$$\nabla \cdot \mathbf{B} = 0 \tag{7.9-2}$$

Under static conditions the $\vec{\mathbf{B}}$ field lines, when extended sufficiently, form closed contours.

7.10 AMPERE'S LAW

Just as Gauss's law related electric flux density to contained charge, similarly, Ampere's law relates magnetic field to current.

Ampere's law states that the line integral of the $\vec{\mathbf{H}}$ *field parallel to the line defining a closed path is equal to the current I enclosed by the path:*

$$\oint \vec{\mathbf{H}} \cdot \overrightarrow{\mathrm{dl}} = \mathrm{I} \tag{7.10-1}$$

For the geometry shown in Figure 7.10-1, consisting of a circle coaxial with a conductor carrying current I, symmetry requires that the H_φ is independent of φ. Since $\mathrm{dl} = \mathrm{r}\,\mathrm{d}\varphi$, integration around the closed circular path, from $\varphi = 0$ to $\varphi = 2\pi$, leads to

$$2\pi \mathrm{r} \mathrm{H}_\varphi = \mathrm{I} \quad \text{or} \quad \mathrm{H}_\varphi = \frac{\mathrm{I}}{2\pi \mathrm{r}} \tag{7.10-2}$$

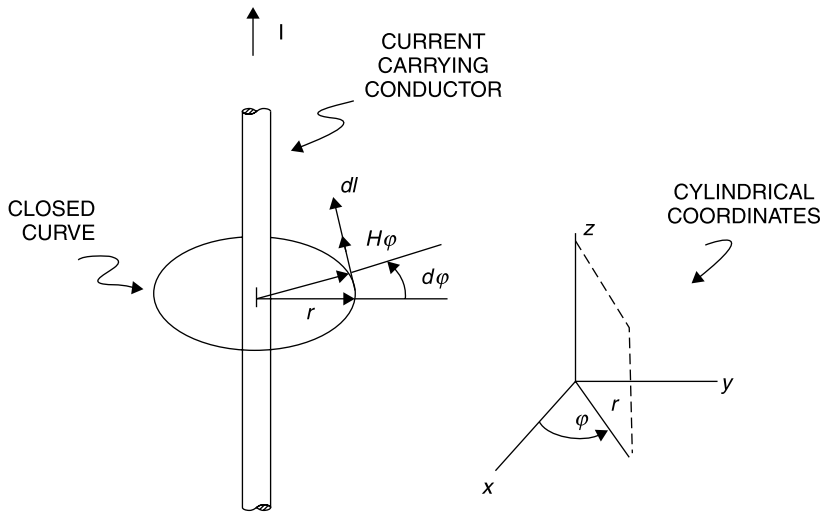

Figure 7.10-1 Closed curve defined about a current-carrying conductor (current I is in the +z direction).

This is a static relation. Both H_φ and I are constant (DC) values that are assumed to have been in place for "an infinite period of time." The effects of time-varying fields and currents will be discussed later.

7.11 VECTOR CURL

Figure 7.10-1 describes a situation wherein the integration of the force field ($\vec{\mathbf{H}}$ in this case) around a closed path *is not zero*. This phenomenon occurs often in the physical world, for example, in the rotational force of a tornado. Thus the integral around the closed path represents the rotational component of the field. This concept can be defined at a point by taking the limit as the path and its enclosed area tend to zero (become infinitesimal). For a given vector field $\vec{\mathbf{A}}$, this mathematical operation is known as the curl of $\vec{\mathbf{A}}$ (*pronounced "del cross* $\vec{\mathbf{A}}$*" or the "curl of* $\vec{\mathbf{A}}$*"*) and is written as $\nabla \times \vec{\mathbf{A}}$.

The curl of a vector $\vec{\mathbf{A}}$ *at a point i is equal to the limit as* $\Delta S_i \to 0$ *of the line integral of* $\vec{\mathbf{A}} \cdot \overrightarrow{\mathrm{dl}}$ *about a path enclosing surface* S_i*:*

$$[\text{curl } \vec{\mathbf{A}}]_i \equiv \nabla \times \vec{\mathbf{A}} \equiv \lim_{\Delta S_i \to 0} \frac{\oint \vec{\mathbf{A}} \cdot \overrightarrow{\mathrm{dl}}}{\Delta S_i} \tag{7.11-1}$$

This curl is a general vector identity and can be applied to any vector field. *The curl of a vector field is also a vector field.* The curl can be envisioned by imagining that the $\vec{\mathbf{A}}$ field represents a wind or water force on a paddle wheel placed in the field [4, Chapter 2]. For example, if a paddle wheel with its rotational axis oriented in the z direction is placed in the xy plane in Figure 7.11-1 and if the line integral of $\vec{\mathbf{A}} \cdot \overrightarrow{\mathrm{dl}}$ is positive for the closed path shown, then the paddle wheel would rotate counterclockwise when viewed from above on the +z axis, as shown.

The paddle wheel test is not as whimsical as it first appears. It might seem that the curl of a vector is simply a measure of its circulation, but this is not so. It is possible to have a field that forms circles, yet whose curl is zero, as will be shown by the coaxial line example later in this section.

The components of the curl can be calculated separately. For example, in rectangular coordinates the z component of the curl is calculated using the geometry of Figure 7.11-1 [6, pp. 96–97] for a surface with sides dx and dy:

$$\oint \vec{A} \cdot \overrightarrow{dl} = dy\, A_y$$

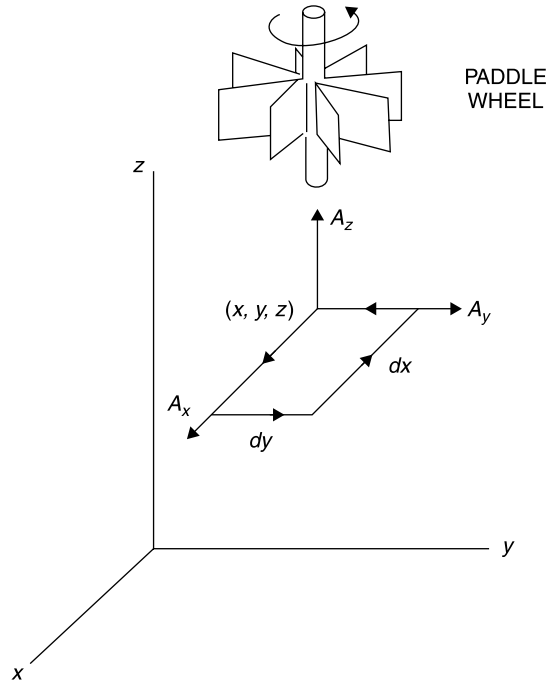

Figure 7.11-1 Determination of the z component of the curl of a vector $\vec{\mathbf{A}}$ [6, p. 96] with paddle wheel model [4, p. 24, reprinted with permissions].

$$\oint \vec{\mathbf{A}} \cdot \overrightarrow{dl} = \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\left(dx\,dy \right. \tag{7.11-4}$$

and applying the definition of the curl gives the z-directed component as

$$(\nabla \times \vec{A})_{\text{z direction}} = \left(\frac{\oint \vec{\mathbf{A}} \cdot \overrightarrow{\mathrm{dl}}}{\mathrm{dx\,dy}}\right)\left(\vec{\mathbf{z}} = \left(\frac{\partial \mathrm{A_y}}{\partial \mathrm{x}} - \frac{\partial \mathrm{A_x}}{\partial \mathrm{y}}\right)\left(\vec{\mathbf{z}}\right. \right. \tag{7.11-5}$$

In like fashion the x- and y-directed components of the curl can be evaluated to produce

$$\nabla \times \vec{\mathbf{A}} = \left(\left(\frac{\partial \mathrm{A_z}}{\partial \mathrm{y}} - \frac{\partial \mathrm{A_y}}{\partial \mathrm{z}}\right)\left(\vec{\mathbf{x}} + \left(\left(\frac{\partial \mathrm{A_x}}{\partial \mathrm{z}} - \frac{\partial \mathrm{A_z}}{\partial \mathrm{x}}\right)\left(\vec{\mathbf{y}} + \left(\left(\frac{\partial \mathrm{A_y}}{\partial \mathrm{x}} - \frac{\partial \mathrm{A_x}}{\partial \mathrm{y}}\right)\left(\vec{\mathbf{z}}\right.\right.\right.\right.\right.\right. \tag{7.11-6}$$

The curl can be written using the shorthand format of the determinant as

$$\nabla \times \vec{\mathbf{A}} = \left|\left(\left(\begin{array}{ccc} \vec{\mathbf{x}} & \vec{\mathbf{y}} & \vec{\mathbf{z}} \\ \dfrac{\partial}{\partial \mathrm{x}} & \dfrac{\partial}{\partial \mathrm{y}} & \dfrac{\partial}{\partial \mathrm{z}} \\ \mathrm{A_x} & \mathrm{A_y} & \mathrm{A_z} \end{array}\right)\right)\right| \tag{7.11-7}$$

Expressions (7.11-6) and (7.11-7) produce the same result. Using the curl the differential form of Ampere's law is written:

$$\nabla \times \vec{\mathbf{H}} = \vec{\mathbf{J}}_C + \frac{\partial \vec{\mathbf{D}}}{\partial t} \tag{7.11-8}$$

where $\vec{\mathbf{J}}_C$ is the *conduction current* density and $\partial\vec{\mathbf{D}}/\partial t$ is the *displacement current density* at the location at which the curl is taken. The displacement current is due to a changing electric field, as, for example, the current that flows between the plates of a capacitor when it is charging or discharging. The displacement current is present with any time-varying $\vec{\mathbf{D}}$ field. When a sinusoidally varying voltage is applied to a capacitor, the changing $\vec{\mathbf{D}}$ field between the capacitor plates produces the AC displacement current density between the plates.

While the definition of the curl has some similarities to the divergence, the two operations are very different in that *taking the divergence of a vector field produces a scalar value* while *taking the curl of a vector field produces another vector field*. Also, the *curl is a differential relationship that applies at a point in space*. The significance of this point relationship can be appreciated from the following example.

Suppose that it is desired to find the $\vec{\mathbf{H}}$ field within the coaxial pair of conductors shown in Figure 7.11-2 [6, p. 94]. Further assume that I is a DC current. Because of the cylindrical symmetry, H_φ is independent of φ; and this allows easy application of Ampere's law. We have two forms for Ampere's law, the line integral relation of (7.10-1) and the point relation of (7.11-8). The line

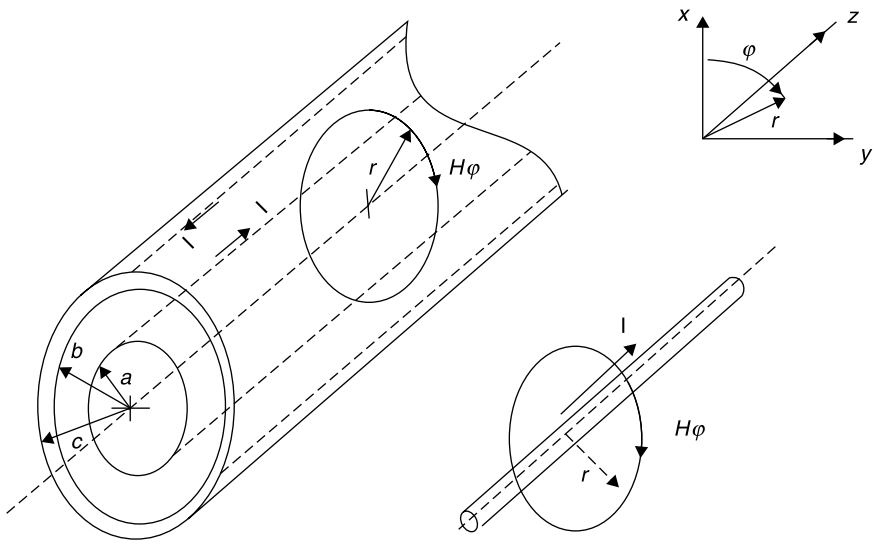

Figure 7.11-2 Magnetic field H_φ about a current-carrying wire, and the result extended to the space within a coaxial transmission line.

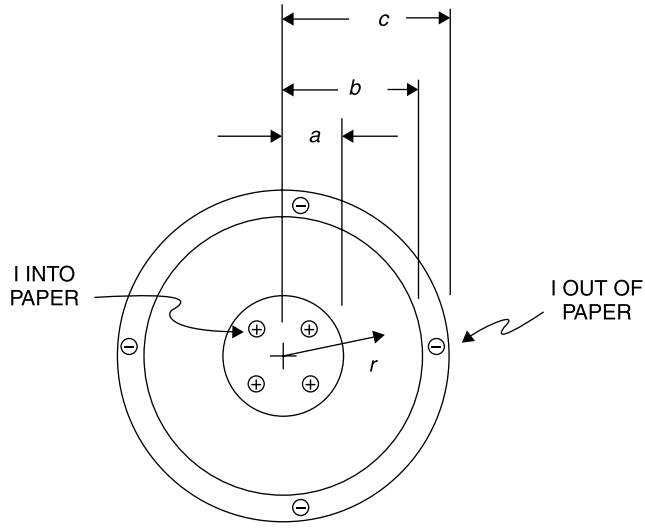

Figure 7.11-3 Cross section of the coaxial line.

integral format requires specification of the *total current within the closed contour* while the point format requires the current density *at the point* at which $\vec{H}$ is to be computed. Suppose the point format is used to calculate H_φ within the center conductor, that is, for $r \leq a$ (Fig. 7.11-3).

Under DC current excitation, I is uniformly distributed over the area of the center conductor, and the current density has only a conduction component given by

$$J_C = \frac{I}{\pi a^2} \tag{7.11-9}$$

Applying (7.11-8),

$$\nabla \times \vec{H} = \frac{\vec{I}}{\pi a^2} \tag{7.11-10}$$

Expressing (7.11-10) in cylindrical coordinates (vector operators in various coordinates are provided in Section 7.14) gives

$$\nabla \times \vec{H} = \vec{r}\left[\frac{1}{r}\frac{\partial H_z}{\partial \varphi} - \frac{\partial H_\varphi}{\partial z}\right] + \vec{\varphi}\left[\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r}\right] + \vec{z}\left[\frac{1}{r}\frac{\partial (rH_\varphi)}{\partial r} - \frac{1}{r}\frac{\partial H_r}{\partial \varphi}\right] = \frac{\vec{I}}{\pi a^2} \tag{7.11-11}$$

In this example, I has only a z component. Thus, equating the z-directed terms in (7.11-14) and noting that, due to symmetry, there can only be a φ component

of $\vec{\mathbf{H}}$ gives

$$\frac{1}{r}\frac{\partial(rH_\varphi)}{\partial r} = \frac{I}{\pi a^2}$$
$$\frac{\partial(rH_\varphi)}{\partial r} = \frac{I}{\pi a^2} r \tag{7.11-12}$$

and integrating gives

$$rH_\varphi = \frac{I}{\pi a^2}\frac{r^2}{2}$$

$$H_\varphi = \frac{I}{2\pi a^2} r \quad \text{for } 0 \le r \le a \tag{7.11-13}$$

To calculate the value of $\vec{\mathbf{H}}$ in the space between the conductors we cannot apply the curl expression because *the current in this region is zero*. This would seem to be a contradiction to what has been studied thus far. Specifically, we know that there is an H_φ between the conductors and this field closes on itself, forming circles, suggesting a finite value of curl. Resolving these two dilemmas is important to understanding the curl and the point form of Ampere's law as expressed by (7.11-8).

First, the value of H_φ was evaluated previously from the line integral form of Ampere's law for an axially directed current in (7.11-2). The result repeated here is

$$H_\varphi = \frac{I}{2\pi r} \quad \text{for } a < r < b \tag{7.11-14}$$

Although there is a finite value for H_φ between the conductors, the value of the curl in this region must be zero, since there is no current density there. To see that the curl is zero, consider that only the fifth term on the left side of (7.11-11) could be nonzero since there is only an H_φ component and it has only an r variation. But this term is

$$\nabla \times \vec{\mathbf{H}} = \vec{\mathbf{r}}\frac{1}{r}\frac{\partial(rH_\varphi)}{\partial r} \tag{7.11-15}$$

and substituting in the value for H_φ gives

$$\frac{1}{r}\frac{\partial(rH_\varphi)}{\partial r} = \frac{1}{r}\frac{\partial(r(I/2\pi r))}{\partial r} = 0 \tag{7.11-16}$$

Thus, even though there is an $\vec{\mathbf{H}}$ field between the conductors, its value cannot be found using the differential form of Ampere's law.

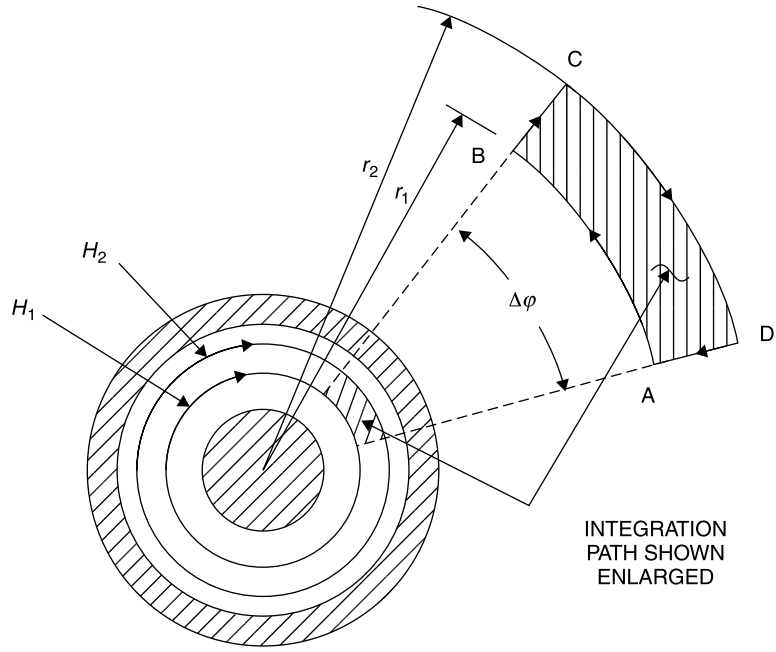

Figure 7.11-4 Closed-path integration to evaluate the curl of H between coaxial conductors.

Next, consider the second *apparent* dilemma: the $\vec{\mathbf{H}}$ field as sketched in Figure 7.11-2 closes on itself and *appears* to have a finite value of curl. Actually, the curl is zero, as was shown by (7.11-16). To arrive at this conclusion in another way, use the line integral method of evaluating the curl about the closed path ABCD shown in Figure 7.11-4. The radial sides of the path, BC and DA produce no contribution to the integral of $\vec{\mathbf{H}} \cdot \mathrm{dl}$ since there is only a circumferential $\vec{\mathbf{H}}$ field component H_φ. The contributions to the circumferential portions of the path, AB and CD, produce

$$\mathrm{H}_\varphi \cdot \mathrm{dl} = \mathrm{H}_1 \mathrm{r}_1 \, \Delta\varphi - \mathrm{H}_2 \mathrm{r}_2 \, \Delta\varphi = \Delta\varphi \left[\frac{\mathrm{I}}{2\pi \mathrm{r}_1} \mathrm{r}_1 - \frac{\mathrm{I}}{2\pi \mathrm{r}_2} \mathrm{r}_2 \right] = 0 \quad (7.11\text{-}17)$$

Thus, it can be seen that the *curvature of the H field and closure on itself does not always imply a finite value of curl*. In this case, the variation of H_φ with r is just sufficient to cancel the curl. Analogously, if a test paddle wheel with rotational axis in the z direction is placed between the conductors, it would not rotate were the $\vec{\mathbf{H}}$ field replaced with a wind current pressure having the same r variation. It might move with the field following a circular orbit about the center conductor, but *it would not rotate about its own axis.*

To compete the analysis of the H_φ in the coaxial line cross section, consider the range $\mathrm{r} > \mathrm{c}$. Applying (7.10-1),

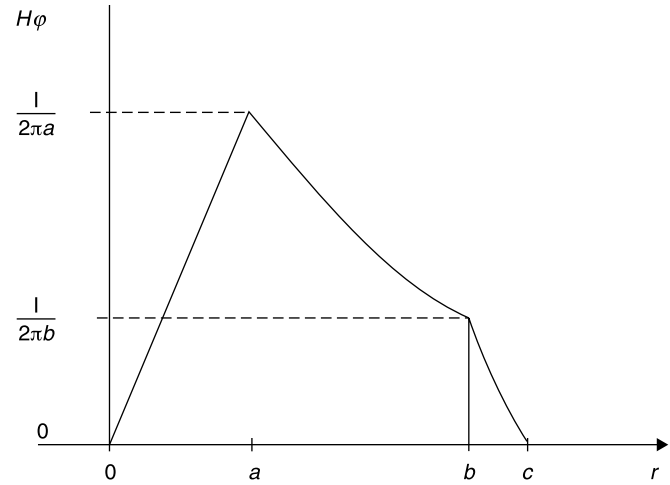

Figure 7.11-5 Magnitude H field of the coaxial conductor set as a function of the radius.

$$\oint \vec{H}_\varphi \cdot \overrightarrow{dl} = I - I = 0 \tag{7.11-18}$$

$$H_\varphi = 0 \quad \text{for } r > c \tag{7.11-19}$$

The field outside the outer conductor is zero because the net current within is zero, a positive current on the center conductor and an equal and opposite current returning on the outer conductor. The field as a function of r is shown in Figure 7.11-5.

7.12 FARADAY'S LAW OF INDUCTION

Faraday's law of induction states that a time-varying magnetic flux passing through a surface produces a voltage around the perimeter of the surface equal to the surface integral of the changing flux through the surface. The polarity of the induced voltage is such as to establish a current to resist the change of magnetic flux:

$$\oint \vec{E} \cdot \overrightarrow{dl} = -\int_S \frac{\partial \vec{B}}{\partial t} \cdot \overrightarrow{dS} \tag{7.12-1}$$

Independently, Lenz duplicated many of Faraday's and Ampere's discoveries. Lenz is credited with the law for determining the polarity of the voltage induced by a changing magnetic field. When Faraday's law is written in differ-

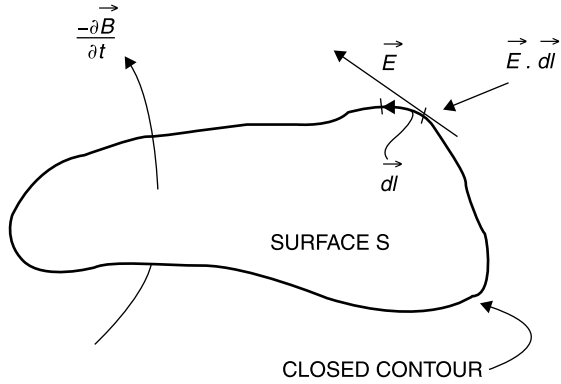

Figure 7.12-1 Diagram for the right-hand sense application of Faraday's law.

ential form using the curl definition, we get the third of Maxwell's equations:

$$\nabla \times \vec{\mathbf{E}} = -\frac{\partial \vec{\mathbf{B}}}{\partial \mathrm{t}} \tag{7.12-2}$$

As defined in (7.12-1) and (7.12-2), the sense of Faraday's law can be remembered by the use of the right-hand rule. If the fingers of the right hand curl in the direction of the path that gives a positive value for the line integral of $\vec{\mathbf{E}} \cdot \vec{\mathrm{dl}}$, then the thumb gives the direction of $-\partial \vec{\mathbf{B}}/\partial \mathrm{t}$, as shown in Figure 7.12-1. It is important to note that in Figure 7.12-1 it is not the direction of the $\vec{\mathbf{B}}$ field that is shown but rather the *direction of the negative rate of change of the* $\vec{\mathbf{B}}$ *field.*

The negative sign in (7.12-2) is the result of *Lenz's law which states that the induced electric field polarity is such as to create a current that opposes a change in the magnetic field.*

7.13 MAXWELL'S EQUATIONS

Maxwell's Four Equations

Understanding propagation of electromagnetic signals begins with understanding physically how electric and magnetic fields interact with one another. The empiric observations just presented that describe the behavior of $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields and their auxiliary $\vec{\mathbf{D}}$ and $\vec{\mathbf{B}}$ fields were interrelated eloquently in a unified theory in about 1882 by James Clerk Maxwell. The theory contained numerous equations and concepts then. However, some time later Oliver Heaviside reduced them to four principal equations together with auxiliary relationships. Maxwell's four equations in both differential and integral formats are summarized as follows:

1. *The divergence of the* $\vec{\mathbf{D}}$ *field equals the volume charge density*. This is Gauss's law:

$$\nabla \cdot \vec{\mathbf{D}} = \rho \tag{7.13-1a}$$

$$\oint_S \vec{\mathbf{D}} \cdot \overrightarrow{dS} = \int_V \rho \, dv \tag{7.13-1b}$$

2. *The divergence of the* $\vec{\mathbf{B}}$ *field is zero:*

$$\nabla \cdot \vec{\mathbf{B}} = 0 \tag{7.13-2a}$$

$$\oint_S \vec{\mathbf{B}} \cdot \overrightarrow{dS} = 0 \tag{7.13-2b}$$

3. *The curl of the* $\vec{\mathbf{E}}$ *field is equal to minus the time rate of change of the magnetic flux density:*

$$\nabla \times \vec{\mathbf{E}} = -\frac{\partial \vec{\mathbf{B}}}{\partial t} \tag{7.13-3a}$$

$$\oint \vec{\mathbf{E}} \cdot \overrightarrow{dl} = -\int_S \frac{\partial \vec{\mathbf{B}}}{\partial t} \cdot \overrightarrow{dS} \tag{7.13-3b}$$

4. *The curl of the* $\vec{\mathbf{H}}$ *field equals the combined conduction and displacement current densities:*

$$\nabla \times \vec{\mathbf{H}} = \vec{\mathbf{J}}_C + \frac{\partial \vec{\mathbf{D}}}{\partial t} \tag{7.13-4a}$$

$$\oint \vec{\mathbf{H}} \cdot \overrightarrow{dl} = \oint_S \overrightarrow{J_C} \cdot \overrightarrow{dS} + \oint_S \frac{\partial \vec{\mathbf{D}}}{\partial t} \cdot \overrightarrow{dS} \tag{7.13-4b}$$

Auxiliary Relations and Definitions

To Maxwell's famous four equations must be added definitions and auxiliary relations to complete the experimental evidence unified by Maxwell. These are:

The Force Law This applies to a charge q moving at velocity $\vec{\mathbf{v}}$ in a region having both electric field $\vec{\mathbf{E}}$ and magnetic flux density $\vec{\mathbf{B}}$:

$$\vec{\mathbf{f}} = q[\vec{\mathbf{E}} + \vec{\mathbf{v}} \times \vec{\mathbf{B}}] \tag{7.13-5}$$

Ohm's Law and the Definition of Conduction Current Density The point form of *Ohm's law* is

$$\vec{\mathbf{J}}_C = \sigma \vec{\mathbf{E}} \tag{7.13-6}$$

where $\vec{\mathbf{J}}_C$ is the current density (in amperes/square meter), σ is the conductivity of the material (in mhos/meter), and E is the electric field (in volts/meter).

Convection Current Density The definition of *convection current density* is

$$\vec{\mathbf{J}}_{\mathrm{CV}} \equiv \rho\vec{\mathbf{v}}_\rho \tag{7.13-7}$$

where $\vec{\mathbf{v}}_\rho$ is the velocity at which the volume charge density ρ is moving [6, Sec. 4.06].

Permittivity (Dielectric Constant) The definition of *permittivity* (*dielectric constant*) is defined by

$$\vec{\mathbf{D}} \equiv \varepsilon\vec{\mathbf{E}} = \varepsilon_0\varepsilon_R\vec{\mathbf{E}} \tag{7.13-8}$$

where $\varepsilon_0 = 8.854186 \times 10^{-12}$ ($\approx 10^{-9}/36\pi$) farads per meter is the permittivity of free space and ε_R is the *relative dielectric constant* used to account for the effects of atomic and molecular dipoles of various materials.

Permeability The definition of *permeability* is defined within

$$\vec{\mathbf{B}} = \mu\vec{\mathbf{H}} = \mu_0\mu_R\vec{\mathbf{H}} \tag{7.13-9}$$

where $\mu_0 = 4\pi \times 10^{-7}$ H/m and μ_R is the *relative permeability* used to account for the magnetic dipole moments of atoms within various materials.

In general μ and ε can be *tensor quantities*, whose values vary with position and direction. For example, a specialty of microwave engineering is the design of artificial (anisotropic) dielectrics [9, Chapter 12], in which metal strips are embedded in an otherwise homogeneous dielectric to produce a composite dielectric having very different relative dielectric constants for different applied electric field directions. Similarly, microwave ferrite materials [10, Sec. 6.6] exhibit quite different relative permeabilities with direction when immersed within a strong biasing magnetic field. However, for many engineering purposes μ and ε can be treated as simple scalar constants.

Visualizing Maxwell's Equations

To review, consider the first of Maxwell's equations:

$$\nabla \cdot \vec{\mathbf{D}} = \rho \quad \text{or} \quad \nabla \cdot \varepsilon\vec{\mathbf{E}} = \rho \tag{7.13-10}$$

This equation states that the "divergence" of the *flux density* $\vec{\mathbf{D}}$ is equal to the charge density in the volume from which the $\vec{\mathbf{D}}$ field is emanating. When the $\vec{\mathbf{D}}$ field lines exhibit a net departure (*divergence*) from any volume (even a microscopic one), there is a net positive electric charge inside. Vice versa, if the lines

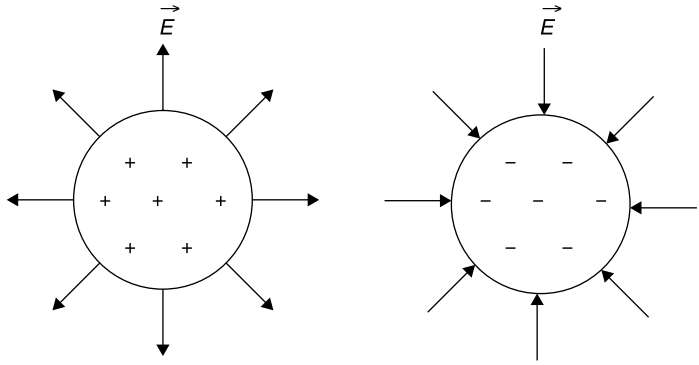

Figure 7.13-1 Divergence of the $\vec{E}$ field is proportional to the contained charge density, and its direction is related to the sign of the net charge, as shown.

enter the volume, there is a net negative charge inside. Departing lines mean that a *positive charge* outside the volume will experience a force that repels it from the volume. That makes sense since *like charges repel*. The $\vec{E}$ field is proportional to the $\vec{D}$ field and the $\vec{E}$ field emerges or enters a volume in proportion to the net positive or negative charge within, respectively, as shown in Figure 7.13-1.

Next consider Maxwell's second equation:

$$\nabla \cdot \vec{B} = 0 \tag{7.13-11}$$

This equation simply indicates that *the* B *field has no divergence* (Fig. 7.13-2). This means that the $\vec{B}$ field lines are continuous throughout space, they do not originate anywhere, and they always close on themselves. For homogeneous regions, the $\vec{H}$ field lines also close on themselves, being proportional to $\vec{B}$. That is, $\vec{B} = \mu\vec{H}$.

