3.9 that an equation of the form (7) has a solution

$$V(z, t) = F_1\left(t - \frac{z}{v}\right) + F_2\left(t + \frac{z}{v}\right) \tag{10}$$

where F_1 and F_2 are arbitrary functions. A constant value of $F_1(t - z/v)$ would be seen by an observer moving in the $+z$ direction with a velocity v so $F_1(t - z/v)$, represents a wave traveling in the $+z$ direction with velocity v. Similarly, $F_2(t + z/v)$ represents a wave moving in the $-z$ direction with velocity v.

To find the current on the line in terms of the functions F_1 and F_2, substitute the expression for voltage given by (10) in the transmission line equation (3)

$$-L\frac{\partial I}{\partial t} = -\frac{1}{v}F_1'\left(t - \frac{z}{v}\right) + \frac{1}{v}F_2'\left(t + \frac{z}{v}\right) \tag{11}$$

This expression may be integrated partially with respect to t:

$$I = \frac{1}{Lv}\left[F_1\left(t - \frac{z}{v}\right) - F_2\left(t + \frac{z}{v}\right)\right] + f(z) \tag{12}$$

If this result were substituted in the other transmission line equation (4), it would be found that the function of integration, $f(z)$, could only be a constant. This is a possible superposed dc solution not of interest in studying the wave solution, so the constant will be ignored. Equation (12) may then be written

$$I = \frac{1}{Z_0}\left[F_1\left(t - \frac{z}{v}\right) - F_2\left(t + \frac{z}{v}\right)\right] \tag{13}$$

where

$$Z_0 = Lv = \sqrt{\frac{L}{C}}\ \Omega \tag{14}$$

The constant Z_0 as defined by (14) is called the *characteristic impedance* of the line, and is seen from (10) and (12) to be the ratio of voltage to current for a single one of the traveling waves at any given point and given instant. The negative sign for the negatively traveling wave is expected since the wave propagates to the left, and by our convention current is positive if flowing to the right.[1]

Example 5.2
Characteristic Impedance and Wave Velocity for a Coaxial Line

Let us find expressions for the characteristic impedance and wave velocity for an ideal coaxial line and examine some typical values. We will assume the conductor spacing is large enough to neglect internal inductance. Using C from Eq. 1.9(4) and L from Eq. 2.5(6) in (14) we find

$$Z_0 = \frac{\ln b/a}{2\pi}\sqrt{\frac{\mu}{\varepsilon}} \tag{15}$$

where a and b are the radii of the inner and outer conductors at the dielectric surfaces, respectively. A common commercial coaxial cable is designated RG58/U. It has a dielectric of relative permittivity 2.26 and the radii are: $a = 0.406$ mm and $b = 1.48$ mm. Substituting these values in (15) and taking $\mu \cong \mu_0$ (Sec. 2.3) one finds that $Z_0 = 51.6\,\Omega$. This is slightly below the published normal value $Z_0 = 53.5\,\Omega$ due in part to the neglect of the frequency-dependent internal inductance of the conductors. There is not much variation of the relative permittivity among the various materials used as the dielectrics in coaxial lines

[1] Since Z_0 as defined here is real, it is more logical to call it a "characteristic resistance," especially since the concept of impedance implies use with the phasor forms appropriate to steady-state sinusoidal excitation. That is an important special case to be considered later, but even for transmission lines used with pulses or other general signals, it is common to refer to the defined Z_0 as *characteristic impedance*.

and since the radius ratio comes in only in a logarithm, one finds that most commercial coaxial lines have characteristic impedance in a limited range, usually $50\,\Omega \lesssim Z_0 \lesssim 80\,\Omega$. The wave velocity (8) becomes

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \tag{16}$$

which is the same as the velocity of plane waves in the same dielectric, Sec. 3.9. This result obtains for all two-conductor transmission lines when the internal inductance can be neglected.[2] The wave velocity is usually between about 0.5 and 0.7 of the velocity of light in vacuum, 3×10^8 m/s, for lines with plastic dielectrics.

5.3 Relation of Field and Circuit Analysis of Transmission Lines

Although we largely utilize the distributed-circuit model for transmission line analysis in this chapter, let us relate the equations obtained in Sec. 5.2 to field concepts of Chapter 3. First, let us take the special case of a parallel-plane transmission line, as indicated in Fig. 5.2, with the conducting planes assumed wide enough in the y direction so that fringing at the edges is not important. If the planes are also assumed perfectly conducting, it is clear that a portion of a uniform plane wave with E_x and H_y, as studied in Chapter 3, can be placed in the dielectric region between the planes and will satisfy the boundary condition that electric field enter normally to the perfectly conducting planes. The Maxwell equations for such a wave [Eqs. 3.9(6) and 3.9(8)] are

$$\frac{\partial E_x(z,t)}{\partial z} = -\mu\frac{\partial H_y(z,t)}{\partial t} \tag{1}$$

$$\frac{\partial H_y(z,t)}{\partial z} = -\varepsilon\frac{\partial E_x(z,t)}{\partial t} \tag{2}$$

If we define voltage as the line integral of $-\mathbf{E}$ between planes at a given z,

$$V(z,t) = -\int_1^2 \mathbf{E}\cdot d\mathbf{l} = -\int_0^a E_x\,dx = -aE_x(z,t) \tag{3}$$

Current for a width b, with positive sense defined for the upper plane, is related to the tangential magnetic field by

$$I(z,t) = -bH_y(z,t) \tag{4}$$

[2] It follows that knowledge of either L or C for such ideal lines determines the other

With these substituted in Eqs. 5.2(3) and 5.2(4), we find the result identical to (1) and (2) if

$$C = \frac{\varepsilon b}{a} \ \text{F/m}, \qquad L = \frac{\mu a}{b} \ \text{H/m} \tag{5}$$

These are, respectively, the capacitance per unit length and inductance per unit length for such a system of parallel-plane conductors, calculated from static concepts. The field and circuit concepts are thus identical in this simple case.

We would also find field and distributed circuit approaches identical if we applied them to the coaxial transmission line with ideal conductors, or to two-conductor systems of other shape so long as conductors are perfect. This is because such systems can be shown to propagate *transverse electromagnetic* (TEM) waves, for which both electric and magnetic fields have only transverse components. The absence of an axial magnetic field means that there are no induced transverse electric fields and no corresponding contributions to the line integral $\int \mathbf{E} \cdot d\mathbf{l}$ taken between the two conductors, so long as the integration paths remain in the transverse plane; thus the voltage between conductors can be taken as *uniquely defined for that plane*. Similarly the absence of an axial electric field means that there is no displacement current contribution to $\oint \mathbf{H} \cdot d\mathbf{l}$ for paths in a given transverse plane, and if such a closed path surrounds one conductor, the integral will be just the conduction current flow in that conductor for that plane at that instant of time. Moreover, the transverse **E** and **H** fields can be shown to satisfy Laplace's equation *in the transverse plane* (Prob. 5.3), thus explaining the appropriateness of using Laplace solutions for the calculation of the L and C of the transmission line.

When the finite resistances of conductors are taken into account, the identity of circuit and field analysis is no longer an exact one, but has been shown to be a good approximation, for practical transmission lines. This field basis for TEM waves will be developed more in Chapter 8.

5.4 Reflection and Transmission at a Resistive Discontinuity

Most transmission-line problems are concerned with junctions between a given line and another of different characteristic impedance, a load resistance, or some other element that introduces a discontinuity. By Kirchhoff's laws, total voltage and current must be continuous across the discontinuity. The total voltage in the line may be regarded as the sum of voltage in a positively traveling wave, equal to V_+ at the point of discontinuity, and a voltage in a reflected or negatively traveling wave, equal to V_- at the discontinuity. The sum of V_+ and V_- must be V_L, the voltage appearing across the junction:

$$V_+ + V_- = V_L \tag{1}$$

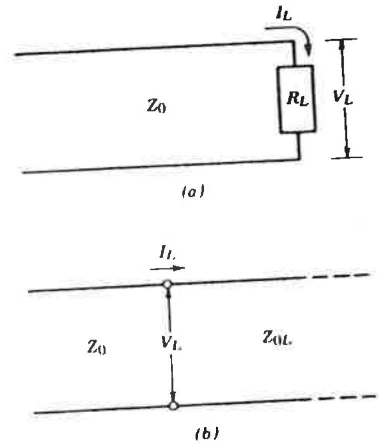


Fig. 5.4 (*a*) Ideal transmission line with a resistive load. (*b*) Ideal transmission line of characteristic impedance Z_0 with a second ideal line, of infinite length and characteristic impedance Z_{0L} as a load.

Similarly, the sum of current in the positively and negatively traveling waves of the line, at the point of discontinuity, must be equal to the current flowing into the junction:

$$I_+ + I_- = I_L \tag{2}$$

The simplest form of discontinuity is one in which a load resistance R_L is connected to the transmission line at the junction, as shown in Fig. 5.4*a*. Another case that is equivalent is that of Fig. 5.4*b* in which the first ideal line is connected to a second ideal line of infinite length and characteristic impedance Z_{0L}; here $R_L = Z_{0L}$. Still other forms of load circuits can produce an effective resistance R_L at the junction. In all these cases, $V_L = R_L I_L$. By utilizing the relations between voltage and current for the two traveling waves as found in section 5.2, equation (2) becomes

$$\frac{V_+}{Z_0} - \frac{V_-}{Z_0} = \frac{V_L}{R_L} \tag{3}$$

By eliminating between (1) and (3), the ratio of voltage in the reflected wave to that in the incident wave (*reflection coefficient*) and the ratio of the voltage on the load to that in the incident wave (*transmission coefficient*) may be found.

$$\rho \triangleq \frac{V_-}{V_+} = \frac{R_L - Z_0}{R_L + Z_0} \tag{4}$$

$$\tau \triangleq \frac{V_L}{V_+} = \frac{2R_L}{R_L + Z_0} \tag{5}$$

The most interesting, and perhaps the most obvious, conclusion from the foregoing relations is this: there is no reflected wave if the terminating resistance is exactly equal to the characteristic impedance of the line. All energy of the incident wave is then transferred to the load and τ of (5) is unity.

In Sec. 5.5 the definitions of reflection and transmission coefficients will be given for the case of sinusoidal signals and will include other than purely resistive loads.

The instantaneous incident power at the load is $W_T^+ = I_+V_+ = V_+^2/Z_0$. The fractional power reflected is, therefore, the constant

$$\frac{W_T^-}{W_T^+} = \rho^2 \tag{6}$$

The remainder of the power goes into the load resistor or the second line so

$$\frac{W_{TL}}{W_T^+} = 1 - \rho^2 \tag{7}$$

Example 5.4a
Pulse on Short-Circuited Line

Let us consider a signal in the form of a pulse having a constant value V_0 between $t = 0$ and $t = t_1/5$ and zero otherwise. The pulse is fed into an ideal transmission line of length l such that $l = vt_1$. We will analyze what happens when the pulse reaches the end of the line, which will be taken as short circuited.

Drawing (i) in Fig. 5.4c shows the pulse moving along the line at $t \approx 0.3t_1$. The arrows connecting the charges on the conductors are electric field vectors. The integral of the electric field is the voltage between conductors. The current flows in the conductors only where there is voltage, and current continuity is accounted for by displacement currents $\varepsilon\,\partial\mathbf{E}/\partial t$ at the leading and trailing edges of the pulse.

At time t_1 the leading edge of the pulse reaches the end of the line. Drawing (ii) in Fig. 5.4c shows the pulse shortly after $t = t_1$. The short circuit requires that the voltage be zero. To maintain the voltage at zero during the time that the incident pulse is at the termination $t_1 < t < t_2$, a negative-z-traveling (reflected) wave having opposite voltage polarity and equal amplitude is generated as shown in drawing (iii). Note that this result is predicted by (4) for $R_L = 0$. Also, the zero voltage on the load agrees in (5) with $R_L = 0$. Notice that the polarity of the current in the reflected wave is the same as in the incident wave, as could be argued from Eqs. 5.2(10) and 5.2(12) with the fact that $V_- = -V_+$. The total voltage on the line at the time used for drawings (ii) and (iii) is their superposition; this is shown in drawing (iv) where it is seen that the voltages are almost

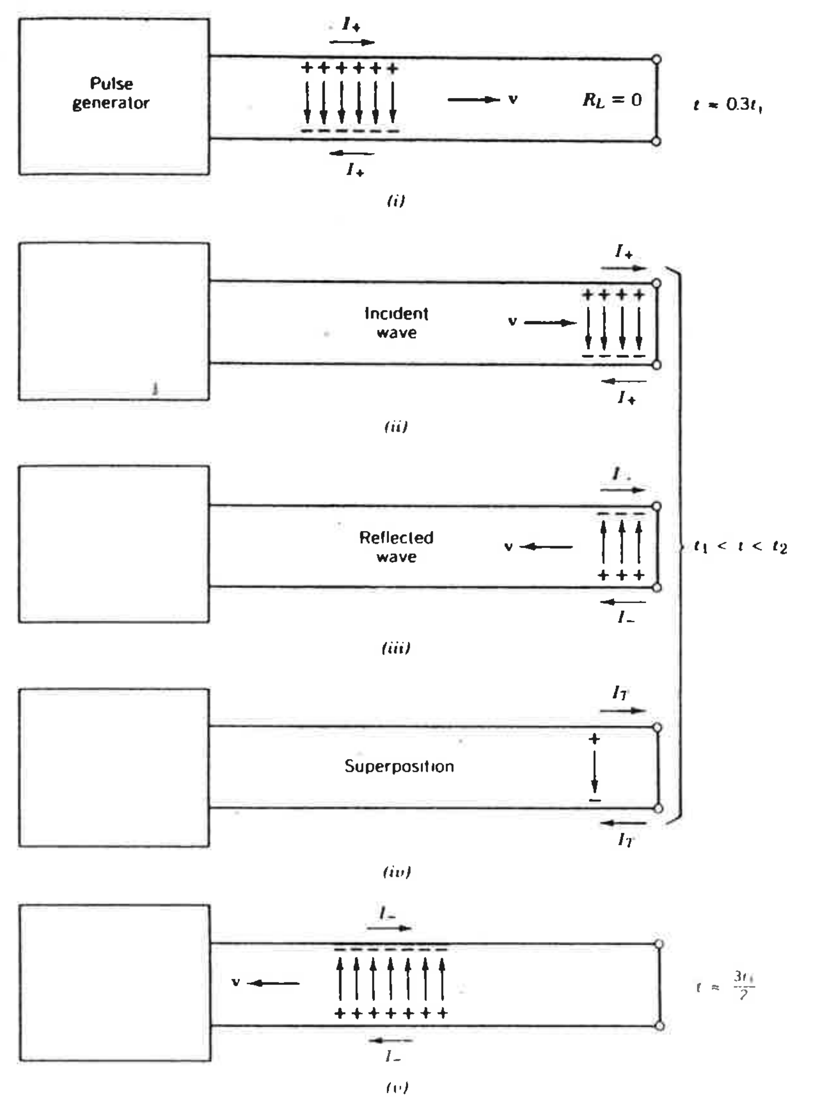


Fig. 5.4c Reflection of a pluse at the end of a shorted line. The times t_1 and t_2 are those for which the leading edge and trailing edges, respectively, reach the end of the line. Drawings (ii), (iii), and (iv) are for various instants during the period in which the reflected wave is generated to maintain the voltage at zero across the short circuit.

completely canceled. At a still later time, the reflected pulse is seen on its way to the generator [drawing (v)]. At $t = 2t_1$, the pulse will reach the pulse generator. What happens there depends on the impedance seen looking into the generator. Typically, the characteristic impedance of the transmission line equals the output impedance of the generator (it is *matched*). In that case, the reflected pulse is absorbed in the generator. Otherwise, another reflection takes place.

Example 5.4b

Pulse Reflections on a Transmission Line Interconnecting Two Computers

The aim of this example is to show the importance of transmission-line matching in controlling reflections for a transmission line used to interconnect two computers. Consider the two computers shown in Fig. 5.4*d* interconnected by a coaxial cable 100 m in length, and with a velocity of propagation 2×10^8 m/s so that there is a time delay of 500 ns for a pulse to propagate from input of the cable to its output. Consider a portion of the digitally coded signal, made up of 10-ns pulses with basic spacing 20 ns as sketched in Fig. 5.4*e*. This is sketched versus distance in Fig. 5.4*f* at 5 ns before the first pulse edge reaches computer no. 2. If input impedance of the second computer matches the characteristic impedance of the line, there is no reflection and the entire signal is accepted. But suppose its input impedance is 100 Ω and the characteristic impedance of the cable is 50 Ω. By (4), reflection coefficient is

$$\rho = \frac{R_L - Z_0}{R_L + Z_0} = \frac{100 - 50}{100 + 50} = \frac{1}{3}$$

So the train of pulses is reflected, at $\frac{1}{3}$ amplitude, toward computer no. 1 as sketched versus distance 110 ns after pictured in Fig. 5.4*f*. If there is impedance mismatch at the terminals of computer no. 1 additional reflection will take place when the signal returns there and a spurious signal will be superposed on whatever desired code is being sent between the computers at that time. Since the

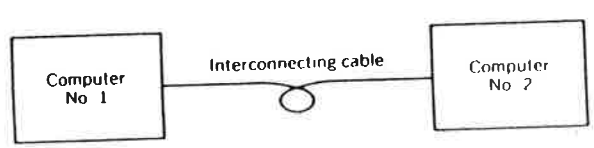


Fig. 5.4*d* Transmission line cable for transmitting digital signals between two computers.

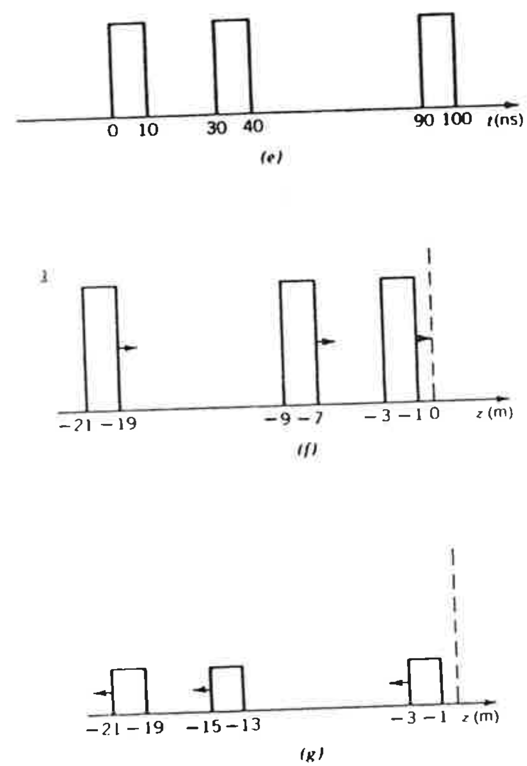


Fig. 5.4 (*e*) A portion of a pulse-coded signal versus time for the computer interconnection of Fig. 5.4*d*. (*f*) Voltage versus distance for a time 5 ns before leading edge of first pulse reaches input of computer no. 2 (defined as $z = 0$). (*g*) Voltage versus distance for reflected signal 110 ns after instant of sketch (*f*).

reflected "echo" is of lower amplitude than the original signal, differentiation is possible on the basis of amplitude level, but there is an obvious advantage in matching impedances well enough that reflected signals are small.

Example 5.4c

Square Wave in z Applied to Infinite Line

As a third example, consider an infinite line suddenly charged at $t = 0$ with a square wave in distance from $z = -1$ m to $z = +1$ m. Voltage distribution at $t = 0$ is then as in Fig. 5.4*h*. This may be considered a superposition of two such square waves, each of amplitude $V_0/2$. One of these moves to the right and the other to the left, each with velocity v. Thus at $t = 1.667$ ns (taking $v = 3 \times 10^8$ m/s) the two partial waves and their sum are shown in Fig. 5.4*i*. Figure 5.4*j* shows the corresponding current distribution, taking into account the different sign relations between current and voltage for positively and negatively traveling waves. Figure 5.4*k* shows voltage and current distributions at $t = 5$ ns.

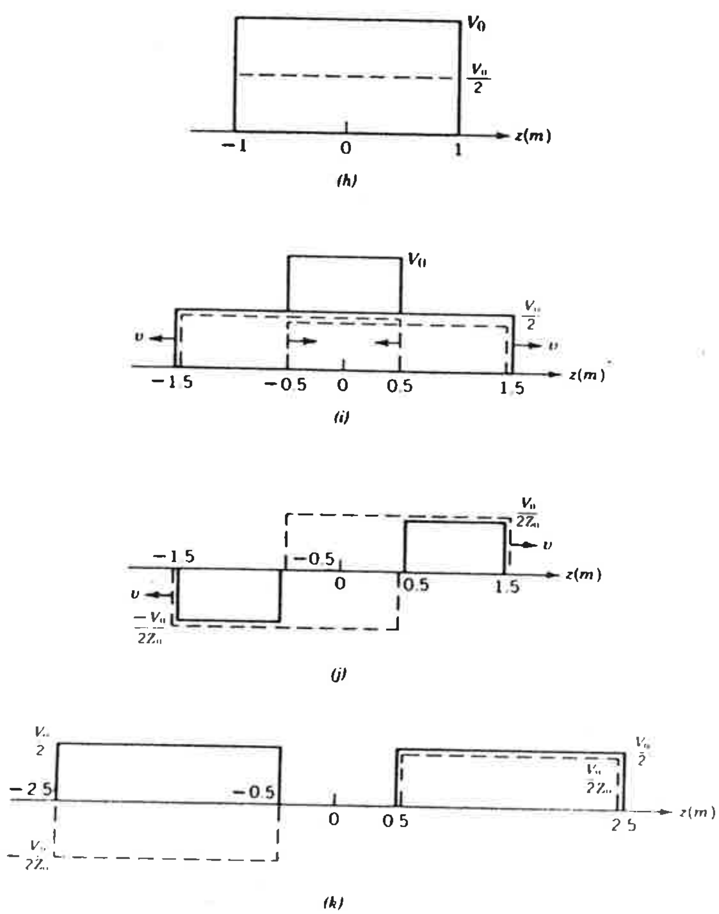


Fig. 5.4 Voltage and current distributions for Ex. 5.4c. (*h*) Voltage distribution at $t = 0$. (*i*) Traveling waves (dashed) and total voltage versus z (solid) at $t = 1.667$ ns. (*j*) Traveling waves of current (dashed) and total current (solid) versus z at $t = 1.667$ ns. (*k*) Voltage (solid) and current (dashed) versus z at $t = 5$ ns.

SINUSOIDAL WAVES ON IDEAL TRANSMISSION LINES WITH DISCONTINUITIES

5.5 Reflection and Transmission Coefficients and Impedance and Admittance Transformations for Sinusoidal Voltages

The preceding discussion has involved little restriction on the type of variation with time of the voltages applied to the transmission lines. Many practical problems are concerned with sinusoidal time variations. If a sinusoidal voltage is supplied to a line, it can be represented at $z = 0$ by

$$V(0, t) = V \cos \omega t \qquad (1)$$

The corresponding wave traveling in the positive z direction is

$$V_+(z, t) = |V_+| \cos \omega\left(t - \frac{z}{v_p}\right)$$

and that traveling in the negative z direction is

$$V_-(z, t) = |V_-| \cos\left[\omega\left(t + \frac{z}{v_p}\right) + \theta_p\right]$$

The total voltage is the sum of the two traveling waves:

$$V(z, t) = |V_+| \cos \omega\left(t - \frac{z}{v_p}\right) + |V_-| \cos\left[\omega\left(t + \frac{z}{v_p}\right) + \theta_p\right] \qquad (2)$$

The corresponding current, from Eq. 5.2(13), is

$$I(z, t) = \frac{|V_+|}{Z_0} \cos \omega\left(t - \frac{z}{v_p}\right) - \frac{|V_-|}{Z_0} \cos\left[\omega\left(t + \frac{z}{v_p}\right) + \theta_p\right] \qquad (3)$$

In Sec. 5.2 we saw that a constant point on a wave described by $F(t - z/v)$ is seen by an observer moving with a velocity v in the positive z direction. The argument of a sinusoid is called its *phase* so the velocity for which phase is constant is called the *phase velocity* v_p.

For sinusoidal time variations, it is useful to rewrite (2) and (3) in phasor form

$$V = V_+ e^{-j\beta z} + V_- e^{j\beta z} \qquad (4)$$

$$I = \frac{1}{Z_0}\left[V_+ e^{-j\beta z} - V_- e^{j\beta z}\right] \qquad (5)$$

where

$$\beta = \frac{\omega}{v_p} \qquad (6)$$

We may take V_+ as the reference for zero phase so that it is real. Then V_- is in general complex and equal to $|V_-|e^{j\theta_\rho}$ with θ_ρ being the phase angle between reflected and incident waves at $z = 0$ as in the instantaneous form (2).

The quantity β is called the *phase constant* of the line since βz measures the instantaneous phase at a point z with respect to $z = 0$. Moreover, voltage (or current) is observed to be the same at any two points along the line that are separated in z such that βz differs by multiples of 2π. The shortest distance between points of like current or voltage is called a *wavelength* λ. By the foregoing reasoning.

$$\beta\lambda = 2\pi$$

or

$$\beta = \frac{2\pi}{\lambda} \tag{7}$$

The expressions for reflection and transmission coefficients in Eqs. 5.4(4) and 5.4(5) can now be written in a special form for sinusoidal waves. It is convenient to choose the origin of the z coordinate at the discontinuity to be analyzed, as shown in Fig. 5.5 for three representative discontinuities. It is assumed that a

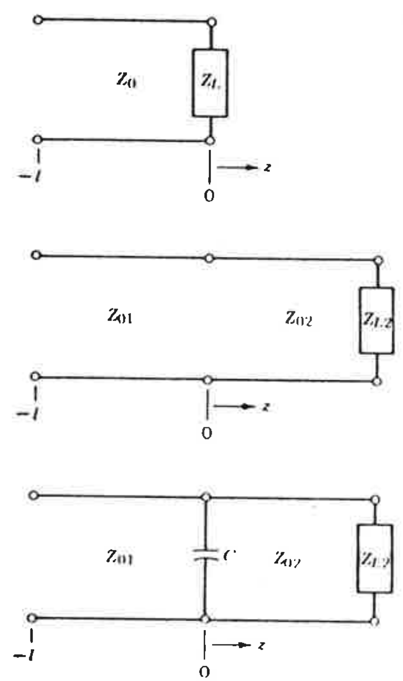


Fig. 5.5 Representative situations where a line of length l with a discontinuity at $z = 0$ is the subject of the analysis.

previous analysis gave the equivalent value of impedance looking in the $+z$ direction at $z = 0$ in the two lower lines. We will see below how this is done. The ratio of total phasor voltage to total phasor current at any point is the definition of impedance. We set the impedance at $z = 0$, the load impedance Z_L, equal to the ratio (4) to (5). Solving for the ratio V_-/V_+, the reflection coefficient for sinusoidal waves is obtained:

$$\rho = \frac{V_-}{V_+} = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{8}$$

Also, (4) and (5) can be combined to give the transmission coefficient:

$$\tau = \frac{V_L}{V_+} = \frac{2Z_L}{Z_L + Z_0} \tag{9}$$

The load voltage V_L is the total voltage at $z = 0$.