Because the $\vec{B}$ field is continuous, even through a boundary dividing regions of different permeability, it follows that the $\vec{H}$ field cannot be continuous through the permeability boundary. Its field strength must reflect the ratio of permeabilities in the two regions, as shown in Figure 7.13-3:

$$\vec{H}_2 = \frac{\mu_1}{\mu_2}\vec{H}_1 \tag{7.13-12}$$

The third of Maxwell's equations introduces the curl concept, the amount of circulation of a vector field:

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{7.13-13}$$

This says that a changing $\vec{B}$ field induces an $\vec{E}$ field at right angles. The $\vec{E}$ field closes on itself in a closed loop or curl. This is the incremental form of Far-

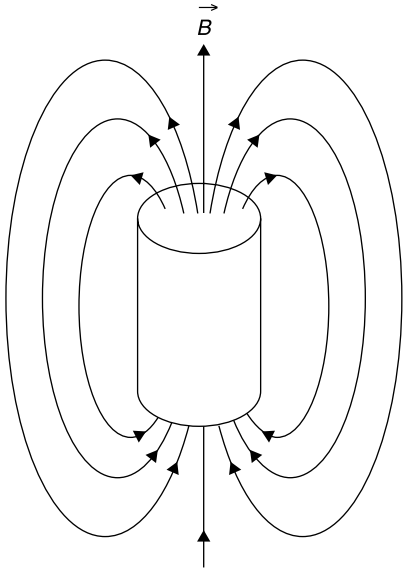

Figure 7.13-2 Bar magnet, illustrating that since the divergence of the $\vec{\mathbf{B}}$ field is zero, the field lines, when sufficiently extended, always close on themselves.

aday's law, which says that the voltage induced in a loop is proportional to the time rate of change of the magnetic field within the loop and (Lenz's law) of such a polarity as to oppose the change in the $\vec{\mathbf{B}}$ field.

The applications of this phenomenon are widespread, including transformers for power distribution and the coils used to generate high voltages for the spark plugs of gasoline engines. We shall see that this and the next equation are responsible for the propagation of radio waves, the essence of wireless signal propagation.

The fourth, and last, of Maxwell's equations, states that the curl of the $\vec{\mathbf{H}}$ field is proportional to the enclosed current:

$$\nabla \times \vec{\mathbf{H}} = \vec{\mathbf{J}}_C + \frac{\partial \vec{\mathbf{D}}}{\partial \mathrm{t}} \tag{7.13-14}$$

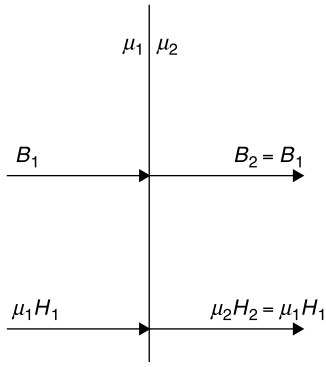

Figure 7.13-3 While the $\vec{\mathbf{B}}$ lines have zero divergence (are always continuous), the $\vec{\mathbf{H}}$ lines change across a boundary to account for a change in permeability from one region to the next.

which has two components: (1) the conduction current density and (2) the time rate of change of the electric flux density $\partial\vec{\mathbf{D}}/\partial t$ (the displacement current density).

The two current forms in this equation have many important practical applications. For example, a conduction current in a vertical wire (such as a monopole antenna) provides a means of exciting a time-varying $\vec{\mathbf{H}}$ field in the space about the antenna. In turn, the time-varying $\vec{\mathbf{H}}$ field produces a time-varying $\vec{\mathbf{E}}$ field perpendicular to the $\vec{\mathbf{H}}$ field (from Maxwell's third equation). We shall see that this combined action generates a *propagating electromagnetic waveform* in which each field is induced by the time derivative of the other. Once one field is created the other is induced by the time rate of change of the first, and the resulting waveforms propagate away from the source as a *radio wave*.

7.14 PRIMARY VECTOR OPERATIONS

In the course of describing the behavior of electromagnetic fields, we have found it useful to define and apply basic vector operations such as the vector dot product, cross product, gradient, divergence, and curl. These vary in form. Some are performed on vectors to yield new vectors, some are performed on vectors to yield scalar quantities, and some are performed on scalar quantities to yield vectors. The fundamental vector operations that we have introduced are summarized in Table 7.14-1.

TABLE 7.14-1 Basic Vector Operators in Rectangular Coordinates (1a,b,c,d,e)[a]

Name	Symbol	Operates on	Yields	Formula
Scalar product (also dot product)	$\vec{\mathbf{A}}\cdot\vec{\mathbf{B}}$	2 vectors	Scalar function	$\vec{\mathbf{A}}\cdot\vec{\mathbf{B}} = A_x B_x + A_y B_y + A_z B_z$
Vector product (also cross product)	$\vec{\mathbf{A}}\times\vec{\mathbf{B}}$	2 vectors	1 vector	$\vec{\mathbf{A}}\times\vec{\mathbf{B}} = \begin{vmatrix} \vec{x} & \vec{y} & \vec{z} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}$
Divergence	$\nabla\cdot\vec{\mathbf{A}}$	1 vector	Scalar function	$\nabla\cdot\vec{\mathbf{A}} = \frac{\partial}{\partial x}A_x + \frac{\partial}{\partial y}A_y + \frac{\partial}{\partial z}A_z$
Curl	$\nabla\times\vec{\mathbf{A}}$	1 vector	1 vector	$\nabla\times\vec{\mathbf{A}} = \begin{vmatrix} \vec{x} & \vec{y} & \vec{z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix}$
Gradient	$\nabla\Phi$	Scalar function	1 vector	$\nabla\Phi = \left(\vec{x}\frac{\partial\Phi}{\partial x} + \vec{y}\frac{\partial\Phi}{\partial y} + \vec{z}\frac{\partial\Phi}{\partial z}\right)$

[a]Where $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ are vectors $\vec{\mathbf{A}} = A_x\vec{x} + A_y\vec{y} + A_z\vec{z}$ and $\vec{\mathbf{B}} = B_x\vec{x} + B_y\vec{y} + B_z\vec{z}$, and $\Phi(x, y, z)$ is a scalar function.

The use of vector operations is simplified by using coordinate systems that exploit the symmetry of the application. In addition to rectangular coordinates, cylindrical and spherical coordinates are often useful. Rather than convert the application to rectangular coordinates, it is more convenient to carry out the vector operations directly in the coordinate system of choice. The vector operations are listed below in cylindrical and spherical coordinates.

Cylindrical Coordinates

$$\nabla\Phi = \vec{\mathrm{r}}\frac{\partial\Phi}{\partial \mathrm{r}} + \vec{\varphi}\frac{1}{\mathrm{r}}\frac{\partial\Phi}{\partial\varphi} + \vec{\mathrm{z}}\frac{\partial\Phi}{\partial \mathrm{z}} \tag{7.14-1a}$$

$$\nabla\cdot\vec{\mathbf{A}} = \frac{1}{\mathrm{r}}\frac{\partial(\mathrm{rA_r})}{\partial\vec{\mathrm{r}}} + \frac{1}{\mathrm{r}}\frac{\partial \mathrm{A}_\varphi}{\partial\varphi} + \frac{\partial \mathrm{A_z}}{\partial \mathrm{z}} \tag{7.14-1b}$$

$$\nabla\times\vec{\mathbf{A}} = \vec{\mathrm{r}}\left[\frac{1}{\mathrm{r}}\frac{\partial \mathrm{A_z}}{\partial\varphi} - \frac{\partial \mathrm{A}_\varphi}{\partial \mathrm{z}}\right] + \vec{\varphi}\left[\frac{\partial \mathrm{A_r}}{\partial \mathrm{z}} - \frac{\partial \mathrm{A_z}}{\partial \mathrm{r}}\right] + \vec{\mathrm{z}}\left[\frac{1}{\mathrm{r}}\frac{\partial(\mathrm{rA}_\varphi)}{\partial \mathrm{r}} - \frac{1}{\mathrm{r}}\frac{\partial \mathrm{A_r}}{\partial\varphi}\right] \tag{7.14-1c}$$

Spherical Coordinates

$$\nabla\Phi = \vec{\mathrm{r}}\frac{\partial\Phi}{\partial \mathrm{r}} + \vec{\theta}\frac{1}{\mathrm{r}}\frac{\partial\Phi}{\partial\theta} + \vec{\varphi}\frac{1}{\mathrm{r}\sin\theta}\frac{\partial\Phi}{\partial\theta} \tag{7.14-2a}$$

$$\nabla\cdot\mathrm{A} = \frac{1}{\mathrm{r}^2}\frac{\partial}{\partial \mathrm{r}}(\mathrm{r}^2\mathrm{A_r}) + \frac{1}{\mathrm{r}\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\mathrm{A}_\theta) + \frac{1}{\mathrm{r}\sin\theta}\frac{\partial \mathrm{A}_\varphi}{\partial\varphi} \tag{7.14-2b}$$

$$\nabla\times\vec{\mathbf{A}} = \vec{\mathrm{r}}\frac{1}{\mathrm{r}\sin\theta}\left[\frac{\partial}{\partial\theta}(\mathrm{A}_\varphi\sin\theta) - \frac{\partial \mathrm{A}_\theta}{\partial\varphi}\right] + \vec{\theta}\frac{1}{\mathrm{r}}\left[\frac{1}{\sin\theta}\frac{\partial \mathrm{A_r}}{\partial\varphi} - \frac{\partial}{\partial \mathrm{r}}(\mathrm{rA}_\varphi)\right] + \vec{\varphi}\frac{1}{\mathrm{r}}\left[\frac{\partial}{\partial \mathrm{r}}(\mathrm{rA}_\theta) - \frac{\partial \mathrm{A_r}}{\partial\theta}\right] \tag{7.14-2c}$$

7.15 THE LAPLACIAN

After applying an operation from Table 7.14-1, it is often useful to apply the same or another operation to the result. One of the most important is the *Laplacian*. It occurs so often that it is considered a separate operator. This operator has two definitions, one for operation on a scalar function and one for operation on a vector field. *When applied to a scalar function* $\Phi(x, y, z)$*, the Laplacian is equal to the divergence of the gradient*. Its value is the sum of the second partial derivatives with respect to x, y, and z of Φ. This can be evaluated by first evaluating the gradient of Φ and then finding the divergence. The result [4, p. 30] is:

Rectangular Coordinates

$$\nabla^2\Phi = \nabla\cdot\nabla\Phi = \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} \tag{7.15-1a}$$

Cylindrical Coordinates

$$\nabla^2\Phi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\left(\frac{\partial\Phi}{\partial r}\right)\right) + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\varphi^2} + \frac{\partial\Phi^2}{\partial z^2} \tag{7.15-1b}$$

Spherical Coordinates

$$\nabla^2\Phi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\left(\frac{\partial\Phi}{\partial r}\right)\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\left(\frac{\partial\Phi}{\partial\theta}\right)\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Phi}{\partial\varphi^2} \tag{7.15-1c}$$

Recall that the electric field $\vec{E}$ is the negative of the gradient of the potential function ($\vec{E} = -\nabla\Phi$) and that the divergence of the static electric field (in the absence of changing $\vec{B}$ field) is equal to the charge density divided by the dielectric constant ($\nabla \cdot \vec{E} = \rho/\varepsilon$), thus

$$\nabla^2\Phi = -\frac{\rho}{\varepsilon} \tag{7.15-2}$$

This relationship is called *Poisson's equation*. It applies to a region of space containing a charge density, ρ. In a region that is free of charge:

$$\nabla^2\Phi = 0 \tag{7.15-3}$$

This relationship is called Laplace's equation. One might ask: If there is no charge, how can there be an electrostatic potential, and accordingly, a need for Laplace's equation? Consider the diagram in Figure 7.15-1.

In region 1 the contained charge results in a finite value of divergence of the $\vec{E}$ field from the region, and Poisson's equation applies. In region 2 there is no

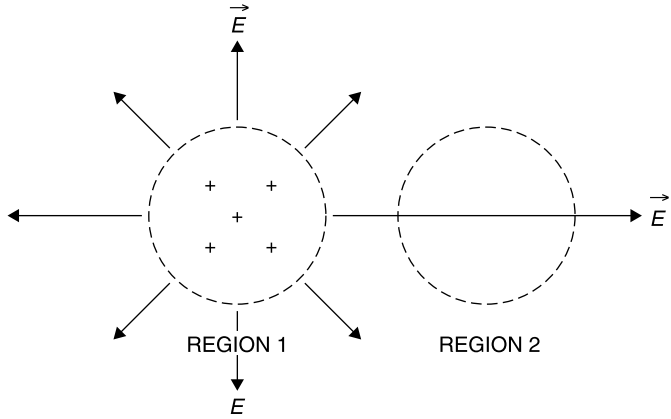

Figure 7.15-1 Poisson's equation applies in a region containing a charge density (region 1), while Laplace's equation is defined in a region without charge (region 2).

charge; the $\vec{E}$ field divergence is zero. However, there is an $\vec{E}$ field due to the charge in the adjacent region 1, and Poisson's equation for region 2 therefore reduces to that of Laplace.

When the Laplacian is applied to a vector field, the result is a vector [4, p. 30]:

$$\nabla^2\vec{A} = \nabla^2(\vec{x}A_x + \vec{y}A_y + \vec{z}A_z) \tag{7.15-4}$$

in which the ∇^2 operator is defined in rectangular coordinates to be

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{7.15-5}$$

Care must be taken that *the entire* ∇^2 *operator is applied fully to each component of* $\vec{A}$. Explicitly,

Rectangular Coordinates

$$\nabla^2\vec{A} = \vec{x}\left(\frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_x}{\partial y^2} + \frac{\partial^2 A_x}{\partial z^2}\right) + \vec{y}\left(\frac{\partial^2 A_y}{\partial x^2} + \frac{\partial^2 A_y}{\partial y^2} + \frac{\partial^2 A_y}{\partial z^2}\right) + \vec{z}\left(\frac{\partial^2 A_z}{\partial x^2} + \frac{\partial^2 A_z}{\partial y^2} + \frac{\partial^2 A_z}{\partial z^2}\right) \tag{7.15-6}$$

Cylindrical Coordinates

$$\nabla^2\vec{A} = \vec{r}\left[\nabla^2 A_r - \frac{2}{r^2}\frac{\partial A_\varphi}{\partial \varphi} - \frac{A_r}{r^2}\right] + \vec{\varphi}\left[\nabla^2 A_\varphi + \frac{2}{r^2}\frac{\partial A_r}{\partial \varphi} - \frac{A_\varphi}{r^2}\right] + \vec{z}[\nabla^2 A_z] \tag{7.15-7}$$

where

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial z^2}$$

Spherical Coordinates

$$\begin{aligned}\nabla^2\vec{A} = \vec{r}&\left[\nabla^2 A_r - \frac{2}{r^2}\left(A_r + \cot\theta A_\theta + \csc\theta\frac{\partial A_\varphi}{\partial \varphi} + \frac{\partial A_\theta}{\partial \theta}\right)\right] \\ &+ \vec{\theta}\left[\nabla^2 A_\theta - \frac{1}{r^2}\left(\csc^2\theta A_\theta - 2\frac{\partial A_r}{\partial \theta} + 2\cot\theta\csc\theta\frac{\partial A_\varphi}{\partial \varphi}\right)\right] \\ &+ \vec{\varphi}\left[\nabla^2 A_\varphi - \frac{1}{r^2}\left(\csc^2\theta A_\varphi - 2\csc\theta\frac{\partial A_r}{\partial \varphi} - 2\cot\theta\csc\theta\frac{\partial A_\theta}{\partial \varphi}\right)\right]\end{aligned} \tag{7.15-8}$$

where

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}$$

7.16 VECTOR AND SCALAR IDENTITIES

The Laplacian operators of the previous section result from the successive application of operators. The scalar and vector forms of the Laplacian are considered operators, themselves. In addition to the Laplacian, there are numerous other such multiple operations that can be defined and found to be useful. When applied in general form, they become *operational identities*. Many of them may appear to have no obvious physical significance, but these identities nevertheless prove useful in deriving certain electromagnetic results, just as the trigonometric identity $\sin^2\theta + \cos^2\theta = 1$ finds numerous applications.

An example of a multiple operation that does have clear physical significance is obtained in first taking the gradient of a scalar function and then taking the curl of the result. Previously, we noted that the value of a scalar function $\Phi(x, y, z)$ is dependent only on the coordinates (x, y, z) and not on how that point is reached. It follows that *the gradient always describes a conservative vector function*, that is to say, the line integral of the gradient about any closed path is zero. Since the gradient is a vector function, we can operate on it to find its curl. But since the curl is the line integral of a vector about a closed path, the curl of the gradient will always be zero. Put simply, *"a field that is the gradient of something has no curl"* [4, p. 29]:

$$\text{Curl of gradient of } \Phi = \nabla \times \text{gradient } \Phi = \nabla \times \nabla\Phi = 0 \qquad (7.16\text{-}1)$$

for any scalar function Φ from which the gradient is derived. This identity has the physical significance of describing how closed-path integrals of conservative functions are zero. The validity of (7.16-1) can be verified by evaluating the gradient of a function and then finding its curl, which leads to equal and opposite partial derivatives. This equality will prove useful in the derivation of the *wave equation*.

A second equality obtained by successive vector operations is that *"a field that is the curl of something has no divergence"* [4, p. 29]:

$$\text{Divergence of curl} = \nabla \cdot \nabla \times \vec{\mathbf{A}} = 0 \qquad (7.16\text{-}2)$$

where $\vec{\mathbf{A}}$ is any vector field $\vec{\mathbf{A}} = A_x\vec{\mathbf{x}} + A_y\vec{\mathbf{y}} + A_z\vec{\mathbf{z}}$. The basic vector operations previously defined were independent of the coordinate system employed. For example, the divergence was defined as the limit of the integral of a vector emerging orthogonal to the surface of a volume as the volume contained by the surface was shrunk to zero. In the previous section, the vector Laplacian was

TABLE 7.16-1 Vector Identities

$\nabla(\Phi + \Psi) = \nabla\Phi + \nabla\Psi$	(7.16-4)
$\nabla \cdot (\vec{A} + \vec{B}) = \nabla \cdot \vec{A} + \nabla \cdot \vec{B}$	(7.16-5)
$\nabla \times (\vec{A} + \vec{B}) = \nabla \times \vec{A} + \nabla \times \vec{B}$	(7.16-6)
$\nabla(\Phi\Psi) = \Phi\nabla\Psi + \Psi\nabla\Phi$	(7.16-7)
$\nabla \cdot (\Psi\vec{A}) = \vec{A} \cdot \nabla\Psi + \Psi\nabla \cdot \vec{A}$	(7.16-8)
$\nabla \cdot (\vec{A} \times \vec{B}) = \vec{B} \cdot \nabla \times \vec{A} - \vec{A} \cdot \nabla \times \vec{B}$	(7.16-9)
$\nabla \times (\Phi\vec{A}) = \nabla\Phi \times \vec{A} + \Phi\nabla \times \vec{A}$	(7.16-10)
$\nabla \cdot \nabla\Phi = \nabla^2\Phi$	(7.16-11)
$\nabla \cdot \nabla \times \vec{A} = 0$	(7.16-12)
$\nabla \times \nabla\Phi = 0$	(7.16-13)
$\nabla \times \nabla \times \vec{A} = \nabla(\nabla \cdot \vec{A}) - \nabla^2\vec{A}$	(7.16-14)
$\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B})$	(7.16-15)

Source: From Ramo and Whinnery [6, p. 114]; reprinted with permission.

defined in three coordinate systems: rectangular, cylindrical, and spherical. Mathematicians do not like to have a fundamental definition dependent upon a coordinate system. To circumvent this impasse, an alternate definition for the vector Laplacian that can be applied to any coordinate system is [6, p. 111, and 4, p. 30]:

$$\nabla^2\vec{A} \equiv \nabla(\nabla \cdot A) - \nabla \times \nabla \times A \tag{7.16-3}$$

This is an identity. Even so, it would appear to be totally esoteric. Remarkably, however, it is employed in several important proofs, as will be seen in the succeeding sections. Its validity can be demonstrated by showing that the left and right sides of (7.16-3) both give the same result when expressed in rectangular coordinates. Since both the divergence and the curl of the curl are both expressible in any orthogonal coordinate system, *the vector Laplacian can be considered to be defined by (7.16-3)* making the definition independent of the coordinate system.

Several vector identities resulting from successive vector operations are listed in Table 7.16-1. They can be verified by performing the indicated operations that their equations represent.

7.17 FREE CHARGE WITHIN A CONDUCTOR

The derivation of the charge distribution in a conductor is a practical example of the application of Maxwell's equations and Ohm's law [6, Sec. 6.03]. Suppose that a quantity of free charge having a volume density ρ is somehow introduced into a conductor. We solve for its distribution by beginning with Ohm's law for the current density:

$$\vec{\mathbf{J}} = \sigma\vec{\mathbf{E}} \tag{7.17-1}$$

where σ is the conductivity of the conductor, and the regular type font indicates that (7.17-1) applies to all time variations. When this value of $\vec{\mathbf{J}}$ is substituted into Maxwell's fourth equation the result is

$$\nabla \times \vec{\mathbf{H}} = \sigma\vec{\mathbf{E}} + \frac{\partial\vec{\mathbf{D}}}{\partial t} \tag{7.17-2}$$

Now take the divergence of both sides of (7.17-2) and note that the divergence of the curl of any vector field is zero:

$$0 = \nabla \cdot \nabla \times \vec{\mathbf{H}} = \frac{\sigma}{\varepsilon}\nabla \cdot \vec{\mathbf{D}} + \frac{\partial(\nabla \cdot \vec{\mathbf{D}})}{\partial t} \tag{7.17-3}$$

From Maxwell's first equation $\nabla \cdot \vec{\mathbf{D}} = \rho$, where ρ in this case is the free charge density, allowing (7.17-3) to be written as

$$\frac{\sigma}{\varepsilon}\rho + \frac{\partial\rho}{\partial t} = 0 \tag{7.17-4}$$

The function that is unchanged by differentiation is the exponential, and therefore the solution to (7.17-4) is

$$\rho(t) = \rho_0 e^{-(\sigma/\varepsilon)t} \tag{7.17-5}$$

This means that any free charge density within a conductor decays exponentially in time with a time constant σ/ε. Since we have not provided for a means by which the charge can be annihilated, it follows that it must flow to the surface of the conductor. The time constant for copper, assuming the dielectric constant of free space within copper, would be

$$\frac{\sigma}{\varepsilon} = \frac{5.8 \times 10^7\ \text{℧/m}}{8.6 \times 10^{-12}\ \text{F/m}} = 7\ \mu\text{s} \tag{7.17-6}$$

It is not practical to measure the relative dielectric constant of materials having high conductivity, such as copper. However, to the extent that the dielectric constant of copper exceeds that of free space, the time constant would be even shorter. Thus, we can conclude that free charge does not linger within a conductor. This makes sense physically. If there were a cloud of charge within a conductor, all of the individual charges would repel one another. Since the conductivity is high, they would move as far apart and as rapidly as possible, taking up positions on the surface when they could go no further. On the surface we could expect that they would distribute themselves in a manner that maximizes their average separation distances from one another.

Another way of estimating the quality of a conductor for sinusoidal currents is to write Maxwell's fourth equation in its phasor form

$$\nabla \times \vec{H} = \vec{J}_C + \frac{\partial \vec{D}}{\partial t} = \vec{E}(\sigma + j\omega\varepsilon) \tag{7.17-7}$$

In a good conductor the displacement current is negligible compared to the conduction current, and this requires that

$$\sigma \gg \omega\varepsilon \tag{7.17-8}$$

Again, since it is impractical to measure the dielectric constant of conductors, let us suppose that for most conductors ε_R is no greater than 10. Further suppose that we wish (7.17-8) to be satisfied by a factor of 100 before we will consider the material a "good conductor." Then a conductor is "good" up to a maximum frequency, $f_{\max}$, given by

$$f_{\max} \leq \frac{\sigma}{2\pi(100)(10)\varepsilon_0} \tag{7.17-9}$$

Applying this criterion to copper, having $\sigma = 5.8 \times 10^7$ $\mho$/m gives

$$f_{\max} \leq \frac{5.8 \times 10^7}{2\pi(10^3)8.854 \times 10^{-12}} = 1 \times 10^{15} \text{ Hz}$$

$$\leq 1000 \text{ GHz} \quad \text{(for copper)} \tag{7.17-10}$$

Conductors satisfying (7.17-9) are called good conductors because for a given applied electric field or voltage the conduction current is much greater than the displacement current.

7.18 SKIN EFFECT

The phenomenon of *skin effect*, by which the current is crowded toward the surface of conductors at high frequencies, was introduced in Section 2.13. Its mathematical derivation [6, Sec. 6.04] provides another example of the application of Maxwell's equations, and the resultant conclusion is an insight into the current distribution *within a conductor*. Having shown that the displacement current is negligible compared to the conduction current in a good conductor, Maxwell's fourth equation can be written for the region within a conductor as

$$\nabla \times \vec{\mathbf{H}} = \sigma\vec{\mathbf{E}} \tag{7.18-1}$$

Taking the curl of both sides, and applying the identity (7.16-3),

$$\nabla \times \nabla \times \vec{\mathbf{H}} = \nabla(\nabla \cdot \vec{\mathbf{H}}) - \nabla^2 \vec{\mathbf{H}} = \sigma(\nabla \times \vec{\mathbf{E}}) \tag{7.18-2}$$

Recognizing that the divergence of $\vec{\mathbf{H}}$ is zero and that the curl of $\vec{\mathbf{E}}$ is related to $\vec{\mathbf{H}}$ through Maxwell's third equation gives

$$\nabla^2 \vec{\mathbf{H}} = \sigma\mu \frac{\partial \vec{\mathbf{H}}}{\partial t} \tag{7.18-3}$$

That this same relationship applies to the $\vec{\mathbf{E}}$ field can be shown by beginning with Maxwell's third equation, taking the curl of both sides and noting that the displacement current is negligible to get

$$\nabla \times \nabla \times \vec{\mathbf{E}} = \nabla(\nabla \cdot \vec{\mathbf{E}}) - \nabla^2 \vec{\mathbf{E}} = -\mu \frac{\partial}{\partial t}(\nabla \times \vec{\mathbf{H}}) = -\mu\sigma \frac{\partial \vec{\mathbf{E}}}{\partial t} \tag{7.18-4}$$

We showed in the previous section that free charge does not remain in a good conductor; hence the divergence of $\vec{\mathbf{E}}$ is zero, and then (7.18-4) can be written as

$$\nabla^2 \vec{\mathbf{E}} = \mu\sigma \frac{\partial \vec{\mathbf{E}}}{\partial t} \tag{7.18-5}$$

Since $\vec{\mathbf{J}} = \sigma\vec{\mathbf{E}}$, it follows that

$$\nabla^2 \vec{\mathbf{J}} = \mu\sigma \frac{\partial \vec{\mathbf{J}}}{\partial t} \tag{7.18-6}$$

Equations (7.18-1) to (7.18-6) apply for all time variations; however, we are usually interested in the sinusoidal case. Then (7.18-3), (7.18-5), and (7.18-6) can be rewritten in phasor form, respectively, as

$$\nabla^2 \vec{H} = j\omega\sigma\mu \vec{H} \tag{7.18-7}$$

$$\nabla^2 \vec{E} = j\omega\sigma\mu \vec{E} \tag{7.18-8}$$

$$\nabla^2 \vec{J} = j\omega\sigma\mu \vec{J} \tag{7.18-9}$$

We find the solution to this equation format by solving for the current distribution in an infinitely thick conductor in the $x > 0$ direction, although it will be seen shortly that all of the interesting distribution of current occurs in a very thin layer near the surface of this conductor. Referring to Figure 7.18-1, suppose that there is a current sheet directed in the $+z$ direction and extending

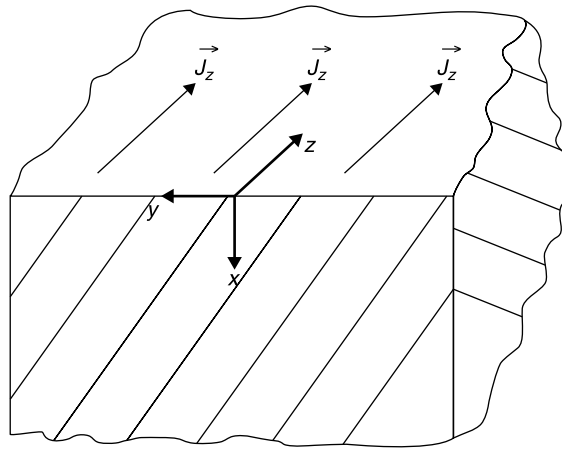

Figure 7.18-1 Infinitely thick conductor in the $x > 0$ direction having a uniform current sheet $\vec{J_z}$ in the yz plane moving in the $+z$ direction. (After Ramo and Whinnery, 6, p. 237, with permission.)

uniformly in the $+y$ and $-y$ directions. In other words, there is a uniform, infinite sheet of current in the yz plane that is moving in the $+z$ direction.

Since the current has no y or z variations by assumption, its Laplacian is simply

$$\frac{d^2 J_z}{dx^2} = j\omega\mu\sigma J_z = K^2 J_z \tag{7.18-10}$$

where J_z is the magnitude of $\vec{J}$ in the z direction and

$$K = \sqrt{j}\sqrt{\omega\mu\sigma} = \sqrt{j}\sqrt{2\pi f\mu\sigma} \tag{7.18-11}$$

Note that

$$\sqrt{j} = (1\angle 90^\circ)^{1/2} = 1\angle 45^\circ = \frac{1+j}{\sqrt{2}} \tag{7.18-12}$$

and therefore

$$K = (1+j)\sqrt{\pi f\mu\sigma} \quad (\text{meters}^{-1}) \tag{7.18-13}$$

Since the constant K has the dimensions of reciprocal meters, it is preferable to cast it in reciprocal form, resulting in

$$K = \frac{1+j}{\delta_S} \quad (\text{meters}^{-1}) \tag{7.18-14}$$

where δ_S is defined as the *skin depth* and given by

$$\delta_S = \frac{1}{\sqrt{\pi f \mu \sigma}} \text{ (meters)} \tag{7.18-15}$$

In general, the differential equation (7.18-10) has two independent exponential solutions, namely

$$\vec{J}_z = Ae^{-Kx} + Be^{+Kx} \tag{7.18-16}$$

However, the positive exponential term cannot apply practically, else current density would increase to infinity with $+x$. Therefore B must equal zero for this situation. Also, A can be set equal to the current density, J_0, at the surface of the conductor at which $x = 0$, then

$$J_z = J_0 e^{-x/\delta_S} e^{-j(x/\delta_S)} \tag{7.18-17}$$

This result indicates that the magnitude of the current density falls off exponentially with distance from the surface, reaching $1/e$ of its surface value when $x = \delta_S$. For good conductors δ_S is an extremely short distance at RF and microwave frequencies. For example, when evaluated for copper at 1 GHz (2.13-4) δ_S is only 2.1 μm. At 25 μm/mil, this is less than 0.1 mil.

The term *skin depth* can be misleading, implying that all current flows within the skin depth. This, of course, is not true. In a conductor with finite conductivity, current penetrates to arbitrary depths. However, practically speaking, in a depth of only three times the skin depth, the current density is only 5% of its surface value.

7.19 CONDUCTOR INTERNAL IMPEDANCE

From the expression of 7.18-17 it can be seen that not only is there a fall off in current density from the conductor surface, there is also a lagging phase in the current relative to the surface value. Since the electric field and current are in phase at the surface, this means that the current density below the surface has an inductive reactance effect. It also means that the total current flow lags the applied electric surface field on the conductor, and that accordingly the conductor has an effective *internal impedance*. We can examine this effect by evaluating the total complex current J_z as follows [6, Sec. 6.06]. The conductor's internal impedance for a unit length and unit width is equal to the ratio of the voltage at the surface divided by the total current per unit width, I_W:

$$I_W = \int_0^\infty J_z\,dx = \int_0^\infty J_0 e^{-(1+j)(x/\delta_S)}\,dx = -J_0 \frac{\delta}{1+j} e^{-(1+j)(x/\delta_S)} \Big|_0^\infty$$

$$= J_0 \frac{\delta_S}{1+j} \tag{7.19-1}$$

The electric field at the surface is related to the current density by

$$E_z(x=0) = \frac{J_0}{\sigma} \tag{7.19-2}$$

and for a unit length of the conductor the voltage V_z has the same numeric value. The internal impedance of the conductor is then

$$Z_S = \frac{V_z(x=0)}{I_W} = \frac{1}{\sigma\delta_S} + j\frac{1}{\sigma\delta_S} \tag{7.19-3}$$

This impedance can be seen to have the form of a resistance in series with an inductance:

$$Z_s = R_s + j\omega L_i \tag{7.19-4}$$

where

$$R_s = \frac{1}{\sigma\delta_S} \tag{7.19-5}$$

and

$$\omega L_i = \frac{1}{\sigma\delta_S} = R_s \tag{7.19-6}$$

The value Z_s is called the *internal impedance* of the conductor to distinguish it from impedances calculated from fields external to the conductors. For many high frequency applications, the internal impedance is neglected, as, for example, when computing the characteristic impedance of a transmission line wherein a lossless model is usually used for the transmission line conductors. However, a more exact model would include the internal impedance.

The results of (7.19-4), (7.19-5), and (7.19-6) indicate that:

1. The surface resistance per square (per unit length and unit width, where any unit may be used), R_s, of a conductor is that which would be found from the material's conductivity and *by assuming that all current flows uniformly in a depth of material equal to the skin depth*. This interesting fact may be what leads the inexperienced to believe incorrectly that no current flows below the skin depth.
2. The internal reactance of a conductor equals the surface resistance R_s *at all frequencies*.

It is interesting to compare the internal conductor impedance (Fig. 7.19-1) with the distributed reactance of, say, a 50-Ω coaxial cable. The inductance per

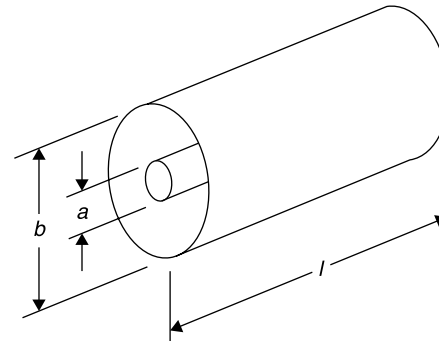

Figure 7.19-1 Coaxial transmission line used to estimate effect of conductor internal impedance.

unit length L of coaxial transmission line will be shown in Section 7.25 to be

$$L = \frac{\mu}{2\pi} \ln \frac{b}{a} = \frac{4\pi \times 10^{-7}}{2\pi} \ln\left(\frac{1 \text{ cm}}{0.43 \text{ cm}}\right) = 166 \text{ nH/m} \qquad (7.19\text{-}7)$$

where the b/a ratio of 2.3 for a 50-Ω air line has been used.

The external inductive reactance, based on the magnetic field between the conductors, at 1 GHz is then

$$\omega L = (6.28 \ \Omega/\text{GHz/nH})(1 \text{ GHz})166 \text{ nH} = 1042 \ \Omega/\text{m} \qquad (7.19\text{-}8)$$

From (7.19-5) the internal resistance of the line is

$$R_S = \frac{1}{\sigma \delta_S} = \frac{1}{(5.8 \times 10^7 \ \text{℧/m}) 2 \times 10^{-6} \text{ m}} = 0.0086 \ \Omega/\text{square} \qquad (7.19\text{-}9)$$

The center conductor has a circumference of $\pi d = (0.43 \text{ cm})\pi = 1.37$ cm. We will consider that this is a flat surface, since its radius of curvature is very large compared to the skin depth. This is the side of a square having the resistance R_s. Then in 1 m, the center conductor has a distributed resistance of

$$R(\text{per meter}) = 0.0086(\Omega/\text{square}) \times \frac{100 \text{ cm}}{1.37 \text{ cm}} = 0.63 \ \Omega/\text{m} \qquad (7.19\text{-}10)$$

The outer conductor also has internal resistance, but it is less by the factor of (1/2.3), or 0.27 Ω/m because of its larger diameter. The total internal conductor resistance is then 0.9 Ω/m.

The internal inductive reactance is numerically equal to the internal resistance, hence it is equal to 0.9 Ω/m, and for most purposes this is negligibly small compared to the external reactance of 1042 Ω/m given by (7.19-8). However, its presence results in the coaxial line having a longer electrical length than an equal length air path of a propagating plane wave (see Exercise 7.19-1).

In general, the average power loss per unit area of a conductor, P_{CL}, is equal to the product of the surface resistivity R_s and the square of the rms current per unit width, I_W. Thus,

$$P_{\mathrm{CL}} = R_s |I_W|^2 \tag{7.19-11}$$

This is a more general expression that can be used to find the conductor losses in cables, waveguides, and cavities.

7.20 THE WAVE EQUATION

To demonstrate that $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields can propagate in free space (more precisely in a charge-free, nonconducting region) we begin with Maxwell's third equation:

$$\nabla \times \vec{\mathbf{E}} = -\frac{\partial \vec{\mathbf{B}}}{\partial \mathrm{t}} \tag{7.20-1}$$

To eliminate $\vec{\mathbf{B}}$, take the curl of both sides, noting that $\vec{\mathbf{B}} = \mu\vec{\mathbf{H}}$ and that the curl and time derivative can be taken in any order:

$$\nabla \times \nabla \times \vec{\mathbf{E}} = -\nabla \times \frac{\partial \vec{\mathbf{B}}}{\partial \mathrm{t}} = -\mu \frac{\partial (\nabla \times \vec{\mathbf{H}})}{\partial \mathrm{t}} \tag{7.20-2}$$

Next, Maxwell's fourth equation applied to any region devoid of conduction currents is

$$\nabla \times \vec{\mathbf{H}} = \frac{\partial \vec{\mathbf{D}}}{\partial \mathrm{t}} = \varepsilon \frac{\partial \vec{\mathbf{E}}}{\partial \mathrm{t}} \tag{7.20-3}$$

Substituting this into (7.20-2),

$$\nabla \times \nabla \times \vec{\mathbf{E}} + \mu\varepsilon \frac{\partial^2 \vec{\mathbf{E}}}{\partial \mathrm{t}^2} = 0 \tag{7.20-4}$$

Next, apply the vector identity (7.16-3)

$$\nabla \times \nabla \times \vec{\mathbf{A}} = \nabla(\nabla \cdot \vec{\mathbf{A}}) - \nabla^2 \vec{\mathbf{A}}$$

and since for any charge-free region, $\nabla \cdot \vec{\mathbf{E}} = 0$, it follows that

$$\nabla^2 \vec{\mathbf{E}} - \mu\varepsilon \frac{\partial^2 \vec{\mathbf{E}}}{\partial \mathrm{t}^2} = 0 \tag{7.20-5}$$

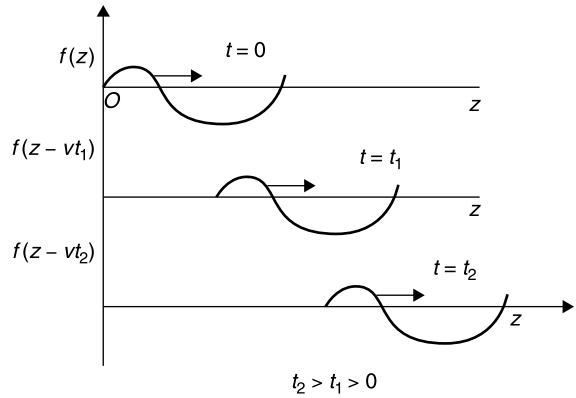

Figure 7.20-1 Electrical disturbance propagating in the $+z$ direction.

This is the *wave equation*. As expressed, it applies in three dimensions. In order to examine this result, let us consider the case for which $\vec{E}$ has only an $\vec{x}$-directed component and only a z variation. For this example (7.20-5) becomes

$$\frac{\partial^2 E_x}{\partial z^2} - \mu\varepsilon \frac{\partial^2 E_x}{\partial t^2} = 0 \tag{7.20-6}$$

This equation describes a propagating wave. It is satisfied by any function of the form [6, p. 273, also 10, p. 26]

$$E_x^+ = f(z - vt) \tag{7.20-7}$$

which describes a disturbance propagating in the $+z$ direction (Fig. 7.20-1), since at the velocity v the quantity $z - vt$ is a constant when z and t both increase together, resulting in a constant value of E_x^+ moving at velocity v in the $+z$ direction with increasing t.

Since (7.20-6) is a second-order differential equation, it has two independent solutions. The second solution is a function $f(z + vt)$ that propagates in the $-z$ direction:

$$E_x^- = f(z + vt) \tag{7.20-8}$$

With either solution, when the indicated differentiations are applied to (7.20-6), the result is

$$f'' + v^2\mu\varepsilon f'' = 0 \tag{7.20-9}$$

which requires that the velocity of propagation is

$$v = \frac{1}{\sqrt{\mu\varepsilon}} = \frac{1}{\sqrt{\mu_r \varepsilon_r}\sqrt{\mu_0 \varepsilon_0}} \tag{7.20-10}$$

For free space, μ_r and ε_r are unity. Therefore

$$v = c = \frac{1}{\sqrt{4\pi \times 10^{-7}\frac{\mathrm{H}}{\mathrm{m}}\frac{1}{36\pi} \times 10^{-9}\frac{\mathrm{F}}{\mathrm{m}}}} \approx 3 \times 10^{8}\frac{\mathrm{m}}{\mathrm{s}}$$

One can imagine Maxwell's excitement on concluding that electromagnetic waves propagated at 300,000 km/s, which he understood to be the approximate speed of light. This probably contributed to his surmise *"we have strong reason to conclude that light itself is an electromagnetic disturbance in the form of waves." (James Clerk Maxwell circa 1863)* [1, p. 4].

Using the same approach, the electric field can be eliminated from Maxwell's third and fourth equations with the result that the $\vec{\mathbf{H}}$ field satisfies the same format as (7.20-5), namely

$$\nabla^2\vec{\mathbf{H}} - \mu\varepsilon\frac{\partial^2\vec{\mathbf{H}}}{\partial t^2} = 0 \tag{7.20-11}$$

Rather than re-solve the wave equation for the $\vec{\mathbf{H}}$ field, it is usually more convenient to solve for either the $\vec{\mathbf{E}}$ or $\vec{\mathbf{H}}$ field and find the other using the appropriate Maxwell third or fourth curl equation.

7.21 THE HELMHOLTZ EQUATIONS

Thus far, the wave equation expressed in terms of $\vec{\mathbf{E}}$ (7.20-5) or $\vec{\mathbf{H}}$ (7.20-11) applies to any time-varying "electrical disturbance," as the generalized waveform sketch of Fig. 7.20-1 describes. However, usually we are interested in sinusoidally varying signals having the implicit time variation $e^{j\omega t}$. Since each differentiation with respect to t produces a factor $j\omega$, the phasor form of the wave equation can be written immediately as

$$\nabla^2\vec{E} + \omega^2\mu\varepsilon\vec{E} = 0 \tag{7.21-1}$$

or

$$\nabla^2\vec{E} + k^2\vec{E} = 0 \tag{7.21-2}$$

and

$$\nabla^2\vec{H} + k^2\vec{H} = 0 \tag{7.21-3}$$

where

$$k \equiv \omega\sqrt{\mu\varepsilon} = \frac{2\pi f}{v} = \frac{2\pi}{\lambda} \tag{7.21-4}$$

$$v = \frac{1}{\sqrt{\mu_R\mu_0\varepsilon_R\varepsilon_0}} = \frac{c}{\sqrt{\mu_r\varepsilon_r}} \tag{7.21-5}$$

Equations (7.21-2) and (7.21-3) are called the *Helmholtz equations*. Note that for uniform plane waves in an unbounded propagating medium k is the *propagation constant*, v is the *velocity of propagation*, and λ is the *wavelength*. For free space propagation, in which $\mu = \mu_0$ and $\varepsilon = \varepsilon_0$, the propagation constant, propagation velocity, and wavelength are denoted k_0, c, and λ_0, respectively. The Helmholtz equations are more convenient to use when the fields have sinusoidal variation.

7.22 PLANE PROPAGATING WAVES

Returning to waves having arbitrary time dependence, suppose we have plane waves that are propagating in the $+z$ and/or $-z$ directions, but have no variations in the x or y directions (Fig. 7.22-1). Then $\partial/\partial x = \partial/\partial y = 0$; and applying Maxwell's third equation produces

$$\nabla \times \vec{\mathbf{E}} = -\mu \frac{\partial \vec{\mathbf{H}}}{\partial t} = \begin{vmatrix} \vec{x} & \vec{y} & \vec{z} \\ 0 & 0 & \dfrac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix} \tag{7.22-1}$$

Note that for this plane wave case, *although the* $\vec{\mathbf{E}}$ *and* $\vec{\mathbf{H}}$ *fields have no variation with respect to x or y, there can be* E_x, E_y, H_x, and H_y *field components*. Equating the separate vector components on each side of (7.22-1),

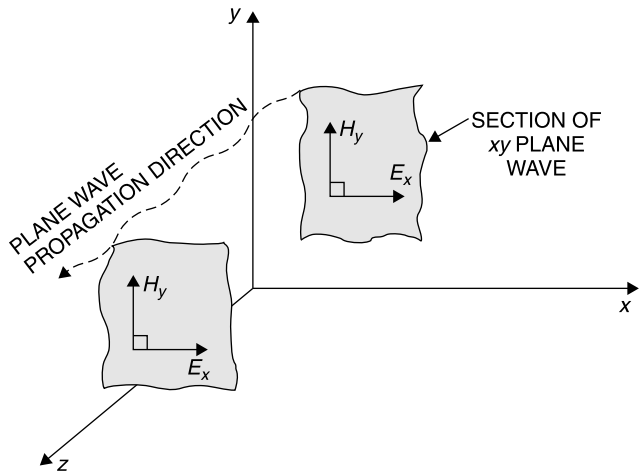

Figure 7.22-1 Plane wave having $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields only in the xy plane and propagation in the $+z$ direction.

$$-\frac{\partial E_y}{\partial z} = -\mu\frac{\partial H_x}{\partial t} \tag{7.22-2}$$

$$\frac{\partial E_x}{\partial z} = -\mu\frac{\partial H_y}{\partial t} \tag{7.22-3}$$

$$0 = -\mu\frac{\partial H_z}{\partial t} \tag{7.22-4}$$

Similarly, from Maxwell's fourth equation

$$\nabla \times H = \varepsilon\frac{\partial E}{\partial t} = \begin{vmatrix} \vec{x} & \vec{y} & \vec{z} \\ 0 & 0 & \dfrac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix} \tag{7.22-5}$$

Again, equating vector components on each side of the equation,

$$-\frac{\partial H_y}{\partial z} = \varepsilon\frac{\partial E_x}{\partial t} \tag{7.22-6}$$

$$\frac{\partial H_x}{\partial z} = \varepsilon\frac{\partial E_y}{\partial t} \tag{7.22-7}$$

$$0 = \varepsilon\frac{\partial E_z}{\partial t} \tag{7.22-8}$$

We first notice, from (7.22-4) and (7.22-8), that the z components of $\vec{E}$ and $\vec{H}$ must be zero, except for a possible time invariant term, but that is of no interest in the case of propagation. Therefore, the time-varying electric and magnetic fields are transverse to the directions ($+z$ and $-z$) of propagation; and the plane wave is said to be a *transverse electromagnetic (TEM) wave*.

We also notice that differentiating (7.22-3) with respect to z and (7.22-6) with respect to t, and substituting the value for $\partial^2 H_y/\partial t\,\partial z$ into (7.22-3) yields the one-dimensional wave equation previously derived:

$$\frac{\partial^2 E_x}{\partial z^2} - \mu\varepsilon\frac{\partial^2 E_x}{\partial t^2} = 0 \tag{7.22-9}$$

We denote the fields that propagate in the $+z$ direction as H^+ and E^+. Specifically, we can relate H_y^+ and E_x^+ to each other [6, p. 278] using (7.22-3). From (7.22-3) we conclude that H_y^+ and E_x^+ have the same functional relationship and differ at most by a constant factor (as well as integration constants which do not vary with z or t, and can be ignored for propagating waves).

We have noted that $E_x^+ = f(z - vt)$. This can be recast as $E_x^+ = g(t - z/v)$. Then on differentiating E_x^+ with respect to z and noting that $v = 1/\sqrt{\mu\varepsilon}$,

(7.22-3) can be written

$$\frac{\partial H_y^+}{\partial t} = -\frac{1}{\mu}\frac{\partial E_x^+}{\partial z} = \frac{1}{\mu}\frac{1}{v}g'\left(t - \frac{z}{v}\right) = \sqrt{\frac{\varepsilon}{\mu}}g'\left(t - \frac{z}{v}\right) \qquad (7.22\text{-}10)$$

Since H_y and E_x have the same functional form, we can integrate (7.22-10) to get

$$H_y^+ = \sqrt{\frac{\mu}{\varepsilon}}E_x^+ \qquad (7.22\text{-}11)$$

$$E_x^+ = \eta H_y^+ \qquad (7.22\text{-}12)$$

where

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} = \sqrt{\frac{\mu_0\mu_r}{\varepsilon_0\varepsilon_r}} \qquad (7.22\text{-}13)$$

Since $\vec{E}$ is in volts/meter and $\vec{H}$ is expressible in amperes/meter, η has the dimension of *ohms*. The factor η occurs frequently as a proportionality constant between $\vec{E}$ and $\vec{H}$ propagating fields and is called the *intrinsic impedance* of the propagating medium. Its reciprocal $1/\eta$ is called the *intrinsic admittance*. In free space

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx \sqrt{\frac{4\pi \times 10^{-7}\ \text{H/m}}{\frac{10^{-9}}{36\pi}\ \text{F/m}}} \approx 120\pi\ \Omega \approx 377\ \Omega \qquad (7.22\text{-}14)$$

A plane wave propagating in either the $+z$ or $-z$ direction can have an E_x component, an E_y component, or both. We determined the relationship of (7.22-14) by beginning with (7.22-3). Using similar reasoning and beginning with (7.22-2) yields [6, p. 279]

$$E_y^+ = -\eta H_x^+ \qquad (7.22\text{-}15)$$

Thus, for plane waves traveling in the $+z$ direction

$$\frac{E_x^+}{H_y^+} = -\frac{E_y^+}{H_x^+} = \eta \qquad (7.22\text{-}16)$$

Similar reasoning for waves propagating in the $-z$ direction shows that

$$\frac{E_x^-}{H_y^-} = -\frac{E_y^-}{H_x^-} = -\eta \qquad (7.22\text{-}17)$$

The relations (7.22-16) and (7.22-17) reveal that

1. $\vec{\mathbf{E}}$ *and* $\vec{\mathbf{H}}$ *are orthogonal to one another in plane propagating waves.*
2. *The instantaneous value of* $\vec{\mathbf{E}}$ *is always* η *times the instantaneous value of* $\vec{\mathbf{H}}$ *for each orthogonal pair of wave components (when propagating in unbounded media).*
3. *The cross product* $\vec{\mathbf{E}} \times \vec{\mathbf{H}}$ *always points in the direction of propagation.* Later we will formalize this result as Poynting's theorem.
4. *The relationships between the transverse* $\vec{\mathbf{E}}$ *and* $\vec{\mathbf{H}}$ *fields of the TEM plane wave apply for any electrical disturbance, regardless of its functional relationship with time.*

For example, the electric field might be an electromagnetic pulse created by a nuclear detonation. Applying a plane wave approximation to the resulting wave when it is sufficiently far from the source of the disturbance, the $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields propagate with their amplitudes related by the intrinsic impedance.

Thus far, this treatment of plane waves applies for any time variation. To determine the solution for $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ in terms of propagation in the +z or −z directions, it would be necessary to substitute the actual form of the time variation into the wave equations and solve for the fields. However, usually we are interested in a sinusoidal time variation, in which case the Helmholtz equations can be applied. Thus the fields for the plane wave case can be written in vector phasor form, namely

$$\vec{E} = \vec{E}_T^{+} e^{-jkz} + \vec{E}_T^{-} e^{+jkz} \tag{7.22-18}$$

$$\vec{H} = \vec{H}_T^{+} e^{-jkz} + \vec{H}_T^{-} e^{+jkz} \tag{7.22-19}$$

where

$$H_T^{+} = \frac{E_T^{+}}{\eta} \quad \text{and} \quad H_T^{-} = -\frac{E_T^{-}}{\eta}$$

where E_T and H_T are the description of the electric and magnetic fields, respectively, in the plane transverse to the direction of propagation. The terms in (7.22-18) and (7.22-19) with superscript "+" represent waves propagating in the $+z$ direction, and those with the superscript "−" represent waves propagating in the $-z$ direction.

7.23 POYNTING'S THEOREM

We now return to the general time variation case to derive Poynting's theorem. As the $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields propagate through space, they carry power. Actually, it

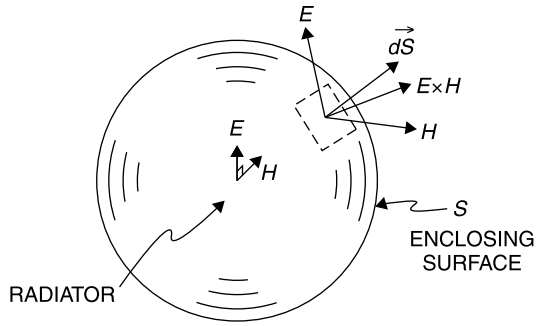

Figure 7.23-1 Poynting's vector represents the power density/unit area radiated out of a closed surface containing a source of propagation.

is impossible to launch a perfectly plane wave because this would require a field that was uniform in, say, the xy plane and extending infinitely far in the $+x, -x, +y$, and $-y$ directions. However, a *directional antenna approximates a plane wave over some aperture*, the larger the aperture dimensions compared to the transmitted wavelength, the more directive (plane-wave-like) the beam and the more gradually it diverges.

Although a practically launched wave that is propagating far from its source can be approximated as a plane wave for the purpose of relating the $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ field components, its wave front actually tends to assume a spherical shape, spreading out so that the amount of power crossing a given area, say, one square meter, becomes less and less with increasing distance from the transmitter. The power density, in watts per square meter, approaches an inverse square law function of distance when measured far from the source that launched the wave (Fig. 7.23-1).

We shall demonstrate that the amount of power passing through a given surface area S of space is given by *Poynting's vector*. This is the cross product of the $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields. The cross product is the magnitudes of the fields multiplied by the sine of the angle between them. But for a wave propagating in space, the $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields are already orthogonal to one another, so their power cross product is simply the product of their amplitudes [6, p. 275].

In this analysis the $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields are their actual instantaneous values, and the power that they convey is the instantaneous power flow regardless of their functional time variation. Later, when applying Poynting's theorem to sinusoidal fields, rms values can be used to determine average power flow.

Poynting's theorem states that the power flow propagating out of a closed surface S is equal to the integral over the surface of the real part of Poynting's vector, $\vec{\mathbf{P}} = \vec{\mathbf{E}} \times \vec{\mathbf{H}}$, where P, E, and H are time-varying functions:

$$\text{Power flow}|_S = \iint_S \vec{\mathbf{P}} \cdot \overrightarrow{dS} \tag{7.23-1}$$

To prove this theorem, we employ the vector identity (7.16-9) into which we have substituted the $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields:

$$\nabla \cdot (\vec{\mathbf{E}} \times \vec{\mathbf{H}}) = \vec{\mathbf{H}} \cdot \nabla \times \vec{\mathbf{E}} - \vec{\mathbf{E}} \cdot \nabla \times \vec{\mathbf{H}} \tag{7.23-2}$$

Next, we substitute values for $\nabla \times \vec{\mathbf{E}}$ and $\nabla \times \vec{\mathbf{H}}$ from Maxwell's third and fourth equations:

$$\nabla \cdot (\vec{\mathbf{E}} \times \vec{\mathbf{H}}) = -\vec{\mathbf{H}} \cdot \frac{\partial \vec{\mathbf{B}}}{\partial \mathrm{t}} - \vec{\mathbf{E}} \cdot \frac{\partial \vec{\mathbf{D}}}{\partial \mathrm{t}} - \vec{\mathbf{E}} \cdot \vec{\mathbf{J}}_c \tag{7.23-3}$$

If ε and μ are constant over the volume of interest,

$$\vec{\mathbf{E}} \cdot \frac{\partial \vec{\mathbf{D}}}{\partial \mathrm{t}} = \frac{1}{2} \frac{\partial (\varepsilon \vec{\mathbf{E}}^2)}{\partial \mathrm{t}} \tag{7.23-4}$$

and

$$\vec{\mathbf{H}} \cdot \frac{\partial \vec{\mathbf{B}}}{\partial \mathrm{t}} = \frac{1}{2} \frac{\partial (\mu \vec{\mathbf{H}}^2)}{\partial \mathrm{t}} \tag{7.23-5}$$

Substituting these values into (7.23-3) and integrating over the volume contained by the closed surface S,

$$-\int_{\mathrm{V}} \nabla \cdot (\vec{\mathbf{E}} \times \vec{\mathbf{H}})\, \mathrm{dV} = \int_{\mathrm{V}} \left[\frac{1}{2} \cdot \frac{\partial (\mu \vec{\mathbf{H}}^2)}{\partial \mathrm{t}} + \frac{1}{2} \cdot \frac{\partial (\varepsilon \vec{\mathbf{E}}^2)}{\partial \mathrm{t}} + \vec{\mathbf{E}} \cdot \vec{\mathbf{J}}_c \right] \mathrm{dV} \tag{7.23-6}$$

Applying the divergence theorem,

$$-\int_S (\vec{E} \times \vec{H}) \cdot \vec{dS} = \int_V \left[\frac{1}{2} \cdot \frac{\partial (\mu \vec{H}^2)}{\partial t} + \frac{1}{2} \cdot \frac{\partial (\varepsilon \vec{E}^2)}{\partial t} + \vec{E} \cdot \vec{J}_c \right] dV \tag{7.23-7}$$

The negative sign on the left side means that instead of diverging from the volume V, the energy is *entering* V. This is consistent with the reasoning that follows.

The first term on the right side has the dimensions of energy per unit volume, henrys $\times$ amperes^2/m^3. When integrated with respect to volume, this is the same form, even to the factor of $\frac{1}{2}$, of the energy stored in an inductor, specifically $\mathrm{LI}^2/2$. As such, this term represents, after integration, the increase in stored magnetic energy per unit time (power inflow) in volume V resulting from the inflow (negative divergence) of Poynting's vector. Similarly, the second term having the dimensions farads $\times$ voltage^2/m^3 also has the dimensions of energy and is equivalent to the energy storage of a capacitor, $\mathrm{CV}^2/2$, having unit volume. This term represents the increase in stored electric energy per unit time (power inflow).

The third term has the dimensions of power dissipated voltage $\times$ current/m^3 and represents, after integration, the amount of power being dissipated within the volume V. Alternatively, if there are mobile charges in V, this term represents the power which increases their motion or, if $\vec{E} \cdot \vec{J}$ is negative, it represents the power contributed by their deceleration [6, p. 275].

All of the net power inflow and dissipation represented on the right side of (7.23-7) comes from the inflow of Poynting's vector. Reversing the sign of this power flow gives the rate at which *energy is leaving* V, that is the power flow out of V.

$$\text{Power flow}|_S = \int_S \vec{P} \cdot \overrightarrow{dS} \quad \text{where } \vec{P} = \vec{E} \times \vec{H} \tag{7.23-8}$$

This is *Poynting's theorem*—the result that we wished to show. Note that P is a power density (watts/meter^2).

Poynting's theorem applies for any time-varying $\vec{E}$ and $\vec{H}$, regardless of the form of time variation (it need not be sinusoidal). *Poynting's theorem gives the instantaneous power flow when instantaneous values are used for the* $\vec{E}$ *and* $\vec{H}$ *fields.*

For sinusoidally varying fields, the average power flow density can be determined using the phasor forms for $\vec{E}$ and $\vec{H}$, specifically

$$\vec{P}_{Av} = \tfrac{1}{2}\,\text{Re}(\vec{E} \times \vec{H}^*) \tag{7.23-9}$$

where * represents the complex conjugate. The imaginary part of $(\vec{E} \times \vec{H}^*)$ represents the ebb and flow of reactive power density through the surface S.

7.24 WAVE POLARIZATION

The electromagnetic wave propagation described in Section 7.22 (Fig. 7.22-3) is known as a linear or plane-polarized wave, since the $\vec{E}$ and $\vec{H}$ fields each lie along an unchanging direction in the plane transverse to the direction of propagation. For example, an electric field oriented in the x direction and propagating in the +z direction has a full period of variation in one wavelength at a given instant of time (Fig. 7.24-1).

The orthogonal magnetic field, in this example in the y direction, has the same pattern along the z axis (Fig. 7.24-2).

The combination of any two or more linearly polarized waves of the same frequency that are in phase and propagating in the same direction results in a linearly polarized wave. Figure 7.24-3 depicts the case of two orthogonal waves, represented by $\vec{E}_x$ and $\vec{E}_y$, that are in phase and of equal amplitude. The resultant wave, $\vec{E}_R$, is oriented at 45° to both $\vec{E}_x$ and $\vec{E}_y$.

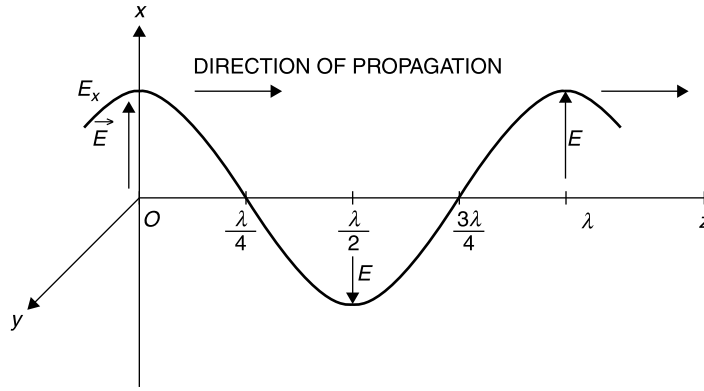

Figure 7.24-1 Amplitude of E_x at an instant in time for a wave propagating in the $+z$ direction. A sinusoid is shown for illustration, but the wave equation applies for all time variations [1, p. 32].

Thus, if

$$\vec{E}_x = \vec{x}E_0 \cos(\omega t - kz) \quad \text{and} \quad \vec{E}_y = \vec{y}E_0 \cos(\omega t - kz) \tag{7.24-1}$$

where $k = 2\pi/\lambda$ is the propagation constant of waves traveling in the $+z$ direction. Then, defining $\vec{E}_R \equiv \vec{E}_x + \vec{E}_y$, the resultant electric field is given by

$$\vec{E}_R(\vec{x} + \vec{y})E_0 \cos(\omega t - kz) \tag{7.24-2}$$

which is simply a vector having the peak magnitude $\sqrt{2}E_0$ and a *direction that is always at an angle of 45° from the x or y axes* (Fig. 7.24-4).

Suppose that instead of two in-phase field components, we excite two space orthogonal components of electric field that have equal amplitudes but are *90° out of phase* with each other.

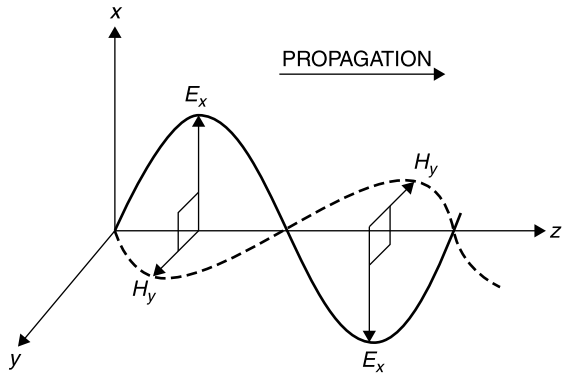

Figure 7.24-2 Orthogonal E and H fields of the wave in Figure 7.24-1 [1, p. 32].