The expressions for power reflected and transmitted at a discontinuity, given for real functions of time in Eqs. 5.4(6)-(7), can be adapted to sinusoidal signals of the complex exponential type. Since power in this case is $VI^*/2$ and for a single wave in a loss-free line this is $VV^*/2Z_0 = |V|^2/2Z_0$, the fractional reflected power is

$$\frac{W_T^-}{W_T^+} = \frac{|V_-|^2}{|V_+|^2} = |\rho|^2 \tag{10}$$

and the remainder goes into the load:

$$\frac{W_{TL}}{W_T^+} = 1 - |\rho|^2 \tag{11}$$

Now let us find expressions for the input impedance and admittance at $-l$. We find impedance by dividing (4) by (5) for $z = -l$:

$$Z_i = Z_0\left[\frac{e^{j\beta l} + \rho e^{-j\beta l}}{e^{j\beta l} - \rho e^{-j\beta l}}\right] \tag{12}$$

Or, substituting (8),

$$Z_i = Z_0\left[\frac{Z_L \cos\beta l + jZ_0 \sin\beta l}{Z_0 \cos\beta l + jZ_L \sin\beta l}\right] \tag{13}$$

By defining admittances $Y_i = 1/Z_i$, $Y_L = 1/Z_L$, and $Y_0 = 1/Z_0$, we can find an expression of the same form for Y_i:

$$Y_i = Y_0\left[\frac{Y_L \cos\beta l + jY_0 \sin\beta l}{Y_0 \cos\beta l + jY_L \sin\beta l}\right] \tag{14}$$

Example 5.5
Cascaded Thin-Film Lines

Suppose the second diagram of Fig. 5.5 represents two thin-film transmission lines in a microwave integrated circuit. The load Z_{L2} represents a device having a real impedance of 20 Ω at the signal frequency, 18 GHz. Line 2, of characteristic impedance $Z_{02} = 30\,\Omega$, has a length $l_2 = 2$ mm. Line 1, of characteristic impedance $Z_{01} = 20\,\Omega$, has a length $l_1 = 1.5$ mm. The phase velocities for both lines are the same, 2×10^8 m/s. Let us find the impedance at the input to line 1.

First we must solve for the way the load impedance Z_{L2} transforms along the line attached to it. To do this we apply formula (13) to that section of line. For both lines (6) gives

$$\beta = \frac{\omega}{v_p} = \frac{(2\pi)(18 \times 10^9)\text{ rad/s}}{2 \times 10^8 \text{ m/s}} = 566 \text{ rad/m}$$

For variety we will take angles in degrees here as both degrees and radians are commonly used in transmission-line calculations

$$Z_{i2} = 30\left[\frac{20\cos 64.9^\circ + j30\sin 64.9^\circ}{30\cos 64.9^\circ + j20\sin 64.9^\circ}\right]$$

$$= 36.7 + j11.8\,\Omega$$

Now we can find Z_{i1} at the input to line 1 using Z_{i2} as the load Z_{L1}:

$$Z_{i1} = 20\left[\frac{(36.7 + j11.8)\cos 48.6^\circ + j20\sin 48.6^\circ}{20\cos 48.6^\circ + j(36.7 + j11.8)\sin 48.6^\circ}\right]$$

$$= 18.9 - j14.6\,\Omega$$

Systems with several lines of different characteristic impedances in cascade can be analyzed as we have in this example. In each case the analysis starts with the point farthest from the signal source, transforming the impedance back successively to the next discontinuity until the input is reached. In general, Z_0, β, and l will be different for each section.

5.6 Standing-Wave Ratio

Let us examine the phases of the voltages in Eq. 5.5(4). One coefficient, say V_+, can be chosen to be real by choice of origin of time. The reflection coefficient, Eq. 5.5(8), which is a complex number, can be written in the form $|\rho|e^{j\theta_\rho}$ so V_- in Eq. 5.5(4) can be replaced with $V_+|\rho|e^{j\theta_\rho}$ giving

$$V = V_+e^{-j\beta z} + V_+|\rho|e^{j(\theta_\rho + \beta z)} \qquad (1)$$

Let us write this as a real function of time with $-z$ replaced by l, the distance from the end of the line:

$$V(t, -l) = V_+\cos(\omega t + \beta l) + V_+|\rho|\cos(\omega t - \beta l + \theta_\rho)$$

Considering any particular instant, say $t = 0$, we can readily see that the argument of the cosine in the incident wave (first term), which is its phase, increases with distance from $z = 0$ and that the phase of the reflected wave (second term) decreases. These phases are shown in the top drawings of Fig. 5.6 where it is clear that at some distance z_0 the phases are the same. At z_π they differ by π rad, one having decreased by $\pi/2$ and the other having increased by $\pi/2$. At $z_{2\pi}$ they differ by 2π rad, and so on. So there are a series of locations where the two sinusoids are in phase and another series of locations where they are π rad out of phase. Where they are in phase, they add directly at each instant and where π rad out of phase, they subtract. At the former locations the total voltage has its maximum amplitude and at the latter, its minimum. Analysis of the sum of the incident and reflected waves given in (1) (see Prob. 5.6f) shows that the total

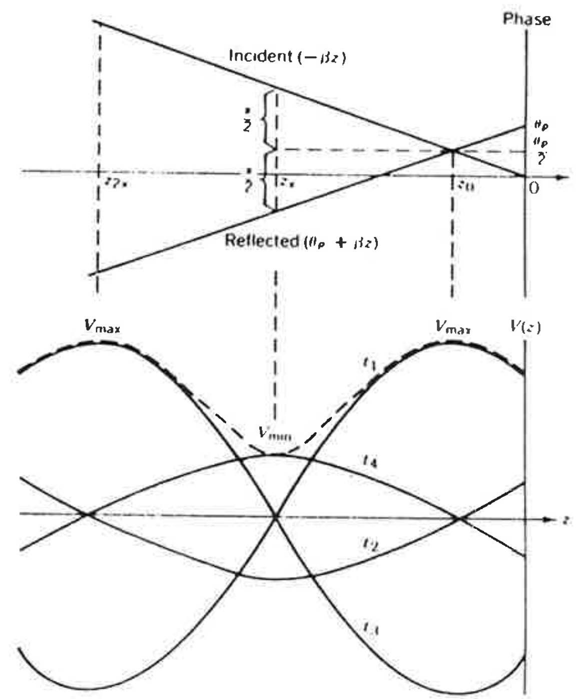


Fig. 5.6 The upper graph shows the phases of the incident and reflected waves at $t = 0$ on a line with a reflection coefficient $\rho = |\rho|e^{j\theta_\rho}$. The total voltage $V(z)$ is shown for selected times in the lower graph. The broken line gives the voltage amplitude along the line.

voltage can also be represented as the sum of a standing wave and a traveling wave. The total voltage is shown in the lower drawing of Fig. 5.6 for several particular times in a cycle selected to show the voltage when it has its maximum and minimum peak values. The broken line shows the voltage *amplitude* along the transmission line.

The maximum voltage is

$$V_{max} = |V_+| + |V_-| \tag{2}$$

and the minimum, found a quarter-wavelength from the maximum, is

$$V_{min} = |V_+| - |V_-| \tag{3}$$

The standing-wave ratio is then defined as the ratio of the maximum voltage amplitude to the minimum voltage amplitude,

$$S = \frac{V_{max}}{V_{min}} \tag{4}$$

By substituting (2) and (3) and the definition of reflection coefficient, Eq. 5.5(8), we find

$$S = \frac{|V_+| + |V_-|}{|V_+| - |V_-|} = \frac{1 + |\rho|}{1 - |\rho|} \tag{5}$$

It is seen that standing-wave ratio is directly related to the magnitude of reflection coefficient ρ, giving the same information as this quantity. The inverse relation is

$$|\rho| = \frac{S - 1}{S + 1} \tag{6}$$

Figure 5.6 is plotted for $S = 3$, corresponding to $|\rho| = \frac{1}{2}$.

Because of the negative sign appearing in the current equation, Eq. 5.5(5), it is evident that, at the position where the two traveling-wave terms add in the voltage relations, they subtract in the current relation, and vice versa. The maximum voltage position is then a minimum current position. The value of the minimum current is

$$I_{min} = \frac{|V_+| - |V_-|}{Z_0} \tag{7}$$

At this position impedance is purely resistive and has the maximum value it will have at any point along the line:

$$Z_{max} = Z_0\left[\frac{|V_+| + |V_-|}{|V_+| - |V_-|}\right] = Z_0 S \tag{8}$$

At the position of the voltage minimum, current is a maximum, and impedance is a minimum and real:

$$I_{max} = \frac{|V_+| + |V_-|}{Z_0} \tag{9}$$

$$Z_{min} = Z_0\left[\frac{|V_+| - |V_-|}{|V_+| + |V_-|}\right] = \frac{Z_0}{S} \tag{10}$$

Example 5.6

Slotted-Line Impedance Measurement

A *slotted-line* is an instrument that can be used to measure impedances. It is a transmission line containing a movable probe with which the standing-wave ratio, and the position of a voltage minimum or maximum can be found. The unknown impedance is connected to the end of the slotted line.

Suppose a measurement on a slotted-line of characteristic impedance $Z_0 = 50\,\Omega$ reveals a standing-wave ratio $S = 3$ and the closest voltage minimum is 0.33λ from the unknown load impedance. Let us see how Z_L is deduced. In Fig. 5.6, we see that at z_x where the minimum is found, the phase of the incident wave, $-\beta z$ attains the value $(\theta_\rho + \pi)/2$ so

$$l_{min} = -z_{min} = \frac{\theta_\rho + \pi}{2\beta}$$

and

$$\theta_\rho = 2\beta(0.33\lambda) - \pi = 2\left(\frac{2\pi}{\lambda}\right)(0.33\lambda) - \pi = 1.0 \text{ rad}$$

where we have used Eq. 5.5(7). Using the given S and (6) we find $|\rho|$ to be 0.5. Then

$$\rho = 0.5e^{j1.0}$$

From Eq. 5.5(8) we can get Z_L in terms of ρ and Z_0. Thus

$$Z_L = Z_0\left[\frac{1 + \rho}{1 - \rho}\right] = 50\left[\frac{1 + 0.5e^{j1.0}}{1 - 0.5e^{j1.0}}\right] = 52.8 + j59.3\,\Omega$$

5.7 The Smith Transmission-Line Chart

Many graphical aids for transmission line computations have been devised. Of these, the most generally useful has been one presented by P. H. Smith,[3] which

[3] P. H. Smith, *Electronic Applications of the Smith Chart*, Krieger, Melbourne, FL, 1983.

8

WAVEGUIDES WITH CYLINDRICAL CONDUCTING BOUNDARIES

8.1 Introduction

A waveguide is a structure, or part of a structure, that causes a wave to propagate in a chosen direction with some measure of confinement in the planes traverse to the direction of propagation. If the waveguide boundaries change direction, within reasonable limits, the wave is constrained to follow it. For example, in a transmission line used to transfer energy from a transmitter to an antenna, the energy follows the path of the line, at least for paths with only small discontinuities. The guiding of the waves in all such systems is accomplished by an intimate connection between the fields of the wave and the currents and charges on the boundaries, or by some condition of reflection at the boundary.

In this chapter we concentrate on cylindrical structures with conducting boundaries. Multiconductor lines can be used for frequencies from dc up to the millimeter-wave range. At the highest frequencies, they are often in the form of metallic films on insulating substrates. Hollow conducting cylinders of various cross-sectional shapes are used in the microwave and millimeter-wave frequency ranges (approximately 1–100 GHz).

Generally, in waveguide analyses we are interested in the distribution of the electromagnetic fields; but of greatest importance is the dependence of the propagation constant upon frequency. From the propagation constant one finds wave velocities, phase variation, and attenuation along the guide, and the pulse dispersion properties of the guide.

Several different types of guide are analyzed in this chapter, including the simple parallel-plate structure (and some more-practical, related forms) and hollow-tube guides of rectangular and circular cross section. An introduction to

means for exciting waves in waveguides is presented. The chapter concludes with a study of the general properties of waves in cylindrical waveguides with conducting boundaries.