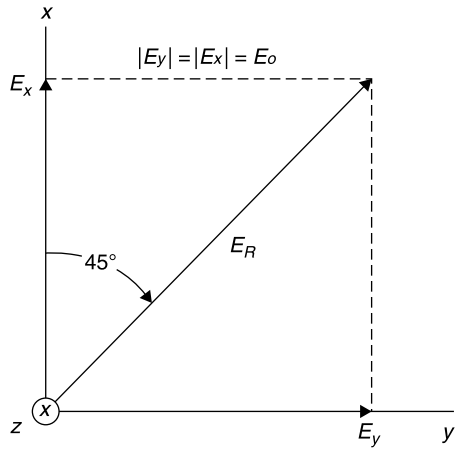

Figure 7.24-3 Sum of two in-phase, linearly polarized waves is also a linearly polarized wave.

In this case

$$\vec{E}_x = \vec{x}E_0 \cos(\omega t - kz) \quad \text{and} \quad \vec{E}_y = \vec{y}E_0 \sin(\omega t - kz) \tag{7.24-3}$$

At $z = 0$ for instance,

$$\vec{E}_R = \vec{x}E_0 \cos \omega t + \vec{y}E_0 \sin \omega t \tag{7.24-4}$$

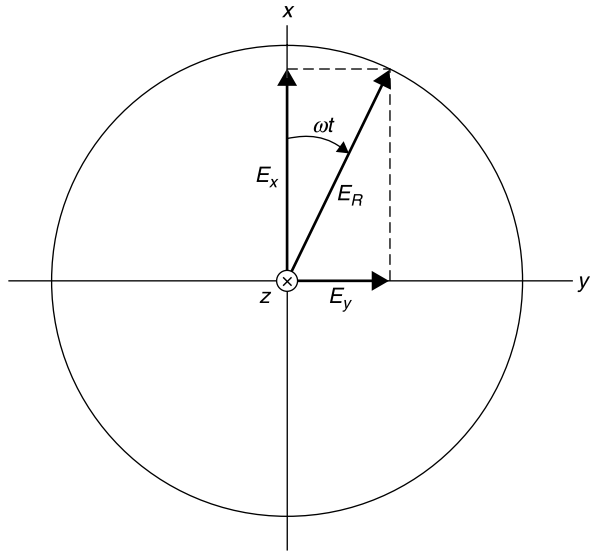

Figure 7.24-4 Circularly polarized plane wave has a *constant amplitude* when projected onto the plane normal to its direction of propagation.

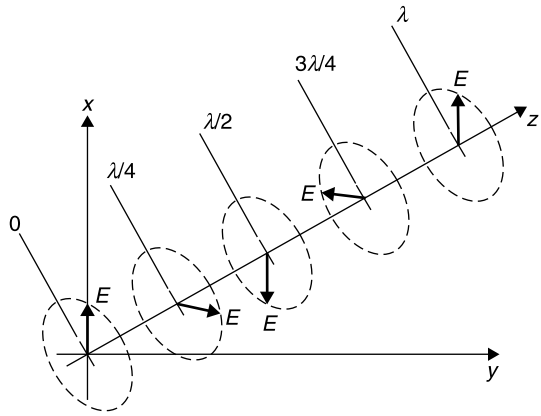

Figure 7.24-5 Left-hand circularly polarized (LHCP) wave propagating in the $+z$ direction. This diagram depicts the steady-state wave amplitudes at quarter wavelength intervals in space at an instant in time ($t = 0$) [1, p. 37].

Since $\cos^2 \omega t + \sin^2 \omega t = 1$, the resultant field, $\vec{E}_R$, has a constant magnitude, E_0, and rotates in the xy plane at an ωt rate. The sense of rotation can be deduced by considering (7.24-4) and Figure 7.24-4. When $\omega t = 0$, $\vec{E}_R$ is oriented in the $+x$ direction. A quarter period later, when $\omega t = \pi/2$, $\vec{E}_R$ is oriented in the $+y$ direction. Thus, for the relative phases of $\vec{E}_x$ and $\vec{E}_y$ in (7.24-3) the rotation shown in Figure 7.24-4 is clockwise. Reversing the 90° phase difference between $\vec{E}_x$ and $\vec{E}_y$ changes the direction of rotation. Thus, for counterclockwise rotation at $z = 0$,

$$\vec{E}_R = \vec{x}E_0 \cos \omega t - \vec{y}E_0 \sin \omega t \tag{7.24-5}$$

If the wave rotates clockwise in a fixed transverse plane when viewed so that it propagates away from you, the polarization is called right-hand circular polarization (RHCP). That is, the advance is the same as that of a right-hand screw.

If it rotates counterclockwise while propagating away from you, it is called left-hand circular polarization (LHCP). Figure 7.24-5 depicts a LHCP wave propagating in the $+z$ direction. Note that for the wave in this figure, *for any fixed transverse plane*, the constant amplitude electric field rotates counterclockwise when viewed along the propagation direction (see Exercise 7.24-3).

If the magnitudes of E_x and E_y are not equal and/or their phase difference is not precisely 90°, then the polarization is called *elliptical polarization* (right or left hand) because the tip of the resultant vector E_R traces an elliptical pattern in the xy plane, rather than a perfect circle.

Circular polarization (CP) transmission is useful in communicating with a linearly polarized receiving antenna (such as a monopole or dipole). The signal will be received no matter how the receiving antenna is tilted. However, a linearly polarized antenna will intercept only one half of the available CP power.

All of the power can be recovered only by using a CP antenna of the same sense as is received. Since the directional sense of propagation is reversed on viewing a wave leaving the transmitter and the same wave arriving at the receiver, if a RHCP antenna was used for transmission, a LHCP antenna must be used for reception.

7.25 EH FIELDS ON TRANSMISSION LINES

Once an electromagnetic wave is launched into space, it diverges and its field strength diminishes due to the spreading out of its power density. This effect can be diminished by using a guiding structure which sets up boundary conditions that reduce the wave's spreading, such as *two conductors* forming a *transmission line*. Because it reduces a propagating wave's spreading, a two-wire transmission line could be considered a form of waveguide. However, the term *waveguide* is generally used to describe hollow pipes or solid dielectrics used for electromagnetic transmission. These terms are not used consistently in the industry, and there are exceptions, such as *coplanar waveguide*, a variation of stripline or microstrip.

For the two-wire line (Fig. 7.25-1) the fields and currents are considerably constrained to travel axially along the line. However, some energy is lost due to

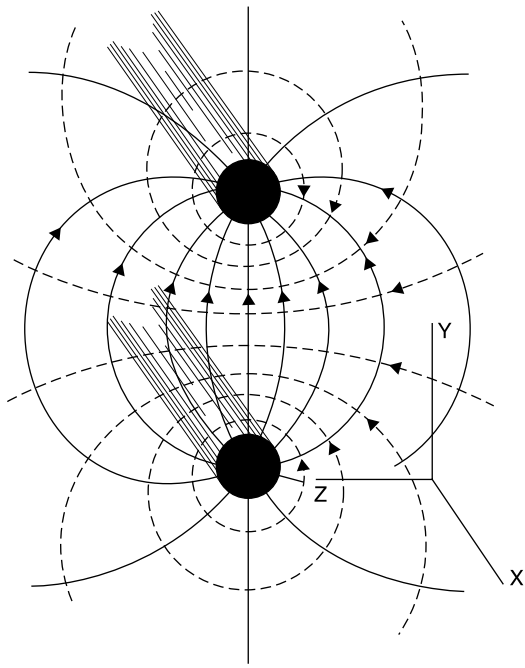

Figure 7.25-1 A two-wire transmission line and its $\vec{\mathbf{E}}$ (solid lines) and $\vec{\mathbf{H}}$ (dashed lines) fields. (After Skilling, 4, p. 160, with permission.)

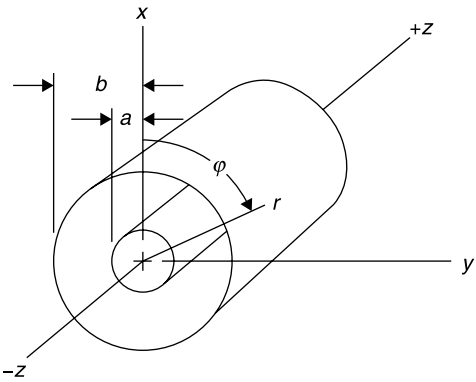

Figure 7.25-2 Coaxial transmission line with cylindrical coordinates r, φ, and z. Note that a is the radius of the inner conductor and b is the radius of the outer conductor.

radiation or leakage of the fields surrounding the line. The principal mode is TEM (transverse electromagnetic), which means that both the electric and magnetic fields are always and everywhere transverse to the direction of propagation. Note that the $\vec{\mathbf{E}}$ field lines terminate on the conductors, and the $\vec{\mathbf{H}}$ field lines are orthogonal to the $\vec{\mathbf{E}}$ lines, as shown in Figure 7.25-2. Both fields fringe outward without limit in the plane transverse to the direction of propagation.

The radiation that leaks from the two-wire line can be reduced to negligible values by causing one conductor to surround the other, as is done with the *coaxial transmission line*. The shielding would be total if the outer conductor could have zero resistivity; but, even with finite resistivity, for all practical purposes the shielding is complete when the outer conductor is a solid metal cylinder of any practical thickness. For analysis we will assume perfect conductors, that is, having infinite conductivity. The following analysis determines the $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ fields in the insulating region of the coaxial transmission line (between the conductors). From this can be derived expressions for the distributed capacitance C and inductance L of the line as well as its characteristic impedance Z_0.

For this transmission line, as for most lines, there are different field patterns by which energy can be propagated. Each is called a *mode* of the transmission line or waveguide. In general, the mode which can propagate at the lowest frequency is called the *dominant mode*. Transmission lines having two separate conductors can propagate all frequencies, theoretically from direct current to light frequencies in the lowest order, or dominant, mode. However, at some frequency, propagation in more than one mode becomes possible, and this is undesirable. Therefore, it is good practice to operate the transmission line at frequencies below the lowest frequency at which the first higher-order mode above the dominant mode can propagate. This is known as the *cutoff frequency of the mode*.

Thus far, we have presented fields and field relationships in rectangular coordinates. However, various symmetries favor the use of other coordinate systems. The most common orthogonal systems are rectangular (also called Cartesian), cylindrical, and spherical. The coaxial transmission line is most easily analyzed using cylindrical coordinates.

To evaluate the fields of the dominant TEM mode we note that by definition both the $\vec{E}$ and $\vec{H}$ fields have components only in the plane transverse to the propagation directions, $+z$ and $-z$. This means that their z-directed components are zero. Since this mode "propagates" at DC (zero frequency), it is possible to evaluate the fields using a static analysis. To solve for the electric field [1, p. 21], apply Gauss's law assuming that there is a charge, $+Q$, on the center conductor and an equal and opposite charge, $-Q$, on the outer conductor of a coaxial line of unit length. Since this is a per-unit-length analysis, we may ignore any fringing fields at the ends of the portion of the line used for analysis.

For any cylindrical surface of radius r defined between the conductors, $a < r < b$,

$$Q = \int_0^L \int_0^{2\pi} \varepsilon |E| r \, d\varphi \, dz \tag{7.25-1}$$

where L is unity, since the analysis is on a per-unit length of transmission line basis. Noting that $\vec{E}$ is independent of z and, due to symmetry, independent of φ, the integration produces

$$Q = \varepsilon E_r (2\pi r) \quad \text{or} \quad E_r = \frac{Q}{2\pi\varepsilon r} \tag{7.25-2}$$

which is the magnitude of the radial $\vec{E}$ field between the conductors (Fig. 7.25-3). The $\vec{E}$ field is the gradient of a scalar function that we will identify as the voltage V having zero potential (voltage) on the outer conductor.

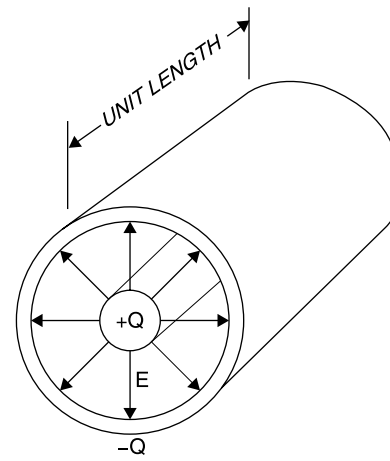

Figure 7.25-3 The $\vec{E}$ field of the TEM mode in coaxial line.

The gradient in cylindrical coordinates is

$$\nabla V = \vec{r}\frac{\partial V}{\partial r} + \vec{\varphi}\frac{1}{r}\frac{\partial V}{\partial \varphi} + \vec{z}\frac{\partial V}{\partial z} \tag{7.25-3}$$

where $\vec{r}, \vec{\varphi}$, and $\vec{z}$ are unit vectors in the r, φ, and z directions. Since the electric field and with it the voltage V vary only in the r direction, only the first term of (7.23-3) is nonzero; thus

$$\begin{aligned} \vec{\mathrm{E}}_r &= -\nabla \mathrm{V} = -\vec{\mathrm{r}}\frac{\partial \mathrm{V}}{\partial \mathrm{r}} \\ \mathrm{V}(\mathrm{r}) &= -\int \vec{\mathrm{E}} \cdot \vec{\mathrm{r}}\, d\mathrm{r} = \frac{-\mathrm{Q}}{2\pi\varepsilon}\int \frac{d\mathrm{r}}{\mathrm{r}} = \frac{-\mathrm{Q}}{2\pi\varepsilon}\ \ln \mathrm{r} + \mathrm{K} \end{aligned} \tag{7.25-4}$$

If we define the value of the voltage to be V_0 on the inner conductor and zero on the outer conductor, the imposition of these boundary conditions requires that

$$\mathrm{K} = \frac{\mathrm{Q}}{2\pi\varepsilon}\ \ln \mathrm{b}$$

$$\mathrm{V}_0 = \frac{\mathrm{Q}}{2\pi\varepsilon}\ \ln\left(\frac{\mathrm{b}}{\mathrm{a}}\right) \tag{7.25-5}$$

$$\frac{\mathrm{Q}}{2\pi\varepsilon} = \frac{\mathrm{V}_0}{\ln(\mathrm{b}/\mathrm{a})} \tag{7.25-6}$$

Since the distributed capacitance C of the coaxial transmission line is Q/V_0, an expression for C is obtained from (7.25-6) as

$$\mathrm{C} = \frac{2\pi\varepsilon}{\ln(\mathrm{b}/\mathrm{a})} \tag{7.25-7}$$

Substituting the value for K and the value for $\mathrm{Q}/2\pi\varepsilon$ into (7.25-4), the expression for the transverse variation of voltage is

$$\begin{aligned} \mathrm{V}(\mathrm{r}) &= \frac{\mathrm{V}_0}{\ln(\mathrm{b}/\mathrm{a})}[-\ln(\mathrm{r}) + \ln(\mathrm{b})] = \frac{\mathrm{V}_0\ \ln(\mathrm{b}/\mathrm{r})}{\ln(\mathrm{b}/\mathrm{a})} \\ \mathrm{V}(\mathrm{r}) &= \frac{\mathrm{V}_0\ \ln(\mathrm{r}/\mathrm{b})}{\ln(\mathrm{a}/\mathrm{b})} \end{aligned} \tag{7.25-8}$$

Then E_r can be obtained by differentiating V(r) with respect to r:

$$\mathrm{E}_r = -\frac{\partial \mathrm{V}}{\partial \mathrm{r}} = \frac{-\mathrm{V}_0}{\ln(\mathrm{a}/\mathrm{b})}\frac{1}{\mathrm{r}} = \frac{\mathrm{V}_0}{\mathrm{r}\ \ln(\mathrm{b}/\mathrm{a})} \tag{7.25-9}$$

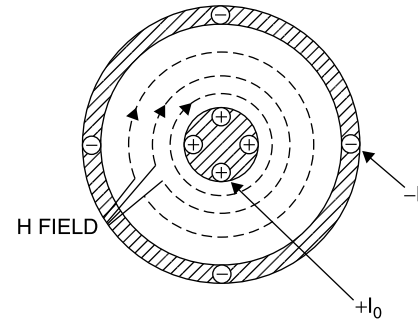

Figure 7.25-4 Magnetic field lines H and the current distribution I_0 for the TEM mode in coaxial line.

where the negative sign in (7.25-9) was eliminated by interchanging the numerator and denominator of the ln expression. Up to this point the expressions apply for any time variation. Usually coaxial lines are used to transmit sinusoidal signals, in which case $\vec{E}$ must also satisfy the Helmholtz equation, and therefore have a variation with z of the form e^{-jkz}. The phasor form of E_r becomes

$$E_r = \frac{V_0}{r \ln(\mathrm{b/a})} e^{-jkz} \tag{7.25-10}$$

Next, we will obtain an expression for the magnetic field between the conductors. The magnetic field lines are concentric with the center conductor as shown in Figure 7.25-4. Its value is determined by using the integral form of Maxwell's fourth equation:

$$\oint \vec{H} \cdot \overrightarrow{dl} = \int_S \vec{J} \cdot \overrightarrow{dS} + \int_S \frac{\partial \vec{D}}{\partial t} \cdot \overrightarrow{dS}$$

The axial $\partial D/\partial t$ term is zero since E and D have no z component. The integral of the current density is just the total current I_0 that flows in the $+z$ direction on the inner conductor and the $-z$ direction on the outer conductor (as is required for the $\vec{H}$ field to be zero outside the coaxial line). Then, since the line integral of $\vec{H}$ is just its magnitude times the circumferential path, the magnetic field for $\mathrm{a} \leq \mathrm{r} \leq \mathrm{b}$ is obtained from

$$2\pi \mathrm{r} \mathrm{H}_\varphi = \mathrm{I}_0$$

Thus

$$\mathrm{H}_\varphi = \frac{\mathrm{I}_0}{2\pi \mathrm{r}} \tag{7.25-11}$$

Equation (7.25-11) is valid for both DC and AC fields. For sinusoidal varia-

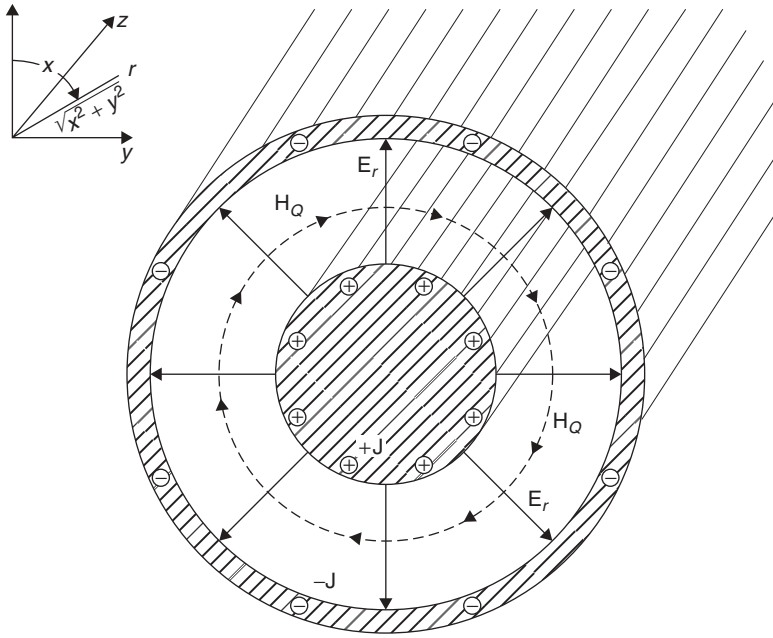

Figure 7.25-5 Coaxial line with transverse E and H fields and axial current density J.

tions, the phasor form of H_φ includes the propagation factor e^{-jkz}. Thus,

$$H_\varphi = \frac{I_0}{2\pi r} e^{-jkz} \tag{7.25-12}$$

For purposes of illustration the current lines in Figure 7.25-5 are shown as filaments into the paper ($+z$ direction) on the center conductor and out of the paper on the inside of the outer conductor. In actuality, however, the currents are distributed uniformly throughout cross sections of the conductors for DC currents and in uniform sheets near the inner surfaces (due to skin effect) of the conductors for high frequency AC currents.

The ratio $V_0/I_0 = Z_0$ is defined as the *characteristic impedance* of the transmission line. This can be evaluated using (7.25-9) and (7.25-11) with the result

$$Z_0 = \frac{V_0}{I_0} = \frac{E_r}{H_\varphi} \frac{\ln(b/a)}{2\pi} \tag{7.25-13}$$

Next, if the value for the intrinsic impedance $\eta = \sqrt{\mu/\varepsilon}$ is substituted for the ratio of the orthogonal fields E_r and H_φ,

$$Z_0 = \frac{1}{2\pi} \sqrt{\frac{\mu}{\varepsilon}} \ln \frac{b}{a} \tag{7.25-14}$$

Substituting the values $\mu = \mu_r\mu_0$, $\varepsilon = \varepsilon_r\varepsilon_0$, and $\sqrt{\mu_0/\varepsilon_0} \approx 120\pi$ gives

$$Z_0 \approx 60\sqrt{\frac{\mu_r}{\varepsilon_r}}\ln\left(\frac{b}{a}\right) \quad (7.25\text{-}15)$$

Previously in (7.25-7) we determined the distributed capacitance C of the transmission line. From (4.11-4) the corresponding distributed inductance is related to Z_0 and C by

$$Z_0 = \sqrt{\frac{L}{C}} \quad (7.25\text{-}16)$$

Substituting C from (7.25-7) and Z_0 from (7.25-14) into (7.25-16), L is determined to be

$$L = \frac{\mu}{2\pi}\ln\frac{b}{a} \quad (7.25\text{-}17)$$

It is interesting to note that we only proved that the ratio of the orthogonal $\vec{E}$ and $\vec{H}$ fields of a propagating wave was equal to η under the conditions that these fields had no variation in the plane transverse to propagation. Although the E_r and H_φ fields for the coaxial line are functions of r, their ratio is equal to the intrinsic impedance η of the insulating region between the conductors. This can be verified by the execution of Exercise 7.25-3.

These derivations demonstrate how Maxwell's equations can be used to solve for the properties of transmission lines. While the principles are the same for all transmission lines, closed-form analytic expressions for the fields about the conductors of various lines such as stripline, microstrip, and coplanar waveguide are much more difficult to formulate that those for coaxial line, one of the few transmission line formats permitting a simple and exact solution. In many cases, a complete analytic approximation is not practical, and electromagnetic (EM) simulation is required to obtain characteristic impedance and other transmission line parameters of sufficient precision for engineering purposes.

7.26 WAVEGUIDES

General Waveguide Solution

The solution for the fields in waveguides (conducting hollow pipes) is different than the solutions for a free space plane wave or a multiconductor transmission line. This is because a TEM wave cannot propagate within a hollow pipe that lacks an axial conductor to carry current, as is available, for example, in a coaxial line. The argument is that a transverse electric field that has a nonzero

curl requires an axial magnetic field. Alternatively, a transverse magnetic field would require either an axial conduction current (which cannot exist since there is no center conductor) or a changing axial electric field. The outcome is that waveguides support *transverse electric (TE)* and *transverse magnetic (TM)* modes of propagation but not a *TEM* mode.

We will solve for the fields of the dominant mode of rectangular waveguide since this is the type most commonly used. The method, described by Ramo and Whinnery [6, Chapters 8 and 9] and attributed by them to Schelkunoff [11], is sufficiently general that it allows solution of the various modes in all types of waveguides. The derivation of the $\vec{E}$ and $\vec{H}$ fields is begun by expanding Maxwell's third and fourth equations, as was done for the plane wave case, but in this case we do not assume that the $\vec{E}$ and $\vec{H}$ fields have no x or y variations, as was assumed for the propagating plane wave case.

On the other hand, we will assume a sinusoidal time variation of the fields since this is the way in which waveguides are used. Indeed, as will be shown, waveguides exhibit a cutoff frequency, below which propagation cannot occur. It should also be noted that even above the cutoff frequency, waveguides are dispersive. That is, they do not have constant propagation delay versus frequency. Consequently, *waveguides cannot propagate absolutely distortion-free representations of arbitrary electrical disturbances*, as can a free space plane wave or the TEM wave in a two or more conductor transmission line.

Waveguides have only a single conductor, and for this reason the transverse electrical field cannot be derived from a static potential function, as was done for the coaxial line. This is because no static potential difference can exist within a charge-free, completely enclosed conductor. Unlike the TEM mode for multiconductor transmission lines, the currents in the waveguide walls have transverse as well as longitudinal components. Waveguides do not have uniquely identifiable voltages and currents associated with propagation, and therefore *waveguides have no uniquely defined characteristic impedance*. As an alternative, a characteristic wave impedance can be defined for waveguide as the ratio of the orthogonal $\vec{E}$ to $\vec{H}$ fields in the transverse plane. However, as we shall see in the study of Green's functions, an absolute impedance for waveguide can be defined under certain conditions.

It will be shown that the characteristic wave impedance of waveguide is approximately equal to the intrinsic impedance of free space, about 377 Ω, for frequencies sufficiently above the cutoff frequency. The fact that waveguides have broad conducting walls results in relatively low ohmic losses. Furthermore, waveguides require no dielectric to support their structure, hence air dielectric is practical, with negligible dielectric loss. Finally, since waveguide propagation is totally enclosed, there is no radiative loss. As a result, waveguides usually provide the lowest loss practical transmission media for guided waves. Their low loss as well as their high power capacity are principal reasons for their use. In addition to being an efficient transmission medium, they find considerable use in the construction of high Q resonator elements, couplers, and power dividers.

We will assume lossless conditions and begin the analysis of waveguides by expanding Maxwell's third and fourth equations, assuming a z dependence of $e^{-\gamma z}$ for waves traveling in the $+z$ direction and $e^{+\gamma z}$ for those traveling in the $-z$ direction. A real value for γ means wave attenuation, while an imaginary value for γ means lossless propagation.

The following analysis assumes $e^{j\omega t}$ time dependence. Taking the curl of $\vec{E}$ and of $\vec{H}$, noting that $\overrightarrow{J_C} = 0$ in a source-free region, performing differentiation with respect to time by multiplying by $j\omega$, and differentiation with respect to z by multiplying by $-\gamma$ (for $+z$-directed waves) gives

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} = \begin{vmatrix} \vec{x} & \vec{y} & \vec{z} \\ \partial & \partial & \partial \\ \overline{\partial x} & \overline{\partial y} & \overline{\partial z} \\ E_x & E_y & E_z \end{vmatrix} \qquad \nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t} = \begin{vmatrix} \vec{x} & \vec{y} & \vec{z} \\ \partial & \partial & \partial \\ \overline{\partial x} & \overline{\partial y} & \overline{\partial z} \\ E_x & E_y & E_z \end{vmatrix}$$

For $+z$-directed propagation, the phasor forms of the above equations become

$$-j\omega\mu H_x = \frac{\partial E_z}{\partial y} + \gamma E_y \tag{7.26-1a}$$

$$-j\omega\mu H_y = -\frac{\partial E_z}{\partial x} - \gamma E_x \tag{7.26-1b}$$

$$-j\omega\mu H_z = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \tag{7.26-1c}$$

$$j\omega\varepsilon E_x = \frac{\partial H_z}{\partial y} + \gamma H_y \tag{7.26-1d}$$

$$j\omega\varepsilon E_y = -\frac{\partial H_z}{\partial x} - \gamma H_x \tag{7.26-1e}$$

$$j\omega\varepsilon E_z = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \tag{7.26-1f}$$

[Equations (7.26-1a) to (7.26-1b) from Ramo and Whinnery, 6, p. 316, reprinted with permission.]

For all of the equations in (7.26-1) and in the following derivation, E_x, E_y, E_z, H_x, H_y, and H_z are functions only of x and y. To recover the time and z dependence, multiply by $e^{(j\omega t - \gamma z)}$ for waves traveling in the $+z$ direction and take the real part thereof. For waves traveling in the $-z$ direction replace $-\gamma z$ with $+\gamma z$.

With diligent algebraic application, the relations given in (7.26-1) can be manipulated to give H_x, H_y, E_x, and E_y in terms of E_z and H_z. For example, the first equation in (7.26-2) is obtained by solving (7.26-1a) for E_y and substituting this value into (7.26-1e). Also note that $k^2 = \omega^2\mu\varepsilon$, as was previously defined. Then, for $+z$-directed propagation:

$$H_x = \frac{1}{\gamma^2 + k^2}\left[\left(j\omega\varepsilon\frac{\partial E_z}{\partial y} - \gamma\frac{\partial H_z}{\partial x}\right)\right. \tag{7.26-2a}$$

$$H_y = -\frac{1}{\gamma^2 + k^2}\left[\left(j\omega\varepsilon\frac{\partial E_z}{\partial x} + \gamma\frac{\partial H_z}{\partial y}\right)\right. \tag{7.26-2b}$$

$$E_x = -\frac{1}{\gamma^2 + k^2}\left[\left(\gamma\frac{\partial E_z}{\partial x} + j\omega\mu\frac{\partial H_z}{\partial y}\right)\right. \tag{7.26-2c}$$

$$E_y = \frac{1}{\gamma^2 + k^2}\left[\left(\gamma\frac{\partial E_z}{\partial y} + j\omega\mu\frac{\partial H_z}{\partial x}\right)\right. \tag{7.26-2d}$$

where $k = \omega\sqrt{\mu\varepsilon}$. [Equations (7.26-2a) to (7.26-2d) from Ramo and Whinnery, 6, p. 317, reprinted with permission.] Note that in this derivation, *k is not the propagation constant*. The equations of (7.26-2) apply to all transmission lines and waveguides since they have been derived directly from Maxwell's equations without any limiting assumptions, other than a sinusoidal time variation. In addition, the electric and magnetic fields must also satisfy the Helmholtz equations. Therefore,

$$\nabla^2\vec{E} = -k^2\vec{E} \quad \text{and} \quad \nabla^2\vec{H} = -k^2\vec{H} \tag{7.26-3a,b}$$

The operator ∇^2 is three dimensional and can be separated dimensionally. Thus

$$\nabla^2\vec{E} = \nabla_{xy}^2\vec{E} + \frac{\partial^2\vec{E}}{\partial z^2} \tag{7.26-4}$$

$$\nabla^2\vec{H} = \nabla_{xy}^2\vec{H} + \frac{\partial^2\vec{H}}{\partial z^2} \tag{7.26-5}$$

The second term is due to the axial field variation along the waveguide, and it has the assumed functional description $e^{-\gamma z}$ in the $+z$ direction. Thus

$$\frac{\partial^2\vec{E}}{\partial z^2} = \gamma^2\vec{E} \tag{7.26-6}$$

$$\frac{\partial^2\vec{H}}{\partial z^2} = \gamma^2\vec{H} \tag{7.26-7}$$

Substituting this format into (7.26-3a) and (7.26-3b) gives

$$\nabla_{xy}^2\vec{E} = -(\gamma^2 + k^2)\vec{E} \tag{7.26-8}$$

$$\nabla_{xy}^2\vec{H} = -(\gamma^2 + k^2)\vec{H} \tag{7.26-9}$$

Some various waveguide types are shown in Figure 7.26-1 along with the $\vec{E}$ field of their *dominant mode*, the mode that propagates at the lowest frequency.

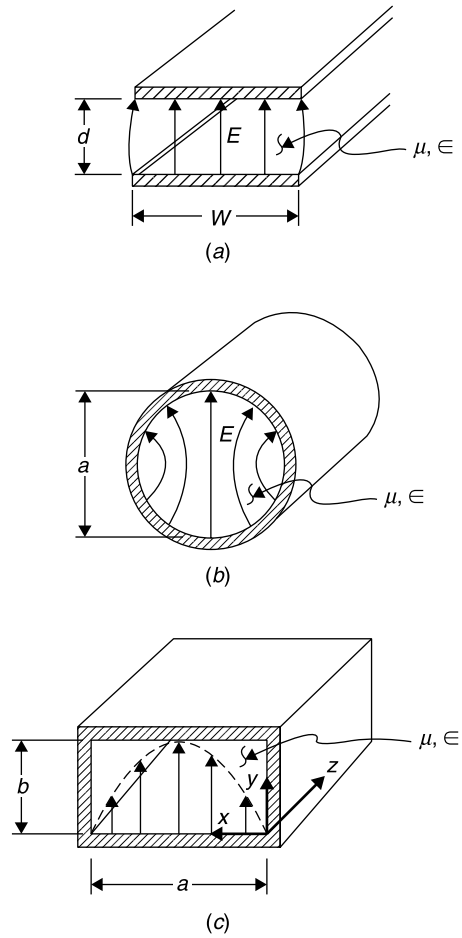

Figure 7.26-1 Waveguide types and the cross-sectional E field pattern of the dominant mode: (*a*) parallel-plate waveguide, (*b*) circular waveguide, and (*c*) rectangular waveguide.

The above equations will be used to determine the propagation characteristics of rectangular waveguide.

Waveguide Types

The parallel-plate waveguide is simple in form and propagates at all frequencies, even direct current. Since it is a two-conductor line, it also can be classed as a transmission line. In principle, it does not radiate because $\vec{E} \times \vec{H}$ is only in the z direction. However, any imperfections or discontinuities in its construction can result in energy leakage. Thus, in practice, it can have radiation loss.

Also, some means must be provided to support the conductors for air dielectric operation since there are no sides.

The round waveguide shown in (Fig. 7.26-1*b*) also can be fabricated simply. In fact, copper water pipe of suitable diameter would serve nicely. The disadvantage of this waveguide is that the polarization of the electric field, shown vertically in Figure 7.26-1*b*, can rotate as the wave propagates due to irregularities in the waveguide. As a result, when the wave reaches the end of the waveguide, a vertically polarized receptor may not be able to intercept all, or even any, of its energy. The rectangular waveguide overcomes these disadvantages and therefore is the most commonly used. The method of solution follows that of Ramo and Whinnery [6].

Rectangular Waveguide Fields

Rectangular waveguide is commonly used in microwave applications and therefore will be the subject of our analysis. Equations (7.26-8) and (7.26-9) are to be satisfied by all of the field components in the waveguide. We have already noted that either TE or TM modes can exist in the waveguide but not TEM modes. Furthermore, we have related the transverse $\vec{E}$ and $\vec{H}$ fields to the axial $\vec{E}$ and $\vec{H}$ fields by the equations in (7.26-2). Thus, it is only necessary to solve for the axial field component for a mode class, TE or TM, and then the transverse fields can be determined from the axial field. Let us consider the TE modes that have an axial $\vec{H}$ field but no axial $\vec{E}$ field. Applying (7.26-9) to the H_z component of the field,

$$\nabla_{xy}^2 H_z = \frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} = -k_c^2 H_z \tag{7.26-10}$$

where $k_c^2 = \gamma^2 + k^2$ and $k^2 = \omega^2 \mu\varepsilon$. Equation (7.26-10) is a partial differential equation that can be solved by the method of separation of variables. To do so we assume it has a solution of the form

$$H_z = f(x)g(y) \tag{7.26-11}$$

The functions $f(x)$ and $g(y)$ describe the field dependency in the plane transverse to the direction of propagation, where $f(x)$ is only a function of x and $g(y)$ is only a function of y. In the transverse (xy) plane the solution is of the form

$$H_z = fg \tag{7.26-12}$$

in which the z variation is implicit. Substituting this into (7.26-10) and performing the indicated differentiations give

$$f''g + fg'' = -k_c^2 fg$$
$$\frac{f''}{f} + \frac{g''}{g} = -k_c^2 \tag{7.26-13}$$

where the double prime indicates the second derivative. Since f and g can be varied independently, each of the terms on the left side of (7.26-13) must equal a constant. There are various forms for the solution of (7.26-13) depending upon whether the ratios are negative, positive, or one negative and one positive [6, p. 355]. If both ratios are assumed to be negative, then we can assign them the values k_x^2 and k_y^2 such that

$$\frac{f''}{f} = -k_x^2 \tag{7.26-14}$$

$$\frac{g''}{g} = -k_y^2 \tag{7.26-15}$$

The solutions for positive values of k_x^2 and k_y^2 are sinusoids:

$$f(x) = C_1 \cos k_x x + C_2 \sin k_x x \tag{7.26-16}$$

$$g(y) = C_3 \cos k_y y + C_4 \sin k_y y \tag{7.26-17}$$

and from (7.26-13)

$$k_x^2 + k_y^2 = k_c^2 = \gamma^2 + \omega^2 \mu\varepsilon \tag{7.26-18}$$

There are also solutions in terms of sinh and cosh functions if the ratios are positive, but for present purposes the sinusoidal solutions suffice.

Applying Boundary Conditions

For the rectangular waveguide the electric field tangential to a conducting waveguide wall must be zero. Therefore, $E_x = 0$ at $x = 0$ *and* a and $E_y = 0$ at $y = 0$ *and* b. Since, for the TE waves, $E_z = 0$, and, referring to (7.26-2c) and (7.26-2d), since E_y and E_x are proportional to $\partial H_z/\partial y$ and $\partial H_z/\partial x$, respectively, it follows that

$$\frac{\partial H_z}{\partial x} = 0 \text{ at } x = 0, a \quad \text{and} \quad \frac{\partial H_z}{\partial y} = 0 \text{ at } y = 0, b \tag{7.26-19a,b}$$

The boundary conditions at $x = 0$ and $y = 0$ require that only the cosine terms in (7.26-16) and (7.26-17) can apply. We will use the remaining boundary conditions at $x = a$ and $y = b$ shortly. Thus

$$H_z = H_0 \cos k_x x \cos k_y y \tag{7.26-20}$$

where H_0 is an amplitude constant. Applying (7.26-2a) thru (7.26-2d) and recognizing that $E_z = 0$,

$$H_x = \frac{\gamma k_x}{k_c^2} H_0 \sin k_x x \cos k_y y \tag{7.26-21}$$

$$H_y = \frac{\gamma k_y}{k_c^2} H_0 \cos k_x x \sin k_y y \tag{7.26-22}$$

$$E_x = \frac{j\omega\mu k_y}{k_c^2} H_0 \cos k_x x \sin k_y y \tag{7.26-23}$$

$$E_y = \frac{-j\omega\mu k_x}{k_c^2} H_0 \sin k_x x \cos k_y y \tag{7.26-24}$$

where $k_c^2 = \gamma^2 + k^2$ and $k^2 = \omega^2\mu\varepsilon$.

Propagation Constants and Waveguide Modes

The boundary conditions that $E_y = 0$ at $x = a$ and $E_x = 0$ at $y = b$ in (7.26-24) and (7.26-23) require, respectively, that

$$k_x = \frac{m\pi}{a} \quad \text{and} \quad k_y = \frac{n\pi}{b} \tag{7.26-25a,b}$$

where m and n are integers $(0, 1, 2, 3 \ldots)$, both of which cannot be zero simultaneously, that denote the mode number. More about modes shortly. Then

$$(k_c)_{m,n} = \sqrt{k_x^2 + k_y^2} = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \tag{7.26-26}$$

For lossless propagation

$$\gamma = \alpha + j\beta \tag{7.26-27a}$$

must be imaginary. Thus, since

$$\gamma^2 = k_c^2 - k^2 \tag{7.26-27b}$$

$$\gamma = j\beta = \sqrt{k_c^2 - k^2} = \sqrt{k_c^2 - (2\pi f)^2\mu\varepsilon} \tag{7.26-27c}$$

For propagation to occur the quantity under the square root must be negative, or

$$f > \frac{k_c}{2\pi\sqrt{\mu\varepsilon}} = \frac{1}{2\pi\sqrt{\mu\varepsilon}} \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \tag{7.26-28}$$

Since the operating wavelength at frequency f is $\lambda = 1/f\sqrt{\mu\varepsilon}$, the condition at which (7.26-28) is just satisfied corresponds to the *cutoff wavelength* and is given by

$$\lambda_C = \frac{2\pi}{k_c} = \frac{1}{\sqrt{(m/2a)^2 + (n/2b)^2}} \tag{7.26-29}$$

Because $f_c\lambda_c = v = 1/\sqrt{\mu\varepsilon}$, a *cutoff frequency* can be defined as

$$f_c = \frac{v}{\lambda_c} = \frac{c}{\lambda_c}\frac{1}{\sqrt{\mu_r\varepsilon_r}} \tag{7.26-30}$$

Thus when $f < f_c$ (or $\lambda > \lambda_c$), the waveguide does not propagate in that mode. Since rectangular waveguide is made with $a > b$, and since m and n cannot both be zero simultaneously, the longest free space wavelength for propagation in rectangular waveguide applies to the TE_{10} mode ($m = 1$ and $n = 0$) given by

$$(\lambda_C)_{10} = 2a \tag{7.26-31}$$

Substituting $k_x = m\pi/a$, $k_y = n\pi/b$, $\gamma^2 + k^2 = k_c^2$, and $\gamma = j\beta$ for propagating waves into (7.26-21) through (7.26-24) gives the transverse field components as

$$E_x = \frac{j\omega\mu}{k_c^2}\frac{n\pi}{b}H_0 \cos\frac{m\pi x}{a}\sin\frac{n\pi y}{b} \tag{7.26-32}$$

$$E_y = \frac{-j\omega\mu}{k_c^2}\frac{m\pi}{a}H_0 \sin\frac{m\pi x}{a}\cos\frac{n\pi y}{b} \tag{7.26-33}$$

$$H_x = \frac{j\beta}{k_c^2}\frac{m\pi}{a}H_0 \sin\frac{m\pi x}{a}\cos\frac{n\pi y}{b} \tag{7.26-34}$$

$$H_y = \frac{j\beta}{k_c^2}\frac{n\pi}{b}H_0 \cos\frac{m\pi x}{a}\sin\frac{n\pi y}{b} \tag{7.26-35}$$

Note that H_z is given by (7.26-20). Generally, a rectangular waveguide is operated in the *dominant mode*, the mode that propagates at the lowest frequency. This is the TE_{10} mode whose cutoff wavelength is $2a$. For the TE_{m0} modes $E_x = 0$ and E_y has a sinusoidal variation in amplitude across the x dimension of the waveguide. The amplitude patterns for the TE_{10}, TE_{20}, and TE_{30} modes are shown in Figure 7.26-2.

Similarly, for the TE_{0n} modes $E_y = 0$ and the E_x field component varies sinusoidally in the y direction as shown in Figure 7.26-3. Field patterns for additional TE and TM modes can be found in [6].

It should be noted that (7.26-20) through (7.26-24) and (7.26-32) through (7.26-35) do not contain the implicit $e^{-j(\omega t - \beta z)}$ factor. To obtain the actual field

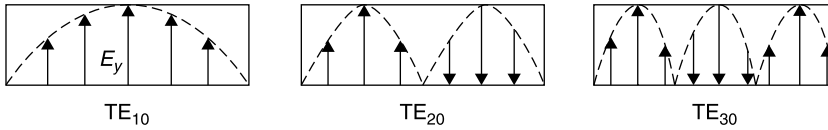

Figure 7.26-2 Electric field magnitude patterns of the TE_{10}, TE_{20}, and TE_{30} modes.

components as a function of z and t, one must reinsert this $e^{-j(\omega t-\beta z)}$ factor and then take the real part thereof. For example,

$$E_y(x, y, z, t) = \frac{\omega\mu k_x}{k_c^2} H_0 \sin\frac{m\pi x}{a} \cos\frac{n\pi y}{b} \cos(\omega t - \beta z) \tag{7.26-36}$$

$$H_x(x, y, z, t) = \frac{-\beta k_x}{k_c^2} H_0 \sin\frac{m\pi x}{a} \cos\frac{n\pi y}{b} \cos(\omega t - \beta z) \tag{7.26-37}$$

Note that E_y and H_x are in phase for all values of t and z and the minus sign preceding H_x. Thus $\vec{E} \times \vec{H}$ yields real power flow and it is in the $+z$ direction.

Usually, it is undesirable to have the waveguide capable of propagation in more than one mode because energy in the dominant mode might be coupled to a *higher order mode*, one having a shorter cutoff wavelength, and then that energy may not be recoverable by the mode launchers and receptors used at the ends of the transmission system. Even if the launchers could recover the higher order mode, the various modes travel at different phase velocities, as will be seen, and different modes would not usually combine in phase.

In short, the best practice is usually to operate in the frequency range for which only one mode, the *dominant mode*, can propagate. Standard waveguides have a ratio of width to height of at least 2. Thus, the next higher propagating mode, the TE_{20}, would have a cutoff wavelength of a, yielding an operating bandwidth ratio of 2 to 1 for the dominant TE_{10} mode only. However, operation very near to cutoff is impractical, and so practically the operating bandwidth is about 1.5 to 1. For example, WR-90 waveguide has a recommended operating band of 8.2 to 12.4 GHz. See Appendix E.

The modes that do not propagate are called *evanescent modes*, evanescent meaning "vanishing." Their propagation constant is real, and this corresponds

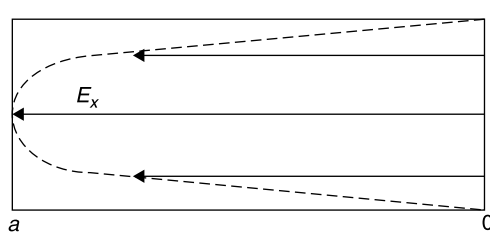

Figure 7.26-3 Electric field magnitude of the TE_{01} mode (for which $\lambda_c = 2b$).

to an attenuating wave amplitude with distance from the site at which they are excited. Irregularities and obstacles in the waveguide serve to excite the field patterns of higher order modes; and, while a waveguide operated only in the dominant mode frequency range may not propagate them, evanescent modes and their resulting fields, which decrease reactively from their excitation sites, store energy, causing them to appear as reactive circuit elements in the waveguide.

Although we expressed the cutoff of the waveguide relative to the operating wavelength λ, the wavelength of propagation *inside the waveguide*, λ_g, is considerably longer than the operating wavelength. This is particularly true as λ approaches the cutoff wavelength λ_c. This is shown by

$$\lambda_g = \frac{2\pi}{\beta} = \frac{2\pi}{\sqrt{(2\pi f)^2 \mu\varepsilon - k_c^2}} \tag{7.26-38}$$

Since the operating wavelength is related to frequency by $f\lambda = 1/\sqrt{\mu\varepsilon}$ and since, from (7.26-29) $\lambda_c k_c = 2\pi$, (7.26-38) can be rewritten as

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (\lambda/\lambda_c)^2}} = \frac{\lambda}{\sqrt{1 - (f_c/f)^2}} \tag{7.26-39}$$

Note that for an air-filled waveguide, $\mu_r = \varepsilon_r = 1$ and $\lambda = \lambda_0 = c/f$.

Characteristic Wave Impedance for Waveguides

The ratio of voltage to current for a traveling wave is Z_0, the characteristic impedance of the transmission line. However, a waveguide does not have unique terminals at which voltage and current can be specified and measured. On the other hand, the ratio of the transverse $\vec{E}$ and $\vec{H}$ fields is unique and is defined as the *characteristic wave impedance*, as was the case for a plane wave propagating in an unbounded medium. In that case, it was shown that the impedance was frequency independent and equal to $\eta = \sqrt{\mu/\varepsilon}$. In waveguide, this ratio approaches η for frequencies well above cutoff but changes rapidly near cutoff with the same functionality as the guide wavelength in (7.26-39).

We find the transverse $\vec{E}$ and $\vec{H}$ fields from (7.26-21) through (7.26-24) and define the characteristic wave impedance for the TE modes as

$$Z_{\text{TE}} = \frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{j\omega\mu}{\gamma} \tag{7.26-40}$$

Substituting $\gamma = j\sqrt{k^2 - k_c^2}$, $k = 2\pi/\lambda$, $k_c = 2\pi/\lambda_c$, and $\omega\mu = k\eta$, (7.26-40) can be rewritten as

$$Z_{\text{TE}} = \frac{\eta}{\sqrt{1 - (\lambda/\lambda_C)^2}} = \frac{\eta}{\sqrt{1 - (f_C/f)^2}} \tag{7.26-41a}$$

where $\eta = \sqrt{\mu/\varepsilon} \approx 120\pi\sqrt{\mu_r/\varepsilon_r} \approx 377\sqrt{\mu_r/\varepsilon_r}\ \Omega$. By a similar analysis one can obtain the characteristic wave impedance for the TM waves, namely

$$Z_{\mathrm{TM}} = \eta\sqrt{1 - \left(\frac{\lambda}{\lambda_C}\right)^2} = \eta\sqrt{1 - \left(\frac{f_C}{f}\right)^2} \tag{7.26-41b}$$

Note that both Z_{TE} and Z_{TM} are real only for frequencies above cutoff ($f > f_C$).

Phase and Group Velocities

Waveguide shows high dispersion near the cutoff frequency. However, pulsed signals on microwave carrier frequencies propagated through waveguides are not likely to have significant distortion since the relative bandwidths are usually fairly small and waveguide paths usually are fairly short, for example, about 100 ft to ascend a communication tower. The definitions and expressions for *phase velocity* and *group velocity* are as follows:

$$\text{Phase velocity} = v_P \equiv \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu\varepsilon}\sqrt{1 - (f_C/f)^2}} \tag{7.26-42}$$

$$\text{Group velocity} = v_g \equiv \frac{\partial\omega}{\partial\beta} = \frac{\sqrt{1 - (f_C/f)^2}}{\sqrt{\mu\varepsilon}} \tag{7.26-43}$$

From (7.26-43) it is seen that group velocity, the speed at which information propagates, is slower than the speed of light in the medium, $v = 1/\sqrt{\mu\varepsilon} = c/\sqrt{\mu_r\varepsilon_r}$, ever more so as the operating frequency approaches the cutoff frequency of the waveguide. The phase velocity is correspondingly faster. This violates no relativity constraints, since no physical matter (or information) travels at this speed, only a *steady-state* constant-phase wave front.

TE and TM Mode Summary for Rectangular Waveguide

The same technique just used to solve for the TE modes can be applied to solve for the fields and related parameters of TM modes. The results can be cast into the format shown below using the various equivalences covered in the TE mode development [after Ramo and Whinnery, 6, Sec. 9.03, reprinted with permission].

TE_{mn} Modes

$$H_z = H_0 \cos\frac{m\pi x}{a} \cos\frac{n\pi y}{b} \tag{7.26-44a}$$

$$E_x = j\eta\frac{n\pi}{bk_c}\frac{f}{f_C}H_0 \cos\frac{m\pi x}{a} \sin\frac{n\pi y}{b} \tag{7.26-44b}$$

$$E_y = -j\eta \frac{m\pi}{ak_c} \frac{f}{f_C} H_0 \sin \frac{m\pi x}{a} \cos \frac{n\pi y}{b} \tag{7.26-44c}$$

$$H_x = -\frac{E_y}{Z_{\text{TE}}} \tag{7.26-44d}$$

$$H_y = \frac{E_x}{Z_{\text{TE}}} \tag{7.26-44e}$$

$$\lambda_C = \frac{2\pi}{k_c} = \frac{1}{\sqrt{(m/2a)^2 + (n/2b)^2}} \tag{7.26-44f}$$

where

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \qquad k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \qquad Z_{\text{TE}} = \frac{\eta}{\sqrt{1 - (\lambda/\lambda_C)^2}}$$

TM_{mn} Modes

$$E_z = E_0 \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} \tag{7.26-45a}$$

$$H_x = j \frac{n\pi}{b\eta k_c} \frac{f}{f_C} E_0 \sin \frac{m\pi x}{a} \cos \frac{n\pi y}{b} \tag{7.26-45b}$$

$$H_y = -j \frac{m\pi}{a\eta k_c} \frac{f}{f_C} E_0 \cos \frac{m\pi x}{a} \sin \frac{n\pi y}{b} \tag{7.26-45c}$$

$$E_x = Z_{\text{TM}} H_y \tag{7.26-45d}$$

$$E_y = -Z_{\text{TM}} H_x \tag{7.26-45e}$$

$$\lambda_C = \frac{2\pi}{k_c} = \frac{1}{\sqrt{(m/2a)^2 + (n/2b)^2}} \tag{7.26-45f}$$

where

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \qquad k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \qquad Z_{\text{TM}} = \eta\sqrt{1 - (\lambda/\lambda_C)^2}$$

TE_{10} Mode For the TE_{10} mode, $m = 1$, $n = 0$, and $\lambda_C = 2a$. The resultant expressions for the $\vec{E}$ and $\vec{H}$ fields of the TE_{10} mode are obtained from the general expressions of (44):

$$E_y = E_0 \sin \frac{\pi x}{a} \tag{7.26-46}$$

$$H_x = -\left(\frac{E_0}{Z_{\text{TE}}}\right) \sin\left(\frac{\pi x}{a}\right) \tag{7.26-47}$$

$$H_z = \frac{jE_0}{\eta}\left(\frac{\lambda}{2a}\right)\cos\left(\frac{\pi x}{a}\right) \tag{7.26-48}$$

$$E_x = H_y = 0 \tag{7.26-49}$$

$$\lambda_g = \frac{\lambda}{\sqrt{1-(\lambda/2a)^2}} \tag{7.26-50}$$

$$Z_{\text{TE}} = \frac{\eta}{\sqrt{1-(\lambda/2a)^2}} \tag{7.26-51}$$

$$v_p = \frac{v}{\sqrt{1-(\lambda/2a)^2}} \tag{7.26-52}$$

$$v_g = v\sqrt{1-(\lambda/2a)^2} \tag{7.26-53}$$

$$\lambda_c = 2a \tag{7.26-54}$$

$$f_c = \frac{1}{2a\sqrt{\mu\varepsilon}} \tag{7.26-55}$$

where $\eta \approx 377\sqrt{\mu_r/\varepsilon_r}\ \Omega$ and $v = 1/\sqrt{\mu\varepsilon} = c/\sqrt{\mu_r\varepsilon_r}$.

The field pattern of the TE_{10} mode can be seen in Figure 7.26-4 to move as a packet, one guide wavelength long. To obtain a complete expression for the field pattern, reinsert the $e^{j(\omega t-\beta z)}$ factor and take the real part of the result using (7.26-46), (7.26-47), and (7.26-48). For example,

$$E_y(x,z,t) = E_0 \sin\frac{\pi x}{a}\cos(\omega t - \beta z) \tag{7.26-56}$$

The pattern shown in Figure 7.26-4 is for $t = 0$. The E_y and H_x field components are in phase, as is necessary for real power flow. Their cross product, Poynting's vector, gives the power density flow in the waveguide's cross section. Integration over the cross-sectional area of the waveguide yields an expression for the total power flow. The conducting waveguide walls constrain the fields to within the hollow waveguide pipe. They do so with a current distribution that confines the fields within the waveguide walls, rather than radiating into the surrounding space. The current distribution can be obtained from the field components within the waveguide using Maxwell's fourth equation.

The conduction current distribution on the inside walls of the waveguide are depicted in Figure 7.26-5. Notice that for the instant of time shown, the currents converge at a point $\lambda_g/4$ from the left end of the guide. At this point the currents continue into the space within the waveguide as *displacement currents*. Note also that at this point and time, $\vec{E}$ is zero but $\partial\vec{E}/\partial t$ is a maximum as the

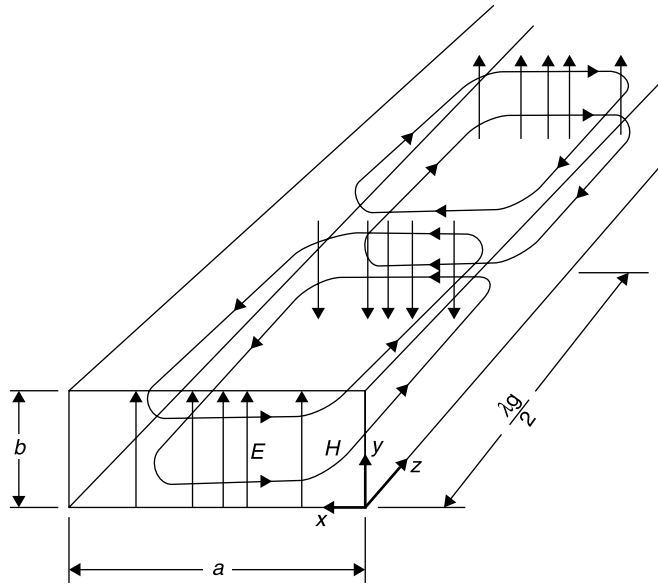

Figure 7.26-4 E and H fields of the TE_{10} mode in rectangular waveguide.

wave propagates. This is true for all sinusoids, namely that when the amplitude is zero, its derivative is maximum. A maximum $\partial\vec{E}/\partial t$ corresponds to a maximum displacement current.

Keep in mind that while the fields and currents are shown as discrete lines in Figures 7.26-4 and 7.26-5 for purposes of illustration, they actually are continuously variable sine functions.

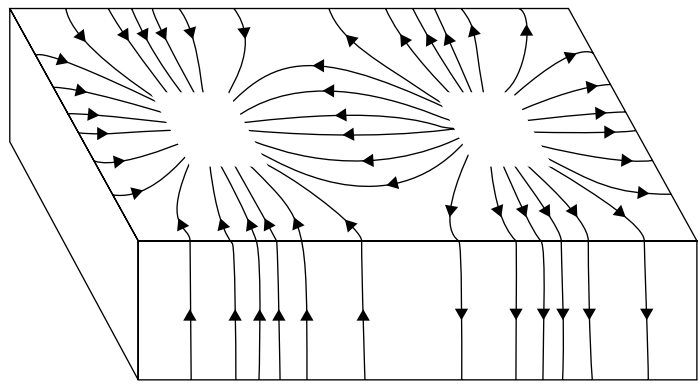

Figure 7.26-5 Current distribution on the inside walls of the waveguide for the TE_{10} mode. [After Ramo and Whinnery, 6, p. 370, with permission.]

7.27 FOURIER SERIES AND GREEN'S FUNCTIONS

Fourier Series

A function can be approximated over an interval by a constant value plus a collection of harmonically related sine and cosine terms:

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx \tag{7.27-1}$$

where

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx \tag{7.27-2}$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx \tag{7.27-3}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx \tag{7.27-4}$$

This equivalence is termed a *Fourier series* representation [2, Sec. 7.11], and the statement that the equivalent Fourier series can be found is called the *Fourier theorem*. By judicious choice of the origin $x = 0$, it may be found that the function has even symmetry, $f(x) = f(-x)$, in which case the series has only cosine terms. Or, if it has odd symmetry, $f(x) = -f(-x)$, the series has only sine terms. In general, the Fourier series only represents $f(x)$ over the interval $0 \le x \le L$ unless $f(x)$ is periodic with period L *for all* x, that is, $f(x) = f(x - L)$ for all x, in which case the series represents $f(x)$ for all x.

As an example, consider the function

$$f(t) = A \tag{7.27-5}$$

for

$$0 \le t \le T \tag{7.27-6}$$

where A is a constant (Fig. 7.27-1). This function has a Fourier series representation:

$$f(t) = \frac{4A}{\pi}\left(\sin\frac{\pi t}{T} + \frac{1}{3}\sin\frac{3\pi t}{T} + \frac{1}{5}\sin\frac{5\pi t}{T} + \frac{1}{7}\sin\frac{7\pi t}{T} + \cdots\right) \tag{7.27-7}$$

in which the independent variable, x, in (7.27-1) has been set equal to $\pi t/T$ in order to match the half period of the sine wave with the interval T over which

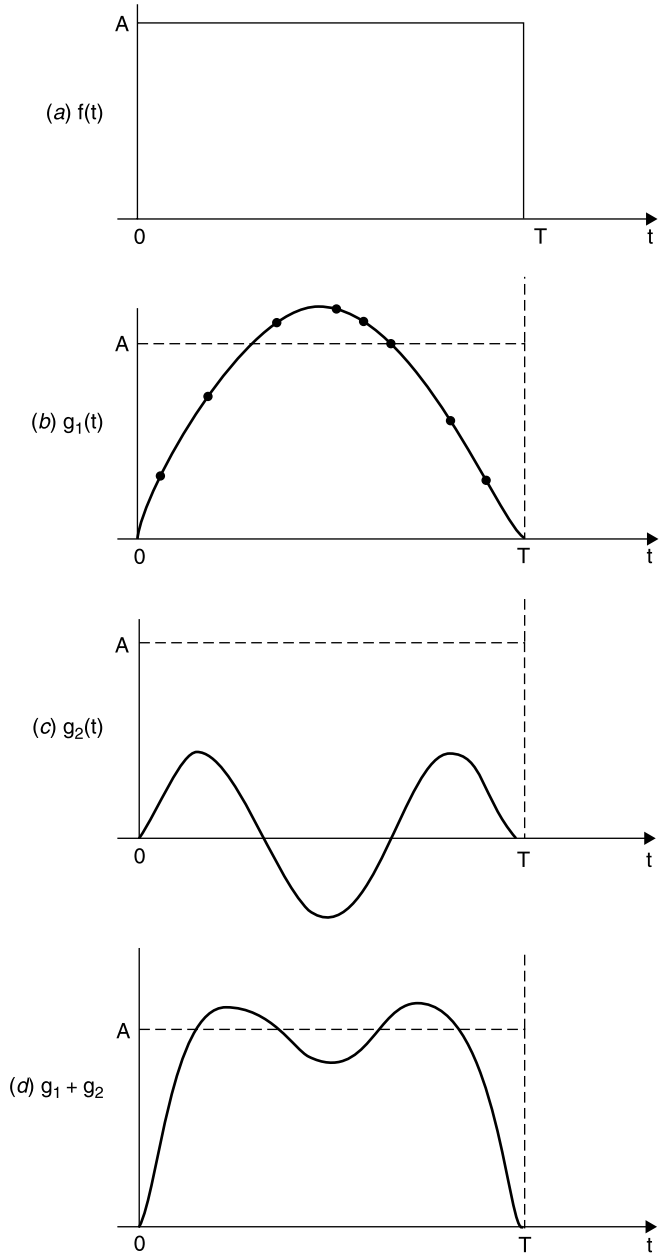

Figure 7.27-1 Representation of a rectangular waveform by the first two terms of a Fourier series.

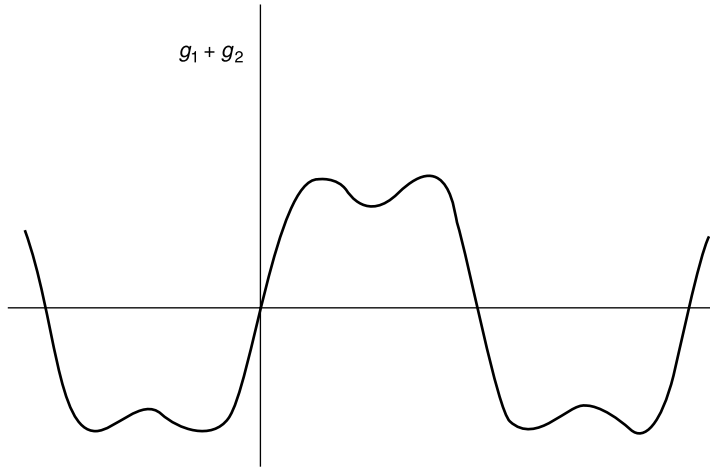

Figure 7.27-2 Periodicity of the function, $g_1 + g_2$, shown in Figure 7.27-1.

the function is defined. An approximation, $g_1 + g_2$, to the result in (7.27-7) using just the first two terms of the series is shown in Figure 7.27-1. The use of more terms would produce a better approximation. This can be seen in Figure 7.27-1. The first term, g_1 shown in Figure 7.27-1*b*, is not as good a representation of $f(t)$ as is the sum of the first two terms, $g_1 + g_2$ shown in Figure 7.27-1*d*. Notice that outside of the interval $0 \leq t \leq T$ the series representation is periodic with a period of $2T$ (Fig. 7.27-2), even though the function $f(t)$ may not be defined outside the range of $0 \leq t \leq T$. Also note that the function $g_1 + g_2$ shown in Figure 7.27-2 is an odd function, having been formed from sine terms.