GENERAL FORMULATION FOR GUIDED WAVES

8.2 Basic Equations and Wave Types for Uniform Systems

We consider here cylindrical systems with axes taken along the z axis. We also consider time-harmonic waves with time and distance variations described by $e^{(j\omega t-\gamma z)}$, as in the study of transmission-line waves. The character of the propagation constant γ tells much about the properties of the wave, such as the degree of attenuation and the phase and group velocities. The fields in the wave must satisfy the wave equation and the boundary conditions. We will assume that there is no net charge density in the dielectric and that any conduction currents are included by allowing permittivity and therefore $k^2 = \omega^2\mu\varepsilon$ to be complex. The wave equations, which reduce to the Helmholtz equations for phasor fields, (Sec. 3.11) are

$$\nabla^2\mathbf{E} = -k^2\mathbf{E}; \qquad \nabla^2\mathbf{H} = -k^2\mathbf{H}$$

The three-dimensional ∇^2 may be broken into two parts:

$$\nabla^2\mathbf{E} = \nabla_t^2\mathbf{E} + \frac{\partial^2\mathbf{E}}{\partial z^2}$$

The last term is the contribution to ∇^2 from derivatives in the axial direction. The first term is the two-dimensional Laplacian in the transverse plane, representing contributions to ∇^2 from derivatives in this plane. With the assumed propagation function $e^{-\gamma z}$ in the axial direction,

$$\frac{\partial^2\mathbf{E}}{\partial z^2} = \gamma^2\mathbf{E}$$

The foregoing wave equations may then be written

$$\nabla_t^2\mathbf{E} = -(\gamma^2 + k^2)\mathbf{E} \tag{1}$$

$$\nabla_t^2\mathbf{H} = -(\gamma^2 + k^2)\mathbf{H} \tag{2}$$

Equations (1) and (2) are the differential equations that must be satisfied in the dielectric regions of the transmission lines or guides. The boundary conditions imposed on fields follow from the configuration and the electrical properties of the boundaries.

The usual procedure is to find two components of the fields, usually the z components of **E** and **H**, that satisfy the wave equations (1) and (2) and the boundary conditions and then the other field components can be found from these by using Maxwell's equations. To facilitate finding the other components, it usually is most convenient to have them explicitly in terms of the z components of **E** and **H**.

The curl equations with the assumed functions $e^{(j\omega t-\gamma z)}$ are written below for fields in the dielectric system, assumed here to be linear, homogeneous, and isotropic.

$$\nabla\times\mathbf{E}=-j\omega\mu\mathbf{H} \qquad \nabla\times\mathbf{H}=j\omega\varepsilon\mathbf{E}$$

$$\frac{\partial E_z}{\partial y}+\gamma E_y=-j\omega\mu H_x \tag{3}$$

$$-\gamma E_x-\frac{\partial E_z}{\partial x}=-j\omega\mu H_y \tag{4}$$

$$\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}=-j\omega\mu H_z \tag{5}$$

$$\frac{\partial H_z}{\partial y}+\gamma H_y=j\omega\varepsilon E_x \tag{6}$$

$$-\gamma H_x-\frac{\partial H_z}{\partial x}=j\omega\varepsilon E_y \tag{7}$$

$$\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}=j\omega\varepsilon E_z \tag{8}$$

It must be remembered in all analysis to follow that these coefficients, E_x, H_x, E_y, and so on, are functions of x and y only, by our agreement to take care of the z and time functions in the assumed $e^{(j\omega t-\gamma z)}$.

From the foregoing equations, it is possible to solve for E_x, E_y, H_x, or H_y in terms of E_z and H_z. For example, H_x is found by eliminating E_y from (3) and (7), and a similar procedure gives the other components.

$$E_x=-\frac{1}{\gamma^2+k^2}\left(\gamma\frac{\partial E_z}{\partial x}+j\omega\mu\frac{\partial H_z}{\partial y}\right) \tag{9}$$

$$E_y=\frac{1}{\gamma^2+k^2}\left(-\gamma\frac{\partial E_z}{\partial y}+j\omega\mu\frac{\partial H_z}{\partial x}\right) \tag{10}$$

$$H_x=\frac{1}{\gamma^2+k^2}\left(j\omega\varepsilon\frac{\partial E_z}{\partial y}-\gamma\frac{\partial H_z}{\partial x}\right) \tag{11}$$

$$H_y=-\frac{1}{\gamma^2+k^2}\left(j\omega\varepsilon\frac{\partial E_z}{\partial x}+\gamma\frac{\partial H_z}{\partial y}\right) \tag{12}$$

For propagating waves, it is convenient to use the substitution $\gamma=j\beta$ where β is real if there is no attentuation. Rewriting the above with this substitution,

$$E_x=-\frac{j}{k_c^2}\left(\beta\frac{\partial E_z}{\partial x}+\omega\mu\frac{\partial H_z}{\partial y}\right) \tag{13}$$

$$E_y=\frac{j}{k_c^2}\left(-\beta\frac{\partial E_z}{\partial y}+\omega\mu\frac{\partial H_z}{\partial x}\right) \tag{14}$$

$$H_x=\frac{j}{k_c^2}\left(\omega\varepsilon\frac{\partial E_z}{\partial y}-\beta\frac{\partial H_z}{\partial x}\right) \tag{15}$$

$$H_y=-\frac{j}{k_c^2}\left(\omega\varepsilon\frac{\partial E_z}{\partial x}+\beta\frac{\partial H_z}{\partial y}\right) \tag{16}$$

$$\nabla_t^2E_z=-k_c^2E_z \tag{17}$$

$$\nabla_t^2H_z=-k_c^2H_z \tag{18}$$

where

$$k_c^2\triangleq\gamma^2+k^2=k^2-\beta^2 \tag{19}$$

In studying guided waves along uniform systems, it is common to classify the wave solutions into the following types:

1. Waves that contain neither electric nor magnetic field in the direction of propagation. Since electric and magnetic field lines both lie entirely in the transverse plane, these may be called *transverse electromagnetic waves* (abbreviated *TEM*). They are the usual *transmission-line waves* along a multiconductor guide.
2. Waves that contain electric field but no magnetic field in the direction of propagation. Since the magnetic field lies entirely in transverse planes, they are known as *transverse magnetic* (*TM*) waves. They have also been referred to in the literature as E waves, or waves of electric type.
3. Waves that contain magnetic field but no electric field in the direction of propagation. These are known as *transverse electric* (*TE*) waves, and have also been referred to as H waves or waves of magnetic type.
4. *Hybrid waves* for which boundary conditions require all field components. These may often be considered as a coupling of *TE* and *TM* modes by the boundary.

The above is not the only way in which the possible wave solutions may be divided, but is a useful way in that any general field distribution excited in an ideal guide may be divided into a number (possibly an infinite number) of the above types with suitable amplitudes and phases. The propagation constants of these tell how the individual waves change phase and amplitude as they travel down the guide, so that they may be superposed at any later position and time to give the total resultant field there. Since it is disadvantageous to have a signal carried by several waves traveling at different velocities because of the resultant distortion, waveguides are normally designed so that only one wave can propagate even if many are excited at the entrance to the guide. (This condition often cannot be met in optical guides, Chapter 14.)

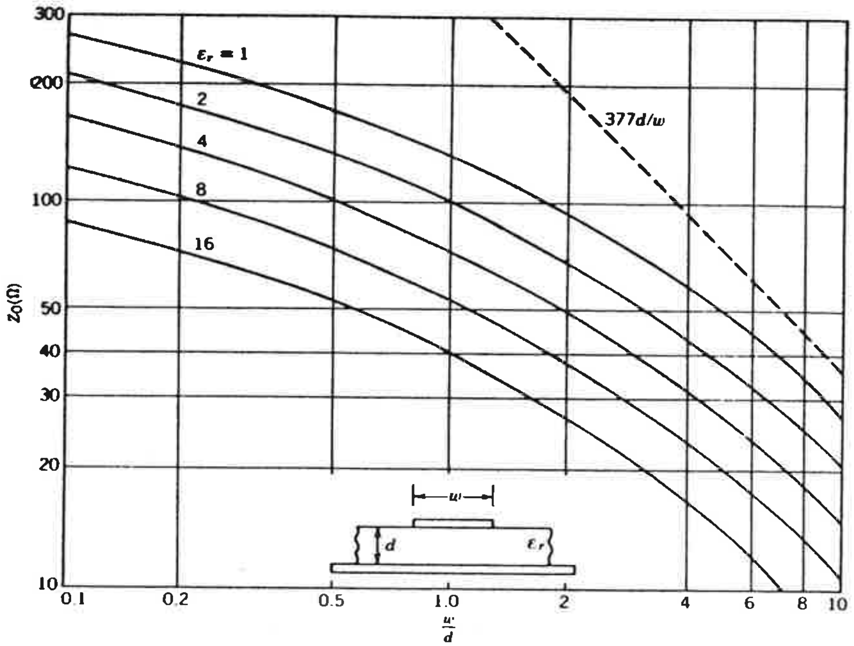


Fig. 8.6c Approximate characteristic impedance of stripline, adapted from Wheeler, Ref. 3. The dashed line is the parallel-plane approximation, for air dielectric. (© 1977 IEEE)

where k_0 and λ_0 are the free-space wave number and wavelength. The variation of the phase constant with frequency is accentuated where high permittivity dielectrics or small w/d ratios are used. Dispersion is discussed further in Sec. 8.16.

There are other types of waveguiding systems employing metal films on dielectrics in which all conductors are on the same side of the dielectric. These include the *slotline* (Fig. 8.6*d*), the *coplanar waveguide* (Fig. 8.6*e*), and *coplanar strips* (Fig. 8.6*f*). The slotline can sometimes be formed in the ground plane of a microstrip system, giving flexibility for forming some kinds of circuit elements. All of these coplanar types of lines afford the opportunity for convenient incorporation of lumped devices and short circuits, which is more difficult in the microstrip or stripline systems. For more details on coplanar, as well as microstrip lines, see Ref. 5.

[5] K. C. Gupta, R. Garg, and I. J. Bahl, *Microstrip Lines and Slotlines*, Artech House, Dedham, MA, 1979.

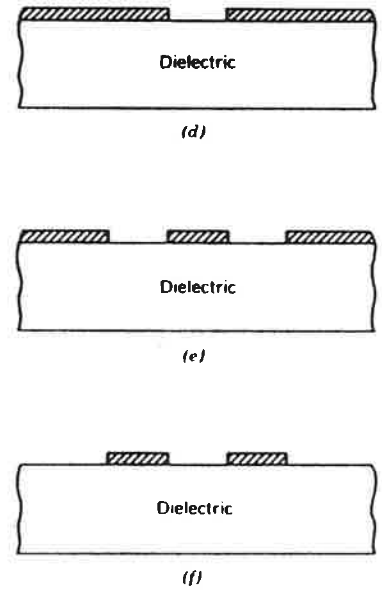


Fig. 8.6 Various forms of strip-type waveguiding systems. (*d*) Slotline. (*e*) Coplanar waveguide. (*f*) Coplanar strips.

8.7 Rectangular Waveguides

The most important of the hollow-pipe guides is that of rectangular cross section. As in Fig. 8.7*a*, a dielectric region of width *a* and height *b* extends indefinitely in the axial (z) direction and is closed by conducting boundaries on the four sides. In the ideal guide, both conductor and dielectric are loss-free. There can be no transverse electromagnetic wave (*TEM*) inside the hollow pipe since, as was

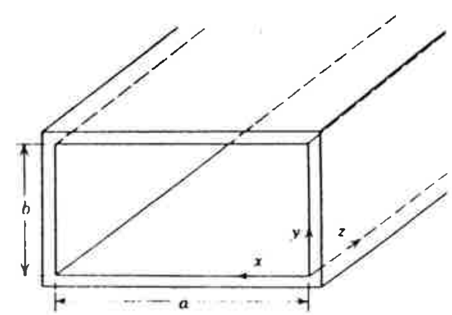


Fig. 8.7a Coordinate system for rectangular guide

shown in Sec. 8.3, *TEM* waves have transverse variations like static fields, and no static fields can exist inside a region bounded by a single conductor. Transverse magnetic (*TM*) and transverse electric (*TE*) waves can exist and will be analyzed below.

***TM* Waves** Transverse magnetic waves have zero H_z but nonzero E_z. The differential equation governing E_z is Eq. 8.2(17), here expressed in rectangular coordinates:

$$\nabla_t^2 E_z = \frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} = -k_c^2 E_z \tag{1}$$

This equation was solved in Sec. 7.19 by separation of variables procedures and found to have solutions of the form

$$E_z = (A' \sin k_x x + B' \cos k_x x)(C' \sin k_y y + D' \cos k_y y) \tag{2}$$

where

$$k_x^2 + k_y^2 = k_c^2 \tag{3}$$

The perfectly conducting boundary at $x = 0$ requires $B' = 0$ to produce $E_z = 0$ there. Similarly the ideal boundary at $y = 0$ requires $D' = 0$. We let $A'C'$ be a new constant A and have

$$E_z = A \sin k_x x \sin k_y y \tag{4}$$

Axial electric field E_z must also be zero at $x = a$ and $y = b$. This can only be so (except for the trivial solution $A = 0$) if $k_x a$ is an integral multiple of π:

$$k_x a = m\pi, \qquad m = 1, 2, 3, \ldots \tag{5}$$

Similarly, to make E_z zero at $y = b$, $k_y b$ must also be a multiple of π:

$$k_y b = n\pi, \qquad n = 1, 2, 3, \ldots \tag{6}$$

So the cutoff condition of the transverse magnetic wave with m variations in x and n in y (designated TM_{mn}) is found from (3):

$$\omega_{c_{m,n}} = \frac{k_{c_{m,n}}}{\sqrt{\mu\varepsilon}} = \frac{1}{\sqrt{\mu\varepsilon}}\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right]^{1/2} \tag{7}$$

Since k_c^2 is $k^2 - \beta^2$ as in Eq. 8.2(19), attenuation for frequencies below the cutoff frequency of a given mode, and phase constant for frequencies above the cutoff frequency have the same forms as for the parallel-plane guiding system:

$$\alpha = k_{c_{m,n}}\sqrt{1 - \left(\frac{\omega}{\omega_{c_{m,n}}}\right)^2}, \qquad \omega < \omega_{c_{m,n}} \tag{8}$$

$$\beta = k\sqrt{1 - \left(\frac{\omega_{c_{m,n}}}{\omega}\right)^2}, \qquad \omega > \omega_{c_{m,n}} \tag{9}$$

Phase and group velocities then also have the same forms as before [Eqs. 8.3(18) and 8.3(19)].