Green's Functions

Green's functions perform a role similar to that of the Fourier series by describing the fields in waveguides and other structures having distinct allowed modes. Such distinct modes are called eigenmodes. The separate terms of the Green's function series are the allowed modes. With the Fourier series we were free to select frequency components that were periodic over any interval desired, although selecting the period of the fundamental sine wave can simplify the analysis. Since the Green's functions incorporate the eigenvalue mode solutions to the geometry in question, their periodicity is predetermined by the form of those modes. Also, Green's functions are written in terms of differential forms of excitation of the modes, such as a current filament.

The use of a Green's function is best demonstrated by an example. In this example the Green's function is used to relate the magnitude of the dominant TE_{10} mode in the waveguide to the current on a thin post installed between the waveguide's broad walls. To simplify the analysis we approximate the behavior

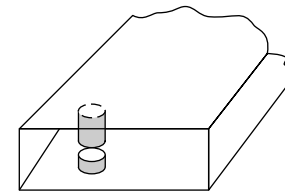

Figure 7.27-3 Thin post with a small gap into which a diode can be installed in rectangular waveguide.

of the post by a thin metal strip (Fig. 7.27-3). Often it is desired to connect a switching PIN diode or microwave source Gunn diode within a gap in the post, thereby coupling the diode to the power that propagates in the waveguide.

The following problem arises: since the waveguide has no uniquely defined characteristic impedance, it is not possible to calculate what impedance is presented to the diode when it is installed in series with the post. In this application example, we will use the Green function to relate the current on the post that is induced by an incident TE_{10} mode wave in the waveguide, and from this infer what absolute characteristic impedance Z_g should be defined for the waveguide. Given this Z_g definition, the load impedance presented to the diode can be calculated with the customary transmission line theory.

In the previous section we solved for the TE and TM modes in rectangular waveguides. These modes are the eigenvalue field solutions for the waveguide. Any electric field distribution in the rectangular waveguide must be representable as a series summation of either the TE_{mn} or the TM_{mn} modes or both since these are the only field patterns that are allowed by the waveguide's boundary conditions. For the rectangular waveguide, these fields are expressed by (7.26-44) and (7.26-45).

Green's functions are written in terms of the field modes that can exist in a region as functions of the primary excitations that can bring them into being, such as an electric field or current distribution in a waveguide. As an example, consider a filamentary unit current in a rectangular waveguide flowing in a vertical metal strip in the waveguide (Fig. 7.27-4).

Such a current filament excites an infinite number of TE_{m0} modes in the waveguide, and the resulting electric field E_y everywhere in the waveguide (at any x, z position) is related to the unit current filament at location (x', z') according to [9, Sec. 5.6]

$$E_y = G(x, z : x', z') = \frac{j\omega\mu_0}{a} \sum_{m=1}^{\infty} \frac{1}{\gamma_m} \sin \frac{m\pi x}{a} \sin \frac{m\pi x'}{a} e^{-\gamma_m |z - z'|} \quad (7.27\text{-}8)$$

where

$$\gamma_m = \sqrt{(m\pi/a)^2 - (2\pi/\lambda)^2} \quad (7.27\text{-}9)$$

and $G(x, z : x', z')$ is the Green's function for a uniform current filament at

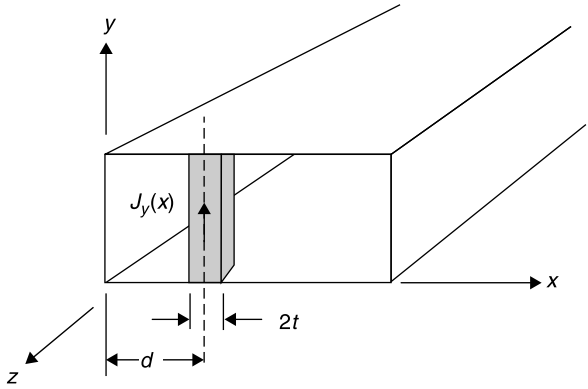

Figure 7.27-4 Vertical current filament serves as an excitation source for rectangular waveguide TE_{m0} modes.

(x', z'). It is an expression for all of the TE_{10} modes in rectangular waveguides that are excited by a unit current filament located at (x', z').

The time dependence, $e^{j\omega t}$, is assumed. Note that in (7.27-8), because of the magnitude sign in the exponential, the modes excited by the current filament propagate equally in both the $+z$ and $-z$ directions away from the filament. The expression for the peak value of E_y in (7.27-8) is for a unit current filament, having a magnitude of one ampere peak. The E_y response is linear; thus, if a 2 A current is applied, the value for E_y would be doubled. Term E_y is expressed in (7.27-8) as a summation of the TE_{m0} modes of the waveguide since these are the only modes excited by a vertical (y-directed) current filament. If there were more than one current filament, the total E_y at any location (x, z) is found by superimposing, or integrating, the current distribution of all such current filaments. Needless to say, this general formulation quickly could become too complex for solution; however, in the example to follow, we assume all of the current is located at one (x', z') location. Notice that in (7.27-8) the primed coordinates (x', z') are the locations of current filaments and the unprimed (x, z) are the locations at which the resulting E_y is evaluated.

A filamentary current in the waveguide is like a radiating antenna. A free space, vertically directed wire antenna, analyzed later in Section 7.32, can radiate power in any aximuthal direction. However, this "waveguide current antenna" located at (x', z') is constrained by the waveguide to propagate energy only in the $+z$ and $-z$ directions and only in the TE_{10} mode, since we assume that the waveguide is operated in the dominant mode frequency range. Note that a current filament having no y variation in the waveguide can only excite TE_{m0} modes, as (7.27-8) indicates.

In this example, the thin post (Fig. 7.27-3) is approximated by the strip geometry in Figure 7.27-4 because the strip exists at a single z' location. For thin posts this is a reasonable approximation. The procedure for determining Z_g is as follows [12]. Assume that there is a TE_{10} mode wave defined by (7.27-10)

and propagating in the $+z$ direction and incident on the conducting strip. The key to defining Z_g is to relate the magnitude of this current, I, on the strip to a given amount of power incident in the TE_{10} mode. The incident TE_{10} wave is described as

$$E_{IN} = E_0 \sin(\pi x/a)e^{-\gamma_1 z} \tag{7.27-10}$$

where a is the waveguide width and γ_1 is defined in (7.27-9). When this incident wave impinges on the strip, the tangential electric field on the strip must be zero since we assume it to be perfectly conducting. As a consequence, this incident wave induces a current distribution, $J_y(x)$, on the strip. The current distribution excites an infinite number of TE_{m0} modes, the sum of whose fields exactly cancel the incident TE_{10} electric field on the surface of the strip. All of the induced TE_{m0} modes so induced, except the TE_{10}, are evanescent since we assume that the waveguide is being operated in the dominant mode frequency range. The evanescent modes affect the energy storage in the vicinity of the strip obstacle, and accordingly its normalized reactance, jx, but for now only the scattered (reflected) TE_{10} dominant mode component need be evaluated. This scattered (reflected) dominant mode, E_R, is the propagating TE_{10} wave that is reflected by the strip. It is defined in (7.27-11) and evaluated using (7.27-8). For this example we assume that the waveguide is air filled and therefore $\varepsilon = \varepsilon_0$, $\mu = \mu_0$, and $\lambda = \lambda_0$. Then

$$E_R = \Gamma E_0 \sin(\pi x/a)e^{\gamma_1 z} \tag{7.27-11}$$

$$E_R = \frac{-j\omega\mu_0}{a\gamma_1}\int_{d-t}^{d+t} J_y(x)\sin(\pi x/a)\,dx \tag{7.27-12}$$

where $\Gamma = E_R/E_{IN}$ is the complex reflection coefficient of the strip to the incident TE_{10} wave. For the narrow strip assumed $J_y(x) \approx J_0 = \text{constant}$. Then $J_0 2t = I$, the peak current induced on the strip. Integrating (7.27-12) under the condition that $t \ll a$ gives

$$E_R = \frac{-j\omega\mu_0}{a\gamma_1}|I|\sin\left(\frac{\pi d}{a}\right) \tag{7.27-13}$$

And, since for a lossless waveguide operated in the dominant propagating mode

$$\gamma_1 = j\beta_1 = j\frac{2\pi}{\lambda_g} \tag{7.27-14}$$

where λ_g is the guide wavelength of the dominant mode, and

$$\omega\mu_0 = 2\pi f\mu_0 = 2\pi\mu_0\frac{c}{\lambda_0} = 2\pi\mu_0\frac{1}{\lambda_0\sqrt{\mu_0\varepsilon_0}} = 2\pi\sqrt{\frac{\mu_0}{\varepsilon_0}}\left(\frac{1}{\lambda_0}\right. \tag{7.27-15}$$

Using these equivalences (7.25-13) can be rewritten as

$$E_R = \frac{-|I|}{a}\sqrt{\frac{\mu_0}{\varepsilon_0}}\frac{\lambda_g}{\lambda_0}\sin\left(\frac{\pi d}{a}\right) \qquad (7.27\text{-}16)$$

Notice that E_R is negative, 180° out of phase with E_{IN}, as must be true, since the reflected wave's polarity must be opposite that of the incident wave in order to contribute, together with higher order excited modes, to the cancellation of the incident field on the surface of the conducting strip.

When Poynting's theorem is applied to the incident TE_{10} wave, its peak propagating power P is related to its peak amplitude E_0 by

$$E_0 = \sqrt{\frac{2P}{ab}\frac{\lambda_g}{\lambda_0}\sqrt{\frac{\mu_0}{\varepsilon_0}}} \qquad (7.27\text{-}17)$$

Since the reflection coefficient Γ is just the complex ratio E_R/E_I, (7.27-17) can be written

$$|E_R| = |\Gamma|\sqrt{\frac{2P}{ab}\frac{\lambda_g}{\lambda_0}\sqrt{\frac{\mu_0}{\varepsilon_0}}} \qquad (7.27\text{-}18)$$

Equating the expressions for E_R in (7.27-16) and (7.27-18), squaring, multiplying both sides of the result by 2, and simplifying gives

$$|I|^2\left[2\frac{b}{a}\sqrt{\frac{\mu_0}{\varepsilon_0}}\frac{\lambda_g}{\lambda_0}\sin^2\left(\frac{\pi d}{a}\right)\right] = 4P|\Gamma|^2 \qquad (7.27\text{-}19)$$

Alternatively, from circuit theory a reactance exposed to an incident peak power P and producing the reflection coefficient Γ when connected across a Z_g transmission line (Fig. 7.27-5) has a peak current I passing through it according to

$$|I|^2 Z_g = 4P|\Gamma|^2 \qquad (7.27\text{-}20)$$

Comparing (7.27-19) and (7.27-20), it follows that the appropriate value of waveguide impedance to produce the same current on the thin strip or post for a given incident power is

$$Z_g = 2\frac{b}{a}\sqrt{\frac{\mu_0}{\varepsilon_0}}\frac{\lambda_g}{\lambda_0}\sin^2\left(\frac{\pi d}{a}\right) \qquad (7.27\text{-}21)$$

The resulting equivalent circuit for a device placed within a gap, AB, in the waveguide post is that shown in Figure 7.27-6. Note that if the post is placed in the center of the waveguide

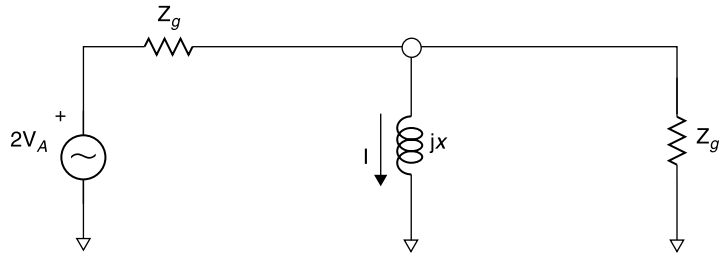

Figure 7.27-5 Equivalent circuit of the waveguide with conducting thin vertical strip (or post).

$$Z_g = 2\frac{b}{a}\sqrt{\frac{\mu_0}{\varepsilon_0}}\left(\frac{\lambda_g}{\lambda_0}\right) \tag{7.27-22}$$

It is interesting to note that this definition for Z_g, which is based on the post current and the TE_{10} mode incident power, is the same value that would be obtained were Z_g defined in terms of the TE_{10} voltage across the waveguide at the location of the post and the propagating power associated with the TE_{10} mode. In fact, (7.27-22) is the commonly defined *voltage-power waveguide impedance* definition. But this definition could not have been advanced a priori because, *with the post present, the voltage across the waveguide at the post location is zero*.

This application of the Green function, which required the evaluation of only one term, has produced a very useful analytical result, allowing the absolute equivalent circuit for the post in a waveguide of Figure 7.27-6. The absolute impedance of the post, $jX = jx(Z_g)$, can be found by determining x either by measurement of or from an analysis of the stored energy of the evanescent fields [13, Sec. 5.11]. The complete circuit, which usually includes a resonant cavity instead of matched load terminations, can then be analyzed using trans-

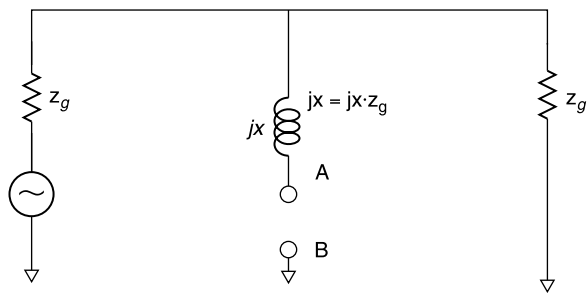

Figure 7.27-6 Equivalent circuit for a device location, AB, in the gap of a waveguide post with absolute impedance values.

mission line theory as described in Chapters 4 and 5. Post coupling of semiconductor devices to the waveguide is a common application, and this equivalent circuit allows calculation of the *absolute load impedance* seen at the device terminals AB.

7.28 HIGHER ORDER MODES IN CIRCUITS

Many engineers have little direct application for waveguides. Circuits, particularly integrated circuits, are small and high frequency interconnections between them usually are made with coaxial transmission lines. Nevertheless, understanding waveguides helps to diagnose problems in circuits that otherwise might seem unexplainable. For example, consider the common situation shown in Figure 7.28-1 in which a circuit board is installed in a metal box.

The rectangular metal box forms a waveguide at frequencies for which its width a is at least $\lambda_0/2$. If its length in the other lateral dimension is $\lambda_g/2$, it forms a resonator in the TE_{10} waveguide mode and has a high Q resonance at some frequency f_R. This is especially problematical if the circuit must operate or can oscillate (has gain) at f_R. Energy can couple to the resonator via the

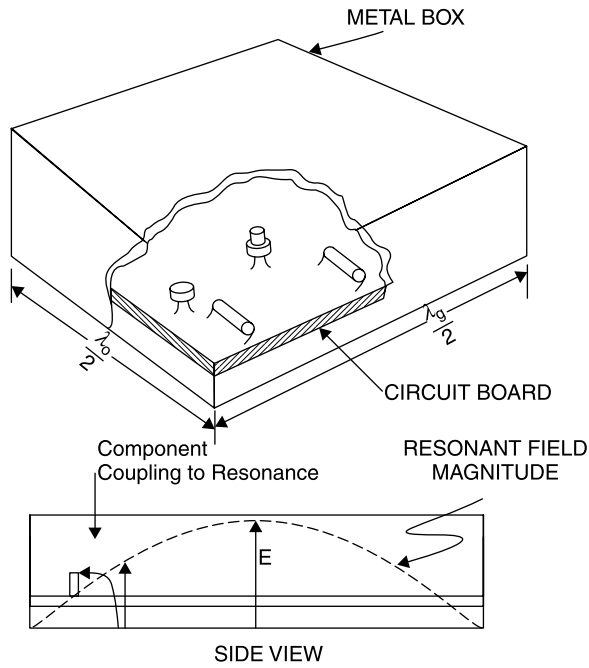

Figure 7.28-1 Circuit housed in a metal box and the TE_{10} waveguide mode resonance field magnitude (side view).

fringing electric field about a component, as shown in the side view of Figure 7.28-1, or through the magnetic field near a component. This coupling will result in the absorption of power from the circuit and dissipation at f_R since practical resonators are not lossless.

Recall from Chapter 3 that the higher the Q the lighter the coupling needed to approach or achieve critical coupling, at which any fraction up to and including all of the circuit's power could be absorbed. Alternatively the circuit may break into oscillation at f_R because the cavity can provide unintended feedback between the output and input circuits of a transistor. Such feedback need not be at the intended operational frequency of the circuit, it can occur at any frequency for which the circuit has gain, and the unintended coupling is appropriate for oscillation.

Without an understanding of the waveguide modes, this problem can be very puzzling and difficult to diagnose. The power absorption or oscillation can be quite sharply defined with frequency since waveguide cavities frequently have unloaded Qs of 1000 or more. Thus, the circuit may perform near perfectly except for a very narrow band of frequencies around f_R.

Moreover, if the box dimensions are considerably larger than $\lambda_0/2$, multiple resonant modes can exist in the box, each of which generally has a different resonant frequency. The simplest means of avoiding such resonances is to ensure that one or both lateral dimensions of the box are less than $\lambda_0/2$ at the highest frequency used in the circuit or at which the circuit has gain. This also assumes that the height of the box is also short compared to $\lambda_0/2$. When the lateral dimensions must be larger than $\lambda_0/2$, *mode suppressors* can be employed. These consist of conductive posts (screws or rivets work nicely) to interconnect the top and bottom covers (Fig. 7.28-2), limiting the free length dimensions in all directions inside the box to less than $\lambda_0/2$.

In this example, we have assumed a rectangular metal circuit enclosure, but higher order modes can exist in a round box structure or in any shape having internal dimensions that extend for a half wavelength in any two of the container's three dimensions.

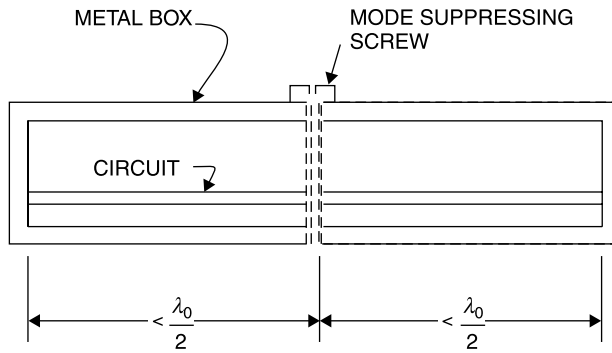

Figure 7.28-2 Use of a conductive post mode suppressor to prevent a box resonance.

7.29 VECTOR POTENTIAL

We demonstrated in Section 7.8 that the static $\vec{E}$ field could be derived from the *static scalar potential function*. This scalar potential was an intermediate calculation whose value was determined by the integration of a charge distribution over a volume. Once determined, the potential function facilitated determination of the $\vec{E}$ field using the gradient function, a derivative process. This process generally poses a simpler task than that of the vector integration that would otherwise be necessary to construct the $\vec{E}$ field directly from a charge distribution.

The scalar potential has obvious physical significance, as the ski slope mountain example demonstrated, and it greatly facilitates the determination of electric field by use of the gradient vector. The success of the scalar potential technique leads to the reasoning that another function might be defined to facilitate determination of the $\vec{B}$ field as well as the time-varying portion of the $\vec{E}$ field [6, Chapter 4]. It will be seen that this can be done by defining a *vector potential*. Unlike the scalar potential, the physical significance of the vector potential is not apparent. It is a mathematical artifice. Nevertheless, it does result in a useful method for routinizing the calculation of the $\vec{B}$ field.

Although Maxwell's equations eloquently describe the physics of electromagnetics, they are not the most convenient format for the solution of some real field situations, such as antennas. It will be seen from the following definition of the vector potential and its use in evaluating the short wire antenna in Section 7.32 that the use of the vector potential neatly derives the near and far fields of the antenna, even taking into account the propagation phase delay between the received signal and its current source.

The newly initiated to electromagnetics may feel overburdened with concepts already, and this feeling is not unjustified, given the complexity of vector mathematics and mastery of concepts necessary to reach this point. It is a little known fact that Maxwell's unified field theory originally was expressed in about 20 equations and included potential functions. Oliver Heaviside eliminated the potential functions and reduced the equations to the now well-known 4 equations [3]. Heaviside was a brilliant, self-taught mathematician and physicist. It was Heaviside who formulated the telegrapher equations that took distributed inductance of transmission lines into account for the first time. Earlier analyses of transmission lines only included resistance and capacitance. If he had not had such an acerbic personality, it is likely that we would be referring to Heaviside's equations instead of Maxwell's. It is therefore noteworthy that no less a theoretician than the brilliant Heaviside is said to have hated the notion of potential functions.

Indeed, few practicing microwave engineers are comfortable enough with the vector potential to employ it. Nevertheless, the vector potential is a concept not much more difficult to apply than the scalar potential already studied. Its format as a vector field has the same form as the distributed current density that produces the $\vec{B}$ field, so no new vector visualization is required.

We begin with a general definition of the vector potential as a three-dimensional vector field, $\vec{A}$, whose curl is equal to the $\vec{B}$ field. Since the divergence of the $\vec{B}$ field is zero in all cases, the divergence of $\vec{A}$ can be defined subsequently in a manner that relates $\vec{A}$ to the time-varying $\vec{E}$ field. In this way the suitably defined $\vec{A}$ allows the determination of both the $\vec{E}$ and $\vec{B}$ fields using only differential processes, not integrals. To this end we define $\vec{A}$ to satisfy

$$\vec{B} \equiv \nabla \times \vec{A} \tag{7.29-1}$$

Then applying Maxwell's third equation, $\nabla \times \vec{E} = -\partial \vec{B}/\partial t$, we can rewrite (7.29-1) as

$$\nabla \times \vec{E} = -\frac{\partial(\nabla \times \vec{A})}{\partial t}$$

$$\nabla \times \left[\vec{E} + \frac{\partial \vec{A}}{\partial t}\right] \equiv 0 \tag{7.29-2}$$

since the time derivative and curl can be taken in any order. The curl of the quantity in brackets is zero, therefore the quantity can be defined as the gradient of a scalar function, Φ, more precisely, the negative of the gradient in order to be consistent with the earlier definition of the scalar potential Φ in Section 7.8. Thus,

$$\vec{E} + \frac{\partial \vec{A}}{\partial t} \equiv -\nabla \Phi \tag{7.29-3}$$

Equivalently,

$$\vec{E} = -\nabla \Phi - \frac{\partial \vec{A}}{\partial t} \tag{7.29-4}$$

The *dynamic scalar potential* Φ defined in (7.29-3) is similar to the static potential function, but they are not the same. The static scalar potential applies to fixed charges. In this section Φ is a *dynamic scalar potential* and applies to moving charges. Generally, no confusion results between the two cases since a given application usually involves either static or moving charge. For example, in propagation situations the presence of fixed charge is of no consequence since the DC fields so created do not propagate. Furthermore, in the event that both static and dynamic charge distributions are of interest, their resulting fields can be added together by superposition if the medium is linear.

Although no assumptions have been made about the medium thus far, potential functions are most useful in linear, isotropic (same behavior in all directions), and homogeneous (uniformity throughout space) media. Accord-

ingly, in using potential functions we assume that μ and ε are scalars of constant value throughout the volume of analysis.

With (7.29-1) and (7.29-4) the $\vec{B}$ and $\vec{E}$ fields are completely specified. Next, let us find expressions for Φ and $\vec{A}$ that will allow their determination from their primary sources, moving charges, or currents. To this end substitute (7.29-4) into Maxwell's first equation, $\nabla \cdot \vec{E} = \rho/\varepsilon$, to obtain

$$-\nabla^2\Phi - \frac{\partial}{\partial t}(\nabla \cdot \vec{A}) = \frac{\rho}{\varepsilon} \tag{7.29-5}$$

Applying Maxwell's fourth equation, $\nabla \times \vec{H} = \overrightarrow{J_C} + \partial\vec{D}/\partial t$, and noting that $\nabla \times \vec{A} = \mu\vec{H}$,

$$\nabla \times \nabla \times \vec{A} = \mu\overrightarrow{J_C} + \mu\varepsilon\left[-\nabla\left(\frac{\partial\Phi}{\partial t}\right) - \left[\frac{\partial^2\vec{A}}{\partial t^2}\right]\right(\tag{7.29-6}$$

Applying the vector identity $\nabla \times \nabla \times \vec{A} = \nabla(\nabla \cdot \vec{A}) - \nabla^2\vec{A}$ to (7.29-6) gives

$$\nabla(\nabla \cdot \vec{A}) - \nabla^2\vec{A} = \mu\overrightarrow{J_C} - \mu\varepsilon\nabla\left(\frac{\partial\Phi}{\partial t}\right) - \mu\varepsilon\frac{\partial^2\vec{A}}{\partial t^2} \tag{7.29-7}$$

We have not used the option of defining the divergence of $\vec{A}$, an action that will serve to define $\vec{A}$ uniquely. Accordingly, we define the divergence to satisfy

$$\nabla \cdot \vec{A} \equiv -\mu\varepsilon\frac{\partial\Phi}{\partial t} \tag{7.29-8}$$

This allows (7.29-5) and (7.29-7) to be written as

$$\nabla^2\Phi - \mu\varepsilon\frac{\partial^2\Phi}{\partial t^2} = -\frac{\rho}{\varepsilon} \tag{7.29-9}$$

$$\nabla^2\vec{A} - \mu\varepsilon\frac{\partial^2\vec{A}}{\partial t^2} = -\mu\vec{J}_C \tag{7.29-10}$$

The expressions of (7.29-9) and (7.29-10) are the basis for the integral expressions by which Φ and $\vec{A}$ are evaluated. Once Φ and $\vec{A}$ are evaluated, $\vec{B}$ and $\vec{E}$ can be derived from the scalar and vector potentials using derivative processes, the vector curl being a derivative operation. These results are summarized in (7.29-11) and (7.29-12):

$$\vec{B} = \mu\vec{H} = \nabla \times \vec{A} \tag{7.29-11}$$

$$\vec{E} = -\nabla\Phi - \frac{\partial\vec{A}}{\partial t} \tag{7.29-12}$$

The dynamic scalar potential is evaluated using an integral expression similar to that employed with the static scalar potential, namely

$$\Phi = \int\!\!\int_V \frac{\rho \, dV}{4\pi\varepsilon r} \tag{7.29-13}$$

Using similar reasoning, the vector potential *for the static or steady-state cases* can be evaluated using a volume integral of its current density sources:

$$\vec{A} = \mu \int\!\!\int_V \frac{\overrightarrow{J_C} \, dV}{4\pi r} \tag{7.29-14}$$

In both (7.29-1) and (7.29-2) the volume V includes all of the respective mobile charge density ρ and current density $\overrightarrow{J_C}$. Even so, these two equations can only be used for steady-state not transient time variation cases because they do not recognize the propagation delay between the sources and the points at which the potentials are evaluated. This limitation is accommodated using the *retarded potentials* discussed in the next section.

7.30 RETARDED POTENTIALS

The corresponding solutions for Φ and $\vec{A}$ in *transient time-varying cases* must take propagation delays into account. This is done by replacing (7.29-13) and (7.29-14) with (7.30-1) and (7.30-2), respectively [6, p. 199]:

$$\Phi = \int\!\!\int_V \frac{[\rho]_{t-(r/v)} \, dV}{4\pi\varepsilon r} \tag{7.30-1}$$

$$\vec{A} = \mu \int\!\!\int_V \frac{[\overrightarrow{J_C}]_{t-(r/v)} \, dV}{4\pi r} \tag{7.30-2}$$

in which $v = 1/\sqrt{\mu\varepsilon}$; $v = c \approx 3 \times 10^8$ m/s in free space. For example, in (7.30-1) the quantity Φ to be evaluated at point P that is a distance r from charge ρ is to use the value of ρ present at time $t - (r/v)$. In this way the time necessary for the effect of ρ to propagate to P is taken into account. The effect travels at velocity $v = 1/\sqrt{\mu\varepsilon}$, the velocity of the plane wave evaluated in 7.22. The same provision for propagation is made in the evaluation of $\vec{A}$ using (7.30-2).

Because of this retardation effect, the potentials evaluated by (7.30-1) and (7.30-2) are called *retarded potentials* because their arrival is retarded by the time of propagation.

7.31 POTENTIAL FUNCTIONS IN THE SINUSOIDAL CASE

For the case of sinusoidal excitations, one is usually interested in the steady-state values of the fields. Therefore, the propagation time from the source to a point a distance r away can be fully accounted for by a phase delay kr, where $k = \omega/v$ is the propagation constant. Given this assumption, the following analysis shows that the expression for $\vec{\mathrm{E}}$ in (7.29-12) can be written exclusively in terms of $\vec{\mathrm{A}}$. To derive this new expression take the gradient of both sides of (7.29-8) to obtain

$$\nabla(\nabla \cdot \vec{A}) = -\mu\varepsilon \frac{\partial}{\partial t} \nabla\Phi \tag{7.31-1}$$

Since time differentiation and the gradient can be taken in any order, and since $\partial/\partial t = j\omega$ for sinusoidal excitations,

$$\nabla\Phi = -\frac{\nabla(\nabla \cdot \vec{A})}{j\omega\mu\varepsilon} = \frac{j\omega}{k^2} \nabla(\nabla \cdot \vec{A}) \tag{7.31-2}$$

where $k = \omega\sqrt{\mu\varepsilon}$ is the propagation constant in the medium. Then, substituting (7.31-2) into (7.29-12), the expressions for $\vec{B}$ and $\vec{E}$ for the sinusoidal case become

$$\vec{B} = \nabla \times \vec{A} \tag{7.31-3}$$

$$\vec{E} = -j\omega \left[\frac{1}{k^2} \nabla(\nabla \cdot \vec{A}) + \vec{A} \right] \tag{7.31-4}$$

where $\vec{A}, \vec{B}$, and $\vec{E}$ are in phasor form. This result is very useful for antenna pattern analysis, as will be seen in the next section.

7.32 ANTENNAS

Short Straight Wire Antenna

Consider a short wire antenna that is excited by a sinusoidal current I_0. The current excites the base of the antenna creating a surrounding $\vec{H}$ field whose time rate of change induces an orthogonal $\vec{E}$ field. The pair of sinusoidally varying fields store energy as well as propagate it.

The simple geometry in Figure 7.32-1 belies the complexity of the resulting electromagnetic field analysis of the short wire antenna, as will be seen from the following analysis. However, the approach is straightforward, and its execution provides an example of the use of the vector potential $\vec{A}$ and the determination of $\vec{E}$ and $\vec{H}$ fields at an arbitrary distance from an antenna [6, Sec. 12.05].