The remaining field components of the TM_{mn} wave are found from Eqs. 8.2(13)–(16) with $H_z = 0$ and E_z from (4):

$$E_x = -\frac{j\beta k_x}{k_{c_{m,n}}^2} A \cos k_x x \sin k_y y \tag{10}$$

$$E_y = -\frac{j\beta k_y}{k_{c_{m,n}}^2} A \sin k_x x \cos k_y y \tag{11}$$

$$H_x = \frac{j\omega\varepsilon k_y}{k_{c_{m,n}}^2} A \sin k_x x \cos k_y y \tag{12}$$

$$H_y = -\frac{j\omega\varepsilon k_x}{k_{c_{m,n}}^2} A \cos k_x x \sin k_y y \tag{13}$$

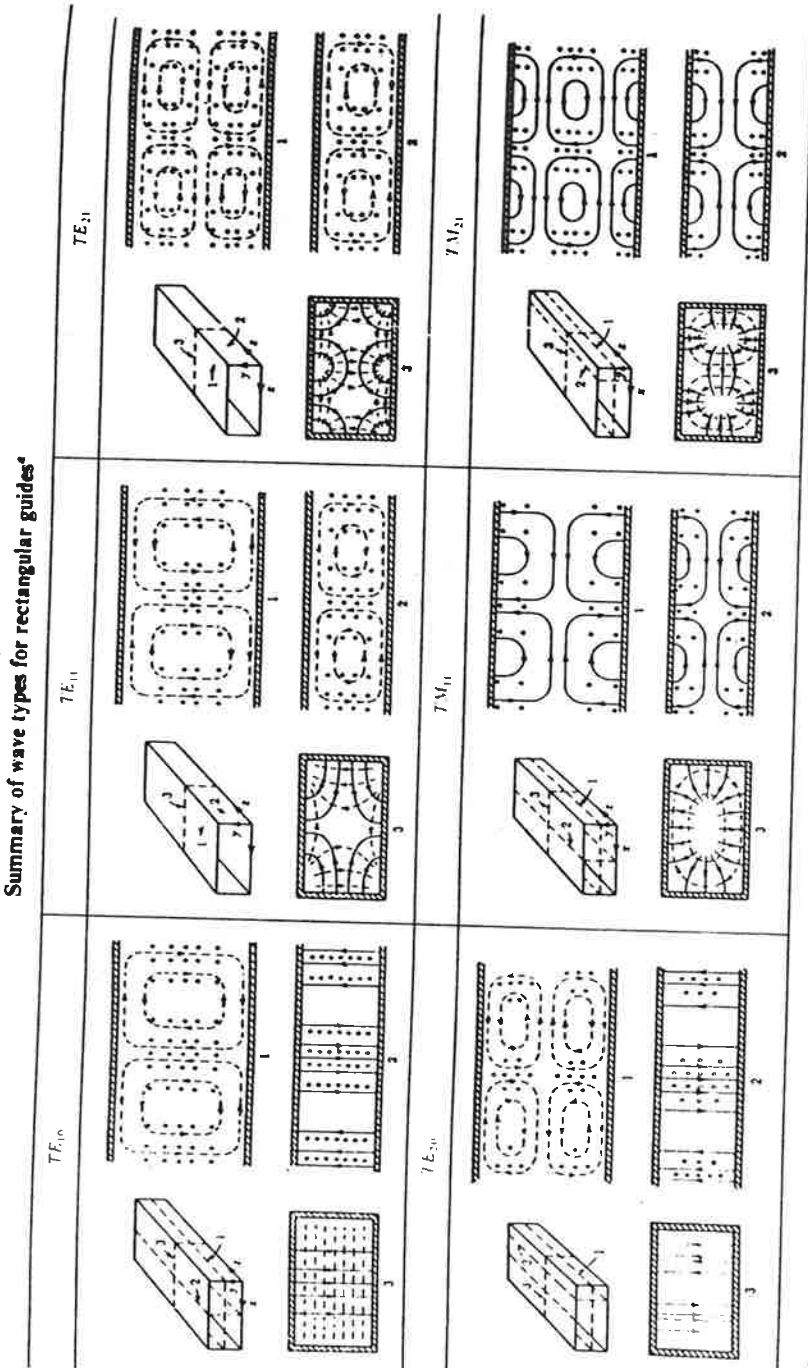
where k_x, k_y, $k_{c_{m,n}}$, and β are defined by (5), (6), (7), and (9), respectively, and all fields are multipled by the propagation term, $e^{-j\beta z}$. Plots of electric and magnetic field lines in the TM_{11} and TM_{21} modes are shown in Table 8.7. Note that electric field lines (shown solid) begin on charges on the guide walls at some fixed z plane, turn and go axially down the guide, and end on charges of opposite sign a half-guide wavelength down the guide. Magnetic field lines (shown dashed) surround the displacement currents represented by the changing electric fields as the pattern moves down the guide with velocity v_p. The pattern for the TM_{21} mode is that of two TM_{11} modes side by side and of opposite sense.

The attenuation resulting from losses in the conducting walls can be calculated for the TM_{mn} wave following the procedure used to find Eq. 8.5(11):

$$(\alpha_c)_{TM_{mn}} = \frac{2R_s}{b\eta\sqrt{1 - (f_c/f)^2}} \frac{[m^2(b/a)^3 + n^2]}{[m^2(b/a)^2 + n^2]} \tag{14}$$

We make two points before leaving this class of waves. Note from (4) and the definitions of k_x and k_y given by (5) and (6) that neither m nor n can be zero for the TM wave without its disappearing entirely. The second point is that we have required as boundary condition that E_z be zero along the perfectly conducting boundary, but should also be sure that other tangential components of electric

Table 8.7
Summary of wave types for rectangular guides[a]



[a] Electric field lines are shown solid and magnetic field lines are dashed.

field are zero there. From (10) and the definitions (5) and (6) we can see that E_x is zero as required at $y = 0$ and $y = b$; from (11) we see that E_y is zero at $x = 0$ and $x = a$. Thus all tangential components do satisfy the boundary conditions at the conductors. It can be shown (Prob. 8.7h) that imposition of the boundary condition on E_z necessarily causes the other tangential components of **E** to be zero on the boundaries because of the form of the relations 8.2(13)–(16).

***TE* Waves** Transverse electric waves have zero E_z and nonzero H_z so that the start is from Eq. 8.2(18), again expressed in rectangular coordinates,

$$\nabla_t^2 H_z = \frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} = -k_c^2 H_z \tag{15}$$

Solution by the separation of variables techniques of Sec. 7.19 gives

$$H_z = (A'' \sin k_x x + B'' \cos k_x x)(C'' \sin k_y y + D'' \cos k_y y) \tag{16}$$

where

$$k_c^2 = k_x^2 + k_y^2 \tag{17}$$

Imposition of boundary conditions in this case is a little less direct, but from Eqs. 8.2(13) and 8.2(14) we find electric field components as

$$E_x = -\frac{j\omega\mu}{k_c^2}\frac{\partial H_z}{\partial y}$$

$$= -\frac{j\omega\mu k_y}{k_c^2}(A'' \sin k_x x + B'' \cos k_x x)(C'' \cos k_y y - D'' \sin k_y y) \tag{18}$$

$$E_y = \frac{j\omega\mu}{k_c^2}\frac{\partial H_z}{\partial x} \tag{19}$$

$$= \frac{j\omega\mu k_x}{k_c^2}(A'' \cos k_x x - B'' \sin k_x x)(C'' \sin k_y y + D'' \cos k_y y)$$

In order for E_x to be zero at $y = 0$ for all x, $C'' = 0$ and for $E_y = 0$ at $x = 0$ for all y, $A'' = 0$. Defining $B''D'' = B$, we have then

$$H_z = B \cos k_x x \cos k_y y \tag{20}$$

We also require E_x to be zero at $y = b$ so that $k_y b$ must be a multiple of π. E_y is zero at $x = a$ so that $k_x a$ is also a multiple of π:

$$k_x a = m\pi, \qquad k_y b = n\pi \tag{21}$$

In contrast to the *TM* waves, one but not both of m and n may be zero without the wave's vanishing. Although we found the boundary conditions by first calculating electric field, we can see from the way in which **E** is related to H_z that the derivative of H_z normal to the conducting boundary must be zero in order for the tangential electric field to be zero there, so boundary conditions can be imposed directly on the form (16) without requiring the explicit forms for E_x and E_y.

The forms of transverse electric field with the derived simplifications to (18) and (19) are

$$E_x = \frac{j\omega\mu k_y}{k^2_{c_{m,n}}} B \cos k_x x \sin k_y y \tag{22}$$

$$E_y = -\frac{j\omega\mu k_x}{k^2_{c_{m,n}}} B \sin k_x x \cos k_y y \tag{23}$$

Corresponding transverse magnetic field components from Eqs. 8.2(15) and 8.2(16) are

$$H_x = \frac{j\beta k_x}{k^2_{c_{m,n}}} B \sin k_x x \cos k_y y \tag{24}$$

$$H_y = \frac{j\beta k_y}{k^2_{c_{m,n}}} B \cos k_x x \sin k_y y \tag{25}$$

Since comparison of (21) with (5) and (6) shows that k_x and k_y have the same forms for TM and TE waves, cutoff frequency and propagation characteristics for a TE_{mn} mode found from (17) are exactly the same as for the same order TM_{mn} mode. That is, the expressions (7), (8), and (9) apply here without change. Modes that have different field distributions but the same cutoff frequencies are said to be *degenerate modes*.

Table 8.7 gives the field distribution for several different TE modes. Since electric field is confined to the transverse plane, we find that for each one of the TE modes shown, electric fields begin on charges for a portion of the boundary and end on charges of opposite sign on another portion, in the same x-y plane. Magnetic field lines surround the displacement currents represented by the changing transverse electric fields. For the TE_{10} mode having no variations in the vertical direction, electric fields go between top and bottom of the guide in straight lines, and magnetic fields lie entirely in planes parallel to top and bottom. The TE_{10} mode is so important that it will be discussed separately in the following section.

Figure 8.7*b* shows a line diagram indicating the cutoff frequencies of several of the lowest order modes referred to that of the so-called *dominant* TE_{10} mode for a guide with a side ratio $b/a = \frac{1}{2}$, which is close to the value used in most practical guides. Normally, such a guide is designed so that its cutoff frequency for the TE_{10} mode is somewhat (say, 30%) below the operating frequency. In this way only one mode can propagate so signal distortion caused by multimode propagation is avoided. Also, by not being too close to the cutoff frequency, dispersion caused by having different group velocities for different frequency components of the signal is minimized for the one propagating mode. Higher

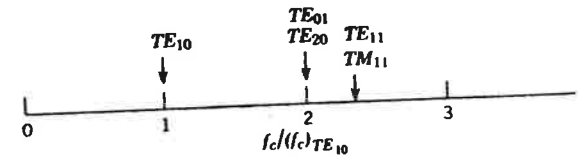


Fig. 8.7*b* Relative cutoff frequencies of waves in a rectangular guide ($b/a = \frac{1}{2}$).

order modes may be excited at the entrance to the guide but they are below their cutoff frequencies and die away in a short distance from the source.

The attenuation constant for TE_{mn} ($n \neq 0$) modes is found using the procedure leading to Eq. 8.5(12):

$$(\alpha_c)_{TE_{mn}} = \frac{2R_s}{b\eta\sqrt{1-(f_c/f)^2}}\left\{\left(1+\frac{b}{a}\right)\left(\frac{f_c}{f}\right)^2 + \left[1-\left(\frac{f_c}{f}\right)^2\right]\left[\frac{(b/a)((b/a)m^2+n^2)}{(b^2m^2/a^2)+n^2}\right]\right\} \tag{26}$$

And for TE_{m0} modes

$$(\alpha_c)_{TE_{m0}} = \frac{R_s}{b\eta\sqrt{1-(f_c/f)^2}}\left[1+\frac{2b}{a}\left(\frac{f_c}{f}\right)^2\right] \tag{27}$$

Figure 8.7*c* shows attenuation versus frequency for TM_{11} and TE_{10} modes in rectangular copper waveguides with various side ratios b/a found using (14) and (27), respectively.

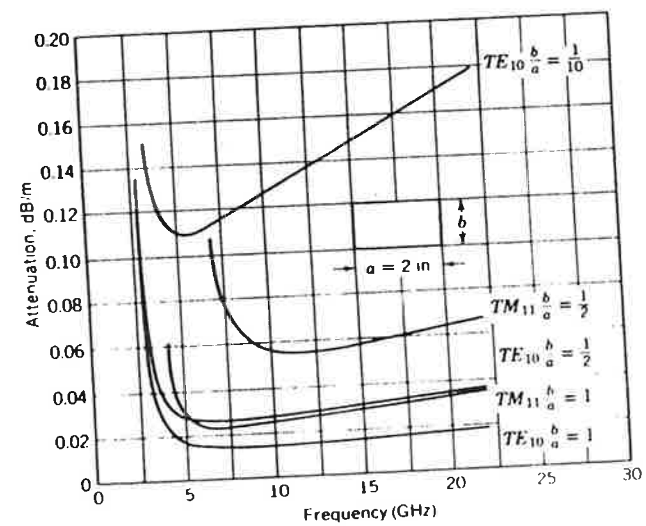


Fig. 8.7*c* Attenuation due to copper losses in rectangular waveguides of fixed width.

8.8 The TE_{10} Wave in a Rectangular Guide

One of the simplest of all the waves which may exist inside hollow-pipe waveguides is the dominant TE_{10} wave in the rectangular guide, which is one of the TE modes studied in the preceding section. This mode is of great engineering importance, partly for the following reasons:

1. Cutoff frequency is independent of one of the dimensions of the cross section. Consequently, for a given frequency this dimension may be made small enough so that the TE_{10} wave is the only wave which will propagate, and there is no difficulty with higher-order waves that end effects or discontinuities may cause to be excited.
2. The polarization of the field is definitely fixed, electric field passing from top to bottom of the guide. This fixed polarization may be required for certain applications.
3. For a given frequency the attenuation due to copper losses is not excessive compared with other wave types in guides of comparable size.

Let us now rewrite the expressions from the previous section for general TE waves in rectangular guides, Eqs. 8.7(20)–(24), setting $m = 1$, $n = 0$, in which case $k_y = 0$ and $k_c = k_x = \pi/a$.

$$H_z = B\cos k_x x \tag{1}$$

$$E_y = -\frac{j\omega\mu B}{k_x}\sin k_x x \tag{2}$$

$$H_x = \frac{j\beta B}{k_x}\sin k_x x \tag{3}$$

All other components are zero. This set may be rewritten in a useful alternate form,

$$E_y = -Z_{TE}H_x = E_0\sin\left(\frac{\pi x}{a}\right) \tag{4}$$

$$H_z = \frac{jE_0}{\eta}\left(\frac{\lambda}{2a}\right)\cos\left(\frac{\pi x}{a}\right) \tag{5}$$

where

$$E_0 = -\frac{j\omega\mu B}{k_x} = -\frac{j2\eta a B}{\lambda} \tag{6}$$

$$Z_{TE} = \eta\left[1-\left(\frac{\omega_c}{\omega}\right)^2\right]^{-1/2} = \eta\left[1-\left(\frac{\lambda}{2a}\right)^2\right]^{-1/2} \tag{7}$$

$$\eta = \sqrt{\frac{\mu}{\varepsilon}}, \qquad \lambda = \frac{v}{f} = \frac{2\pi}{\omega\sqrt{\mu\varepsilon}} \tag{8}$$

Cutoff frequency, wavelength, and wavenumber are

$$f_c = \frac{1}{2a\sqrt{\mu\varepsilon}}, \qquad \lambda_c = 2a, \qquad k_c = \frac{\pi}{a} \tag{9}$$

Phase and group velocities, and wavelength measured along the guide, are

$$v_p = \frac{1}{\sqrt{\mu\varepsilon}\sqrt{1-(\lambda/2a)^2}}, \qquad v_g = \frac{1}{\sqrt{\mu\varepsilon}}\sqrt{1-\left(\frac{\lambda}{2a}\right)^2} \tag{10}$$

$$\lambda_g = \frac{v_p}{f} = \frac{2\pi}{\beta} = \frac{\lambda}{\sqrt{1-(\lambda/2a)^2}} \tag{11}$$

The attenuation arising from an imperfect dielectric is obtained by replacing ε with $\varepsilon' - j\varepsilon''$ in the equation for γ. Since k_c in the TE_{m0} modes is of the same form as in the parallel-plane modes, Eq. 8.5(3) applies here and leads to a result equivalent to Eq. 8.5(4)

$$\alpha_d = \frac{k\varepsilon''/\varepsilon'}{2\sqrt{1-(\lambda/2a)^2}} \tag{12}$$

To find attenuation if the conductor is imperfect, we first calculate the power transferred by the wave from the Poynting theorem,

$$W_T = \frac{1}{2}\,\mathrm{Re}\int_0^a\int_0^b(-E_y H_x^*)\,dx\,dy \tag{13}$$

Utilizing the forms (4) we have

$$W_T = \frac{E_0^2 b}{2Z_{TE}}\int_0^a \sin^2\frac{\pi x}{a}\,dx = \frac{E_0^2 ba}{4Z_{TE}} \tag{14}$$

Next we find approximate losses in the walls by using currents of the ideal mode in material of surface resistivity R_s. Current in a conductor is related to the tangential magnetic field H_z at the side walls $x = 0$ and $x = a$ so there is current per unit width $|J_{sy}| = |H_z|$ there. Both components H_x and H_z are tangential at top and bottom surfaces giving rise to surface current densities $|J_{sz}| = |H_x|$ and $|J_{sx}| = |H_z|$. Thus power loss per unit length is

$$(w_L)_{\text{SIDES}} = 2\left(\frac{bR_s}{2}|H_z|^2_{x=0}\right) = \frac{bR_s E_0^2\lambda^2}{4\eta^2 a^2} \tag{15}$$

$$\begin{aligned}(w_L)_{\text{TOP AND BOTTOM}} &= 2\,\frac{R_s}{2}\int_0^a(|H_x|^2 + |H_z|^2)\,dx \\ &= R_s\int_0^a\left[\frac{E_0^2}{Z_{TE}^2}\sin^2\frac{\pi x}{a} + \frac{E_0^2\lambda^2}{4\eta^2 a^2}\cos^2\frac{\pi x}{a}\right]dx \\ &= \frac{a}{2}R_s\left(\frac{E_0^2}{Z_{TE}^2} + \frac{E_0^2\lambda^2}{4\eta^2 a^2}\right)\end{aligned} \tag{16}$$

Adding the two contributions (15) and (16) and substituting Z_{TE} from (7),

$$w_L = \frac{R_s E_0^2}{\eta^2}\left[\frac{b\lambda^2}{4a^2} + \frac{a}{2}\left(1 - \frac{\lambda^2}{4a^2} + \frac{\lambda^2}{4a^2}\right)\right] = \frac{R_s E_0^2}{2\eta^2}\left(a + \frac{b\lambda^2}{2a^2}\right)$$

The attenuation from conductor losses, Eq. 5.9(4), is then

$$\alpha_c = \frac{w_L}{2W_T} = \frac{R_s Z_{TE}}{\eta^2 ba}\left(a + \frac{b\lambda^2}{2a^2}\right) \tag{17}$$

or

$$\alpha_c = \frac{R_s}{b\eta\sqrt{1 - (\lambda/2a)^2}}\left[1 + \frac{2b}{a}\left(\frac{\lambda}{2a}\right)^2\right] \tag{18}$$

A study of the field distributions (1) to (3) or (4) and (5) shows the field patterns for this wave sketched in Table 8.7. First it is noted that no field components vary in the vertical or y direction. The only electric field component is the vertical one E_y passing between top and bottom of the guide. This is a maximum at the center and zero at the conducting walls, varying as a half-sine curve. The corresponding charges induced by the electric field lines ending on conductors are

1. Charges zero on side walls.
2. A charge distribution on top and bottom with $\rho_s = \varepsilon E_y$ on the bottom, $-\varepsilon E_y$ on the top.

The magnetic field forms closed paths surrounding the vertical electric displacement currents arising from E_y, so that there are components H_x and H_z. Component H_x is zero at the two side walls and a maximum in the center, following the distribution of E_y. Component H_z is a maximum at the side walls and zero at the center. Component H_x corresponds to a longitudinal current flow down the guide in the top, and opposite in the bottom; H_z corresponds to transverse currents in the top and bottom and vertical currents on the side walls. These current distributions are sketched in Fig. 8.8*a*.

This simple wave type is a convenient one to study in order to strengthen some of our physical pictures of wave propagation. Electric field is confined to the transverse plane and so passes between opposite charges of equal density on the top and bottom. The electric field E_y and the transverse magnetic field H_x are maximum at planes marked A. Halfway between those planes is the maximum rate of change of E_y and therefore the location of maximum displacement current. The currents concomitant to the tangential magnetic field feed into the displacement currents so there is continuity of current. The magnetic fields surround the electric displacement currents inside the guide and so must have an axial as well as a transverse component.

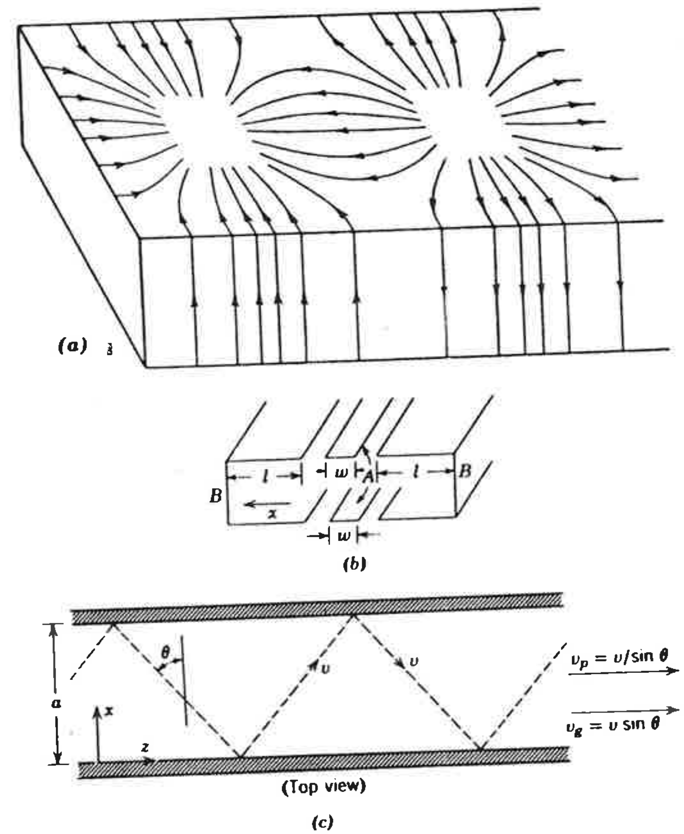


Fig. 8.8 (*a*) Current flow in walls of rectangular guide with TE_{10} mode. (*b*) Guide roughly divided into axial- and transverse-current regions. (*c*) Path of uniform plane wave component of TE_{10} wave in rectangular guide.

As a fairly crude way of looking at the problem, one might also think of this mode being formed by starting with a parallel plate transmission line A of width w to carry the longitudinal current in the center of the guide, and then adding shorted troughs B of depth l on the two sides to close the region, as pictured in Fig. 8.8*b*. Since one would expect the lengths l to be around a quarter-wavelength to provide a high impedance at the center, the overall width should be something over a half-wavelength, which we know to be true for propagation. The picture is only a rough one because the fields in the two regions are not separated, and propagation is not purely longitudinal in the center portion or transverse in the side portions.

A third viewpoint follows from that used in studying the higher order waves between parallel planes. There it was pointed out that one could visualize the

TM and *TE* waves in terms of plane waves bouncing between the two planes at such an angle that the interference pattern maintains a zero of electric field tangential to the two planes. Similarly, the TE_{10} wave in the rectangular guide may be thought of as arising from the interference between incident and reflected plane waves, polarized so that the electric vector is vertical, and bouncing between the two sides of the guide at such an angle with the sides that the zero electric field is maintained at the two sides. One such component uniform plane wave is indicated in Fig. 8.8*c*. As in the result of Sec. 8.4, when the width a is exactly $\lambda/2$, the waves travel exactly back and forth across the guide with no component of propagation in the axial direction. At slightly higher frequencies there is a small angle θ such that $a = \lambda/2\cos\theta$, and there is a small propagation in the axial direction, a very small group velocity in the axial direction $v\sin\theta$ and a very large phase velocity $v/\sin\theta$. At frequencies approaching infinity, θ approaches 90 degrees, so that the wave travels down the guide practically as a plane wave in space propagating in the axial direction.

All the foregoing points of view explain why the dimension b should not enter into the determination of cutoff frequency. Since the electric field is always normal to top and bottom, the placing of these planes plays no part in the boundary condition. However, this dimension b will be important from three other points of view.

1. It is desired to have a large spacing between the cutoff frequency of the TE_{10} mode and that of the next higher mode. If $b/a > \frac{1}{2}$ the TE_{01} mode has the next lowest f_c and that is raised by decreasing b.
2. The smaller b is (all other parameters constant), the greater is the electric field across the guide for a given power transfer, as is seen from (14), and so the danger of voltage breakdown is greater.
3. The smaller b is (all other parameters constant), the greater is the attenuation due to conductor losses. This may be seen by keeping fields constant as b approaches zero. Power transfer, by (14), and loss in the sides, by (15), approach zero, but loss in top and bottom remains constant at the value calculated in (16).