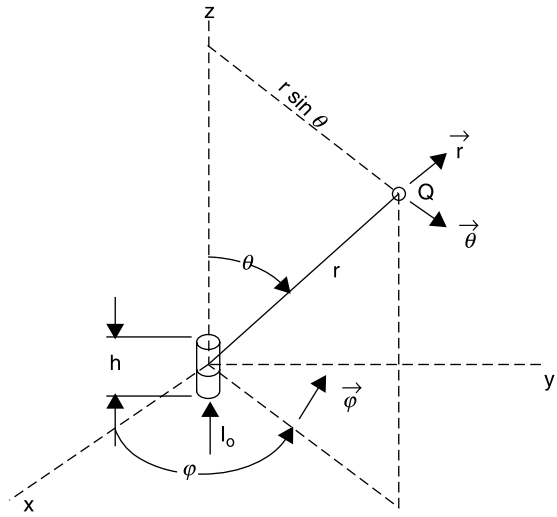

Figure 7.32-1 Geometry for analysis of the short straight wire antenna. [6, pg 497 Reprinted with permission]

The procedural convenience of using the vector potential can be appreciated in this analysis of the short straight wire antenna excited by a sinusoidal current phasor $\vec{I}_0$, for which the factor $e^{j\omega t}$ is implicit. The expression (7.30-2) for evaluating $\vec{A}$ indicates that each differential current contributes a part of $\vec{A}$ that is in the same direction as that current element. In other words, in this example of the short wire, since $\vec{I}_0$ points only in the z direction, so must $\vec{A}$. The fact that the vector potential has the same vector form as the currents that excite $\vec{E}$ and $\vec{H}$ fields underlies its usefulness.

For this analysis we assume that the length h and the diameter of the short straight wire antenna are very small compared to the operating wavelength, hence there is negligible change of phase in $\vec{I}_0$ over its length and diameter. The integration for $\vec{A}$ produces, for any point Q at radius r from the antenna, gives

$$\vec{A} = \vec{z}A_z \tag{7.32-1}$$

$$A_z = \mu \frac{hI_0}{4\pi r} e^{-j(\omega r/v)} \tag{7.32-2}$$

To evaluate $\vec{A}$ as a function of the radial distance r from the current source, it is best to express it in spherical coordinates as

$$A_r = A_z \cos\theta = \mu \frac{I_0 h}{4\pi r} e^{-jkr} \cos\theta \tag{7.32-3}$$

$$A_\theta = -A_z \sin\theta = -\mu \frac{I_0 h}{4\pi r} e^{-jkr} \sin\theta \tag{7.32-4}$$

where A_r and A_θ are the components of A_z in the $\vec{r}$ and $\vec{\theta}$ directions, respectively, and the medium propagation constant k is given by

$$k = \frac{\omega}{v} = \omega\sqrt{\mu\varepsilon} = \frac{2\pi}{\lambda} \tag{7.32-5}$$

Notice that, due to the symmetry of this case, all $\partial/\partial\varphi \equiv 0$, and there is no A_φ component. The utility of the vector potential can now be appreciated by observing that the integration needed to evaluate it results in a vector having the same direction as the current that produces it. To evaluate $\vec{H}$, note that when the curl of $\vec{A}$ is evaluated for this case in spherical coordinates only the φ component is nonzero. Thus, using (7.29-1)

$$\vec{H} = \frac{1}{\mu}\nabla \times \vec{A}$$

$$\vec{H} = \vec{H_\varphi} = \frac{1}{\mu}\frac{\vec{\varphi}}{r}\left[\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right] = \frac{1}{\mu}\frac{\vec{\varphi}}{r}\frac{\mu h I_0}{4\pi}\left[(-jk)e^{-jkr}\sin\theta + \frac{1}{r}e^{-jkr}\sin\theta\right]$$

$$\vec{H_\varphi} = \vec{\varphi}\frac{hI_0}{4\pi}e^{-jkr}\left[\frac{jk}{r} + \frac{1}{r^2}\right]\sin\theta \tag{7.32-6}$$

Next find $\vec{E}$ using (7.31-4) which is repeated below:

$$\vec{E} = -j\omega\left[\frac{1}{k^2}\nabla(\nabla\cdot\vec{A}) + \vec{A}\right] \tag{7.32-7}$$

To evaluate $\vec{E}$, first calculate $\nabla\cdot\vec{A}$, noting that $A_\varphi = 0$:

$$\nabla\cdot\vec{A} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2A_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta A_\theta)$$

$$\nabla\cdot\vec{A} = \frac{\mu I_0 h}{4\pi}\left[\frac{\cos\theta}{r^2}\frac{\partial}{\partial r}(re^{-jkr}) - \frac{e^{-jkr}}{r^2\sin\theta}\frac{\partial}{\partial\theta}(\sin^2\theta)\right] \tag{7.32-8}$$

$$\nabla\cdot\vec{A} = \frac{-\mu I_0 h}{4\pi}\left[\frac{jk}{r} + \frac{1}{r^2}\right]e^{-jkr}\cos\theta$$

Since $\nabla\cdot\vec{A}$ is not a function of φ,

$$\frac{1}{k^2}\nabla(\nabla\cdot\vec{A}) = \frac{1}{k^2}\left\{\vec{r}\frac{\partial(\nabla\cdot\vec{A})}{\partial r} + \vec{\theta}\frac{1}{r}\frac{\partial(\nabla\cdot\vec{A})}{\partial\theta}\right\}$$

$$= \frac{-\mu I_0 h}{k^2 4\pi}\left\{\vec{r}\cos\theta\frac{\partial}{\partial r}\left[\frac{jke^{-jkr}}{r} + \frac{e^{-jkr}}{r^2}\right]\right.$$

$$\left. + \vec{\theta}\frac{e^{-jkr}}{r}\left[\frac{jk}{r} + \frac{1}{r^2}\right]\frac{\partial(\cos\theta)}{\partial\theta}\right\} \tag{7.32-9}$$

Performing the indicated steps and simplifying gives

$$\frac{1}{k^2}\nabla(\nabla\cdot\vec{A}) = \frac{-\mu I_0 h}{k^2 4\pi}\left\{\vec{r}\cos\theta e^{-jkr}\left[\frac{k^2}{r} - \frac{2jk}{r^2} - \frac{2}{r^3}\right] - \vec{\theta}e^{-jkr}\left[\frac{jk}{r^2} + \frac{1}{r^3}\right]\sin\theta\right\} \tag{7.32-10}$$

Next substitute (7.32-10), (7.32-3), and (7.32-4) into (7.32-7) to get (7.32-11), (7.32-12), and (7.32-13):

$$\begin{aligned} E_r &= j\omega\frac{\mu I_0 h}{k^2 4\pi}e^{-jkr}\left\{\cos\theta\left[\frac{k^2}{r} - \frac{2jk}{r^2} - \frac{2}{r^3}\right] - \frac{k^2\cos\theta}{r}\right\} \\ &= -j\omega\frac{\mu I_0 h}{k^2 4\pi}e^{-jkr}\left[\frac{2jk}{r^2} + \frac{2}{r^3}\right]\cos\theta \end{aligned} \tag{7.32-11}$$

Recognizing that $k^2 = \omega^2\mu\varepsilon$ and $\eta = \sqrt{\mu/\varepsilon}$, (7.32-11) can be written

$$E_r = \frac{I_0 h}{4\pi}e^{-jkr}\left(\frac{2\eta}{r^2} - \frac{2j}{\omega\varepsilon r^3}\right)\cos\theta \tag{7.32-12}$$

Similarly,

$$E_\theta = \frac{I_0 h}{4\pi}e^{-jkr}\left[\frac{\eta}{r^2} + \frac{j\omega\mu}{r} + \frac{-j}{\omega\varepsilon r^3}\right]\sin\theta \tag{7.32-13}$$

Again noting that $\vec{A}$ has no φ component and no φ variation, the $\vec{H}$ field is obtained as

$$\begin{aligned} \vec{H} &= \frac{1}{\mu}(\nabla\times\vec{A}) = \frac{1}{\mu}\vec{\varphi}\left[\frac{1}{r}\frac{\partial}{\partial r}(rA_\theta) - \frac{1}{r}\frac{\partial A_r}{\partial\theta}\right] \\ H_\varphi &= \frac{I_0 h}{4\pi}e^{-jkr}\left[\frac{jk}{r} + \frac{1}{r^2}\right]\sin\theta \end{aligned} \tag{7.32-14}$$

Certainly, the development of expressions (7.32-12), (7.32-13), and (7.32-14), while straightforward, is mathematically tedious. But once expressed, they reveal the field behavior of the short antenna quite handily. First, the terms in $1/r^2$ and $1/r^3$ fall off rapidly for large r, leaving the $1/r$ terms.

The $1/r$ terms of E_θ and H_φ are in phase with and orthogonal to each other, hence they give peak propagating power density equal to Poynting's vector $\vec{P}_D$. Their ratio, $E_\theta/H_\varphi = \eta$, is the same as that for the plane propagating wave analyzed in Section 7.18. Taking their directions into account, $\vec{\theta}\times\vec{\varphi} = \vec{r}$, the power propagates in the $\vec{r}$ direction, away from the antenna. The peak propagating power density is given by

$$\vec{P}_D = \vec{E}\times\vec{H}^* = \vec{r}E_\theta H_\varphi = \vec{r}\eta|H_\varphi|^2 \text{ watts/meter}^2 \tag{7.32-15}$$

where $\eta \approx 377\ \Omega$ for free space. Since peak values are used for the $\vec{E}$ and $\vec{H}$ fields, the time-average power flow is equal to one half the peak value of $\vec{P}_D$, or

$$\vec{P}_D(r) = \vec{r}\frac{1}{2}\eta\left(\frac{kI_0h}{4\pi r}\sin\theta\right)^2 \text{ watts/meter}^2 \tag{7.32-16}$$

where h and r are in meters and k is in $(\text{meters})^{-1}$. The total propagating power across a spherical surface of radius r is

$$P = \int_S \vec{P}_D \cdot \overrightarrow{dS} = \int_0^{2\pi}\int_0^{\pi} P_D r^2 \sin\theta\, d\theta\, d\varphi$$

$$P = \pi\eta\left(\frac{kI_0h}{4\pi}\right)^2 \int_0^{\pi} \sin^3\theta\, d\theta \tag{7.32-17}$$

Since $\int_0^{\pi} \sin^3\theta\, d\theta = \frac{4}{3}$, and for free space $\eta \approx 120\pi\ \Omega$, and $k = 2\pi/\lambda_0$, (17) reduces to

$$P = 40\pi^2 I_0^2\left(\frac{h}{\lambda_0}\right)^2 \text{ watts} \tag{7.32-18}$$

which is the total power radiated out of a sphere of radius r. Note that while P is independent of r, the power density P_D is inversely proportional to r^2.

Radiation Resistance

The radiation resistance of an antenna is defined as the resistance that would dissipate the same power as is radiated by the antenna when supplied with the same current. For the short, straight wire antenna radiating into free space

$$\frac{1}{2}I_0^2 R_R = 40\pi^2 I_0^2\left(\frac{h}{\lambda_0}\right)^2$$

$$R_R = 80\pi^2\left(\frac{h}{\lambda_0}\right)^2 \text{ ohms} \tag{7.32-19}$$

For example, a 1-in.-long antenna at 1 GHz would have a radiation resistance of

$$R_R = 80\pi^2\left(\frac{1 \text{ in.}}{11.8 \text{ in.}}\right)^2 \Omega = 5.7\ \Omega \tag{7.32-20}$$

By itself, 5.7 Ω would be a large mismatch on a 50-Ω feed line; however, the mismatch is exacerbated further by the fact that there is a large reactance in series with the radiation resistance. Calculation of the reactance is more diffi-

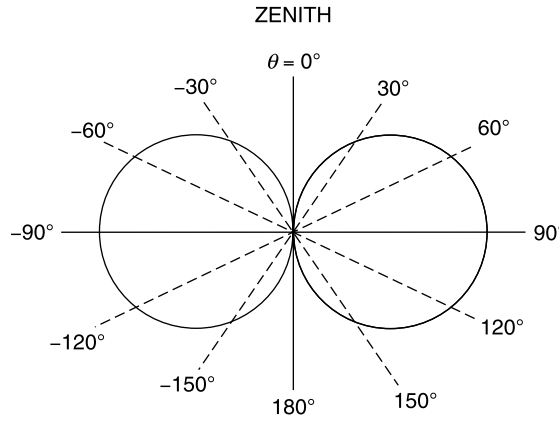

Figure 7.32-2 Radiation pattern of a dipole antenna.

cult because it is a volume integral over all space. One cannot use the integral limits $r = 0, \infty$ because the fields have a singularity at $r = 0$. Rather the integration must extend from the actual surface of the antenna to $r = \infty$. While radiation resistance is not a sensitive function of the diameter of the wire antenna, the reactance is very sensitive to the diameter because very thin wires have much stronger surrounding fields.

Radiation Pattern

The short wire antenna just analyzed (Fig. 7.32-1) is a dipole because the "horizon" plane bisects its vertical axis. Since, by assumption, there is no ground plane nearby, it radiates equally above and below the horizontal plane ($\theta = 90°$), and the $\vec{E}_\theta$ and $\vec{H}_\varphi$ fields both vary as $\sin\theta$. Thus, the radiated power is maximum in the horizontal plane and is zero in the vertical direction, or zenith, as shown in Figure 7.32-2.

The radiation is maximum at right angles to the wire antenna in the azimuth direction (on the horizon). The pattern is circularly symmetric about the antenna's axis. The pattern shown in Figure 7.32-2 is an *elevation plane cut*. This pattern applies when the antenna is located well above the earth, which would serve as a *ground plane*. In three dimensions the pattern resembles a bagel with a very small center hole (Fig. 7.32-3).

Half-Wavelength Dipole

Since the radiation resistance for a short monopole is very low, most wire antennas are made longer, but then the "short" approximation no longer holds, and a new analysis must be made, this time assuming a current distribution

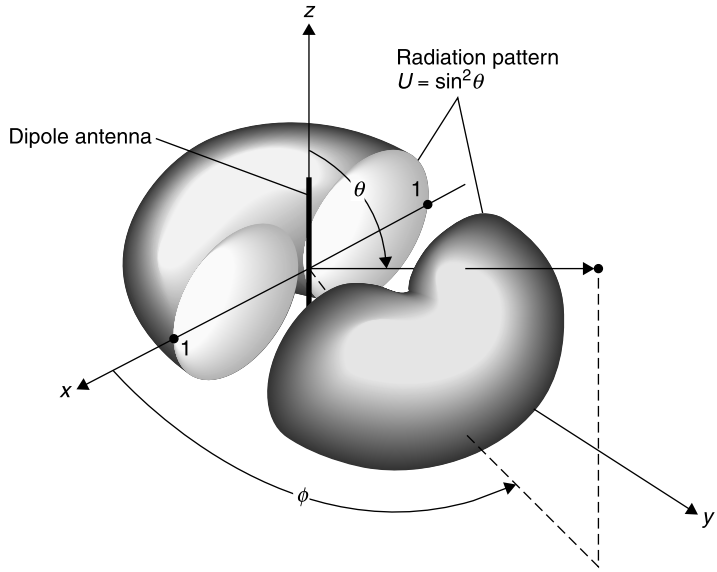

Figure 7.32-3 Sketch of three-dimensional radiation pattern of short wire dipole antenna [after Balanis, 15, Fig. 4.3j, with permission].

which is sinusoidal and goes to zero at the open ends. For the half-wave dipole, for which $l = \lambda/4$ (Fig. 7.32-4), the electric field $\vec{E}_\theta$ is found by integrating the field contributions from infinitesimal lengths of straight antennas of the type just covered and assuming a sinusoidal current distribution with the maximum I_m at the dipole's center feed point. The result of this analysis [6, Sec. 12.7] is that for the center-fed half-wavelength dipole antenna,

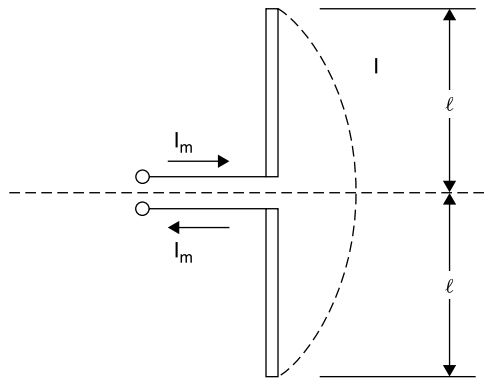

Figure 7.32-4 Center-fed dipole antenna.

$$|E_\theta| = \frac{60I_m}{r}\left[\frac{\cos[(\pi/2)\cos\theta]}{\sin\theta}\right] \text{volts/meter} \tag{7.32-21}$$

$$P_D(r) = \frac{15I_m^2}{\pi r^2}\left[\frac{\cos[(\pi/2)\cos\theta]}{\sin\theta}\right]^2 \text{watts/meter}^2 \tag{7.32-22}$$

$$R_R = 73.09\ \Omega \tag{7.32-23}$$

where I_m is the maximum current, presumed equal to the value at the center feed point. The driving point impedance is nearly equal to the radiation resistance of 73 Ω since the antenna is nearly resonant when its length is $\lambda/2$. Interestingly, the half-wavelength dipole and the short wire antenna have nearly the same radiation patterns [6, Secs. 12.05 to 12.07]. However, the short wire antenna is much more difficult to drive because of its low impedance and high reactance.

A practical problem arises when it is necessary to interconnect the half-wave dipole antenna to a transmitter whose output port is an unbalanced transmission connector or transmission line, such as coax. The antenna requires a balanced feed at terminals *AB* in Figure 7.32-5. That is, both of the *AB* terminals are to be separated from ground by the same impedance. A coaxial output connector or a coaxial cable used to interconnect transmitter and antenna is an unbalanced transmission system. However, this problem is easily corrected by using a *balun*. On transmission, the balun converts the unbalanced line from the transmitter to the balanced format of the antenna, as shown in Figure 7.32-5. It does this by creating an infinite impedance to ground at point *C* by virtue of the quarter wavelength collar attached to the end of the coaxial line. The balun's operation is reciprocal. On receiving a signal, the balun converts the balanced format of the antenna to the unbalanced format of the receiver. The function of the balun is limited in bandwidth by the quarter wavelength requirement of the collar.

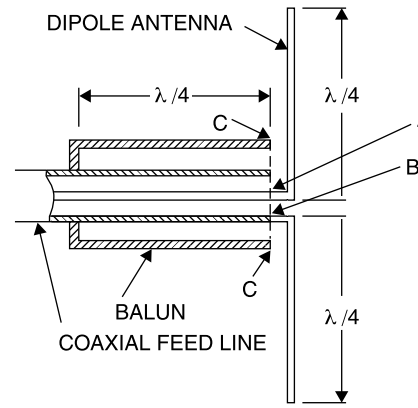

Figure 7.32-5 Balun used to connect a coaxial cable to a balanced, half-wavelength dipole antenna.

Antenna Gain

For a given radiated power an antenna that radiates more power in certain directions and less in others, as all practical antennas do, can place more power at a receiver than would an *isotropic radiator* that distributes the same input power uniformly over a spherical surface. Actually, it is impossible to build a perfectly isotropic radiator, but the concept is useful for defining the performance of practical antennas.

The power density delivered in a given direction by a practical antenna compared to what would be delivered by an isotropic radiator is called *antenna gain*. It is expressed in dBi, for decibels above (or below, if negative) the level that would be achieved with an isotropic radiator. Gain can be defined in any direction, but customarily, unless otherwise noted, it is defined for the direction in which the antenna radiates the greatest power density. For the short wire antenna the maximum radiated power density occurs on the horizon ($\sin^2 \theta = 1$) and has the value

$$P_{D,\max} = \frac{1}{2}\eta|H_\varphi|^2 = \frac{1}{2}\eta\left(\frac{kI_0h}{4\pi r}\right)^2$$

where H_φ is the radiated component of the H field at r. Then, since $k = 2\pi/\lambda_0$ and $\eta \approx 120\pi\ \Omega$, this reduces to

$$P_{D,\max} \approx 15\pi\left(\frac{h}{\lambda_0}\right)^2\frac{I_0^2}{r^2} \tag{7.32-24}$$

Since the surface area of a sphere at radius r is equal to $4\pi r^2$, an isotropic radiator produces a power density at radius r of

$$P_{D,i} = \frac{1}{4\pi r^2}\frac{I_0^2}{2}R_R = \frac{1}{8\pi r^2}I_0^2 80\pi^2\left(\frac{h}{\lambda_0}\right)^2 = 10\pi\frac{I_0^2}{r^2}\left(\frac{h}{\lambda_0}\right)^2 \tag{7.32-25}$$

Dividing (7.32-24) by (7.32-25), the gain of the short dipole antenna is [6, p. 504]

$$G_{\text{short dipole}} = \frac{P_{D,\max}}{P_{D,i}} = 1.5 = 1.76\ \text{dB} \tag{7.32-26}$$

Using the same method, the gain of the half-wavelength dipole antenna is

$$G_{\text{half wavelength dipole}} = 1.64 = 2.14\ \text{dB} \tag{7.32-27}$$

It should be noted, however, that although the short dipole and the half-wavelength dipole have nearly identical radiation patterns and gains, the half-

wavelength dipole is much easier to excite due to its higher radiation resistance and its near zero reactance.

Antenna Effective Area

In calculating the overall loss between transmitter and receiver, we need to know the *effective area* A_E of the receiving antenna. Aperture antennas such as dishes and waveguide horns have an effective signal capture area that is approximately equal to their physical area in the direction of the received or transmitted signal, but even the small wire antenna, whose physical area is infinitesimal, has an appreciable effective area. This effective area is related to its gain.

The fact that receiving antennas have varying gains and hence capture relatively more or less power is treated by considering that the power they capture from the radiation field is proportional to their respective effective areas [6, p. 560, and 14, Sec. 11.12]. But the received power is also proportional to the gain of the receiving antenna. Therefore, it follows that there is a proportionality constant relating gain G to effective area A_E. *Once this factor is found for any one receiving antenna, the same factor can be applied to all receiving antennas.* Furthermore, the network between transmitter and receiver terminals, including the free-space propagation medium, is linear and reciprocal. Accordingly, what applies to all transmitting antennas, by reciprocity, applies to all receiving antennas.

To determine the proportionality factor between gain and effective area, we use as a receiving antenna the short wire dipole previously evaluated. We define that the power delivered to the receiver terminals by the receiving antenna is given by

$$P_{\text{rec}} \equiv P_D A_E \tag{7.32-28}$$

where P_{rec} is the power delivered to the receiver connected to the receiving antenna, P_D is the power density (in watts/meter^2) of the wave incident on the receiving antenna and A_E is the effective area of the receiving antenna.

In the vicinity of the receiving antenna, P_D is related to the rms electric field strength E parallel to the antenna by

$$P_D = \frac{E^2}{\eta} = \frac{E^2}{120\pi} \tag{7.32-29}$$

where η is the impedance of free space. The available power absorbed in a matched load connected to the receiving antenna is given by

$$P_{\text{rec}} = \frac{V_{\text{OC}}^2}{4R_R} = \frac{(Eh)^2}{4R_R} \tag{7.32-30}$$

where R_R is the radiation resistance of the receiving antenna given by (7.32-19).

It might seem that the factor of 4 should not appear in the denominator in (7.32-30), that instead the full voltage Eh delivers power to the resistance R_R; but this is not so. The transmitter, its antenna, the free-space propagation path,and the receiving antenna taken together represent a linear network. The voltage Eh induced at the receiving antenna terminals is an "open-circuit voltage" *that exists in the absence of the receiving antenna or when the receiving antenna terminals are open circuited*, that is, when the incident radiation is not "loaded." When the R_R matched load is applied, the voltage at the receiving antenna terminals drops to half its open-circuit value, and the power delivered to the load at the receiver terminals is that given by (7.32-30). Substituting the value for R_R from (7.32-19) into (7.32-30) gives

$$P_{\text{rec}} = \frac{(Eh)^2}{4} \frac{\lambda_0^2}{80\pi^2 h^2} = E^2 \frac{\lambda_0^2}{320\pi^2} \tag{7.32-31}$$

Solving (7.32-28) for A_E and substituting the values for P_D from (7.32-29) and P_{rec} from (7.32-31) into it,

$$A_E = \frac{P_{\text{rec}}}{P_D} = \frac{E^2 \lambda_0^2}{320\pi^2} \frac{120\pi}{E^2} = \frac{3}{8} \frac{\lambda_0^2}{\pi} = \frac{\lambda_0^2}{4\pi} \frac{3}{2} \tag{7.32-32}$$

But from (7.32-26) the gain of the short dipole antenna is $G = 1.5 = \frac{3}{2}$, therefore (7.32-32) can be written

$$A_E = \frac{\lambda_0^2}{4\pi} G \tag{7.32-33}$$

Monopole Antenna

The fact that the dipole antenna radiates into the lower hemisphere as well as the upper hemisphere is a disadvantage for terrestrial communications since most receptors are on a line with the horizon or above it. Consequently, the *monopole antenna* with ground (Fig. 7.32-6) has the advantage of an additional 3 dB of gain compared to the dipole because it radiates all of its power into only the upper hemisphere. The monopole antenna is often called a *whip antenna* because it was often used for very high frequency (VHF) police radios and connected to the rear bumper of an automobile. The "whip action" resulted as the car accelerated and decelerated. Due to its symmetry with the half-wavelength dipole, its driving point impedance is one half that of the dipole.

For the quarter-wave monopole antenna

$$R_r = 36.5\ \Omega \tag{7.32-34}$$

$$G_{\max} = 3.28 = 5.16\ \text{dB} \tag{7.32-35}$$

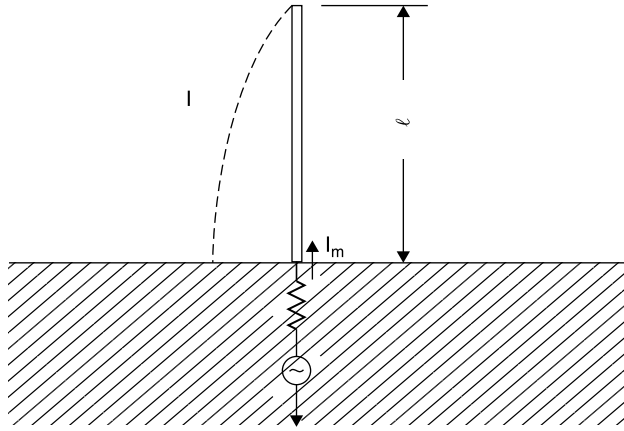

Figure 7.32-6 Monopole antenna with ground plane.

It should be noted that *the presence of a good ground plane is as important to achieving this gain improvement as the antenna*. For example, with a small hand-held monopole, the radiation pattern is likely to be that of the short wire antenna, radiating both above and below the horizon plane that contains the antenna, unless that plane coincides with a good earth ground plane at the frequency of operation. Large AM broadcast antennas are examples of monopole antenna applications. Their monopole function is usually ensured by installing *ground radials* at the base of the antenna, consisting of wires that run radially in all directions from the base of the antenna. Their lengths in the earth are comparable to the antenna's height. The conventional automobile antenna consisting of a straight wire mast also may approximate the monopole, achieving conducting ground plane by means of the sheet metal of the automobile near the aerial's base.

Aperture Antennas

The larger the area of an equiphase front created by an antenna the less the propagating beam, called the *main beam*, radiated by it will diverge, hence the greater its "gain" in its preferred propagation direction, and the more power it can place on a target receptor compared to an isotropic radiator transmitting the same power. By reciprocity, the same comments apply to the receiving antenna.

It was shown that the effective area of an antenna is directly proportional to its maximum gain in what is usually termed the *main beam* or *bore site* direction, the direction orthogonal to the equiphase plane of the antenna. The term *bore site* derives from the practice of installing a small telescope on high gain antennas as an aid to aiming them in the desired direction. For most aperture

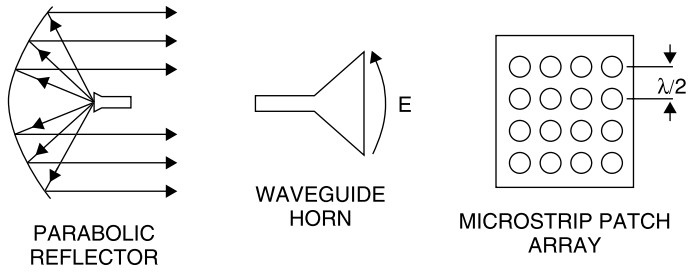

Figure 7.32-7 Aperture antennas.

antennas (Fig. 7.32-7) the area of the aperture would be approximately the effective area if the antenna were *uniformly illuminated*, that is, if it radiated a uniform power density over its entire aperture.

However, uniform illumination is often undesirable because it results in an abrupt discontinuity in the $\vec{E}$ and $\vec{H}$ fields near the edges of the antenna. This in turn results in auxiliary radiation patterns, called *sidelobes* that are separate from and radiate in different directions than that of the main beam. Sidelobes waste power and therefore correspond to signal loss. Often they are even more objectionable because they may intercept spurious signals that interfere with the desired reception. To minimize sidelobe strengths, an *illumination taper* is employed whereby the field strength is maximum near the center of the aperture and diminishes near the periphery.

Since the illumination taper requires that not all of the aperture is energized to the fullest, the result is that the effective area of the antenna is correspondingly less than its physical cross-sectional aperture, A_{CS} (the area projected in the propagation direction). In fact, any reduction in the signal strength over the aperture reduces the antenna's effective area, and with it the main beam gain. Nevertheless, this may be more desirable than encountering interfering signals through sidelobes. Other causes of aperture reduction include dielectric covers to protect the antenna, called *radomes*, surface irregularities such as roughness or actual holes made in the antenna surface to reduce its weight or wind resistance and structural supports of the feed horn.

Equation (7.32-33) can be rewritten as

$$G = \frac{4\pi}{\lambda_0^2} A_E \tag{7.32-36}$$

To account for an effective area A_E less than the actual cross-sectional area A_{CS} of the antenna, a factor to relate them is defined according to

$$A_E = \varepsilon_{AP} A_{CS} \tag{7.32-37}$$

where ε_{AP} is the *aperture efficiency* and $0 \leq \varepsilon_{AP} \leq 1$ [15, Sec. 12.5.3]. As already noted, the aperture efficiency can be less than unity for a variety of reasons.

Phased Arrays

Antennas used for point-to-point communications are fixed in position. These include terrestrial communication link applications and Earth to synchronous-satellite communications. Those used for scanning generally are moved mechanically to cover a given surveillance area, as for radio ranging and detection, or *radar*, applications. However, antennas that can scan electronically without need of mechanical motion have appeal in applications for which mechanical slewing is either too slow or impractical, as in some satellite-borne applications or for tracking rapidly moving radar targets such as missiles. For these applications the electronically scanned *phased array antenna* may be employed (Fig. 7.32-8).