8.9 Circular Waveguides

Hollow-pipe waveguides of circular cross section are used in a number of instances, for example, when circular polarization is to be transmitted to certain classes of antennas. Also, as will be shown, the class of TE_{0n} modes (called *circular electric*) is interesting because of the low attenuation in this class at high frequencies. As before, we start with ideal dielectric and conducting boundary and make approximate modifications to these solutions when the materials have small losses. Before treating separately the *TM* and *TE* classes of waves, it is

desirable to have the set of equations 8.2(13)–(16) transformed to circular cylindrical coordinates. A straightforward transformation gives

$$E_r = -\frac{j}{k_c^2}\left[\beta\frac{\partial E_z}{\partial r} + \frac{\omega\mu}{r}\frac{\partial H_z}{\partial\phi}\right] \tag{1}$$

$$E_\phi = \frac{j}{k_c^2}\left[-\frac{\beta}{r}\frac{\partial E_z}{\partial\phi} + \omega\mu\frac{\partial H_z}{\partial r}\right] \tag{2}$$

$$H_r = \frac{j}{k_c^2}\left[\frac{\omega\varepsilon}{r}\frac{\partial E_z}{\partial\phi} - \beta\frac{\partial H_z}{\partial r}\right] \tag{3}$$

$$H_\phi = -\frac{j}{k_c^2}\left[\omega\varepsilon\frac{\partial E_z}{\partial r} + \frac{\beta}{r}\frac{\partial H_z}{\partial\phi}\right] \tag{4}$$

where

$$k_c^2 = \gamma^2 + k^2 = k^2 - \beta^2 \tag{5}$$

***TM* Waves** The transverse part of the Laplacian in Eq. 8.2(17) for E_z is expressed in circular cylindrical coordinates for this configuration, with the coordinate system as shown in Fig. 8.9*a*.

$$\nabla_t^2 E_z = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial E_z}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial\phi^2} = -k_c^2 E_z \tag{6}$$

Separation of variables techniques in Sec. 7.20 led to the solution

$$E_z(r,\phi) = [A'J_n(k_c r) + B'N_n(k_c r)][C'\cos n\phi + D'\sin n\phi] \tag{7}$$

where J_n and N_n are nth order Bessel functions of first and second kind, respectively. The second kind, $N_n(k_c r)$, is infinite at $r = 0$ for any n and so cannot be included in the interior solution which includes the axis. Also, for simplicity, we choose the origin of ϕ so that we have just the $\cos n\phi$ variation. Letting $A'C' = A$,

$$E_z = AJ_n(k_c r)\cos n\phi \tag{8}$$

From (1) to (4) with $H_z = 0$, the remaining field components are

$$E_r = Z_{TM}H_\phi = -\frac{j\beta}{k_c}AJ_n'(k_c r)\cos n\phi \tag{9}$$

$$E_\phi = -Z_{TM}H_r = \frac{j\beta n}{k_c^2 r}AJ_n(k_c r)\sin n\phi \tag{10}$$

where

$$Z_{TM} = \frac{\beta}{\omega\varepsilon} \tag{11}$$

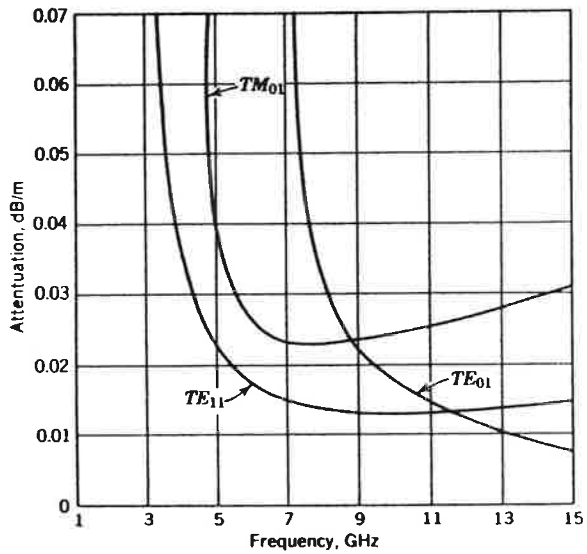


Fig. 8.9c Attenuation due to copper losses in circular waveguides; diameter = 2 in.

coupled to the guide walls at high frequencies. However, other modes may propagate, as shown by Fig. 8.9*b*, so that there are problems with mode conversion when such guides bend to go around corners. Practical ways of solving such problems have been developed[6] and this system has been used as a low-loss guiding system for millimeter waves.

8.10 Higher Order Modes on Coaxial Lines

The lowest-order mode on a coaxial line is a *TEM* wave; this was assumed implicitly in the transmission-line treatment of Ex. 5.2 where we used the capacitance and inductance found from static fields. As in the parallel-plane guide, higher order (*TM* and *TE*) modes can also exist. Normally, the line is designed in such a way that the cutoff frequencies of the higher order modes are well above the operating frequency. Even in that case, these modes can be of importance near discontinuities.

The general forms useful for the *TM* and *TE* modes in circular cylindrical coordinates are listed in Sec. 7.20. The boundary conditions require that E_z for *TM* waves be zero at r_0 and r_i (Fig. 8.10*a*).

[6] See, for example, S. E. Miller, *Bell System Tech. J.* 33, 1209 (1954).

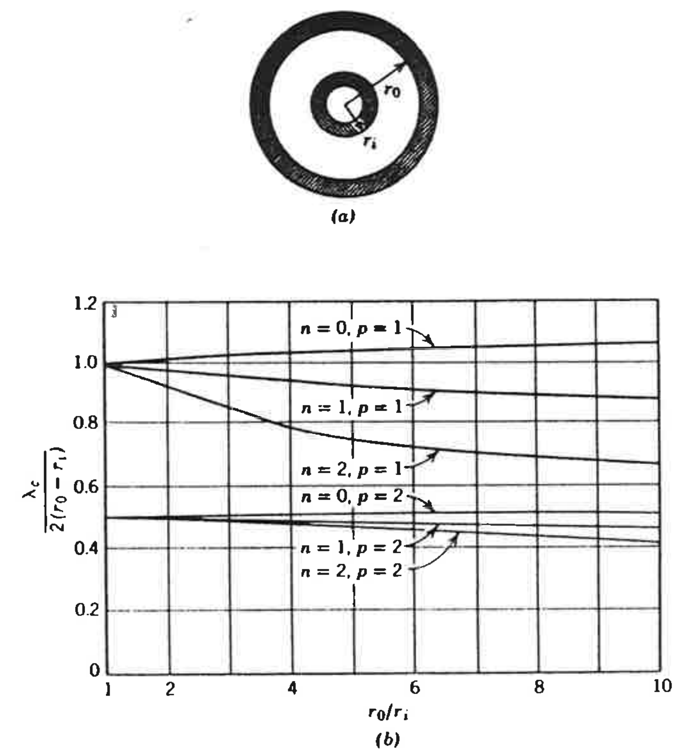


Fig. 8.10 (*a*) Cross section of a coaxial line. (*b*) Cutoff wavelength for some higher order *TM* waves in coaxial lines.

For *TM* waves,

$$A_n J_n(k_c r_i) + B_n N_n(k_c r_i) = 0$$

$$A_n J_n(k_c r_0) + B_n N_n(k_c r_0) = 0$$

or

$$\frac{N_n(k_c r_i)}{J_n(k_c r_i)} = \frac{N_n(k_c r_0)}{J_n(k_c r_0)} \tag{1}$$

For *TE* waves, the derivative of H_z normal to the two conductors must be zero at the inner and outer radii. [See discussion following Eq. 8.7(21).] Then, in place of (1),

$$\frac{N_n'(k_c r_i)}{J_n'(k_c r_i)} = \frac{N_n'(k_c r_0)}{J_n'(k_c r_0)} \tag{2}$$

Solutions to the transcendental equations (1) and (2) determine the values of k_c and hence cutoff frequency for any wave type and any particular values of r_i and r_0. Solution of the transcendental equations is accomplished by graphical or numerical methods. By analogy with the parallel-plane guide, we would expect to find certain modes with a cutoff such that the spacing between conductors is of the order of p half-wavelengths.

$$\lambda_c \approx \frac{2}{p}(r_0 - r_i), \qquad p = 1, 2, 3, \ldots \tag{3}$$

This is verified by Fig. 8.10*b* for values of r_0/r_i near unity.

Probably more important is the lowest order TE wave with circumferential variations. This is analogous to the TE_{10} wave of a rectangular waveguide, and physical reasoning from the analogy leads one to expect cutoff for this wave type when the average circumference is about equal to wavelength. The field picture of the TE_{10} mode given in Sec. 8.8 should make this reasonable. Solution of (2) reveals this simple rule to be within about 4% accuracy for r_0/r_i up to 5. In general, for the nth order TE wave with circumferential variations,

$$\lambda_c \approx \frac{2\pi}{n}\left(\frac{r_0 + r_i}{2}\right), \qquad n = 1, 2, 3, \ldots \tag{4}$$

There are, of course, other TE waves with further radial variations, and the lowest order of these has a cutoff about the same as the lowest order TM wave.

8.11 Excitation and Reception of Waves in Guides

The problems of exciting or receiving waves in a waveguide are not simple field problems. In this section we will give only a qualitative introduction to the manners of excitation of fields in various kinds of guides. Approaches to analysis and measurement of these junctions are given in Chapter 11. Reception of the energy of a wave uses the same kind of structure as excitation and is just the reverse process. To excite any particular desired wave, one should study the field pattern and use one of the following concepts.

1. Introduce the excitation in a probe or antenna oriented in the direction of electric field. The probe is most often placed near a maximum of the electric field of the mode pattern, but exact placing is a matter of impedance matching. Examples are shown in Figs. 8.11*a* and 8.11*b*.
2. Introduce the excitation through a loop oriented in a plane normal to magnetic field of the mode pattern (Fig. 8.11*c*).

3. Couple to the desired mode from another guiding system, by means of a hole or iris, the two guiding systems having some common field component over the extent of the hole. An example of coupling between waveguides using a large iris is shown in Fig. 8.11*d*. The coupling is sometimes done with a small hole as for coupling to resonant cavities (Sec. 10.10).
4. Introduce currents from one kind of transmission line into another, as in coupling from a coaxial line to microstrip shown in Fig. 8.11*e*.
5. For higher order waves combine as many of the exciting sources as are required, with proper phasings (Fig. 8.11*f*).
6. Gradually taper a transition between two types of guides, as for a TE_{10} wave in a rectangular guide to a TE_{11} in a circular guide.

Since most of these exciting methods are in the nature of concentrated sources, they will not in general excite purely one wave, but all waves that have field components in a favorable direction for the particular exciting source. That is, we see that one wave alone will not suffice to satisfy the boundary conditions of the guide complicated by the exciting source, so that many higher order waves must be added for this purpose. If the guide is large enough, several of these waves will then proceed to propagate. Most often, however, only one of the excited waves is above cutoff. This will propagate down the guide, and (if absorbed somewhere) will represent a resistive load on the source, comparable to the radiation resistance of antennas which we shall encounter further in Chapter 12. The higher order waves that are excited, if all below cutoff, will be localized in the neighborhood of the source and will represent purely reactive loads on the source. For practical application, it is then necessary to add, in the line that feeds the probe or loop or other exciting means, an arrangement for matching to the load that has a real part representing the propagating wave and an imaginary part representing the localized reactive waves.

Example 8.11

Excitation of a Waveguide by a Coaxial Line

Let us look in more depth at the structure in Fig. 8.11*b* where a coaxial line is inserted in the center of the broad side of a waveguide of rectangular cross section in order to excite a TE_{10} mode. The waveguide is short circuited at a distance l from the probe to aid in matching the coaxial line to the waveguide. The fields associated with the probe excite both the desired TE_{10} mode and other higher order modes. The latter are cutoff and do not propagate, but they store reactive energy and therefore constitute a reactive component of the load on the coaxial line. Proper choice of the size and location of the probe for a given frequency and guide dimensions make the standing wave between the probe and the shorted end

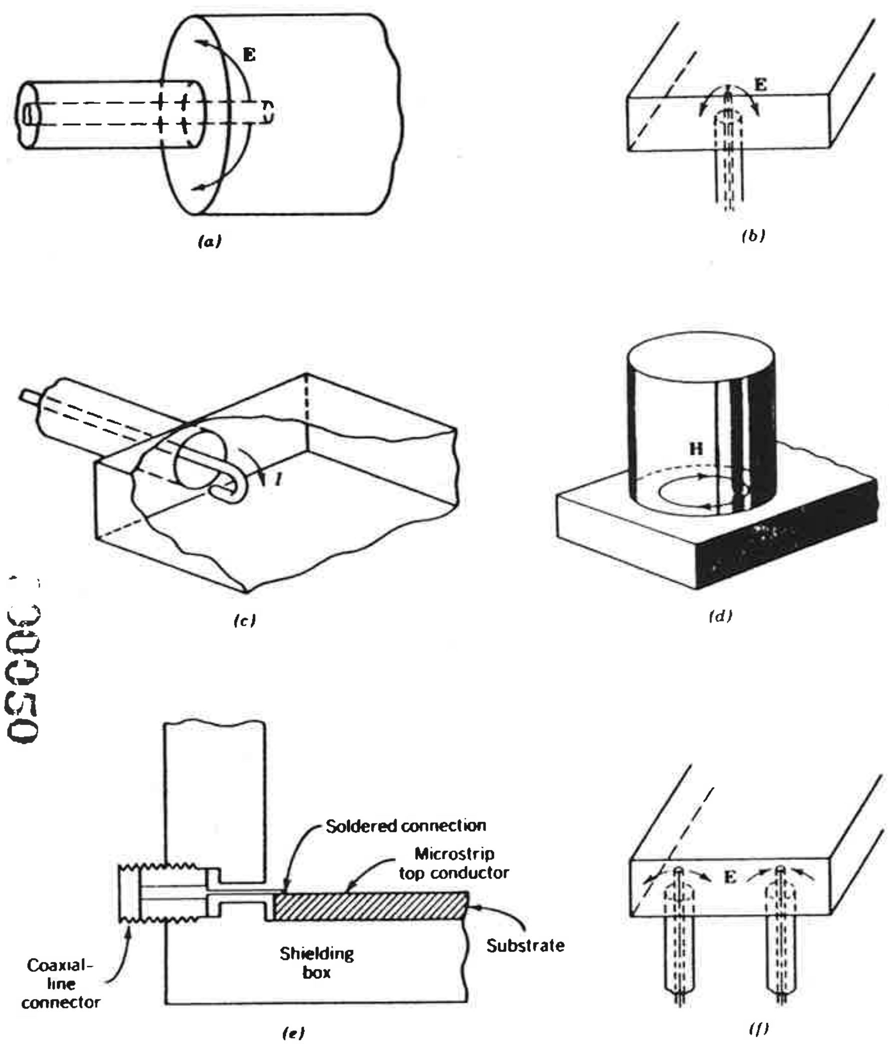


Fig. 8.11 (*a*) Antenna in end of circular guide for excitation of TM_{01} wave. (*b*) Antenna in bottom of rectangular guide for excitation of the TE_{10} wave. (*c*) Loop in end of rectangular guide for excitation of TE_{10} wave. (*d*) Junction between circular guide (TM_{01} wave) and rectangular guide (TE_{10} wave); large-aperture coupling. (*e*) Coaxial line coupling to microstrip. (*f*) Excitation of the TE_{20} wave in rectangular guide by two oppositely phased antennas.

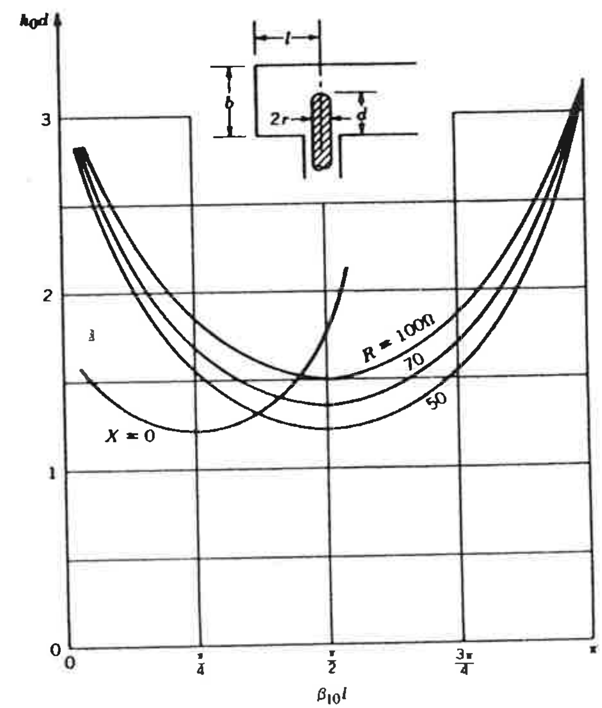


Fig. 8.11g Impedance parameters for coaxial line-to-waveguide transition in inset. $X = 0$ and R = constant contours in $k_0d - \beta_{10}l$ plane for $a = 2.29$ cm, $b = 1.02$ cm, $r = 0.05a$, and $\lambda_0 = 3.14$ cm. From Ref. 7.

contain reactive energy of opposite sign and equal magnitude so that the net reactive component of the input impedance is zero. These adjustments are used to make the real part of the load impedance on the coaxial line equal to its characteristic impedance so that perfect matching is achieved and all the power is coupled into the guide. Figure 8.11g shows the calculated results for a probe of given radius in a guide of dimensions appropriate for use at about 10 GHz (X band).[7] The probe length and location required to make the reactive load X zero and for a selected resistive load R can be read from the graph. Similar graphs of R and X can be calculated using the formulas in Ref. 7 for other guide sizes and probe radii.

[7] R. E. Collin, *Field Theory of Guided Waves*, McGraw-Hill, New York, 1960, Sec. 7.1.

Matrices for which the transpose is the conjugate of the inverse matrix are called ***unitary matrices***. Use of the product rule for matrices show that they have the following properties:

$$\sum_{n=1}^{N} S_{in} S_{in}^* = 1 \tag{18}$$

$$\sum_{n=1}^{N} S_{in} S_{jn}^* = 0, \qquad i \neq j \tag{19}$$

The above relations may be derived directly from conservation of energy, (18) by applying an incident wave to terminal i with all terminals matched, and (19) by applying incident waves to terminals i and j with all terminals matched. Note that reciprocity was not required in the derivation. The relations have important consequences as to what can or cannot be done with loss-free junctions, as will be illustrated with two examples.

Example 11.10a
Limitations on Loss-Free Three Ports

The three-port Y junction pictured in Fig. 11.10*b* is useful as a power divider or power combiner. It is assumed that it satisfies reciprocity. If sources are introduced at terminals 1 and 2 with the combined power obtained at 3, one might wish to have $S_{12} = 0$ in order to eliminate direct interaction between the two sources. But condition (19), with $i = 1, j = 2$, gives

$$S_{11}S_{21}^* + S_{12}S_{22}^* + S_{13}S_{23}^* = 0 \tag{20}$$

Thus if $S_{12} = 0$, either S_{13} or S_{23} is zero also, eliminating one of the two desired couplings. The junction will act as a power combiner, but there is interaction between the two sources, and if the two sources are not identical in magnitude and phase, one source will tend to feed power to the other.

Example 11.10b
Limitations on Ideal Isolating Networks

The ideal isolator would be a loss-free, one-way transmission line with $S_{12} = 0$ but $S_{21} \neq 0$. It was noted that the unitary property (17) applies to nonreciprocal as well as reciprocal networks. Equation (19) with $i = 1, j = 2$ gives

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

so that if $S_{12} = 0$, either $S_{21} = 0$ or $S_{11} = 0$, so this ideal also is impossible. We shall see useful isolators employing nonreciprocal elements in Chapter 13, but because of the limitation shown here, they will have dissipative elements to absorb the reflected wave. (Note that $S_{11} = 0$ when $S_{12} = 0$ follows from (18).)

Some consequences of the unitary property to directional couplers and four-port hybrid networks are discussed in the following section.

11.11 Directional Couplers and Hybrid Networks

One of the most important four ports is the directional coupler, designed to couple in a separable fashion to the positively and negatively traveling waves in a guide. Figure 11.11*a* gives the simplest conception of this device. Imagine a main waveguide with two small holes placed a quarter-wave apart coupling to an auxiliary guide terminated at each end by a matching resistance and meter as shown. If wave A progresses toward the right, coupled waves from the two holes at terminal 4 follow paths B and C of equal lengths, and the contributions add in that load, its meter indicating the strength of A. The couplings through the two holes cancel at terminal 3, however, since the paths E and D differ in length by a half-wavelength, and the couplings through the two holes are substantially the same in amount if the holes are small. By symmetry, a wave flowing to the left will register at terminal

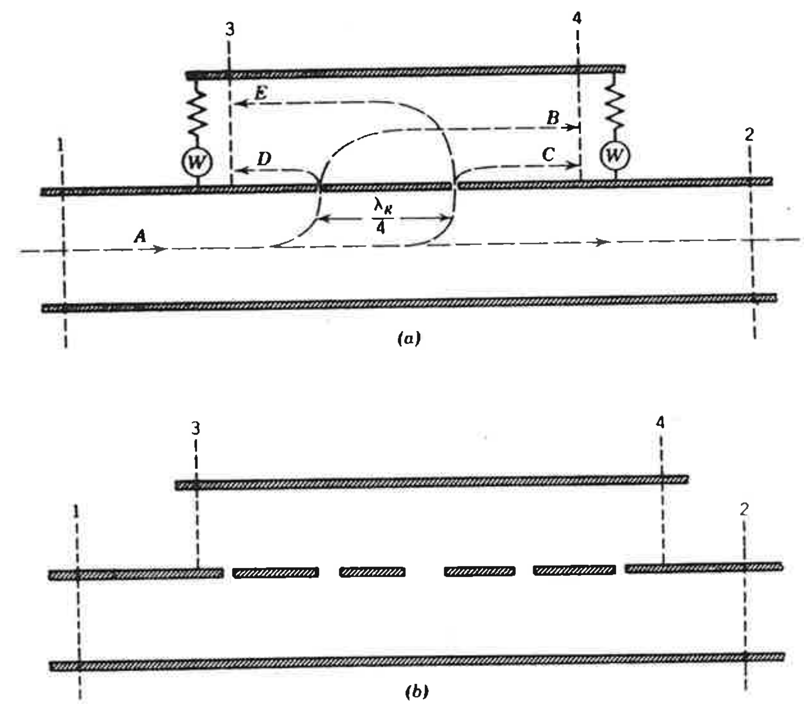


Fig. 11.11 (*a*) Basic directional coupler. (*b*) Broadband directional coupler.

3 but yield coupled waves which cancel at 4. Thus meter 4 reads the strength of the wave to the right, and meter 3 that to the left.

This simple coupler is frequency sensitive since it depends on the quarter-wave spacing of holes. A like effect with greater bandwidth may be obtained by supplying several holes with properly graded couplings, as illustrated in Fig. 11.11*b*. Still other embodiments are described in the references. All these couplers may be considered as four ports with the four reference planes as shown in Fig. 11.11*a* or Fig. 11.11*b*. Losses may normally be neglected. Several important general properties follow.

To study the properties of the coupler, it is most convenient to use the scattering matrix form of Eq. 11.10(3). It is desired not to couple between 1 and 3 with 2 and 4 matched, so $S_{13} = S_{31} = 0$. It is also desired to have no coupling between 2 and 4 with 1 and 3 matched, so $S_{24} = S_{42} = 0$. Moreover, the ideal directional coupler should be matched so that all the power entering at one terminal divides between the other two for which there is coupling, leaving no reflections at the input. Thus S_{11}, S_{22}, and so on, are zero. The network is also assumed to satisfy reciprocity so that $S_{12} = S_{21}$, $S_{14} = S_{41}$, and so on. Thus the scattering matrix for an ideal directional coupler has been specialized to

$$[S] = \begin{bmatrix} 0 & S_{12} & 0 & S_{14} \\ S_{12} & 0 & S_{23} & 0 \\ 0 & S_{23} & 0 & S_{34} \\ S_{14} & 0 & S_{34} & 0 \end{bmatrix} \tag{1}$$

For negligible loss within the network, power conservation leads to the unitary property of the scattering matrix as shown in the preceding section. Equation 11.10(18), for different values of the index i, gives

$$i = 1: \quad S_{12}S_{12}^* + S_{14}S_{14}^* = 1 \tag{2}$$

$$i = 2: \quad S_{12}S_{12}^* + S_{23}S_{23}^* = 1 \tag{3}$$

$$i = 3: \quad S_{23}S_{23}^* + S_{34}S_{34}^* = 1 \tag{4}$$

$$i = 4: \quad S_{14}S_{14}^* + S_{34}S_{34}^* = 1 \tag{5}$$

Comparison of (2) and (3) shows that $|S_{14}| = |S_{23}|$ and comparison of (2) and (5) shows that $|S_{12}| = |S_{34}|$. Moreover, reference plane 2 may be selected with respect to 1 so that S_{12} is real and positive, and similarly 4 with respect to 3 so that S_{34} is real and positive. Then

$$S_{12} = S_{34} \triangleq a \tag{6}$$

There remains the Eqs. 11.10(19), also following from the unitary nature of $[S]$. The specializations already made in (1) satisfy these identically except for

$$i = 1, \quad j = 3: \quad S_{12}S_{23}^* + S_{14}S_{34}^* = 0 \tag{7}$$

$$i = 2, \quad j = 4: \quad S_{12}S_{14}^* + S_{23}S_{34}^* = 0 \tag{8}$$

Use of (6) in either of the above requires $S_{14} = -S_{23}^*$. Reference plane 4 may then be selected with respect to 1 so that S_{14} is real and

$$S_{23} = -S_{14} \triangleq b \tag{9}$$

Thus we have reduced the scattering matrix to the very simple form

$$[S] = \begin{bmatrix} 0 & a & 0 & -b \\ a & 0 & b & 0 \\ 0 & b & 0 & a \\ -b & 0 & a & 0 \end{bmatrix} \tag{10}$$

Note that b gives the coupling from the main guide to the auxiliary guide and is known as the coupling factor (often expressed in decibels). The coefficient a may be called the transmission factor and the two are related by any of the energy relations (2) to (5)

$$a^2 + b^2 = 1 \tag{11}$$

Although a and b are the only two parameters of an ideal directional coupler, real units give some coupling to the terminal for which zero coupling is desired. That is, S_{13} and S_{24} will not be exactly zero for a real coupler. The coupling to the desired terminal in the auxiliary guide, as compared with the undesired terminal, is defined as the *front-to-back ratio* or the *directivity*, usually expressed in decibels.

Several important theorems may be proved for loss-free reciprocal four ports.

1. A four port with two pairs of noncoupling elements is completely matched. That is, the setting of S_{11}, S_{22}, and so on equal to zero was not a separate condition but followed from $S_{13} = 0$, $S_{24} = 0$ because of power relations resulting from the complete Eqs. 11.10(18) and 11.10(19).
2. Any completely matched junction of four waveguides is a directional coupler. (Note that this does not mean that an arbitrary four port may be made into a directional coupler by externally introducing matching transformers, since the adjustment of one of these in such a case disturbs matching for the other ports; it must be an internal property giving $S_{11} = S_{22} = S_{33} = S_{44} = 0$.)
3. A four port with two noncoupling terminals matched is a directional coupler. That is, if $S_{13} = 0$, $S_{11} = 0$, and $S_{33} = 0$, the other properties defined earlier follow.

The Magic T and Other Hybrid Networks The special case of a directional coupler with $a^2 = b^2 = \frac{1}{2}$ is of particular interest in that it may be used as a bridge or "hybrid" network. (The latter name is taken from the properties of the classical hybrid coil.[15]) One of the most common configurations used for this

[15] C. G. Montgomery, R. H. Dicke, and E. M. Purcell, Ref. 4, p. 307