The phased array antenna uses small-aperture radiating elements that, because of their small area, necessarily and desirably have a wide radiation pat-

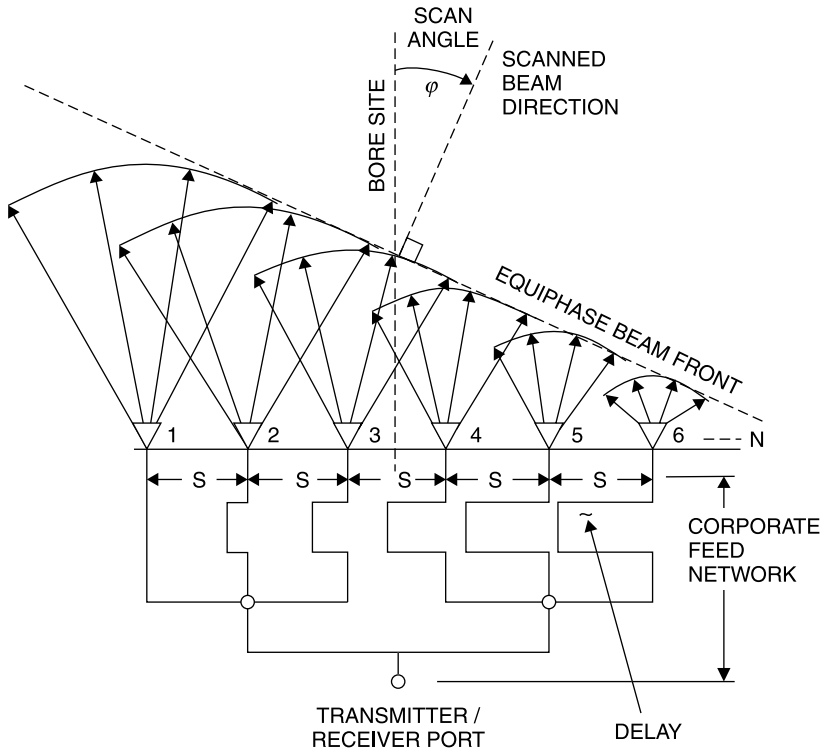

Figure 7.32-8 Schematic diagram of a phased array antenna using elements with time delay to scan the radiated beam.

tern in the bore site direction of the array. A one-dimensional, linear array is shown in Figure 7.32-8 for illustration of the principles, but phased arrays more usually are two dimensional. Steering is accomplished by adjusting the phase of the signals applied to each element so that the aggregate equiphase *wave front* representing the summation of the fields of all of the elements is in a plane at an angle skewed from the bore-site direction. This is called the phased array *scan angle*. In Figure 7.32-8 the steering shown uses *time delay* rather than *phase shift*. When time delay is used, the scan angle is independent of frequency, and the antenna can be used with arbitrarily wide bandwidth, subject to the limits of the radiating elements. The amount of delay, τ_N (in wavelengths) required for the Nth linear element is given by

$$\tau_N = NS \sin \theta \tag{7.32-38}$$

where S is the spacing between elements (in wavelengths) and θ is the scan angle. For two-dimensional scanning of an $M \times N$ array, *in azimuth and elevation*, (7.32-38) must be applied separately for both directions and the results totaled for the delay τ_{MN} to be applied to the MNth element. The element spacing is usually about one-half wavelength since radiating elements necessarily occupy about this minimum dimension on a side. Furthermore, wider spacing would give rise to *grating lobes*, spurious radiation beams for which the wider spacing permits redundant equiphase planes at large scan angles. Note that in Figure 7.32-8 the delay is shown as a fixed line length for purposes of illustration. Array scanning requires that variable delay be selectable. Also note that one wavelength of transmission line produces one period ($\omega t = 360°$, where $t = T$) of time delay.

For large phased array antennas whose apertures are many wavelengths across, the use of time-delay steering is impractical to apply to each element. For example, consider a square antenna 10 wavelengths on a side with half-wavelength spacing between elements in azimuth and elevation directions. This antenna would have 400 elements and from (7.32-33) up to 31 dB of gain in the bore site direction, neglecting losses and sidelobes and assuming an illumination efficiency of 100%. Phased arrays usually are limited to about $\pm 45°$ steering. This limit is imposed by the beamwidth of the subelements as well as the fact that for larger steering angles the array's projected area would be too small reducing its gain too much. But to accomplish *45°* steering for this example array, the time delay required of the edge elements for each scanning direction would be, from (7.32-38), equal to 14.14λ. For two-dimensional scanning this is doubled and therefore a delay of 28.28λ would be required.

Time delay is usually arranged using electronic switches that provide discrete delay steps. Thus a given required delay is managed only approximately, and there is a delay error equal to as much as one half the smallest switched step, or *delay quantization*. To provide for continuous scan angle steering of this size antenna, a delay quantization of 0.125λ might be employed corresponding to a phase error of no more than $\pm 0.0625\lambda$ or $\pm 22.5°$, one half the smallest delay

step. Then time-delay devices capable of delay adjustment from about 0.125λ to 28λ in 0.125λ steps (corresponding to 45° phase steps) would be required. Thus a control device having steps of 0.125λ, 0.25λ, 0.5λ, 1.0λ, 2.0λ, 4.0λ, 8.0λ, and 16λ is needed. This would be termed an *8-bit time-delay device* providing control to 31.875λ in 0.125λ steps.

Such a device is difficult to build. These delay lines, being several wavelengths long at the center frequency, have multiple resonances within the operating bandwidth. This is because diodes are characterized by a high reactance (small capacitance) in their "off state." A long line with small capacitors at each end becomes a high-Q resonator. When the line is long enough, it is certain to have resonant frequencies in the operating band of the antenna, and at these resonances high absorption of the signal power occurs. When the switches do not turn off the long line completely. Even without the resonances, high insertion loss can be expected due to the long line lengths to be switched and the losses of a cascade of switches required to select them.

For these reasons, phase shift rather than time delay is usually applied to each element. The required phase shift is obtained by computing the time delay required at the center frequency of operation and subtracting integer multiples of 2π. For example, the most remote element of the 20λ by 20λ array for a 45° scan in *both azimuth and elevation* would require 28.28λ of delay. This could simulated using a phase shift corresponding to 0.28λ or 100°. In turn, the phase shifters also would be realized in binary form. A 4-bit phase shifter would provide 0° to 337.5°. In the phase domain 360° is equivalent to 0° of control in 22.5° steps. The nearest step to 100° is 90°, and this would be selected. The fact that phase shift control is only correct at the frequency for which the time-delay-to-phase-shift conversion is made means that the radiated antenna beam will *squint* off the intended scan angle as the frequency of operation departs from this center frequency.

Unlike mechanically scanned antennas, phased array antennas have reduced effective area, consequently less gain, as they are steered off the bore site direction because their projected area in the scanned direction is correspondingly less.

Because of these considerations and other complexities not described, large-aperture phased array antennas are employed only when their motionless, high-speed scanning properties are deemed essential. However, small apertures of a few wavelengths in diameter may prove practical in some communication applications.

7.33 PATH LOSS

Although Maxwell's equations are fundamental to radio engineering, their application requires an analytic description of the $\vec{E}$ and $\vec{H}$ fields, and this can be quite complex as was seen for the short wire antenna analysis. However, some propagation problems can be solved with a much simpler approach once the

principles of wave propagation are appreciated. One of these is the *path loss* calculation, which is the determination of how much power reaches a radio receiver from a distant transmitter. In fact, the path loss calculation is usually one of the first in the design of a radio communication system.

Previously we introduced the isotropic radiator assumption. The isotropic radiator is defined to be a 0 dB gain antenna, and the power radiated from it is called the *effective isotropic radiated power (EIRP)*. As was noted earlier, it is impossible to build an isotropic radiator, but the concept is useful for calculating how much power would be available at a distance were one to have been employed. Then it is a simple matter to adjust the isotropic result so obtained to that which applies given the relative gains of the actual antennas used for transmission and reception. *The isotropic antenna uniformly "paints" the surface of a sphere having a radius equal to the distance R between transmitter and receiver.*

Figure 7.33-1 depicts a transmitter at the center of a sphere of radius R. If the transmitter power P_T is radiated by an isotropic antenna, then, since the surface area of a sphere is $4\pi R^2$, the power density at R is $P_T/4\pi R^2$. If the isotropic antenna is replaced by a transmitting antenna of gain G_T, then the power density in the direction of its maximum intensity (the direction of its main beam) is given by

$$P_D = G_T \frac{P_T}{4\pi r^2} \tag{7.33-1}$$

If a receiving antenna at distance R is oriented to accept maximum power from

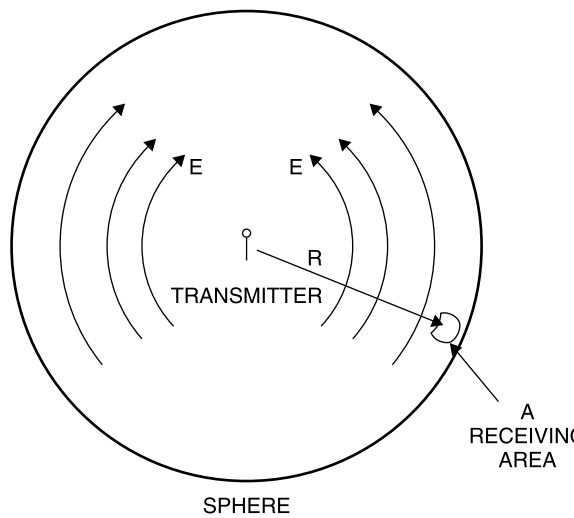

Figure 7.33-1 Effect of power reduction as a result of spreading out isotropically (painting an imaginary sphere) at a distance R from the transmitter.

this direction (it is aimed at the transmitter), the power it will receive is

$$P_R = A_E G_T \frac{P_T}{4\pi R^2} \tag{7.33-2}$$

where A_E is the effective capture area of the receiving antenna. Note that P_R is actually the available power at the terminals of the receiving antenna. Applying (7.32-33) this can be written

$$P_R = \left(\frac{\lambda_0}{4\pi R}\right)^2 G_T G_R P_T \tag{7.33-3}$$

Usually the radio system engineer is interested in determining the loss between the terminals of the transmitting antenna and those of the receiving antenna. This is called the *path loss*. For this purpose (7.33-3) can be cast in an insertion loss format as

$$\text{IL}_{\text{PATH}} = \frac{P_T}{P_R} = \left(\frac{4\pi R}{\lambda_0}\right)^2 \frac{1}{G_T G_R} = \frac{16\pi^2}{G_T G_R}\left(\frac{R}{\lambda_0}\right)^2 \tag{7.33-4}$$

Note that numeric values (not dB) must be used for G_T and G_R and that the same length units must be used for A_E, R, and λ_0 in (7.33-1) to (7.33-4).

As an example, consider the case of a 1-W, 900-MHz cellular telephone transmitter radiating its signal with a circularly polarized antenna having a gain factor of 1.64 above isotropic. The receiving monopole antenna on the exterior of an automobile is vertically polarized and has a gain factor of 3.28. The automobile is 3 miles from the transmitter, and there is a line-of-sight transmission path. What is the path loss between transmitter and receiver?

First note that since the receiving antenna is linearly polarized it can capture only half of the circularly polarized power incident upon it. To take this into account, the gain of the receiving antenna is reduced by one half to 1.64.

Substituting the system values into (7.33-4) yields

$$\text{IL}_{\text{PATH}} = \frac{16\pi^2}{G_T G_R}\left(\frac{R}{\lambda_0}\right)^2 = \frac{16\pi^2}{(1.64)(1.64)}\left(\frac{3 \times 1610}{0.33}\right)^2 = 1.23 \times 10^{10} = 101 \text{ dB}$$

since 1 mile = 1610 m and $\lambda_0 = 0.33$ m at 900 MHz. Note that since $P_T = 1$ W = 1000 mW = 30 dBm, $P_R = 30$ dBm − 101 dB = −71 dBm.

The path loss calculation is only a first step in the design of a radio communication link. Implicit in the path loss calculation is the assumption that the transmitter and receiver are on a line-of-sight with each other, that is, that there are no intervening obstacles. For real radio systems (Fig. 7.33-2), this is rarely the case. Buildings, foliage, hills, and other obstacles occupy the propagation path.

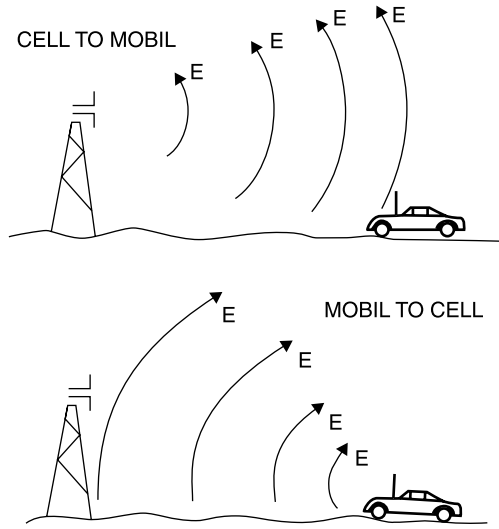

Figure 7.33-2 Cellular radio system depicting cell-to-mobile and mobile-to-cell transmissions. For illustration, a dipole antenna is depicted at the base station, but customarily a circularly polarized antenna is employed.

For a cellular telephone communication system there is almost never an unobstructed path. The transmitted signal may undergo several reflections on its roundabout path to the receiver, and the strongest signal reaching the receiver may actually be coming from a direction opposite that of the transmitting antenna. The resultant signal may arrive at the receiver as the phasor addition of several separate signals with random phases, resulting from multiple paths, a process called *multipathing*.

When this occurs, the signals combine with random phases at the receiver. Often, multiple signals result in a combined signal lower than any of its constituents, a process called *Rayleigh fading*. Rayleigh was a physicist who, among other accomplishments, formulated the distribution associated with random events. To overcome deep fades, a cellular base station may switch between different antennas (*diversity switching*) to create a more favorable addition of the multipath signals at the receiver. This can occur hundreds or thousands of times during a cellular conversation conducted between the cell site and a moving automobile containing the cellular system client.

In specifying a communication system, once the *minimum detectable signal level* is determined (more about this in Chapter 10), the minimum required transmitter power can be determined. However, a transmitted power well beyond that required by the path loss calculation is generally specified in order to create a *system margin* and a more robust communication link. *The system margin can overcome the obstacles to delivering a detectable signal much of the time, but never always*. The greater the system margin, the less the probability

that the communication link will prove to be unacceptable, and the greater the percentage of time that the link will operate satisfactorily. However, no matter how great the allowance, there will always be some probability, however small, that a sufficiently deep fade will be encountered to render the link unusable for some period of time or at all times in some locations. Cellular telephone users operating from moving vehicles experience *drop-outs* regularly and can attest to this fact. *Wireless communication links are never completely reliable.*

7.34 ELECTROMAGNETIC (EM) SIMULATION

When AC analysis is used to evaluate a component, we assume that it behaves according to its equivalent circuit model. However, this is an idealization. Real circuits have physical sizes that often are appreciable compared to the operating wavelength, causing a change of voltage and current over their dimensions. This is not included in the equivalent circuit used by a circuit simulator or in the manual application of AC analysis. The resulting circuit performance estimates are based on this idealization; consequently, they will contain some errors as a result. Sometimes the error is quite significant, as the following example demonstrates.

Suppose that we wish to employ a 25 Ω characteristic impedance, open-circuited shunt stub on a 50-Ω transmission line as a filter. The filter is to block the second harmonic of a 1920-MHz cellular system transmitter, so that the spurious requirements of its radio license will not be violated. Therefore, it is desired that the open-circuited stub presents a short circuit (or very low impedance) to the transmission line at 3840 MHz. The stub's length is to be chosen equal to 90° at 3840 MHz, resulting in a short circuit across the 50-Ω line and blocking the transmission of spurious power at 3840 MHz.

Using an ideal equivalent circuit, the open-circuited stub is modeled as shown in Figure 7.34-1, in which the 57.5° line lengths on either side of the stub correspond to 50-Ω line sections used later in the EM simulation modeling. These lengths have no effect on the stub's insertion loss performance. The calculated performance using a circuit simulator is shown in Figure 7.34-2, from

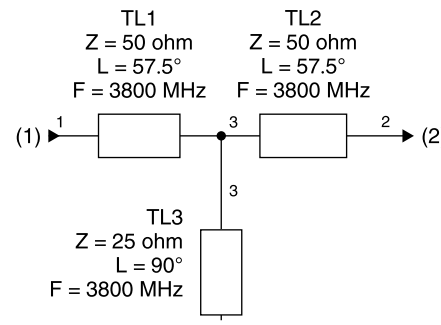

Figure 7.34-1 Ideal electrical schematic of an open-circuited, 25 Ω, shunt stub on a 50-Ω transmission line.

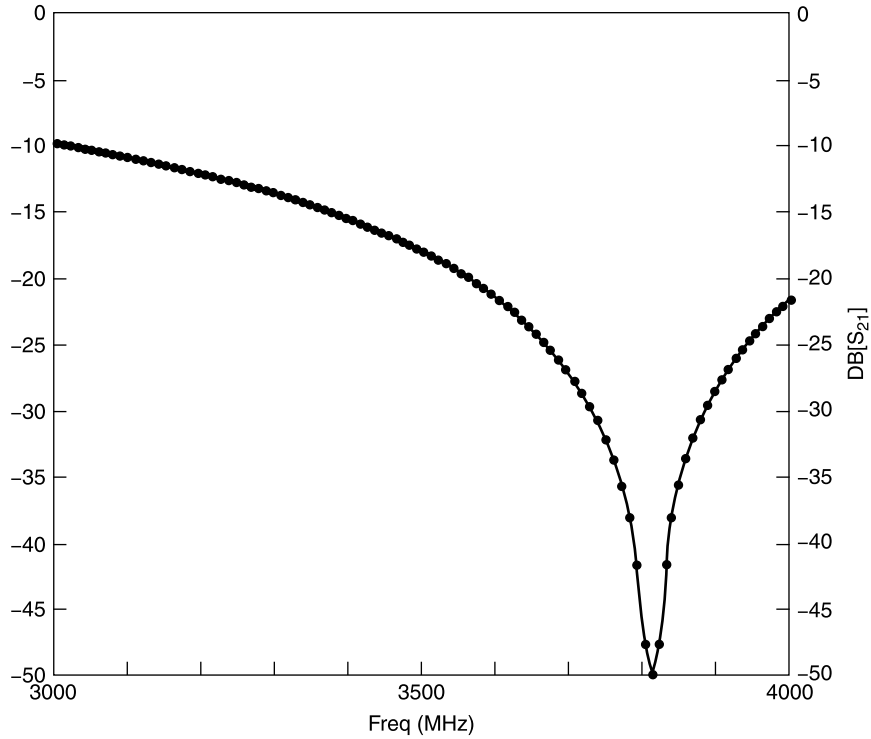

Figure 7.34-2 Calculated response of an open-circuited 25-Ω stub obtained using ideal electrical schematic.

which it is seen that 40 dB or more of isolation can be obtained over a 50-MHz bandwidth.

To build this circuit it is necessary to create a circuit pattern. The determination of the geometry is accomplished by determining the width of a 50-Ω transmission line and a 25-Ω stub. We also require determination of the wavelength on the stub at the operating frequency in the medium in which the stubs will be realized. Using the Genesys auxiliary program TLINE and selecting a design medium consisting of a microstrip line having a 0.031-thick substrate with dielectric constant 2.22, the wavelength at 3840 MHz (at which we want the stub to block transmission) is 2.165 in. (5.50 cm) and the width of a 25-Ω stub is 240 mils (0.240 in. or 61 mm) (Fig. 7.34-3).

The width of a 50-Ω line in this medium is 93 mils (0.093 in. or 2.36 mm). The wavelength at 3840 MHz is 2.5506 inches on the 50-Ω line, considerably different that the wavelength at the same frequency on the 25-Ω stub. This is a consequence of the nonhomogeneous nature of microstrip line. Were stripline used, the wavelength would be the same for both impedances.

When we attempt to draw the center conductor pattern layout for this circuit, some questions arise, such as: Where does the stub begin? Does it begin at

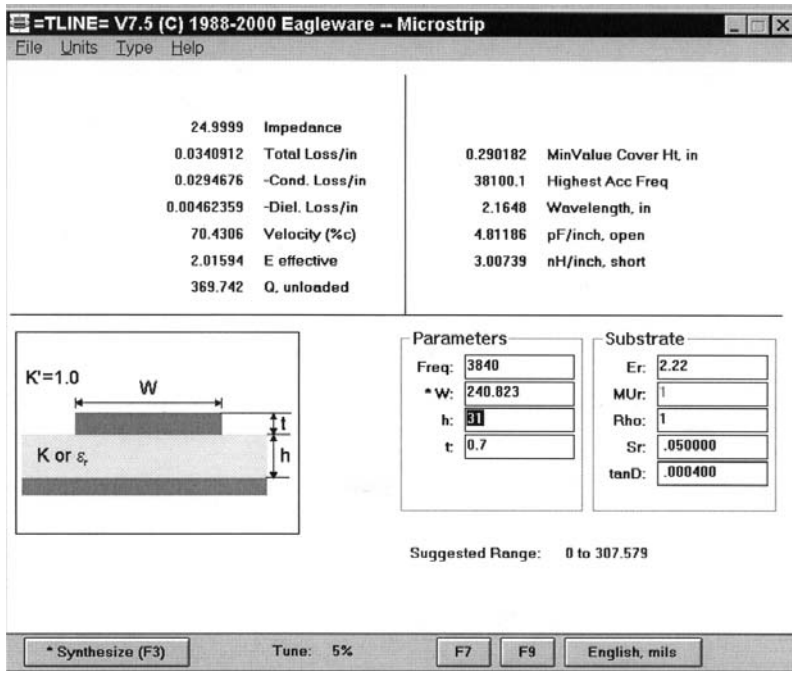

Figure 7.34-3 Calculated dimensions and wavelength for a 25-Ω microstrip line using 0.031-in.-thick substrate with relative dielectric of 2.22.

the edge of the 50-Ω main transmission line, or should one measure from the center line of the 50-Ω line? Should an allowance be made for the fringing electric field at the end of the stub, the effect of which would make the electrical length of the stub appear longer?

To answer these questions experimentally, as was the practice prior to the availability of EM simulation, it was customary to build and test the stub, usually with various geometries, and then to deduce design rules that could be used to approximate the behavior of future designs. The availability of *electromagnetic simulators (EM simulators)* eliminates the need for time-consuming model building by permitting the experimental "construction and test of the circuit" on the computer.

The detail in Figure 7.34-4 represents dimensioning using the edge of the transmission line as the beginning of the stub and no length allowance for any fringing electric field at the open-circuited end of the stub. The 25-Ω stub impedance requires a 0.240-in.-wide line. At 3840 MHz the wavelength of 2.165 in. results in 0.541 in. for a quarter wavelength stub. Notice that the width of the stub is nearly half its length! The horizontal conductor at the top is a 50-Ω line and has a 0.093-in. width.

The EM *simulation grid* used in Figure 7.34-4 is 20 mils square. It is over this granularity that the $\vec{E}$ and $\vec{H}$ fields of adjacent cells, measuring $0.020 \times$

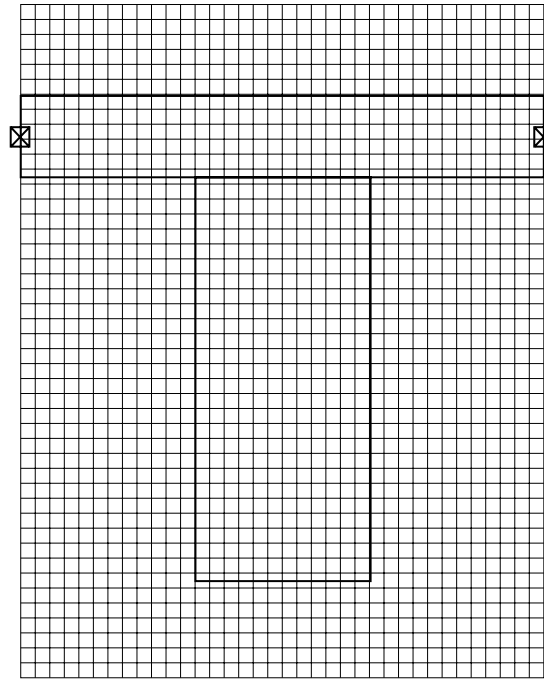

Figure 7.34-4 Layout of the 25-Ω, open-circuited shunt stub used for EM simulation. A theoretical perfectly conducting "box" containing the circuit measures 0.720 in. wide by 0.900 in. high is assumed in the simulation. The analysis also assumes a perfectly conducting box cover separated by 0.150 in. of air above the circuit. The *simulation grid* shown is 20-by-20 mils.

0.020×0.020 in. are forced to be continuous by the simulator. Basically, the simulation software begins by assuming starting values for the $\vec{E}$ and $\vec{H}$ fields, matching the boundary conditions that require that the tangential $\vec{E}$ field and the normal $\vec{H}$ field on conductors must be zero. Then the program determines the $\vec{E}$ and $\vec{H}$ fields within each 20 mil^3 cell and checks the continuity of the fields on the cell boundaries from cell to cell.

Usually the initially assumed values for the fields are in error, and the computer must guess a modified set of field values in all of the cells and retry the solution. This process is repeated until the combined errors are below a tolerance set by the software, and satisfactory agreement is obtained for the field continuity at all cell-shared boundaries. Using a cell size of $(0.020 \text{ in.})^3$ in a box size of $0.720 \times 0.900 \times 0.180$ in.^{3} results in $36 \times 45 \times 9$ grids or 14,580 cells for each evaluation frequency. For each calculation vector field calculations must be performed and fields on the boundaries of all adjacent cells matched.

This is a formidable task, even for a computer capable of millions of mathematical operations per second. A typical execution analysis for this model can

take several minutes to converge to a satisfactory field distribution throughout the box containing the circuit, and this process must be redone at each evaluation frequency, in the case of this example for 101 frequencies between 3000 and 4000 MHz.

The final result, once the fields sufficiently match on the cell boundaries, permits an evaluation of the fields at the input and output ports of the network. These fields are used to calculate the voltages and currents at the ports when matched source and loads are connected to them. From these voltage and current values, the scattering parameters of the network are calculated for each frequency of evaluation. Given the scattering matrix, the performance of the stub when embedded in a transmission line network is easily calculated using the network simulator portion of the software.

The EM simulation results are shown in Figure 7.34-5 and compared with the initial circuit simulation response. As can be seen, there is a significant difference in the two simulations. Since this stub was to form part of a filter in a 1920-MHz cellular system to remove the second harmonic at 3820 MHz, the discrepancy shown in Figure 7.34-5 suggests that a design based upon only a

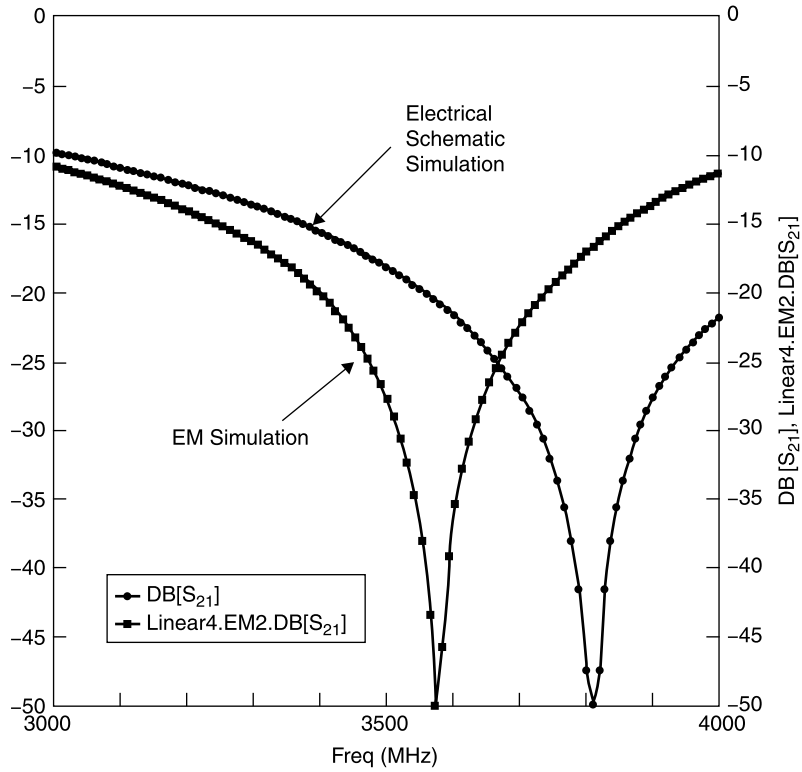

Figure 7.34-5 Comparison of EM and electrical schematic performance simulations.

network simulation would likely prove inadequate, necessitating additional experimental trials. However, with the EM simulation a design likely to be usable after the first design iteration would much more likely be obtained.

In this case, the error in using only an electrical schematic to model the filter would result in a stub resonance error of about 240 MHz at 3800 MHz, a 6% error. While this is a small percentage, it results in a difference in the null depth of 30 dB at the stub's stopband center, demonstrating the usefulness of electromagnetic simulation.

REFERENCES

1. Peter A. Rizzi, *Microwave Engineering, Passive Circuits*, Prentice-Hall, Englewood Cliffs, NJ, 1988. *Excellent microwave engineering textbook covering theory and design of transmission lines, couplers, filters, and numerous other passive devices.*
2. Simon Ramo, John R. Whinnery, and Theodore Van Duzer, *Fields and Waves in Communication Electronics*, 3rd ed., Wiley, New York, 1994. *This is a revision of [6].*
3. Paul J. Nahian, *Oliver Heaviside: Sage in Solitude*, IEEE Press, New York, 1988.
4. Hugh Hildreth Skilling, *Fundamentals of Electric Waves*, Wiley, New York, 1942. *This is a superbly written text on field theory, with many well-illustrated drawings of fields and vector functions.*
5. Sears and Zemansky, *University Physics, Vols. 1 and 2*, Addison-Wesley, Reading, MA, 1956.
6. Simon Ramo and John R. Whinnery, *Fields and Waves in Modern Radio*, Wiley, New York, 1944, 1953 (later revised with a third author, Van Duzer, in 1965, 2nd ed., 1984, and 3rd ed., 1994). *This is a classic introductory text describing fields and Maxwell's equations.*
7. George B. Collins, *Microwave Magnetrons, Radiation Laboratory Series*, McGraw-Hill, New York, 1948.
8. Jerome L. Altman, *Microwave Circuits*, D. van Nostrand Company, Inc., Princeton, NJ, 1964.
9. Robert E. Collin, *Field Theory of Guided Waves*, McGraw-Hill, New York, 1960. *Collin is a master of field theory as is demonstrated by this summary work.*
10. Robert E. Collin, *Foundations of Microwave Engineering*, McGraw-Hill, New York, 1966.
11. S. A. Schelkunoff, "The electromagnetic theory of coaxial transmission lines and cylindrical shields," *Bell System Technical Journal*, Vol. 13, October, 1934, pp. 532–579.
12. Joseph F. White, "Simplified theory for post coupling Gunn diodes to waveguide," *IEEE MTT Trans.*, Vol. MTT-20, No. 6, June, 1972, pp. 372–378.
13. N. Marcuvitz, *Waveguide Handbook*, Vol. 10 of the Radiation Laboratory Series, McGraw-Hill, New York, 1951.

14. Edward C. Jordan and Keith G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1968, Section 11.12.
15. Constantine A. Balanis, *Antenna Theory—Analysis and Design*, 2nd ed., Wiley, New York, 1997.

EXERCISES

E7.3-1 **a.** Find the electric field due to an isolated charge of 1 electron at a distance of 1 m in free space.

b. How many electrons would be required to provide an electric field of 1 V/m at this same distance?

E7.3-2 Assume that there are exactly 10^{23} atoms/cm^3 of solid copper metal. (a) What is the size of the side of a copper cube large enough to provide the charge in Exercise 7.3-1b if every copper atom is ionized by gaining one electron? (b) What is the electric field if each copper atom gives up one electron?

E7.3-3 An electric dipole consists of two point charges, $+q$ and $-q$, spaced a distance 2d apart along the x axis. The midpoint between them is at $x = 0$. Show that for $x \gg d$, the E field along the x axis is inversely proportional to x^3.

E7.4-1 A television set with a screen height of 0.5 m has a 525-line picture. It is turned on and has a beam accelerating voltage of V = 20,000 V. Once the electron beam passes the accelerating electric field, its velocity, v (assume v is in the $-$z direction), is constant and given by $v = \sqrt{2eV/m}$ where m and e are, respectively, the mass and charge of the electron. When the beam is undeflected, the spot appears at the center of the screen. The distance from the electron gun to the screen is 0.7 m. A long wire carrying 10 A (assume this is in the +z direction) of peak AC current is located under the TV. It is 0.5 m below the beam and parallel to it.

a. In which direction will the beam be deflected when the current is traveling in the +z direction?

b. What will be the peak-to-peak deflection of the beam due to the AC current?

c. The viewable height of the screen is 0.5 m. The picture consists of 525 lines. Will the deflection caused by the 10-A AC current be noticeable?

d. If the deflection is noticeable, comment on how frequently this occurrence might be expected.